Studies in Distributed Intelligence

Mohamed Elhoseny
Xiaohui Yuan
Salah-ddine Krit *Editors*

Distributed Sensing and Intelligent Systems

Proceedings of ICDSIS 2020

Springer

Studies in Distributed Intelligence

The ubiquitous sensing, multi-modality data, and various computing platforms enable rapid, personalized, dedicated services for our communities. The global perspectives on the distributed intelligence showcase cutting-edge research that embraces a number of fields from system design to deployment and from data acquisition to analysis in cross-disciplinary areas that require discovery of valuable information from dynamic, massive, diversified data. The research to be included in this series focuses on the pressing, contemporary aspects of the decentralized, computational intelligence techniques for geotechnology applications including environment, resource management, transportation, health, robotics and autonomous vehicles, and security, to name a few.

The series "Studies in Distributed Intelligence" publishes new developments and advances in the areas of distributed intelligence and covers the theories, methods, and applications of distributed intelligence, as embedded in the fields of engineering and computer science as well as the cross-disciplinary fields. The series contains monographs, lecture notes, and edited volumes in distributed intelligence spanning the areas of environmental sensors, big geoscience data analysis, smart sensing networks, distributed systems, edge computing, supercomputing for climate problems, artificial intelligence in environmental monitoring, urban intelligence, self-organizing systems, intelligent transportation systems, soft computing, smart vehicular communication, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the worldwide distribution, which enable both wide and rapid dissemination of the research outputs.

If you are interested in contributing to the series, please contact the Publisher: Aaron Schiller [Aaron.Schiller@springer.com].

More information about this series at https://link.springer.com/bookseries/16393

Mohamed Elhoseny • Xiaohui Yuan •
Salah-ddine Krit
Editors

Distributed Sensing and Intelligent Systems

Proceedings of ICDSIS 2020

Springer

Editors
Mohamed Elhoseny
Mansoura University
Mansoura, Egypt

Xiaohui Yuan
Computer Science and Engineering
University of North Texas
Denton, TX, USA

Salah-ddine Krit
Faculty of Sciences
Ibn Zohr University
Agadir, Morocco

ISSN 2662-3706 ISSN 2662-3714 (electronic)
Studies in Distributed Intelligence
ISBN 978-3-030-64257-0 ISBN 978-3-030-64258-7 (eBook)
https://doi.org/10.1007/978-3-030-64258-7

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Elliptic Curve Cryptography (ECC) for a Lightweight Public Key Infrastructure (PKI)

Jalal Laassiri, Yousra Berguig, Sanae Hanaoui, and Sara Hassi

Abstract In recent years, the information access, use, and control have become progressively an interesting area of research. The increase in the number of cyber-attacks makes network and information security situation bleak. In this chapter, we have presented an overview of cryptography that complements strongly security, we describe and compare symmetric encryption systems and asymmetric encryption systems. Therefore, we give a brief introduction to the elliptic curve cryptography (ECC), performance, characteristics, advantages, and challenges. Also, we investigate the PKI (Public Key Infrastructure) technology for certificate revocation. Then we propose a new approach for a lightweight PKI based in elliptic curve cryptography (ECC); furthermore, we present a public key infrastructure model for certificate revocation, as well we elaborate the deployment of the presented Model.

Keywords Public Key Infrastructure (PKI) · Cryptography · Digital signatures · Elliptic curve cryptography (ECC) · Security · Block ciphers

1 Introduction

Recently a set of new security concerns and issues have appeared due to the fast growth of electronic activities and information growth, hence the need for security and protection increase. Cryptography remains one of the best solutions to gain security and protection in a distributed system. In cryptography, we have symmetric encryption which is the oldest and well-known technique. It uses a secret key that can be a number, word, or random letters. All the parties, the sender, and the receiver need to have the key in their possession. However, there is a problem with the concept of a secret key. Having access to this secret key allows decrypting the encrypted message. Thus the solution to this issue is to use a key-pair, the use of

J. Laassiri (✉) · Y. Berguig · S. Hanaoui · S. Hassi
Laboratory of Computer Science, Systems and Optimization (ISO Lab), Department of Computing, Ibn Tofail University, Kénitra, Morocco
e-mail: LAASSIRI@uit.ac.ma; sanae.hanaoui@uit.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_1

a public key, that anyone can hold, and a private key. A PKI system allows both encryption and digital signatures by a wide breadth and diverse set of applications. Public key cryptography has proved to be a practical safeguard against identity theft perpetrated through the Web, email, and Virtual Private Networks. Nonetheless, the main challenge is the key size, in order to guarantee a high level of security, it is necessary to have a larger key, which requires a great deal of execution time. For this, we propose a new approach for a lightweight PKI based on elliptic curve cryptography (ECC), which is an asymmetric algorithm that has a key to encrypt and other to decrypt. ECC can offer a high level of encryption power for much shorter keys, providing better security while reducing computing requirements, which makes it an interesting option for devices with limited storage and processing power.

2 Asymmetric Encryption System

2.1 Definition of Cryptography

Cryptography is a conversion from plain message to the ciphertext in order to protect the information and make it secure from attacks and intruders. Achieve a strong understanding of cryptography requires knowledge of these four main areas of primitives, random number generation, symmetric encryption, asymmetric encryption, and hash functions. All those primitives are very complementary [1]. Cryptographic systems attempt to guarantee three main objectives. Frist confidentiality, if a user can decrypt the message without having a key, then this objective is not achieved. The second objective is integrity, to assure the data has not been modified. Finally, authentication, which is assuring that the source of the data is verified [1].

2.2 Symmetric and Asymmetric Encryption System

Asymmetric encryption uses private and public keys to encrypt and decrypt data as illustrated in Fig. 1. Symmetric encryption or single key encrypt and decrypt information as shown in Fig. 2. The benefit of using single-key encryption is speed. Accordingly, symmetric key encryption used with asymmetric encryption also makes a quick transaction [1]. Symmetric encryption can be used to send confidential information and guarantee integrity. Compared to symmetric encryption, asymmetric encryption requires two keys, a private key and a public key. The public key is available to anyone who wants to use it. Asymmetric encryption is slower than symmetric encryption because it uses number theory to increase its strength. Regardless of its slow speed, it does an excellent job of the difficulty of sharing keys. Also, asymmetric encryption is more secure and more complex than symmetric

Fig. 1 Asymmetric encryption

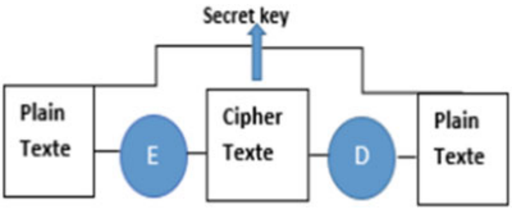


Fig. 2 Symmetric encryption

encryption [2]. Digital signatures and hashes both use asymmetric encryption. A digital signature message uses authentication to prove a certain person sent the message and used to guarantee important security goals [3]. Hash functions are used to gain enhanced performance when large blocks of data are using asymmetric encryption. Hashes have three properties that make them valuable. It is doubtful that the same messages can hash identically. With a known digest, it would be nearly impossible for a second message to create the same digest finally, it would also be almost impossible to locate the original message that created the known Digest. Hashing is also valuable in storing passwords and hashing ensuring integrity. A message remains the same regardless of the number of times it is computed [1].

The main asymmetric encryption algorithms used for digitally signing data and encrypting are: RSA, ECC, and ELGAMAL.

3 Elliptic Curve Cryptography

3.1 ECC Definition

The ECC is an asymmetric algorithm it is an alternative of RSA which is the most common used for SSL certificates. These two types of master keys share the same important property of having a key to encrypt and other to decrypt. However, ECC can offer the same level of encryption power for much shorter keys, providing better security while reducing computing requirements. The shorter keys make ECC a very interesting and attractive option for devices with limited storage processing power (Table 1).

Table 1 Comparison of an elliptic curve cryptography key and RSA key

Symmetric	RSA key length (bit)	ECC key length (bit)
80	1024	160
112	2048	224
128	3072	256
192	7680	384
256	15,360	521

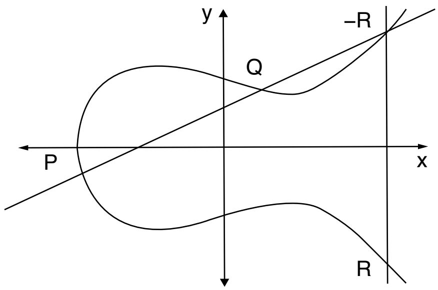


Fig. 3 Point addition

– An elliptic curve E is curve given by a Weierstrass equation:

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \tag{1}$$

3.2 *Proposition*

Let L be the vertical line passing through R. We define $P + Q \in E\ (K)$ as the second point of intersection of L with E. With the law of composition, $(E(K), +)$ is an abelian group whose neutral element is the point to infinity (O) (Fig. 3).

- *Point addition*: With two distinct points, P and Q, the addition is defined as the negation of the point resulting from the intersection of the curve, E, and the line defined by the points P and Q, giving the point, R.

$$\begin{aligned} &P + Q = R \rightarrow \left(x_p, y_p\right) + \left(x_q, y_p\right) = (x_r, y_r) \\ &x_r = \lambda^2 - \left(x_q + y_q\right) \\ &y_r = \lambda \times \left(x_p - x_r\right) - y_p \text{ with } \lambda = \frac{(y_p - y_q)}{x_p - x_q} \end{aligned}$$

- *Point doubling*: When the points P and Q are coincident, the addition is similar, except that there is no straight line defined by P and Q, so the operation is closed using the limit case, the tangent to the curve E, to P and Q. This is calculated as above but with a (Fig. 4):

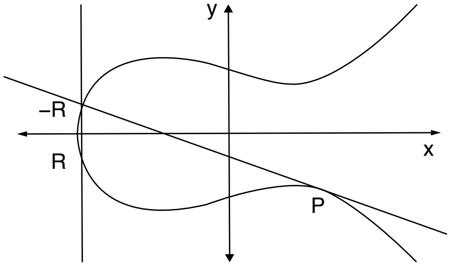


Fig. 4 Point doubling

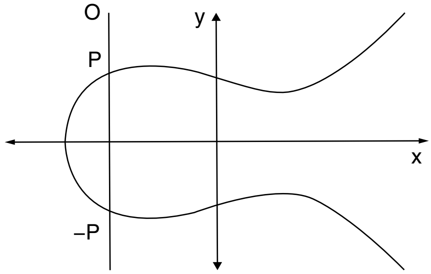


Fig. 5 Vertical point

$$\lambda = \frac{\left(3x_p^2 + a\right)}{2y_p}$$

- *Vertical point*: The straight line joining any point P and its symmetrical relative to the horizontal axis, noted $-P$, is a vertical line, the third point of intersection with the curve is the point at infinity (which is its own symmetrical with respect to the abscissa axis), hence $P + (-P) = 0$ (Fig. 5).

Double-and-Add The simplest method is the double-and-add method, similar to multiply-and-square in modular exponentiation. The algorithm works as follows: To compute DP, start with the binary representation for:

$$d = d_0 + 2d_1 + 2^2 d_2 + \cdots + 2^m d_m \text{ with } [d_0 \ldots d_m] \text{ in } [0, 1]\,.$$

– Analysis of the encryption and decryption process: To compare the performance of the encryption and decryption process, the ECC-160, ECC-192, ECC-224, RSA-512, RSA-1024, RSA-2048, Elgmal-1024, Elgamal1536, and Elgamal-2048 public key cryptosystems are used. The comparison table for performance analysis is given in Table 2.

Table 2 Performance analysis of encryption and decryption process

	RSA key length (s)	ECC key length (bit)
ECC-160	4	1:6.4
Elgamal-1024	7	
RSA-1024	8	
ECC-192	5	1:8
Elgamal-1536	9	
RSA-1536	12	
ECC-224	6	1:9.14
Elgamal-2048	11	
RSA-2048	22	

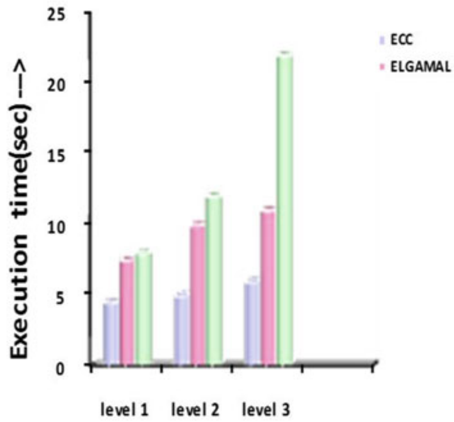


Fig. 6 Execution time comparison

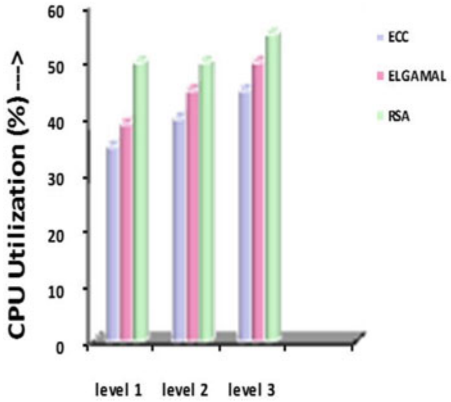
– Comparison of execution time between ECC and other asymmetric algorithms.

From Table 2 and the above graph in Fig. 6, it is clear that ECC is much faster than RSA and Elgamal cryptosystem, and also the key size ratio of the ECC is much less than the RSA and Elgamal cryptosystem, and at the same time, ECC provides the same level of security as RSA and Elgamal cryptosystems. Hence, the performance of the ECC is much better in terms of execution time than these two public key cryptosystems [4].

– Comparison of CPU utilization between ECC and other asymmetric algorithms:

Based on the results shown by the graph presented in Fig. 7, we notice that CPU utilization of ECC never exceeds 50% utilization of CPU, the other two public key algorithms RSA and Elgamal exceed the 50% utilization of the CPU. Thus, ECC also uses low resources. Hence, the performance of the ECC is much better than these two public key cryptosystems [4].

Fig. 7 CPU utilization



4 Public Key Infrastructure for Certificate Revocation Model

4.1 PKI Workflow and Lifecycle

PKI is an all-inclusive system that affords the use of digital signatures and public key encryption; it is a system that manages certificates and keys. The use of PKI allows the creation and management of a loyal and trustworthy networking environment [5]. PKI is totally synonymous with asymmetric encryption because it is more secure than symmetric encryption, and use a public and private key, one for encryption and the other for decryption. Everyone knows the public key, and only the owner knows the private key. The issue appears if a public key is used by a party that is not the actual owner of the public key. To avoid this problem, digital certificates verify that the owner is making use of a CA (Certificate Authority) [6] (Fig. 8).

To create a certificate, a PKI-enabled application is required. The CR (Certificate Request) contains a public key and information used to submit the type of certificate requested [1]. The organization needs to decide the types of certificates required, and this would determine whether the implementation of a public or private CA. If an organization chooses a private CA, then a choice has to be made of roles that are a Root CA, Intermediate CA, or an Issuing CA. It is prudent to design a system with multiple cases to assure available and secure distributing points [1].

4.2 A Detailed Process for Certification Revocation

Revocation is another critical component of PKI. Essentially, if the private key corresponding to a certificate is compromised or stolen, then the certificate might be revoked before its expiry. If not revoked properly, intermediate certificates with compromised/stolen private keys can result in session hijacking and unauthorized issuance of new certificates, respectively. Generally, there are two used methods for

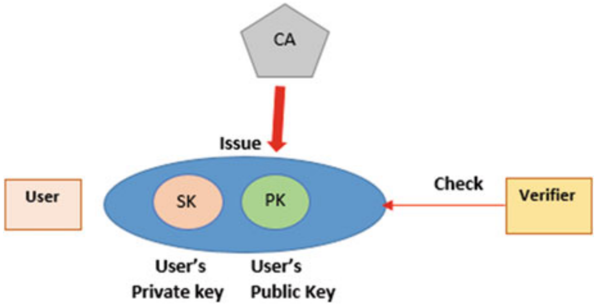


Fig. 8 Public key infrastructure

propagating information about certificate revocation: CRL and OCSP. Correctly, as part of certificate validation, the certificate revocation status must be checked [7].

As we can see in Fig. 9, a CA issued a certificate to an Application Server (Step I). If requested, the CA can revoke an issued certificate and publishes a CRL (Step II), which is managed by the Revocation Manager. In our model, we assume that the revocation manager manages both a CRL server and an OCSP server [8]. A Client (a web browser) sends a Client Hello message to request a new SSL/TLS connection. The application server responds with a Server Hello message followed by a server Certificate. After receiving a certificate, the client validates the certificate. However, some clients miss the certificate revocation, to solve this problem, we propose CRG that aims at addressing problems associated with CRL and OCSP methods. As we can see, both certificate issuance (Step I) and CRL publication (Step II) stay the same. Besides, there is no change in the SSL/TLS flow.

That is, we do not require any modification in the SSL/TLS; however, CRG intercepts the Certificate message of the SSL/TLS protocol (Step 3) and consults the Certificate DB (Step 3a) before forwarding the Certificate Revocation Request to the Certificate Validator (Step 3b).

The Certificate Validator requests the certificate status (Step 4), gets the status (Step 5), and sends the response to the Certificate Checker (Step 5a). The Certificate Checker updates the Certificate DB and forwards the decision together with the certificate (Step 5c) to the Certificate Modifier, which can send the original or faked certificate to the client (Step 6). If the certificate has been revoked, then CRG has to communicate this information to the client. It is important to mention that we do not modify the underlying protocol between the client and the application server. Instead, we generate a fake certificate to indicate that the certificate is not valid anymore. In other words, the fake certificate is such that its validation on the client will fail.

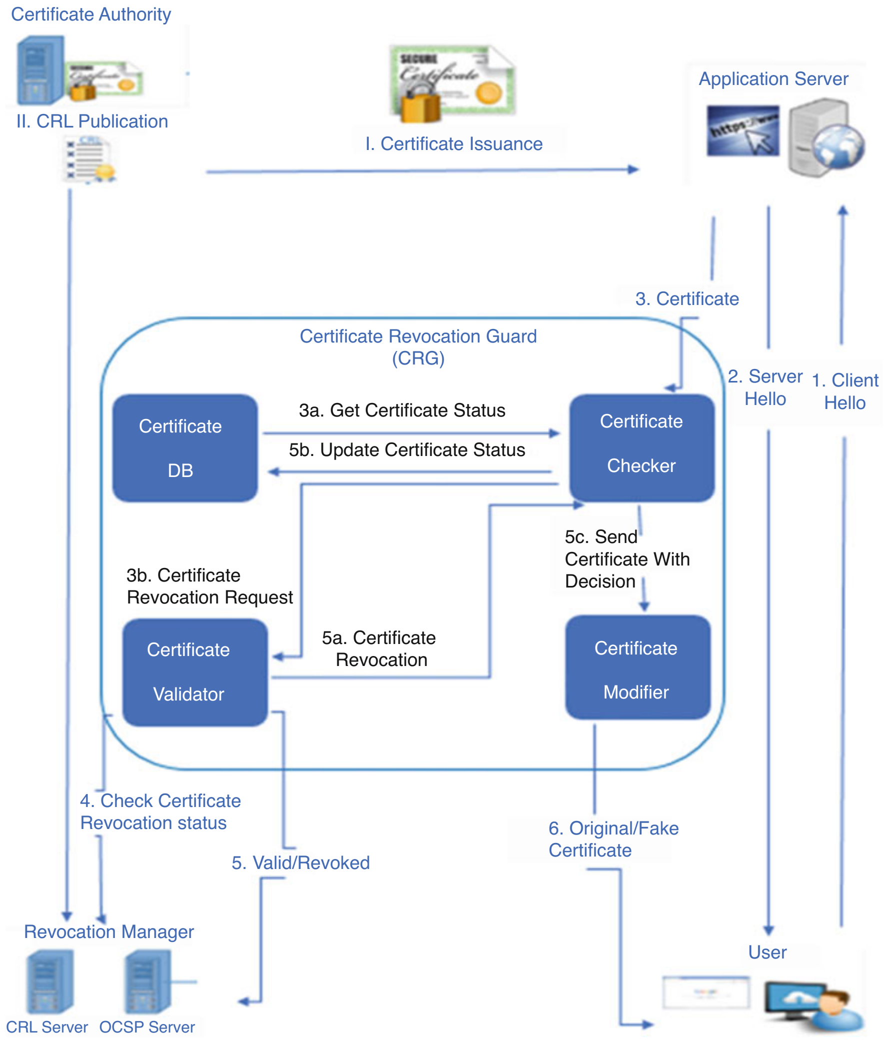


Fig. 9 A detailed view of CRG

4.3 Advantages and Disadvantages Associated with PKI

PKI has various advantages; however, it has several disadvantages. One of the major limits of PKI is the possibility of violation from the company which had issued the CA, another advantage is when companies that cannot afford the cost or not willing to yield authority to the third party of a public CA would have to invest in an in-house solution. However, PKI is one of the best solutions for security problems [9]. Considering the increased use and number of digital transactions by consumers, a high assurance requires confidence [5]. PKI enables the business community to convey to the consumer that their transactions and privacy are secure and safe. A CA like a trusted third party can increase consumer confidence. Another significant advantage is cost.

According to researchers, PKI will decrease costs attenuating the support of labor-intensive transactions. High speed and increased volume of transactions will reduce the cost per transaction. This would more than likely increase the business' market share. PKI is also one of the best methods of creating an infrastructure that is safe for Internet transactions and from hackers, theft of Personal Information (PI), and virus injection. These may include an environment where users can meet in private. PKI is usually not required in a single-user environment. Also, PKI does not replace symmetric encryption but to augment it to make it more secure.

5 Deploying an ECC PKI

This section examines various topics relating to the deployment of a PKI based upon ECC.

5.1 Public Key Algorithm Selection

The first issue in PKI deployment is to decide about which class of crypto systems are to be used for key generation. Many cryptographic algorithms are available, e.g., RSA, DSA, ECC. Selection of one of these algorithms depends on many factors such as speed, ease of implementation, and public acceptance. In the Internet environment, PKI technology is almost exclusively based on RSA cryptosystem but in wireless world a newer, alternative cryptographic technology called elliptic curve cryptography (ECC) is becoming popular. ECC can perform the same basic function as RSA but with fewer CPU resources and smaller key size. This makes signatures computationally inexpensive [10]. As we have seen, ECC has some theoretical advantages over RSA. But some practical considerations have to be taken into consideration, especially the challenges of transitioning from an RSA infrastructure to one based upon ECC.

5.2 Transitioning to ECC

Before planning a transition from a security architecture based on RSA to one based on ECC, a various number of factors should be considered. This section highlights some of these factors. Most modern security infrastructures are designed to be independent of algorithms. So, in theory, it should be possible to switch from one algorithm to another with relative ease. However, in reality, a number of details have to be worked out: first, the key sizes and curves must be chosen; then the applications to be supported must be cataloged and investigated; and the necessary infrastructure products, application software and hardware must be identified, sourced, tested, and deployed.

Second, the operational environment should be simulated in a laboratory setting in order to confirm the compatibility of the chosen products and rehearse the transition. Also, the software based on ECC that encrypts and verifies signature has to be deployed to all users before ECC keys are issued and data is encrypted or signed. Then, ECC should be authorized for users progressively, so that teething troubles can be identified before a large user population is affected. Preparations such as helpdesk training and updating affected policies and practices should be completed prior to the system going live. Because the deployment of ECC-enabled software must take place before any ECC keys are issued, it is necessary that all new application software be capable of processing the existing crypto algorithms in addition to the ECC ones [11]. And if the planned architecture contains devices that only support ECC, then these cannot be introduced until the later stages of the rollout. Once all end-user software and hardware are deployed, we are ready to make the transition. This can be achieved as a normal key-rollover operation. CAs typically accomplish this by using the old (RSA) root key to create a link certificate containing the new (ECC) root key and similarly using the new root key to create a link certificate containing the old root key [11]. As we have cited, the transition from a PKI-based RSA to a PKI-based ECC has a large number of challenges, which will make this transition difficult to achieve in the real case. However, the theoretical study of this transition will facilitate the task of real technical realization.

6 Conclusion

This chapter gave an overview of public key infrastructure (PKI) and elliptic curve cryptography (ECC), in which we investigated the PKI technique, definition, workflow and lifecycle, advantages, and limits. Hence, we exposed the problem of certificate revocation in PKI. Therefore, we present the elliptic curve cryptosystem, and we proposed the idea of a transition from a public key infrastructure-based RSA to public key infrastructure-based ECC. The proposed approach allows us to have a lightweight PKI with a high-security level but also with a much shorter key, hence reduced computing requirements and low CPU utilization. In theory, it has to be possible to swap from one algorithm to another with relative ease. Nonetheless, in reality, the number of factors and details should be considered, so in the above, we present the main factors and limits of the deployment of the proposed approach. As a perspective, we will further study the integration of ECC in the PKI in order to realize an adequate model, simulate it, and compare the performance of this lightweight PKI with other PKI based on the different asymmetric algorithms.

References

1. Cole, E. (2008). *Network security fundamentals*. Wiley.
2. Barukab, O. M., Khan, A. I., Sharief Shaik, M., Ramana Murthy, M., & Ali Khan, S. (2012). Secure communication using symmetric and asymmetric cryptographic techniques. *Int. J. Inf. Eng. Electron. Bus., 4*(2), 36–42.
3. Villanueva, J. (2015). *What is a digital signature?* http://www.jscape.com/
4. Khatua, S. (s.d.). *International Journal of Latest Trends in Computing*, *l2*(2), *2*(2).
5. Uchimi, A. (2003). What is PKI. *Kyokai Joho Imeji Zasshi/Journal Inst. Image Inf. Telev. Eng., 57*(5), 585–588.
6. Braun, J., Volk, F., Classen, J., Buchmann, J., & Mühlhäuser, M. (2014). CA trust management for the web PKI. *Journal of Computer Security, 22*(6), 913–959.
7. Zhang, L., Choffnes, D., Levin, D., Dumitras, T., Mislove, A., Schulman, A., et al. (2014). Analysis of SSL certificate reissues and revocations in the wake of Heartbleed. In *Proceedings of the conference on internet measurement conference* (pp. 489–502). ACM.
8. Liu, Y., et al. (2015). An end-to-end measurement of certificate revocation in the web's PKI. In *Proceedings of the ACM conference on internet measurement conference* (pp. 183–196). ACM.
9. Oppliger, R. (2014). Certification authorities under attack: A plea for certificate legitimation. *IEEE Internet Computing, 18*(1), 40–47.
10. Gura, N., Patel, A., Wander, A., Eberle, H., & Shantz, S. C. (2004). Comparing elliptic curve cryptography and RSA on 8-bit CPUs. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 3156*, 119–132.
11. Bafandehkar, M., Yasin, S. M., Mahmod, R., & Hanapi, Z. M. (2013). Comparison of ECC and RSA algorithm in resource constrained devices. In *2013 international conference on IT convergence and security*. ICITCS.

Smart Business Process Modeling: Toward an IoT to Detect the Data Flow Anomalies in Ad Hoc Mesh Network

Najat Chadli, Mohamed Elhoseny, Mohammed Issam Kabbaj, and Zohra Bakkoury

Abstract The smart business processes modeling SBPM in a new technology used an Internet of Things IoT in the big companies. The IoT contribute by the owners of these keys such as a sensor and both physical and virtual Universe and smart camera etc. This IoT technology's coin mechanism provides an optimal routing from the device to the device. However, this smart technology cannot detect the data flow anomalies when modeling. Indeed, the sensorial of IoT define the spread of smart detection features that resolve the issues of data flow. The aim of this chapter is to control data flow modeling in SBPM when the IoT concept release and forwards the data in a meshed network at routing of data from activity to the activity. The main reason for using IoT is to implement smart keys in an active help method and an ad hoc mesh network. However this method need a database storage concept to serve the method by recording the data and their read, write/destroy state at modeling data during routing. This database is also used to clean up data when it is detected by active sensors and the smart camera. The approach has a goal to find data flow anomalies as missing data, redundant data applying a computation method of the average speed during a prescribed distance. This average speed is compared to the threshold of speed at a given time instances when the intelligent business process modeling (SBPM) applying an IoT.

Keywords Smart BPM · IoT · Smart camera · Active sensor · Active help

N. Chadli (✉) · M. I. Kabbaj · Z. Bakkoury
AMIPS Research Group, Ecole Mohammadia d'Ingénieurs, Mohammed V University, Rabat, Morocco
e-mail: najatchadli@research.emi.ac.ma

M. Elhoseny
Mansoura University, Mansoura, Egypt

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_2

1 Introduction

In recent years, the internet of thing IoT that is invented by Kevin Ashton back in 1999 and the term are able to connect and exchange data. This novel paradigm of technologies has rapidly gained a population in the estate of communication with the wireless and wired [1]. Indeed, the foundation smart business process management has a challenge [2].

Making a deep in the modeling of the business process in an information system in a big enterprise required to manage the risk and has a new technology to control flow and data flow. The IoT devices in a central business process models can accumulate, and filter uses data [3]. The smart business process modeling introduced in modeling IoT devices for an exchange data in an information system using a device active sensor. This way, the routing of data in an information system from activity to activity required to be in phase modeling to detect the anomalies at the time of processing. The smart supports used are the smart camera and speed camera to detect the data flow at routing in a smart workflow [4]. The IoT system is a smart kind to permits an ad hoc method to have a means a correction the errors just as soon as the detection of data anomalies. The active help method chosen in an ad hoc network is used to assist detection at the time of modeling in each process fragmentation [5–7]. The business process modeling extension IoT [8] in a wireless sensor of the intelligent camera (speed camera) in autoroute is extending an internet of everything IoE [9].

The aim of this challenge of new technology integration the IoT in business process modeling is a concept to assist the correction the anomalies of data flow in the AutoRoute system. Indeed, the use of a database storage [10] to record the data flow, and also a concept to correct data also to clean the server of the no used data. The goal of this chapter is to define the role of the IoT concept of smart business processes modeling SBPM such as used an ad hoc approach applying active help at each instance of time in the fragmented process to detect the data flow anomalies when the data are captured (vehicle identification) through an active sensor in the road traffic on the highway. In this sense, the errors are caused while the active sensor detects the identification that is made by many springs during image treatment such as missing data, conflicting data, redundant data, and so on. Thus, the work is focused in missing data and redundant data in one process have a several activities [11]. The type for the model is a sequential process used in data flow exchange of information system depended in smart keys of internet of thing IoT.

The reminder of this chapter is structured as described below in Sect. 1 is the introduction, Sect. 2 is a related work to cite a certain approach to the Internet of Things and for modeling business processes and data flows. The Section 3 allows some definition of the subject. The Sect. 4 provides the motivation and use case with the new approach. The ending is the conclusion.

2 Related Work

Most organizations use an information system. This system needs data that make a transition from one task to another or from one activity to the next, that's why this data needs a routing that requires a network mesh that transforms the data from one passover to another. Indeed, the process determines a modeling of these tasks from the beginning of the data entry to the end of the data output. We try to have an overview in business process modeling with the data flow when errors are detected. This literature is an introduction to a new technology that gives an intelligent business process modeling that has been enriched by the Internet for Iot things with an active sensor and camera intelligence. This technology is used to control vehicles on the highway. At the heart of this study we cite some research using this new technology in several areas. Some article has suggested a sensor clustering method for dynamically structuring a heterogeneous WSN through GA called "DCHGA." DCHGA provides a framework for incorporating multiple heterogeneity and clustering factors [12]. Thanks also to the use of the new and emerging generation of computer technologies, such as Internet, IoT, Internet and other services, etc. It to provide a high level of family lifestyle and high quality services and family life. A correspondingly high quality of comfort, convenience, interaction, enhanced safety and energy efficiency as show in [13]. In the aim of optimizing the life of the entire network, the genetic algorithm GA is used to search for the most suitable sensor nodes to relay signals to BS, such as CHs. Using the selected CHs, the sensor aggregates are formed in such a way as to optimize the total transformation of the internal node concentration into CH distance as in [14]. The paper survey to analyze the secure clustering techniques that exist and indicate the extent to which each method is applied. They used S - CH, S - CF, S - CF, S - DA and S - DR to indicate to the four phases respectively: 1-secure cluster Head selection, 2-secure rotting to Base-station, 3-secure data aggregation,4-secure cluster formation as in [15]. Some paper fairytale has looked at the most significant aspects of IoT, with an emphasis on what is being done and what issues need more in-depth research. That's right, the current technologies are making the IoT concept realizable but do in fact not adapt well to requirements for extensibility and efficacy [1]. The article also shows how implementation workflows used to provide conceptual information at different semantic layers for smart workflows can be used [3]. In this work, they use the standard BPMN only to determine both central and IoT behavior of business processes and also use the resource element BPMN to incorporate information from IoT peripheral devices into the model as in [4]. This document provides a solution to the current problem that the role of IoT devices as a resource type of a business process is not directly represented in standard process models to date. This refers to the following: IoT devices and the non-standard software constituents do not have any process components from the point of view of ERP systems! In fact, this work looks at the way in which the "IoT device" element and its native services can be articulated as assets in an IoT aware process model as in [8]. In the following paper, the authors presented an alternative view of a design-based in decentralized private IoT. The first is a peer-to-peer network (P2P) in which data from IoT devices is privately stored (instead of being entrusted) to central companies [16]. The second, A peer-to-peer cloud storage network using end-to-end encoding would provide a means for users to transmit and exchange data without the need for a third-party data provider [17]. This literature has to improve our work on how to use advanced technology IoT by a smart camera and an active sensor to detect data flow anomalies in intelligent modeling of business processes on the highway.

3 The Concept Internet of Thing IoT and the Internet of Everything IoE

3.1 Internet of Thing IoT

The internet of thing IoT is a new paradigm of technologies in the world, that why is mean the interconnection between the internet and object. The concept of IoT is a native software [9]. The IoT can amount, and filter data and to construct the locally decision [3]. Though the sensor technologies, wireless technologies may also contain on IoT. One important step, the object of IoT significative could reproduce itself digitally, that why, grow to be more important than the object by itself. Absolutely, the object identified just to its user. In fact, it's connected to the object and database.

3.2 Internet of Everything IoE

The Internet of Everything IoE is an extension of the internet of thing IoT. The IoT is interested in the physic object and the IoE [18] is newly trends by a concept of thing then process, people and data, this IoE is invented by Gartner in 2015. The structure of the internet of everything:

People: Connecting people in more relevant, valuable ways.

Data: The create a better decision by Converting data into intelligence.

Process: The correct information exchange in the right person or machine at the best time.

Things: The internet and the other thing be connected with the object and physical devices that each other make the right intelligent decision, called the internet of thing IoT.

3.3 Database Storage

A major challenge for IoT application developers is to clean, process and analyze the large amount of data collected by active sensors. There is a recommended solution for analyzing information called wireless sensor networks [19, 20] these networks share data between sensor nodes that are sent to a network system for sensory data analysis.

4 Motivation and Use Case for New Approach

4.1 Problematic

In the highway, the traffic monitoring uses the servers contains databases to record data and assist in the detection of vehicle images during traffic. This detection will be at the registration number of the vehicle for several reasons either to calculate the acceleration speed of the vehicle or security applications such as restricted area access control and vehicle tracking, etc. However, the license plate is the unique identifier of a vehicle, and its image provides important information to recognize its owner. Often, it is necessary to identify vehicle license plates for safety reasons. The extracted information can be used for several purposes, such as access and flow control, border and tollgate surveillance, search for suspected vehicles or the fight against criminality, etc. That is why their reading is very important and certain in all these areas to read vehicle license plates.

In conformity with Moroccan registration standards [21]. License plates in Morocco have changed style in early 2000. This is now the third "generation" of plates, the change is caused by the inadequate and overloaded old system (######-###|#) The last number was a complete generation of plates, and the change is due to the unsuitability and overloading of the previous system because the last one included a complete region. (At that point in the case study the plates were identified differently according to the eight regions that the die had) The new system contains letters and numbers (in Arabic). The plates are now of the style ####### | A | | ##. They have a white background and the letters are in black. When the one letter series is out of print, we move on to the following letter of the Arabic alphabet. The third part of the plaque is made up of a number of one or two digits, this time symbolizing not the regions of the kingdom, but its prefectures and provinces. The vehicle registration problem has resulted in a disruption in the server that tends not to be able to write it to the database, so this problem is due to the first level of the checkpoint. Nevertheless, the identification number was read at the second level of the server upon its arrival, which gives us a missing data. There is also a further complication, namely the return of the vehicle to the highway in the opposite direction. For this case, it is assumed that the image handling has written the identifier into the database given at the beginning, but when the vehicle has disappeared, so redundant data or also in the case of troubleshooting the vehicle it can take a path that the arrival point.

4.2 The New Approach

The new technology has simplified and solved many problems in various environments. Therefore the use of our architecture with the Internet of the IoT object which includes an intelligent camera and an active sensor that simplifies the task in Auto-routes of monitoring vehicle traffic speeds at a specified distance and required by traffic monitoring services. As well as a restricted accelerating speed that is the threshold for the comparison of vehicle speeds. In addition to the prescribed distance, the modeler added a chronometer for time measurement when the vehicle passes through the prescribed distance in the meantime. This measurement is a series of decreasing time (tn to t0) that will result between start time and finish time as a function of speed threshold. Moreover, at every instance of the time printed and computed by active sensor, the control is performed by an active help, which was suggested to help the modeler proceed with the verification when the anomalies are detected through the ad hoc approach. While the other vehicles in circulation, the system are locking to correct the errors of data flow at modeling.

These vehicles in circulation need to be controlled, that why the concept database storage is used to record the data. Therefore, the database storage of data gives record and cleaning that play the concept roles and to record the data with their current state (Read, Write, Destroy) detected by active sensor. Indeed, the role of the Internet of all that IoE is an extension of the Internet of Things that IoT has been conceptualized for Smart Business Process Modeling ("SBPM"). As long as this IoE is a concept of many other things, data, processes, servers and people. Inside the system in the first server and at the first detection by the smart camera, it is deduced that the active sensor cannot describe the vehicle identification(id) due to other external factors that were mentioned in the problem during the distance covered. After the distance required by the traffic monitoring service, the active sensor reads the vehicle id detected by the smart camera that has been recorded in stored in the data base as show in Fig. 1a, b.

As much as the system checks of process, will calculate the average speed of the vehicle's acceleration and will compare it against the threshold speed V0. For this reason, the server will not retrieve the data until it has been entered into the server. Thus, to calculate the average speed with respect to time, the data must have a written state W, read R or destroy D (R, W, D). Remain in the same direction of control, another case that arises during vehicle traffic, when the vehicle id is written in all sensors during the prescribed distance with the calculation of the arrival time at this end point. At the arrival time at the terminal described, the system couldn't read the id by the sensor, can be caused by other external event which causes redundant data. The goal of this method is to reduce the risk of full up in an infinite loop when the verification didn't apply until the modeler finish modeling.

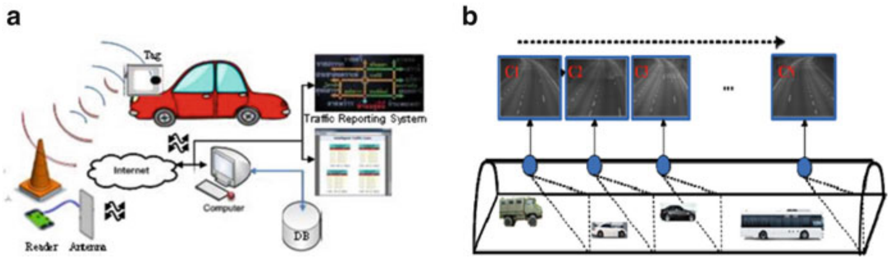


Fig. 1 (**a**) The concept IoE used in AutoRoute. (**b**) Detection of vehicle by smart camera

$$\text{The speed formula is speed}(v) = \frac{\text{distance}\ (d)}{\text{time}(t)}$$

The average speed is:

$$\text{final distance} = df$$

$$\text{starting distance} = di$$

$$\text{final time} = tf$$

$$\text{departure time} = ti$$

$$\text{average speed} = Vav$$

$$Vav = \frac{(df - di)}{(tf - ti)}$$

Let x, y, z be elements that describe the data state
x: data is reading R
y: data is written W
z: data is destroyed D

4.3 Approach Rules

In this approach, we introduce Rules in relation of anomalies that are deduced by these anomalies to make the verification in the sense to correct the errors when the data are either the data is written, read or destroyed from the database stored.

Missing Data If the active sensor couldn't sense the data d at the starting point, data didn't write in database, entail data not initialized, that $y = 0$. This data d is read in database $x \neq 0$. Hence, detection of missing data.

At the final distance, the second active sensor and the second smart camera detect the vehicle identification (id) with their speed of vehicle traffic v', while $y \neq 0$.

Additionally, the filter of the internet of things IoT has not yet destroyed the data. The data is always in the database stored, so $z = 0$, while this is a missing data.

Redundant Data Following the same procedure, the data is saved in the database of the first server. So, the sensor described the data and wrote them in the server of data base. During traffic, the vehicle is attacked by problems from the outside, causing a reading difficulty when this happens will always remain in writing until arrive in the full stop. This leads to data redundancy. Therefore, $x = 0$ at the beginning to the end of the distance traveled $y \neq 0$ and $z = 0$.

4.4 The Use Case

The configuration of the IoT network is driven by business process modeling and added as a component. Therefore, business process modeling requires a workflow to use their component and route data with input and output to each activity. However, IoT devices are often low-powered, which leads to problems with the software they can use. In our case, we try to use the smart camera, with active sensor in road traffic to detect vehicle identification in image processing apply the ad hoc network mesh approach with active help at each time and servers contain databases that do the design with data stored Let us also add a server that collects all the data and belongs to the traffic monitoring. This monitoring is done for Vehicle Control when the speed required on this road exceeds the threshold speed or when vehicles are in the opposite direction of traffic and so forth. In this way, both the communication for the smart camera and vehicles, the vehicle identification recording server for the vehicles traveling on the highway was realized through the sharing of the information system data using the Internet for IoT things. In effect, the communication to other objects and the filtering of the data and correction of data stored or destroyed in a specific database. Under this architecture, the information systems used data flow exchanges from the first server to the other components as requested during traffic flow. When switching from one server to an another, a routing path is used in which the appropriate network mesh has fulfilled its role. The required time to detect the id of vehicle by smart camera in image processing,

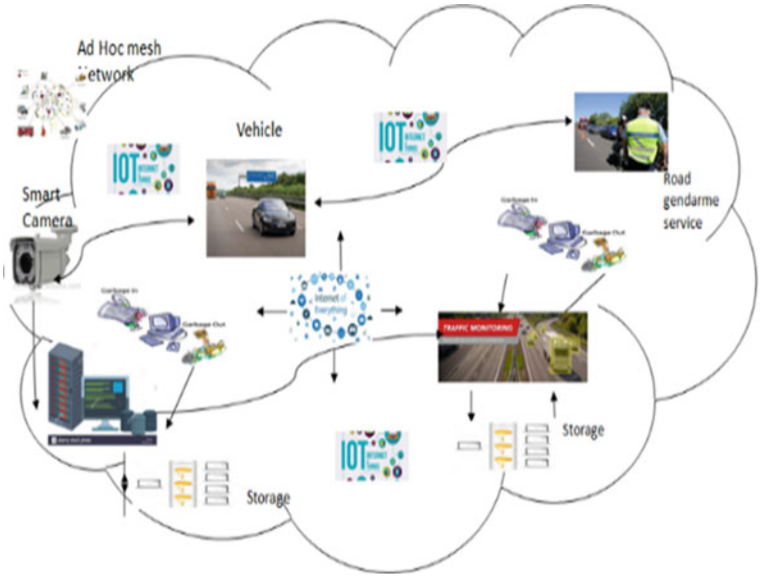


Fig. 2 Use case of road traffic in smart IoT

there are errors that occur which have been discussed in the problem section, but the value of road surveillance is to calculate the average acceleration speed of vehicles for comparison with the prescribed threshold speed and also to check the road. After errors at the data routing court in the information system, traffic controllers submit a penalty to the traffic police for the owner of the stopped vehicle. Therefore, the IoT Internet object with a design and tools improved this road surveillance study by using the Internet to help any IoE to perform a data exchange as shown in Fig. 2.

4.5 Description of Architecture

For drawing the model of this architecture, it needs to draw two lanes one of business process modeling to route the data in a sequence activity, the second lane is to modeling the internet of thing with a filter of data in the garbage in and garbage out and a smart service home system [19] to send the penalty. Therefore, this modeling of data flow in the business process and in IoT are communicating with each other by a mesh of ad hoc networking. Therefore, this modeling of data flow in the business process modeling applying an IoT's keys to communicating with the things, data, process and database by using the approach of a mesh of ad hoc networking using a data exchange of information system required by an active help.

In this mean, data is controlled in two modes, by modeling IoT and mesh ad hoc network. The stored of data on each record in the database is not easier to be

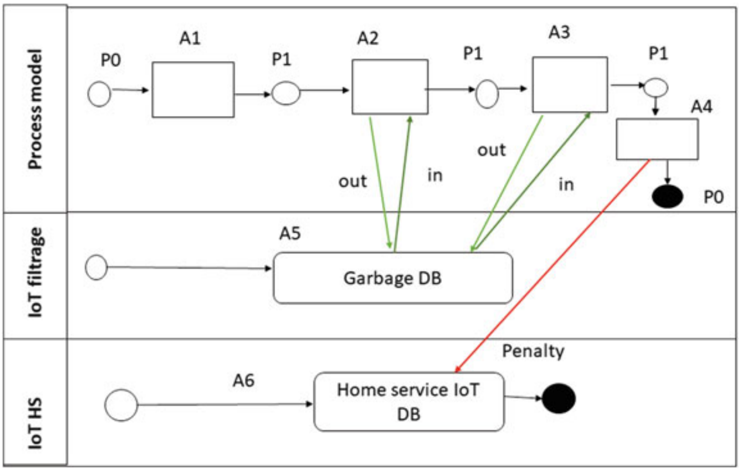


Fig. 3 Designing the process model and IoT modeling

Table 1 Description of activities

Activities and sever	description	Role
A1	Smart camera1	Detect the id vehicles, image processing
A2	Record server DB	Recorded the data id
A3	Active sensor	Sensor of id number of vehicles
A4	Traffic monitoring DB	Analyze of failing traffic by DB
A5	Garbage DB	Filter data DB
A6	Service home system DB	Penalty (traffic monitoring)

in guard until to end of processing. Thus, the data record change in the passage of activity to activity, because the time when the data are stored in the database the garbage triggered by filtering data in lane IoT as show in Fig. 3 and Table 1.

This description of activities used is considered as a process model that showing in Fig. 3.

The models' uses have four activities and five servers two to filtrate data (garbage) and two for storage of data in each concerned service and one for smart service home system to send the penalty of wrong traffic road.

The architecture used many things for everything in order to discuss how the use internet IoT in business process modeling to be smart as in show in Fig. 4. Indeed, in Fig. 5 the summary of the Fig. 4.

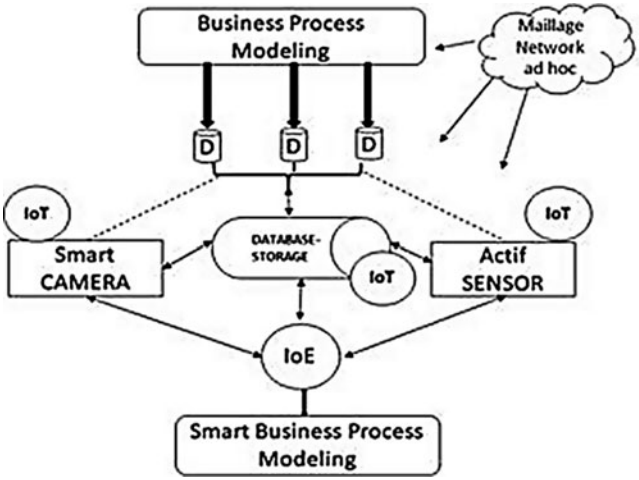
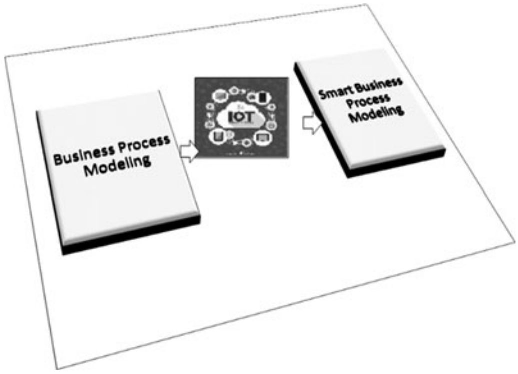


Fig. 4 Architecture of business process modeling to smart BPM by IoT and IoE

Fig. 5 The summary architecture of BPM to be smart by the IoT



5 The Approach Solution

5.1 Verification

At the time of detection of the data flow anomalies, the system locks in order to correct them. While the process is still in continuous function for processing. In this case the data captured by the camera are recorded in a database storage so as not to lose them at the system locking time. So, the sensors from the Internet of things IoT have been proposed for some distance between the initial distance to the final distance demand by the traffic monitoring. Therefore, at each times ti the verification and the control are made by the system to find the errors while the capture of vehicle identification by an active sensor recording the state of the vehicle id in write, read and destroyed as (read: R, write: W, destroyed: D) = (x, y, z) as show in Table 2.

Table 2 Compute a strange distance

Chronometer 1: way one			Chronometer 2: way two			Chronometer 3: way three		
Time	Distance	Threshold	Time	Distance	Threshold	Time	Distance	Threshold
tn	$d0$	$V0$	*tm*	$d0$	$V'0$	*tq*	$d0$	$V''0$
tn-1	$d1$	$V0$	*tm*-1	$d1$	$V'0$	*tq*-1	$d1$	$V''0$
ti	*di*	$V0$	*tj*	*Dj*	$V'0$	*tk*	*dk*	$V''0$
$t\ldots$	$d\ldots$	$V0$	$t\ldots$	$d\ldots$	$V'0$	$t\ldots$	$d\ldots$	$V''0$
$t0$	*dn*	$V0$	$t0$	*Dm*	$V'0$	$t0$	*dq*	$V''0$

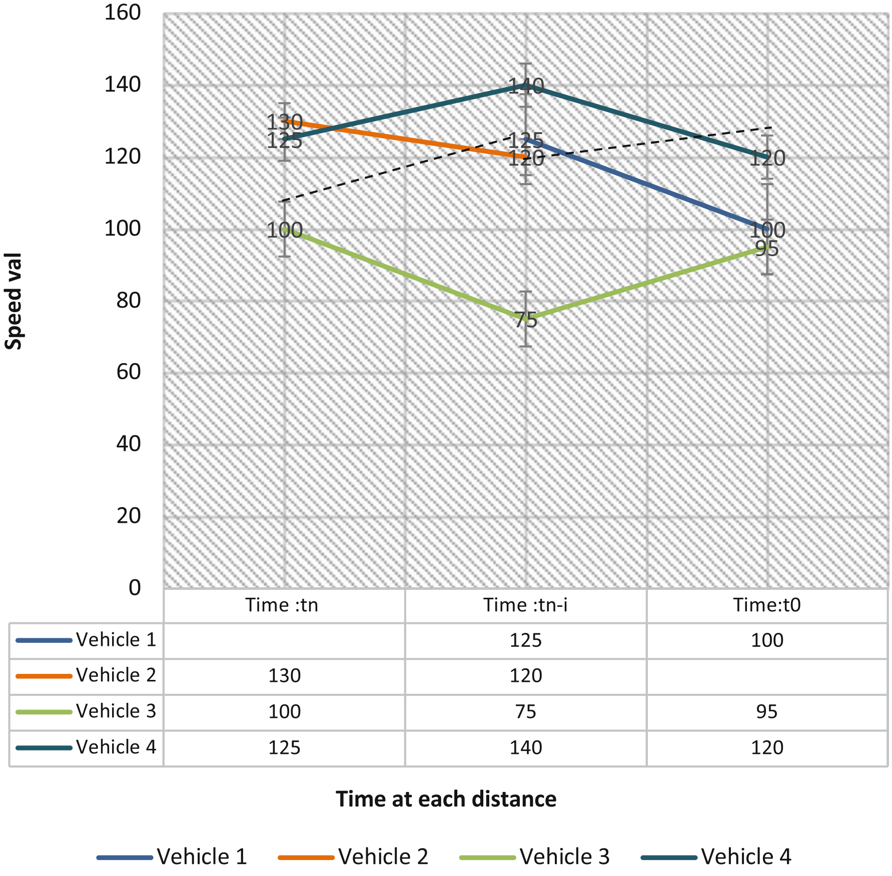


Fig. 6 The graph with correction

To compute the stranger distance *di*, *dj*, *dk* when,*j*, $k \epsilon \mathbb{N} - \{0\}$. The function of speed is $v\,(t_i) = \frac{d_i}{t_i}$

We use the formulate that show also in graph of Fig. 6.

$$tn - i * v0 = dn - i$$

$$tm - j * v'0 = dm - j$$

$$tq - k * v''0 = dq - k$$

The same when the anomalies of redundant data, the data can't be read also at last the final distance and when the autoroute use many ways.

Missing Data Rule The Solution of missing data errors, either the data found in the service record of the traffic monitoring or data rejected by the IoT filter becomes out garbage. The designer locks the system until the error is corrected. This operation is done during modeling in an active help method. Thus, the correction of errors of the vehicle identification be write or reject. For this case, the id will be written when the server of treatment of image correct the id or found it in the memory of server.

The Redundant Data Rule The solution is either the data will be destroyed, or the data becomes read in the second server.

5.2 Interpretation

The graph below shows how the ride crosses the distance described for four vehicles applied the Rules. In this case, we explain the graph of Fig. 6, we define four vehicles in use, each vehicle has an average speed in comparison according the time arrived in these times.

Hence, for the blue curved line 1 correspond to the first vehicle 1, the sensor can't detect the first speed in the input time tn, but at the next time ti, the sensor can describe the write speed, and during traffic the other sensors can read the data from the internet of thing IoT. Thus, the missing data. In this case the modeler proposed during the correction that the data is written, and the speed has a value for vehicle 1. If we look at the orange curved line 2 of vehicle 2, we see that the last data was not read from the sensor, the remained recording are still in state write note read. Therefore, the data problem is redundant. The modeler proposes the solution to read the data, mean well the tends of the sensor has detected the written data as show in Table 3 and Fig. 6.

Table 3 Interpretation Camera one in the final distance *df*

Camera two in a final distance *df* at time $t0$		
Time	Final speed	Compares the speed relative to the limit value
tn	$V\,1$	Vav $1 > V\,0$
tn-1	$V\,2$	Vav $2 > V\,0$
ti	$V\,3$	Vav $3 < V\,0$
$t\ldots$	V	Vav $\ldots > V\,0$
$t0$	Vn	Vav $n < V\,0$

6 Conclusion

In fact, we are trying to use intelligent business process modeling in an Internet of IoT things in highway traffic monitoring to detect anomalies in data flow modeling when exchange data is used by servers to calculate the average travel speed prescribed by surveillance traffic. Indeed, the stored concept database used to record data and data also to correct, according to a smart camera and an active sensor to detect data at the time of the vehicle is on the way. This chapter will cover an adaptation of the business process modeling in a new Internet of Things technology and an extended Internet of everything that is increasing wealth by our approach in a mesh of an ad hoc network use an active help method. This method requires at the time of detection of the anomaly to lock the system in order to correct the data in a database stored as mentioned at the top of the page.

References

1. Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer Networks, 54*(15), 2787–2805.
2. Mendling, J., Baesens, B., Bernstein, A., & Fellmann, M. (2017). Challenges of smart business process management: An introduction to the special issue. *Decision Support Systems, 100*, 1–5.
3. Martins, F., & Domingos, D. (2017). Modelling IoT behaviour within BPMN business processes. *Procedia Computer Science, 121*, 1014–1022.
4. Wieland, M., Kaczmarczyk, P., & Nicklas, D. (2008). Context integration for smart workflows. In *2008 sixth annual IEEE international conference on pervasive computing and communications (Per Com)* (pp. 239–242). IEEE.
5. Kabbaj, M. I., Bétari, A., Bakkoury, Z., & Rharbi, A. (2015). Towards an active help on detecting data flow errors in business process models. *IJCSA, 12*(1), 16–25.
6. Chadli, N., Kabbaj, M. I., & Bakkoury, Z. (2018). Detection of dataflow anomalies in business process an overview of modeling approaches. In *Proceedings of the 12th international conference on intelligent systems: Theories and applications* (p. 37). ACM.
7. Chadli, N., Kabbaj, M. I., & Bakkoury, Z. (2020). *An enhanced adhoc approach based on active help to detect data flow anomalies in a loop of a business modeling*. Springer.
8. Meyer, S., Ruppen, A., & Magerkurth, C. (2013). Internet of things-aware process modeling: Integrating IoT devices as business process resources. In *International conference on advanced information systems engineering* (pp. 84–98). Springer.
9. Abdelwahab, S., Hamdaoui, B., Guizani, M., & Rayes, A. (2014). Enabling smart cloud services through remote sensing: An internet of everything enabler. *IEEE Internet of Things Journal, 1*(3), 276–288.
10. Klein, Y., & Cohen, A. L. (2019). *Computer storage systems and methods of managing database server applications*. U.S. Patent No. 10,324,900. 18 June 2019.
11. Danilov, M., Buinov, K., Rakulenko, A., Skripko, G., & Zakharov, K. (2019). U.S. Patent No. 10,289,488. U.S. Patent and Trademark Office.
12. Elhoseny, M., Elleithy, K., Elminir, H., Yuan, X., & Riad, A. (2015). Dynamic clustering of heterogeneous wireless sensor networks using a genetic algorithm, towards balancing energy exhaustion. *International Journal of Scientific & Engineering Research, 6*(8), 1243–1252.

13. Acharjya, D. P., & Ahmed, N. S. S. (2017). Recognizing attacks in wireless sensor network in view of internet of things. In D. P. Acharjya & M. K. Geetha (Eds.), *Internet of things: Novel advances and envisioned applications* (pp. 149–150). Springer. ISBN 9783319534725.
14. Elhoseny, M., Yuan, X., El-Minir, H. K., & Riad, A. M. (2014). Extending self-organizing network availability using genetic algorithm. In *Fifth international conference on computing, communications and networking technologies (ICCCNT)* (pp. 1–6). IEEE.
15. Elhoseny, M., Elminir, H., Riad, A. M., & Yuan, X. I. A. O. H. U. I. (2014). Recent advances of secure clustering protocols in wireless sensor networks. *International Journal of Computer Networks and Communications Security, 2*(11), 400–413.
16. Conoscenti, M., Vetro, A., & De Martin, J. C. (2017). Peer to peer for privacy and decentralization in the internet of things. In *2017 IEEE/ACM 39th international conference on software engineering companion (ICSE-C)* (pp. 288–290). IEEE.
17. Wilkinson, S., Boshevski, T., Brandoff, J., & Buterin, V. (2014). Storj a peer-to-peer cloud storage network.
18. Shojafar, M., & Sookhak, M. (2019). Internet of everything, networks, applications, and computing systems (IoENACS). *International Journal of Computers and Applications*. https://doi.org/10.1080/1206212X.2019.1575621
19. Xu, H., Xu, Y., Li, Q., Lv, C., & Liu, Y. (2012). Business process modeling and design of smart home service system. In *2012 international joint conference on service sciences* (pp. 12–17). IEEE.
20. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems, 29*(7), 1645–1660. https://doi.org/10.1016/j.future.2013.01.010
21. Anoual, H. (2012). *Détection et localisation de texte dans les images de scènes naturelles: Application à la détection des plaques d'immatriculation marocaines*.

IT Project Risk Management Model

Vitalina Babenko, Koniaieva Yalyzaveta, Nadiia Shylovtseva, Tetiana Marenych, Olha Myrna, and Olha Serdiuk

Abstract The chapter develops a knowledge-based IT project risk management model, which, unlike the existing ones, allows to take into account the dynamic and incremental way of executing IT projects, whereby the requirements for the information product and the ways of their realization are refined taking into account new information and experience gained. According to the model, the inputs to the risk management decision-making process should be stored in the knowledge base of the project, based on which the rules for the operation of the expert system are formulated.

It is suggested to use such components of the knowledge base as the risk database and the risk management knowledge repository. The Risk Database contains information on the tasks of IT project implementation, a concise description of problem situations, directions for solving the problem, quantification of risks, and the effectiveness of risk management measures. Updated risk database information, along with implicit knowledge (project stakeholder experience and qualifications, domain laws, etc.), is used to replenish the risk management knowledge repository to derive conclusions and patterns reflecting key project risk management policies.

Keywords IT project · Risk management · Buffer management · Scrum methodology · Risk management models

V. Babenko (✉) · K. Yalyzaveta · N. Shylovtseva
V.N. Karazin Kharkiv National University, Kharkiv, Ukraine
e-mail: vitalinababenko@karazin.ua

T. Marenych
Kharkiv Petro Vasylenko National Technical University of Agriculture, Kharkiv, Ukraine

O. Myrna · O. Serdiuk
Poltava State Agrarian Academy, Poltava, Ukraine
e-mail: olga.myrna@pdaa.edu.ua; olga.serdiuk@pdaa.edu.ua

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_3

1 Introduction

In the article, a knowledge-based Risk Management Models (RMM), which, unlike the existing, allows for considering a dynamic and incremental manner of an IT project's execution implying that requirements to the information product and the means of their fulfillment are adjusted based on new information and analysis of experience, is developed. According to the model, input data for the Risk Management (RM) decision-making process should be stored in the project's knowledge base, based on which the rules for the expert system are formulated.

IT projects are characterized by unclear initial customer requirements, constant changes related to intangible deliverables, a high level of complexity in terms of the use of intellectual and technological resources, and therefore uncertainty due to lack of information to make project management decisions. In addition, as numerous studies show, the implementation of IT projects often violates the deadlines, exceeds the budget and develops functionality that does not meet user requirements. A high level of uncertainty increases the risk of failure to achieve the intended results and threatens the successful implementation of the project.

Therefore, effective RM is a prerequisite for proper implementation of IT projects. In the context of project management, risk is associated with events that have certain implications for the project. Accordingly, RM includes processes that maximize the likelihood of achieving project goals by actively managing threats (risks that can adversely affect a project) and opportunities (risks that can positively impact it). These processes involve the use of specific RM methods and models, the choice of which depends on the particularities of the implementation of the IT project.

In the Sect. 3, it is advisable to consider the project as a system that can be in several states that meet a certain degree of risk to quantify the risks of an IT project. It is suggested to use Markov chains as a simulation tool to determine the probability of a system. It shows properties of potential states of IT project implementation, conceptual model of IT project RM and project implementation states.

In the Sect. 4, there is a such model, where inputs are based on information resources such as historical data, experience, stakeholder feedback, observations, forecasts and expert opinions.

2 Related Works

A wide range of RM issues is covered in the works of Kuznetsov et al. [1], Ramazanov et al. [2], Kaminski [3] and others. Some aspects of project management are covered in the works of Galitsyn and Suslov [4], Teslenko [5] and Danchenko [6]. The issues of project RM are devoted to the work of Rach [7], Sviridova [8] and Skopenko et al. [9]. Foreign experts in IT project management, such as Archibald [10], Boehm [11], Schwalbe [12], Demarko, and Lister [13], attach great importance to RM processes. Goldratt [14] and Leach [15]. Recent publications on the use of methods and models in IT project RM include the works of foreign authors Machak [16], Chotkirtikul [17], Jeon [18] and Kumar [19]. Among Ukrainian scholars, the works of Tesli [20], Babenko et al. [21, 22], Shorikov et al. [23], Rishniak [24, 25], Melnyk [26], Kolesnikov [27], Onishchenko [28], and others.

At the same time, insufficient attention is paid to the application of methods and models in RM, taking into account the particularities of the implementation of IT projects, in particular the methodologies used in the creation of software. In this regard, an urgent scientific and applied challenge is to improve the RM processes of information technology projects by developing appropriate methods and models that combine the benefits of different methodological approaches to software development.

3 Possible States of IT Project Implementation

In order to quantify the risks of an IT project, it is advisable to consider the project as a system that can be in several states that meet a certain degree of risk. In this regard, it is suggested to use Markov chains as a simulation tool to determine the probability of a system being in a certain state in n steps:

- state $\omega 1$. The implementation of the project does not cause any threats that may adversely affect the expected results and objectives of the project. The deviations from the targets are not significant and the risk level in the project is within acceptable limits. The quantitative risk assessment indicator (time or cost buffer) is in the green (safe) zone. Staying the project in state $\omega 1$ at the time of completion means that from a budget and schedule standpoint, the project has been completed successfully, so this status is desirable for the project. If positive events occur and the effects of negative events are negligible, the project continues to be in state $\omega 1$, while as a result of significant negative events, the project may go into state $\omega 2$ or $\omega 3$;
- state $\omega 2$. There are threats in the implementation of the project that can adversely affect the expected results and objectives of the project. Negative events occur at the level of the circumstances of the project implementation and may adversely affect the achievement of the project objectives in terms of obtaining the expected results. This situation indicates the need for additional risk analysis and possible steps to return the project to the desired state $\omega 1$. Buffer depletion is in the yellow (relatively safe) zone. Staying the project in the state of $\omega 2$ at the time of completion means that from a budget and schedule standpoint, the project has been conditionally successful since no time and cost savings were used. If positive events occur and the effects of negative events are negligible, the project continues to be in state $\omega 2$ or return to state $\omega 1$, while as a result of significant negative events the project may go into state $\omega 3$;

Table 1 Properties of possible states of IT project implementation

State properties	Possible states of project implementation		
	State$\omega 1$	State$\omega 2$	State$\omega 3$
The nature of the state	Desirable	Acceptable	Unacceptable
The nature of threats	Almost nonexistent	Can negatively affect the expected results	Negatively affect the expected results and strategic goals
The feasibility of risk management measures	Taking measures is not appropriate	Additional risk analysis, possible action	Action is necessary
Buffer usage rate	Low	Acceptable	High (unacceptable)
Part of the heat chart	Safe	Conditionally safe	Unsafe
The success of the project	The project was successfully completed	The project was conditionally successful	Project failed

- state $\omega 3$. The implementation of the project creates threats that adversely affect the expected results and objectives of the project. Negative events affect the expected results and strategic goals. This situation indicates the need to take measures to return the project to the desired state $\omega 1$ or intermediate state $\omega 2$. The rate of buffer depletion is in the red (dangerous) zone. Staying the project in the state of $\omega 3$ at the time of completion means that from the budget and schedule positions the project was not completed successfully, as there were violations of time limits and exceeding the budget limits. If positive events occur and / or RM measures are taken, the project may return to $\omega 2$ or $\omega 1$ in one step, while new significant negative events result in the project remaining in $\omega 3$ [29]. The main properties of the possible states of the project are specified in Table 1 [1–4].

The form of the IT project RMM describes the processes associated with it according to the project implementation status, depending on the buffer usage indicator (Fig. 1).

According to the developed conceptual model, IT project RM (detection, analysis, strategy development and control) involves continuous communication between project stakeholders, periodic risk control and prompt response depending on the degree of risk determined by the project buffer utilization indicator (time and/or costs) [5, 28].

Within these processes, the identification and assessment of risks in the project are determined. The result is a risk response plan, which is part of the project and input management plan for the risk monitoring and management process. Starting with the implementation of an IT project, it is possible to realize the identified and the emergence of unidentified risks [6, 27]. IT project risk management (PRM) is carried out by analyzing the current status of the project and making decisions on managing its parameters based on the information received; Managed project

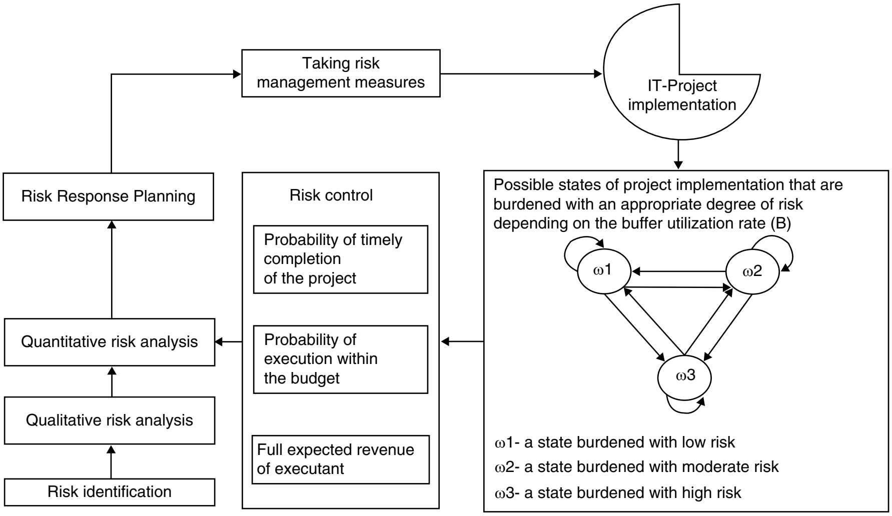


Fig. 1 Conceptual model of IT project risk management. (Source: developed by the author)

parameters are aspects or constraints that affect the end result, such as lead time, budget and project content.

Information on the emergence of new and the implementation of identified risks is also input to the risk control process. Based on the analysis of the information received, a decision is made on the feasibility of taking measures to minimize threats and maximize opportunities [7]. If such actions, in the management's view, are necessary for the success of the project, risk control measures are taken. However, if there are good reasons to believe that the risk response measures provided by the plan will not be sufficient to ensure the success of the project, changes to the risk response plan may be amended. As a result of refraining from taking corrective action or taking action to maximize opportunities and minimize threats, the project can move from one execution state to another. Accordingly, the main task in the RM process is to provide the conditions under which threats have minimal impact on the status of the project, while the opportunities—the maximum [8, 9, 25, 26].

Thus, within the framework of an IT PRM model, the use of Markov chains enables one to estimate the likelihood of the project being in the desired state. The estimates obtained, in turn, provide information and reason for taking RM measures in case of unfavorable deviations from the expected level or strengthening of positive trends, if such deviations are favorable. In Fig. 2 shows the presence of a conditional project in three possible states, depending on the values of the buffer usage indicator during its execution [29].

According to the above state distribution, we have a set of project states that form a complete group of events:

$$\Omega = (\omega_1, \omega_2, \omega_3)\,,$$

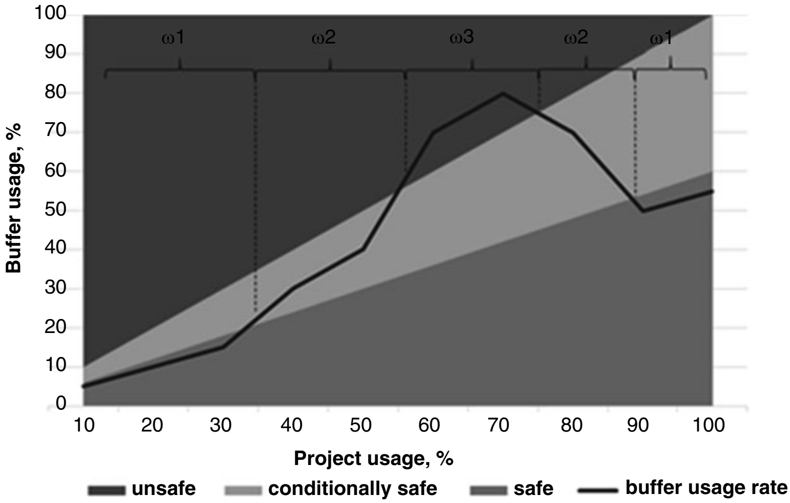


Fig. 2 Project implementation states (Source: developed by the author)

where ω_1—state 1, ω_2—state 2, ω_3—state 3.

The transition probabilities from one state to another correspond to the elements of the matrix:

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix},$$

where p_{ij}—the probability of the system going from state i to state j, $i, \ j = 1, 2, 3$.

The probability of finding a project in a certain initial state is determined by the vector of initial probabilities:

$$\overrightarrow{\pi_0} = \left(\pi_1^0, \pi_2^0, \pi_3^0\right),$$

where π_i^0—the probability that the project is in an initial stage ω_i, $i = 1, 2, 3$.

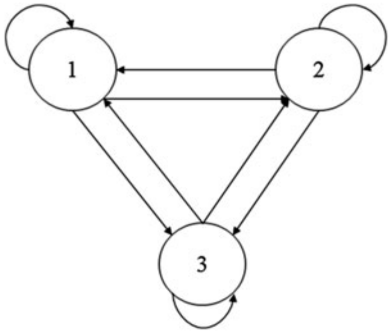
Then, according to the Markov property, the probability of finding the project in a certain state in n steps ($n = 1, 2, \ \ldots, N$) is obtained as the product of the vector of initial state on the matrix of one-step transition in degree n:

$$\overrightarrow{\pi^{(n)}} = \overrightarrow{\pi_0} \cdot P^n, \tag{1}$$

where $\overrightarrow{\pi^{(n)}}$—vector of project state probabilities in n steps [12, 13].

Links and transitions between project states can be displayed as an oriented graph (Fig. 3) [14, 15].

Fig. 3 Graph of states and transitions between project states. (Source: developed by the author)



All vertices of the graph shown in Fig. 3, are significant, that is, they mingle with any vertex that follows it. Therefore, at the interval corresponding to the time interval between system observations, the project can move from one state to another. All three vertices correspond to significant states. A chain that contains only significant states forms a single equivalence class and is homogeneous (ergodic). However, given that all vertices have a loop ($i \to i$), the chain is regular [30].

4 A Knowledge-Based IT Risk Management Model

To exploit the benefits of using knowledge management techniques to reduce uncertainty, it is advisable to develop a knowledge-based IT RMM. The practical implementation of the model involves the use of knowledge management methods and the development of a decision support system based on information obtained with such methods. The proposed knowledge-based IT RMM is shown in Fig. 4 [23].

According to the knowledge-based IT PRM model, which is part of the Information Analysis and Risk Response block (Fig. 1), IT PRM is based on explicit and implicit knowledge. The input to the RM decision-making process should be stored in the project knowledge base [24]. The project participants use information from the knowledge base about the nature of the tasks and their place in the project.

Inputs are based on information resources such as historical data, experience, stakeholder feedback, observations, forecasts and expert opinions:

1. explicit knowledge available, which includes lessons learned, databases, historical data, etc., is updated and entered into a risk database. The risk database must contain qualitative and quantitative indicators regarding the classification, criticality, and likelihood of risks;
2. updated information from the risk database, together with implicit knowledge (experience and qualifications of project stakeholders, domain laws, etc.), is

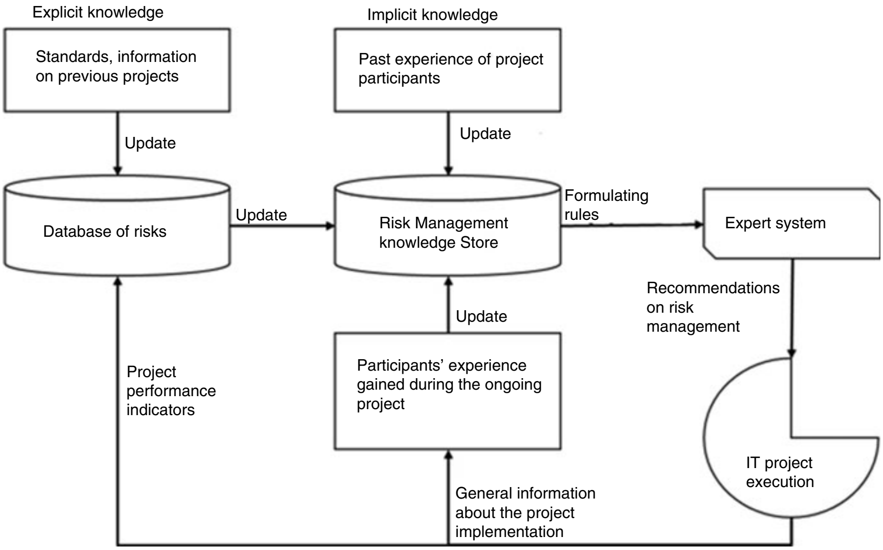


Fig. 4 A knowledge-based IT risk management model. (Source: developed by the author)

used to derive conclusions and patterns reflecting basic RM policies, as well as relevant assessments and assumptions;
3. based on the conclusions and patterns obtained, rules are formulated for the operation of an expert system, which helps the user to identify and solve problems through a series of consecutive questions and answers;
4. the recommendations provided by the expert system can be applied to PRM;
5. starting a project based on information at its execution (explicit knowledge, which includes data on the value of the buffer utilization indicator) and stakeholder feedback (implicit knowledge), the knowledge-based RM system receives new data and the cycle is repeated [10, 20–22].

5 Discussion of the Results

In order to update the data received from stakeholders regarding the identification of patterns and the accuracy of the estimates provided, it is advisable to keep statistics of performers by such parameters as the accuracy of the performed assessments, experience, task complexity, role in the project, involvement in the execution of critical tasks, contribution to the knowledge base.

Another important issue is the development of an integrated approach to the development of a knowledge-based system in terms of knowledge utilization.

Handzic and Durmic [31] distinguish four generations of knowledge management models:

1. technocratic. Technocratic models focus on formalized knowledge bases with an emphasis on information and communication technologies. The disadvantages of this approach are to balance research and use of knowledge, to choose the exact content and, accordingly, to make significant efforts to ensure evolutionary development, flexibility and usability;
2. people and organization oriented. Such knowledge management systems view knowledge as a competitive advantage that determines a firm's strategy;
3. context-oriented. As a departure from earlier approaches, the use of context-oriented systems implies that the effectiveness of knowledge management methods is determined by the context of knowledge use;
4. integrated. Integrated models acknowledge the evolutionary and contextual nature of knowledge management, in which knowledge management is considered both socially and technologically.

The proposed knowledge-based IT RMM integrates technocratic aspects (knowledge base), approach to knowledge as a competitive advantage and context-oriented (RM). The model also allows to take into account the dynamic and incremental nature of the implementation of IT projects, whereby the requirements for the information product and the ways of their implementation are refined taking into account new information and experience gained [16, 18, 19].

Therefore, in order to implement the proposed methodological provisions, a model was developed that, unlike the existing ones, takes into account the features of IT PRM (reserves, timely execution, communication, knowledge accumulation, software development methodologies) and allows solving the following problematic issues:

- the probability of successful execution of an IT project in terms of terms and budget is determined by calculating the probability of the project being in the desired Markov state through a set number of steps;
- the number of mark states corresponds to the number of intervals of the project buffer usage indicator and the number of steps to the number of sprints required to complete the project;
- the model takes into account the features of cascading (project buffer formation) and flexible (using sprints) methodologies used in software development [11, 17];

In order to improve information support for decision-making, a knowledge-based RMM has been developed in the project, which takes into account the dynamic and incremental nature of IT project implementation.

In order to put this model into practice, it is advisable to perform modeling of IT PRM processes based on data on real IT projects.

6 Conclusions

In order to implement the proposed methodological provisions, a model was developed that, unlike the existing ones, takes into account the features of IT PRM (reserves, timely execution, communication, knowledge accumulation, software development methodologies) and allows solving the following problematic issues:

- the probability of successful execution of an IT project in terms of terms and budget is determined by calculating the probability of the project being in the desired Markov state through a set number of steps;
- the number of mark states corresponds to the number of intervals of the project buffer usage indicator and the number of steps to the number of sprints required to complete the project;
- the model takes into account the features of cascading (project buffer formation) and flexible (using sprints) software development methodologies;
- in order to improve information support for decision-making, a knowledge-based RMM has been developed in the project, which takes into account the dynamic and incremental nature of IT project implementation.

References

1. Kuznetsov, A., Kavun, S., Smirnov, O., Babenko, V., Nakisko, O., & Kuznetsova, K. (2019). Malware correlation monitoring in computer networks of promising smart grids. In *2019 IEEE 6th international conference on energy smart systems, ESS 2019 – Proceedings* (Vol. 8764228, pp. 347–352). https://doi.org/10.1109/ESS.2019.8764228
2. Ramazanov, S., Antoshkina, L., Babenko, V., & Akhmedov, R. (2019). Integrated model of stochastic dynamics for control of a socio-ecological-oriented innovation economy. *Periodicals of Engineering and Natural Sciences, 7*(2), 763–773. https://doi.org/10.21533/pen.v7i2.557
3. Kaminsky, A. B. (2006). *Economic and mathematical modeling of financial risks*. Monograph, Publishing and Printing Center "Kyiv University".
4. Galicin, V. K., Suslov, O. P., & Galicina, O. V. (2016). Modeling the mechanism of management of organizational projects. *Business Inform, 2*, 160–164.
5. Teslenko, P. A. (2010). Evolutionary theory and synergetics in project management. *Project Management and Production Development, 4*(36).
6. Danchenko, A. B. (2014). Modern approaches to deviation management in projects. *Management of Complex Systems Development, 19*.
7. Rach, D. V. (2013). Management of uncertainty and risks in the project: Terminological basis. *Project Management and Production Development, 3*(47).
8. Sviridova, S. S. (2011). Risk management of the innovation project lifecycle. *Economics: Realities of Time, 1*, 103–108.
9. Skopenko, N. S., Evseeva, I. V., & Moskalenko, V. O. (2013). Risk management in project management. *Investments: Practice and Experience, 24*, 41–44.
10. Archibald, R. (2010) Managing high-tech programs and projects (Trans. with English, 3rd ed., revised. and ext. ITT Company). DMK Press.
11. Boehm, B. A. (2019). *Prioritized top-ten list of software risk items*. http://users.humboldt.edu/smtuttle/s12cs435/435lect06-2/435lect06-2-boehm-top10-risks-to-post.pdf

12. Schwalbe, K. (2012). *Information technology project management*. Cengage Learning.
13. Demarco, T., & Lister, T. (2003). *Waltzing with bears: Managing risk on software projects*. Dorset House.
14. Goldratt Eliyahu, M. (1997). *Critical chain*. The North River Press-Publishing Corporation.
15. Leach, L. (2010). *On time and within budget: Critical-chain project management/Lawrence Leach*. Alpina Publishers.
16. Machac, J., & Steiner, F. (2014). Risk management in early product lifecycle phases. *International Review of Management and Business Research, 2*(3), 1151–1162.
17. Choetkiertikul, M., Dam, H. K., & Tran, T. (2015). Predicting delays in software projects using networked classification (T). In *Automated software engineering (ASE), 30th IEEE/ACM international conference* (pp. 353–364).
18. Jeon, C., Kim, N., & Peter, H. (2015). Probabilistic approach to predicting risk in software projects using software repository data. *International Journal of Software Engineering and Knowledge Engineering, 6*(25), 1017–1032.
19. Kumar, C., & Yadav, D. K. (2015). A probabilistic software risk assessment and estimation model for software projects. *Procedia Computer Science, 54*, 335–361.
20. Teslya, Y. M., & Kubiavka, L. B. (2014). The concept of construction and function of the project risk management system in informatization programs. *Management of Development of Complex Systems, 19*, 93–97.
21. Babenko, V., Lomovskykh, L., Oriekhova, A., Korchynska, L., Krutko, M., & Koniaieva, Y. (2019). Features of methods and models in risk management of IT projects. *Periodicals of Engineering and Natural Sciences, 7*(2), 629–636. https://doi.org/10.21533/pen.v7i2.558
22. Babenko, V. A. (2013). Formation of economic-mathematical model for process dynamics of innovative technologies management at agroindustrial enterprises. *Actual Problems of Economics, 139*(1), 182–186.
23. Shorikov, A. F., & Babenko, V. A. (2014). Optimization of assured result in dynamical model of management of innovation process in the enterprise of agricultural production complex. *Economy of Region, 1*(37), 196–202. https://doi.org/10.17059/2014-1-18
24. Rishnyak, I. V. (2011). Risk management system for IT projects. In *Information systems and networks. Collection of scientific papers* (pp. 250–259). Publisher of Lviv Polytechnic National University.
25. Rishnyak, I. V. (2010). Application of imitation models for risk management of IT projects. In *Information systems and networks. Collection of scientific papers* (pp. 171–181). Publisher of National University "Lviv Polytechnic".
26. Melnik, G. V. (2013). Modeling the information risk management system in the corporate information system. *Business Inform, 9*, 95–99.
27. Kolesnikova, E. V. (2013). Modeling of poorly structured systems of project management. *Proceedings of the Odessa Polytechnic University, 3*, 127–131.
28. Onishchenko, I. I. (2016). Cognitive modeling as a method of qualitative risk analysis of IT projects. *Bulletin of the National Technical University of KhPI. Series: Strategic Management, Portfolio, Program and Project Management, 2*, 88–81.
29. The Kanban Way. *Speed up project delivery using Critical Chain*. http://www.kanbanway.com/speed-up-project-delivery-using-critical-chain#.VdrP4_ntmko
30. Renard, L. *Essential frameworks and methodologies to maximize the value of IT*. http://www.isaca.org/Journal/archives/2016/volume-2/Pages/essential-frameworks-and-methodologies-to-maximize-the-value-of-it.aspx,last
31. Handzic, M., & Durmic, N. (2015). Knowledge management, intellectual capital and project management: Connecting the dots. *Electronic Journal of Knowledge Management, 1*(1), 51–61.

Traditional and Alternative Assets in Portfolio Management: ETF Using Approach

Andrii Kaminskyi, Maryna Nehrey, and Denys Butylo

Abstract The fast-growing development of ETFs provides new opportunities for portfolio management. We have realized portfolio optimization procedures for traditional and alternative asset classes based on ETFs. There was grounded for analysis sample of the ETFs corresponding seven indices of traditional asset classes and seven indices of alternative asset classes. Alternative asset classes were expanded by cryptocurrency index CRIX. Portfolio creation was based on optimization problems involving coherent risk measures: measure of Fischer with different coefficients of risk attitude and CVaR. The analysis of obtained results is presented. The investigation illustrates the effectiveness of using ETFs for involving in portfolio construction procedures on the "asset classes" level. A combination in portfolio traditional and alternative assets shows the risk-reducing effect. Numerical results indicate that alternative assets "replace" stocks in optimal portfolios. That is combinations of bonds and alternatives estimated as optimal portfolios. Results also point to the vital importance of risk attitude for portfolio creation. Cryptocurrencies index illustrates high risk, high return and low correlation with other considered asset classes. Verification results have been also considered.

Keywords Exchange trade funds · Portfolio management · Alternative asset classes · Traditional asset classes · Risk measuring · Cryptocurrencies

1 Introduction

Financial investments have been become enriched by the new tool at the average 1990th. It was Exchange Traded Fund (ETF) construction which realizes the fruitful investment idea of tracking indices. This tool integrated several very important

A. Kaminskyi (✉) · D. Butylo
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

M. Nehrey
National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_4

components, which caused a real interest for investors and turned it into the fastest-growing segment of the asset management universe (an overview is presented, for example, in [1]).

Gaining momentum (increasing capitalization, which currently exceeds \$5.64 Tn [2], of the \$4 Tn in the US market [3]), this tool is becoming more and more in the focus of portfolio management. As large portfolio players as small portfolio investors have a tool with which they can construct new strategies of both passive and active types.

In our opinion, using strategies in portfolio management, which include combinations of ETFs of traditional and alternative asset classes, is very interesting. Alternative asset classes illustrate the dramatic growth of recent decades—according to PWC estimates [4] their capitalization will exceed seven trillion in 2021. Using alternative assets in combination with traditional assets creates new opportunities for investment managers and servicers. ETFs provide good possibilities for the effective implementation of such new portfolio strategies.

The purpose of this article is an empirical analysis of portfolio strategies that involve using ETFs, both traditional asset classes and alternative ones. We also consider the inclusion into alternative assets cryptocurrency index CRIX [5] in the sample as an innovative element. In our opinion, in recent years, the capitalization of cryptocurrencies has increased significantly and their use as an alternative asset is very interesting.[1] At the same time, there is no ETF for this index yet, therefore we used for CRIX real values of this index.

Our analysis involved a special selection of ETFs tracking for indices of assets of both types. This selection forms the base for portfolio creation at the "asset classes level." Proposed in this article approach of portfolio construction based on minimizing a chosen risk measure. We focus our consideration for applying risk measures with coherency property [6]. The coherency of risk measure is an important condition for the correctness of optimization from allocation of capital point of view. First of all, because indices return, especially for alternative asset classes, do not correspond to hypothesis of normality. For realizing our purpose, we have considered the application two such coherent risk measures—Conditional Value-at-Risk (CVaR)and Fischer's risk measure. At the same time, a little more attention is paid to the latter, because it includes a coefficient that reflects a risk attitude. Goal functions of considered optimization problems based on these risk measures. We provide the empirical calculation of optimal portfolio weights on the base of real data of values corresponding to ETFs. The criterion of verification of the resulting portfolios was estimation risk measures and of constructed portfolios an independent sample of data. As addition we have estimated some risk-return indicators.

[1] Cryptocurrency ETFs are already trading in some countries, but so far regulators in the U.S. have denied multiple attempts of entering on stock exchanges.

The basic result is based on the analysis of three types of strategies: (1) portfolios constructed only from traditional assets, (2) only from alternative assets, (3) portfolios constructed on the base of a combination of traditional and alternative assets. The results show some characteristic features of these portfolio strategies. In particular, alternative assets expressed by ETFs for alternative assets show risk which is comparable to the risk of ETF tracking classical stock indices. Their weak correlation with most traditional assets and among themselves gives rise to an additional effect of reducing risk and, in part, profitability. It is interesting that the basis of the combined portfolio is bonds and alternative assets, which we consider as a main empirical result. The cryptocurrency index shows an almost zero level of correlation with both traditional and alternative assets. There is also a high level of expected profitability and risk, which determines results of involving it in the portfolio structure.

Such results may be applied for the portfolio construction of passive investments.

This paper is organized as follows. In Chap. 2, we provide literature review about basic scientific spheres areas presented in research. It includes recent publications devoted to ETF, alternative investments, coherence risk measures and estimation performance of portfolios. In Chap. 3, we present our methodology. Proposed methodology involves consideration of ETF as tools for analysis risk and return of different asset classes. This chapter includes description of our forming sample of traditional and alternative asset classes and data sources for calculations. Also, we explain risk measures which are used as goal functions in using optimization problems. In Chap. 4, we present results of our research. Results include optimal portfolio structures and risk-return correspondences for created portfolios. Results of verification procedures also considered in this chapter. This chapter accumulates conclusions from our research and possible ways for their development.

2 Literature Review

Key trends and discussing ideas about using ETF in modern investment fund management are presented in [7]. Moreover, different ideas for apply ETF in portfolio forming are tackled there in sufficient depth. One great advantage of this book is description ETFs as for indices of traditional assets as for alternatives.

It must be noted book [8], which focuses on implementation ETFs into investment strategies. The crucial benefit of this book is dealing with practical cases of applying ETF.

The authors of [9] are discussing the ETFs portfolio selection problem using different performance measures and propose some new performance measure that is consistent with the choices of non-satiable risk-averse investors. The authors also consider CVaR application for the measurement risk of a portfolio from ETFs.

The performance of ETFs is analyzing in [10]. The essence of alternative assets, their characteristics and the problematic of using at investment management very widely present in recent publications. This is a consequence of the faster growth of this segment and the potentials of their using. Overview of alternative investments for institutional asset allocators and construction portfolios containing both traditional and alternative assets is present in the book [11]. Authors analyzing alternative assets, alternative strategies, and alternative portfolio management. At the 17 chapter of this book authors are grounded enhanced return and reduced risk achieved by combining alternatives with traditional assets.

The impact of cryptocurrencies (bitcoin first of all) on portfolio construction is illustrated in [12].

Risk measurement and portfolio optimization are topics which deeply analyzing in [13]. The problematic of coherency is raised from the ideas presenting in a fundamental article [14]. The development and application of coherency conception are presenting in overview [6]. The nature of coherent risk measures which we use in our research—CVaR and Fischer measures are described in articles [15, 16]. The logic of empirical comparison of performance portfolios constructed by different risk measures presents in [17–19].

3 Methodology and Data

3.1 Methodology of ETF Sample Forming

The first cornerstone of our research methodology is the issue of identifying alternative asset classes. For our study, this is necessary to generate an ETFs representative sample. Issues related to the identification of alternative asset classes are conditioned by the lack of a well-defined alternative assets concept in modern investment literature. This is due to the fact that alternative assets are dynamically developing and their use depends on the type of investor. Most often, scientists use a methodological approach to their definition through traditional assets. That is, alternative investment assets include those that cannot be attributed to traditional asset classes. The advantage of this approach is rather a clear definition of the set of traditional assets. Traditional assets consist of assets of two main types: stocks, bonds (sometimes add a third class which is bank certificates of deposit). In reality, this approach to the definition is quite broad ("all accept stocks and bonds") and it is difficult to introduce the classification of alternative assets.

As another methodological approach, one may consider an approach based on the identification of characteristic features inherent in a particular type of asset to attribute particular assets to the class of alternative. The main feature highlighted by most scholars is the low or negative correlation with traditional investment assets and the overall stock market situation. Simplified, the impact of cyclicality on the economy as well as adverse economic factors on alternative assets is much smaller than on traditional ones. This feature is widely used by investors to diversify their portfolio by including alternative assets. By the way, we observed this feature in our study as well. Another difference between alternative assets and traditional assets is the economic nature of return generation. Very often return of alternatives is difficult to connect with a particular asset, as is the case with stocks or bonds (good examples are an investment in antiques or collection of wines). However, it is worth noting that some of the alternative assets are still closely related to tangible assets. In addition, often alternative assets are often characterized by low liquidity and some of them are speculative.

In our study, we have used the following approach to generate a sample for the study. We have structured alternative assets into several categories. The first is alternative assets related to physical assets—goods, real estate, infrastructure, and others. In particular, this category includes forward and futures contracts for commodities such as oil, metals, gas, agricultural products. Their popularity with investors is related to the significant role of these commodities in the world economy.

In the second category, we included alternative assets related to investments in hedge funds, private equity (stocks and bonds of companies not traded in the stock market). Venture capital investments and bad debt have also become widespread.

The third category of alternative assets is other types of assets that are not in the first two. These are primarily structured products and financial derivatives. The most classic example of this category of investment assets is secured debt, a typical example of which are credit default swaps (CDS). In addition, currency, index and other swaps are also included in this category of alternative assets.

We review cryptocurrencies separately as an alternative asset class. As a fourth class. Cryptocurrencies are becoming entrenched in investing and have not a correlation with any other assets. Of course, it may be regulative problems involving cryptocurrencies into investment portfolios of institutional investors. But consideration such investment is very promising for portfolio forming.

We have structured traditional assets into stocks and bonds. More detail we structured stock indices for indices of large, small and medium-sized US companies. They were selected for sample in the stock aspect. The bonds were structured according to government, corporative and inflation-indexed government bonds.

Based on this, we formed the following ETFs sample.

- SPDR's SPY correspond to the index Standard & Poor's 500 and is the oldest and largest of all ETFs;
- OEF. The iShares S&P 100 ETF seeks to track the investment results of an index composed of 100 large-capitalization U.S. equities;
- SPDR's MDY that tracks the Standard & Poor's 400 to model the mid-cap equities, while being smaller than iShares IJH it has about the same turnover but offers a longer time series;
- iShares IJR follow the small-cap companies, it tracks the Standard & Poor's 600 indexes and is much larger and liquid than the corresponding SPDR fund SLY;
- iShares IEF retrace a balanced portfolio of Treasury bonds, the choice of this particular government bond fund is motivated by its duration 7.6 years that is comparable to the duration of other bond funds analyzed in this paper;

- iShares LQD correspond to a balanced portfolio of investment-grade corporate bonds, it's one of the oldest bond ETFs and its duration (8.5 years) is approximately the same as for the IEF fund mentioned above, so we can contrast government and corporate bonds;
- iShares TIP tracks inflation-linked bonds, an asset class that should have quite a distinct characteristic, however, its duration (7.6 years) aligned to LQD and IEF;
- IQ Hedge's QAI follow-up the investments in hedge funds, as QAI is the largest of the ETFs that correspond to a blend of different investment styles, such as long/short equity, global macro, market neutral, event-driven, fixed income arbitrage, and emerging market;
- Vanguard's VNQ tracks a market-cap-weighted index of companies involved in the ownership and operation of real estate in the United States;
- SPDR's GLD trace the investments into gold, as the fund holds only gold bullion;
- Investco's PSP correspond to private equity investments, as it's one of a few funds that invest in a global portfolio of private equity firms and has the highest asset base of those;
- USCI (United States Commodity Index Fund) is an exchange-traded security that is designed to track the price movements of the Summer Haven Dynamic Commodity Index Total Return;
- WYDE is an actively managed fund, primarily invests in credit default swap (CDS) indexes to express bearish views on the overall credit quality of the high-yield market in North America;
- SPDR's GLD retrace an investments into gold, as the fund holds only gold bullion;
- WTMF tracks an index that takes long and short positions in currency, commodity, and Treasury futures;
- CRIX. This is a cryptocurrency index. The index includes basic cryptocurrencies which are weighted with its market capitalization. There are absent ETF corresponds to this index. After that, we use the real values of this index in our research.

3.2 Methodology of Data Sample Choosing

We have used data from resource Investing.com which contains information about all considered ETF. Data for CRIX was used from site [5]. The time period which was considered for modeling is January 2016–December 2018. Daily returns of ETFs were included for analysis. The verification period includes January 2019–July 2019.

3.3 Methodology of Optimization Procedures

Portfolio constructions are based on consideration of optimization problems, which include identification variables, goal function and constrains. Variables are present parts of capital investing in ETF from described above sample. Constraints are usual which include non-negatives conditions and equality to 100% total amount. We consider some risk measures as goal function. As a rule, such goal function corresponds to minimization criteria. Of course, it is possible to introduce a number of criteria that form multi-objective optimization (adding expected return, liquidity and so on). We use in our research one-criteria optimization: risk measure minimization. There are different risk measures that can be used for portfolio risk optimization. Classically it is possible to use standard deviation as risk measure which traces back to H. Markowitz's approach. Other well-known approach involves the minimization of VaR (Value-at-Risk). The problem of applying these measures and many others is absence properties, which adequate to the logic of capital allocation. These properties called now "Coherency properties" [6, 14].

Let us consider X as random variable of asset return (return of ETF in our case) or portfolios created from such assets. The coherent risk measure is mapping ρ from set of X to R (real numbers) with the following properties: monotonicity, subadditivity, positive homogeneity, and translation invariance.

Monotonicity: For all $X1$, $X2$, it is supposed that $\rho(X1) \geq \rho(X2)$, if $X1 \leq X2$. Economically monotonicity means that a position $X2$ with a higher value than position $X1$ has a lower risk.

Subadditivity: For all $X1$, $X2$ it is supposed inequality $\rho(X1 + X2) \leq \rho(X1) + \rho(X2)$.

Economically this means diversification effect, which is the cornerstone of portfolio creation.

Positive Homogeneity: For all X and for all real numbers $\lambda \geq 0$ should be $\rho(\lambda X) = \lambda\rho(X)$.

Positive Homogeneity ensures that the risk of a position depends linearly on the size of the position.

Translation Invariance: For all X and for all real numbers a should be $\rho(X + a) = \rho(X)$—a. This property states that risk free asset (with constant return) to a position reduces the risk by the same amount.

There are two basic measures that have coherency property. First is Fisher measure [15]:

$$\rho_{F,a}(X) = -E(X) + a\sigma^{-}(X),$$

where $\sigma^{-}(X)$—lower semi-standard deviation, $a \in [0; 1]$ is a coefficient which presents "weight" of $\sigma^{-}(X)$. We interpret a as risk attitude coefficient. If this coefficient close to 0 it is interpreted as risk seeking attitude. If a close to 1 it is interpreted as risk averse attitude.

The second is CVaR [16]:

$$CVaR_{\alpha} = E\,(X|\,X < VaR_{\alpha}(X))$$

where $VaR_{\alpha}(X)$ (Value-at-Risk) measures the maximum potential loss over a holding period at a given confidence level α (typically 0.95–0.99).

We apply in our research these two coherent risk measures.

4 Results: Portfolio Optimization and Verification

The logic of empirical research was structured on portfolio optimization and verification procedures. In turn we divided portfolio optimization into three parts. First part includes portfolio optimization for whole sample. Second part surrounds exclusively traditional assets and third part focuses on portfolio optimization of alternative assets. In our view such approach reinforces results understanding by comparative analysis. We consider verification as comparison of estimations different risk and return indicators for constructed portfolios at another time interval. It is not supposed reforming portfolios on the new data.

First direction of our empirical research was concerning portfolio optimization with Fisher measure at the place of goal function. The distinguishing features of this measure are combination expected return and risk estimation together with risk attitude. The structure of the optimal portfolios for the whole sample of ETFs is presented in Table 1. It can be found portfolios for different risk attitude coefficient a.

We found results are interesting by following issues. First and foremost, we can see "replace" stocks for alternative asset classes in optimal portfolios. This can be viewed for moderate levels coefficient a and for strong risk aversion (coefficient close to 1, which mean higher "weight" of semi-standard deviation. Clue of this can be found in correlation analysis. The average correlation through SPY, OEF and MDY is 0.87. The average correlation through alternative asset classes is 0.07. So, alternative asset classes provide good possibilities for diversification effect. Second interesting result concerns portfolio with 0 or very small coefficient a. At this case optimal portfolio includes one asset—cryptocurrency (CRIX). It can be explained by high expected return, high risk of this asset but low weight of semi-standard deviation when coefficient close to 0.

Results of verification procedures are presented in the Table 2. Logic of verification tied with preservation corresponding for risk-return estimations in different time frames. It is interesting, that for all indicated parameters except ER/STD correlation coefficient (through values coefficient a) between calculated and verification values very close to 1 (~0.99). Our point of view for this is similarity. For ER/STD verification (expressed by correlation) is -0.34. But, if we exclude cases $a = 0$ and $a = 0.1$ (where optimal portfolio includes 100% CRIX) we will receive correlation 0.97. This means that ER/STD can be changed significantly

Table 1 Optimal portfolios for traditional and alternative asset classes with Fischer measure in the place of goal function

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Coefficient a	← Risk seeking					Risk averse →					
SPY	0	0	0.094	0.034	0.014	0.005	0.000	0.000	0.000	0.000	0.000
OEF	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MDY	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
IJR	0	0	0.003	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
IEF	0	0	0.245	0.229	0.218	0.216	0.216	0.215	0.215	0.214	0.214
LQD	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TIP	0	0	0.331	0.304	0.284	0.279	0.277	0.279	0.278	0.278	0.277
QAI	0	0	0.157	0.237	0.264	0.274	0.280	0.277	0.275	0.273	0.272
VNQ	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PSP	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
USCI	0	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WYDE	0	0	0.049	0.109	0.115	0.123	0.127	0.131	0.134	0.136	0.138
GLD	0	0	0.082	0.032	0.017	0.009	0.004	0.002	0.002	0.000	0.000
WTMF	0	0	0.007	0.033	0.071	0.078	0.082	0.084	0.085	0.086	0.087
CRIX	1	1	0.032	0.016	0.011	0.009	0.008	0.007	0.006	0.005	0.005
Structure of optimal portfolios by asset classes											
Stocks	0.00%	0.00%	9.75%	4.06%	1.97%	1.13%	0.61%	0.62%	0.63%	0.64%	0.64%
Bonds	0.00%	0.00%	57.61%	53.27%	50.20%	49.56%	49.29%	49.40%	49.23%	49.29%	49.14%
Alternative	100%	100%	32.64%	42.67%	47.83%	49.31%	50.10%	49.98%	50.14%	50.07%	50.22%

Table 2 Results of verification for optimal portfolios traditional and alternative asset classes

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Coefficient *a*	← Risk seeking					Risk averse →					
Risk-return characteristics. Estimations based on basic data											
Expected return	0.434	0.434	0.017	0.006	0.003	0.001	0.000	0.000	−0.001	−0.001	−0.001
Fischer measure	−0.434	−0.095	0.016	0.029	0.040	0.050	0.060	0.070	0.080	0.090	0.100
SEMI-STD-	3391	3391	0.164	0.116	0.106	0.102	0.100	0.099	0.099	0.099	0.098
STD	4750	4750	0.228	0.159	0.144	0.139	0.137	0.135	0.135	0.134	0.134
ER/STD	0.091	0.091	0.076	0.039	0.019	0.008	0.001	−0.004	−0.007	−0.009	−0.011
Risk-return characteristics. Estimations based on verification data											
Expected return	0.532	0.532	0.052	0.034	0.027	0.025	0.023	0.022	0.022	0.021	0.021
Fischer measure	−0.532	−0.257	−0.026	−0.005	0.006	0.015	0.023	0.031	0.039	0.046	0.054
SEMI-STD-	2746	2746	0.133	0.094	0.083	0.079	0.077	0.076	0.075	0.075	0.074
STD	3985	3985	0.196	0.137	0.120	0.114	0.111	0.109	0.108	0.108	0.107
ER/STD	0.133	0.133	0.266	0.247	0.230	0.218	0.209	0.203	0.199	0.196	0.193

for CRIX through the time. This asset in some extent “derailed on” correspondence between risk and return.

Table 3 presents results of applying optimization procedures exclusively for traditional stocks. The main specificity is dominating bonds in optimal portfolios. Moreover, approximately 55% weights cover ETF which correspond to inflation protected bonds. Verification analysis present in the Table 4. Corresponding is present if we exclude case $a = 0$. It can be explained the fact that in this case optimal portfolio was found for maximum return, without considering lower semi-standard variation.

Tables 5 and 6 illustrate applying portfolio optimizing procedures for alternative asset classes. Numerical results illustrate that 3 ETFs are not included into the structure of optimal portfolios under all values of coefficient a.

Optimization which based on risk measure CVaR provides results which present at the Table 7. It is interesting that lower risk (higher value of CVaR) presents at the optimal portfolio constructed on the base traditional and alternative assets. But verification procedures show that optimal portfolio based on alternative assets has better CVaR. The explanation of this bases on potential high frequency of changing directions of evolution returns.

5 Conclusions

Carried out research indicates that adding alternatives to portfolio construction essentially changes its optimal structure. First of all, optimal portfolio consists mainly from bonds and alternatives. So, alternatives in some extent “replace” stocks in the portfolio structure. This effect we have indicated by application both risk measures: Fischer measure and CVaR. The explanation of this arises from low correlation of returns as among alternatives themselves as with traditional classes. Second, cryptocurrency index CRIX which considered in this paper as specific alternative asset class indicates high return and high risk. This class is dominated in portfolios constructed by risk seeking approach. Returns of this index also have zero correlation as with traditional as with alternative asset classes. Taking into account frequent changing relations between risk and return for cryptocurrencies this class a little bit spoils the verification of our results. But after exclusion cases where CRIX is dominated (risk seeking situation with close to 0 coefficient in Fischer risk measure) verification is good and illustrates similarity. Third, not all alternative asset classes are presenting in optimal portfolios. ETFs, which present hedge funds and funds based on investing in CDS indexes, take up a lot of portfolio structure.

We think that approach proposed in our paper has good potential for development. The development may be realized by following ways. First way consists in more deeply analyzing different types of ETFs. It is bear in mind that ETFs tied now with wide variety of indexes. At the beginning ETFs were almost exclusively seeking to track broad value-weighted equity indices (for example, the S&P 500), but ETFs today track to equities, bonds, broad-market, currencies, sectors, and other

Table 3 Optimal portfolios for traditional asset classes with Fischer measure in the place of goal function

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Coefficient a	← Risk seeking					Risk averse →					
SPY	0	0.149	0.066	0.048	0.042	0.018	0.011	0.000	0.000	0.000	0.000
OEF	1	0.043	0.055	0.054	0.051	0.068	0.072	0.080	0.078	0.076	0.075
MDY	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
IJR	0	0.000	0.004	0.006	0.007	0.007	0.008	0.008	0.008	0.009	0.009
IEF	0	0.241	0.317	0.336	0.344	0.350	0.353	0.356	0.358	0.359	0.360
LQD	0	0.000	0.000	0.001	0.001	0.001	0.001	0.007	0.007	0.007	0.007
TIP	0	0.567	0.557	0.555	0.555	0.555	0.555	0.549	0.549	0.549	0.550
Portfolios structure by classes											
Stocks	100%	19.2%	12.58%	10.84%	9.99%	9.44%	9.09%	8.77%	8.6%	8.46%	8.35%
Bonds	0.00%	80.8%	87.42%	89.16%	90.01%	90.56%	90.91%	91.23%	91.4%	91.54%	91.65%

Table 4 Results of verification for optimal portfolios traditional asset classes

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Coefficient *a*	← Risk seeking					Risk averse →					
Risk-return characteristics. Estimations based on basic data											
Expected return	0.030	0.005	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001
Fischer measure	−0.030	0.010	0.023	0.036	0.048	0.061	0.073	0.086	0.098	0.111	0.123
SEMI-STD-	0.604	0.147	0.129	0.126	0.125	0.125	0.125	0.124	0.124	0.124	0.124
STD	0.805	0.203	0.178	0.174	0.173	0.172	0.171	0.171	0.171	0.171	0.171
ER/STD	0.037	0.025	0.015	0.011	0.009	0.008	0.008	0.007	0.007	0.006	0.006
	Risk-return characteristics. Estimations based on verification data										
Expected return	0.098	0.044	0.039	0.038	0.038	0.037	0.037	0.037	0.037	0.037	0.036
Fischer measure	−0.098	−0.031	−0.014	−0.001	0.012	0.026	0.038	0.051	0.064	0.077	0.090
SEMI-STD-	0.515	0.129	0.124	0.125	0.125	0.125	0.125	0.126	0.126	0.126	0.126
STD	0.723	0.195	0.183	0.182	0.182	0.182	0.182	0.182	0.183	0.183	0.183
ER/STD	0.135	0.226	0.215	0.210	0.206	0.204	0.202	0.201	0.200	0.200	0.199

Table 5 Optimal portfolios for alternative asset classes with Fischer measure in the place of goal function

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Coefficient *a*	← Risk seeking					Risk averse →					
QAI	0.000	0.000	0.623	0.574	0.560	0.553	0.549	0.546	0.544	0.542	0.540
VNQ	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PSP	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
USCI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WYDE	0.000	0.000	0.105	0.173	0.193	0.203	0.210	0.215	0.218	0.220	0.222
GLD	0.000	0.000	0.142	0.095	0.082	0.074	0.070	0.067	0.065	0.063	0.062
WTMF	0.000	0.000	0.092	0.138	0.151	0.158	0.162	0.164	0.166	0.168	0.169
CRIX	1000	1000	0.037	0.020	0.015	0.012	0.010	0.009	0.008	0.007	0.006

Table 6 Results of verification for optimal portfolios alternative asset classes

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Coefficient *a*	← Risk seeking					Risk averse →					
Risk-return characteristics. Estimations based on basic data											
Expected return	0.434	0.434	0.018	0.008	0.004	0.003	0.002	0.001	0.000	0.000	0.000
Fischer measure	−0.434	−0.095	0.020	0.036	0.050	0.063	0.076	0.089	0.102	0.115	0.127
SEMI-STD-	3391	3391	0.189	0.145	0.136	0.132	0.130	0.128	0.128	0.127	0.127
STD	4750	4750	0.264	0.200	0.187	0.181	0.178	0.177	0.176	0.175	0.174
ER/STD	0.091	0.091	0.068	0.040	0.021	0.017	0.011	0.006	0.000	0.000	0.000
Risk-return characteristics. Estimations based on verification data											
Expected return	0.532	0.532	0.041	0.025	0.021	0.018	0.016	0.015	0.014	0.014	0.013
Fischer measure	−0.532	−0.257	−0.016	0.000	0.010	0.017	0.025	0.032	0.038	0.045	0.052
SEMI-STD-	2746	2746	0.127	0.085	0.075	0.070	0.068	0.067	0.066	0.065	0.065
STD	3985	3985	0.194	0.128	0.111	0.103	0.099	0.096	0.094	0.093	0.092
ER/STD	0.133	0.133	0.213	0.198	0.185	0.173	0.164	0.158	0.152	0.147	0.143

Table 7 Portfolios weights for optimization problem with CVaR in the place of goal function

	Traditional assets	Traditional and alternative assets	Alternative assets
SPY	0.058	0.022	0.000
OEF	0.000	0.000	0.000
MDY	0.022	0.011	0.000
IJR	0.003	0.005	0.000
IEF	0.419	0.253	0.000
LQD	0.000	0.000	0.000
TIP	0.497	0.318	0.000
QAI	0.000	0.155	0.536
VNQ	0.000	0.000	0.000
PSP	0.000	0.000	0.000
USCI	0.000	0.000	0.000
WYDE	0.000	0.182	0.215
GLD	0.000	0.001	0.050
WTMF	0.000	0.053	0.194
CRIX	0.000	0.000	0.004
CVaR for optimal portfolios. Estimations based on basic data			
	−0.395	−0.318	−0.401
CVaR for optimal portfolios. Estimations based on verification data			
	−0.35053	−0.21827	-0.18122

indices. Now it is possible to involve ETFs corresponded to indices from different countries and regions. These indices include both developed and emerging markets across Asia, Europe and Latin America. It is more adequate to approach based on global portfolio optimization.

Second way of development is corresponding to extending alternatives asset classes. Typical example is creation of ETF for cryptocurrencies [20]. Broadening of alternative asset classes which presented by ETFs will facilitate the development portfolio optimization practice. Of course this arise risks which connected with using such asset classes.

Third way reveals oneself in building more complex portfolio optimization problems. Such problems should include formalization criteria, which more adequate to investor goals. As example, actual criteria for using ETFs in portfolio construction is liquidity maximization. Because many ETFs have low capitalization. Other criteria may concern more specific goal functions.

References

1. Lettau, M., & Madhavan, A. (2018). Exchange-traded funds 101 for economists. *Journal of Economic Perspectives, 32*(1), 135–154.
2. *ETFGI reports assets invested in the Global ETF and ETP industry reached a record US$5.64 trillion at the end of June 2019*. https://etfgi.com/news/press-releases/2019/07/etfgi-reports-assets-invested-global-etf-and-etp-industry-reached

3. Rosenbluth, T. (2019). *US ETF assets hit new milestone*. https://www.etf.com/sections/blog/us-etf-assets-hit-new-milestone
4. Brashaw, N. (Ed.). (2016). *ETFs: A roadmap to growth* (36 p). PricewaterhouseCoopers.
5. *The Crix*. https://thecrix.de
6. Karabey, U. (2012). Risk measures and risk capital allocation. *İstatistikçiler Dergisi, 5*, 32–42.
7. Stevenson, D., & Tuckwell, D. (2019). *The ultimate ETF guidebook: A comprehensive guide to the world of exchange-traded funds - Including the latest innovations and ideas for ETF Portfolios* (Kindle Edition, 252 p). Harriman House LTD.
8. Ullal, A. (2013). *ETF investment strategies: Best practices from leading experts on constructing a winning ETF portfolio* (256 p). McGraw-Hill Education.
9. Petronio, F., Lando, T., Biglova, A., & Ortobelli, S. (2014). Optimal portfolio performance with exchange-traded funds. *Ekonomická Revue – Central European Review of Economic Issues, 17*, 6–12.
10. Hassine, M., & Roncalli, T. (2013). Measuring performance of exchange-traded funds. *The Journal of Index Investing, 4*(3), 57–85.
11. Chambers, D., Black, K., & Lacey, N. (2018). *Alternative investments: A primer for investment professionals* (p. 183). CFA Institute Research Foundation.
12. Shashwat, G. (2016). Analyzing the effects of adding bitcoin to portfolio. *World Academy of Science, Engineering and Technology International Journal of Economics and Management Engineering, 10*(10), 14 p.
13. Szego, G. (Ed.). (2004). *Risk measures for the 21st century* (521 p). Wiley.
14. Artzner, P., Delbaen, F., Eber, J., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance, 4*, 203–228.
15. Fischer, T. (2003). Risk capital allocation by coherent risk measures based on one-sided moments. *Insurance: Mathematics and Economics, 32*, 135–146.
16. Rockafellar, R., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *The Journal of Risk, 2*(3), 21–41.
17. Malyarets, L. M., Babenko, V. O., Nazarenko, O. V., & Ryzhikova, N. I. (2019). The modeling of multi-criteria assessment activity in enterprise management. *International Journal of Supply Chain Management, 8*(4), 997–1004.
18. Polbennikov, S., & Melenberg, B. (2005). *Mean-coherent risk and mean-variance approaches in portfolio selection: An empirical comparison*. Discussion papers of NETSPAR, 2005–2013 July 2005, 54 p.
19. Ramazanov, S., Antoshkina, L., Babenko, V., & Akhmedov, R. (2019). Integrated model of stochastic dynamics for control of a socio-ecological-oriented innovation economy. *Periodicals of Engineering and Natural Sciences, 7*(2), 763–773.
20. *First #1 crypto ETF*. https://firstcryptoetf.com/

Influence of Management Aspects on the Performance of Energy-Efficient Activities

Valeriia Dykan and Liliia Bilous

Abstract The present chapter discusses the trends in global economies toward energy conservation and energy efficiency. The steps taken by the Ukrainian state in the development and implementation of the SDGs are analyzed, the role and influence of the smart city concept in this process is noted. The analysis of the energy efficiency of the project in the town of Babai, Kharkiv region, Ukraine on the modernization of street lighting is carried out. The dynamics of changes in the volume of electricity consumption by street lamps and the total costs of street lighting in the period from 2015 to 2018 is estimated. Methods are proposed for optimizing energy costs by using alternative sources of energy production.

Keywords Management · Energy efficiency · Alternative energy sources · Energy management · Regional development management

1 Introduction

Global trends of world economies are currently aimed at energy conservation and energy efficiency. This is reflected in the Global Sustainable Development Goals 2016–2030, which were approved in 2015 at the United Nations (UN) Summit on Sustainable Development, and the development strategy of individual states is traced, which were approved at the same summit.

At the initiative of the Government of Ukraine and with the assistance of the UN in Ukraine, an open and inclusive process of adapting sustainable development goals (SDGs) took place throughout the year. Using a wide range of information, statistical and analytical materials, a national SDG system was developed, including 17 development goals, 86 goals and 172 indicators to monitor their implementation [1].

V. Dykan · L. Bilous (✉)
Management and Administration Department, V.N. Karazin Kharkiv National University, Kharkiv, Ukraine
e-mail: valeriya.dykan@karazin.ua; l.bilous@karazin.ua

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_5

These steps are the response of the Ukrainian government to the request of civil society, which, in turn, is developing projects aimed at conserving natural resources, reducing energy consumption, reusing consumer products and recycling. But, without proper management and support of the administrative resource, all these developments cannot be massively implemented.

The development of the smart city concept has been supported for many years by international technology companies and the European Union, as well as national and municipal governments. For example, in Sweden, projects with smart city technologies, such as urban smart grids, have shown the potential to restructure relationships between utilities, energy consumers, and other actors by changing the configuration of the dynamics of energy supply and demand. These projects are not only technological innovations, but also a whole system that needs competent management, in the absence of such management, the expended resources can not only fail to pay off, but also bring much more costs in the process of use than before their implementation.

Despite the popularity of projects related to the exchange of climate data [2], questions arise about whether urban projects can realize their potential. They are implemented in existing cities and must combat the inertia of existing socio-technical systems that go beyond experiment and city [3].

2 Management and Efficiency of Energy Activities in Ukraine

The main internal document regulating the ecological use of resources, as well as achieving the optimal state of the environment by introducing an ecosystem approach to all areas of socio-economic development, is the Law of Ukraine "On the Main Ambush (Strategy) of the State Ecological Policy of Ukraine for the Period until 2030" adopted on February, 28, 2019 and effective from January, 1, 2020. According to the plan for implementing the state environmental policy of Ukraine for the period until 2030 [4], it is planned to introduce modern monitoring systems, market mechanisms to create a green economy, as well as reduce greenhouse gas emissions and emissions from stationary pollution sources by 15% compared to 2015. This document from the date of its introduction terminates the previous law [5].

At the same time, the trends of environmental modernization find their response not only in laws and programs, but also locally in the united communities and the not yet united territorial units. Activities such as the use of more economical and environmentally friendly diode lamps instead of incandescent lamps are common both among foreign countries and among households [6] in Ukraine. Having assessed the experience of foreign colleagues and local households, regional territorial associations and communities are switching to new, more economical

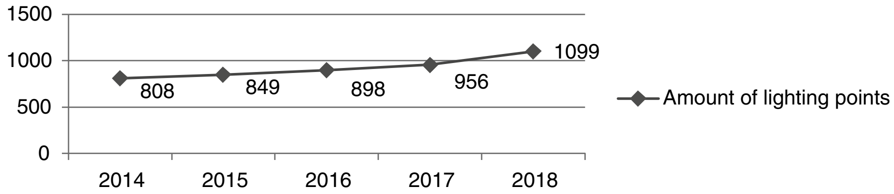


Fig. 1 Dynamics of installation of lighting points in the period from 2014 to 2018

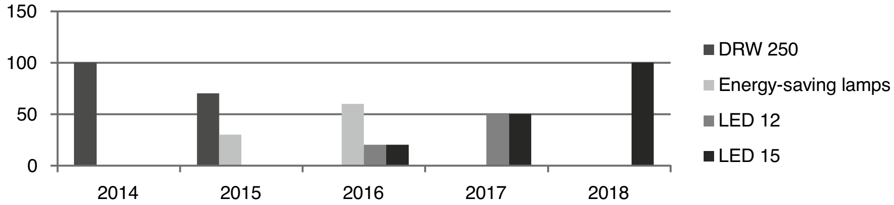


Fig. 2 The dynamics of these changes in the structure of light sources in the village of Babai 2014–2018

lamps, directing the saved money to solving an important problem in the sector—reducing the number of unlit plots.

So, for example, consider the situation in which the town Babai in the Kharkiv region found itself. The composition of the Babaevsky town council includes the town Babai and also the village Zatyshne. The population of these points is 7 thousand people with an area of almost 6 km^2.

Since 2015, the heads of the village council decided to increase the number of street lamps and, as a result, reduce unlit areas in the settlements. In the period from 2015 to 2018, the number of lamps increased by 33%. The dynamics of the installation of lighting points in the period from 2015 to 2018 is shown in Fig. 1.

Adapting the experience of foreign countries since 2015, standard 250 W DRV lamps were initially replaced by energy-saving lamps with 65 W at first, and in the period 2017–2018, they completely switched to LED lamps at first at 10 W and in 2018 at 15 W. The dynamics of these changes is shown in Fig. 2.

Due to such changes, despite the increase in the number of lighting points, electricity consumption in kilowatts did not increase, but rather decreased. So, in 2015, this indicator amounted to 176,897 kW per year, and in 2018 with the amount of 1099–114,963 kW per year, i.e. by 61,934 kW per year, given that the number of lamps increased by 33%.

With a difference in energy consumption of almost 46%, there is no cost savings. Due to a significant increase in the tariff rate and the absence from January 2017 of separate tariffication of day and night lighting, the amount paid for consumed kilowatts has almost doubled. The dynamics of reducing consumption and rising cost of services is shown in Fig. 3.

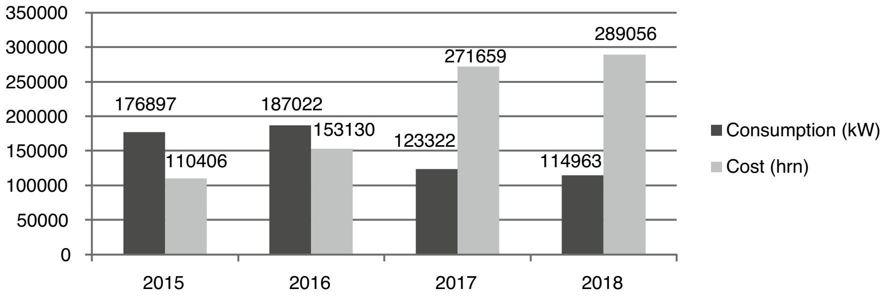


Fig. 3 Dynamics of consumption and cost of street lighting services 2015–2018

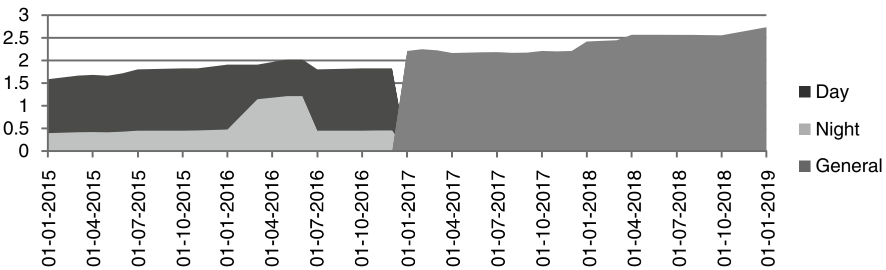


Fig. 4 Dynamics of changes in the cost of electricity

Figure 4 demonstrates that despite the reduction in consumption, due to the state policy and management specificities, the final cost of energy consumed neither remained unchanged nor decreased, but almost doubled. So, in January 2015, a kilowatt hour of energy in the daytime cost 1.58532 hrn, and at night it cost 0.39633 hrn. In December 2018, the cost of a kilowatt of energy, regardless of the time of day, amounted to 2.553312 hrn. The dynamics of the change in the cost of energy is shown in Fig. 4.

This example confirms the thesis put forward in the article "Energy Efficiency Left Behind?" [2], in which the author raised the fear that all activities aimed at energy conservation could be offset by the lack of competent management. In this particular example, tariffs completely destroy the value of all modernizations, although if such modernizations were not carried out, the volume of consumption would be ten times higher, as well as the cost of services.

A similar situation existed in the Chkalovska territorial unit in the Kharkiv region, the total number of street lighting units increased, as did the cost of lighting services, and the total consumption decreased due to energy-efficient measures.

With the development of technology and the availability of several sources of financing, this problem can be solved by the following methods:

1. construction of a waste recycling plant;
2. replacement of all street lighting devices with those capable of self-generating energy;
3. obtaining energy from plants.

The most ambitious project among the proposed ones is the construction of a waste recycling plant near the settlement. Such a construction, obviously, will solve not only the problem of street lighting, but also a number of others. We can consider the positive and negative consequences of the implementation of this project.

So, the positive results include the fact that the generated energy will not only be sufficient to provide street lighting, the converted energy from processing will be able to feed the entire village. This project can be implemented by investor's funds or for grant money. Large foreign funds such as GIZ, EU4Energy and others are already investing in large environmental projects in Ukraine.

In addition, it is possible to take out a loan at low interest rates from international organizations. In Ukraine, international financial institutions are represented mainly by financial institutions of the World Bank Group (IFC, EBRD, EIB, World Bank) and the American Foreign Private Investment Corporation (OPIC). In addition, the Chinese Committee for the Development of Enterprises Abroad (CODA), and the Chinese Development Bank (CDB) became interested in Ukrainian green projects, which are becoming important players in the field of investment in renewable energy projects, the European Bank for Reconstruction and Development (EBRD). Also, public–private partnership (PPP) may be one of the options for implementing such a project [7].

In addition, the issue of waste disposal in the village, the nearest settlements and the city of Kharkiv will be resolved, and as we know, the problem of solid waste disposal is relevant for the whole of Ukraine [8]. The presence of such a plant will create new jobs for residents and provide an inflow of capital to the local budget.

However, the construction of such a structure can cause a number of negative consequences.

Firstly, this project is a long-term one; its construction will take years and a significant amount of financial resources. Secondly, during construction, there is a high probability of violation of construction standards or non-compliance with environmental requirements, which will lead to a significant deterioration of the environmental situation in the village after the commissioning of the plant. In addition, a constant flow of raw materials for the plant from the region, and its storage may degrade the ecology of the village and expose the inhabitants of the town dangers of respiratory diseases. Such a structure should be located at a distance of at least 10 km. from the town and have its own road entrances.

However, even with all the rules and mitigation of risks, the construction of a waste recycling plant near the town Babai requires refinement and coordination activities since it may be associated with protests of the local population. In 2016, the administration of the town council proposed the placement of such a plant 7 km from residential areas, but the towners came out with protests and the construction was never started. In 2018, the construction of a waste recycling plant in the Kharkiv region [9]. At the enterprise, which will be built in the Dergachevsky district of the Kharkov region (43 km. from the town of Babai), advanced technologies will be applied, which are used today in the EU and the USA. The plant will receive, sort and recycle waste. At the new and old landfills—to produce gas. Also, the installation is activated at the site, allowing to generate electricity from gas (generator power—24 MW).

According to the engineer of the production and technical department company representative Vekto, the plant will process 400 thousand tons of waste per year. The new training ground is designed for more than 10 years of operation.

Next option of solving was proposed to replace all street lighting capable of self-generating energy, thus eliminating the need for in external recharge. Power is provided by an autonomous solar station, consisting of a photo module, battery and charge controller.

Independent mount allows maximum photomodule performance—orient it as far south as possible, and it is easy to change the angle of inclination depending on the time of year.

Batteries of various manufacturers provide from 5 to 11 h lamp work at night. A special controller that controls both battery charging and flashlight operation, supports several operating modes:

– automatic on and off depending on the light;
– continuous work throughout the night;
– work at certain intervals;
– shutdown a few hours after sunset and turn on a few hours before sunrise.

The cost of one unit of this installation is from 300 to 3500 USD. That at the current rate will be from 7943 to 92,668.59 UAH. Thus, while maintaining the current tariffs and adopting the foreign number of lighting points 2500 with an annual increase of 30% from the current indicator—1099 (Fig. 1), payback periods will look like this: at an average purchase price and without calculating replacement supplies the payback of this project will be approximately 8–10 years.

However, for the implementation of such projects grantors funds usually has been attracted which significantly reduces the cost of the project and payback periods for community. Last option will be relevant for more remote places where is problematic to stretch high voltage cables or for companies or open space zones to attract the visitors attention.

The author of this technology is a Dutch company Plant-e [10], which first presented it in the fall of 2014 in one of the parks of Hamburg. The project was called Starry Sky, and the mail idea was that 300 led downlights will receive electricity from living plants. This was demonstrated to all comers, at the presentation.

Plant-e founders are sure in the revolution of technology, since the method is completely environmentally friendly, it is possible to use vast areas of swamps for industrial power generation where there is a shortage of it. And here we are talking about whole countries.

On the company's website, the principle of the technology is described as follows: "A plant produces organic material through photosynthesis. Part of this organic material is used by the plant for its own growth, but not everything is used by the plant. This remaining part is excreted via the roots to the soil. In the soil, around the roots of the plant, bacteria break down the organic material. In this degradation process electrons are released as a waste product. The P-MFC technology ensures that we can use these electrons as electricity" [11].

The company's management claims that 1 m^2 of garden area, equipped in this way will be able to produce 28 kWh of electric energy per year, and this is quite suitable for areas of 100 m^2 or more, whether it is a garden plot, or similarly equipped greenhouses.

Due to the lack of prices on the company's website, the calculation of services for enterprises is impossible, but this service is definitely technological, environmentally friendly and innovative, and if such an initiative will be approved, it can find its co-financing in domestic programs and foreign grants.

3 Conclusions

The trends in global economies examined in the chapter showed that they are moving toward energy conservation and energy efficiency. Ukraine has also taken this path by developing, with UN support, a national SDG system. One of the ways of implementation applied in Ukraine was smart city concept. Using the smart city concept has proved its effectivity both in countries with advanced energy efficiency development such as Sweden, so in Ukraine, that just beginning to introduce elements of this concept in use.

An example of a successful project is the town Babai in Kharkiv region where, thanks to energy-efficient measures, electricity consumption by street lamps was reduced by 40% in the period 2015 to 2018 with an actual increase in lighting points by 30%.

At the same time, was noted that a decrease in actual consumption does not always lead to a reduction in costs. In some situations, this directly depends on the state regulation. So, in the period from 2017 to 2018, the night tariff was canceled and electricity prices increased, which actually leveled the savings due to energy-efficient measures.

However, these measures cannot be considered ineffective. Firstly, the number of lighting points in the village was increased by 30%. Babai, and secondly, despite the increase in lighting points, the amount of energy consumed was reduced by 40%. Thus, all the same, it reduced the possible costs of electricity for this locality. Indeed, if the indicators remained unchanged, then in 2018, for energy consumption at the 2015 level, one would have to pay 444,779.73 UAH instead of 289,056.41 UAH. paid upon, saving almost 50%.

The chapter notes that the example of the town Babai and Chkalovska UTC in the Kharkiv region showed that a decrease in resource consumption does not always lead to a reduction in costs. In this case, cost reduction has become impossible due to aspects of public administration aimed at raising tariffs.

The authors of the chapter proposed a number of measures that can reduce energy costs, considered options for their implementation, and also evaluated the economic feasibility. The chapter also notes that without competent and coordinated management at all levels of government, most energy-efficient measures may not be economically viable, although they will reduce the actual consumption of resource.

References

1. *Sustainable Development Goals (SDG)*. http://www.un.org.ua/ua/tsili-rozvytku-tysiacholittia/tsili-staloho-rozvytku
2. Bulkeley, H., & Castán Broto, V. (2013). Government by experiment? Global cities and the governing of climate change. *Transactions of the Institute of British Geographers, 38*(3), 361–375. https://doi.org/10.1111/j.1475-5661.2012.00535.x
3. *About the fundamental principles (strategy) of the State environmental policy of Ukraine for the period up to 2030*. https://zakon.rada.gov.ua/laws/show/2697-viii
4. Hodson, M., Geels, F., & McMeekin, A. (2017). Reconfiguring urban sustainability transitions, analysing multiplicity. *Sustainability, 9*(2), 299. https://doi.org/10.3390/su9020299
5. *About the fundamental principles (Strategy) of the state environmental policy of Ukraine for the period up to 2020*. https://zakon.rada.gov.ua/laws/show/2818-17
6. *Energy 2020: A strategy for competitive, secure, and sustainable energy [COM(2010)639]*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52010DC0639&from=EN
7. Bilous, L. B., & Dikan, V. V. (2017). *Public-private partnership as a way to attract investment in Ukrainian economy from the Baltic to the Black Sea: The formation of modern economic area* (pp. 37–38).
8. Bilous, L. B., & Dikan, V. V. (2019). Regional peculiarities of the formation of MSW in Ukraine and the methods of their processing. *International Scientific and Practical Conference World Science, 1*(9), 21–28.
9. *At the Dergachevsky landfill began the construction of a waste recycling plant*. https://kharkivoda.gov.ua/ru/news/92220
10. *Plant-e*. https://www.plant-e.com/
11. *Plant-e information and technology*. https://www.plant-e.com/en/informatie/

Detection of the Black Hole Attack on SDN-Based VANET Network

Mohammed Erritali, Badreddine Cherkaoui, Hanane Ezzikouri, and Abderrahim Beni-hssane

Abstract To ensure drivers safety many investments are carried on vehicular- ad hoc network (VANET). However adopting practically a VANET technology faces several issues linked to the attacks that can occur and may affect the efficiency and the reliability of the system. To get rid of the restriction and make VANET systems more efficient Software-Defined Networking (SDN) technology is introduced (SDN-based VANET). The communication of data that provide information to drivers and enhance road safety is based on the transmission between vehicles and Road Side Unit (RSU) connected by a wireless. The attackers may cause the dropping of incoming packets and creates one of the most security challenges: the black hole. Proposed study aims to identify and exclude attackers from the SDN-based VANET.

Keywords VANET · Black hole attack · SDN

1 Introduction

Vehicular ad hoc network is a set of autonomous and cooperatives mobile nodes (vehicles) that move and communicate by wireless transmission and do not assume preexisting management infrastructure. The ad hoc network is formed sponta-

M. Erritali (✉) · H. Ezzikouri
TIAD Laboratory, Computer Sciences Department, Sciences and Technics Faculty, University of Sultan Moulay Slimane, Béni Mellal, Morocco
e-mail: m.erritali@usms.ac.ma

B. Cherkaoui
LAROSERI Laboratory, Higher School of Technology, Chouaib Doukkali University, Sidi Bennour, Morocco
e-mail: b.cherkaoui@ucd.ac.ma

A. Beni-hssane
LAROSERI Laboratory, Computer Sciences Department, Faculty of Sciences, Chouaib Doukkali University, El Jadida, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_6

neously and provisionally, as soon as several nodes are within radio range of each other. A vehicle can communicate directly with another or by relying on the cooperation of neighbors to route packets to the destination [1]. The routing process in Vanet must be both short delivery delay time and low overhead. To meet these constraints, several studies [2, 3] have proposed to incorporate Software-Defined Networking (SDN) into VANET.SDN [4] is an architecture that decouples control and data transfer functions in the network in order to have a physical infrastructure completely free of any network service. In this model, network devices simply implement rules, injected by applications, for processing data flows. An intelligent entity, called “controller” can provide a global view to compute optimal packet routing paths for vehicle-to-vehicle communications. The centralized control of SDN provides an optimal routing solution according to [5], but we think that the entire network could be compromised if the controller is under attack. For this reason, we decided to study the effect of the black hole attack on SDN-based VANET network. To our best knowledge, this chapter is the first to propose a study of the black hole in SDN-based VANET. We implemented our studied black hole attack in ns-3 and we compared the performance with normal functioning in urban settings. The rest of this chapter is organized as follows: Sect. 2 describes the Software-Defined VANET routing architecture. Section 3 gives an overview of the black hole attack in VANET and SDN. Section 4 presents the simulation results of the network under black hole attack. Finally, Sect. 5 concludes this chapter.

2 SDN-Based VANET Architecture

In VANET the vehicles have both mode of communication vehicle-to-vehicle and vehicle-to-base-station. Besides, the Internet access service. So, in this chapter, we assume that vehicles have at least two modes, one of which can support vehicle-to-vehicle communication and another one can that be used to access the central controller which is connected to the Internet. In the architecture of SDN-based VANET, illustrate in Fig. 1, there is a remote central controller connected to the Internet and a routing server application running on it. Vehicles establish a routing client application to treat the packet sending event.

To set up an SDN-based VANET system, the architecture can have the following SDN components:

- SDN Controller: The SDN controller is responsible for determining the overall performance of the network.
- SDN Wireless Node: The SDN controller controls the elements of the data plane. And in this case, it is the vehicles that receive the control message from the SDN controller to perform actions.
- SDN RSU: These are the infrastructures that are deployed along the road segments.

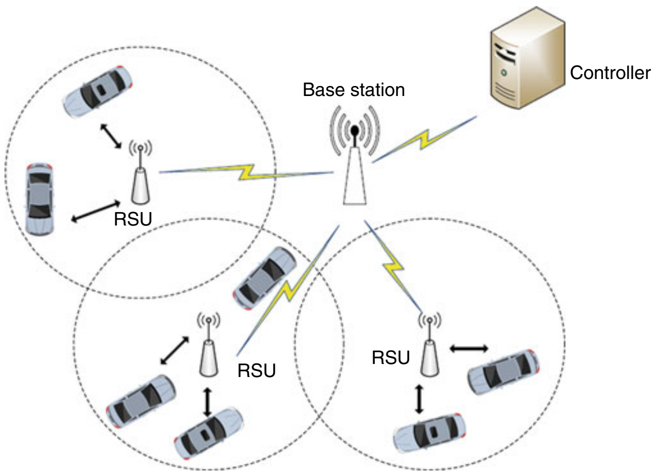


Fig. 1 SDN-based VANET architecture

In the SDN-based VANET architecture, various wireless technologies are chosen for the control plane such as LTE/Wimax for control plane, and wireless broadband connection, Wi-Fi for data plane [2].

3 Background Study of Black Hole Attack in Vanet and SDN

There are many techniques for black hole attack detection in SDN or in VANET architecture. Those works approaches depending on their anomaly detection methods are as follows: machine learning, traffic pattern analysis, and the use of SNORT [6]. Table 1 lists the details of those studies. However, these detection studies are either for the SDN or the vehicular network.

4 Simulation Experiments

This section presents the evaluation of the effect of DDOS attack on SDN-based in VANET using the network simulator NS-3. To evaluate the performance, we firstly use SUMO to generate vehicular traces in urban environment. And then use the trace as test scenarios. The Fig. 2 illustrates the process of simulation.

The simulation scenario is an area that was extracted from the OpenStreetMap of the ElJadida city in Morocco. We used SUMO to generate the movements of the vehicle nodes. We input the map extracted from the OpenStreetMap into SUMO and we specify the speeds limits. Simulation parameter is shown in Table 2.

Table 1 Black hole detection methods

Domain	Techniques
SDN	In [7] the authors present ez-Segway, a decentralized mechanism to update the network state while preventing forwarding anomalies like loops and black holes and avoiding link congestion
VANET	In [8], authors propose a cooperative cross layer-based intrusion detection schemes (IDSs) to enhance the performance of the watchdog detection technique in tackling black hole attack. Two detection schemes were proposed, the first one uses the information among physical and network layers, while the second one relies on the physical MAC and network layers to enable an efficient and reliable detection
	The authors in [9] propose a mechanism utilizes watchdog mechanism in order to detect selfish behaviors as well as black hole attacks against QoS-OLSR [10]. The detection technique is a five-phase and it is based on Tit-for-Tat concept and Dempster-Shafer theory of evidence
	In [11] to detect black hole attacks a dynamic entity-centric trust model based on weight is proposed according to the types of applications and the authority levels of nodes. Each vehicle uses three trust parameters initialized with default values

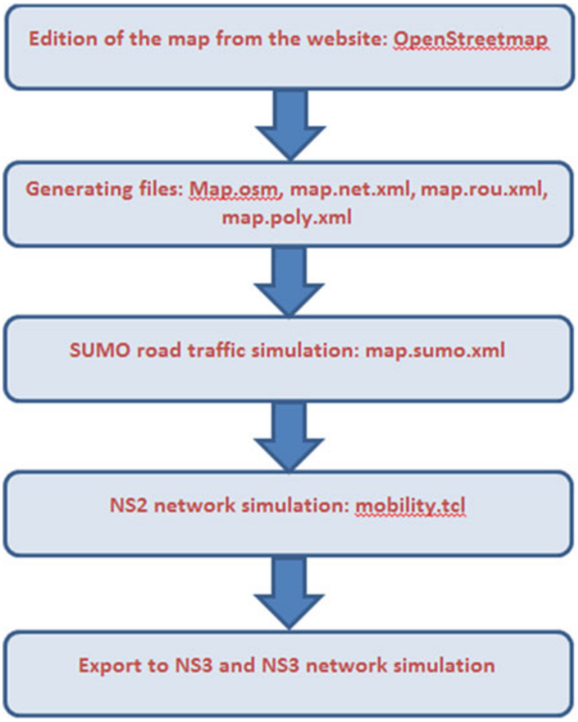


Fig. 2 Simulation process

Table 2 Simulation parameters

Parameter	Value
Number of vehicles	50, 100, 150, 200, 250, 300 and 350
Version of the simulator	NS-2.35 and NS-3.29
Vehicles speed	From 0 m/s to 25 m/s
Packet size	64 bytes
Transmission rate	2 kbps
Number of source nodes	5
Routing protocol	AODV
Number of attackers	1
Simulation time	100 s

4.1 Design of a Black Hole Attack

To perform a black hole attack, we established communication between several nodes (vehicles) in order to have a good amount of data circulating within the network. We assumed that a receiving node and 4 transmitting nodes are sufficient to calculate the performance of the network in question. We will analyze the behavior of the network in the normal case, then the behavior when the network is under a black hole attack. For this, we activate a malicious node that will perform the attack at the time of the establishment of a connection between the transmitter and the receiver. The main mechanism of the attack is to forge the routing information, such as the sequence number and the hop count, in order to illegally acquire the route. Then the attacker begins to receive the data packets and then they delete them all. This puts the receiving node out of service.

4.2 Performance Analysis

After observing the impact of the black hole attack on a network that operates in a traditional VANET environment [12], we will see that it is the impact on communication if we use an SDN approach. In this experiment, we used a simple Layer 2 learning switch controller application for testing. RSU act as learning switches that are all connected to the base station where the SDN controller is integrated. Then we analyze the behavior of the network in the normal case, then in the case of a black hole attack performed by a malicious node.

As we can see in the Fig. 3, the throughput at the receiving node in an SDN architecture is much better than in a traditional decentralized architecture de- spite the network's scalability in terms of the number of nodes. This proves that the architecture based on an SDN controller is more efficient compared to the traditional architecture. The same observation can be deduced from the Fig. 4 for the analysis of the packet delivery ratio. With the SDN controller, the PDR exceeds 90% even if the

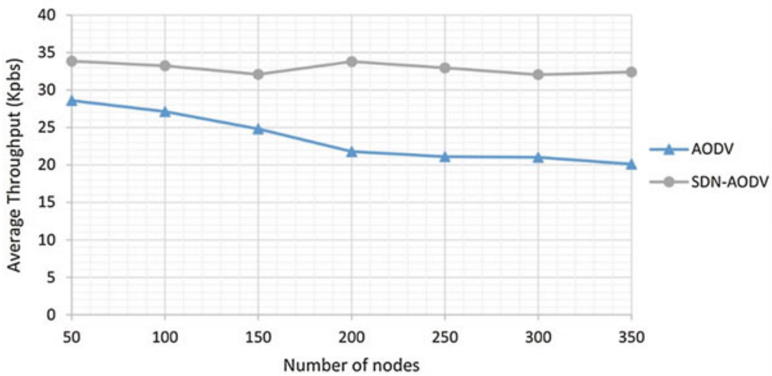


Fig. 3 Throughput of normal network

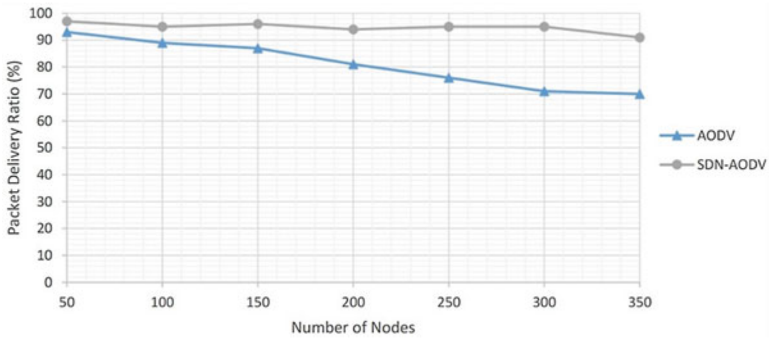


Fig. 4 PDR of normal network

number of nodes increases considerably. However, in the traditional architecture the PDR is affected by the increase of the number of nodes circulating in the network.

The results shown in the Figs. 5 and 6 Illustrate the throughput and the PDR of the receiving node in the case of a black hole attack. As noted in these figures, the throughput of the receiver in both use cases (with and without SDN) decreases considerably when a black hole attack is performed. Also, the PDR decreases by more than 70% in the architecture based on the SDN controller whereas the PDR in the case of the traditional architecture has decreased by a little near 40% when activating the hole attack black. To sum up, as noticed that during normal network behavior, architectures based on SDN controllers are more efficient compared to traditional architectures. However, in the attack case, architectures based on the SDN controller are more affected by the black hole attack compared to traditional architecture. This reveals fragility and also a major inconvenient of SDN architectures in vehicular environments.

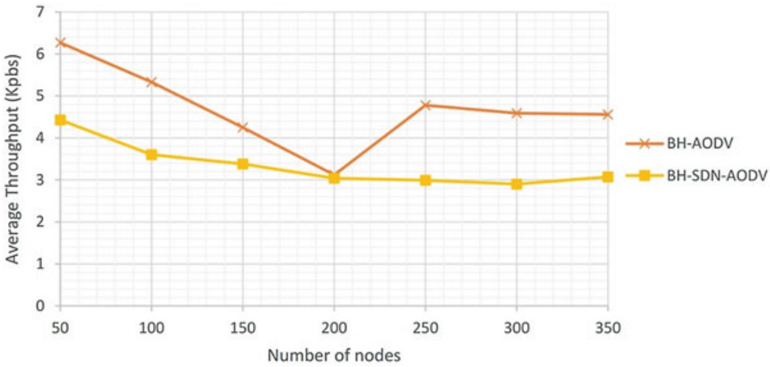


Fig. 5 Throughput of the network under black hole attack

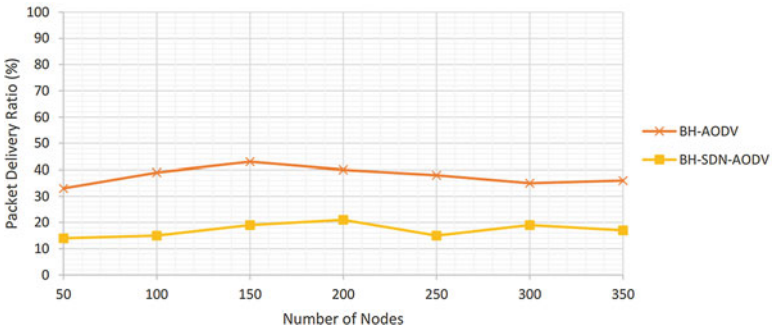


Fig. 6 PDR of the network under black hole attack

5 Conclusion

VANET networks, which have the objective to provide solutions for security and traffic management, need several security measures to ensure reliable communication. Among the security threats that currently exist, the black hole attack. In this chapter, we have studied the impact of the black hole attack in the SDN-based VANET architecture. Our future work will involve further research in the field of vehicular ad hoc networks in order to design a method to eliminate this attack in SDN-based VANET.

References

1. Erritali, M., & El Ouahidi, B. (2013). Performance evaluation of ad hoc routing protocols in VANETs. In *Proceedings of 3rd international symposium on automatic Amazigh processing (SITACAM'13), IJACSA* (pp. 33–40).
2. Ku, I., Lu, Y., Gerla, M., Gomes, R. L., Ongaro, F., & Cerqueira, E. (2014). Towards software-defined VANET: Architecture and services. In *Med-hoc-net* (pp. 103–110).

3. Zhu, M., Cao, J., Pang, D., He, Z., & Xu, M. (2015). SDN-based routing for efficient message propagation in VANET. In *International conference on wireless algorithms, systems, and applications* (pp. 788–797). Springer.
4. Durand, J. (2015). *Le SDN pour les nuls*. JRES.
5. Li, J., Chang, X., Ren, Y., Zhang, Z., & Wang, G. (2014). An effective path load balancing mechanism based on SDN. In *2014 IEEE 13th international conference on trust, security and privacy in computing and communications* (pp. 527–533). IEEE.
6. Erritali, M. (2013). *Contribution à la sécurisation des réseaux ad hoc véhiculaires*.
7. Nguyen, T. D., Chiesa, M., & Canini, M. (2017). Decentralized consistent updates in SDN. In *Proceedings of the symposium on SDN research* (pp. 21–33). ACM.
8. Baiad, R., Alhussein, O., Otrok, H., & Muhaidat, S. (2016). Novel cross layer detection schemes to detect blackhole attack against QoS-OLSR protocol in VANET. *Vehicular Communications, 5*, 9–17.
9. Wahab, O. A., Otrok, H., & Mourad, A. (2014). A Dempster–Shafer based tit-for-tat strategy to regulate the cooperation in VANET using QoS-OLSR protocol. *Wireless Personal Communications, 75*(3), 1635–1667.
10. Wahab, O. A., Otrok, H., & Mourad, A. (2013). VANET QoS-OLSR: QoS-based clustering protocol for Vehicular Ad hoc Networks. *Computer Communications, 36*(13), 1422–1435.
11. Yao, X., Zhang, X., Ning, H., & Li, P. (2017). Using trust model to ensure reliable data acquisition in VANETs. *Ad Hoc Networks, 55*, 107–118.
12. Cherkaoui, B., Beni-hssane, A., & Erritali, M. (2017). A clustering algorithm for detecting and handling black hole attack in vehicular ad hoc networks. In *Europe and MENA cooperation advances in information and communication technologies* (pp. 481–490). Springer.

Modeling the System of Farmland Taxation

Iryna Koshkalda, Nataliia Demchuk, Anastasiy Suska, Tetiana Anopriienko, and Alona Riasnianska

Abstract The rapid development of economic systems leads to the modeling and as a result of the involvement of modern information technologies. The land tax system is no exception, as land is a special real property and is included in the list of items of taxation in each country. In international practices, there are three basic concepts of the taxation system, based on the land as a subject property: the land as a separate item of taxation; the land is taxed with the real estate located on it; the land is subject to multiple taxes. The most widespread system is the land tax along with the real estate located on it.

Ukraine has introduced a model of a combined system of land taxation. Land taxpayers are the owners of the land, land shares (land units) and they are land users. The tax base is the regulatory monetary valuation of land plots and their area. Local self-government bodies set rates of the land tax and tax preferences. Establishing discriminatory rates of the land tax, the category of land, functional use, composition of farmlands, the qualitative status of the land, etc. are taken into account. Particularly relevant is the issue of modeling of the taxation of agricultural land—farmlands, which occupy the largest share of the land fund of the country, and the priority of their use for the intended purpose.

Models of the general system of land taxation in Ukraine and farmland taxation have been developed. Concerning it, the classification of farmland taxation has been proposed and the variants of establishing diversified rates of the land tax have been considered. They consider the qualitative state of specific land plots and the level of implementation of measures to improve their quality using the Thomas

I. Koshkalda (✉) · T. Anopriienko
V. V. Dokuchayev Kharkiv National Agrarian University, Kharkiv, Ukraine
e-mail: iryna.koshkalda@springer.com

N. Demchuk
Dnipro State Agrarian and Economic University, Dnipro, Ukraine
e-mail: nataliia.demchuk@springer.com

A. Suska · A. Riasnianska
Petro Vasylenko Kharkiv National Technical University of Agriculture, Kharkiv, Ukraine
e-mail: anastasiy.suska@springer.com; alona.riasnianska@springer.com

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_7

Saati Hierarchy Analysis Method. The introduction of a diversified taxation model will encourage taxpayers to allocate their funds for activities aimed at improving, rationalizing and protecting agricultural land.

Keywords System of taxation · Land tax · Farmland taxation

1 Introduction

In international practices, there are three basic concepts of the taxation system, based on the land as a subject property: the land as a separate item of taxation; the land is taxed with the real estate located on it; the land is subject to multiple taxes. Each country chooses its model of the real estate taxation system, farmland included, which depends on the economic development, political orientation, historical and cultural values of the country, as well as the structure of its land fund. At the same time, one of the tasks that are solved during the modeling is to ensure the conditions for efficient and rational use of the land, introduce a system of measures for improving its quality indicators.

2 Background

2.1 Types of Taxation Systems in the World

The most common system is the taxation of land along with the real estate located on it. In particular, this system is widely used in the USA, Canada, Portugal, Spain, and other countries. This system has two types: (1) the tax rate is the same for the whole land plot along with the real estate; (2) different tax rates for the land and the real estate located on it. For example, F. Dye Richard and W. England Richard, in their work entitled "Assessing the Theory and Practice of Land Value Taxation," give an example of the successful application of the Pennsylvania tax system: "The City of Scranton, Pennsylvania, has a two-rate tax with an estimated value of land taxed at 10.3% and improvements at 2.2%. Given that land value and improvements value represent 24% and 76% of the total tax base respectively, the equivalent of a single property tax rate on both land and improvements would be 3.7%" [1].

However, the world's land taxation systems differ not only in terms of the land as an object of the separate value or the value together with buildings but also in terms of other indicators. Thus, researchers R. Bjord and E. Slack in their scholarly work "International Handbook of Land and Property Taxation" emphasize that land taxation systems in different countries of the world differ not only in terms of the subject to taxation: "... there are differences in the determination of the tax base, the setting of tax rates, and the ability to levy and collect the tax. In some countries, one property tax covers all types of property. In others, there are different taxes for different components of real property. Countries may, for example, have separate taxes on land and buildings; separate taxes on the residential and non-residential property; or separate taxes in urban and rural areas" [2, p. 4].

Thus, we can also distinguish two types of taxation systems, based on the tax base: based on the value of the taxable object; based on the taxable area.

A very important function of land taxation is to fill local and state budgets. In particular, in the United States, bodies of local self-government adjust the land tax rate each year so that the proceeds from the land tax payment can cover the budgeted costs of financing specific social events [3].

It is worth noting the system where the rental value of the land is the tax base. This system is applicable in the UK, Spain, France, Portugal, Greece, India, Thailand, and Tunisia. Due to the complex process of determining the estimated rental value, countries such as the United Kingdom, Portugal, and Indonesia abandoned such a system and switched to the market value of the land as a tax base [4].

The taxation system in which the cadastral value of the land acts as a tax base is fundamentally different. The peculiarity of this system is that the state cadastral authority collects the necessary information and evaluates the real estate, while in countries where the market value is used as a tax base, special institutions (registration, valuation, insurance companies) perform these activities. Countries with such a system are France, Germany, and Sweden.

In countries with low levels of land cadaster, the land-based cadaster is applied in the countries of Central and Eastern Europe, Latin America, and Africa. As a rule, in these countries, the land and real estate are taxed separately [4–7].

Particularly special is the system of assessment and taxation in Maryland, which was the first in the United States to formally adopt a policy that included a reduction in the value and tax rates for agricultural land in order to comply with the principles of the State Agricultural Conservation Program [8].

Among the 28 EU countries, only nine countries have a land tax as a separate tax, while in others there are combined tax options. Thus, in post-socialist countries, land taxation depends on the form (type) of the land use and takes into account the real estate assets on it. At the same time, real estate taxes, including the land tax, play an important fiscal function when filling local budgets. Local authorities set tax rates independently based on the market value or cadastral value of the assets. Many researchers consider it is advisable to use a rented land taxation system.

World experience shows that the differentiation of land payments introduced by the state very often gives positive results in the form of more efficient and rational use of the land, the introduction of a system of measures to improve its quality.

In some countries, the legislation sets out the mechanisms, employing which the state exercises the incentive functions of rational land use. For example, in the case of irrational land use, the tax rate increases, and when introducing effective land use methods, preferential conditions are granted, up to a zero tax rate, or for low-productive lands, rates are reduced to 50%. Although in some countries in Eastern Europe, the tax rate depends on the value of the area, not on the object.

2.2 Land Taxation in Ukraine

Ukraine has a combined system of land taxation, particularly in agriculture, land is a specific subject of taxation [9]. Thus, the Tax Code of Ukraine [10] establishes that the tax base is the normative monetary value of land plots, taking into account the indexation coefficient and the area of land plots where the normative monetary evaluation have not been carried out. The rates of land tax for agricultural land where the normative monetary evaluation have been carried out do not depend on the location of the plot and they are:

– for the agricultural land—not less than 0.3% and not more than 1% of their regulatory monetary value;
– for the public land—not more than 1% of their regulatory monetary value;
– for land plots that are in constant use of economic entities (except for the state and municipal property)—not more than 12% of their regulatory monetary value;
– for all other types of land—not more than 3% of their regulatory monetary value.

One of the most effective levers for improving the quality of land, preserving its natural fertility and potential is to use economic mechanisms, including the taxation system.

Methods of economic and mathematical modeling and practical use of optimal processes for predicting the efficiency of economic systems have been widely adopted recently, this will further promote the rational use of land resources and improving the food security of the country as a result [11, 12]. The processes of evaluation and taxation of agricultural land should be considered as the main means of production in agriculture in this context.

2.3 Land Taxation in European Countries

In European countries, the price scale is quite high: from 4 thousand euros per hectare in Finland to 10–12 thousand euros per hectare in France and Germany, and 25 thousand euros in the Netherlands [13]. Their assessment includes such factors as the level of agricultural intensity, the structure and type of soil, and so on. In the United States, agricultural land prices range from $2–3 thousand per hectare in arid regions up to 30–40 thousand dollars per hectare in California.

2.4 Estimation Methods with the Storie Index Determination: US Experience

The most known land evaluation technique in the United States is the Storie Method [14–16], which provides for the classification of arable land with the definition of a Storie Index that affects the land value.

$$\text{Storie Index} = A \cdot B \cdot C \cdot X, \tag{1}$$

where A is the different characteristics of the soil profile; B—mechanical soil composition; C—ground slope; X—other characteristics (drainage, erosion, fertility rate, micro-relief).

Thus, arable land is found to be divided into five classes, wherein the higher the class the higher is the Storie Index. For the first class, the Storie Index is in the range of 80–100%, for the second class—60–79%, for the third class—40–59%, for the fourth class—20–39%, for the 5th—less than 20%. According to this classification, the value of US agricultural land ranges within \$300–1200 per year.

3 Research Approach

3.1 Adaptation of the Storie Method to Determine the Land Tax in Ukraine

Using the Storie method, the formal model of determining land tax in Ukraine will be as follows:

$$\text{LT} =< F, R, \text{CFT}, Q, B, t >, \tag{2}$$

LT—land tax; F—the legal framework of the land taxation system; R—the legal regime of the land plot (property, permanent use, the subject of taxation); CFT—land category, functional use, and type of land; Q—the quality of the land; B—benefits of the land taxation system; t—period (time) of land taxation.

The Tax Code of Ukraine [10] stipulates that the land tax rate for the farmland (irrespective of its location) should be not less than 0.3% of its normative monetary evaluation and not more than 1% of its normative monetary evaluation, taking into account the indexation coefficient.

Therefore, the model of the land tax system for farmland in Ukraine is as follows:

$$\text{LT}_{\text{al}} =< F, R, \text{FT}, Q, B, t >, \tag{3}$$

LT_{al}—land tax on farmland; F—the legal framework of the system of land taxation of farmland; R—the legal regime of the land plot (property, permanent use, subject of taxation); FT—functional use and type of farmland; Q—the quality of the land; B—benefits of the land taxation system; t—period (year, time) of land taxation.

In the Ukrainian land taxation system, including the farmland taxation, there are constant changes and improvements in the regulatory framework (F), so the time factor plays an important part of the current model of land taxation (t).

The Tax Code of Ukraine [10] stipulates that land taxpayers are the owners of the land, land shares and they are the land users. Differentiated land tax rates are set out in Articles 273–277 of the Tax Code of Ukraine [10], which consider the land category, functional use, type of land (FT) and regulatory monetary evaluation of the respective plots of land and shares. The Land Code of Ukraine [17] provides for five types of agricultural land—arable land, lay land, perennial crop fields, hayfields, and pastures. These types of land can be provided for agricultural commodity production as well as for individual gardening, horticulture and the like.

Local bodies of self-government, in their territories, independently set land tax rates and benefits (*B*) for tax payment. When establishing differentiated land tax rates, the quality of the land (*Q*) is also taken into account.

3.2 Introduction of Differentiated Taxation System in Ukraine

In Ukraine, to increase the concernment of landowners and permanent land users in the land use based on measures aimed at improvement, rational use and protection of the land, it is advisable to introduce a classification of farmland during taxation, namely a diversified system of taxation of farmland based on five classes:

first class—non-eroded and slightly eroded land and eroded land on slopes over 3°, with no measures taken by the subject of taxation to improve its quality;
second class—slightly and moderately eroded land on slopes up to 3° with no measures taken by the subject of taxation to improve its quality;
third class—slightly eroded land on slopes up to 7°, the subject of taxation is introducing measures to improve the quality;
fourth class—moderately and heavily eroded land on slopes up to 7°, the subject of taxation is introducing measures to improve the quality;
fifth class—eroded lands on slopes over 7°, the subject of taxation is introducing measures to improve the quality.

4 Results

To establish diversification in the system of taxation of agricultural land (based on five classes), we consider two options for determining the tax rate based on the quality of land and measures taken directly by the subject of taxation to improve its quality, setting a minimum rate for the fifth class (0.3% of the regulatory monetary evaluation) and the maximum rate for the first class (1.0% of the regulatory monetary evaluation).

While grouping, these classes imply a certain qualitative composition of the soil surface and slope of the land, the worse the quality indicators, the more the necessity to attract resources, especially financial ones, to preserve and improve the quality, the more complex the measures for achieving the optimum result may be.

The first option. We assume to set equal intervals for increasing the tax rate from 0.3% to 1.0%.

(1.0% − 0.3%) /4 = 0.175%.

Round to the tenths, we get an interval of 0.2%.

Then the land tax rates will be:

fifth class: 0.3% of regulatory monetary evaluation.
fourth class: 0.3% + 0.2% = 0.5% of regulatory monetary evaluation.
third class: 0.5% + 0.2% = 0.7% of the regulatory monetary evaluation.
second class: 0.7% + 0.2% = 0.9% of regulatory monetary evaluation.
first class: 0.9% + 0.2% = 1.1% of regulatory monetary evaluation. Since the maximum amount of the land tax is set at 1.0% of the normative monetary evaluation, and given that landowners and land users do not take measures to improve the land, the land tax rate for the first class is set at 1.0% of normative monetary evaluation of these types of land.

The second option. Setting land tax rates for different classes, we propose to apply the Thomas Saati Hierarchy Analysis Method, which has been widely used in evaluation activities while selecting alternatives and finalizing the results of the evaluation [18].

Therefore, to determine the land tax rates for farmland of different quality and where the measures to improve its quality can be introduced by landowners and land users (subjects of taxation), we suggest to use paired comparisons and the following scale to evaluate the results of the comparison of alternatives:

1—equivalence (equal importance);
3—moderate (slight) advantage;
5—great (significant) advantage;
7—a very (obvious) advantage;
9—the highest (extreme, absolute) advantage;
2, 4, 6, 8 are intermediate values.

We make paired comparisons of alternatives, the results of which are presented in Table 1.

We change a simple fraction into a decimal and calculate the term sums of the results of paired comparisons of alternatives, the normalized value (which sum

Table 1 Paired comparisons of alternatives

Class	5	4	3	2	1
5	1	2/1	3/1	6/1	9/1
4	1/2	1	4/1	5/1	6/1
3	1/3	1/4	1	2/1	3/1
2	1/6	1/5	1/2	1	2/1
1	1/9	1/6	1/3	1/2	1

Table 2 Calculation of the weighting coefficients using the Hierarchy Analysis Method based on the results of paired comparisons of alternatives

Class	Sum by line	Normalized value	Calculated weight, %	Defined weight, %
5	21.00	0.4195	41.95	40
4	16.50	0.3296	32.96	30
3	6.58	0.1315	13.15	20
2	3.87	0.0772	7.72	10
1	v2.11	0.0422	5	0
Total	50.06	1.0000	100	100

should be equal to 1.0), and define the weighting coefficients (Table 2). According to the Saati method, the normalized values are accepted as evaluations of alternatives. The calculation of the weighting coefficients is approximated to 10%.

Then the land tax rates (rounded to tenths) will be:

first class: $1.0\% - 0.7\% \cdot 0\% = 1.0\%$ of normative monetary evaluation.
second class: $1.0\% - 0.7\% \cdot 10\% = 0.9\%$ of normative monetary evaluation.
third class: 0.9% - 0.7% $\cdot$ 20% $= 0.8\%$ of normative monetary evaluation.
fourth class: $0.8\% - 0.7\% \cdot 30\% = 0.6\%$ of normative monetary evaluation.
fifth class: $0.6\% - 0.7\% \cdot 40\% = 0.3\%$ of normative monetary evaluation.

Thus, as in the first case, the difference between the minimum and maximum rates of land tax is 0.7%, with the maximum rate being defined for the first class 1.0% of the normative monetary evaluation (Table 3).

Economic indicators of the profitability of the land of different quality influence their demand from potential landowners and land users. Eroded land on slopes over 7° requires significant investment to undertake measures to improve the quality of such plots. A reduced rate of land tax will allow economic entities to use more of their saved money for land protection measures and to increase the land fertility. Non-eroded and poorly eroded land on slopes up to 3° requires measures mainly to maintain its condition and prevent its deterioration. Such measures require small investments and are mostly ensured with the crop rotation, hay and pasture systems maintained by the economic entity.

The results obtained for the two variants of the calculations shown in Table 3 indicate the similarity of the results. It is worth noting that the data obtained by the Hierarchy Analysis Method are more accurate to reflect the level of investment for the implementation of measures to improve the quality of eroded land and to create sustainable environmentally friendly agro-landscapes and more.

5 Conclusions

Taking into account the experience of agricultural land taxation studied abroad two approaches of establishing differentiated land tax rates in Ukraine are considered in the chapter. The classification of agricultural land taxation by the differentiated

Table 3 Classification of farmland taxation at different rates

Class	Characteristics of farmland	Land tax rate, % of normative monetary evaluation	
		Option 1	Option 2
1-st	Non-eroded and slightly eroded land and eroded land on slopes over 3°, with no measures taken by the subject of taxation to improve its quality	1	1
2-й	Slightly and moderately eroded land on slopes up to 3° with no measures taken by the subject of taxation to improve its quality	0.9	0.9
3-й	Slightly eroded land on slopes up to 7°, the subject of taxation is introducing measures to improve the quality	0.7	0.8
4-й	Moderately and heavily eroded land on slopes up to 7°, the subject of taxation is introducing measures to improve the quality	0.5	0.6
5-й	Eroded lands on slopes over 7°, the subject of taxation is introducing measures to improve the quality	0.3	0.3

rates according to the two variants is offered, herewith the second variant (data obtained by the method of hierarchy analysis) is recommended for introduction into the national system of taxation of agricultural land.

Consequently, the introduction of a diversified taxation model will encourage taxpayers (landowners and land users) to allocate their funds for the activities aimed at improving, rationalizing and protecting the farmland.

References

1. Dye, R. F., & England, R. W. (2010). *Assessing the theory and practice of land value taxation*. Lincoln Institute of Land Policy. 34 p. https://www.lincolninst.edu/sites/default/files/pubfiles/assessing-theory-practice-land-value-taxation-full_0.pdf
2. Bird, R., & Slack, E. (2003). *International handbook of land and property taxation*. Oxford University Press. 320 p. https://www.e-elgar.com/shop/international-handbook-of-land-and-property-taxation
3. *Innovative land and property taxation: United Nations human settlements*. https://www.ucl.ac.uk/qaser/pdf/publications/UN_HABITAT
4. Bird, R. M., & Slack, E. (2005). Land and property taxation in 25 countries: A comparative review. *CESifo DICE Report*, (3), 34–42. https://core.ac.uk/download/pdf/6631080.pdf
5. Olubode-Awosola, F. (2010). Agricultural land tax and farm-level land-use changes and output supply response. *China Agricultural Economic Review, 2*(1), 79–93. https://doi.org/10.1108/17561371011017513

6. *SDAT agricultural transfer tax*. https://dat.maryland.gov/realproperty/Documents/agtransf.pdf
7. Chigome, J. (2015). A reflection on agricultural land tax in Zimbabwe. *Research Journali's Journal of Economics, 3*(2) https://www.researchgate.net/publication/280732976_A_Reflection_On_Agricultural_Land_Tax_In_Zimbabwe
8. Skinner, J. (1991). *The World Bank Economic Review, 5*(1), 113–133. https://www.jstor.org/stable/3989972?seq=1#page_scan_tab_contents
9. Gershman, Y. V. (2018). Land as a specific object of taxation in agriculture. *Eastern Europe: Economics, Business and Management, 6*(17), 641–646.
10. Tax Code of Ukraine (*Podatkovyi kodeks Ukrainy*). https://zakon.rada.gov.ua/laws/show/2755-17#n6750
11. Babenko, V. (2019). Formalization of the model of management of the technological innovations. *CEUR Workshop Proceedings, 2393*, 595–602. Online, http://ceur-ws.org/Vol-2393/. http://ceur-ws.org/Vol-2393/paper_431.pdf
12. Panukhnyk, O., Popadynets, N., & Fedotova, Y. (2019). Analysis and modeling of factor determinants of food provision at consumer market of Ukraine. *Global Journal of Environmental Science and Management, 5*(SI), 215–226.
13. United States Department of Agriculture. https://www.nrcs.usda.gov/wps/portal/nrcs/main/national/landuse/
14. Storie, R. E. (1978). Storie index soil rating. *Division of Agricultural Sciences, 3203*, 1–4. https://anrcatalog.ucanr.edu/pdf/3203.pdf
15. *A revised storie index for use with digital soils information–ANR catalog* (pp. 1–11). https://anrcatalog.ucanr.edu/pdf/8335.pdf.
16. National Soil Survey Handbook (NSSH). (2018). United States Department of Agriculture. Natural Resources Conservation Service. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2_054242
17. Land Code of Ukraine (*Zemelnyi kodeks Ukrainy*). http:/zakon2.rada.gov.ua/laws/show/2768–14
18. Saaty, T. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences, 1*, 83–98. https://pdfs.semanticscholar.org/e3c5/61049eb532e328fc2b8288c490986cd9403f.pdf

Partner Relationship Assessment Methodology

Iryna Perevozova, Nadiia Daliak, Andriy Mokhnenko, Tetiana Stetsyk, and Vitalina Babenko

Abstract The mechanism of marketing activity of an enterprise that implements a specific policy of interaction with partners takes into account the specifics of the organisation of such interaction, defined by the regulatory field of the relevant market. This approach is defined by the functioning of a set of structural elements as the organisation of a single mechanism. But, in turn, corresponds to the goal orientation of the economic mechanism of the enterprise, that is, the structure and internal processes that occur in the system of marketing relationships of the enterprise, and the enterprise as a whole must be mutually consistent. In the conditions of high level of instability and uncertainty of the market it is necessary to constantly develop and improve the mechanisms for enhancing stability, effective management of enterprises, forming new relationships. Increasing the potential for interaction in today's socioeconomic environment requires the expansion of a methodological apparatus and tools to study the relationship between counterparties. The theory and methodology of marketing activity is built around the measurement of the effectiveness of the interaction of actors, because marketing is increasingly seen as a process of interaction. This approach describes and defines the laws under which the mechanism of interaction of market entities operates, the nature of which determines the level of viability and competitiveness of the company, and relations at different levels of interaction are far from homogeneous. The

I. Perevozova (✉) · N. Daliak
Entrepreneurship and Marketing Department, Ivano-Frankivsk National Technical University of Oil and Gas, Ivano-Frankivsk, Ukraine

A. Mokhnenko
Department of Finance, Accounting and Entrepreneurship, Kherson State University, Kherson, Ukraine

T. Stetsyk
Department of Philology and Translation, Ivano-Frankivsk National Technical University of Oil and Gas, Ivano-Frankivsk, Ukraine

V. Babenko
V.N. Karazin Kharkiv National University, Kharkiv, Ukraine
e-mail: vitalinababenko@karazin.ua

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_8

organisational mechanism for conducting the research of the system of relations with the counterparties has been developed and the universal method of estimating the satisfaction of the counterparties and the system of interaction with them has been improved. To increase the degree of involvement of all divisions of the enterprise in the achievement of the common goal—to strengthen the relationship with counterparties and the corresponding increase of profit in the results of business agreements with them, it is proposed to bring the results of the study to the management of all services of the enterprise and to use as planned indicators in the further work of the services.

Keywords Partner relationship system · Harrington's desirability function · Dynamic model

1 Introduction

Modern demands of the competitive environment require enterprises to form an interaction system with their partners on the principles of flexibility, transparency and adaptability. "There is only one legitimate justification for the purpose of business: creating a satisfied customer" [1, c. 198]. It is the study of consumers' needs, satisfaction and guessing of their expectations faster than competitors [2, p. 145] that serve as an innovative basis for business development in the conditions of hypercompetition.

Enterprises are forced to quickly and effectively respond to a client's needs using all possible communication channels. Thus, more attention has recently been paid to the creation of mutually beneficial, long-term relations between market players, where consumers and manufacturers work together for creating and providing consumer value: the transition from the classical client-oriented concept to the partnership concept has begun.

2 Theory of the Matter

The formation of "partnership" or network approach in marketing was started by an international scientists group from Europe (IMP Industrial Marketing and Purchasing), including scientists from France, Germany, Italy, Sweden and the UK. The group emerged in the late 1970s and early 1980s as a result of a research program based on the hypothesis that marketing theory is incomplete and not suitable for understanding important aspects of industrial marketing in practice.

The transition of enterprises to the B2B type of relationship requires the development of a model of partner relationship management system (PRM-system) that would connect the whole complex of relationships with different subjects of interaction. It should be noted that a limited number of research works deals with the formation of a common system of relationships between companies and different groups of partners. Analysing the authors' work it becomes evident that there is no consensus on the partner relationship (PR) formation. One group of authors [3–6] focus on the relationship with only one group of subjects' interaction and Ph. Kotler [7], G. Morgan and S. Hunt [8], J. Egan [9] and others include in the relationship system different groups of partners as participants in the relationship, they consider customers (consumers), distributors, suppliers, employees and other partners relationship.

3 Detailed Presentation of the Methodology of the Study, Characterisation of Materials and Methods of Analysis, Statistical Processing of Results

It should be emphasised that there are a large number of classification features regarding the types of enterprise-to-customer cooperation. For example, D. Ford and E. Jones [10] divide business-to-consumer relationships into: symmetrical, where the two parties benefit from mutual learning and sharing of knowledge and technology, and customers in a symmetrical relationship hope to gain from the seller the support beyond the contract; asymmetric, where the supplier (seller) is dominating, and customers rely on the unique capabilities of one or more suppliers; and asymmetric, buyer-dominated relationships, and the supplier's technological, financial and resource capabilities are limited by emphasis on cost reduction and streamlining, interaction is managed by the customer, suppliers depend on one or more customers in the network. Ph. Kotler and G. Armstrong [11] propose to differentiate the enterprise-to-customer relationship by the levels of trust and strength of these relations, depending on the type of customers they are divided into: those who keep a close eye on the enterprise; potential customers; one-time customers; customers who have re-purchased products; customers in general; lawyers; members; partners. The first two groups are not in contact with the organisation. Consumers, who keep a close eye on a company or product, require more spending. Lawyers, members and partners cover resources spent by the value brought, in other words, they spread a positive opinion about the company and attract new customers at no charge.

Going along with A. Bolotnaya [12], it should be noted that the relationship system is based on building relationship not only with customers, but also with all enterprise's partners. The basis of CR and PR building is taking into account their expectations and needs as well as the principles of mutual respect. Bilateral or multilateral communication is possible in the relationship system (Fig. 1).

The system is called integral if its structure includes all the necessary elements and connections; each of the elements performs certain functions, as does the system; according to the W.-R. Ashby [13], principle of emergence, the characteristics of the system are different from the totality of the elements and relationship

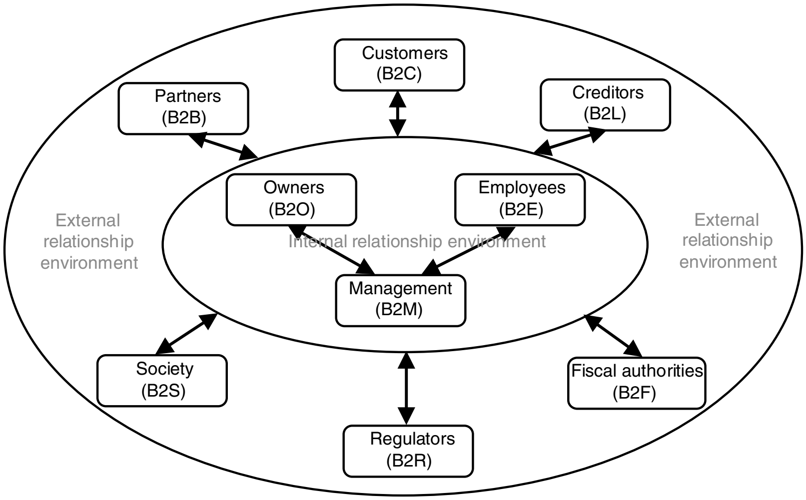


Fig. 1 Partner relationship system environment

characteristics. The peculiarities of how the system functions as a holistic formation are determined by the elements and connections nature, that is, the structure, and the system itself influences the elements, causing changes in their properties and qualities.

The system of partner interaction refers to the subsystem of the company governance (management), which brings the processes of the internal relationship environment in accordance with the company's target plans or programs in relation to the external relationship environment. Regulation is based on the analysis of partners' reactions, with great attention being paid to innovative approaches. The process of the relationship system implementation means creation, adjustment and maintenance of the mechanism in the company's management system, which ensures not only the implementation of the company's strategic and tactical goals, but also creates conditions for their achievement. The first stage of this process is to build a system that implements customer service in accordance with common corporate standards. The next step is to incorporate all other business technologies, including marketing technologies, into the system. The ultimate goal of the project is to increase the flow of new customers. Related results are marketing costs optimisation, improvement of customer order flow management, marketing efficiency overall improvement.

The first stage of this process is to build a system that implements customer service in accordance with common corporate standards. The next step is to incorporate all other business technologies, including marketing technologies, into the system. The ultimate goal of the project is to increase new customers' inflow. Related results are as follows: marketing costs optimisation, improvement of customer order flow management and marketing efficiency overall improvement.

Taking into account A. O. Ustenko's [14] approach regarding the structural components of the enterprise management system, we consider it expedient to clarify the directions of interaction between them, since the target subsystem is initial for both the controlling and the controlled ones, at the same time, given the author's approach, the controlling subsystem has no influence on the controlled one. In addition, the author misses the feedbacks. The system of partner-to-enterprise relations is compositionally a complex holistic system that includes many elements and links, the processes in the system are probable, the system has a large number of parameters and characteristics, some of them are criteria that take extreme values in the process of functioning with regard to restrictions (Fig. 2). By the evaluation indicator of the PR subsystems we mean an integral indicator (multiplier), which quantitatively determines the qualitative characteristics of a process, plan. Indicators are defined as the parameters of the boundaries in which the system, including organisational mechanisms, technological links, monetary financial flows, can steadily function and develop [15].

Scientists offer various stages, methods and scales of assessments, but limited number of research works is devoted to the development of PR evaluation system, which can simplify work with consumers and partners as well as help to identify the enterprise's strengths and weaknesses that require improvement to increase loyalty and further marketing development.

In most methodological approaches respondents are asked to evaluate a company's products, services and performance against a set of criteria. These criteria can be considered as characteristics of the PR quality category. Thus, in 1995 R. Carter offered a list of generalised characteristics that are taken into account in complex assessments of companies' compliance with the partners' specifics and needs. This list is called "Ten C" [16]: competence—the supplier's competence to solve the tasks; capacity—the supplier's capacity to meet the buyer's needs; commitment—the supplier's commitment to the consumer regarding quality, price and service; control system—inventory, costs, budgets, personnel and information control systems; cash resources and financial stability—cash resources and financial stability, attesting to the supplier's financial health and its ability to continue business in the foreseeable future; cost—price in accordance with the quality and level of service; consistency—supply stability, product improvement and services quality; culture—the supplier and consumer have a compatible culture and common values; clean—suppliers and their products meet legislative and environmental safety requirements; communication—the ability to communicate with the supplier using modern information technologies.

One of the most well-known approaches is N. Kano's model of three levels of satisfaction, according to which customer's satisfaction is proportional to the level and type of product quality [17]. Expected product quality characterises what may or may not prompt the purchase. Desired product quality means that if the product

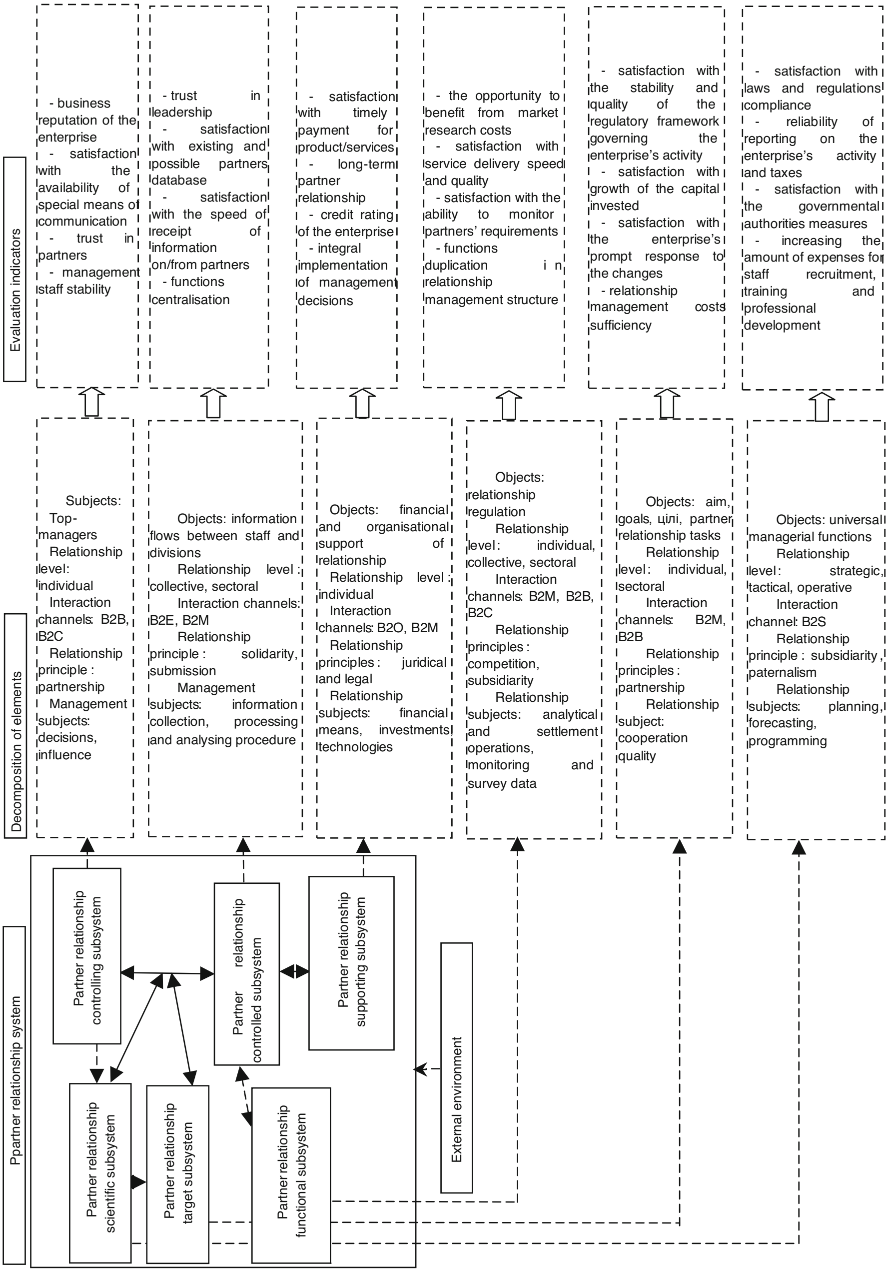


Fig. 2 Partner relationship system

is improved, then satisfaction increases. Attractive or exciting product quality is an unexpected level of service. In such cases, consumers are pleasantly surprised, delighted and even shocked. At the same time, such exciting quality soon turns into the expected.

A description of the general procedure for assessing the level of partners' satisfaction or dissatisfaction is provided by J.-J. Lamben [18, p. 165]. It is based on the concept of a multi-attribution model and includes three steps: first, it evaluates the average value of overall satisfaction with the product or supplier; second, it calculates mean value of satisfaction and importance for each characteristic; in the conclusion, it assesses the intentions to make a repeat purchase or to make a deal. The analysis evaluates satisfaction/importance ratio to determine satisfaction with the most important characteristics, which has a decisive influence on the counterpart choice.

The Gap model developed in 1985–1991 by American researchers A. Parasuraman, V. A. Zeithaml, L. L. Berry suggests to measure partners' satisfaction with service by analysing the gaps between expectations and actually provided level of service. Gap means exceeding the consumers' expectations over the evaluation of the service received in reality [19, 20]. The Gap model makes it possible to see the process of service provision as a whole, to identify a possible source of its unsatisfactory quality.

G. Likert scale, proposed in 1932, is a method of multivariate estimation in which respondents evaluate specified judgments using the answers from one critical position through neutral and to another critical position, e.g. "completely satisfied", "somewhat satisfied", etc. The rating scale is assigned to each criterion [21]. This is a simple but not very reliable approach as it does not take into account the relative importance for partners of the criteria by which the company is evaluated that complicates the identification of aspects requiring priority improvements from the customer's point of view.

As shown by B. Mittal and W. M. Lassar study [22], dissatisfaction guarantees disloyalty, whereas satisfaction does not guarantee loyalty, and only maximum satisfaction ensures it. Therefore, full partners' satisfaction is a significant factor in creating loyalty, which requires regular research to track the dynamics of customers' satisfaction.

In the scientific literature, the practical application of the above models is more often reflected independently, which significantly limits the comprehensive understanding of the PR system. In addition, in most methodologies, work on customer satisfaction assessment is limited to assessing the products or services quality. This is not enough to obtain a qualitative assessment of partners' satisfaction and conduct a detailed analysis of the customers' opinion about the company. Combining different strategies allows a more detailed approach to the analysis.

The authors' methodology is based on theoretical approaches of J.-J. Lamben, A. Parasuraman, V. A. Zeithaml and L. L. Berry. The modified 10-point Staple scale is used as analysis tools in the methodology.

The organisational mechanism for conducting research of the PR system involves several stages.

Stage 1. Setting the survey purpose: it can be identifying of critical indicators that have led to products demand decrease; determining buyers' expectations to maintain a leading position in the market; segmenting customers by specific indicators and more.

Stage 2. Developing a list of indicators, which are important for both the company and partners and allows to answer the set research goal, by all the company's services.

Stage 3. Preparing a database of respondent companies.

Stage 4. Developing the survey questionnaire: formulating questions, selecting an evaluation scale.

Stage 5. Personal questionnaires are sent by fax or e-mail to the respondent companies. Their routing is clearly tracked.

Stage 6. Carrying out the analysis on the basis of the questionnaire data collected and evaluating each PR subsystem.

The author's method of PR system assessment is presented in Table. 1.

To increase the degree of involvement of all divisions in achieving the common goal of strengthening PR and making a profit, it is proposed to bring the study results to the attention of the management of all the company's units and use in the further units' work. In order to increase the respondents' motivation to fill in questionnaires and establish feedback with counterparts, it was proposed to send

Table 1 Methods of evaluation of subsystems of the partner relationship system

Indicator name	Indicator symbol	Stage content: formula
Average value of all indicators significance $P_1, \ldots, P_k$ for one respondent C_j. Calculated for each respondent	$I^k_{C_j}$	$I^k_{C_j} = \frac{C_j^{P_1} + \cdots + C_j^{P_k}}{k}$, (1) where C_j ($j = 1, \ldots, m$)—respondent; $P_1, \ldots, P_k$—indicators analysed; $C_m^{P_1}, \ldots, C_m^{P_1}$—respondents' assessment C_m of indicators $P_1, \ldots, P_k$; k—number of indicators analysed; m—number of respondents
Average value of one indicator significance P_i for all respondents $C_1, \ldots, C_m$. Calculated for all indicators.	$I^m_{P_i}$	$I^m_{P_i} = \frac{P_i^{C_1} + \cdots + P_i^{C_m}}{m}$, (2) where P_k—indicators analysed; $C_1, \ldots, C_m$—respondent; $P_i^{C_1}, \ldots, P_i^{C_m}$—respondents' assessments $C_1, \ldots, C_m$ of indicator P_i ($i = 1, \ldots, k$); k—number of the parameters analysed; m—number of respondents
Overall average value of all indicators significance $P_1, \ldots, P_k$ for all respondents $C_1, \ldots, C_m$	I_{PM}	$I_{PM} = \frac{\sum_{j=1}^{m} I^k_{C_j}}{m} = \frac{\sum_{i=1}^{k} I^m_{P_i}}{k}$, (3) where $I^k_{C_j}$—average value of all indicators significance $P_1, \ldots, P_k$ for one respondent C_j ($j = 1, \ldots, m$); $I^m_{P_i}$—average value of one indicator significance P_i ($i = 1, \ldots, k$) for all respondents $C_1, \ldots, C_m$; k—number of indicators analysed; m—number of respondents

written notifications on the results of each questionnaire on the measures that were developed and taken on the basis of the questionnaires received.

Despite the considerable amount of research on partner interaction analysis, it remains to be seen that, in order to maximise customer satisfaction, increasing attention should be paid to examining their needs and expectations. In this case, the partners' satisfaction becomes one of the most effective marketing and management tools, which allows to evaluate the company's effectiveness and to predict how the company's market share may change depending on the current state of its customers and partners' satisfaction.

For the overall assessment of the PR system we propose to use Harrington's desirability function [23], which allows us to model the process of concerted behaviour of individual subsystems of the whole, to consider the relationship and the impact between them.

The basis for constructing and prioritising this generalised function is the transformation of the natural values of the partial parameters of different physical entities and dimensions into a single dimensionless scale of desirability (preference). The purpose of the scale is to establish the correspondence between the physical and psychological parameters of optimisation.

The desirability function can be used as an accessory function because d [0,1]. It emerged from the observations of the respondents' real decisions and has such useful properties as continuity, monotony and smoothness. In addition, this curve conveys well the fact that in the dimensions, preferably close to 0 and 1, its "sensitivity" is significantly lower than in the middle zone (Fig. 3). Actually it is a logistic (S-shaped) system efficiency curve.

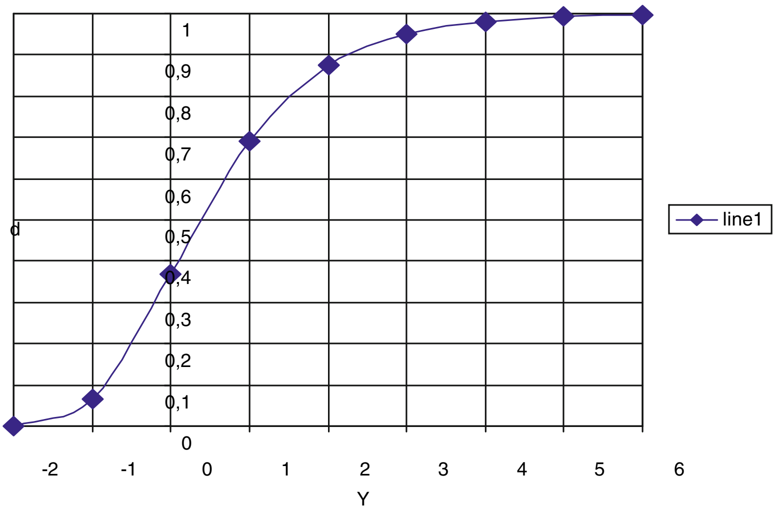


Fig. 3 Harrington's desirability function graph with one-way constraint

The value of the partial variable translated into a dimensionless desirability scale, is denoted by di ($i = 1, 2, \ldots, n$) and is called partial desirability. The desirability scale has an interval from zero to one. The value of $di = 0$ corresponds to an absolutely unacceptable level of the property, and the value of $di = 1$ corresponds to the best value of the property. From a mathematical point of view, the author of the approach recommends assigning a desirability value of d at point of 0.37 to the normative values of the indicators. The number 0.37 is an approximate result of dividing 1 by the number e, where e is the basis of the natural logarithm. The second such point is the desirability value of 0.63, which is the result of the difference (1–1/e). The generalised desirability index is calculated by the formula:

$$D = \sqrt[6]{d(1)d(2)\ldots.d(6)}, \tag{4}$$

where 6 is the number of subsystems of the counterparts relations system. The root of the sixth degree "smoothes" the arising deviations, and the received result allows to estimate systems (with a certain degree of accuracy, so to speak, "mathematically".

In turn, the desirability index d for each individual given characteristic and Y for a group of indicators with a unilateral restriction is determined by the formula:

$$d = e^{-\left(e^{-Y'}\right)} \tag{5}$$

where e is the logarithmic constant, which is approximately equal to 2.71828 . . . ;

Y is the result of linear transformation of Y relationship subsystem estimation.

This formula represents a special case of the Gomperz growth function, which is shown in Fig. 1.

In order to further transform this metric into a dimensionless form, Harrington E. recommends first introduce two pairs of values (Y_{d1}, $_{d1}$) and (Y_{d2}, $_{d2}$). After that, the indicators Y_{d1}, Y_{d2} are standardised by the following formula:

$$Y' = -[\ln(-\ln d)] \tag{6}$$

Using a pair of reduced values Y we can calculate two constants $b0$ and $b1$, which are necessary for the standardisation of other indicators with a one-way constraint. These constants are found from the linear equation:

$$Y' = b_0 + b_1 Y \tag{7}$$

After translating all the values into dimensionless form, a generalised desirability index is calculated as the geometric mean of all partial desirability indicators. The logic behind using the geometric mean is that if at least one of the parameters is zero, in other words, is undesirable, then the state of the entire object of assessment is undesirable.

4 Discussion of Results

As can be seen from the above procedure for calculating a generalised index using the desirability function, standardisation of indicators requires determining their normative values, determining a pair of numbers (Y_{d0}, $_{d0}$) to calculate the parameter n, as well as two more pairs of values (Y_{d1}, $_{d1}$) and (Y_{d2}, $_{d2}$) to standardise unilateral indicators. It is these stages that give rise to the biggest disadvantage of Harrington's approach—subjectivism. Therefore, to reduce the impact of subjective evaluation, you should involve a team of experts to determine these pairs of numbers.

The interpretation of the estimate obtained is a dynamic model of the levels of the PR system (Fig. 4), the zones of which are correlated with Harrington's "desirability scale" (Table 2).

Intrasubjective level of the PR system implies the enterprise's passive behaviour towards its partners and vice versa. Autonomous or intrasubjective quasi-communication (the subject is both the addresser and the addressee (internal dialogue), in other words, is the object to himself or herself directly) and pseudo-communication (cooperation with inanimate objects, etc.) is present.

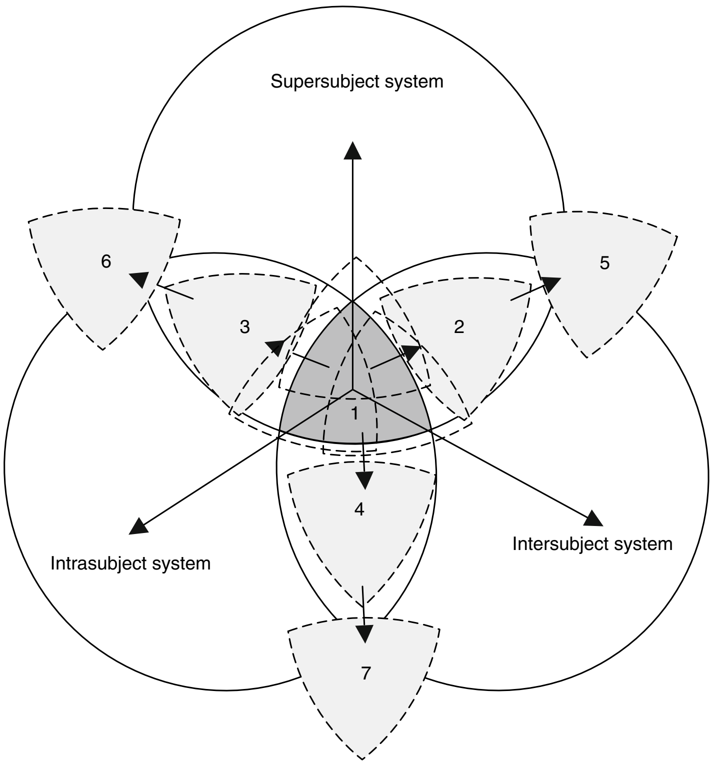


Fig. 4 Dynamic model partner relationship system levels

Table 2 Harrington's "desirability scale"

Desirability	Zone	Qualitative mark on the desirability scale "*d*"	Qualitative characteristic of the scale "*d*"
Absolute	1	1.00	Displays the extreme level of PR excellent level that is meaningless to improve
Very good	2, 3	1.00–0.80	Acceptable at the "excellent" level. Displays an unusually good level of PR system
Good	4	0.80–0.63	Acceptable at the "good" level. Displays a level that exceeds the best level that corresponds to $d = 0.63$
Satisfactory	5	0.63–0.37	Acceptable at "satisfactory" level. The PR system is acceptable to the maximum permissible level but needs improvement.
Bad	6	0.37–0.20	Limit level. If standard specifications exist, individual products will fall outside of these specifications. (if the characteristic accurately corresponds to the set minimum or maximum, then the value of "*d*" should be equal to 0.36788 = 1/e)
Very bad	7	0.20–0.00	Unacceptable level

The intersubjective level of the PR system is aimed at changing the overall atmosphere of the relationship. Intersubjective is a question of all three subjects "what I need from them", "what they expect from me", "how can I help the situation to turn it on the right track" and so on. The answers they give to these questions shape the dynamics of intersubjective space. Intersubjectivity posits that we need to understand others in order to understand ourselves.

The supersubjective level of the PR system characterises the development of social-communication forms and lifestyles of society and aims to provide stable networks of social interactions (institutions).

Thus, Zone 1 provides a balance of interests and benefits of all participants in the PR system (enterprise–partners–society).

Zones 2, 3, 4 take into account specific multiple participants' wishes and requirements (partners–society, enterprise–society and enterprise–partners, respectively).

Zones 5, 6, 7 in the elements beyond the boundaries of Euler circles take into account the uncertainty index, where it is not possible to neatly form and evaluate the relationship system.

5 Conclusion

Thus, the PR system should be considered as a complex phenomenon in which all components are interdependent and interacting. The main components of this system are: subjects, which are responsible persons at different levels (e.g. the Head of customer service, Marketing Director, CEO); objects (processes of customer service and interaction between staff to solve the problems of this service); goals and tasks (e.g. achieving a high level of coordination of customer service processes and their effectiveness); methodology (methods of administrative, organisational, disciplinary, and other impacts on system objects to qualitatively complete the task); tools of influence (technical, technological, organisational and regulatory means of monitoring, planning, organising, motivating, and adjustments, due to which the system is operating); resources (material, monetary, intellectual, informational and other tools that are used in the system); parameters (qualitative and quantitative measurements of the system functioning).

References

1. Tracy, B. (2004). *Victory!* Popuri.
2. Charles, D. (2003). *Schewe, Alexander, Watson Hiam: The portable MBA in marketing*. Publisher House.
3. Narodytskii, I. (2010). Partner relationship management (PRM). *Financial Newspaper, 16*, 45–49.
4. Newell, F. (2004). *Why CRM doesn't work? Customer relation management: How to win by letting customers manage the relationship*. Publisher House.
5. Payne, A. (2007). *Handbook of CRM: Achieving excellence in customer management*. Publisher House.
6. Noland, C. (2003). Beyond strategic sourcing: Strategic supplier relationship management. In *ISM's 88th annual international supply management conference, 2003* (pp. 45–51). Print House.
7. Kotler, P. (2006). *A framework for marketing management* (2nd ed.). Piter.
8. Morgan, R., & Hunt, S. (2004). The commitment-trust theory of relationship marketing. *Russian Management Journal, 2*(2), 73–110.
9. Egan, J. (2008). *Relationship marketing exploring. Relational strategies in marketing* (2nd ed.). Uniti-Dana.
10. Ford, D., & Johnsen Rhona, E. (2000). Asymmetrical and symmetrical customer-supplier relationship: Contrasts, evolution and strategy. *International Journal of Entrepreneurial Behaviour and Research, 5*, 297–312.
11. Kotler, P., & Armstrong, G. (2003). *Marketing essentials* (9th ed.). Viljams.
12. Bolotnaya, O. *Relationship marketing as a modern marketing concept*. https://cyberleninka.ru/article/v/marketing-vidnosin-yak-suchasna-kontseptsiya-marketingu
13. Ashby, W. R. (1959). *An introduction to cybernetics*. Inostrannaja lileratura.
14. Ustenko, O. O. (2014). Enterprise management system. In: *Bulletin of the pre-carpathian university. Series: Economics*. (10th ed., pp. 96–103).
15. Voronkova, V. G., Belichenko, A. G., & Popov, O. M. (2006). *Human resources management: Philosophical background*. Profesional.
16. Carter, R. (1995). The seven Cs effective supplier evaluation. *Purchasing and Supply Chain Management*.
17. Kano, N., Seraku, N., Takahashi, F., & Tsuji, S. (1984). Attractive quality and must-be quality. *Journal of the Japanese Society for Quality Control, 14*.
18. Lamben, J.-J. (1996). *Strategic marketing*. European perspective: Transl. from French. Nauka.
19. Zeithaml, V. A. A., Berry, L. L., & Parasuraman, A. (1988). Communication and control processes in the delivery of service quality. *Journal of Marketing, 52*.

20. Parasuraman, A., Berry, L. L., & Zeithaml, V. A. (1985). Conceptual model of service quality and its implication for future research. *Journal of Marketing, 49*.
21. Predvoditeleva, M. D., & Balaeva, O. N. (2005). Service quality management approaches: Focus on the consumer. *Management in Russia and Abroad*, (2).
22. Mittal, B., & Lassar, W. M. (1998). Why do customers switch? The dynamics of satisfaction versus loyalty. *Journal of Services Marketing, 12*.
23. Harrington, E. C., Jr. (1965). The desirability function. *Industrial Quality Control, 21*, 494–498.

Anomaly Detection-Based Negative Selection Algorithm

Hanane Chliah, Amal Battou, and Omar Baz

Abstract The real-valued negative selection algorithm and binary-valued negative selection algorithm are an anomaly detection approach, motivated by the negative selection immune system principle. In the training stage, solely positive (normal) samples are obtainable of some anomaly detection applications. Further, during the training phase traditional classification algorithms require samples for total classes. This chapter begins with a review of anomaly detection. Then, we present a description of negative selection algorithms.

Keywords Anomaly detection · Negative selection algorithm · Real-valued negative selection · Binary-valued negative selection

1 Introduction

In late years, the field of anomaly detection problem can be appointed as a one-class classification problem, whereas just normal samples are obtainable at training stage [1]. The role of anomaly detection method is identical to the biological immune system because both of them essay to spot the abnormal information [2]. Among several mechanisms in the immune system that are scout about for AIS, negative selection is the highest defeated model [3]. In 1994, Forrest et al. [4] suggested a method based on the generation of T cells in the immune system, and this method is named negative selection algorithm (NSA). NSA has abstracted the regard of many researchers and practitioners [5]. It has approved to be effective for anomaly detection problems [3]. Several modified versions of NSA algorithms had in the event also qualified solutions for problems of anomaly detection [6], fault diagnosis

H. Chliah (✉) · A. Battou
IRF-SIC Laboratory, Faculty of Science, Ibn Zohr University, Agadir, Morocco

O. Baz
Higher School of Technology of Agadir, Ibn Zohr University, Agadir, Morocco
e-mail: o.baz@uiz.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_9

[7], computer security [8, 9]. Binary coding put to use in the first NSA to depict self and non-self-samples [4]. The hyper sphere detectors adopted with constant radius, and a real-valued NSA (RNSA) was presented [10]. A little after, a real-valued with variable-sized detector [11], hypercube detector [2], hyper-ellipsoid detector [12], and multi-shaped detector [13] were suggested. The hyper sphere detector is also extensively used, considering it has a simple mathematic version. Various methods were engaged which orderly degrade the detector number and accomplish enough detector coverage [14]. Though the methods brought up on top can remove the holes and enhance the detection rate. The detectors are generated at haphazard of these methods for the identical training data, size, the position and quantity of detectors generated in every time are dissimilar. The NSA has been sparsely attention paid to with constant detectors. NSA detectors namely quantity, position, and size are constant. All of these detectors are just associated to learning samples, and have no relation to training times. This chapter focuses on negative selection algorithm in anomaly detection. The following of this chapter is ordered as ensuing. The second section, we present the principals of anomaly detection. In section three, we provide a detailed discussion about a negative selection algorithm and the different presentations as binary-valued negative selection and real-valued negative selection. Finally, the fourth section draws conclusions and future research directions. Sample Heading (Third Level). Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

2 Anomaly Detection

Offered an element of the space, tabulated it as normal or abnormal are the two-class problem which present the anomaly detection problem. Various terminologies are utilized in dissimilar classification applications, like newness [15] or surprise [16] detection, fault detection [17], and outlier detection. Correspondingly, statistical techniques have been cited in several approaches [18], machine learning [19], data mining [20] and immunological inspired techniques [2].

However, the negative samples (abnormal) in many anomaly detection applications are not convenient at the training phase. For example, it is arduous to obtain information about any feasible attacks in computer security applications. The absence of abnormal class samples in machine learning approaches leads to difficulties in the application of supervised techniques like classifying. Accordingly, use an unsupervised algorithm is the evident machine learning.

The approaches of anomalies detection inspired by artificial immune systems have been successfully painstaking, not totally in the field of computer security [21], but as well as in other areas such as generous hardware failures [1], function optimization [21], tool breaking detection [15].

3 Negative Selection Algorithm

The variation and anomaly detection has been used extensively in the negative selection algorithm (NSA) [4]. The algorithm is motivated by the selection process that takes place within the thymus. The T cells are suspended which admit body's own cells (self-cells) in this process; this guarantees that the residual T cells will acknowledge just extraneous molecules. D'haeseleer et al. [22] suggested a powerful implementation of the negative selection algorithm recognized as the greedy algorithm, which is specifically selected for a binary representation of the self/non-self-space and the r-contiguous matching rule. Although this algorithm has been used disinterested change/anomaly detection problems, certain confines, moreover, have prohibited it from be applied more widely:

- Scalability: a great number of detectors has to be generated (depending on the size of the self), in order to warranty right levels of detection. Certain problems, the generation the number of detectors could be difficult [23].
- The illustration of the base (binary) level detector precludes the extraction of significant domain acquaintance. However, it is difficult to interpret the reason to report an abnormality.
- Considering the acute discrimination into normal and abnormal. This separates the space between two subsets: self (normal) and non-self (abnormal). An element of space called abnormal if exists a detector that accords to it. Actuality, normality is not a fuzzy concept. A natural way of typify personal space is to specify a point of normality.
- Some immune-inspired algorithms use elevated level representation (e.g., real-valued vectors) and a small level illustration (e.g., binary-valued vectors) makes difficult the NS algorithm with diverse immune algorithms.

The negative selection algorithm, is inspired by the process of thymocytes (i.e., Young t-lymphocytes) maturation: only those lymphocytes outwear which do not acknowledge any self molecules [4]. Formally, let u be a universe, i.e. The set of all possible molecules. The subset s of u represents collection of all self molecules and its complement n in u represents all non-self molecules. Let $d \subset u$ represent a set of detectors and let match(d, u) be a function (or a procedure) indicate if a detector $d \in d$ recognizes the molecule $u \in u$. Habitually, match(d, u) is patterned by a distance metric or a similarity measure, i.e. We say that match(d, u) = true only if dist(d, u) $\leq \delta$, where dist is a distance and δ is a pre-specified threshold. Various matching function are debated in [3, 8]. The problem relies about manufacturing the set d in such a way that for all detector $d \in d$.

$$\text{Match } (d, u) \begin{cases} \text{False} & \text{if } u \in S \\ \text{True} & \text{if } u \in N \end{cases}$$

A simple solution to this problem, implicit by biological mechanism of negative selection, consists of four steps:

(a) Initialize d as void set, $d = \varnothing$.

(b) Randomly generating a detector d.

(c) If math(d, s) = false for all $s \in S$, append d to the set d.

(d) Reiterate steps (b) and (c) until an adequate number of detectors are generated.

The binary and real representations of the problem are described below.

3.1 Binary Valued Negative Selection Algorithm

Forrest et al. [4, 24] expanded a negative selection algorithm based on the principles of self/non-self discrimination in the immune system. The collection s of strings of length l over a finite alphabet has been specified by the negative selection algorithm, this collection that needs to be supervised. Thus the algorithm generates a set r of detectors each of which does not correspond to any chain of s. The NSA s monitor looks for changes by continuously matching the detectors from r to s. When a detector matches, a change is recognized to have occurred because the detectors are composed to not match any of the original strings S.

There are different versions of the algorithm with different degrees of computational complexity to efficiently generate detectors; they use a binary representation and run-in linear time with the size of the individual, depending on the choice of matching rule [4, 21, 22].

However, the temporal (and spatial) complexity of the algorithm is exponential with respect to the size of the corresponding window (the number of bits to be used in the comparison of two bit strings). This represents a problem of scalability, because the expression of complex interaction clouds requires large corresponding windows. Another important problem is the comprehensibility of the problem space representation. The binary coding is low level representation for certain applications. This makes complicate or impossible in many cases to return binary sensors to the original problem space. Therefore, the approach get unnecessary as an extracting learning.

3.1.1 Matching Rule For Binary Representation

- R-contiguous matching rule

  In r-contiguous matching rule, divide string of arbitrary extent into smaller segments of predefined length. Matching requirement is defined as r-contiguous matching symbols in corresponding positions.

- R-chunk matching rule

  The r-chunk matching rule has been proposed for the first time by balthrop [25] and is an improved variant of the r-contiguous matching rule developed by Percus. The rule of r-contiguous concordance is one of the first rules that was based on the biological immune system and shaped the binding between antibody and antigen. In formally, two elements, of the same length, correspond under the r-contiguous rule, if the least contiguous characters are identical. The r-chunk matching rule performs the highest match performance against other matching rules of the binary alphabet. Since the number of detectors that can be generated strongly influences the matching performance, we focus on the number of detectors that can be generated on arbitrary sizes.
- Hamming distance

  Hamming distance is a matching rule for strings representation. The hamming distance of two strings, $s1$ and $s2$, is determined as the least number of point mutations consisted to change $s1$ into $s2$, where a point mutation is to change a letter, to insert a letter, or to delete a letter. Levenshtein distance is a generalization of the Hamming distance, the latter can be designed by Levenshtein distance.

3.2 *Real-Valued Negative Selection Algorithm*

Rather of utilizing a binary coding in the negative selection algorithm, a new approach uses a real-valued representation to typify the auto-non_self space and progress a set of detectors that can be placed in the (non_self) complementary subspace [26].

The real-valued vector representation is therefore another important type of representation. It executes each data as a vector of real numbers. The matching rules and the measure of variance or similarity must be based on the numerical elements of the vector. The negative selection algorithm (RNS) was proposed by Gonzalez et al. [26]. A detector (antibody) is defined by an n-dimensional vector that corresponds to the center and by a real value that represents its radius. Therefore, a detector can be considered as a hypersphere. The real-valued representation implies that the data is normalized in the range [0,1] before further processing.

3.2.1 Matching Rule For Real-Valued Representation

- Euclidean Distance

  Euclidean distance is tantamount to the projection of an inferior-dimensional space from the original space. In matching rule Euclidean distance is not studied again total the elements of the vector. In a matching rule Euclidean distance is not studied over all the elements of the vector. In a lower-dimensional space merely certain elements are utilized to determine Euclidean distance. Ji and Dasgupta

[3] measure can be selected contiguously as incidentally. In both instances, the choice of positions must correspond within the two points whose distance is calculated.

- Manhattan Distance

  Manhattan distance is the distance between two points measured alongside axes at right angles. For example, gifted two points $F1$ and $F2$ in a two dimension plane at $(A1, B1)$ and $(A2, B2)$ respectively, the Manhattan distance within $F1$ and $F2$ is determined by $\mid A1 - A2 \mid + \mid B1 - B2 \mid$.
- Minkowski Distance

  The Euclidean distance and the Manhattan distance can be considered the Minkowski distance as a generalization of the both of them. It's a metrical in a normed vector space. Minkowski distance of order p between two points

$$X = (X_1, X_2, \ldots, X_n) \quad \text{and} \quad Y = (Y_1, Y_2, \ldots, Y_n) \in R^n$$

$$\text{Is defined as : } \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}}$$

4 Conclusion

In surveying on negative selection algorithm based on artificial immune system, NSA gaining the popularity and attraction. The majority of works suggested are based on variation of NSA. Its performance is based on the interaction among the detector generation algorithm and matching technique adopted for use. Matching techniques and rules used are different in different works. Relying on the type of data representation, either for strings or real valued, the proper detection algorithm must be assigned. To solve all these problems and progress in this research, our future work should focus on in-depth experiments to measure the real power of the NSA algorithm.

References

1. Patcha, A., & Park, J.-M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks, 51*(12), 3448–3470.
2. Dasgupta, D., & Gonzalez, F. (2002). An immunity-based technique to characterize intrusions in computer networks. *IEEE Transactions on Evolutionary Computation, 6*(3), 281–291.
3. Ji, Z., & Dasgupta, D. (2007). Revisiting negative selection algorithms. *Evolutionary Computation, 15*(2), 223–251.
4. Forrest, S., Perelson, A. S., Allen, L., & Cherukuri, R. (1994). Self-nonself discrimination in a computer (pp. 202–212).
5. Dasgupta, D., Yu, S., & Nino, F. (2011). Recent advances in artificial immune systems: Models and applications. *Applied Soft Computing, 11*(2), 1574–1587.

6. Stibor, T., Mohr, P., Timmis, J., & Eckert, C. (2005). Is negative selection appropriate for anomaly detection? In *Proceedings of the 2005 Conference on Genetic and Evolutionary Computation—GECCO'05*, Washington DC, USA (p. 321).
7. Dasgupta, D., Krishnakumar, K., Wong, D., & Berry, M. (2004). Negative selection algorithm for aircraft fault detection. In G. Nicosia, V. Cutello, P. J. Bentley, & J. Timmis (Eds.), *Artificial immune systems* (Vol. 3239, pp. 1–13). Springer.
8. Harmer, P. K., Williams, P. D., Gunsch, G. H., & Lamont, G. B. (2002). An artificial immune system architecture for computer security applications. *IEEE Transactions on Evolutionary Computation, 6*(3), 252–280.
9. Gonzalez, F., & Dasgupta, D. (2005). Artificial immune systems in intrusion detection. In *Enhancing computer security with smart technology* (pp. 165–208). Auerbach Publications.
10. Dasgupta, D., & Gomez, J. (2003). An evolutionary approach to generate fuzzy anomaly signatures. In *Proceedings of IEEE Systems, Man and Cybernetics Society, IEEE, 2003* (pp. 251–259).
11. Ji, Z., & Dasgupta, D. (2009). V-detector: An efficient negative selection algorithm with "probably adequate" detector coverage. *Information Sciences, 179*, 1390–1406.
12. Shapiro, J. M., Lamont, G. B., & Peterson, G. L. (2005). An evolutionary algorithm to generate hyper-ellipsoid detectors for negative selection. In *Proceedings of the 2005 Conference on Genetic and Evolutionary Computation—GECCO'05*, Washington DC, USA (p. 337).
13. Balachandran, S., Dasgupta, D., Nino, F., & Garrett, D. (2007). A framework for evolving multi-shaped detectors in negative selection. In *2007 IEEE Symposium on Foundations of Computational Intelligence*, Honolulu, HI, USA (pp. 401–408).
14. Costa Silva, G., Palhares, R. M., & Caminhas, W. M. (2012). Immune inspired fault detection and diagnosis: A fuzzy-based approach of the negative selection algorithm and participatory clustering. *Expert Systems with Applications, 39*(16), 12474–12486.
15. Dasgupta, D. (1996). *Stéphanie, ForrestDept, Novelty detection in time series data using ideas from immunology.*https://www.semanticscholar.org/me/library
16. Keogh, E., Lonardi, S., & 'Yuan-chi' Chiu, B. (2002). Finding surprising patterns in a time series database in linear time and space. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002 (pp. 550–556).
17. Yairi, T., Kato, Y., & Hori, K. (2001) Fault detection by mining association rules from housekeeping data (p. 7).
18. Denning, D. E. (1987). An intrusion-detection model. *IEEE Transactions on Software Engineering, SE-13*(2), 222–232.
19. Omar, S., Ngadi, A., & Jebur, H. H. (2013). Machine learning techniques for anomaly detection: An overview. *International Journal of Computer Applications, 79*(2), 33–41.
20. Lee, W., & Stolfo, S. J. (2000). *Data mining approaches for intrusion detection*. Defense Technical Information Center.
21. Hofmeyr, S. A., & Forrest, S. (2000). Architecture for an artificial immune system. *Evolutionary Computation, 8*(4), 443–473.
22. D'haeseleer, P., Forrest, S., & Helman, P. (1996). An immunological approach to change detection: Algorithms, analysis and implications. In *Proceedings 1996 IEEE Symposium on Security and Privacy*, Oakland, CA, USA, 1996 (pp. 110–119).
23. Kim, J., & Bentley, P. J. An evaluation of negative selection in an artificial immune system for network intrusion detection (p. 8).
24. Gonzalez, F., Dasgupta, D., & Kozma, R. (2002). Combining negative selection and classification techniques for anomaly detection. In *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)*, Honolulu, HI, USA, 2002 (Vol. 1, pp. 705–710).
25. Saurabh, P., & Verma, B. (2014). A novel immunity inspired approach for anomaly detection. *International Journal of Computer Applications, 94*(15), 14–19.
26. Gonzalez, F. A., & Dasgupta, D. (2003). Anomaly detection using real-valued negative selection. *Genetic Programming and Evolvable Machines, 4*, 383–403.

Optimized Flow Aggregation for Energy-Efficiency and QoS Routing in Ad Hoc Wireless Network

Mustapha Boukrim and Jilali Antari

Abstract Ad Hoc wireless network is a type of network that suffers from various types of limitations. Among these limitations is energy consumption. The proposed flow aggregation technique allows the user to maximize the number of nodes that could be turned off, in the network, to increase energy saving while meeting QoS provisions. The proposed solution was formulated as an Integer Linear Programming (ILP) problem using a set of energy and QoS constraints. To evaluate the efficiency of the proposed model, a performance-based comparison was made with anther routing model. The simulation results show that the proposed model has a better performance in terms of global energy consumption and network load.

Keywords Ad Hoc wireless networks · ILP · Flow's aggregation · Global energy consumption · QoS provisions · Routing

1 Introduction

An Ad Hoc wireless network involves the interconnection of wireless nodes without using any fixed infrastructures to achieve flexibility in the network structure. Nodes in Ad Hoc wireless networks are usually battery-operated and are mostly deployed in critical environments such as military zones, disaster, recovery [1–3] and in an emergency healthcare situation where it is almost impossible to replenish the batteries [4]. This makes it necessary to conserve battery energy for the sustainability of operation and for the network longevity to be prolonged. The rates at which nodes in the network consume energy differ depending on whether the nodes are in transmitting, receiving, listening or sleeping state [5]. The least energy

M. Boukrim (✉)
Laboratory of Metrology and Information Processing, Faculty of Sciences, Ibn Zohr University, Agadir, Morocco
e-mail: mustapha.boukrim@edu.uiz.ac.ma

J. Antari
Polydisciplinary Faculty of Taroudant, Taroudant, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_10

is consumed when a node is in sleeping state, but all nodes will not always be in the sleeping state. Also, power consumption is a major concern to find network topology that meets QoS requirements. Several studies have already been gone through this direction [6–8]. The majority of these studies have proposed algorithms that make numerous assumptions about the computing power of the nodes to reduce the energy consumption locally. However, it cannot be optimal in terms of global energy savings of the network. Usually, the classical approaches are based on the equitable distribution of the load between all nodes so that their batteries exhaust equivalently to prevent a node from dying prematurely because of its overuse [9].

Most of the work involved in constructing and maintaining a network topology with the required connectivity has used minimal energy consumption. The chapter [10] proposes two centralized algorithms for use in static networks. These algorithms aim to construct a connected and two-connected networks with the minimization of the maximal transmitting power for each node. For mobile networks, two other distributed heuristics were further presented, such as Local Information No Topology and Local Information Link State Topology [10], to adaptively adjust node transmit powers in response to topological changes. The issue of finding an optimal topology for Bluetooth Wireless Personal Area Networks is discussed in [11]. The optimal topology that provides complete network connectivity and meets traffic requirements and system specification constraints is found. Moreover, this topology minimizes the traffic load of the most congested node in the network.

Other studies have discussed the problem of energy consumption for the Ad Hoc wireless network. Many energy-aware routing algorithms are developed and their effects on various QoS parameters intrinsic to traditional networks such as delay, bandwidth, packet loss are discussed in [12]. This study shows that energy-aware algorithms usually increase the overall longevity of the network by efficiently using the energy resources of every node. To determine the optimal transmission power of nodes, Rashmi et al. [13] propose a clustering-based scalable topology control method (Optimized Power Control), which aims at keeping the network connected throughout the communication. The effects of this method on network parameters, such as average transmission power per node, total energy consumption, and network delay, have been studied. The proposed approach is based on topology control, and its simplicity makes it suitable to be used for performance enhancement of other network scenarios such as WSNs, VANETs, and CRNs.

In [14], the author proposes a solution to determine the transmission power for each node to minimize the consumed energy by all nodes in the network while meeting the end-to-end QoS constraints in terms of delay and bandwidth. In [9], the authors prove that the distribution of the load over a large number of nodes forces them to remain active along the communication process, which results in very low global energy consumption. Then it is necessary to exploit mechanisms that allow to switch off unused nodes for good management of network resources.

The present work aims at including a flows' aggregation technique to increase the number of unused nodes, hence significantly reducing global energy consumption in the network. The goal of the study is to aggregate flows over a sub-set of nodes in the network, to minimize the number of nodes required to route a set of flows.

Given a set of nodes in a two-dimensional surface and QoS requirements of flows, this idea consists of finding a network topology that minimizes the global energy consumption while meeting QoS requirements which are the same to be respected in [14]. Integer Linear Programming is used to formulate the problem. A Branch-and-Bound mathematics' technique was exploited to get the optimal solution.

The remaining sections of this chapter are organized as follows. The second section recalls the model used for Ad Hoc wireless networks. The third section describes the proposed formulation with its specifications. The simulation setup and results are detailed in Sect. 4. The fifth section summarizes the work presented in this chapter.

2 Network model

In this chapter, we adopt the widely used transmitting power model for radio networks: $P_{ij} = (d_{ij})^{\alpha}$, where P_{ij} is the transmitting power needed for source node i to reach destination node j, d_{ij} is the distance between nodes i and j and α is a parameter typically taking a value between 2 and 4. This parameter depends on the characteristics of the communication medium.

The Ad Hoc wireless network is modeled by a graph $G = (V,E)$, where V is the group of N nodes and E is the group of undirected links between those nodes. A pair of nodes (i,j), where i and j belong to V, is in E if there exists a link between nodes i and j. Thus, a link (i,j) exists if j is in the Transmission Range (TR) of node i.

We assume also that each node can send signals to its neighbors in a conflict-free fashion. Thus, signal interference is not considered in this chapter. Some many protocols and methods have been proposed in the chapters [15, 16] taking into account interference in radio transmissions.

Let p_i denotes the transmitting power of node i. Each node is assumed to be able to adjust its power level, but not exceed the maximum power $P_{\max}$. The communication between two nodes depends on their distance. There is a link $(i,j) \in E$ if and only if $p_i \geq (d_{ij})^{\alpha}$ and $p_j \geq (d_{ij})^{\alpha}$. The total energy cost, P_{total} is the consumed energy by all active nodes in the network. The P_{total} formula is $P_{\text{total}} = \sum_{i=1}^{N} p_i$.

Let B_i, W_{ij}, λ_f, and D_f denote the total available bandwidth on node i, the total available bandwidth on the link (i,j), the traffic requests and Maximally allowed delay for each flow $f \in F$ respectively, where F is the set of flows. The problem of our concern is the following: given a set of N nodes and their locations, B_i and W_{ij} for each node i and each link (i,j) respectively, λ_f and D_f for each flow $f \in F$ respectively, we need to find the transmission power p_i for each node i and minimize the global energy consumption P_{total} of Ad Hoc wireless networks. To reach this objective, we have to minimize the number of used nodes to switch off a maximum number of unused nodes.

3 Problem Formulation

In order to formulate our problem, a set of decision variables are defined as follows:

- x_i: a Boolean variable which is equal to 1 if the node i is used, and 0 otherwise;
- X_{ij}: a Boolean variable which is equal to 1 if there is a link from node i to node j, and 0 otherwise;
- X_{ij}^f: a Boolean variable which is equal to 1 if the link (i,j) is used to route flow f, 0 otherwise;
- M_{ij}: The reserved bandwidth on the link (i,j);
- p_i: The transmitting power consumed by node i.

The following problem formulation is proposed as an optimal solution for routing a set of flows minimizing the number of nodes used while meeting the end-to-end QoS constraints in terms of delay and bandwidth.

$$\min \sum_{i=1}^{N} p_i$$

s.t.

$$X_{ij} = X_{ji}; \quad \forall (i, j) \in E \tag{1}$$

$$X_{ij} = \begin{cases} 1 \text{ if } d_{ij} \le \mathrm{TR} \\ 0 \text{ otherwise} \end{cases}; \quad \forall (i, j) \in E \tag{2}$$

$$M_{ij} \le W_{ij}; \quad \forall (i, j) \in E \tag{3}$$

$$0 \le p_i - \sum_{j=1}^{N} P_{ij}.X_{ij}^f \le P_{\max}; \quad \forall f \in F, \forall i \in V \tag{4}$$

$$\sum_{i=1}^{N} X_{ij}^f - \sum_{k=1}^{N} X_{ji}^f = 0; \quad \forall f \in F, \ \forall j \in V - \left(S_f, D_f\right) \tag{5}$$

$$\sum_{j=1}^{N} X_{ij}^f - \sum_{j=1}^{N} X_{ji}^f = \begin{cases} 1 & \text{if } S_f = i \\ -1 & \text{if } D_f = i \\ 0 & \text{otherwise} \end{cases}; \quad \forall i \in V \tag{6}$$

$$\sum_{f \in F} \lambda_f X_{ij}^f = M_{ij}; \quad \forall (i, j) \in E \tag{7}$$

$$\sum_{j=1}^{N} X_{ij}^{f} \leq 1; \quad \forall f \in F, \ \forall i \in V - D_f \tag{8}$$

$$\sum_{(i,j)\in E} X_{ij}^{f} \leq \Delta_f; \quad \forall f \in F \tag{9}$$

$$\sum_{f\in F} \sum_{i=1}^{N} \lambda_f X_{ij}^{f} + \sum_{f\in F} \sum_{k=1}^{N} \lambda_f X_{jk}^{f} \leq B_j; \quad \forall j \in V \tag{10}$$

$$x_i = \begin{cases} 0 \ if \ M_{ij} = 0 \\ 1 \text{ otherwise} \end{cases}; \quad \forall (i,j) \in E \tag{11}$$

$$u_i^f - u_j^f + |V| \,.X_{ij}^{f} \leq |V| - 1; \quad \forall f \in F, \forall (i,j) \in E \left[i \neq S_f, j \neq D_f \right] \tag{12}$$

$$\begin{aligned} \text{with } x_i &\in \{0,1\}; \quad \forall i \in V \\ X_{ij} &\in \{0,1\}; \quad \forall (i,j) \in E \\ M_{ij} &\in \Re^+; \quad \forall (i,j) \in E \\ u_i &\in \Re^+; \quad \forall f \in F, \forall i \in V - S_f \end{aligned}$$

The first constraint is symmetric's links. It corresponds to two directed links. Constraint 2 establishes the links between nodes in the network. Constraint 3 ensures that the reserved bandwidth on the link does not exceed the total available bandwidth on the same link. Constraint 4 defines the relations between the transmitting power consumed by each node and the decision variable X_{ij}^{f}. Constraint 5 ensures the flow conservation. It means that every intermediate node receiving a flow retransmits it to one of its neighbors. Constraint 6 ensures that all links on flow f should meet flows conservation. Constraint 7 focuses on aggregation: the reserved bandwidth on a link is the sum of the reserved bandwidth by each flow on that link. Constraint 8 represents the single path constraint that is a flow should not be split into several paths. Constraint 9 ensures that the hop-count for each flow does not exceed the pre-specified bound. Constraint 10 ensures that the total transmission and reception of flows at a node do not exceed the total available bandwidth on this node. Constraint 11 determines the state of each node (used or unused). The last constraint is added from the Traveling Salesman Problem [17] to prevent any loops in the final solution.

4 Simulation Results

In this section, simulation experiments are conducted to highlight the pertinence of the proposed solution in comparison with the performance of another model that is Jia et al. [14]. On the other hand, the performances of flows' aggregation are shown in terms of global energy consumption. The results obtained are presented and discussed by performing the proposed and Jia's models in the case of non-splittable traffics.

Random connected graphs are generated and the number of nodes used is counted when a set of flows is inserted. Then each flow has requirements in terms of delay and bandwidth, the source S_f and destination D_f of flow are selected randomly so that they are more than one-hop neighbors. *Thirty* (30) nodes random networks were carried out in a two-dimensional free space area of $40 \times 40\ \text{m}^2$ where nodes have a Transmission Range TR = 20 m and links have a capacity of $W_{ij} = 5 \cdot 10^3$ units. All network nodes have the same total bandwidth $B_i = 5.5 \cdot 10^4$ units.

To assign traffic requests λ_f for each flow, a random function of a normal distribution with variance equal to $0.5\mu_\text{m}$ was used (μ_m is the mean value of the normal distribution function). The average bandwidth requirement per flow is μ_m. The maximum value which we set for the flow hop-count is $\Delta_f = 2N/15$. The parameter α in the power model is set to 2 and 4, i.e.,$P_{ij} = (d_{ij})^\alpha$, for $2 \leq \alpha \leq 4$.

The first experiment is to compare the resulting topologies for the proposed solution with another model. Four flows are inserted in the input network. The delay constraint Δ_f for the flow is set to four. The average traffic amount per flow is $0.02B_i$. The routing information for each model is reported in Table 1.

Table 1 shows a given case where four flows are routed using Jia's and proposed models. When routing is performed with Jia's model, the routes are more scattered and thus more nodes have to be used. This is due to Jia's model which does not take into account the constraint of total available bandwidth on links. Its problem is just to find the network topology such that all flows can be routed and the total energy consumed by all nodes is minimized. With the proposed solution based on flows' aggregation, the same set of flows have routed over a limited number of nodes while providing the same QoS. In the example, it can be switched off around 26% of nodes used with Jia's model.

Another interesting point to note is that the value of the objective function P_total is different in both models as displayed in Table 1. This is due to how the power cost of nodes is calculated in both formulations. In the current study's formulation, the power cost is the corresponding element from the power matrix (constraint 4 in Sect. 3) which defines the relations between the transmitting power p_i and the variable X_{ij}^f. In Jia's formulation, the same constraint is used, but using a variable X_{ij}.

Table 1 Flows and routing information

S_f	D_f	λ_f	Jia's model	Proposed model
2	19	1100.8	$2 \to 24 \to 15 \to 27 \to 19$	$2 \to 5 \to 15 \to 16 \to 19$
7	11	1097.6	$7 \to 6 \to 15 \to 27 \to 19$	$7 \to 6 \to 15 \to 16 \to 11$
16	18	1101.4	$16 \to 11 \to 4 \to 12 \to 18$	$16 \to 18$
1	8	1102.1	$1 \to 22 \to 30 \to 8$	$1 \to 8$
P_total [J]			$2.04 \cdot 10^3$	$0.8 \cdot 10^3$

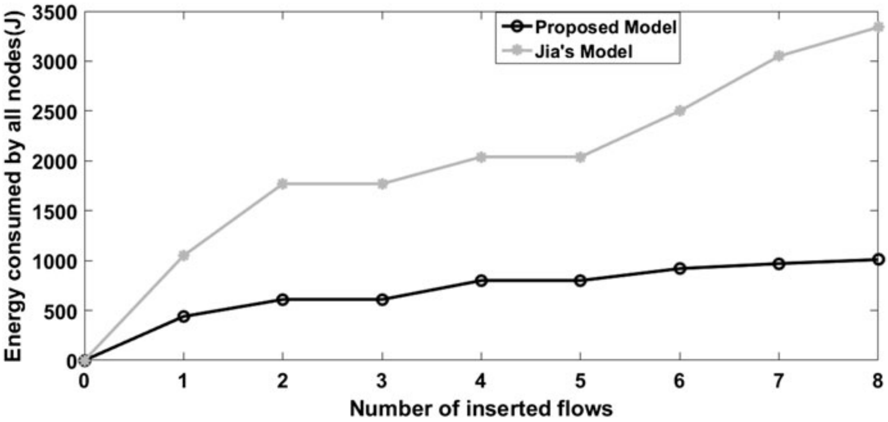


Fig. 1 Global energy consumed when the number of flows increases

In the second and last experiments, the present study's proposed model is compared with Jia's model, in terms of global energy consumption and of a number of nodes that can switch off as a function of the number of inserted flows in the network, respectively.

Figure 1 plots the total energy consumption in the network according to the number of inserted flows. It can be noticed that the graph associated with the proposed model is under that of Jia's model. The total energy consumption of both solutions increases with the increase of the number of inserted flows. The reason is when the number of flows increases; the network topology must be funded such that all flows can be routed with respecting the QoS provisions. The latter caused much more energy costs. Furthermore, the performance of the attained solution is better than Jia's model.

As it is shown in Fig. 2, the number of unused nodes decreases according to the number of inserted flows. The results obtained by both models are expected because inserting a new flow will force its source and its destination to be activated; this means the gain is achieved just on intermediate nodes and how routes pass through already used nodes. This is corroborated by the case where the number of inserted flows is equal to four as shown in Table 1. From this table, it can be noticed that the number of source and destination nodes is the same in both models, and the number of intermediate nodes in Jia's model is equal to eight, unlike the current study's proposed model which is equal only to three. That means that the newly proposed model designed for this study is able to turn off 62% of the used intermediate nodes in the case of Jia's model.

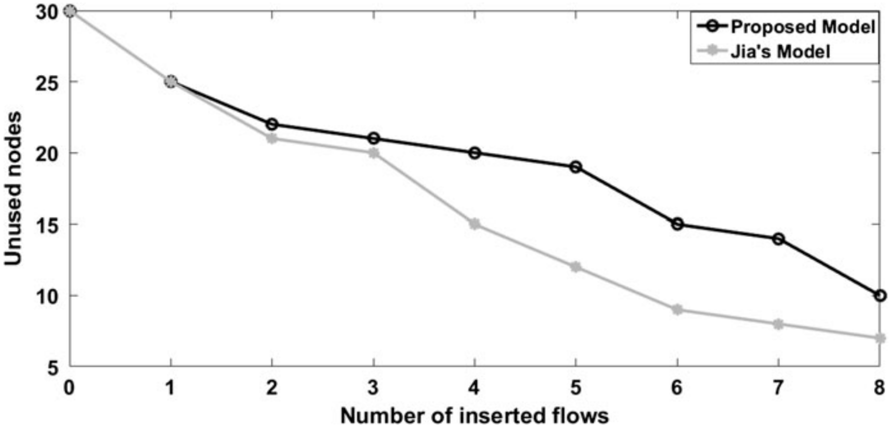


Fig. 2 Number of unused nodes when the number of flows increases

5 Conclusion

In this chapter, we discussed a proposed solution based on flow's aggregation technique which aims at turning off more nodes used to route inserted flows and then minimizing the global energy consumption in Ad Hoc wireless networks. The solution is modeled as an Integer Linear Program, with a set of linear constraints to maximize the number of nodes unused in the network while meeting the QoS provisions of flows inserted in terms of delay and Bandwidth. A Branch-Bound algorithm has been used to solve it to optimality. It has shown by simulation that the proposed solution can significantly increase the global energy saving in the network compared to another Jia's solution.

References

1. Meena, K. S., & Vasanthi, T. (2014). Reliability analysis of mobile ad hoc networks using universal generating function. *Quality and Reliability Engineering International, 32*(1), 111–122.
2. Liu, J., & Kato, N. (2016). A Markovian analysis for explicit probabilistic stopping-based information propagation in postdisaster ad hoc mobile networks. *IEEE Transactions on Wireless Communications, 15*(1), 81–90.
3. Dharmaraja, S., Vinayak, R., & Trivedi, K. S. (2016). Reliability and survivability of vehicularad hoc networks: An analytical approach. *Reliability Engineering and System Safety, 153*, 28–38.
4. Ibukunola, A. M., Oludayo, O. O., & Abiodun, M. (2013). Minimizing energy consumptionin wireless ad hoc networks with meta heuristics. *Procedia Computer Science, 19*, 106–115.
5. Feng, W., Alshaer, H., & Elmirghani, J. M. H. (2009, August). Optimization of energy consumption in rectangular ad-hoc wireless networks, Xian, China (pp. 1–5).

6. Thenmozhi, D. S., & Lakshmipathi, R. (2010). Highly ensured QoS routing in mobile adhoc networks based on multiple constraints. *IJCA*, *8*(4), 23–29.
7. Bang, W., Qiang, Y., Laurence, T. Y., & Chunsheng, Z. (2017). On minimizing energy consumption cost in green heterogeneous wireless networks. *Computer Networks, 129*, 522–535.
8. Karmakar, S., Koutris, P., Pagourtzis, A., & Sakavalas, D. (2017). Energy-efficient broadcasting in ad hoc wireless networks. *Journal of Discrete Algorithms, 42*, 2–13.
9. Laube, A., Martin, S., Quadri, D., & Al Agha, K. (2016). Optimal flow aggregation for global energy savings in multi-hop wireless networks. Ad-hoc-Now. In *Lecture notes in computer science* (Vol. 9724, pp. 124–137). Springer.
10. Ramanathan, R., & Rosales-Hain, R. (2000). Topology control of multi-hop wireless networks using transmit power adjustment. In *IEEE Xplore* (pp. 404–413).
11. Marsan, M. A., Chiasserini, C. F., Nucci, A., Carello, G., & Giovanni, L. D. (2002). Optimizing the topology of Bluetooth wireless personal area networks. In *IEEE Xplore.*
12. Charu, V. A. (2013). *A quality of service analysis of energy aware routing protocolsin mobile ad hoc networks*. IEEE.
13. Chaudhry, R., & Tapaswi, S. (2018). Optimized power control and efficient energy conservation for topology management of MANET with an adaptive Gabriel graph. *Computers and Electrical Engineering, 72*, 1021–1036.
14. Xiaohua, J., Deying, L., & Dingzhu, D. (2004). QoS topology control in ad hoc wireless networks. In *INFOCOM 2004* (pp. 1264–1272). IEEE.
15. Gawedzki, I., Benfattoum, Y., Martin, S., & AlAgha, K. (2010). I2ASWP: Routing with intraflow interference consideration in ad hoc network. Research report, CNRS University of Paris Sud-LRI.
16. Walrand, J., Jia, Z., Gupta, R., & Varaiya, P. (2005, June). Bandwidth guaranteed routing for ad hoc networks with interference consideration. In *IEEE Symposium.*
17. Tucker, A., Miller, C., & Zemlin, R. (1960). Integer programming formulation of travelling salesman problems. *Journal of Alternative and Complementary Medicine, 7*, 326–329.

Management of Enterprise's Advanced Development for Its International Competitiveness

Viktoriia Shkola, Oleksandra Olshanska, Tetiana Kasianenko, and Maryna Domashenko

Abstract This chapter is devoted to the exploring the theoretical fundamentals of the enterprise's advanced development management in order to provide its international competitiveness. The essence of international competitiveness and production drivers of leading innovative development are examined in the context of the innovation advancement paradigm. The system approach to evaluation of the economic security management mechanism based on the marketing principles by means of correlating changes in the environmental safety level with the costs of forming, the operation maintenance, as well as the mechanism development are proposed. Assessment of the directions and variants for the enterprise innovative development is proposed to be determined by the indicators of expected environmental and economic efficiency for enterprises, consumers and society as a whole, as well as the expected commercial efficiency of the innovation development direction and variant during the ecological and economic cycle of innovation. The comprehensive approach to estimation of the enterprise's economic security taking into account internal and external factors influencing the economic security is proposed. Not only theoretical significance has the obtained results, but also both the rationale for the threat of responsible investment in innovations and the applied value for market participants in terms of investment decisions making.

Keywords Management · Competitiveness · Enterprise · Innovation

V. Shkola (✉) · T. Kasianenko · M. Domashenko
Sumy State University, Sumy, Ukraine
e-mail: v.shkola@macro.sumdu.edu.ua; m.domashenko@macro.sumdu.edu.ua

O. Olshanska
Kyiv National University of Technology and Design, Kyiv, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_11

1 Introduction

The fourth industrial revolution and new technologies stimulate the development of new production technologies and business models that fundamentally transform production. The speed and scale of technological changes, coupled with the emergence of other trends, complicate the task of developing and implementing industrial strategies that promote productivity and inclusive growth. Moreover, recent changes have identified the paradigm of low-cost export competitiveness as a vehicle for growth and development at risk. Awareness of this causes the need to address the issue of strengthening existing market positions and support long-term competitive advantages in its activities of the enterprise in this new production paradigm. Besides, the study of the factors and conditions that have the greatest impact on the transformation of production systems, further evaluating their readiness for the future, will enable appropriate measures to be taken to overcome potential gaps in their readiness for future production, increase their competitiveness in the future and increase the level of economic security of enterprises.

2 The Essence of International Competitiveness in the Context of the Innovation Advancement Paradigm

The notion of competitiveness was researched in the works of national and foreign scholars [1–12]. Thus, competitiveness is defined as an opportunity to operate in open market conditions and to remain profitable for a long time [13].

Fathutdinov [12] explains competitiveness as the property of an object that is characterized by a degree of real or potential satisfaction of specific needs in comparison with similar objects in this market. Yaroshenko [14] defines competitiveness as the ability of an enterprise to operate in a market economy, gaining profits sufficient for scientific and technical improvement of production, employee stimulation and product quality support at a high level. It means that the competitiveness of production is mainly reduced to competitiveness of products. A similar view is adhered to [15], considering competitiveness as the ability to provide goods to a greater extent than other products, the level of satisfaction with consumers' needs and demands at all stages of their life cycle.

In the study [16] the competitiveness of the enterprise is proposed to be understood as the ability of the enterprise to study demand (market), the ability to design, fabricate and implement a commodity that, by its properties, will better meet the needs of consumers than competitors. It is characterized by its ability to adapt to environmental conditions. Separately by the author [16] is defined the concept of international competitiveness as an achievement of the enterprise's competitive advantages in rivalry in the international market.

Given the pace of economy globalization, the consequence of which is the expansion of market boundaries for the national framework, the term of international competitiveness, which is more capacious than its own content, allows you to comprehensively capture and reflect the essence of ongoing processes. In general understanding international competitiveness reflects the enterprise position, both in the national and foreign markets, compared to rivals, as well as retention facilities market positions in conditions of increasing competition from national, international and foreign enterprises.

Accepting that the competitiveness of an enterprise in the foreign market is determined by the level of competitiveness of its products, the conformity of the product nomenclature to the needs of the market, as well as the efficiency of its marketing system in the foreign market, which, in particular, determines the directions of development of the enterprise, taking into account its innovative potential, technical, economic and organizational conditions for the creation, production and marketing of high quality products that meets the requirements of specific groups of consumers, should be noted separately in the prominence of another component of international competitiveness—the competitiveness of production systems, which should be considered as technical and technological advantages of the enterprise, which ensure the possibility of production at low cost.

Considering the competitiveness of production systems, which corresponds to the competitiveness paradigm of low-cost export production [17], it should be agreed with the author's opinion [15] that it is necessary to adhere to a certain compromise between the level of satisfaction and the cost of consumption, which should correspond to this level, that is, to be fair, first of all from the point of view of consumers.

Consequently, taking into account all of the above, ensuring international competitiveness of the enterprise is possible provided the continuous and steady implementation of various types of innovations, that is, the fact at the expense of a leading innovative development of the enterprise. In the research [18], leading development is defined as a radical, innovative and cyclical process of continuous improvement of the enterprise potential and the search for promising directions of its realization, which results in radical changes of the enterprise.

It should be noted separately that in view of the fact that Ukraine, like most other countries of the world, has accepted the course on ensuring sustainable development, it is expedient to introduce innovations that at all stages of their life cycle there will not be ecodestructive impact on the environment or there will be an ecoconstructive impact. Taking into account this, another important characteristic of the leading innovative development is its compliance with the principles of the Concept of Sustainable Development.

Consequently, leading innovative development will be considered as a process of introducing innovations that are ahead of scientific and technological development, the formation on this basis of the production system and enterprise business portfolio of goods and services, which are at different stages of the life cycle and satisfy not only the existing, but also potential needs that in general will ensure the stability of the enterprise competitive position in the international market and the possibility of their growth, as well as high level of its economic and ecological security. Comprehensive approach to providing ecological security is explored in chapter [19]. Issues of enterprise's economic security management at the foreign marketing are considered in the article [20].

3 Production Drivers of Leading Innovative Development: The Essence and Structure

According to the report [17], the countries' success in the conditions of the current production changes is determined by the production drivers that are explained as the key factors that position the country to take advantages of the fourth industrial revolution for accelerating the transformation of national production systems.

The main factors that position the enterprise for using new technologies and opportunities in the future are [17]: technologies and innovations, human capital, global trade and investments, institutional frameworks, stable resources and demand environment.

Technologies and innovations reflect the level of innovations implementation and commercialization that are potentially used in production, as well as the level of development and infrastructure security to support the new technologies implementation in production. The model of the technological innovations in the form of a discrete dynamic system is explored in chapter [21].

Human capital is the ability to respond to changes in the labor market caused by the Fourth Industrial Revolution, considering both the current workforce and the long-term ability to develop skills and talent in future workforce.

Global trade and investments is the participation in the international trade for facilitating the exchange of products, knowledge and technologies; the establishment of global connections, the existence of financial resources for development investment related to production, as well as the infrastructure quality for the activity provision connected to production.

Institutional structure is effectiveness of state institutions, rules and regulations in the field of technological development, new enterprises and leading production. Especially it should be delighted the role of institutions in field of environment protection and waste management including hazardous waste (in more detail see [22]).

Sustainable resource is the production impact on the environment, including the usage of natural resources and alternative energy sources.

The demand environment is access to external and local demand for production, the complexity of the consumer base, as it can lead to diverse activities of industry and new products.

Besides above mentioned factors, on authors thought, the crucial driver of the enterprise's innovative development is competition. As it is explored in chapter [2], competition has a positive effect on long-term economic growth. In conditions of growing competition those enterprises which follow the strategy of leading innovative development are in a stronger market position.

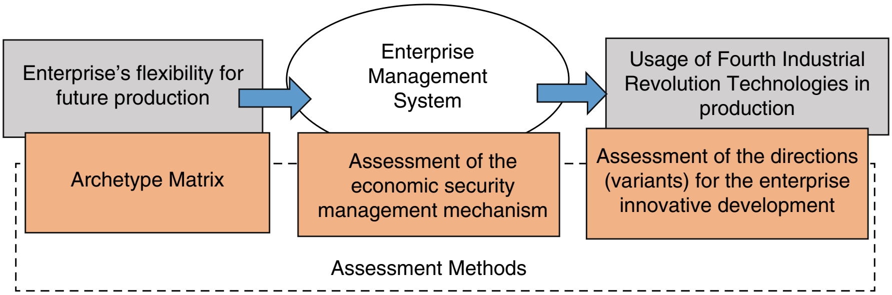


Fig. 1 Model of production transformation (developed by the authors)

Leading development involves the production transformation to meet the consumers' needs and demands at the international markets in the conditions of global competition, which will ensure its economic security.

Model of production transformation from the study phase of the enterprise's flexibility for changing, which is determined on the basis of the archetype matrix, till the implementation of the changes in the enterprise, presented by the author in Fig. 1, shows the role of the enterprise management system in ensuring its development, and consequently the growth of its ecological and economic efficiency.

In the given model, the management system is the key link between the production system and the market, and therefore it is considered to be the ability of the enterprise to ensure its constant competitiveness on the basis of: quality management and production organization and enterprise marketing as a whole; effective usage of human capital; application of the latest marketing tools; rational usage of available financial and material resources.

The management system, which has the task to ensure the international enterprise competitiveness in the future under the current conditions, is represented by the factor of "*Human Capital.*"

A detailed Ukrainian profile [17], which allows identifying concrete opportunities and challenges for countries, as they are oriented toward the production future, was allowed to conclude that human resources and human capital are a strong position of Ukraine. At the same time, the possibility of using other growth factors depends on the efficiency of the management system of the economic system and its marketing potential. Taking into account the fact that the enterprise is the basic unit of the national production system, in our opinion, it is appropriate to study the key factors of the enterprise efficiency growth.

We suggest implementation of the assessment of the economic security management mechanism based on the marketing principles by means of correlating changes in the environmental safety level (as a result of this mechanism functioning) with the costs of forming, the operation maintenance, as well as the mechanism development. Target function and restrictions that correspond to the above-described description concerning mechanism assessment of the economic security management, are presented as a system:

$$\begin{cases} \sum_{j=1}^{m} \frac{R_j}{C_{1j}+C_{2j}+C_{3j}} \to \max, \\ R_j = f\left(\Delta K_{nj}, \mathrm{UC}_j, \Delta K_s\right), \\ 0 \le \Delta K_{nj} \le 1 \end{cases} \tag{1}$$

where R_j—the result of the economic security management mechanism according to the j-type activity ($j = 1,...,m$), currency unit; C_{1j}, C_{2j}, C_{3j}—the mechanism action costs (respectively: functioning, operation maintenance, development) according to the j-type activity ($j = 1,...,m$), currency unit; ΔK_{nj}—change indicator of the economic security level according to j-type activity; UC_j—unit costs caused by changes of the economic security level to the elementary measurement unit on the scale $0 \le \Delta K_{nj} \le 1$, currency unit/production unit; ΔK_{s}—change indicator of the enterprise economic security level caused by the synergetic effect of this mechanism.

4 Assessment of the Directions (Variants) for the Enterprise Innovative Development

An important criterion for selecting areas, and within them, for innovation development options is the efficiency level during the ecological and economic cycle of innovation that covers the life cycle of innovation (LCI) and customization cycle of innovation (CCI), which is proposed to be determined by the indicators of expected environmental and economic efficiency for enterprises, consumers and society as a whole, as well as the expected commercial efficiency of the innovation development direction (variant).

The expected environmental and economic efficiency should be determined in the assessment of the directions (variants), reflecting the results of innovation activities for the manufacturer, consumers and society as a whole, and the expected commercial efficiency given the market optimality, taking into account the interests and economic benefits of the innovative enterprise.

The authors propose to consider the ecological and economic efficiency E as a indicators system which reflects the general results and costs for the implementation of the chosen innovation development direction (variant) for innovative enterprises, consumers and society in general, including both direct results and costs, and external effects in related sectors of the economy, including environmental and social eco-innovations during the ecological and economic cycle. The following formula is proposed for its calculation:

$$E = \frac{\sum_{t=1}^{T} \left(R_{\mathrm{E}t} \cdot g_t - E_{\mathrm{E}t} \cdot k_t\right) \cdot s_t^m \cdot (1+r)^{-t}}{\sum_{t=1}^{T} E_{\mathrm{E}t} \cdot k_t \cdot (1+r)^{-t}}, \tag{2}$$

where R_{Et}—the expected direct (variant) ecological and economic result of the enterprise innovative activity in the t-period, currency unit; E_{Et}—expected expenses of the t-period for realization of measures in the direction (variant) under consideration, currency unit; r—discount rate, relative unit; k_t—correction coefficient taking into account the level of expenses change for measures implementation within the innovation development direction depending on the type of implemented innovation, the corresponding stage of its evolutionary development and the priority ecologization concept of the enterprise innovation activity; g_t—coefficient of society's flexibility for the innovation acceptance; s_t^m—coefficient that takes into account the synergy of social, economic and environmental effects and the synergy of adding the ecological and economic effect in each t-period to the previous ones (with the intensifying nature of action m takes on a meaning $+1$, with the decreasing one goes to -1); T—the ecological and economic cycle duration, years. The value of the s_t, k_t, g_t coefficients and m index are established on the basis of empirical analysis of the output data. It was developed the table of values for the k_t, g_t coefficient, the application of which allows to increase the accuracy of predictive calculations taking into account exo- and endogenous factors.

The expected commercial efficiency of the innovation development direction (variant) is determined by the criterion of the market optimality and by the indicators of net present value NPV, internal rate of return IRR, profitability index PI and payback period PP taking into account risk in three forecast variants (pessimistic, most probable and optimistic).

The NPV calculation taking into account the R_t risk is proposed to be carried out according to the formula:

$$\mathrm{NPV} = \sum_{t=1}^{T} \left(\frac{E_t}{(1+r)^t} - \sum_{j=1}^{d} \frac{RV_{tj}}{(1+r)^t} \right), \tag{3}$$

where E_t—the expected effect from the direction (variant) realization in the t-period, currency unit; RV_{tj}—the expected absolute value of the loss from the j risk type in the period t, $j \in [1; d]$, currency unit; r—discount rate, relative unit; t—the realization period of the direction taking into account the eco-economic innovation cycle (T), the year. The methodology for risk assessment of innovation activity is presented in [23–25]. The calculation of IRR, PI, PP is carried out according to the procedure outlined in [26, 27].

The market optimality of the variant is based on the comparison of the evolutionary development stages of innovation (S_1), enterprise (S_2) and market (S_3) (see in more detail [28]).

The matrix for combination of the innovation, enterprise and market development stages (Table 1) has been developed to determine the market optimality of the innovation development variant.

It allows determining the realizing enterprise possibility of the investigated variant and the expediency of investing investment resources taking into account possible refinement and approximation to the existing market situation and possible

Table 1 The matrix for combination of "the innovation, enterprise, and market system" (developed by the authors)

S (S_1, S_2, S_3)		The development stage of innovation, S_1											
		I			II			III			IV		
		The development stage of enterprise, S_2											
		I	II	III	I	II	III	I	II	III	I	II	III
The development stage of market, S_3	0	A	A	C	E	E	E	E	E	E	E	E	E
	I	C	C	C	C	B	C	C	B	C	E	E	E
	II	E	E	E	D	B	B	D	B	B	D	B	B
	III	E	E	E	D	D	D	D	C	A	D	C	A
	IV	E	E	E	E	E	E	D	B	A	C	A	A

A—the combination is absolutely market-optimal; B—the combination is market-optimal provided compliance with additional recommendations; C—the combination may be optimal under certain conditions and provided compliance with certain recommendations; D—the combination is acceptable under certain conditions; E—combination is unacceptable

scenarios of its development based on forecasting changes in the enterprise marketing environment and market conditions.

Establishment of absolute market optimality (segment A of Table 1) allows you to go directly to the assessment of the commercial effectiveness of the option, taking into account the risks identified in accordance with the specificity of the analyzed version and implemented innovation. When the combination of the "innovation-enterprise-market" system is unacceptable (segment E of Table 1), it is necessary to determine which of the components makes it impossible to implement the investigated variant and determine the possibility of its adjustment. The presence of certain opportunities for the enterprise to implement the research variant (segments B, C, D of Table 1) necessitates an additional analysis based on the specific risks of the analyzed variant and the implemented innovations.

5 Estimation of Economic Security of the Enterprise

For the complex consideration of internal and external factors influencing the economic security of the enterprise when entering the external market, a three-component indicator of the level of the enterprise economic security K_{nj} is proposed:

$$K_{nj} = f\left(I, P_n, D_j\right),$$
$$I, P_n, D_j = \begin{cases} 1, & \text{if } I, P_n, D_j \geq I_{\text{suf}}, P_{\text{suf}}, D_{\text{suf}}, \\ 0, & \text{if } I, P_n, D_j \geq I_{\text{suf}}, P_{\text{suf}}, D_{\text{suf}}, \end{cases} \tag{4}$$

where I—value of estimating the potential of an enterprise for the implementation of an enterprise; P_n—country's risk level for a country n; D_j—indicator of the level of market opportunities of the enterprise for the implementation of j type of activity;

Table 2 Values of indicators for assessing the level of economic security of the enterprise (developed by the authors)

The level of enterprise potential, I		Country risk level, P_n		Level of market opportunities of the enterprise, D_j	
Value	Characteristic	Value	Characteristic	Value	Characteristic
$0{,}95 \leq I \leq 1$	Absolutely safe	$75 < P_n \leq 100$	Low	$0{,}75 \leq D_j \leq 1$	High
$0{,}75 \leq I < 0{,}95$	Acceptable	$30 < P_n \leq 75$	Middle	$0{,}5 \leq D_j < 0{,}75$	Middle
$0{,}5 \leq I < 0{,}75$	Unstable				
$0{,}25 \leq I < 0{,}5$ $0 \leq I <{,}25$	Low Inadmissible	$0 \leq P_n \leq 30$	High	$0 \leq D_j < 0{,}5$	Low

I_{suf}, P_{suf}, D_{suf}—sufficient value of indicators I, P_n, D_j (see Table 2 where area of the sufficient indicators is highlighted by the eclipse).

The value of the indicator I is proposed to be determined by the formula:

$$I = \sum_{i=1}^{n} B_i \cdot \frac{F_i}{G_i}, \tag{5}$$

where B_i—the ponderability coefficient of the i security component; F_i—actual value of the i security component; G_i—sufficient value of the i security component; n—the number of security components.

The main components of potential I: financial (characterizes the financial sustainability of the enterprise); industrial-technical (characterizes the efficiency of using the main enterprise production assets); intellectually-cadre (shows the efficiency of the labor resource usage); marketing (reflects the stability of the enterprise in the sectoral market); legal (characterizes the degree of the enterprise interests and its workers protection); interface (characterizes the reliability of interaction with contractors); innovative-technological (defines the technological potential of the enterprise); raw materials and energy (reflects the supply of raw materials and energy resources); ecological (characterizes the ability of the enterprise to carry out production activities in accordance with environmental standards).

The country risk level P_n is determined on the basis of the BERI index, which is calculated four times per year using the expert judgment estimation method. The structure of the analyzed parts of the indicator includes: efficiency of the economy; level of political risk; level of indebtedness; availability of bank loans; availability of short-term financing; availability of long-term loan capital; likelihood of the occurrence of force majeure circumstances; the level of creditworthiness of the country; the amount of outstanding debt repayment obligations [29].

There are a number of methodological approaches, the most common of which are: SWOT-analysis [29, 30], M. Porter's strategies model [30–33], Peter T. FitzRoy's competitive advantage matrix, Boston Consulting Group (BCG) matrix [29, 30], the GE-McKinsey matrix [30, 34], the Gap-analysis [29, 30], the Shell-DPM matrix [30], the matrix of R. Cooper [30]. The analysis of their advantages and disadvantages has allowed to reveal insufficiently complete objectivity of the results

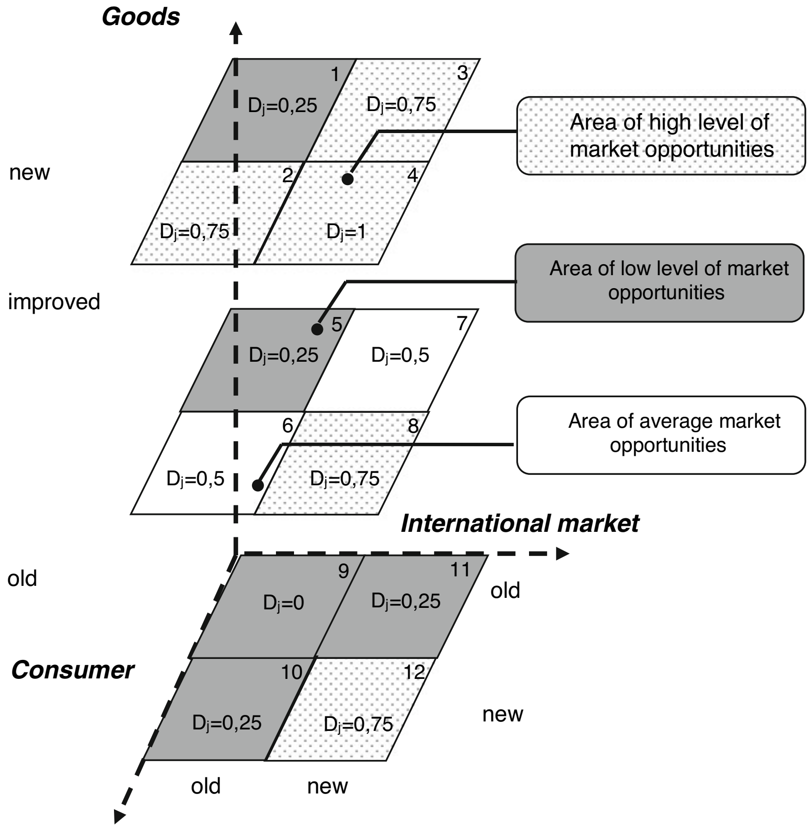


Fig. 2 The matrix of levels of the enterprise market opportunities (developed by the authors)

obtained on their basis, since all of them are based on qualitative or quantitative estimates of individual components of the competitiveness of the enterprise itself and its closest competitors. In order to solve this problem, the authors offered a methodology for assessing the enterprise market opportunities, based on the comprehensive study of the system "product–market–consumer" (see Fig. 2).

Characteristics of the levels of the enterprise market opportunities, allocated using the corresponding matrix (see Fig. 2), is presented in Table 3

Theoretically possible there are eight values of the three-component index K_{nj} corresponding to the four zones of economic security of the enterprise international economic activity (Table 3). For each of the selected zones of economic security, competitive strategies are proposed. Their characteristics are given in Table 4. The groups of activities that have relevance to the Book Index are formed within the framework of each strategy (Table 5). The implementation of these groups can be complex or prioritized, based on the enterprise financial capabilities. It should be

Table 3 Characteristics of market opportunities (developed by the authors)

Characteristics	Area of market opportunities		
	High	Middle	Low
	Quadrant 2, 3, 4, 8, 12	Quadrant 6, 7	Quadrant 1, 5, 9, 10, 11
Purpose of marketing policy	Optimization of the sales network abroad	Optimization and sales intensification	Sales optimization
Purpose of commodity policy	Extension of the range, basic innovation	Improved modifications, a wide range of models	Elimination of overweight
Purpose of the promotion policy	Consumers' belief in the need to purchase products	Maximum consumer awareness	Maintaining the distinctive advantages of products
Pricing strategy	Establishing a price at the level of compensation of cost and contractual prices	Setting the price at the level of cost, sliding falling and contract prices	Setting of contract prices, flexible prices and preferential prices. Price elasticity

Table 4 Zones of economic security of the enterprise (developed by the authors)

(I,P_n, D_j)		The level of the enterprise potential,			
		sufficient		insufficient	
		Country risk level, P_n			
		sufficient	insufficient	sufficient	insufficient
Level of market opportunities, D_j	sufficient	(1;1;1)	(1;1;0)	(1;0;1)	(1;0;0)
	insufficient	(0;1;1)	(0;1;1)	(0;0;1)	(0;0;0)

– Zone I (absolute security);
– Zone II (acceptable security);
– Zone III (unstable security);
– Zone IV (inadmissible security).

noted that the enterprise can apply different strategies at different markets and for different activities simultaneously.

The choice of measures aimed at ensuring an adequate level of economic security of the enterprise is proposed to be carried out in the following sequence: (1) identification of indicators for assessing the level of economic security (see Table 2); (2) determination of zones of economic security of the enterprise based on the values of the three-component indicator (see Table 4); (3) choice of strategies for ensuring economic security of the enterprise in the world market (see Table 5); (4) selection and implementation of measures to ensure the economic security of the enterprise (see Table 5); (5) assessment of the effectiveness of measures to ensure the economic security of the enterprise.

Table 5 Strategies for ensuring economic security of the enterprise foreign economic activity (developed by the authors)

Zone	Strategy	Characteristics of the strategy	K_{nj}	Action groups
Zone I	Support	Maintaining a sufficient level of economic security, preventing the emergence of threats to the economic interests of the enterprise	(1;1;1)	1. Support for core production assets; 2. Effective use of resources; 3. Attraction of investments; 4. Preparation of foreign trade agreements; 5. Supporting a sufficient amount of sales and ensuring its stable growth; 6. R & D support; 7. Improvement of conditions for trade in export goods
Zone II	Strengthening	Conduct measures to strengthen one of the indicators of a three-component indicator of enterprise economic security whose value is insufficient	(1;1;0)	1. Analysis of market opportunities; 2. Formation of own sales network abroad; 3. Market monitoring; 4. Improving the quality of goods; 5. Extension of product range; 6. Development of foreign trade ties
			(1;0;1)	1. Risk Insurance; 2. Search for new sales markets
			(0;1;1)	1. Modernization of production; 2. Motivation of the personnel; 3. Reduced resource costs; 4. Increasing the competitiveness of products; 5. Increase sales; 6. Involvement of new resource providers on more favorable terms
Zone III	Adaptation	Adaptation of the type of activity to a certain market and vice versa, as well as adaptation of the components of economic security to the requirements of the environment	(1;0;0)	1. Search for other sales markets; 2. Updating the product range; 3. Improving product quality; 4. Drawing up of new contracts
			(0;0;1)	1. Improvement of results of financial and economic activity; 2. Budget financing; 3. Search for other sales markets; 4. Risk Insurance
			(0;0;1)	1. Improvement of the results of financial and economic activity; 2. Adaptation of main productive assets to the type of activity of the enterprise; 3. Promotion of product sales for export; 4. Tax privileges; 5. Budget financing; 6. Increasing the competitiveness of products; 7. Involvement of new suppliers of material resources at more favorable conditions for the enterprise
Zone IV	Change	Conduct changes in the type of activity and the external market	(0;0;0)	1. Formation of a new portfolio of options for Foreign Economic Activity; 2. Search for new sales markets

6 Conclusions

The developed theoretical and methodological approach to the management of innovative progress allows to take into account: positive results determined by changes in the ecodestructive load on the environment and recipients; the possible enterprise cost level for the implementation of the direction (option) depending on the type of innovation, the stage of the ecological and economic cycle of innovation, which increases the level of substantiation of management decisions at the initial stages; market optimality of the direction (option), which allows to reduce investment risks at the early stages and determine its market perspective in the long- and short-term benefits.

The theoretical and methodological approach suggested by the authors to the formation of strategies, as well as to the system of measures ensuring the enterprise economic security, allows to increase the overall level of the enterprise economic security, stimulate the management of the enterprise to implement measures aimed at the rational use of the enterprise resources, the choice of optimal market for distribution of their products, as well as the choice of the kind of activity that will fully meet the existing needs of the international market. Further research should be aimed at developing approaches to assessing the effectiveness of the proposed strategies and measures to ensure the enterprise economic security.

Acknowledgment This chapter was written according to budget money from the Ministry of Education and Science of Ukraine, given to develop scientific-research topic № 0120U102003 "Process of formation of the novel ecologically safe fertilizers with prolonged action based on the phosphorite deposits raw material."

References

1. Budd, L., & Hirmis, A. K. (2004). Conceptual framework for regional competitiveness. *Regional Studies, 38*(9), 1015–1028. https://doi.org/10.4324/9780203607046
2. Charles, V., & Zegarra, L. F. (2014). Measuring regional competitiveness through Data Envelopment Analysis: A Peruvian case. *Expert Systems with Applications, 41*(11), 5371–5381. https://doi.org/10.1016/j.eswa.2014.03.003
3. Yankovy, O. (2013). *Competitiveness of the enterprise: Assessment of the level and enhancement directions*. Atlant. (in Ukrainian).
4. Calabrese, A., Campisi, D., Capece, G., Costa, R., & Di Pillo, F. (2013). Competiveness and innovation in high-tech companies: An application to the Italian biotech and aerospace industries. *International Journal of Engineering Business Management, 5*(1), 40. https://doi.org/10.5772/56755
5. Mulatu, A. (2016). On the concept of 'competitiveness' and its usefulness for policy. *Structural Change and Economic Dynamics, 36*, 50–62. https://doi.org/10.1016/j.strueco.2015.11.001
6. Levytska, A. O. (2013). Methods for assessing competitiveness: National and foreign approaches to classification. *Mechanism of Economic Regulation, 4*, 155–162. (in Ukrainian).
7. Legkyi, V. I. (2015). Stages of assessment competitiveness of enterprises and its improvement types of strategies. *Scientific and Information Bulletin I van-Frankivsk University named after Danylo Halytsky Economics, 11*, 267–273. (in Ukrainian).

8. Mavrova, V. V. (2016). Enterprise's competitiveness management in Ukraine. *Young Scientist 7*, 78–80. (in Ukrainian).
9. Sivachenko, I. Y., Kozak, Y. G., & Logvinova, N. S. (2010). *Enterprise's international competitiveness management*. Center for Educational Literature. (in Ukrainian).
10. Porter, M. E. (2018). *On thinking about deregulation and competition. The telecommunications revolution: Past, present and future* (pp. 39–44). Routledge. https://doi.org/10.4324/9781351115704
11. Faskhiev, H. A. (1998). Assessment of the new equipment competitiveness. *Marketing, 6*, 25–35. (in Russian).
12. Fathutdinov, R. A. (2000). *Competitiveness: Economy, strategy, management*. INFRA-M. (in Russian).
13. Zavadsky, Y. S., Osovska, T. V., & Yushkevich, O. O. (2006). *Economic dictionary*. Condor. (in Ukrainian).
14. Yaroshenko, S. P. (1998). *Reserves for enhancing the competitiveness of the meat subcomplex in Ukraine*. Cossack Val. (in Ukrainian).
15. Illyashenko, S. M. (2005). *Marketing commodity policy*. University Book. (in Ukrainian).
16. Zayats, R. (2017). *Managing the international competitiveness of the enterprise*. Ternopil National Economic University. (in Ukrainian).
17. The readiness for the future of production report. (2018). http://www3.weforum.org/docs/GCR2018/05FullReport/TheGlobalCompetitivenessReport2018.pdf
18. Illyashenko, N. S. (2018). *Advanced innovative development: Theory, methodic, practice*. Trytoria. (in Ukrainian).
19. Domashenko, M., Kotenko, O., Shkola, V. Y., & Kuchmiyov, A. (2017). Innovative marketing strategies to provide ecological safety at regional and global levels. *Marketing and Management of Innovations, 4*, 367–373. https://doi.org/10.21272/mmi.2017.4-33
20. Prokopenko, O., Domashenko, M., & Shkola, V. (2014). Management features of economic security of foreign economic activity of Ukrainian machine-building enterprises. *Actual Problems of Economics, 10*(160), 188–194. ISSN: 19936788.
21. Babenko, V. (2019). Formalization of the model of management of the technological innovations. *CEUR Workshop Proceedings, 2393*, 595–602. http://ceur-ws.org/Vol-2393/paper_431.pdf
22. Brauweiler, H.-C., Shkola, V. Y., & Markova, O. O. (2017). Economic and legal mechanisms of waste management in Ukraine. *Marketing and Management of Innovations, 2*, 359–368. https://doi.org/10.21272/mmi.2017.2-33
23. Illyashenko, S. M. (2004). *Economic risk*. Center for Educational Literature. (in Ukrainian).
24. Illyashenko, S. M. (2010). *Innovative management*. University Book. (in Ukrainian).
25. Babenko, V., Lomovskykh, L., Oriekhova, A., Korchynska, L., Krutko, M., & Koniaieva, Y. (2019). Features of methods and models in risk management of IT projects. *Periodicals of Engineering and Natural Sciences, 7*(2), 629–636. https://doi.org/10.21533/pen.v7i2.558
26. Kozmenko, S. (2005). *Investment decisions and management of scientific and technical progress*. University Book, LLC Consulting Publishing Company "Business Perspectives". (in Russian).
27. Lipsits, I. V., & Kosov, V. V. (1996). *Investment project: Preparation and analysis methods*. BEK. (in Russian).
28. Kasyanenko, T. V. (2012). *Economic justification of ecologically oriented innovative development*. Donetsk State University of Management. (in Ukrainian).
29. Dolzhanskyy, I., & Tkachuk, S. (2007). Investment risk as threats of economic security. *Development Management, 8*, 20–22. (in Ukrainian).
30. Illyashenko, S. (2005). *Management of innovative development*. University Book/Princess Olga. (in Ukrainian).
31. Porter, M. E. (1990). Competitive advantage of nations. *Competitive Intelligence Review, 1*(1), 14–14. https://doi.org/10.1002/cir.3880010112
32. Porter, M. E. (2008). The five competitive forces that shape strategy. *Harvard Business Review, 86*(1), 79–93. ISSN: 00178012.

33. Illyashenko, S., & Olefirenko, O. (2008). *Management of a portfolio of orders of the research-and-production enterprise*. University Book. (in Ukrainian).
34. Meskon, M. H., Hodoury, M., & Albert, F. (1992). *Fundamentals of management*. Case. (in Russian).

Conception of a Novel Drone Based on the Multispectral Camera Dedicated to Monitoring of Vital Parameters in Agricultural Fields

Kamal Dabali, Rachid Latif, and Amine Saddik

Abstract In agricultural applications, the use of robots and Unmanned Aerial Vehicle (UAVs) becomes the most tools for developing the control of agriculture while the ground tests take more time in analyzing data from the field and then send this data to a laptop and use another software for analysis. These steps are costly and doesn't improve the future and remote sensing of agriculture control. This paper proposes the conception implementation of a drone using an open source software and recent developed hardware. This work consists on the design and dimensioning of a Quadcopter implementing a Parrot Sequoia multispectral camera onboard. The Data captured by this camera will be processed by the heterogeneous system Odriod-XU4 type CPU-GPU board. We implemented the methodology for the development of a control model of agriculture parameters to improve the production. This approach is based on the assembly of a quadcopter with APM 2.8 Ardu-Pilot Mega flight controller with a laptop ground station controlled via Mission Planner open source software and different sensors to stabilize the flight. The method integrates the combination of data processing algorithms from multispectral camera and GPS coordinates. Finally, we present results of this work in different agriculture areas. The objective of this framework is to develop a new concept of analyzing real time data in the quadcopter. The development of such methods will help agronomists to achieve efficient processing techniques in remote sensing agriculture.

Keywords UAVs · Multispectral camera · Heterogeneous system · APM2.8 · Agriculture

K. Dabali (✉) · R. Latif · A. Saddik
Laboratory of Systems Engineering and Information Technology LISTI, National School of Applied Sciences, Ibn Zohr University, Agadir, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_12

1 Introduction

The low-cost open source software and modern microprocessors has made electronic and even completely autonomous control of quadcopters feasible for commercial, agricultural purposes [1]. In [2], they proposed frame using metal rod for the quadcopter and with this structure there is the issue of vibration. In the research done by [3], a quadcopter controlled manually by RC transmitter and receiver where manually flight parameters are conducted to errors. In [4], PID controller is used to control the quadcopter while Extended Kalman Filter (EKF) is used to reject the noise from IMU sensor. In [5], he presented how to build own quadcopter and use PID controller for stabilized flights of quadcopter. In [6], the author implemented and tested the APM 2.6 autopilot and the Mission Planner software on a fixed wing UAV. The testing verified the system flexibility and proper working in different environments. Reference [7] has designed a quadcopter with Arduino UNO board as a flight controller. But Arduino board is not suitable for real time control of quadcopter. Hence the quadcopter requires tuned PID parameters. Furthermore, the flight characteristics of quadcopter is not good. Reference [8] they used APM 2.6 controller which leads the quadcopter using onboard sensors. But it is not having the inbuilt compass required for navigation. Also, it is designed as a manually controlled quadcopter. Unmanned aerial vehicles (UAVs) are very promising instruments in agricultural sciences [9, 10]. Surveys by Satellite and manned aircraft have reported limitations concerning resolution, operational costs, and unfavorable climate conditions [11]. In contrast, the continuous development of unmanned aerial vehicles (UAVs) designs, navigation systems, portable image sensors and cutting-edge machine learning methods allow unobtrusive, accurate, and versatile surveillance tools in precision agriculture and biosecurity [12–15]. Many studies have positioned UAVs for the collection of aerial imagery in applications such as weed, disease, and pest mapping and wildlife monitoring [16–18]. More recently, unmanned aerial systems (UASs) have been deployed in cluttered and global positioning system (GPS)-denied environments [19]. The authors in [20] used UAV equipped with camera for site-specific vineyard management. Normalized Differential Vegetation Index (NDVI) values acquired by the Tetracam ADC-lite camera mounted on board were compared to ground-based NDVI values measured with the FieldSpec Pro spectroradiometer to verify the precision of the ADC system. The vegetation indices obtained from UAV images are in excellent agreement with those acquired with a ground-based high-resolution spectroradiometer. Reference [21] addresses the design of an autonomous unmanned helicopter systems for remote sensing missions in unknown environments. Focuses on the dependable autonomous capabilities in operations related to Beyond Visual Range (BVR) without a backup pilot by providing flight services. Utilizes a method called Laser Imaging Detection and Ranging (LIDAR) for object detection which are applicable in real world development. The study in [22] used a team of UAVs to take images in order to create a full map, applying mosaicking procedures for post processing and automatic task partitioning management which is based on negotiation among the UAVs, considering their state and capabilities. Thus, they combine a strategy which encompass multi robot task planning, path planning and UAV control for the coverage of a crop field for data collection. Precision agriculture is also known as precision ag or precision farming. Perhaps the easiest way to understand precision is to think of it as everything that makes the practice of farming more accurate and controlled when it comes to the growing of crops and raising livestock. In our case, agriculture requires good soil quality. This quality directly affects agricultural production, which requires a permanent monitoring of the vital parameters namely the standardized difference vegetation index (NDVI), the normalized difference water index (NDWI) and other problem like the bad ones. Herbs, for a good irrigation of agricultural fields and especially in our region SOUSS MASSA. In this work we have opted for the XU4 heterogeneous CPU-GPU architecture based on the work of [23, 24]. Using the previous specifications, we have developed a Drone using Arducopter APM 2.8, the latest version of APM series which has many sensors onboard. The mathematical modeling of quadcopter is done using MATLAB Simulink model using PID controller for stabilized flights of quadcopter. Pitch roll and yaw movement of quadcopter responses are obtained through MATLAB Simulink.

2 Component and Architecture

2.1 A Quadcopter Platform

The function of the main structure is to be the framework of the quadcopter, gathering all the components. The frame used is S500 strong lightweight glass fiber with landing gear. It is an x-shape quadcopter with about 15 min flying time and 150 m remote distance. This frame has 480 mm motor centers so is ideal for 9 and 10 in. props and it also has an adjustable battery mount to achieve the perfect weight distribution and the bottom frame is ready to take a whole host of camera mounts (Fig. 1).

Fig. 1 Quadcopter frame

2.2 Ground Control Station (GCS)

The Mission Planner software is installed on a windows-based laptop computer and exploited as GCS. The purpose of the GCS is to plan the flight path. The features of mission planner [11] are: Point-and-click waypoint entry using Google Maps.

2.3 Ardu-Pilot Mega Board

The Ardu-Pilot Mega (APM 2.8) is a complete autopilot system. It is used in this work to make the drone fully autonomous; capable of performing programmed GPS missions with waypoints. It includes 3-axis gyro, accelerometer, along with a high-performance barometer. Onboard 4 MegaByte Dataflash chip for automatic datalogging. Optional off-board GPS, LEA-6H module with Compass. One of the first open source autopilot systems to use Invensense's 6 DoF Accelerometer/Gyro MPU-6000. Includes a barometric pressure sensor upgraded to MS5611-01BA03, from Measurement Specialties, Atmel's ATMEGA2560 and ATMEGA32U-2 chips for processing and usb functions, respectively (Fig. 2).

2.4 Motors

Quadcopter brushless motors are more efficient, more reliable and quieter than a brushed motor. We have used 1000 rpm/Kv A2212-1000 Kv brushless motor.

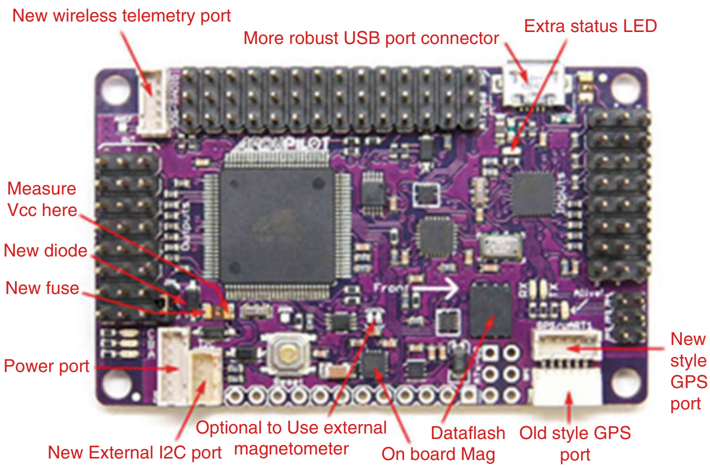


Fig. 2 APM 2.8 Flight Controller board

2.5 Blades

Blades are the quadcopter element in charge of producing the lifting and propelling forces. We selected plastics blades, since they are cheaper and their effects in performance are negligible respect to the use of carbon fiber.

2.6 Electronic Speed Controllers

The aim of the Electronic Speed Controllers (ESC) is adapting the motor velocity according the PWM (Pulse Wide Modulation) inputs that they are receiving.

2.7 Battery

Drones commonly use LiPo battery as energy source. The main reasons of this tendency are their lightweight,. Capacity is an indication about how much power can be stored by the battery, and it is measured in mAh (miliAmperes per hour). For ensuring flight missions of 15 min, LiPo batteries of 3000 mAh have been used.

2.8 Radio Controller

It is the device that connects the pilot with the aircraft. It has a minimum of four channels, and it transmits at frequencies in the range of 2.4 GHz. The radio controller used can select a maximum of three modes. One of them is going to be reserved for AUTO mode. It will be selected when it will be being done an automatic mission. On this way, it can be recovered the quadcopter control as soon as it is required.

2.9 Telemetry and GPS

The function of the telemetry is proportionate flight data at real time. The GPS is in charged to estimate the global position of the aircraft by measuring the relative positions with respect to several satellites. We chose external GPS/compass module: to Ublox-NEO 6 M GPS module with built in HMC5883L compass.

Fig. 3 Parrot Sequoia Camera

Table 1 Parrot Sequoia Camera Performance Data

Diagonal fields of view	73.5°
Vertical fields of view	50.1°
Horizontal fields of view	63.9°
Weight	72 g
Energy	5 W (~12 W en pic)
Dimensions	59 mm × 41 mm × 28 mm

2.10 Multispectral Camera

An off-the-shelf, lightweight multispectral camera has been downward-faced and mounted to the bottom of the platform. The camera was set to capture (Fig. 3 and Table 1).

2.11 XU4 Board

ODROID-XU4 is a new generation of computing device with more powerful, energy-efficient hardware and a smaller form factor. Offering open source support, the board can run various flavors of Linux, including the latest Ubuntu 16.04 and Android 4.4 KitKat and 7.1 Nougat. By implementing the eMMC 5.0, USB 3.0 and Gigabit Ethernet interfaces, the ODROID-XU4 boasts amazing data transfer speeds, a feature that is increasingly required to support advanced processing power on ARM devices.

3 System Implementation

Implementing a quadcopter mission is not simple. It requires some system preparations such as: software installation/configuration, stabilizing the quadcopter and maneuvering trough waypoint designation and hardware interfacing. Also, the step of creating waypoint in mission planner requests knowledge with map utilization including GPS longitude/latitude readings.

3.1 Preparation Phase

We connect the autopilot to the computer and we load the firmware and we select the robot and frame type. After that we connect the autopilot to APM Planner 2 using the 3DR Radio which allows us to communicate with the quadcopter in flight and provides more mobility during calibration. We select the connect box and set the rate to 57,600 db as port. Finally, we start the calibration with the condition of removing the propellers.

3.2 Compass and Radio Calibration

Based on the above discussion, a timer will start for 60 s where we hold the quadcopter in the air and we perform rotation slowly so that each side front, back, left, right, top and bottom) points down toward the earth for a few seconds in turn. One time the calibration period ends, the APM will display the resulting offsets. These values are between -150 and 150. The disadvantage of the internal compass, the compass will have more magnetic interference. In the radio calibration we verify that it is in airplane mode and all trims are centered. We opt for Mode 1 transmitters, the left stick will control pitch and yaw, the right stick will control throttle and roll. The transmitter's three-position switch should be attached to channel 5 and will control flight modes. We select Radio Calibration to order the autopilot to work with the RC transmitter and receiver. Resulting normal values are around 1100 for minimums and 1900 for maximums.

4 Modeling

The mathematical modeling of the quadcopter is studied from [2]. Quadcopter can move in six directions (Fig. 4).

4.1 Kinematics

We define the position and velocity of the quadcopter in the inertial frame as $x = (x, y, z)^{\mathrm{T}}$and $\dot{x} = (\dot{x}, \dot{y}, \dot{z})^{\mathrm{T}}$, respectively. Similarly, we define the roll, pitch,

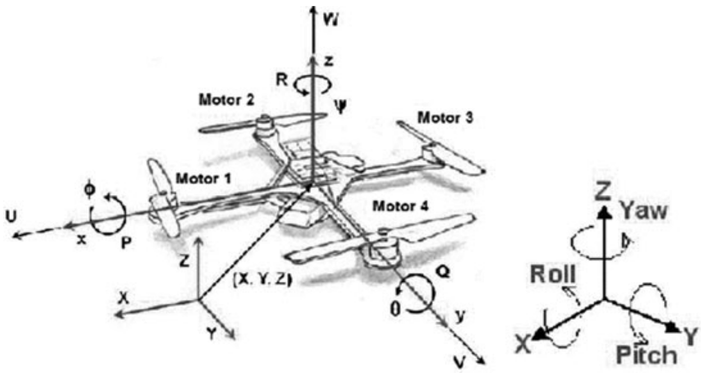


Fig. 4 Quadcopter Modeling

and yaw angles in the body frame as $\theta = (\phi, \theta, \psi)^{\mathrm{T}}$, with corresponding angular velocities equal to $\dot{\theta} = \left(\dot{\phi}, \dot{\theta}, \dot{\psi}\right)^{\mathrm{T}}$.

Relation between the body and inertial frame by the rotation matrix R.

$$R = \begin{bmatrix} c_\phi c_\psi + c_\theta s_\phi s_\psi & -c_\psi s_\phi - c_\phi c_\theta s_\psi & s_\theta s_\psi \\ c_\theta c_\psi s_\phi + c_\phi s_\psi & c_\phi c_\theta c_\psi - s_\phi s_\psi & -c_\psi s_\theta \\ s_\phi s_\theta & c_\phi s_\theta & c_\theta \end{bmatrix} \dot{\theta} \tag{1}$$

4.2 Motors

Brushless motors are used for all quadcopter applications. For our electric motors, the torque produced is given by

$$\tau = K_t \left(I - I_0\right) \tag{2}$$

After simplifying all calculation, we obtained the final simplified equation for power:

$$P \approx \frac{K_v}{K_t} \tau \omega \tag{3}$$

4.3 Forces

The total thrust on the quadcopter is given by

$$T_B = \sum_{i=1}^{4} T_i = \kappa \begin{bmatrix} 0 \\ 0 \\ \sum \omega_i^2 \end{bmatrix} \tag{4}$$

In addition, global drag forces will be modeled by an additional force term

$$T_B = \begin{bmatrix} -\kappa_d \dot{x} \\ -\kappa_d \dot{y} \\ -\kappa_d \dot{z} \end{bmatrix} \tag{5}$$

.

4.4 Torques

The torques equation in the body frame are

$$\tau_{\mathrm{B}} = \begin{bmatrix} L\kappa \left(\omega_1^2 - \omega_3^2\right) \\ L\kappa \left(\omega_2^2 - \omega_4^2\right) \\ b \left(\omega_1^2 - \omega_2^2 + \omega_3^2 - \omega_4^2\right) \end{bmatrix} \tag{6}$$

The model is supposed too simple ignoring a lot of effects like rotational drag forces, blade flapping, wind, etc.

4.5 Equations of Motion

The body frame rotational equations of motion is:

$$\dot{\omega} = \begin{bmatrix} \tau_\phi I_{xx}^{-1} \\ \tau_\theta I_{yy}^{-1} \\ \tau_\psi I_{zz}^{-1} \end{bmatrix} - \begin{bmatrix} \frac{I_{yy}-I_{zz}}{I_{xx}} \omega_y \omega_z \\ \frac{I_{zz}-I_{xx}}{I_{yy}} \omega_x \omega_z \\ \frac{I_{xx}-I_{yy}}{I_{zz}} \omega_x \omega_y \end{bmatrix} \tag{7}$$

4.6 Design Calculation

The weight of the entire drone is estimated to be 1000 g, the total thrust required to lift the drone is twice the weight of the quadcopter.

Mathematically $T = 2 \times W$, where T is thrust and W is the weight if the quadcopter. Hence $T = 2 \times 1000 = 2000$ g, since the quadcopter has 4 rotors, each rotor will have a thrust of $T = W/4$, therefore each motor should be capable of delivering a thrust of 500 g.

4.7 Control Strategy

PID controller design is shown below Fig. 5. Reference signal is pitch or roll or yaw to PID controller and its output is given to quadcopter. Quadcopter output signal is given as negative feedback to adder block. According to error, K_p, K_i and K_d are chosen. Where K_p is the proportional gain, K_i is the integral gain, K_d is the derivative gain, $e(t)$ is the error signal and $u(t)$ is the PID controller output.

5 Results and Discussion

5.1 Software and Hardware Simulation Results

The quadcopter dynamic motions are simulated by using MATLAB software using PID controller. The MATLAB results are shown in Figs. 6, 7, and 8 which are pitch, roll and yaw responses of quadcopter respectively. The PID gains are chosen as 25, 0.2, and 80, to stabilize the quadcopter motions. The quadcopter was assembled,

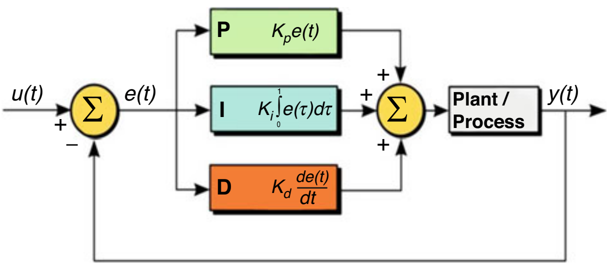


Fig. 5 Quadcopter model using PID controller

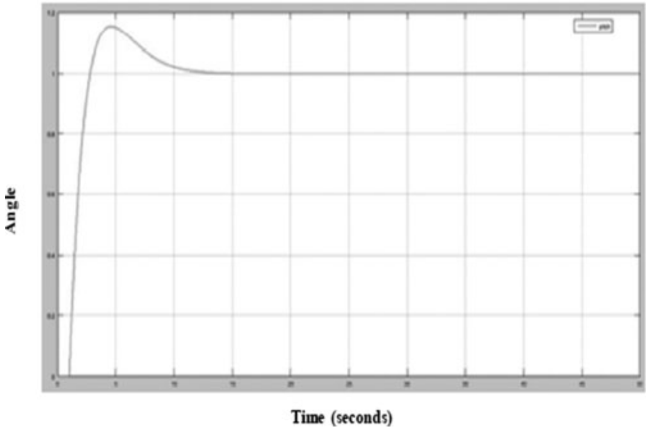


Fig. 6 Pitch movement of the quadcopter

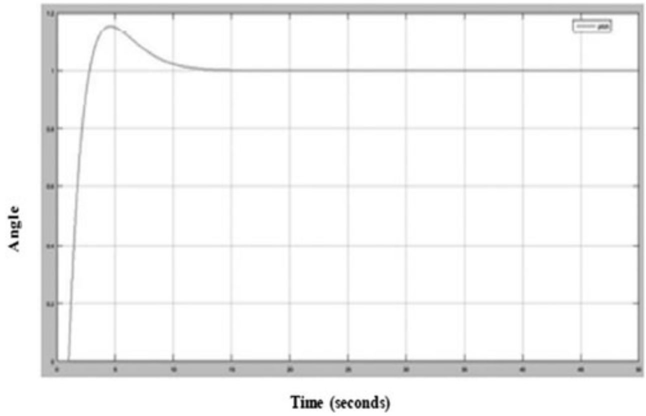


Fig. 7 Roll movement of the quadcopter

each component was tested to ensure if they were in good conditions. The flight controller was tested by implementing the new PID parameters and calibrated. The quadcopter was tested by several flight. With the battery chosen, the estimated time of flight was about 10 min.

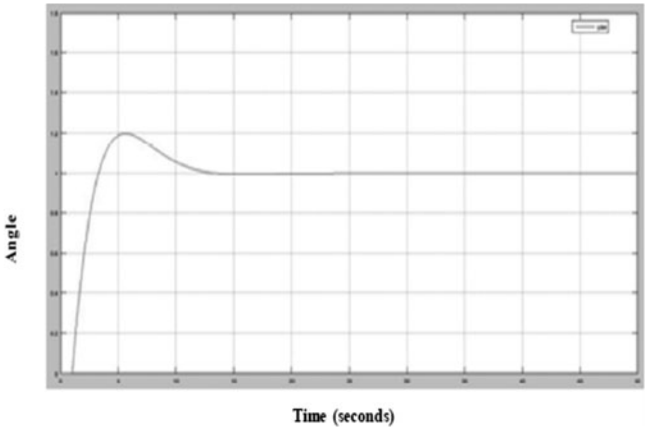


Fig. 8 Yaw movement of the quadcopter

References

1. Widodo, N. S., Yudhana, A., & Sunardi. (2014). Low cost open source based UAV for aerial photography. *International Journal of Innovative Research in Advanced Engineering (IJIRAE), 1*(10).
2. Khajure, S., Surwade, V., & Badak, V. (2016). Design of a quad copter and fabrication. *International Advanced Research Journal in Science, Engineering and Technology*, *3*(2).
3. Kumar, N. S., & Sankaralingam, R. K. (2016). Design and control implementation of quadcopter. *International Journal of Mechanical and Production Engineering*, *4*(5).
4. Hassan Tanveer, M., Faiz Ahmed, S., Hazry, D., Warsi, F. A., & Joyo, M. K. (2013). Stabilized controller design for attitude and altitude controlling of quad-rotor under disturbance and noisy conditions. *American Journal of Applied Sciences, 10*(8), 819–831.
5. Norris, D. (2014). *Build your own quadcopter. Power up your designs with the parallax Elev-8*. McGraw Hill Professional.
6. Figueiredo, D. L. (2014). *Autopilot and ground control station for UAV*. Instituto Superior Técnico.
7. Rajpoot, A. S., Gadani, N., & Kalathia, S. (2016). Development of arduino based quadcopter. *International Advanced Research Journal in Science, Engineering and Technology*, *3*(6).
8. Abdulmajeed, W. R., Athab, O. A., & Sattam, I. A. (2016). Implementation and evaluation of APM 2.6-controlled quadcopter with aerial imagery as a case study. *ARPN Journal of Engineering and Applied Sciences, 11*(19), 11724–11734.
9. Floreano, D., & Wood, R. J. (2015). Science, technology and the future of small autonomous drones. *Nature, 521*, 460–466.
10. Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., Miklas, P. N., Carter, A. H., Pumphrey, M. O., Knowles, N. R., et al. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *European Journal of Agronomy, 70*, 112–123.
11. Salami, E., Barrado, C., & Pastor, E. (2014). UAV flight experiments applied to the remote sensing of vegetated areas. *Remote Sensing, 6*, 11051–11081.
12. Glassock, R., Hung, J. Y., Gonzalez, L. F., & Walker, R. A. (2008). Design, modelling and measurement of a hybrid powerplant for unmanned aerial systems. *Australian Journal of Mechanical Engineering, 6*, 69–78.

13. Whitney, E., Gonzalez, L., Periaux, J., Sefrioui, M., & Srinivas, K. (2004). A robust evolutionary technique for inverse aerodynamic design. In *Proceedings of the European Congress on Computational Methods in Applied Sciences and Engineering*, Jyvaskyla, Finland, 24–28 July 2004 (Vol. 2, pp. 1–2).
14. Gonzalez, L., Whitney, E., Srinivas, K., & Periaux, J. (2004). Multidisciplinary aircraft design and optimisation using a robust evolutionary technique with variable fidelity models. In *Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany, NY, USA, 30 August–1 September 2004 (Vol. 6, pp. 3610–3624).
15. Ken, W., & Chris, H. C. (2014). Remote sensing of the environment with small unmanned aircraft systems (UASs), part 1: A review of progress and challenges. *Journal of Unmanned Vehicle Systems, 2*, 69–85.
16. Gonzalez, L., Montes, G., Puig, E., Johnson, S., Mengersen, K., & Gaston, K. (2016). Unmanned Aerial Vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. *Sensors, 16*, 97. https://doi.org/10.3390/s16010097
17. Sandino, J., Wooler, A., & Gonzalez, F. (2017). Towards the automatic detection of pre-existing termite mounds through UAS and hyperspectral imagery. *Sensors, 17*, 2196. https://doi.org/10.3390/s17102196
18. Vanegas, F., Bratanov, D., Powell, K., Weiss, J., & Gonzalez, F. (2018). A novel methodology for improving plant pest surveillance in vineyards and crops using UAV-based hyperspectral and spatial data. *Sensors, 18*, e260. https://doi.org/10.3390/s18010260
19. Vanegas, F., & Gonzalez, F. (2016). Enabling UAV navigation with sensor and environmental uncertainty in cluttered and GPS-denied environments. *Sensors, 16*, 666. https://doi.org/10.3390/s16050666
20. Primicerio, J., Di Gennaro, S. F., Fiorillo, E., Genesio, L., Lugato, E., Matese, A., & Vaccari, F. P. (2012). A flexible unmanned aerial vehicle for precision agriculture. *Precision Agriculture, 13*(4), 517–523.
21. Merz, T., & Chapman, S. (2011). Autonomous unmanned helicopter system for remote sensing missions in unknown environments. In *Conference on Unmanned Aerial Vehicle in Geomatics (UAV-g)* (2011) (pp. 14–16).
22. Barrientos, A., Colorado, J., Cerro, J. d., Martinez, A., Rossi, C., Sanz, D., & Valente, J. (2011). Aerial remote sensing in agriculture: A practical approach to area coverage and path planning for fleets of mini aerial robots. *Journal of Field Robotics, 28*(5), 667–689.
23. Latif, R., Saddik, A., & Elouardi, A. (2019). Evaluation of agricultural precision algorithms on UAV images. In *2019 International Conference of Computer Science and Renewable Energies (ICCSRE)*, Agadir, Morocco (pp. 1–4).
24. Saddik, A., Latif, R., Elhoseny, M., & El Ouardi, A. (2020). Real-time evaluation of different indexes in precision agriculture using a heterogeneous embedded system. *Sustainable Computing: Informatics and Systems, 30*, 100506. https://doi.org/10.1016/j.suscom.2020.100506

Metrics in Precision Agriculture Using Multispectral Images: Review and Evaluation

Rachid Latif, Amine Saddik, and Abdelhafid Eouardi

Abstract Precision agriculture has recently seen a huge technological revolution, which affects the quality of production in agricultural fields. Among the different tools used in precision agriculture, we find unmanned aerial vehicles, Micro Aerial Vehicles, and soil robots. These tools are equipped with multispectral or hyperspectral cameras to solve a variety of problems in agriculture. In this work, we present an overview of precision agriculture based on field monitoring using the different normalized indices. In our contribution, we also present an evaluation of an algorithm that processes the different multispectral images collected by different tools. The results obtained have shown that we can process 400 frames/s, in the case of low resolution and 30 frames/s in the case of high resolution, which will allow us to process a sequence of images in real time.

Keywords Precision agriculture · UAV · MAV · Soil robot · Real time

1 Introduction

Agriculture has recently benefited from technological evolution to better make decisions and controls. These technologies have spread in different areas such as fertilizer rate control, irrigation systems and parameter control related to agricultural soil. Today, agriculture precision is based on the collection of information issued by specific sensors, as an example soil moisture control is based on humidity sensors. This method remains useless when we talk about agricultural soils of hundreds of hectares [1, 2]. Precision agriculture is defined as a management method that uses

R. Latif · A. Saddik (✉)
Laboratory of Systems Engineering and Information Technology LISTI, National School of Applied Sciences, Ibn Zohr University, Agadir, Morocco
e-mail: amine.saddik@edu.uiz.ac.ma

A. Eouardi
Université Paris-Saclay, ENS Paris-Saclay, CNRS, Systèmes et Applications des Technologies de l'Information et de l'Energie, Orsay, France

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_13

information technology to integrate data from a variety of sources into agricultural production decisions [3]. The tools traditionally used to obtain images remotely of agricultural soil are satellites [4], but the main constraint remains in the low resolution of images [5, 6]. These constraints can be solved by using unmanned aerial platforms known as unmanned aircraft systems, or remotely operated aircraft systems. Agricultural fields are generally divided into three types, large, medium and greenhouse, monitoring of large agricultural fields requires an effective tool that can treat a large area in a specific time. The most commonly used tool, in this case, is satellite imagery or UAVs. In the case of medium fields, we can use micro aerial vehicles equipped with multispectral cameras to monitor different agricultural problems [6]. For small farms or a greenhouse, we can use soil robots [7, 8]. All these tools allow us to monitor problems in different farms.

The main goal of precision agriculture is planted improvement, based on solving different problems. The most common problems encountered in agriculture can be summarized as diseases, weeds and irrigation system management. In this context, the work of P. Jiang et al. presents a method of disease detection based on convolutional neural networks (CNN). The algorithm proposed in this paper deals in real time with five types of the disease most commonly encountered in plants, Rust, Brown spot, Alternaria leaf spot, Mosaic and Gray spot. The results show that the use of the SSD model with Inception module and Rainbow concatenation gives a detection accuracy of 78.80% for different maladies, treat 23 frames/s [9]. P. Bosilj et al. presents a method of weed classification based on vegetation per pixel, the results of this paper show an improvement compared to the different methods existing in the state of the art [10]. S. Gobbo presents the surface energy balance algorithm land for the estimation of the yield of maize fields, to provide sufficient water. The study is based on the collection of the standardized vegetation index and the use of humidity sensors to map the irrigation area using the aforementioned Management Zone Analyst software [11].

In this article, we present in Sect. 2 a state of art on precision agriculture based on the different problems encountered in traditional agriculture. In Sect. 3, we propose an algorithm for monitoring the different agricultural fields using Normalized Difference Water Index NDWI and Normalized Difference Vegetation Index NDVI. Section 4 shows the evaluation of the algorithm based on the implementation of the proposed algorithm in an embedded system using the parallel programming language OpenMP.

2 Precision Agriculture: An Overview

Precision agriculture is characterized by the use of intelligent tools to improve agricultural production. In this context, the algorithm part of precision agriculture is based on the use of machine learning, deep learning or segmentation and classification algorithms such as Object-Based Image Analysis (OBIA) [12]. These algorithms are more and more complicated, which requires a detailed study to obtain an embedded implementation of these algorithms. On the other side, the image acquisition is done by multispectral cameras like Parrot Sequoia +, TetraCam, or RededgeMicasense. These cameras are devised in two types, for example, red-edge Micasense gives separate multispectral images for each color band, but TetraCam gives images with the three bands R, G, and B in the same image, in this case, a band separation is recommended especially in applications that are based on the normalized index. The standardized indices in agriculture are divided into several types, soil index, vegetation index, or water index. This variety of indices allows us to monitor in detail the differences characteristic of soil and plant.

2.1 Indices Based Precision Agriculture

The calculation of indices in precision agriculture is divided into two tools; the first tool consists of the use of an aerial or soil robot equipped by a multispectral camera. The choice of the robot depends on the desired application, for example, agriculture in the SOUSS-MASSA region in Morocco is characterized by the use of a greenhouse which makes the use of UAVs very difficult, in this case, the used of soil robots are recommended. The second tool is based on the satellite image, the satellite image database gives images with different bands varies between 11 and 12 bands. Generally, we find satellite images such as Landsat 8 OLI with 11 bands including R, G, B, NIR, and Short-Wave Near-Infrared (SWIR) with a resolution between 30 and 100 m and a wavelength between 0.43 and 12.5 μm [13, 14]. There is also the Sentinel 2 MSI satellite dataset which has 12 bands with a resolution ranging from 10 to 60 m and a wavelength from 0.4427 to 2.2024 μm. The problem of surveillance using satellite images is summarized in the real-time boarding of algorithms that process the different indices. In this case, the use of robots remains the best choice. In digital agriculture we find a variety of indexes, each one is dedicated to an application, for example, the NDVI (The Normalized Difference Vegetation Index) dedicated to the detection of vegetation. Generally, this index is used for the detection of weeds, we can also use it in the detection of diseases. The use of this index is very important especially in applications that are based on deep learning or machine learning to avoid processing all the images we focus only on vegetation. For water and humidity control the NDWI (The Normalized Difference Water Index) and NDMI (The Normalized Difference Moisture Index) are recommended. These three indices are the most well-known but we can find other indices such as the NPCRI (The Normalized Pigment Chlorophyll Ratio Index) which is an index that determines the chlorophyll content of crops or vegetation. We also find BSI (The Bare Soil Index) index dedicated to soil mapping and crop identification. Table 1 shows the different indices [15–18].

Table 1 Index overview

Index	NDVI [15]	NDWI [16]	NDMI [16]	NPCRI [17]	BSI [18]
Equation	$\frac{\mathrm{NIR}-R}{\mathrm{NIR}+R}$	$\frac{G-\mathrm{NIR}}{G+\mathrm{NIR}}$	$\frac{\mathrm{NIR}-\mathrm{SWIR}}{\mathrm{NIR}+\mathrm{SWIR}}$	$\frac{R-B}{R+B}$	$\frac{(R+\mathrm{SWIR})-(\mathrm{NIR}+B)}{(R+\mathrm{SWIR})+(\mathrm{NIR}+B)}$

2.2 Algorithms Study

Precision agriculture can solve different problems in traditional agriculture, but it can also be used in various applications, for example, J. F. González et al. used drones to produce orthomosaic plans in NDVI to determine the existence and identify the Roman road in the areas studied. The system used in this work is based on a drone equipped with a multispectral camera that calculates the NDVI vegetation index. This study covers ten distinct areas of 88 ha. In addition to the multispectral camera, they added another Canon PowerShot S100 camera with high resolution and low weight. In general, flight heights vary between 120 and 25 m, but they found that 120 m flights have a low resolution. On the other hand, flights at 25 m give a high resolution but the construction of orthomosaic images becomes very difficult. For the algorithmic part, they processed the images emitted by the drone on the Qgis 2.14 software to determine the NDVI and analyze the results. After processing the image in the software part, the orthomosaic images were rendered in gray, a level that highlights the differences in plant health. The higher the NDVI index, the whiter it will appear on the images [19].

The use of the different standardized indices cited in the work of A. Khaliq et al., this article shows a comparison between images collected by drones and satellite images for the evaluation of the variability of the vineyard. In this study, the authors used cloud-free reflectance images of level 2A Sentinel-2 and were then derived from level 1 using the algorithm offered in the SNAP toolbox (6.0) and the sen2core processor (2.5.5) for atmospheric classification and correction. Multispectral imagery based on drones was created with AgisoftPhotoScan software, processing image blocks of more than 1000 aerial images obtained with an airborne Parrot Sequoia multispectral camera. The trajectory of the drone was designed to maintain the flight height at nearly 35 m. To compare the image captured by the drone with the satellite image, the authors followed a preliminary procedure to subsample the high-resolution images of the drones. He named the bumblebee image matrix $G\ (i, j)$ where i is the number of rows and j is the number of columns that is related to the pixels of the satellite $S\ (i, j)$. They then calculate three NDVI indices to perform a detailed analysis after a comparison with the one based on satellite images [20].

The work of R. PUDELKO et al. shows an evaluation of the ability to monitor agricultural soils using a remote flight model. The work is divided into two parts; the first part aims to show the strong and weak bridges of motor glider flight as a platform. The second part is based on the collection and development of photographs taken with this type of construction. The work also presents an example of how to detect areas where differences are visible within the vegetation cover structure. The example shows the presence of weeds and domestic diseases [21]. In 2018, the authors of C. Ju. et al. worked on a system that contains several unmanned aerial vehicles (UAVs) on agricultural fields to reduce fatigue and processing time. In this article, they replaced the systems used in the first single drone work with a centralized controller. They developed a multi drone system for agriculture using the distributed control algorithm. The proposed work was evaluated and analyzed according to the following performance measures:

- The total time of the system processing.
- Time of preparation (what an operator does before the UAV performs an agricultural task).
- The flying time (Time during which the drone performs agricultural tasks and which is directly related to the drone's energy consumption).
- Battery consumption.
- Coverage rate [22].

U. Mogili et al., 2018, offers an overview of the applications of UAV systems in precision agriculture. He described the different techniques used, such as the type of unmanned aerial vehicle (Quad Copter, Hexa Copter, Octo Copter, Octo Copter, Fixed Wing, Single Rotor Helicopter) and the equipment components used to perform agricultural work, such as the calculation of the NDVI vegetation index [23]. On the other side specifically in weed detection. I. Sa et al., present a classification approach based on multispectral images collected by aerial microvesicles to detect weeds. They use neural networks to obtain a dataset, and also, they have used a field with different herbicide levels, which gives it plots that contain only crops or weeds. Subsequently, they have used the NDVI vegetation index to characterize the plots. The work is divided into three parts, the first part is the database construction; the second part is related to the pre-processing of this database and the last one for Dense Semantic Segmentation Framework [6].

2.3 Embedded Systems Based Agriculture Applications

The real-time processing of these algorithms remains a challenge, especially when talking about applications that require large numbers of frames per second. The most commonly used methods today are based on different software related to precision agriculture. The boarding of these algorithms requires some study to implement these algorithms in embedded systems. In this context, we notice that we have a lack of boarding in precision agriculture. In this part, we will discuss some work that is based on the use of embedded systems. J. Chen et al., process his new approach on the SPOT VGT-S product which includes 8849×5601 pixels and 48 layers. This study is based on a DELL desktop computer that has 4 CPUs running at 1800 MHz and 1 GB RAM. The processing time, in this case, is 22 h [24]. The following table shows a synthesis of different algorithms (Table 2).

Table 2 Synthesis of different algorithms

Reference	Application	Algorithm	System
Chen et al. [24]	Filtering NDVI values	Savitzky–Golay	DELL desktop @ 1800 MHz
Jin et al. [25]	Classification of Remote Sensing Images	Object-Oriented Method Combined with Deep Convolutional	NVIDIA Tesla K80 GPU and Intel Xeon CPU E5-2630 @2.30 GHz
Sa et al. [6]	Weed detection	CNN-based dense semantic	Jetson TX2 GPU and Titan X @ 1.3 GHz
Peña et al. [26]	Weed detection	OBIA	Software eCognition Developer 8.9
da Silva et al. [27]	Failures identification in plantations	—	FPGA DE2 card and the Raspberry Pi 2
			Card @ 1.6 GHz
Sanchez et al. [28]	Weed Management	OBIA	Software AgisoftPhotoscan and JMP
Zhou et al. [29]	Spatial recognition of maize plants	SDST Otsu method	Software MATLAB, and LabVIEW 2014.

B. Jin et al. present a Land-Use-Type classification of remote sensing images. They used a Windows operating system with an NVIDIA Tesla K80 GPU for acceleration, an Intel Xeon(R) CPU E5-2630 and 128 GB of memory [25]. I. Sa et al. present work for weed detection by classification based on an embedded computer equipped with a Jetson TX2 GPU, 2 GHz Hexa-CPU cores, 256 GPU @1.3 GHz. For this application, they used another embedded system based on Titan X to make a comparison between the two systems. The final results showed that Titan X is 3.6 times faster than TX2 but with high energy consumption rates [6]. Another proposal that appeared in the work of [30] is the use of a heterogeneous system to develop an autonomous inspection system for agriculture precision based on the use of one or more drones. Authors in E. Nunes et al., 2017, use an FPGA card for MAS involving reconfigurable hardware in farming. The card used is a Xilinix FPGA with a soft-core processor, which can be easily extended with peripherals via the Xilinx Integrated Development Kit (EDK) [31]. Authors in [27] proposed an embedded system for aerial image processing. This system is based on the FPGA DE2 card and the Raspberry Pi 2 card. They used the Open Source Computer Vision (OpenCV) library. The system is divided into two blocks, the first for general processing through an Intel Atom N2600 @1.6 GHz, the second block ensures high flexibility for projects through an Altera Cyclone IV FPGA.

3 Use Case

The proposed algorithm is dedicated to agricultural soil monitoring to help farmers make subsequent decisions. This algorithm is based on receiving images and calculating the two plant indices NDVI and NDWI. Generally, the algorithm is divided in two parts. The first part for multispectral images with bands in the same image. This concept consists of separating the bands from the image then sending them to the processing station. The overall architecture of the algorithm is divided into three blocks. The first block dedicated to the separation of the bands to have images with different bands, Red, Green, and Blue. The second block is based on the reception of these images with different bands and calculates it from the two indices NDVI and NDWI. The third functional block applied a thresholding operation to classify the values of NDVI and NDWI.

The images used in this work are divided into three databases, the first [32] with a wavelength of 550 nm for images that have a green band, 790 nm for NIR and 660 nm for the red band, with a resolution of 1280 $\times$ 960 pixels, this database is the most usable and contains images collected by different cameras. In our case, we have chosen the parrot sequoia+ camera which gives multispectral images of UAV with separate bands. The other database [6] offers a resolution of 1280 $\times$ 960 pixels, this database also contains separate bands collected by a MAV; these images are generally valid for medium agricultural fields. The third database contains images collected by the TetraCam which gives only two images, one contains an RGB image and the other a NIR image. This database requires a separation of the bands with the calculation of the indices, the database has a resolution of 1296 $\times$ 966 pixels. Our work based on an Intel Core i5-3820 (8) CPU @ 2.60 GHz desktop computer with 8 GB RAM and a GeForce GTX 560 with 4046 MB RAM memory, and a heterogeneous embedded CPU-GPU card.

The results obtained showed that we can process 30 frames/s using the desktop and with the XU4 card we can process 20 frames/s. These results show an improvement in the recently published article [33]. This improvement is due to the use of the OpenMP parallelization language, the evaluation of these databases shows that the algorithm can process the different agricultural fields, which shows the performance of this algorithm. We have also evaluated the algorithm on several resolutions, we can process up to 400 frames/s using QVGA resolution images. Among the characteristics of this algorithm, we also find the low complexity which makes the algorithm easy to implement on different architectures.

4 Result and Discussion

In this work, we used the three databases to evaluate the different cases possible in agriculture. The images used in this work represent large area farms collected by a UAV, the second type concerns medium area farms, in this case, we used

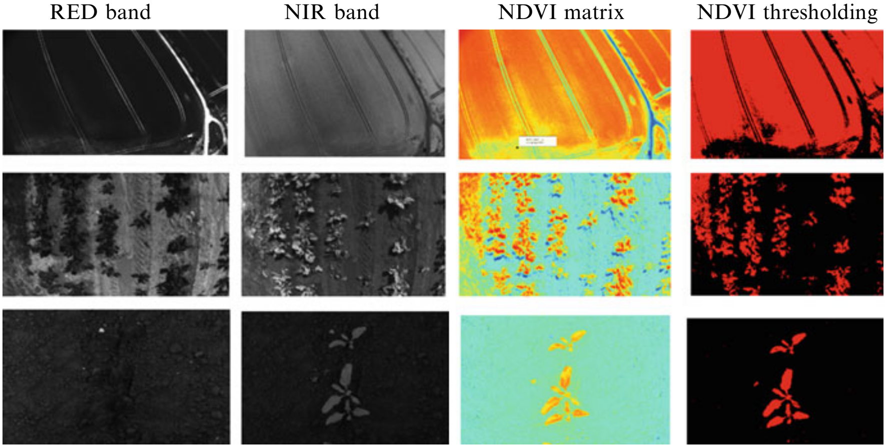


Fig. 1 Evaluation of the algorithm with the three databases using NDVI

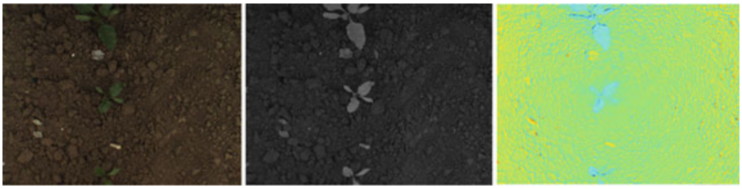


Fig. 2 Evaluation using NDWI

MAV images, the third concerns small area greenhouse, in this case, we used images collected by a soil robot. The results showed that our algorithm can process all three cases without any modification. This shows flexibility in the algorithm, in addition, the algorithm has low complexity. This complexity depends on the number of images used and also the resolution. Generally, the evaluation showed that our algorithm processes high-resolution images in real time. Figure 1 shows an evaluation of the proposed algorithm.

Figure 1 shows images that contain the *R* and NIR band, then an image that contains the NDVI value matrix. In this evaluation we used a threshold of 0.6 for images collected by UAVs, 0.4 for MAVs, and 0.2 for soil robots. The thresholding part in this assessment is very important for accurate monitoring it is necessary to know the exact values of the NDVI index to determine if the plant requires vegetation or not. Figure 2 shows an evaluation using this time the water index. The image used collected by a soil robot, we used the value 0.1 as a threshold.

After the validation of the algorithm, an embedded implementation was proposed in order to have an application that monitors agricultural fields based on a drone. The time results showed processing of 30 frames/s in a desktop and 20 frames/s using the embedded system XU4 with the parallel programming language OpenMP. The processing on the desktop shows that the results are much better than the XU4 card with ten more processing images. But the problem here is the portability of the XU4 card with its low weight and consumption, which is not the case in the desktop. Conventional machines such as laptops and desktops are powerful than embedded systems but the problem in applications that require low energy consumption and low weight. In our case, the choice of the XU4 card depends on its low weight of 60 g and its low consumption with 5 W. The execution times obtained after the evaluation shows that with the XU4 card we can process an image in a range between 0.047 s using OpenMP and 0.23 using the predefined OpenCV library. In the case of desktop, the time is 0.031 s using C/C++ and 2.76 s with the Matlab.

5 Conclusion

There are several algorithms applied to precision agriculture. These algorithms are more and more complexes in terms of processing times. If their validation on conventional machines such as PCs is verified, we are far from their implementation on embedded systems that can help making decision in real time. The future work consists in developing an algorithm providing several parameters, to control the ground state, implemented on an UAV embedded architecture. This objective requires a hardware software co-design approach that can help to validate a system with experimental data in real time.

References

1. Eason, G., Santhosh, K. S., Soizik, L., Grant, M. C., & George, A. S. (2003). Remote sensing applications for precision agriculture: A learning community approach. *Remote Sensing Environment, 88*, 157–169.
2. David, J. M. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering, 114*, 358–371.
3. National Research Council. (1997). *Precision agriculture in the 21st century* (p. 149). National Academy Press.
4. Nebiker, S., Annen, A., Scherrer, M., & Oesch, D. (2008) A light-weight multispectral sensor for micro UAV—Opportunities for very high resolution airborne remote sensing. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences; International Society for Photogrammetry and Remote Sensing (ISPRS):* Beijing, China, 2008 (Vol. 37(B1)).
5. Anderson, K., & Gaston, K. J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment, 11*(3), 138–146.
6. Sa, I., et al. (2018). weedNet: Dense semantic weed classification using multispectral images and MAV for smart farming. *IEEE Robotics and Automation Letters, 3*(1), 588–595.
7. Lottes, P., Behley, J., Milioto, A., & Stachniss, C. (2018). Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. *IEEE Robotics and Automation Letters, 3*(4), 2870–2877.
8. Maxwell, J. C. (1892). *A treatise on electricity and magnetism* (Vol. 2, 3rd ed., pp. 68–73). Clarendon.

9. Jiang, P., Chen, Y., Liu, B., He, D., & Liang, C. (2019). Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *IEEE Access, 7*, 59069–59080.
10. Bosilj, P., Duckett, T., & Cielniak, G. (2018). Analysis of morphology-based features for classification of crop and weeds in precision agriculture. *IEEE Robotics and Automation Letters, 3*(4), 2950–2956.
11. Gobbo, S., Lo Presti, S., Martello, M., Panunzi, L., Berti, A., & Morari, F. (2019). Integrating SEBAL with in-field crop water status measurement for precision irrigation applications—A case study. *Remote Sensing, 11*, 2069.
12. de Castro, A. I., Torres-Sánchez, J., Peña, J. M., Jiménez-Brenes, F. M., Csillik, O., & López-Granados, F. (2018). An automatic random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery. *Remote Sensing, 10*, 285.
13. Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment, 185*, 46–56.
14. Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., & Nõges, T. (2016). First experiences in mapping lake water quality parameters with sentinel-2 MSI imagery. *Remote Sensing, 8*, 640.
15. Jackson, R. D., & Huete, A. R. (1991). Interpreting vegetation indices. *Preventive Veterinary Medicine, 11*(3–4), 185–200.
16. McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing, 17*(7), 1425–1432.
17. Bastarrika, A., Alvarado, M., Artano, K., Martinez, M., Mesanza, A., Torre, L., Ramo, R., & Chuvieco, E. (2014). BAMS: A tool for supervised burned area mapping using landsat data. *Remote Sensing, 6*, 12360–12380.
18. Rikimaru, A., Roy, P. S., & Miyatake, S. (2002). Tropical forest cover density mapping. *Tropical Ecology, 43*, 39–47.
19. Fuldain González, J. J., & Varón Hernández, F. R. (2019). NDVI identification and survey of a Roman road in the Northern Spanish Province of Álava. *Remote Sensing, 11*, 725.
20. Khaliq, A., Comba, L., Biglia, A., RicaudaAimonino, D., Chiaberge, M., & Gay, P. (2019). Comparison of satellite and UAV-based multispectral imagery for vineyard variability assessment. *Remote Sensing, 11*, 436.
21. Rafał, P., Stuczynski, T., & Borzecka, M. (2012). The suitability of an unmanned aerial vehicle (UAV) for the evaluation of experimental fields and crops. *Zemdirbyste, 990014*, 431–436.
22. Ju, C., & Son, H. I. (2018). Multiple UAV systems for agricultural applications: Control, implementation, and evaluation. *Electronics, 7*, 162.
23. Mogili, U. M. R., & Deepak, B. B. V. L. (2018). Review on application of drone systems in precision agriculture. *Procedia Computer Science, 133*, 502–509.
24. Chen, J., Jönssonc, P., Tamurab, M., Gua, Z., Matsushitab, B., & Eklundhd, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment, 91*, 332–344.
25. Jin, B., Ye, P., Zhang, X., et al. (2019). Object-oriented method combined with deep convolutional neural networks for land-use-type classification of remote sensing images. *Journal of Indian Society of Remote Sensing, 47*, 951–965.
26. Peña, J. M., Torres-Sánchez, J., Serrano-Pérez, A., De Castro, A. I., & López-Granados, F. (2015). Quantifying efficacy and limits of unmanned aerial vehicle (UAV) technology for weed seedling detection as affected by sensor resolution. *Sensors, 15*, 5609–5626.
27. Silva, J. F. d., Brito, A. V., Lima, J. A. G. d., & Moura, H. N. d. (2015). An embedded system for aerial image processing from unmanned aerial vehicles. In *2015 Brazilian Symposium on Computing Systems Engineering (SBESC),* Foz do Iguacu, 2015 (pp. 154–157).
28. Torres-Sánchez, J., López-Granados, F., De Castro, A. I., & Peña-Barragán, J. M. (2013). Configuration and specifications of an unmanned aerial vehicle (UAV) for early site specific weed management. *PLoS One, 8*(3), e58210.

29. Zhou, C., Yang, G., Liang, D., Yang, X., & Xu, B. (2018). An integrated skeleton extraction and pruning method for spatial recognition of maize seedlings in MGV and UAV remote images. *IEEE Transactions on Geoscience and Remote Sensing, 56*(8), 4618–4632.
30. Doering, D., Benenmann, A., Lerm, R., de Freitas, E. P., Muller, I., Winter, J. M., & Pereira, C. E. (2014). Design and optimization of a heterogeneous platform for multiple UAV use in precision agriculture applications. *IFAC Proceedings Volumes, 47*(3), 12272–12277.
31. Nunes, É., Behnck, L., & Pereira, C. E. (2017). Multi-agent based implementation of an embedded image processing system in FPGA for precision agriculture using UAVs. In M. Götz, G. Schirner, M. Wehrmeister, M. Al Faruque, & A. Rettberg (Eds.), *System level design from HW/SW to memory for embedded systems. IESS 2015. IFIP advances in information and communication technology* (Vol. 523). Springer.
32. https://www.sensefly.com/education/datasets
33. Latif, R., Saddik, A., & Elouardi, A. (2019). Evaluation of agricultural precision algorithms on UAV images. In *2019 International Conference of Computer Science and Renewable Energies (ICCSRE),* Agadir, Morocco (pp. 1–4).

Deep Morphological Gradient for Recognition of Handwritten Arabic Digits

Mouhssine El Atillah and Khalid El Fazazy

Abstract Arabic is among the most spoken languages in the world today. Despite this, the optical recognition of Arabic manuscript characters by the algorithms of deep learning remains insufficient. Recently, some studies are moving towards this side and give remarkable results either for the recognition of alphabets or Arabic numbers. We highlight in this work a deep morphological gradient for the problem of recognition of Arabic manuscript digits. We use the multilayer perceptron (MLP) network used in the conference paper (Ashiquzzaman, A., & Tushar, A. K. (2017). *Handwritten Arabic Numeral Recognition using Deep Learning Neural Networks*), which is preceded by the morphological gradient algorithm to detect the contours of the digits. This model is applied to the database of Arabic manuscript digits available on Kaggle (Arabic Handwritten Digits Dataset), which consists of 70,000 images. The classification accuracy of the model was 99.9% with a very minimum loss of 0.09%.

Keywords Deep learning · Multilayer perceptron · Morphological gradient · Optical character recognition

1 Introduction

The OCR (Optical Character Recognition), created by the German engineer Gustav Tauschek in 1929, operated with a photosensitive detector that compared the characters to the template contained in his memory. The idea has made its way since, until digitization processes are improved thanks to online systems: today, OCR technology makes it possible to photograph handwritten or typed documents

M. E. Atillah (✉)
Computer Science, Signals, Automation and Cognitivism Laboratory (LISAC), FSDM, USMBA, Fes, Morocco

K. El Fazazy
Research Team Complex Systems and Interactions-Faculty of Applied Sciences, Ait Melloul, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_14

Table 1 The difference between Arabic and Roman digits

Arabic digits	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
English digits	0	1	2	3	4	5	6	7	8	9

and to organize the information by categories [1]. We distinguish two handwriting recognition systems, offline and online systems [2]. The OCR is interested in old documents through the transcription automation of their textual contents, given the complex and irregular nature of writing [3]. The recognition of Arabic manuscripts is slowly developing compared to other languages [4] despite the fact that Arabic is the main language of North Africa and the Middle East. Statistically, it is the fifth language spoken in the world today [5, 6]. There are around 250 million Arabic speakers [7]. In the last two years, some scientific research in the field of recognition of manuscript characters by deep learning algorithms is moving towards Arabic manuscripts [8, 9]. The difference between Arabic and the main Roman languages lies in the fact that the Arabic words, as well as the characters contained in the words, are written from right to left. This difference is shown in Table 1. The machine learning methods usually apply a combination of the following feature extractors by classifiers to wait for the targeted classes. The convolutional neural networks (CNNs) are the most used in the recognition of manuscripts in general thanks to their efficiencies. One study is based on the two databases of Arabic characters AIA9k and AHCD and uses a CNN with regularization parameters such as batch normalization to avoid over-fits and that gives classification accuracies of 94.8% and 97.6% [10], respectively, for the two databases. Another study is based on a database of 16,800 manuscript Arabic characters in which the optimization methods are implemented to increase the performance of CNN. The results of this method give an average misclassification error of 5.1% on the test data [11].

The recognition of Arabic numerals is also a case study because of its difference to Roman numerals. We will highlight two studies in this side. The first proposes a new approach based on the restricted Boltzmann machine (RBM) and the convolutional neural network (CNN). RBM is a deep learning technique that can extract extremely useful features from raw data. This study is given the highest accuracy rate compared to the CMATERDB 3.3.1 database for handwritten Arabic numerals [12]. The second proposes a new model based on convolutional neural networks using an appropriate activation function and a regularization layer, which gives considerably improved accuracy compared to the existing Arabic numerical recognition methods. The proposed model gives an accuracy of 97.4%, which is the highest recorded accuracy of the dataset used in the experiment. The study also proposes a modification of the method described in [13], where it gives a precision with a value of 93.8%.

In this project, we focus on the recognition of Arabic numerals using the low data available on Kaggle [14].

2 Method

2.1 Motivation

Studies in the field of recognition of Arabic manuscript digits [5, 12] are numerous and remain insufficient. The results obtained by the modules that deal with Arabic digits are remarkable and tend towards 97.4% for the study in [13]. These results are insufficient because of the binary nature of the processed images.

2.2 Architecture

We highlight our method shown in Fig. 2 and how it improves the accuracy of database recognition processed. It is based on the MLP method used in [13] preceded by the morphological gradient method. The latter detects the contours of processed images using the difference between morphological dilation and erosion and is preceded by the inverted binary thresholding method. Figure 1 shows the morphological gradient of each Arabic number.

The dilation and erosion have almost the same principle, except that the first expands the white part of image (it is the body of alphabet). A 2×2 core containing one is used. In the same way, the second erodes the white part of the image. The deference between them gives us the outline of the alphabet. The result of the gradient method occurs at the three-layer multilayer perceptron (MLP) (input, output and masked layer). The input layer receives the densities of the pre-segmented image pixels (morphological gradient images) as a size vector 1024. The first and the second layers are composed of 512 neurons with ReLU as affinity activation function to avoid the negative values at the exit of each neuron. The last layer composed of 10 neurons is linked to the target classes (the ten numbers) with softmax as an activation function for each neuron to render the results as categorical probabilities between 0 and 1 taking a predicate for a class of the ten classes (the numbers treated) is true. The Adam Optimizer is used during the training process to get the best model possible.

In general, the process of a deep learning model goes through two stages: forward propagation and backpropagation in order to obtain the best peas and rough values.

In our case, the treatment process begins with the morphological gradient method before moving to the MLP as shown in Fig. 2. In this part, we focus on the mathematical side of our proposed model.

Beginning with the morphological gradient written as follows:

$$Grad(X) = (X \oplus B) - (X \ominus B) \tag{1}$$

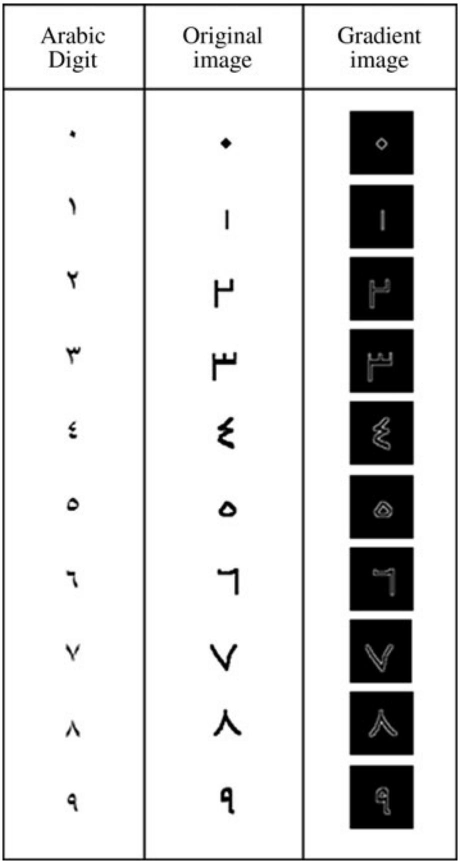


Fig. 1 Morphological gradient images of Arabic digits

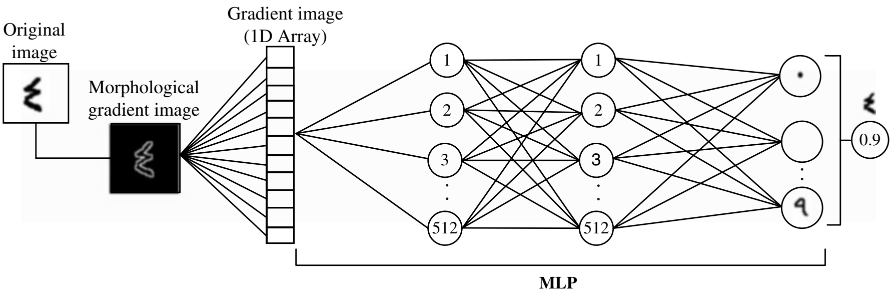


Fig. 2 The description of deep morphological gradient method

where X: E→R is a grayscale image, E a Euclidean space or a discrete grid E (such as R^2 or Z^2) in the real line, b(x) a structuring element in grayscale and ⊕ and ⊖, respectively, denote the dilatation and erosion.

2.2.1 Forward Propagation

As the name suggests, input data is routed downstream over the network. Each masked layer accepts the input data, processes it according to the activation function and moves on to the next layer. The ReLU activation function is used (the last layer uses softmax). Starting with neuron 1 of the first layer, we have

$$f^{in} = \sum_{i=1}^{784} w_i Grad(x_i) + b_1 \tag{2}$$

$$f^{out} = Relu(\sum_{i=1}^{784} w_i Grad(x_i) + b_1) \tag{3}$$

For all the neurons of the first layer, we write this in matrix form, and we will have

$$h^{in} = \begin{bmatrix} x_1 \cdots x_{784} \end{bmatrix} \begin{bmatrix} w_1^1 & \cdots & w_1^{512} \\ \vdots & \ddots & \vdots \\ w_{784}^1 & \cdots & w_{784}^{512} \end{bmatrix} \begin{bmatrix} b_1 \cdots b_{512} \end{bmatrix} \tag{4}$$

If we represent the entries as matrix I, weight W and bias B, we obtain

$$f^{in} = I * W + B \tag{5}$$

$$f^{out} = Relu(f^{in}) \tag{6}$$

Generally, we can write

$$f_0^{out} = I \tag{7}$$

$$f_i^{in} = f_{i-1}^{out} * W_i + B_i \tag{8}$$

$$f_i^{out} = F_i(f_i^{in}) \tag{9}$$

where i is the layer number and F is the activation function for a given layer.

2.2.2 Backpropagation

This is the essence of neural network formation. It is the practice of refining the weights of a neural network according to the error rate (i.e. loss) obtained in the previous epoch (i.e. one iteration). Proper weight adjustment guarantees lower error rates, making the model reliable by increasing its generalization. We use the function Cross Entropy as loss function for our module which is defined as follows:

$$E = -\sum_{j}^{n} y_j * ln(\hat{y}_j) \tag{10}$$

where n is the number of classes, y the vector of the network output and y the vector of true tags.

In order to find the error gradients with respect to each variable, we will start with the last layer, and taking a partial derivative of the loss in relation to the weight of the neurons, we get

$$\frac{\partial E}{\partial W_3} = \frac{\partial E}{\partial f_3^{in}} * \frac{\partial f_3^{in}}{\partial W_3} \tag{11}$$

Knowing that in case of $softmax$ activation function and $crossentropy$ loss, we have

$$\frac{\partial E}{\partial f_3^{in}} = \hat{y} - y \tag{12}$$

Now, we can find a gradient for the last layer as

$$\begin{aligned} \frac{\partial E}{\partial W_3} &= (\hat{y} - y)\frac{\partial f_3^{in}}{\partial W_3} \\ &= (f_2^{out})^T \sigma_3 \quad \text{where } \sigma_3 = (\hat{y} - y) \end{aligned} \tag{13}$$

Similarly, for layers 2 and 1, we find successively

$$\frac{\partial E}{\partial W_2} = (f_1^{out})^T \sigma_2 \tag{14}$$

$$\frac{\partial E}{\partial W_1} = (f_0^{out})^T \sigma_1 \tag{15}$$

The same thing for the bias, we have

$$\begin{aligned} \frac{\partial E}{\partial B_3} &= (\hat{y} - y)\frac{\partial f_3^{in}}{\partial B_3} \\ &= (\hat{y} - y) \\ &= \sigma_3 \end{aligned} \tag{16}$$

and σ_2 and σ_1 for layers 2 and 1 successively.

We can now follow a common pattern, which can be generalized as follows:

$$\frac{\partial E}{\partial B_i} = \sigma_i \tag{17}$$

With these equations, we can calculate the error gradient as a function of each weight/bias. To reduce the error, we need to update our weights/bias in a direction opposite to the slope.

3 Experience

We apply our proposed model to the Arab Numbers database available on Kaggle [14] of 60,000 training images and 10,000 for validation. The total size of the parameters of our proposed model is 969706. The execution is formed for 10 iterations. The optimizer Adam is used as an optimization function, while the binary entropy is used to calculate the loss. The model is implemented in the Python programming language with Keras [15] and backend Tensoflow [16] libraries. The experimental configuration computer had an Intel CPU i3-3120M CPU @ 2.50 GHz with 4 Gb of RAM.

4 Results and Discussion

Our proposed method gives very effective results compared to those of MLP method in [13]. The database accuracy of training tends to 99.9%, which is unusual in image recognition. The validation samples also give satisfactory results that tend towards 99.4%, which shows the restriction of knowledge of the new images outside the training database. The loss of data during the execution process is almost zero at 0.09% for the trapping data. Figures 3 and 4 show the variation of precision/loss with respect to the number of iterations performed for the two databases (training and validation).

In order to obtain the best model for forming the network, many experimental configurations have been tested to improve network performance. A 2×2 core is used as a filter for the morphological gradient method to obtain clear contours of the numbers processed. The neural network is increased by the addition of a layer of 128 neurons, but the precision is diminished.

The results obtained by our model are more efficient than the MLP method in the conference paper [13] as shown in Table 2.

Table 2 Precision of the proposed method and that in [13]

Method	Accuracy
MLP method in [13]	93.8%
Deep morphological gradient	99.9%

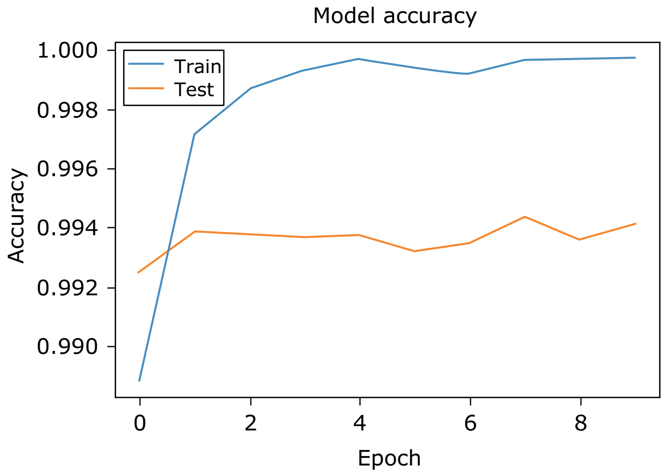


Fig. 3 The variation of precision for deep morphological gradient method

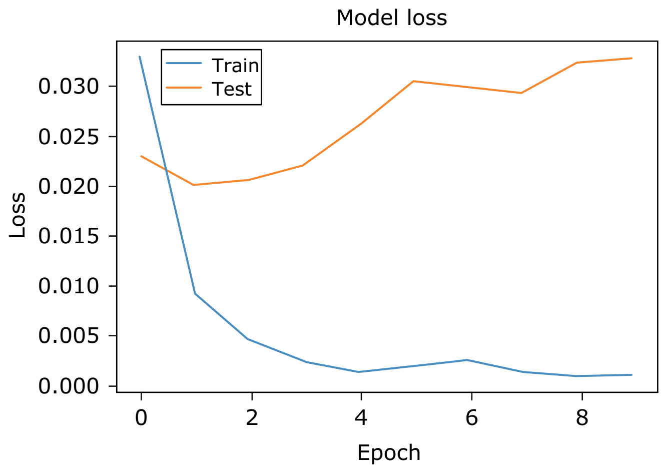


Fig. 4 The variation of loss for deep morphological gradient method

5 Conclusions and Perspectives

Automated software systems used in huge applications in many industrial sectors could be enriched by the remarkable development of recognition of Arabic manuscripts. In this study, we present a deep morphological gradient based on a multilayered neural network (MLP) proceeded by the morphological gradient algorithm. It is able to classify Arabic manuscript figures with a classification accuracy of 99.9% for the Kaggle database [14].

In the same place and as future work, we plan to develop a model that recognizes Arabic alphabets and numbers at the same time.

References

1. Jouini, K. (2019). *Dématérialisation: : La reconnaissance optique de caractères (OCR) doit accélérer sa disruption et élargir son champ d'application*, Economie Matin, Accessed 29 08 2019. http://www.economiematin.fr/news-reconnaissance-optique-application-ocr-innovation-industrie-jouini.
2. Plamondon, R., & Srihari, S. N. (2000). On-line and Off-line handwriting recognition: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*, 63–84.
3. Belaïd, A., & Ouwayed, N. (2011). Guide to OCR for Arabic Scripts. Märgner, V., & El Abed, H. (pp.103–122). London: Springer-Verlag.
4. Abandah, G., Khedher, M., & Younis, K. (2008). Handwritten arabic character recognition using multiple classifiers based on letter form. In *IEProceedings of the 5th IASTED International Conference on Signal* (vol. 22, pp. 128–133).
5. Das, N., Mollah, A. F., Saha, S., & Haque, S. S. (2006). Handwritten arabic numeral recognition using a multi layer perceptron. In *IEProceedings of the 5th IASTED International Conference on Signal*.
6. Sherif, A., & El-Sherif, E. (2008). *Arabic Handwritten Digit Recognition, 11*, 127–141. https://doi.org/10.1007/s10032-008-0073-5.
7. IPAC, Team. (2019). Quelles sont les 10 langues les plus parlées au monde? Accessed 28 07 2019. https://www.ipac-traductions.com/blog/10-langues-plus-parlees.
8. El-Sawy, A., EL-Bakry, H., & Loey, M. (2016). *CNN for handwritten arabic digits recognition based on LeNet-5*. Springer International Publishing.
9. AL-Saffar, A., Awang, S., Al-Saiagh, W., Tiun, S., & Al-khaleefa, A. S. (2018). Deep learning algorithms for Arabic handwriting recognition. *International Journal of Engineering and Technology (UAE), 7*, 344–353.
10. Younis, K. S. (2017). Arabic handwritten character recognition based on deep convolutional neural networks. *Jordanian Journal of Computers and Information Technology (JJCIT), 3*.
11. El-Sawy, A., Loey, M., & EL-Bakry, H. (2017). *Arabic Handwritten Characters Recognition using Convolutional Neural Network*.
12. Alani, A. A. (2017). *Arabic Handwritten Digit Recognition Based on Restricted Boltzmann Machine and Convolutional Neural Networks, 8*. https://doi.org/10.3390/info8040142.
13. Ashiquzzaman, A., & Tushar, A. K. (2017). *Handwritten Arabic Numeral Recognition using Deep Learning Neural Networks*. https://doi.org/10.1109/ICIVPR.2017.7890866.
14. Arabic Handwritten Digits Dataset. Kaggle. https://www.kaggle.com/mloey1/ahdd1.
15. Chollet, F. (2015): *Keras: Deep Learning for humans*. Accessed 05 09 2019 https://github.com/keras-team/keras.
16. Google Tensorflow (2015). Accessed 12 09 2019. https://www.tensorflow.org/.

Strategic Prediction of Sustainable Development Territory on the Basis of Convergence Processes

Halyna Raiko, Anzhela Hryhorova, Denis Khapov, and Maryna Fedorova

Abstract The essence of theoretical principles and some analysis indicator methods of territorial socioeconomic development are analyzed in the chapter. The authors conduct considerable research of a territory using a large number of different database values, which were later used in the studies. They present a practical implementation of indicators value forecasting methods for the socioeconomic development of the territory. They propose information technology modules for the automation of socioeconomic analysis and develop the appropriate software.

Keywords Development · Territory · Strategy · Sustainable Development Goals

1 Problem Statement

The European vector of Ukraine's development focuses on the integration of processes aimed at promoting the stabilization of economy, technology, science, attracting foreign investment. The state of Ukraine's development as a whole and as individual communities (territories) today depends on the development of effective strategies and the implementation of high quality projects. It should be noted that most transformations in all areas happen for the first time, and so they require specialists to develop and implement new methods and models in decision-making and territorial management.

The importance of the research topic of the article on the implementation of the 17 Sustainable Development Goals (MDGs) in Ukraine and 169 implementation tasks is emphasized, approved at the UN Sustainable Development Summit in 2015. In the area of sustainable development, the program is expected to achieve its goals by 2030, covering all levels: global, national and territorial (regional and local).

H. Raiko (✉) · A. Hryhorova · D. Khapov · M. Fedorova
Department of Information Technology, Kherson National Technical University, Kherson, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_15

Most of the countries in the world will have implemented the CDC into Sustainable Development Program by 2030.

Major international financial and charitable organizations have chosen CSR as their "road map." This vector has directed a significant amount of global, public and private investment to meet CSD indicators at all levels of management. In turn, investments should contribute to innovative transformations across the four dimensions of CSD: economic, social, environmental and institutional.

In Ukraine, the United Nations System and the United Nations Development Program (UNDP), in particular, support the implementation of the Procedure by 2030 and CSDs as national, regional and local levels. Today, two important tasks are needed to ensure sustainable development for the people of Ukraine: the creation of an institutional mechanism and the localization of CSF in regions and small communities.

Among the scholars who are engaged in convergence in the regional economy, we can distinguish the works of well-known neoclassicists, such as R. Barro, H. Sala-i-Martin, D. Qua, J. Borts, J. Stein, who see the end result of convergence (approx.) in revenue levels and regional development rates. Many supporters of the theory of cumulative causality are also known, namely J. Friedman, P. Krugman, M. Fujita, P. Martin, and J. Ottaviano, who "predict" the divergence of regions according to these indicators.

The socioeconomic development of local communities is the foundation upon which primary resources, financial flows, and bases of regional division of labor and industrial cooperation, form stable connections and interdependencies in socioeconomic processes. Local economic development reproduces the links that form internal regional and national markets, strengthening the unity of the national economy [1].

Economic growth is only possible in a community if there is a strategy that defines the development vector and the smart specialization of the territory. The development of an effective and well-grounded strategy is based on a complex and long-lasting process of socioeconomic analysis that enables the determination of specificity, priorities, goals, and development scenarios. Using information technology to automate this process will allow one to optimize the time spent and obtain scientifically valid criteria values for the socioeconomic development of the territory, which can be used to set strategic and operational goals for local development. Therefore, the topic of the study is relevant.

2 Methodology

The invariance of current regional governance methodologies indicates that modern communication tools, technologies, the increased availability of information, and accelerated rates of transformation enable the quick adaptation of different methodologies to meet management challenges under decentralized conditions. In this regard, the processes of convergence and integration of the elements of such a system into existing methodologies are updated to ensure their further harmonization.

The methods for determining the convergences of systems and their types have become more relevant in the context of decentralization and comparisons between communities or regions. Convergences may be:

– Absolute (unconditional), when only the level of initial development is included in the composition of exogenous variables;
– Conditional, which implies taking into account the influence of endogenous and exogenous factors (geopolitical position; natural resources, ecological potential; human, innovative, and financial capital; the level of capitalization of enterprises, development of infrastructure, including innovative);
– Sigma-convergence, which implies a decrease in the time differences in convergence levels;
– Beta-convergence, where there is a negative relationship between the rate of economic growth and the initial level of development of the territory.

Erwin Laszlo defined the concept of the "age of bifurcation": "Fortunately, the boundaries of scientific disciplines are not eternal. These borders are a legacy of the past and they are now obsolete. Each of the new major fields of research—the new physics, the new biology and the new systems sciences—seek and find the traits of unity in the diversity of the world" [2]. This statement can be taken as the basis of a scientific approach to conceptualizing the formation of convergent thinking, which can create a desired image and forecast the future.

The interpretation of divergence in economic vocabulary is interpreted as movement in divergent vectors: the widening of the gap between levels of development of territories (macro-, subterritories) and increases in differences between economic models, their structures and mechanisms. The term "divergence" is also used to characterize the deviation of individual territories from average macroeconomic indicators [3].

Convergence is the opposite of divergence—increases in closeness between the economies of different territories. Integration and globalization processes drive convergence. Planners of effective territory strategy should consider divergence processes.

We describe the approaches used to determine convergence. The use of Barro regression is appropriate for the analysis of absolute or conditional convergence on the basis of spatial sampling: in the sample, the convergence growth rate of average per capita income should be negatively dependent on some level of average per capita income [4]:

$$\frac{\ln y_{iT} - \ln y_{i0}}{T} = a + b \ln y_{i0} + \varepsilon_i, \tag{1}$$

$$\frac{\ln y_{iT} - \ln y_{i0}}{T} = a + b \ln y_{i0} + X_i' \gamma + \varepsilon_i, \tag{2}$$

where y_{i0} is the actual GVA per person-territory i at the beginning of the analyzed period; y_{iT} is the real GVA per person-territory i at the end of the analyzed period; T is the duration of the analyzed period; X is a set of additional repressors that are the factors of conditional convergence.

The hypothesis of a model of β-convergence with minimal conditions or minimal-conditional convergence [5] assumes that the territories are on different trajectories of proportional development. This situation arises if the dynamics of economic development of a given territory is determined by the dynamics and/or the neighbors' level of development. This model allows us to test the hypothesis of the presence of spatial clustering of average development rates with the help of an endogenous spatial lag of average VDV development rates per person. The model of minimal-conditional β-convergence in the specification of a spatial lag model is as follows [6]:

$$\frac{\ln\left(y\gamma_{i,T}/\ln y\gamma_{i,0}\right)}{T} = a - \beta \ln yr_{i,0} + \delta Sg_{i,T} + \varepsilon_i, \tag{3}$$

where $\frac{\ln\left(y\gamma_{i,T}/\ln y\gamma_{i,0}\right)}{T}$ is the logarithm of average growth rate of real GVA per person; $\ln yr_{i,0}$ is the logarithm of GVA per person.

This model allows us to account for the spatial autocorrelation of residues by including the quality of an explanatory variable of endogenous spatial lag on the logarithm of the average rate of development of GVA per person (Sgi, T).

It is essential for the convergent-divergent approach in territory management to use adaptive forecasting methods: the moving average forecast method, the exponential average method, the Brown method, the Holt and Holt-Winters models, the error correction method (Trigg method), the Trigg-Leach method, and the Chow method. When choosing a forecasting method, preference is given to a method that, under similar other conditions, enables updating (corrections) the forecast with less cost [7].

The Brown method uses exponential averages of different orders to construct polynomial models. The theoretical basis of this method is the Mayer-Brown theorem, the essence of which is that the value of a function and its derivatives can be expressed with given accuracy using exponential averages of different orders. The minimizing function of the model is as follows:

$$F\left(a_0, a_1\right) = \sum_{j=1}^{\infty} \Big(y_{t-j} - (a_0 - a_1 j)^2 \beta^j \to \min_a, \tag{4}$$

system of normal equations:

$$\begin{cases} \sum\limits_{j=0}^{\infty} \beta^j a_0 - \sum\limits_{j=0}^{\infty} j\beta^j a_1 = \sum\limits_{j=0}^{\infty} y_{t-j}\beta^j \\ \sum\limits_{j=0}^{\infty} j\beta^j a_0 - \sum\limits_{j=0}^{\infty} j^2\beta^j a_1 = \sum\limits_{j=0}^{\infty} jy_{t-j}\beta^j \end{cases}. \tag{5}$$

Auxiliary ratios:

$$\begin{cases} \sum\limits_{j=0}^{\infty} \beta^j = \frac{1}{1-\beta} = \frac{1}{\alpha}, e\alpha + \beta = 1(0 < \beta < 1) \\ \sum\limits_{j=0}^{\infty} j^2 \beta^j = \frac{\beta}{(} 1 + \beta)(1-\beta)^3 = \frac{\beta}{\alpha}^3 (1+\beta) \end{cases}.$$

System (5) is transformed into the form:

$$\begin{cases} a_0 - \frac{\beta}{\alpha} a_1 = \alpha \sum\limits_{j=0}^{\infty} y_{t-j} \beta^j \\ \beta a_0 - \frac{\beta}{\alpha} (1+\beta) a_1 = \alpha^2 \sum\limits_{j=0}^{\infty} j y_{t-j} \beta^j \end{cases}. \tag{6}$$

After substituting expressions and making simple groupings, we will have:

$$q_t^{(2)} = \alpha \sum_{j=0}^{\infty} j y_{t-j} \beta^j + \alpha q_t^{(1)}, \quad \text{or} \quad \alpha^2 \sum_{j=0}^{\infty} j y_{t-j} \beta^j = q_t^{(2)} - \alpha q_t^{(1)}.$$

Thus, system (6) can be rewritten as follows:

$$\begin{cases} a_0 - \frac{\beta}{\alpha} a_1 = q_t^{(1)} \\ \beta a_0 - \frac{\beta}{\alpha} (1+\beta) a_1 = q_t^{(2)} - \alpha q_t^{(1)} \end{cases}. \tag{7}$$

Brown's model predicts that by the time $t = \eta$, the values will change according to a linear law with increment. As a result, Brown's model looks like:

$$\hat{y}_{n+L} = a_0(n) + a_1(n) L. \tag{8}$$

Estimation of the forecast:

$$s_e^2 = \frac{\sum\limits_{t=1}^{n} \left(y_t - \hat{y}_t\right)^2}{n-1} = \frac{\sum\limits_{t=1}^{n} e_t^2}{n-1}. \tag{9}$$

The variance of the predicted values is calculated using the formula:

$$S_{\hat{y}}^2 = \gamma(\beta, L) s_e^2. \tag{10}$$

The estimation of value variance $\hat{y}_{t+L}$ is calculated as follows:

$$S_{\hat{y}_{t+L}}^2 = S_e^2 + \gamma(\beta, L) s_e^2 = (1\gamma + (\beta, L)) s_e^2. \tag{11}$$

The confidence interval will be:

$$\left(\hat{y}_{t+L} - t_a(k)s_{y_{t+L}}; \hat{y}_{t+L} + t_a(k)s_{y_{t+L}}\right), \tag{12}$$

where a is the level of significance; k is the number of degrees of freedom, $k = \eta - 2$.

In Holt's model, the parameters $a_0(t)ia_1(t)$ are estimated using two moving averages with different independent smoothing parameters. The coefficient $a_1(t)$ is estimated as the exponential mean of increments of the parameter $a_0(t)$. Then, according to Holt's model:

$$a_1 = \alpha_1 p(t) + (1 - \alpha_1)\,\alpha_1\,(t - 1)\,, \tag{13}$$

where $0 \le \alpha_1 \le 1$ is the first smoothing parameter [8].

$a_0(t)$ is the exponential mean of the levels of the series, calculated with the correction for the previous gain $a_0(t{-}1)$:

$$a_0(t) = \alpha_0 y(t) + (1 - \alpha_0)\,\alpha_0\,(t - 1) + (1 - \alpha_0)\,\alpha_1\,(t - 1)\,, \tag{14}$$

where $0 \le \alpha_1 \le 1$ is the second smoothing parameter, which does not depend on ax.

The forecast at time t for L steps ahead is created using the formula:

$$\hat{y}_{t+L} = a_0(t) + a_1(t)L. \tag{15}$$

Marking the forecast error made at the moment $(t - 1)$ on moment t from e_t as$e_t = y_t - \hat{y}_t = y_t - a_0\,(t - 1) - a_1\,(t - 1)$, we can rewrite Eqs. (13) and (14) as:

$$a_0(t) = \alpha_0\,(t - 1) + \alpha_1\,(t - 1) + a_0 e_t, \tag{16}$$

$$a_1(t) = \alpha_1\,(t - 1) + \alpha_0 a_1 e_t. \tag{17}$$

The Holt model uses two independent smoothing parameters—a_0 and a_1. Holt used the value $a_0 = 0$, 1 and $a_1 = 0.01$. The value of a_0 should be chosen as the average of several initial values of row levels, and the value of a_1, as the average of several initial values for first level differences. For the general case, the estimation of the prediction error when using the Holt model is a rather laborious task. Its approximate value, which is adequate when using the model in short-term forecasting, can be obtained using a generalized Brown model:

$$\operatorname{var}\left(\hat{y}_{t+L}\right) \cong \left(1, 25\left(1 - \beta_1^2\right) + \left(1 - \beta_1^2\right)^2 L\right)\sigma_\varepsilon^2, \tag{18}$$

where $\beta_1 = \alpha_0\alpha_1$ is the generalized smoothing parameter.

Chow's method is based on the idea of adapting forecasts by using three predictions at a time. These projections are based on exponential means at different parameter values. In terms of model sensitivity, they can be interpreted as predictions made on weak, medium, and highly sensitive models. For example, the forecasts may be obtained at $\alpha_1 = 0.1$; $\alpha_2 = 0.2$; $\alpha_3 = 0.3$. In his experiments, Chow used the following three-parameter scheme: $\alpha_1 = \alpha_1 - h$; $\alpha_2 = \alpha$; $\alpha_3 = \alpha + h$, where $h = 0.05$.

The average forecast is the main one. This projection is considered to be the actual forecast. If, at some stage, one of the extreme forecasts improves by the selected forecast criterion, the parameter values change automatically. In this case, the model with the best value of criterion at the moment becomes main and it is based on forecast for the next period. Other values of α change depending on the direction of the forecast. If the forecast is better with $\alpha_3 = 0.3$, then in the new forecasting system, starting from now, the following parameter values are taken: $\alpha_1 = 0.2$; $\alpha_2 = 0.3$; $\alpha_3 = 0.4$, at h—0.1. If the best forecast was at $h = 0.1$, then there may be new values $\alpha_1 = 0.05$; $\alpha_2 = 0.1$; $\alpha_3 = 0.2$ [9].

Suppose at time t the best forecast is at values $\tilde{q}_{t-1}$i $\tilde{\alpha}$. Then, the forecasts for the next period are based on the formula:

$$\hat{y}^i_{t+1} = q^i_t = \tilde{\alpha}^{(i)} y_t + \left(1 - \tilde{\alpha}^{(i)}\right) \tilde{q}_{t-1}. \tag{19}$$

Initially, the central value of the parameter may be arbitrary. As experience shows, it is best to take it equal to 0.2. The parameter step may also change depending on the situation.

The choice of the criterion for estimating the forecast is important in the Chow method. Here are several options for the criteria: minimum current absolute error, minimum average error in recent periods, minimum average weighted error, etc. In his experiments on the forecast (sale) of goods, Chow showed that the use of his model reduces the error of the forecast by 10% on average [10].

3 Research Results

The practical implementation of aforementioned modeling methodology was implemented using the Vinogradov United Territorial Community (OTG), formed by decision of the Kherson Regional Council No. 1266 from 05.06.2015. The administrative center of the Vinogradov community is the village of Vinogradovo.

The Vinogradov OTG occupies 20.6% of the Oleshkiv district and 1.3% of the Kherson region. The population of community is 16.5% of total population of the district and 1.1% of population of the region. The population density is 31 persons/km^2, while the average in Ukraine is 73 persons/km^2.

The researchers have carried out an analysis of indicators of socioeconomic development of the Vynogradska ATG. Figure 1 presents the population structure of the Vynogradiv OTG, taking into account the population of the villages that are part of it.

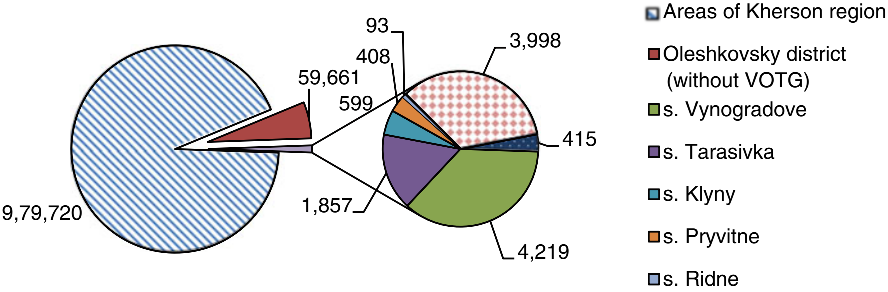


Fig. 1 Population structure of Vynogradiv OTG, rayon, district for 2017

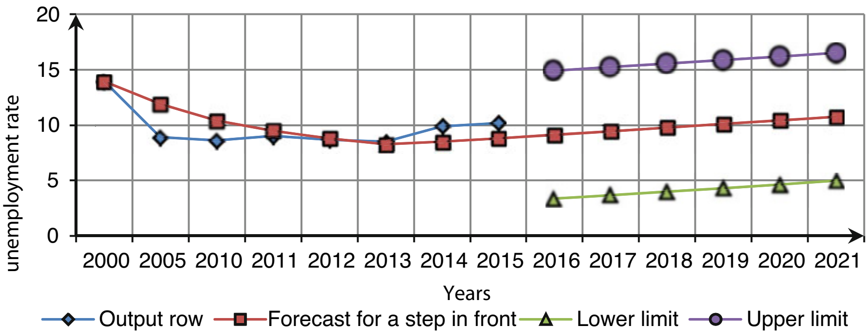


Fig. 2 Unemployment forecast for the linear Brown model

The ratio of births and deaths for the years 2010–2017 shows a declining trend in the number of deaths in the community during the period 2010–2017, but the death rate is much higher than the number of births per year.

According to the analysis, the largest average annual absolute increase in local budget revenues over 6 years (2012–2017) is in the payment for land—380,92 thousand hryvnias. This significantly increased the tax revenues together and tax on agricultural producers, specifically—223,62 and 203,67 thousand UAH in accordance.

The analysis of social indicators is important for the further development of the territory. The social development of the territory has not only social but also economic importance for the development of society. The economic value manifests in the fact that the social sphere provides the services necessary for the reproduction of the workforce, the growth of its labor productivity. The social role of this field is to increase the earnings level and improve the living conditions of people. According to the Brown model forecast (Fig. 2), there is a trend of increases in unemployment. The forecast value for 2016 is 9.13%, and by 2021, the unemployment rate will reach 10.74%. The value of the indicator can vary from 4.95% to 16.52%.

Performing the Holt model forecast, we set the following coefficient values $a_0(0)$ and $a_1(0)$: $a_0(0) = 10.5$, $a_1(0) = -1.6$, the value of the smoothing parameters α_0

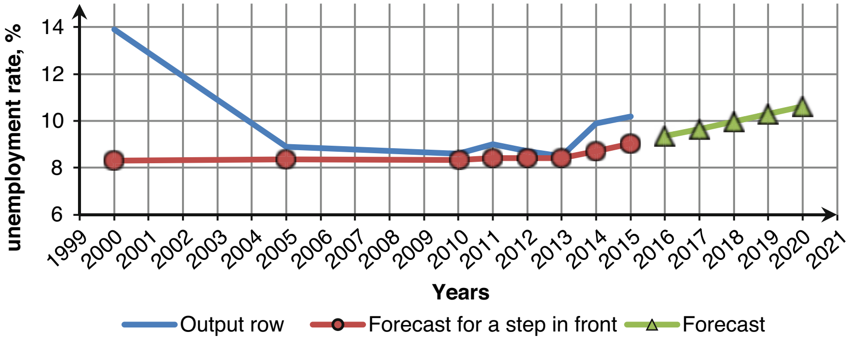


Fig. 3 Unemployment forecast for the Holt model at $a_0 = 0.2$ and $a_1 = 0.1$

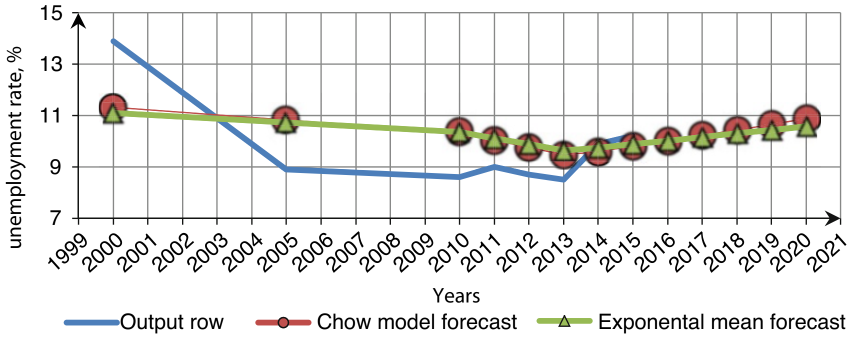


Fig. 4 Chow unemployment rate forecast

and α_1, as 0.2 and 0.1 in accordance. We calculate the coefficients $a_0(t)$ and $a_1(t)$, a step forward forecast, the deviation between the actual series levels and the forecast values, and the square of these deviations (Fig. 3). According to the forecast for the next five periods, the unemployment rate is increasing: in 2016 it was 9.34%., and in 2020, it will reach 10.61% unless appropriate measures are taken to improve the situation.

We predict the value of the Chow unemployment rate. The initial values of the smoothing parameters and the exponential mean are $\alpha_1 = 0.1$; $\alpha_2 = 0.2$; $\alpha_3 = 0.3$; $q_0 = 10.5$, forecast by the exponential mean we construct at $\alpha = 0.2$ and $q_0 = 10.5$. In 2020 this value may reach to 10.9% by the Chow method and 10.58% by the exponential average unemployment rate (Fig. 4).

We analyze the relationship between social indicators such as household income, unemployment, and average monthly nominal wage using a 3D surface graph (Fig. 5).

After analyzing the graph from Fig. 5, we can say that the income of population and the average nominal wage are directly proportional, and the level of unemployment and income are inversely proportional.

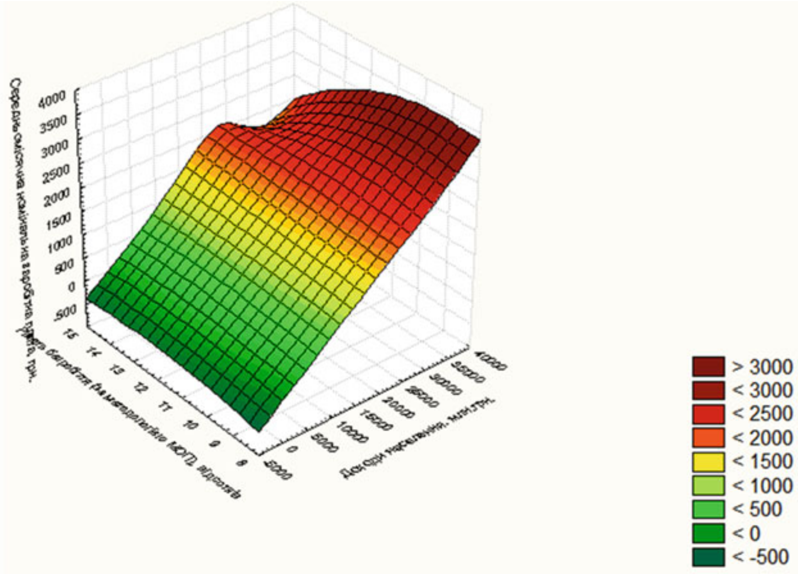


Fig. 5 Graph of correlation between unemployment rate, household income and average monthly wage

The performed analysis will be the basis of information technology used for the analysis of social indicators of the development of the territory, which can be used for the analysis of indicators of the Vynogradiv CTS, which allows to automate the process of analysis. This technology has been designed with the goal of analyzing metrics. The following software modules of the information system have been implemented: “Criteria for the development of Vynogradiv ATS,” “Historical background,” “Summary of the community,” “Summary of the region and district,” “Official site of the community,” “Database of Vynogradiv ATS,” “Development Strategy 2018–2023,” “User’s Guide.” The main menu interface of the system is shown in Fig. 6.

The module “Database of Vynogradiv OTG” has been implemented in MS Access 2010, contains data on the population of the community, the distribution of population by age and age by settlements, and statistics of natural and migratory movements of the population. After analyzing the indicators of the development of the territory, the user can immediately make adjustments to the community development strategy, using the appropriate module “Development Strategy 2018-2023.”

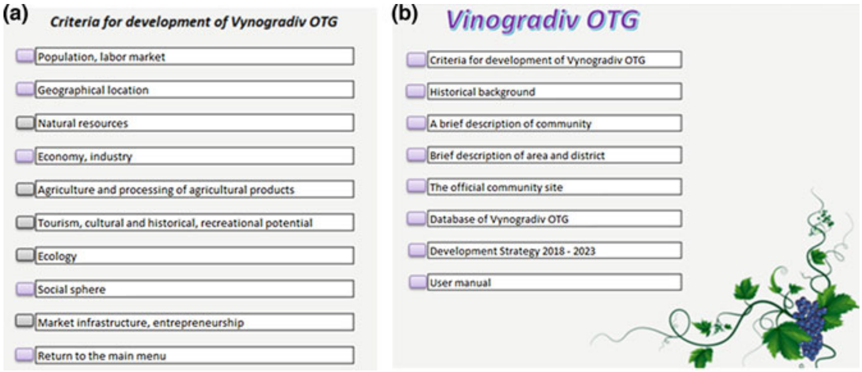


Fig. 6 The main windows of developed program for Vynogradiv OTG: (**a**) program selection window of criteria for development of Vynogradiv OTG, (**b**) program main menu

4 Conclusions

The analysis of indicators of socioeconomic development of a territory is a critical segment of the work, which requires accuracy, timeliness and reliability and is a rather labor-intensive process. Therefore, the use of information technology to analyze indicators will greatly improve the productivity of the analyst of the integrated territorial community.

As a result of the conducted research, the central of theoretical foundations and some methods of analysis of indicators of socioeconomic development of the territory are analyzed, practical implementations of methods of forecasting values indicators of socioeconomic development the territory are presented, using the Brown, Holt, Chow methods, and some information technology modules for automation of socioeconomic analysis are proposed. In a further study, we plan to refine and extend the system modules in the functional plan in order to identify the relationships between indicators to increase the number of visualized analysis results.

References

1. Chaikin, O., & Kirieieva, E. (2020). Branch approach to sustainable development and inclusive growth: Ukraine case. *Scientific Horizons, 6*(91), 19–25.
2. Bila, S. O., Shevchenko, O. V., Kushnir, M. O., & Zhuk, V. I. (2016). *Stimulating economic growth at the local level: An analytical report* (Vol. 54). NISD.
3. *State policy, legislative developments, interim results of the first stage of decentralization of power in Ukraine*. http://decentralization.gov.ua/about
4. Ramazanov, S., Antoshkina, L., Babenko, V., & Akhmedov, R. (2019). Integrated model of stochastic dynamics for control of a socio-ecological-oriented innovation economy. *Periodicals of Engineering and Natural Sciences, 7*(2), 763–773. https://doi.org/10.21533/pen.v7i2.557

5. *Economic development of the region as a component of economic development of the state*. http://www.cpk.org.ua/index.php
6. Dorosh, M. S. (2015). Convergence of system parameters in the formation of project management methodologies. *Bulletin of the National Technical University "KPI", 2*(1111), 112–120.
7. Laszlo, E. (1995). *The age of bifurcation: Understanding the changing world* (Vol. 1, pp. 3–129). Gordon & Breach Science Publication.
8. Malyarets, L. M., Babenko, V. O., Nazarenko, O. V., & Ryzhikova, N. I. (2019). The modeling of multi-criteria assessment activity in enterprise management. *International Journal of Supply Chain Management, 8*(4), 997–1004. http://ojs.excelingtech.co.uk/index.php/IJSCM/article/view/3342
9. Barro, R. J., & Sala-I-Martin, H. (2010). Economic growth. BINOM. Laboratory of Knowledge, 824.
10. Babenko, V. (2019). Formalization of the model of management of the technological innovations. *CEUR Workshop Proceedings, 2393*, 595–602. Online, http://ceur-ws.org/Vol-2393/. http://ceur-ws.org/Vol-2393/paper_431.pdf

Deep Learning in Smart Farming: A Survey

Hicham Ridany, Rachid Latif, and Amine Saddik

Abstract The artificial intelligence (AI) in agriculture field is deployed in several ways. First, it helps farmers to monitor what is happening on their farms. The latter have drones equipped with cameras, machine learning (ML) and deep learning (DL) to analyze the data collected by these instruments with greater precision. Deep learning has undergone a vast revolution since the appearance of powerful computers and databases for training and testing, which have given added value to this type of processing.

In this chapter, we will discuss the general scheme of learning systems citing the difference between machine learning and deep learning, and then we will discuss the different applications based on the DL algorithms used in precision agriculture, and we will also make comparisons of the different techniques used mainly in binary or multi-class classification, and we will evaluate the possibility of using embedded systems to accomplish the tasks solved by DL.

Keywords Artificial intelligence · Machine learning · Deep learning · Precision agriculture · Classification

1 Introduction

Hard human work is not always profitable and is over, but smart farming, optimized by Deep Learning with its high-precision algorithms, is a new concept that is emerging today. Aiming to increase the quantity and quality of products, this advanced movement allows sustainable productivity growth for all those working in the agricultural sector [1].

The practice, called precision agriculture or smart farming, uses historical and real-time data and deep learning algorithms to make specific decisions for small

H. Ridany (✉) · R. Latif · A. Saddik
Laboratory of Systems Engineering and Information Technology LISTI, National School of Applied Sciences, Ibn Zohr University, Agadir, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_16

areas of the application, rather than performing the same task for a very large area of traditional technology. For example, instead of applying large amounts of pesticides over a large area, you can apply smaller amounts to specific trees, shrubs or even leaves, reducing costs and waste while avoiding the damage that a chemical can cause when used in large quantities [2].

Artificial intelligence(AI) is already a reality in agriculture, according to numerous research has been done by researchers and agricultural specialists have shown the AI utilities in this vital field. In particular, the increasing use of vision technologies by computer for agricultural applications and the growing demand for continuous monitoring and analysis of crop health contribute to the growth of the market for AI solutions based on computer vision [3].

Developments in AI have followed different trends in the last few years, systems experts correspond to an approach based on the ability to reproduce logical reasoning. These challenges provide answers to questions or problems in a very specific field, where the machine is used to implement logical rules described by human experts and implemented in knowledge bases. This type of approach is still quite similar to traditional where all the information to be provided to the system must be identified in advance. It can quickly be limited to complex cases where multiple parameters must be taken into account [4–7].

Deep learning has undergone a revolution in the last 10 years due to the emergence of powerful computers, especially multi-core CPUs, GPUs, and also the availability of training databases.

Among the applications in crop production, Deep Learning can be used to solve problems of [7–10]:

- Detection Plant and recognition.
- Detection and Recognition of diseases [8–10].
- Classification Land occupation and crops [11].
- Yield estimation.
- Fruit counting.
- Weather prediction.
- Prediction of soil characteristics (humidity...) (Fig. 1)

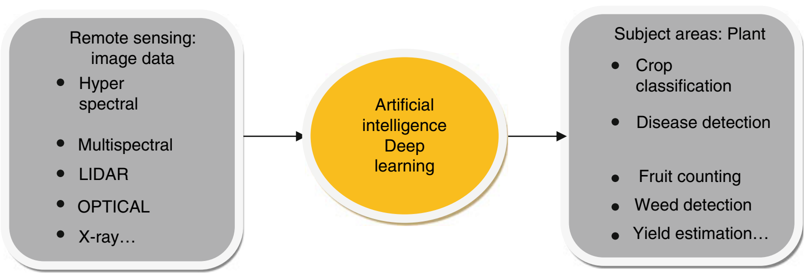


Fig. 1 General approach deep learning in precision agriculture

2 Deep Learning Block Diagram

The block diagram of Deep learning system consists five major systems. They are namely: Image acquisition, Pre-processing, Deep learning (classification), Data training, and classes [12].

Image acquisition: The images are the core of DL is taken from a culture, using an RGB or multispectral and hyperspectral camera with high resolution for more accuracy carried by an unmanned aerial vehicle (UAV) or ground robot. Each image obtained is stored in the respective size and in digital format.

Preprocessing: The resulting images are subject to various disturbances such as noise, lighting variations, poor image resolution, unwanted background, and camera movement. In preprocessing, some tools are used for enhancement using filtering techniques to remove noise and unwanted objects from the background noise.

Classification: The classification process is based on Deep learning, this system classifies each pixel of the collected image according to classification attribute, after the extraction of the features and data training to learn the weights of the neural networks used by minimizing classification errors according to the algorithms defined for this purpose.

Data training: Training data are essential to the process, they can be considered as the food used by the system, having more data implies more accuracy. The performance of Deep Learning algorithms, tends to improve exponentially when they are given more data to train and then analyze.

Classes: A class therefore represents a category of objects have properties (attributes) common to a set of objects, it is the output of a classifier, as an example of a crops class and a weeds class (Fig. 2).

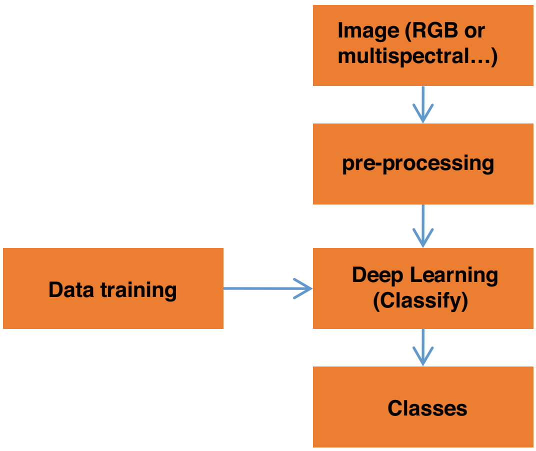


Fig. 2 Block diagram of the system

3 Deep Learning Versus Machine Learning

The latest developments in artificial intelligence may seem amazing, but if it is learning the basics that interests you, you can summarize many AI innovations in two concepts: machine learning and in-depth learning. These terms often seem to be substitutable words of moderation, hence the importance of knowing the differences. And these differences should be known from examples of machine learning and in-depth learning are everywhere. Using the example of facial recognition, automatic learning algorithms will focus on distinctive elements such as nose size, mouth shape, eye color, position of facial features, etc. However, humans are not always able to identify all the determining characteristics of a situation. And that's where the deep learning comes [2].

Indeed, since the 1980s, machine learning (ML) has been the application of statistical methods to algorithms to make them more intelligent. The challenge of the ML is to build curves close to the data and allowing an easy generalization. It is therefore based on the ability of algorithms to receive a lot of data and to learn.

Machine learning is a vast field that includes many algorithms. Among the most famous:

- Regressions (linear, polynomial...) these are curves that approach the data.
- Clustering: Still using mathematics, we will group the data into packets so that in each packet, the data is as close as possible to each other.
- Decision trees: by answering a number of questions and following the branches of the tree that carry these answers, we arrive at a result with a probability.

The idea of the birth of Deep Learning (DL) is inspired by the functioning of our brain (with neural networks) to push the analysis further and know how to extract the data itself.

The DL which is therefore a subdomain of the ML is therefore based on what are called artificial (deep) neural networks, i.e., a set of neurons: small machines that perform mathematical operations that reflect numbers according to their links to each other to the output neurons. Thanks to this architecture, the DL is able to recognize faces, synthesize texts or even drive an autonomous car.

In fact, the algorithm will adapt the links between its neurons, so that at the output we will have a good approximation of the input data [3].

Here are some examples of deep learning algorithms:

- Artificial neural networks (ANRs): these are the simplest and are often used as a complement because they extract information efficiently.
- Convolutional Neural Networks (CNNs): specializing in image processing, they apply filters to data to bring out new information.
- Recurrent Neural Networks (RNNs): which have the ability to store information and use it for a short time.

In brief deep learning, it makes it possible to do without a human expert to sort the data, since the algorithm will find its correlations on its own. To use the example

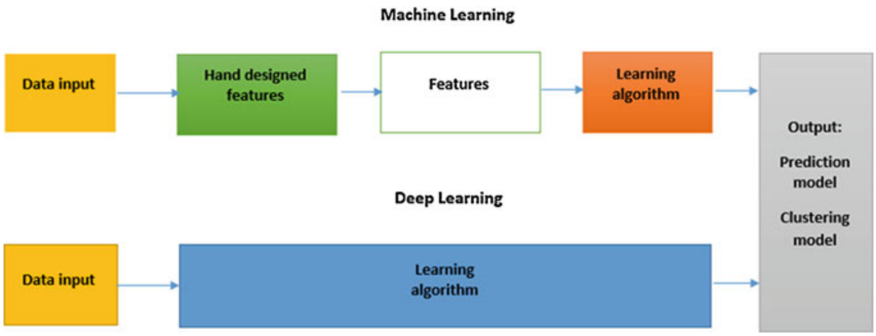


Fig. 3 Difference between deep learning and machine learning [2]

of facial recognition, the DL algorithm will determine for itself whether it should take into account the distance between the eyes (between the pixels) or whether this information is not decisive enough in relation to the others (Fig. 3).

4 Literature Survey

A great deal of research has led to the development of the classification based on artificial intelligence is particularly DL by developing the following techniques algorithm segmentation, feature extraction, representation, and classification.

In this work, we are mainly interested in one famous application it is the weed detection.

M. Dyamann et al. describes "Using a fully convolutional neural network (CNN) for detecting locations of weeds in images from cereal fields," how they can detect and separate the area affected by weeds from the crops using a fully convolutional neural network (CNN). The network was trained on more than 13,000 weed records in high-resolution RGB images from the Danish wheat and rye fields. This weed detection network has been evaluated on Nvidia Titan X, on which it is able to process a resolution 5MPx image in 0.02 s, making the method applicable for real-time field operation with an average accuracy (AP) of 76% [11] (Fig. 4).

A. Milioto et al. proposed "real-time blob-wise sugar beets vs. weeds classification for monitoring fields using convolutional neural networks" are of most important technical, the classification system is based on convolutional neural networks (CNNs). This approach is purely vision-based and uses 4 RGB + NIR channels to calculate the NDVI vegetation index and also thresholding in the post NDVI segmentation operation. In their experiments, we use the threshold segmentation of the NDVI-transformed image. The network was formed by two dataset A and B, calculating the accuracy for each dataset. The classifier architecture has been implemented on three computers Jetson TX2 SoC, Geforce GTX 940MX, and i7

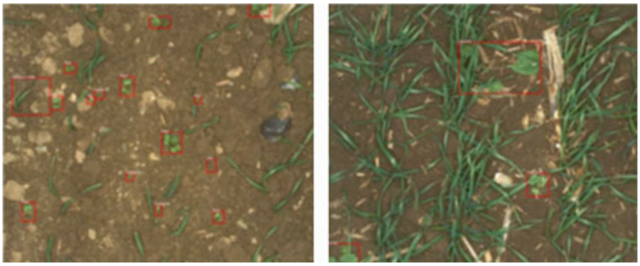


Fig. 4 Weed detection [11]

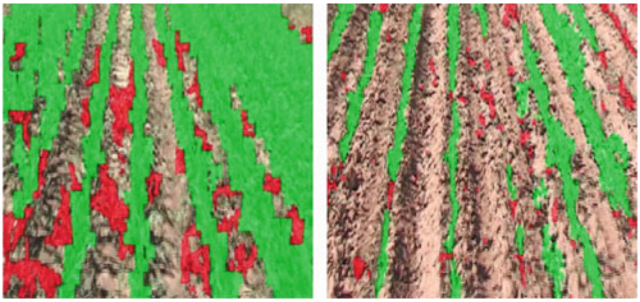


Fig. 5 Weed detection: weed (red) and crop (green) discrimination [12]

Processor in order to calculate the runtime for each machine, this implementation has an accuracy of more than 95% [12] (Fig. 5).

C. Potena et al. "Fast and Accurate Crop and Weed Identification with Summarized Train Sets for Precision Agriculture" introduces a perception system for robotics that allows an unmanned ground vehicle (UGV) equipped with a multispectral camera to automatically perform crop and weed detection and classification tasks in real time. This architecture includes two different convolutional neural networks (CNNs) applied to multispectral images (RGB + near infraredNIR) input. The first less deep CNN allows a fast and robust binary segmentation of the image, pixel by pixel, to extract the pixels that represent the projections of 3D points belonging to the green vegetation, the second a deeper CNN is then used to classify the extracted pixels between crop and weeds. The network was formed by real datasets taken from a farm robot moving in a sugar beet field, the average accuracy obtained is 91.3% [13] (Fig. 6).

C. McCool et al. "Mixtures of Lightweight Deep Convolutional Neural Networks: Applied to Agricultural Robotics" have developed a new approach for in-depth training of convolutional neural networks (CNNs) that allows us to compromise complexity and precision, so they have divided this approach into three phases to have less complexity and more precision, comparing with traditional CNN networks, this model is considered as a compression of the classical model for less parameters, for this architecture the accuracy comes to 93.3% [14] (Fig. 7 and Table 1).

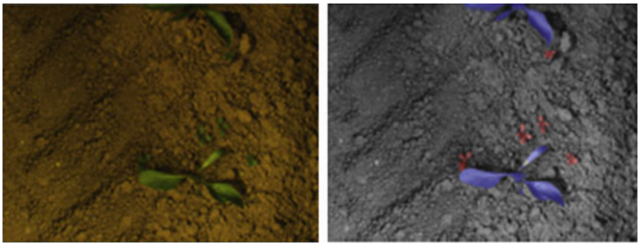


Fig. 6 Weed detection: weed (red) and crop (blue) discrimination [13]

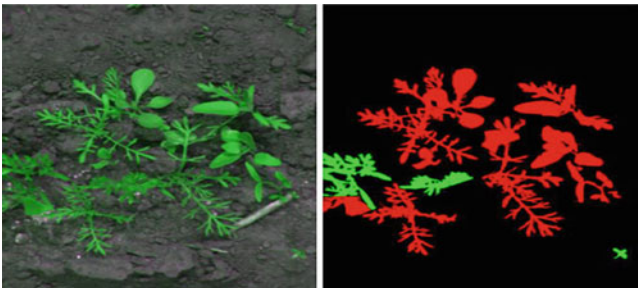


Fig. 7 Weed detection [14]

Table 1 Accuracy comparison table

Authors	Application	Accuracy (%)
M. Dyamann et al.	Detecting locations of weeds in images from cereal fields	76
A. Milioto et al.	Sugar beets vs. weeds classification	95
C. Potena et al.	Fast and Accurate Crop and Weed Identification	91.3
C. McCool et al.	Mixtures of Lightweight Deep Convolutional Neural Networks	93.3

5 Conclusion and Future Work

Most of the recent advances in agriculture made to improve agricultural productivity, plant disease reduction and preparedness, strengthening modern agriculture and agro-industry. Deep learning is generally used for image recognition or data classification, which can be summarized in four phases: data collection and data analysis, data pre-processing, neural network training, model testing, and final analysis of results. For the first stage, the combination of DL with other advanced technologies such as unmanned aerial vehicles (UAVs), radars and radar information systems. The Internet of Things can provide high quality data sets of images and images of other forms. This data greatly improves the applications of DL in the field of agriculture and to improve the accuracy of the resulting tools. For new algorithms and new learning methods can improve the efficiency of precision learning, especially for applications that require high accuracy. In this paper, we have given an overview on one of the applications of DL especially CNN in smart agriculture is the detection of weeds, despite the large number of weights and its complexity by comparing it with other DL algorithms such as Recurrent neural network (RNN) and Generative adversarial networks (GANs), it remains the best technique in this framework. Our next work is to implement the various deep learning algorithms CNN, GAN and RNN on embedded systems in real time, especially multi-core CPUs and GPUs, after an optimization on architectures to have a good compromise between accuracy and complexities.

References

1. Schönfeld, M. V., Heil, R., & Bittner, L. (2018). *Big data on a farm, smart farming*. Springer. https://doi.org/10.1007/978-3-319-62461-7_12
2. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436.
3. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117.
4. Haykin, S. (2008). *Neural networks and learning machines*. Pearson Schweiz AG.
5. Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70–90.
6. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Proceedings of Advances in Neural Information Processing Systems.*
7. Chen, Y., Lin, Z., Zhao, X., Wang, G., & Gu, Y. (2014). Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7*(6), 2094–2107.
8. Amara, J., Bouaziz, B., & Algergawy, A. (2017). *A deep learning-based approach for banana leaf diseases classification* (pp. 79–88). BTW Workshop.
9. Deng, L., & Yu, D. (2014). Deep learning: Methods and applications. *Foundations and Trends in Signal Processing, 7*(3–4), 197–387.
10. Christiansen, P., Nielsen, L. N., Steen, K. A., Jørgensen, R. N., & Karstoft, H. (2016). DeepAnomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field. *Sensors, 16*(11), 1904.
11. Dyrmann, M., Jørgensen, R., & Midtiby, H. (2017). Robot WeedSupport—Detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. *Advances in Animal Biosciences, 8*, 842–847. https://doi.org/10.1017/S2040470017000206
12. Milioto, A., Lottes, P., & Stachniss, C. (2017). Real-time blob-wise sugar beets vs weeds classification for monitoring fields using convolutional neural networks. In *Proceedings of the International Conference on Unmanned Aerial Vehicles in Geomatics*, Bonn, Germany.
13. Potena, C., Nardi, D., & Pretto, A. (2016). Fast and accurate crop and weed identification with summarized train sets for precision agriculture. In *International Conference on Intelligent Autonomous Systems* (pp. 105–121). Springer.
14. McCool, C., Perez, T., & Upcroft, B. (2017). Mixtures of lightweight deep convolutional neural networks: Applied to agricultural robotics. *IEEE Robotics and Automation Letters, 2*(3), 1344–1351. https://doi.org/10.1109/lra.2017.2667039

The Influence of Training Dataset Size on the Performance of EFDT: Application in the Field of Healthcare

Mariam Benllarch, Salah El Hadaj, and Meriem Benhaddi

Abstract Advancement in information and communication technologies will service in healthcare decision-making of patients. Nowadays, diverse medical and information communication technology is working on the building of effective classification/prediction techniques, especially in medical data stream analytics. Online decision tree algorithms can be very helpful in this case. The big challenge for this type of algorithms is that we build a classifier model based on the full set of data. In other words, the training dataset characteristics can influence the predictive capability of the model. One of the most important characteristics is the size; according to the results of a great number of studies, a large amount of training data plays a critical role in making the online decision trees successful. Nevertheless, a good model should give a high accuracy whatever the size of the training dataset. The objective of this chapter is to show that with a small dataset trained we can also achieve better results in an online decision tree. The classification algorithms tested in our work are VFDT (Very Fast Decision Tree, the parent of online decision tree algorithms) and EFDT (Extremely Fast Decision Tree, a recent model based on VFDT). Our results in the area of heart diseases show that with parameter optimization in the models and data balance, we can get similar performances of a large data in a smaller dataset. It will be important that future research should focus on the other impacts of data volume in streaming algorithms of decision trees.

Keywords Data analytics · Training dataset size · Online decision trees · Healthcare · Data streams · Classification · Machine learning

M. Benllarch (✉) · S. El Hadaj · M. Benhaddi
LAMAI Laboratory, Cadi Ayyad University, Faculty of Sciences and Techniques, Marrakech, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_17

1 Introduction

The deep secret of building smart technologies and advanced communication systems is the science of data and analytics. Data analytics (DA) is a process of inspecting, cleansing, transforming, and analyzing data with the aim of discovering useful information, proposing conclusions, and making decisions. Among various fields, healthcare is one where data analytics can be applied to make a big change [1, 2]. Healthcare analytics [3, 4] has the potential to cure diseases, reduce costs of treatment, predict outbreaks of epidemics, and ameliorate the quality of health in general. Moreover, the type of medical datasets which characterizes by its heterogeneity and massive quantity [5] is another cause to use the science of data analytics. In the modern world, heart diseases are considered as one of the most critical diseases that affect human life extremely badly. According to the World Health Organization, cardiac diseases are causative for taking 17.7 million lives every year, 31% of all global deaths. Hence, there is a need to develop a decision support system on the phase of analyzing data, by using data analytics techniques for predicting heart disease of a patient. Classification is a powerful solution that is commonly used for prediction. Traditionally, the data in classification algorithms is moved in batches. Batch learning can give a good analysis for a vast quantity of data, but with a long period of latency. For instance, the model is run every 20 h. While this can be an efficient way to manage big data, it does not work with data that need to be streamed. However, a new breed of classification called stream-based classification can manage data streams, which are continuous, infinite, and unstructured, for instance, heart biosignal datasets. Online decision trees [6] are one type of these algorithms that can improve the accuracy of predicting heart disease risk in real-time. The great difficulty to improve a model of online decision trees with this type of data is that we should always have a large dataset in order to build a good model because there is a relation between the size of the decision tree model and the accuracy of prediction. To put it another way, data volume means a varied model and then a better level of prediction. Despite this, when we want to make changes or develop this type of models, it is better to start with a small dataset in favor of minimizing costs and time series. So, what are the techniques that we can use to get the same results of big data accuracy in a smaller dataset? This chapter presents some solutions that attend this question, by using two models of online decision trees for heart disease prediction. A brief introduction about the methods used in this study: Very Fast Decision Tree and Extremely Fast Decision Tree are given in the following subsection, followed by the description of heart dataset used in our study in Sect. 3. Section 4 deals with our proposed techniques to improve the model by using small dataset. Section 5 deals with results and discussions. We will conclude our final remarks in Sect. 6.

2 Methods

The decision tree model is more powerful for classification problems. It is a representation of a decision procedure for determining the class of a given instance. The process of this model is divided into two steps: the training phase and the prediction phase. In the training phase, we talk about the building of a decision tree model by using a training data set of records. In the prediction phase, the decision tree is used to predict the response to the given data.

Online decision trees [7] are the type of real-time analytics, based on the decision tree approach. This section presents two models of these algorithms: Very Fast Decision Tree and Extremely Fast Decision Tree.

2.1 Very Fast Decision Tree

In a stream-based classification, the VFDT presented by Domingos and Hulten [8], is the parent of online decision trees. It is an incremental, anytime model that is able to learn from massive data streams. The VFDT is based on the Hoeffding bound. This inequality confirms that a small sample can often be enough to determine an optimal splitting attribute. Expect we use n independent observations of a random variable r with range R, where r is an attribute selection measure such as Information Gain or Gini Index. The Hoeffding bound affirms that with probability $1 - \delta$, the true mean $\overline{r}$ of r will be at least $E[r] - \epsilon$, with.

$$\epsilon = \sqrt{\frac{R^2 \ln 1/\delta}{2n}} \tag{1}$$

By this inequality, the VFDT algorithm can decide, with great probability, the number n of examples needed at a node when selecting a splitting attribute. The pseudo-code of VFDT is shown below (Fig. 1):

We can explain some particular aspects of the algorithm as follows:

- Sufficient statistics (line 4), for each sample after sorting the observations, the statistics of the dataset should be sufficient to calculate Information Gain or Gini Index.
- Nmin or grace period (line 6), dictates how many examples should be seen in a leaf before to start the calculations.
- The heuristic measure G (line 7), is used to determine split attributes at a node. It is calculated by using information gain or Gini index method.
- Pre pruning Xø (line 11), represents the null attribute that approves the merit of no-split. It is used for pruning the tree.

Very Fast Decision Tree Algorithm

1: Let HT be a tree with a single leaf (the root)
2: **for** all training examples **do**
3: Sort example into leaf l using HT
4: Update sufficient statistics in I
5: Increment nl the number of examples seen at l
6: **if** nl mod nmin = 0 and examples wen at I not all of same class **then**
7: Compute $\overline{Gl(x\iota)}$ for each attribute
8: Let Xa be attribute with highest $\overline{Gl}$,
9: Let Xb be attribute with second highest $\overline{Gl}$,
10: Compute Hoeffding bound $\epsilon = \sqrt{\frac{R^2 \ln 1/\delta}{2n}}$
11: **if** Xa ≠ Xø and ($\overline{G(xa)}$ - $\overline{G(xb)}$ > ϵ or ϵ< τ) **then**
12: Replace I with an internal node that splits on Xa
13: **for** all branches of the split **do**
14: Add a new leaf with initialized sufficient statistics
15: **end for**
16: **end if**
17: **end if**
18: **end for**

Fig. 1 VFDT algorithm

- Tie breaking τ (line 11), is used to indicate in between two attributes that have a very close heuristic measure. If the difference is so small, then $\epsilon < \tau$.

2.2 Extremely Fast Decision Tree

Extremely Fast Decision Tree [9] is a new algorithm based on VFDT. The idea of this online classifier is to construct an incremental model with a high quality of prediction at any given point in the instance stream, by adding a new procedure in the VFDT algorithm, which verifies the validity of split attribute at a node.

To put it another way, suppose that we use the information gain as heuristic measure, with the function of Hoeffding bound, a VFDT algorithm can guarantee that $G(xa) > G(xb)$ by the condition: $\overline{G(xa)} - \overline{G(xb)} > \epsilon$ and then indicate that *xa* is the best attribute to split a node.

The problem is that with time the model can increment internal node statistics. So, it is increasingly likely that some other split will turn out to be superior as the total number of attributes increases. There is no recourse to alter the tree in such a scenario.

The EFDT algorithm resolved this problem with the check validity function which is presented in Fig. 2.

Extremely Fast Decision Tree Algorithm

Let HT be a tree with a single leaf (the root)
for all training examples **do**
    Sort example into leaf l using HT
    Update sufficient statistics in I
    Increment nl the number of examples seen at l
    if node = l **then**
        TryToSplit (l)
    else
        ReEvaluateBestSplit(node)
    end if
end for

Fig. 2 EFDT algorithm

TryToTest function

if nl mod nmin = 0 and examples wen at I not all of same class **then**
    Compute $\overline{Gl(x\iota)}$ for each attribute
    Let Xa be attribute with highest $\overline{Gl}$,
    Let Xb be attribute with second highest $\overline{Gl}$,
    Compute Hoeffding bound $\epsilon = \sqrt{\frac{R^2 \ln 1/\delta}{2n}}$
    if Xa ≠ Xø and ($\overline{G(xa)}$ - $\overline{G(xb)}$ > ϵ or ϵ< τ) **then**
        Replace I with an internal node that splits on Xa
        for all branches of the split **do**
            Add a new leaf with initialized sufficient statistics
        end for
    end if
end if

Fig. 3 The function: TryToTest of EFDT algorithm

The start is as VFDT algorithm, after incrementing the number of examples seen on a leaf, we check if the leaf is still a leaf, that means before the calculations and the choice of the split attribute. If it is true, we call the TryToTest function. Else if the leaf is a node (after the split and the determination of split attribute), we call the ReEvaluateBestSplit function. The two functions are presented below (Figs. 3 and 4):

The first function TryToTest is as the last pseudo-code of VFDT; we do the same thing to attempt the split of a leaf node.

If EFDT determines at a node that there is another attribute better than the current split attribute, it kills the node and replaces it with a leaf. Then, it uses the new attribute selected to divide the leaf and transform it into a node.

ReEvaluateBestSplit function

Compute $\overline{G(x\iota)}$ for each attribute Xi − {X∅} using the Sufficient statistics updated at the node.
Let Xa be the attribute with the highest $\overline{G(x\iota)}$
Let Xcurrent be the current split attribute
Compute Hoeffding bound $\epsilon = \sqrt{\frac{R^2 \ln 1/\delta}{2n}}$
if ($\overline{G(xa)}$ - $\overline{G(xcurrent)}$ > ϵ) **then**
  if Xa = X∅ **then**
    Replace internal node int with a leaf (kills subtree)
  else if Xa ≠ Xcurrent **then**
    Replace int with an internal node that splits on Xa
    for all branches of the split **do**
      Add a new leaf with initialized sufficient statistics
    end for
  end if
end if

Fig. 4 The function ReEvaluateBestSplit of EFDT algorithm

Table 1 Attributes of heart dataset

Parameter	Description
Age	Age in years
Sex	1 = male, 0 = female
cp	Pain type
trestbps	Blood pressure (in mmHg on admission to the hospital)
chol	Cholesterol in mg/dL
fbs	(Fasting blood sugar >120 mg/dL) (1 = true; 0 = false)
restecg	Electrocardiographic results
thalach	Heart rate achieved
exang	Induced angina (1 = yes; 0 = no)
oldpeak	Depression induced by exercise relative to rest
slope	The slope of the peak exercise ST segment
ca	Number of major vessels (0–3) colored by flourosopy
thal	3 = normal; 6 = fixed defect; 7 = reversable defect
target	1 (abnormal) or 0 (normal)

3 The Data

The dataset used in this study contains information concerning heart disease diagnosis. The dataset was collected from the Cleveland Clinic Foundation, and it is available at the UCI machine learning Repository. It contains 600 records and 14 parameters, such as age, sex, and cholesterol, with some domain values associated with them, considered to predict the probability of heart disease as shown in Table 1.

We note that the dataset contains two classes to predict the heart situation of patients: class 1 (abnormal situation of the heart) and class 0 (normal situation of the heart).

4 Proposed Techniques

Our application of online decision tree algorithms shows that there is a relationship between accuracy, decision tree size, and time. A decision model with large size gives always high accuracy to predict instances.

In order to approve this relation, we implemented the VFDT algorithm with big data of the same dataset of heart disease by using the MOA framework [10]. The VFDT is noted as Hoeffding tree algorithm in the MOA framework (Fig. 5).

We remark that with more data subsets to train the model, the size of the tree grows, and then the accuracy increases. Our results indicate also that if we use a smaller data set, the level of accuracy is not good because the size of the tree is so small. In this section, we present two techniques to improve the results of VFDT and EFDT in a smaller dataset. The techniques can be applicated in two steps: Step 1, optimization of the parameter nmin; Step 2, balancing of classes in the dataset.

4.1 Optimization of the Parameter nmin

As we demonstrate, with a big dataset, we can build a large tree size and then a high prediction of instances. So, the first question that we can request when we work with

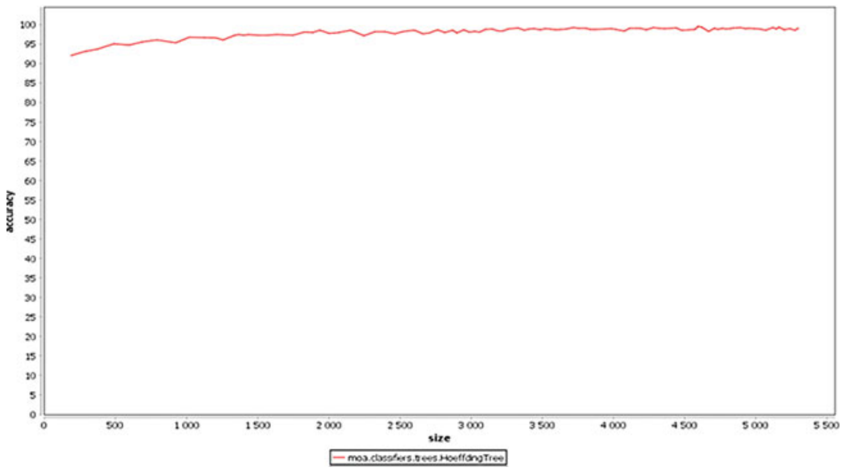


Fig. 5 The evolution of VFDT accuracy by the size of tree

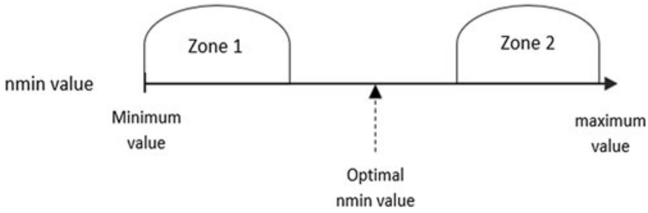


Fig. 6 Proposed technique to choose the optimal value of nmin

Table 2 Test of nmin value in VFDT algorithm with a small dataset

nmin value	10	20	50	100	150	200
VFDT accuracy	60%	70.66%	78%	78%	73%	70%

smaller data is: How we can maximize the size of the decision tree with a limited dataset?

To solve this problem, we must choose the optimal value of the parameter nmin which controls the tree growth speed. Since it presents the number of instances, a leaf should observe between split attempts.

In general, the default value of nmin used to applicate a VFDT algorithm is nmin = 200. However, the selection of an optimal value for nmin depends on the dataset nature.

In the following, we propose a strategy to update the size of the tree with the dataset size in order to choose this optimal value (Fig. 6):

- Zone 1: Even if the number of nodes is high (important tree size), the information updated at nodes are not diverse. That means, the accuracy is not good, and the time of the model increases.
- Zone 2: Least number of nodes in the tree and then minimizing the accuracy of prediction.

To explain how we can use this strategy with small dataset size, we suppose that the number of instances in the dataset is 500. So, if we take the default value of nmin (nmin = 200), we are in zone 2 and if we choose a minimum value in a relation of this dataset such as 10 or 20, we will be in zone 1. The optimal nmin value of this case should be between {50–100} (Table 2).

4.2 Balancing of Classes in the Dataset

Another technique to improve an online decision tree model in a smaller dataset is to make a balance with classes. In our situation, we have two classes in the dataset: normal heart situation and abnormal heart situation. For example, if the percentage of normal situations in the dataset is 80%, that means, the model tree is not diverse, so the accuracy will not be good.

5 Results and Discussions

In this section, we investigate the performance of the techniques proposed in this chapter. We implemented the two algorithms (VFDT and EFDT) in python, making use of the sklearn package [11]. The first result presents the impact of the parameter nmin on the model's accuracy.

We conducted a series of studies to evaluate the effectiveness value of parameters in the two algorithms. The models were run with $\delta = 0.01$, $\tau = 0.5$.

Based on Fig. 7, we can clearly see that the default value used in nmin (nmin = 200) is not optimal on the two models. To improve the accuracy of our small dataset, we should use the value of nmin = 100.

The next result shows the effect of data balance, the models were run with $\delta = 0.01$, $\tau = 0.5$, and nmin = 100.

As shown in Table 3, the performance of the classification models is good when there is a balance between the two classes of the small dataset. In general, the results of EFDT are better than VFDT, because EFDT is extremely fast.

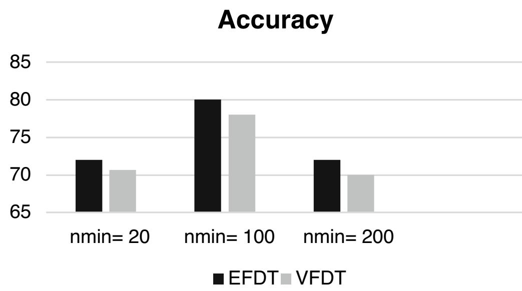


Fig. 7 The impact of the parameter nmin on the accuracy of models

Table 3 Comparison of the predictive capability of the models with the percentage of classes in heart dataset

Percentage of classes	Accuracy	Prediction time (s)
(+): 50%	Vfdt = 100%	Vfdt = 0.000994
(−): 50%	Efdt = 100%	Efdt = 0.00099
(+): 60%	Vfdt = 99.66%	Vfdt = 0.000935
(−): 40%	Efdt = 99.99%	Efdt = 0.000000
(+): 40%	Vfdt = 99.88%	Vfdt = 0.001004
(−): 60%	Efdt = 100%	Efdt = 0.000000
(+): 80%	Vfdt = 92.55%	Vfdt = 0.000994
(−): 20%	Efdt = 93%	Efdt = 0.00099
(+): 90%	Vfdt = 74%	Vfdt = 0.000002
(−): 10%	Efdt = 80%	Efdt = 0.000000

(+): the class of normal heart situation. (−): the class of abnormal heart situation

6 Conclusion

Real-time classification with online decision trees for the early prediction of abnormal heart situations needs a large training dataset to involve the accuracy of prediction. Whatever, managing an algorithm in order to minimize costs and time series requires a small smart dataset. Our solution for this problem can be resumed in two points: adapting the parameter nmin of online decision trees with the size of the dataset and balancing of classes in the dataset. With the application of our proposed techniques, the EFDT can attend 99.99% in the accuracy of prediction, with nmin = 100, 60% of normal heart instances, and 40% of abnormal heart instances. These results suggest that among the data analytics algorithms, EFDT model has the potential to significantly improve the stream classification methods for use in healthcare, even if the dataset used to train the model is small. The effect of the size in the training dataset on the prediction accuracy should overcome when thinking a new generation of online decision trees in the future.

References

1. Summary, E. (2010). The value of analytics in healthcare.
2. Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. *International Journal of Medical Informatics, 77*, 81–97. https://doi.org/10.1016/j.ijmedinf.2006.11.006
3. Mehta, N., & Pandit, A. (2018). Concurrence of big data analytics and healthcare: A systematic review. *International Journal of Medical Informatics, 114*, 57–65. https://doi.org/10.1016/j.ijmedinf.2018.03.013
4. Barati, E., Saraee, M., Mohammadi, A., Adibi, N., & Ahamadzadeh, M. R. (2011). A survey on utilization of data mining approaches for dermatological (skin) diseases prediction.
5. GraciaJacob, S., & Geetha Ramani, R. (2012). Data mining in clinical data sets: A review. *International Journal of Applied Information Systems, 4*, 15–26. https://doi.org/10.5120/ijais12-450774
6. Utgoff, P. E. (1989). Incremental induction of decision trees. *Machine Learning, 4*, 161–186. https://doi.org/10.1023/A:1022699900025
7. Gama, J., Rocha, R., & Medas, P. (2003). Accurate decision trees for mining high-speed data streams. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 523–528). https://doi.org/10.1145/956750.956813
8. Domingos, P., & Hulten, G. (2000). Mining high-speed data streams. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 71–80).
9. Manapragada, C., Webb, G. I., & Salehi, M. (2018). Extremely fast decision tree. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1953–1962). https://doi.org/10.1145/3219819.3220005
10. Bifet, A., Holmes, G., Kirkby, R., & Pfahringer, B. (2010). MOA: Massive online analysis. *Journal of Machine Learning Research, 11*, 1601–1604.
11. Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., & Mueller, A. (2015). Scikit-learn. *GetMobile: Mobile Computing and Communications, 19*, 29–33.

Security and Privacy Issues in Smart City

Abdessamad Badouch and Salah-Eddine Krit

Abstract The concept of a smart city is becoming more relevant to the policy-makers, developers, and to people. While smart cities enable the establishment of great social and economic opportunities, based on many smart technologies like the internet of things, augmented reality, and the artificial intelligence, numerous challenges are facing the rollout of a smart city in the world. Security and privacy of data are significant challenges in smart cities. Security includes illegal access to information or malicious attacks, and privacy consists of protecting all types of citizen's data, especially their activities, identities, queries, and locational data. This chapter will first introduce briefly the concept of smart city and its applications, then elaborate on the various challenges that are posed by the security and confidentiality of a person's data. Furthermore, we underline on the multiple open research methodology that is used to target the security and private areas in the smart cities and addresses the existing solutions.

Keywords Privacy · Security · Data · Smart city · IoT

1 Introduction

Nowadays, information technologies contribute in our daily lives. Cities are not an exception. The increased use of IoT gadgets has made the concept of smart city regular. It covers many domains from energy consumption to traffic management. A smart city encompasses a suite of technologies, including Internet of things, artificial intelligence, and virtual/augmented reality.

Security and data privacy are fundamental to the safety of a person. The information security in the smart city supersedes the relations of people within the city and their safety. The security of a smart city aims at protecting the relationship between the persons and the infrastructure in the city. The Cambridge Analytica

A. Badouch (✉) · S.-E. Krit
Polydisciplinary Faculty of Ouarzazate, Ibn Zohr University, Agadir, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_18

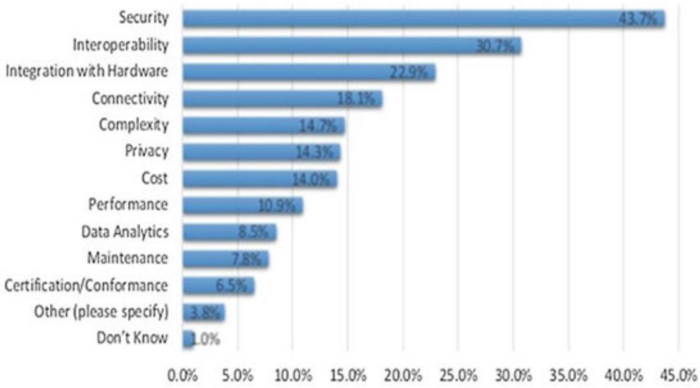


Fig. 1 Top concerns for developing smart city solutions (2016–2018)

scandal already gave us a glimpse of what is possible when a third hostile part gets access to citizen data [1]. The technologies used in smart cities generate an enormous lot of data.

Security and privacy of information is a significant concern in the design and development of smart cities.

This chapter examines these challenges as well as the illegal access to information and attacks that might disrupt operations and services, addresses a set of real issues, and discusses about papers published targeting identified challenges (Fig. 1).

2 Smart City

Smart city is an urban center that utilizes technology to gather information and supplies data which is used to improve the quality of public services [2]. Use of technology has revolutionized the development of cities. The new design model is shaping up the world. Smart cities have led to the growth of urban centers using technology, thus enhancing the quality and performance of services. It has led to minimize wastage and reduce resource consumption, thus improving the living hood of its citizens and the environment through smart technology.

A smart city is an intricate ecosystem that is characterized by the utilization of information and communications technologies (ICT) that aims at making the cities more alluring and sustainable in numerous areas and applications (Fig. 2 and Table 1).

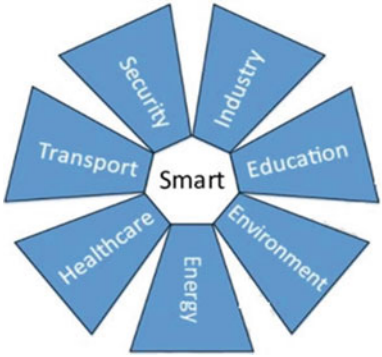


Fig. 2 Smart city applications

Table 1 Smart city domains and applications

Domain	Example of applications
Government	E-government systems, transactions, dashboards
Emergency	Centralized control rooms, surveillance, response management
Transport	Smart transport, traffic intensity, smart cards, bikeshare, real-time information, parking
Energy	Smart grids, smart lighting, energy waste
Environment	Pollution, smart temperature; water needs
Homes	Smart house, domotic system, smart appliances

3 Data in a Smart City

Development of a smart city majorly relies on data collection, analysis, and processing enormous information from sensors. There will be large sizes of sensor information reason for an immense number of IoT gadgets. While a few information can be made open and shared freely, as with the info discharged through urban dashboards, most are viewed as private and should be held safely and kept secure. Given the value of information to cybercriminals for fraud and coercion, to organizations in having insider facts, and country states for security and cyber war, they are much sought after [3–5] (Fig. 3).

Locational data and information can be a crucial security concern. Many set the GPS starting location from their homes. Access to this information details the exact site of an area. Getting this information without the consent of a legal request is unauthorized.

Data sharing and access control has been a significant challenge in securing the smart city (Fig. 4 and Table 2).

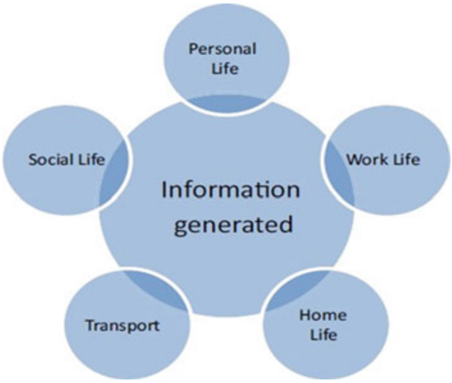


Fig. 3 Source nodes of activities and services producing data [4]

Fig. 4 The recursive cycle of data in the smart city [4]

Table 2 Data in smart city [6]

Sector	Domain	Kind of data	Example of application
Infrastructure	Transport and asset management, built environment	Monitoring and registration data	Traffic and congestion, real-time dashboards
Sustainability	Energy, water, environment, weather	Monitoring data, sensor	Air quality monitoring and pollution warnings
Health	Health quality, well-being	Health and survey data	Social and health problems
Cohesion	Education, migration	Survey data	School quality, integration
Commerce	Opportunities, marketing	Social media data, open data	Attracting new business
Experience	Events, leisure, tourism	Social media data, sensor data	Real-time social media analytics

4 Privacy

The data and information about a person is a critical issue in the smart cities. The locational data can reveal much information about a person with his knowledge.

Privacy manifests in many other forms:

Identity privacy: It protects personal and confidential information about a person.
Bodily privacy: It refers to the integrity protection of the physical person.
Territorial privacy: It deals with the protection of the personal space and property one has.

Table 3 General principles to respect in a privacy scheme

General principle	General description
Notification	Citizens are informed that data are being generated and the goal to which the data will be put
Choice	Citizens have the choice to validate or cancel as to whether and how their data will be used or disclosed
Permission	Data are only used with the permission of citizens
Security	Data are protected from loss, unauthorized access and attacks
Accessibility	Citizens can access and verify their own data
Use	Data are only used for the goal for which they are generated and citizens are informed about any change

Locational privacy: It aims at protecting against tracking of spatial behavior.

Communication privacy: Deals with the protection of any form of surveillance of personal conversations.

A smart city is vulnerable against data leakage and information interfered by the outside attackers. Since private data is gathered, transmitted, and processed, development of individual protection mechanism in a smart city is becoming a challenge (Table 3).

5 Security

There are two key security subjects concerning smart city technology: the security of the infrastructure required and the security of the information stored and shared over the various technologies and frameworks.

Shodan, a web search tool on the Internet that gathers data of around 500 million devices and services every month, has discovered that many devices and systems in the smart city have no or minimal security features. This puts a higher risk to the person since the information can be used maliciously [7].

5.1 A Secure IoT Architecture

Many global cities are presently on their way to full network connectivity (Internet of Things). With an increase in the number of sensors, IoT is gathering and managing private data, and turning it into a goldmine of data for malicious use; hence, the security concern is becoming a priority in the deployment of the IoT network.

The infrastructure development that will overcome the security concerns of the IoT is becoming a primary concern. Any IoT engineering ought to address recently referenced security issues but also deal with the IoT gadgets deployed over Software Defined Networks (SDN) and cloud infrastructure. Moreover, the detection of malicious network traffic is becoming a setback in the full development of smart city technology [8].

5.2 Authentication

In IoT domain, authentication permits incorporation of IoT devices that are deployed in various systems. Efficient management of key and cryptographic key exchange constitutes a significant challenge in the authentication of IoT devices [8].

In [9], a solid remote user authentication and key management protocol for e-governance applications in smart cities. The given protocol is based on XOR and hash operations and includes a password and smart card, user anonymity, mutual authentication, and shared session key.

5.3 Attacks

Cyber-attacks in IoT are generally categorized into one of these three types:

Availability attacks: This refers to the attacks that deny access to some services or cause system breakdown.

Confidentiality attacks: This attack causes unauthorized access and monitoring of information.

Integrity attacks: These attacks are characterized by altering the data. They change with configurations of data by inhibiting the services and infecting the network system with viruses (Table 4).

Cyberattacks in smart cities are still moderately uncommon; despite resources being available for sustaining and development of smart cities, this technology is still at their pre-commercial stages because cities lack technology-related skills and capacity.

Table 4 Categories of attacks [3]

Attack surface	Vulnerability
Local data storage	Unencrypted data, lack of data integrity checks
Device–web interface	SQL injection, cross-site scripting, non-strong passwords
Device network services	Information disclosure, week encryption, DoS
Mobile application	Non-efficient trust by devices, transport encryption

6 Security and Privacy Solutions

However, despite smart cities earning a lot of credit globally and having a lot of benefits, it also has some shortcomings. Data privacy, security, and hacking pose a significant threat in the design development of smart cities. Security is becoming a substantial setback in the development of state-of-the-art cities. Many security paradigms and strategies are being developed [10]. Transfer of data and information from one IoT device to another has become a significant setback in the deployment of smart cities. There has been the development of intrusion detection system which detects the routing attacks and then forwards the attacks. The system provides a link and a better route request rate with low downtime delay as compared to the previous system [11].

RFID tags are generally utilized in smart city infrastructure. This RFID has enabled detection of many security issues. Security solutions that have been geared toward discovery and countering this threat have been developed. RFID algorithm helps to secure multiple tags, and an RFID reader deals with low computational power [12].

Various security schemes have been proposed for the wireless sensor networks which can be applied to IoT devices. The Simple Anonymity Scheme (SAS) and the Cryptographic Anonymity Scheme (CAS) help in establishing and maintaining a secure link in multiple wireless sensor networks by utilizing dynamic designation [13]. Information security and data security within the system can be enhanced using encryption techniques or by aggregating the datasets.

Smart city devices utilize diverse programming and applications to perform their activities. The bugs in the systems act as a backdoor to potential cyber attackers who compromise the functionality of smart city systems. Developers should be trained and take into consideration security concepts at each stage of the system development cycle of a software [14].

Different testing phases should be implemented to enable detection of any security loophole and potential vulnerabilities that might exist. Various mechanisms are being applied to change the Software-Defined Network in order to improve the security of the system by managing the exploitation loopholes of the attackers. In [8], a mechanism is suggested to change attack surface in SDN-based IoT networks to raise attacker's efforts for a successful exploitation.

Sustainability of smart cities has been a significant concern to many developers. There should be educating and engaging the communities, in the overall decision-making, thus making them to feel inclusive and encouraging others to use it. A lot of money and resources are being invested in a security and protection of data. State-of-the-art infrastructures should be put in place and also user education to local business about the global market. The use of artificial intelligence system like cameras can help to combat security. Smart cities have enabled real-time monitoring of environmental condition enabling control of global warming.

7 Conclusion

A smart city is a complex technological system that has the ability and power to make our lives better and easier through technology. The methods used are vulnerable to many security issues. Therefore, proper feasibility should be done before the implementation of such a megaproject. Developers should also consider its effects on the people rather than just focusing on what impact it will have on the people. The Internet of things, blockchain, big data, and artificial intelligence will play a significant role in the design, development, and sustainability of secure exponential smart technologies. This chapter presents the smart city concept and privacy and security issues, and has addressed the existing solutions.

References

1. https://fr.wikipedia.org/wiki/Cambridge_Analytica
2. Noor, B. (2016). *Determining the Internet of Things (IOT) challenges on smart cities: A systematic literature review. Faculty of Computing*. Universiti Teknologi Malaysia.
3. Kitchin, R. (2016). *Getting smarter about smart cities: Improving data privacy and data security*. Data Protection Unit, Department of the Taoiseach.
4. Elmaghraby, A. S., & Losavio, M. M. (2014). Cyber security challenges in smart cities: Safety, security and privacy. *Journal of Advance Research, 5*(4), 491–497.
5. Khalifa, E. (2019). Smart cities: Opportunities, challenges, and security threats. *Journal of Strategic Innovation and Sustainability, 14*(3). https://doi.org/10.33423/jsis.v14i3.2108
6. van Zoonen, L. (2016). Privacy concerns in smart cities. *Government Information Quarterly, 33*(3), 472–480.
7. https://money.cnn.com/2013/04/08/technology/security/shodan/index.html
8. Conti, M., Dehghantanhab, A., Frankec, K., & Watsond, S. (2018). Internet of Things security and forensics: Challenges and opportunities. *Future Generation Computer Systems, 78*(2), 544–546.
9. Sharma, G., & Kalra, S. (2017). A secure remote user authentication scheme for smart cities e-governance applications. *Journal of Reliable Intelligent Environments, 3*, 177–188.
10. Raza, S., Wallgren, L., & Voigt, T. (2013). Svelte: Real-time intrusion, detection in the internet of things. *Ad Hoc Networks, 11*(8), 2661–2674.
11. Popescul, D., & Radu, L. (2018). Data security in smart cities: Challenges and solutions. *Informatica Economica, 20*(1/2016), 29–38. https://doi.org/10.12948/issn14531305/20.1.2016.03
12. Dong, Y., Chim, T. W., Li, V. O., Yiu, S.-M., & Hui, C. (2009). ARMR: Anonymous routing protocol with multiple routes for communications in mobile ad hoc networks. *Ad Hoc Networks, 7*(8), 1536–1550.
13. Evangelos, T., Georgios, M., & Ioannis, C. (2013). *Developing an IoT Smart City framework*. IEEE.
14. Butt, T., & Afzaal, M. (2017). The security and privacy in smart cities: Issues and current solutions.

Migration Strategies and Refactoring Methodology When Moving a Legacy ERP System to Cloud Platform

Majda El Mariouli and Jalal Laassiri

Abstract Migration to the cloud is something most enterprises are doing right now, and Enterprise Resource Planning (ERP) applications form the bulk of these projects. Essentially, the migration of critical business applications is happening because on-premise solutions are becoming very occur to move to the next generation of ERP. Also, cloud computing brings major improvements to the ERP; these improvements include mobility and its benefits, data analysis, and collaboration and sharing in government and industry. Furthermore, factors such as time, training, and extensive re-engineering activities make the migration process time consuming and error-prone. Although many research discuss Cloud ERP system adoption and advantages and disadvantages of switching to a Cloud ERP system, only a few specifically address the resolution of these constraints. This chapter presents the process of migrating applications to a different type of cloud. In addition, on the basis of a review of the literature and related emerging methodologies, we give several methodologies used to migrate from an on-premise ERP system to Cloud ERP. And based on refactoring strategy, we explain why companies need to do refactoring before moving to the cloud, we also present ERP refactoring strategies and migration feasibility. Finally, we propose one generic methodology for migrating legacy ERP systems to the cloud.

Keywords Enterprise Resource Planning (ERP) · ERP Cloud · ERP migration · Refactoring legacy systems · Refactoring applications

1 Introduction

Modernizing legacy systems is one of the most challenging problems we often face when engineering information systems. With new technologies emerging and

M. El Mariouli (✉) · J. Laassiri
Informatics Systems and Optimization Laboratory, Faculty of Science, Ibn Tofail University, Kenitra, Morocco
e-mail: majda.elmariouli@uit.ac.ma; laassiri@uit.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_19

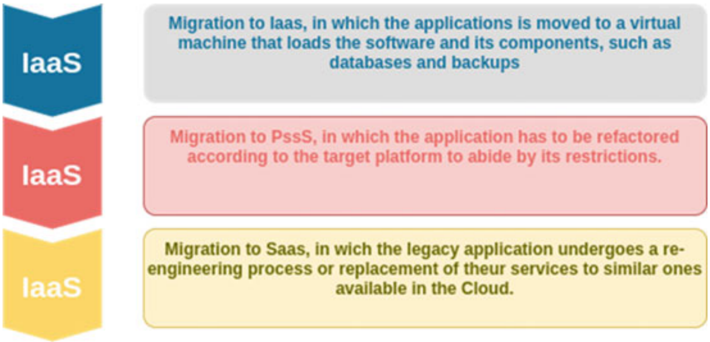


Fig. 1 Migration types of legacy application to the cloud

application domains evolving, legacy systems need to be migrated into new systems at some point, to support enhanced functionality and re-engineered business models [1]. Moving a meaningful portion of organization's existing IT assets, services or resources to the cloud is considered a "migration." The decision to migrate to the cloud can be driven by several factors, including data center lease expiration, required hardware upgrades, software license renewals, location requirements to meet regulatory compliance, global market expansion, increased developer productivity, or the need for a standard architecture [2]. A cloud migration process is a set of migration activities carried out to support an end-to-end cloud migration; it defines a comprehensive perspective, capturing business and technical concerns. There are several common components found in each successful migration, and there is no one-size-fits-all solution to deciding on the best approach. Three types of migration, considering the cloud service models, are shown in Fig. 1 [3].

Lee et al. [4] used cloud technology to move the existing on-site ERP system into a cloud environment using SaaS and proposed an effective elastic model view controller (MVC) e-ERP architecture and implement the application layers for cloud-based ERP and information systems.

Gaining an understanding of cloud migration processes from on-premise architectures is our aim here. Studying various migration approaches and choosing the appropriate approach are key to successful execution of a migration project. The chapter is organized into five sections with the introduction being Sect. 1. Section 2 provides the theoretical background of the subjects related to this work and covers many strategies and solution to migrate applications to the cloud. Section 3 explains in detail the process of refactoring methodology. Section 4 is reserved to explain our generic refactoring methodology to move legacy ERP system to the cloud. Then Sect. 5 provides the conclusion.

2 Background

Legacy system migration represents a complex task and can be a lengthy process. The effort involved depends on a number of factors, including the amount of code that needs to be converted, separation of the data access layer from the user interface layer, and availability of migration tools depending on the source technology.

A common understanding of cloud migration processes was assumed by AWS as shown in Fig. 2.

C. Pahl, H. Xiong, R. Walshe have summarized in their work [2] the process architecture by the four different migration solution provider (SaaS, PaaS, IaaS) with their essential activities shown in Fig. 3.

Regarding classification of migration studies, the authors of [5] classify studies in a reference model called Cloud Reference Migration Model (Cloud-RMM), dividing them into four processes (Fig. 4).

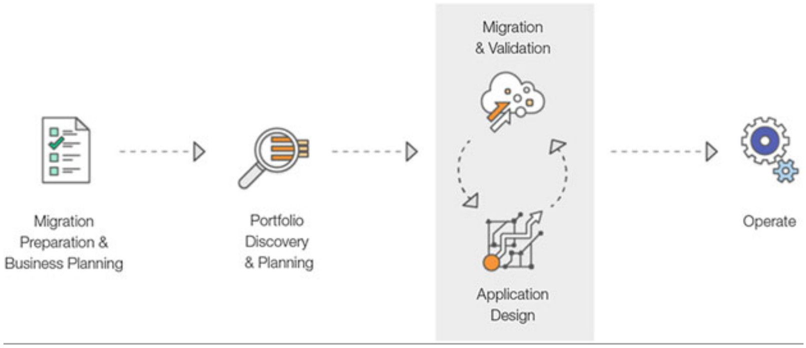


Fig. 2 AWS five-phase migration process diagram

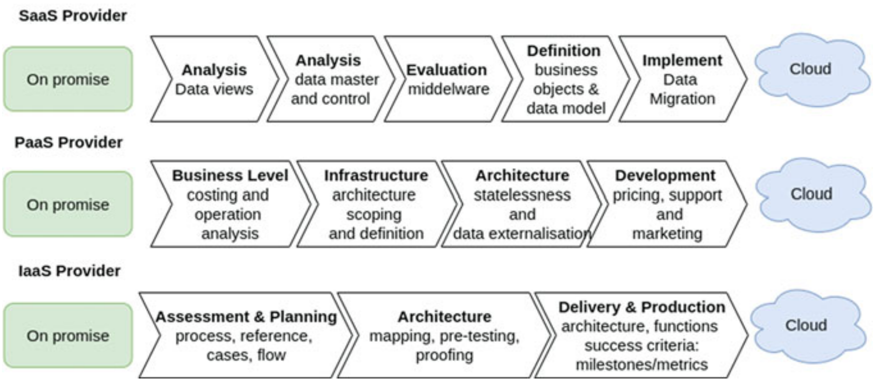


Fig. 3 Migration processes framework: use cases and selected activities

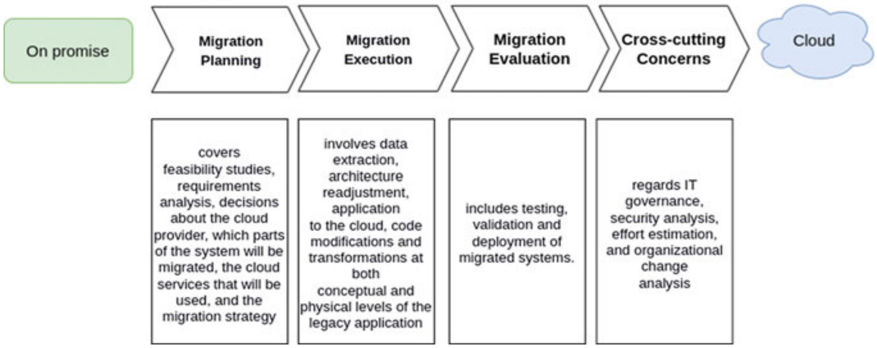


Fig. 4 Processes of Cloud Reference Migration Model

According to many sources (V. Andrikopoulos et al. in [6], AWS . . .), the most common approaches to migrate legacy applications to the cloud are as follows:

Rewriting, Replacement, Converting, Migration/Modernization, Re-platforming, Refactoring, Web service (SOA), Hybrid migration approach.

In this work, we are focusing on refactoring methodology to move legacy ERP system to the cloud.

3 Refactoring

Refactoring is a process of modifying internal design of the system while preserving its external behavior [6, 7]. The goal of refactoring is to improve a certain quality while preserving others. In another definition [8], refactoring is a feature that supports code transformation, which allows you to modify or add new features to already working code. New features, which can replace old features, are also not affected. Refactoring only affects the common set of business requirements, so that it will also decrease the number of possible objects. Refactoring only concerns existing business processes, new business requirements certainly are not [3].

There have been numerous examples where refactoring is applied in different cases and scenarios with migrating legacy applications to cloud platforms [3].

Refactoring can be a simple combination of two or more simple refactorings (move, add, rename, remove, pull-up, substitute, and complex). It is not always the same, two different categories can be defined: (1) code refactoring and (2) architecture refactoring.

3.1 Code Refactoring

Code refactoring aims modifying the code without changing its behavior with the main scope of increasing the overall software quality. One refactoring is usually a small task, but several refactorings applied to the code can considerably increase its quality. Several books and works have been written about code smells and refactoring patterns. Bad code smells can be a duplicated code, large class, long method, lazy class, feature envy, long parameter list, shotgun surgery, and so on [9].

There are many automatic tools that can detect code bad smells such as FindBugs, PMD, and CheckStyle. Also tools capable of automatically refactoring, such as the NetBeans IDE. In this context, István Orosz et al. [8] introduced a new method of encapsulating and identifying the software parts, which can be later reused in a cloud SaaS environment. The result is a Model-Driven Architecture. This MDA development approach is based on models, serving as the foundation for the design, development, and lifecycle of operation. It separates the core business logic from the technical implementation by identifying the parts of code to refactor. Thus, Isong Bassey et al. [9] provide an analysis of existing Refactoring Opportunity Identification (ROI) approaches in the context of quality metrics. They also analyzed 16 existing empirical studies to provide insight into software refactoring approach mode of operations, refactoring activities, programming languages mostly refactored, and the impact of software metrics on code quality. Marcos Borges and his collaborators [3] have proposed a new semi-automatic approach, called Cloud Restriction Solver (CRS), for migrating applications to a PaaS environment that avoids restrictions in the cloud through user-defined refactoring. This approach relies on two open and extensible tools. The first one, called CRSAnalyzer, identifies the code elements that violate the restrictions of the chosen PaaS platform, while the second one, CRSRefactor, modifies these elements by equivalent services compatible with the cloud.

Code refactoring process can be performed as follows: (1) Identifying the system's source code to identify which segments of the code need to be refactored, (2) Determining which suitable refactorings are to be applied which is done by examining the code bad smells identified. (3) Assuring that the refactorings preserve the external behavior of the software. In this case, for a given input value, the corresponding output values are expected to be the same before and after the refactoring has been conducted. (4) Applying the chosen refactorings on the ERP source code. (5) Evaluating the impact it has on quality characteristics or the processes.

3.2 Architecture Refactoring

Architecture refactoring aims changing the software architecture, reorganizing the existing code into new logical layers with the main scope of increasing the overall

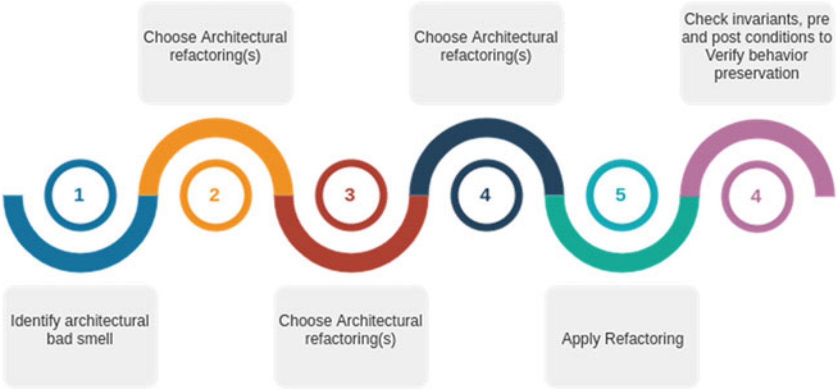


Fig. 5 Generic refactoring steps by Kesavulu et al. [13]

software quality and surpassing architecture/infrastructure limitations that critically mine systematic qualities such as scalability, extensibility, evolvability, testability, and robustness of the actual solution system. A major challenge of this kind of refactoring is trying to reuse as much as possible existing code, to avoid writing a new software.

Nikola Petković et al. [10] present an innovative refactoring strategy named RefStratERP for large-scale ERP systems. This strategy aims at improving quality of architecture and extending the lifetime of the ERP systems. Also Kesavulu et al. [7] present a generic architectural refactoring approach that facilitates the process of transforming on-premise applications to use cloud-based services. Figure 5 presents the generic refactoring steps proposed by Kesavulu et al. [13].

Automating some or all of the refactoring processes is becoming more and more popular. There is also a wide variety of refactoring shortcuts and tools to choose from to make refactoring less painful. Many of them can be learned by reading Martin Fowler's book [11]. Moreover, container development such as Docker and Kubernetes deployment and operations is also a common way of refactoring.

It is often better to refactor an application as part of a migration, but sometimes organizations do it retroactively. This can happen when performance does not meet expectations after a lift and a pass, and the setting does not solve the problem. A migrated application can also benefit from refactoring when invoices are raised unexpectedly due to application or database inefficiencies or when security vulnerabilities occur because the application cannot integrate with native security systems, such as identity and access management tools.

4 Refactoring ERP System Before Moving to the Cloud

4.1 Why We Need Refactoring Before Moving to the Cloud?

Today's companies need to be much more agile, and users expect more simplicity and elegance in the applications they use. But sometimes the technology fails to meet the needs of users and the team fails to deliver on the needs of the organization. Refactoring enables a company to break down a specific business process into smaller chunks and support those chunks with new-generation, cloud-native applications that enable IT to quickly deploy new functionality the company and users need.

Refactoring an application for cloud typically means updating the code to make use of the cloud platform's APIs and endpoints (such as network addresses and storage media) instead of those built for the enterprise datacenter. It also means ensuring that the application architecture is updated to make the best use of the target environment's specific capabilities in terms of virtual machine sizes and types of storage and network resources. There can be a couple of reasons for a refactoring, first of all we need to check if the current application is not fit to be run in the cloud. So we have to make significant code revisions in order to migrate it and have the application work properly. Second, if we need to add features, improve performance, or scale the application once in the cloud, and these improvements cannot be achieved in the application's existing state.

4.2 ERP Refactoring Methodologies and Migration Feasibility

According to Nikola Petković et al. [10], refactoring can be used with ERP systems, having architectural issues, with a purpose to improve the quality of system's architecture and thus prolong its life cycle. By re-coding some portion of an existing ERP system, organizations are able to take full advantage of cloud-native features and maximize operational cost efficiency in the cloud. It also opens up doors for the company to add important new functionalities such as artificial intelligent, analytics, and DevOps quickly and easily.

While adopting the cloud, some organizations will prioritize by separating and migrating some ERP modules or functions to the cloud first (such as human resources management), while the remainder continue to stay on-premise. Other organizations will move a critical mass of their business applications to the cloud, finding little sense in retaining core systems on-premise. Both these paths are viable, selecting one of them will depend on the specific context for the industry and company.

There are some important steps that need to be checked to determine the possibility as well as feasibility of application migration to the cloud. According to Q. H. Vu et al. [12], some applications have specific requirements that are above standard requirements, and hence cannot be satisfied by IaaS. Three typical examples of such applications are as follows:

- Applications contain sensitive or important data that cannot be exported to outside the organization or outside the country where the organization is situated.
- Applications require special hardware devices or special physical configurations that cannot be configured remotely.
- Applications process data streams such as sensor data, computer network traffic, banking transactions, stock information, media information, and social network information. This type of data processing typically consumes large network bandwidth and requires intense computation.

In addition, PaaS usually requires applications to meet certain demands of the platform when being deployed on it. Below are important examples.

- Checking whether the PaaS supports the programming language used to implement the application. It is because existing PaaSes only support a set of limited programming languages. For example, Google Apps Engine (GAE) simply supports two programming languages: Python and Java while Microsoft Azure supports a set of. NET programming languages.
- Checking which databases are supported by the PaaS. This checking step is important only if the application depends on a specific database to store its data. For example, if an application uses SQL Server as its database, and there are many actions done in triggers or stored procedures, it cannot be easily moved to another database, and hence looking for a PaaS supporting SQL Server is a must.
- Checking restrictions and limitations of the selected PaaS. This checking step is required to fully understand limitations of PaaS candidates before selecting the final one for application migration. For example, if the application requires a long processing time, GAE is not suitable. It is because GAE restricts the processing time to only 30 s.

4.3 Proposed Methodology for Migrating ERP System to the Cloud

In our work we are combining multiple strategies of moving applications into the cloud to apply them into a small chunks of our ERP system modules. The small chunks of the module are named patterns, where a pattern might model all or part of our module. The methodology aims to determine whether an existing module is compatible with the cloud, can run in the cloud, or can be optimized for the cloud.

The methodology involves first evaluating whether a pattern is incompatible with the cloud, can run in the cloud, or is prevented to run in the cloud. Second, assessing whether it is practical to optimize the pattern and move it to the cloud.

The first step in the process is assessing the existing instance, the subsequent steps involve assessing the pattern's functions and whether they can be decomposed and perhaps refactored to more fully optimize the pattern for use in the cloud.

There are several possibilities for how pattern is deployed into the cloud:

- The pattern lands after an initial assessment, with no structural changes made. The pattern instance may have operational changes made to allow it to run in the cloud, including packaging it as a virtual machine, possibly basing it on an appliance, changing binding to be dynamic, and utilizing the self-service deployment nature of the cloud.
- The pattern can be optimized so that it can more fully utilize the benefits of the cloud. Optimization often involves decomposing a pattern at a functional level and recomposing it in a new way that more fully exploits the nature of the cloud computing.

  There are many levels at which a pattern can be optimized for the cloud.

  – It can be adapted to use the massive parallelism of the cloud.
  – It can very closely match its resource consumption characteristics to its workload.
  – It can be refactored to automatically deploy and scale itself, and to completely redeploy itself when changes are made to the application.
  – It can be substituted where appropriate.

- An anti-pattern, illustrating the fact that there is some roadblock that prevents a pattern movement to the cloud, specifically one that should be respected. The possibilities include business risk, architectural or technical roadblocks, regulatory or compliance issues.

Below, our methodology applied to an ERP module pattern illustrated in Fig. 6. In the example noted as 1, a pattern is determined to run in the cloud with only some packaging changes, and it can be refactored further to optimize its use of the cloud. In the example noted as 2, a pattern that is initially assessed as incompatible with the cloud can later be refactored so that it can run in the cloud. In the example noted as 3, a pattern that is assessed as incompatible with the cloud is prevented from running in the cloud and noted with an anti-pattern. An example of this situation is government or industry regulation that places restrictions on the use of data. If a cloud provider cannot provide a service-level agreement that satisfies the regulatory requirements, then that data cannot be stored by the provider.

This chapter presents an effective methodology for migrating an enterprise's ISs to cloud ERP. In addition, this method could assist decision-makers and help ERP–cloud integrators and developers understand how such projects should be prepared and realized. ERP–cloud integrators and IT developers and practitioners should explore and investigate current and future web recommendations using the detailed solutions of a developmental diary [4].

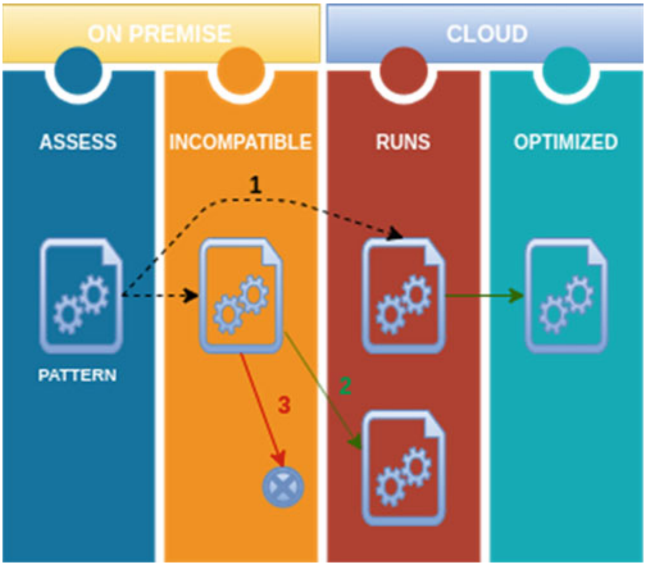


Fig. 6 The more than one pattern can be moved to the right, the more that it can exploit the benefit of the cloud computing

5 Conclusion

Most companies invest in modern ERP systems to get access to the digital world and analytics. Staying updated would give them the flexibility and insights required for growth and success. But, for companies which build a customized or personalized solution, they are facing some obstacles to modernize their system. They want to upgrade the ERP software system or gradually move parts of their business to the cloud. Regardless of the refactoring they opt for, it is essential for them to go through the features thoroughly. In the future, the ERP market will need transition to assistive and conversational user experiences (including chatbots, automation, and human–machine interaction). ERPs will be capable of processing, analyzing, and acting on vast volumes of usage data in real time, using in-memory computing technologies. By learning about user preferences, they will be able to adapt business processes iteratively—increasing user engagement and satisfaction. It is impossible to know what technology will be used after 10 years from now, so how flexible is the solution in the long term?

For that purpose, we are ensuring that companies must think about “refactoring” when moving to the cloud. Our work gives a number of considerations to make when deciding whether and how to move an existing ERP system to the cloud. In addition, the generic methodology presented in this chapter will help developers, IT managers, and cloud architects make informed decision and follow the strategy that will help them make the fullest use of their ERP system in the cloud with the least risk.

References

1. Weng, F., & Hung, M.-C. (2014). Competition and challenge on adopting cloud ERP. *International Journal of Innovation, Management and Technology, 5*(4), 309–314.
2. Pahl, C., Xiong, H., & Walshe, R. (2013). A comparison of on-premise to cloud migration approaches. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), 8135 LNCS* (pp. 212–226).
3. Borges, M., Barros, E., & Maia, P. (2018). Cloud Restriction Solver: A refactoring-based approach to migrate applications to the cloud. *Information and Software Technology, 95*, 346–365.
4. Lee, H., & Wang, N. (2019). Computer Standards & Interfaces Cloud-based enterprise resource planning with elastic model—View—Controller architecture for Internet realization. *Computer Standards & Interfaces, 64*, 11–23.
5. Jamshidi, P., Ahmad, A., & Pahl, C. (2013). Cloud migration research: A systematic review. *IEEE Transactions on Cloud Computing, 1*(2), 142–157.
6. Andrikopoulos, V., Binz, T., Leymann, F., & Strauch, S. (2012). How to adapt applications for the Cloud environment. *Computing, 95*(6), 493–535.
7. Kesavulu, M., Bezbradica, M., & Helfert, M. (2017). Generic refactoring methodology for cloud migration position. In *The 7th International Conference on Cloud Computing and Services Science (CLOSER 2017)* (pp. 664–667).
8. Orosz, I., & Orosz, T. (2017). Code reusability in cloud based ERP solutions. In *INES 2017, 21st International Conference on Intelligent Engineering Systems,* October 20–23, 2017.
9. Bassey, I., Dladlu, N., & Bassey, E. (2016). Object-oriented code metric-based refactoring opportunities identification approaches: Analysis. In *2016 4th International Conference on Applied Computing and Information Technology/3rd International Conference on Computational Science/Intelligence and Applied Informatics/1st International Conference on Big Data, Cloud Computing, Data Science & Engineering.*
10. Petković, N., & Håkansson, A. (2017). *RefStratERP—A refactoring strategy for ERP systems*. Degree Project in Information and Communication Technology. KTH Royal Institute of Technology.
11. Fowler, M., et al. (2018). *Refactoring: Improving the design of existing code*. Addison-Wesley Professional.
12. Vu, Q. H., & Asal, R. (2012). Legacy application migration to the cloud: Practicability and methodology. In *2012 IEEE Eighth World Congress on Services*.
13. Kesavulu, M., Helfert, M., & Bezbradica, M. (2016). *Towards Refactoring in Cloud-Centric Internet of Things for Smart Cities*. Dublin.

Markov Decision Processes with Discounted Reward: Suboptimal Actions

Abdellatif Semmouri and Mostafa Jourhmane

Abstract A large number of practical problems from different fields can be seen as Markov decision problems (MDPs) that they give a natural framework for modeling sequential decision making under uncertainty. These problems can, in principle, be solved via tools of dynamic programming. For this goal, it is necessary to develop efficient computational procedures for finding an optimal policy and the corresponding supreme value of the objective function. Next the Bellman optimality equation is the basic entity in MDP theory.

When the discounted reward criterion is considered to solve such problems, we are interested to determine lower and upper bounds of the optimal return in order to reduce the computational complexity if it is possible. Thus, we propose a novel test to identify suboptimal actions mainly in the infinite planning horizon case and at discrete times points. Next, we compare our previous test with those provided by MacQueen and Porteus to highlight the degree of its novelty and to show its strong point.

Our action elimination procedure can be applied in a wide range of subject areas including financial management, economics, bioinformatics, social networking, queuing, inventory control, and artificial intelligence.

Keywords Markov decision processes · Discounted gain · Suboptimality · Optimization

1 Introduction

In action elimination procedures, different results are proposed and developed in the same background for giving bounds on the performance functional. This allows reducing the computational complexity by reducing the size of action spaces at

A. Semmouri (✉) · M. Jourhmane
Faculty of Sciences and Techniques, Sultan Moulay Slimane University, BeniMellal, Laboratory TIAD, Beni-Mellal, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_20

each iteration if it is possible and accelerating the used iterative methods to find an optimal policy and the corresponding value vector. At this regard, many authors were interested to focus their efforts in this literature. The most of these scientific researchers have inspired theirs works from MacQueen [1, 2] (in 1966; 1967) and Porteus [3] (in 1971) results. By determining lower and upper bounds on the optimal value utility, they proposed various tests to eliminate suboptimal actions with are not a part of optimal decisions.

To keep up this challenge, the main goal of this manuscript is to establish a new test to identify nonoptimal actions by providing bounds of the performance system. Moreover, we compare the presented test with MacQueen and Porteus tests.

The rest of this chapter is organized as follows. In Sect. 2, we introduce the control model with finite state and action spaces and the discounted criterion to maximize over infinite planning horizon. Next, we give an overview on the main schemes of Value Iteration Algorithm. Section 3 presents a new MDP and describes our novel test applying to MDPs for identifying nonoptimal decisions. This requires giving a numerical example in order to illustrate the comparison between our test and MacQueen and Porteus tests. Finally, the conclusion is given in Sect. 4.

2 Preliminaries

2.1 Markov Decision Problem Model Control

MDPs are modeled by Markov decision processes as a stochastic dynamic programming which have been used specifically to model and solve dynamic decision-making situations.

In this subsection, we describe the underlying stationary model and its properties using the terminology of Semmouri and Jourhmane [4]. We consider a Markov decision problem that is formally modeled by the tuple

$$\left(S, A = \underset{s\in S}{\cup} A(s), Q = \left\{p\left(s'/s, a\right)\right\}, r = \{r\left(s, a\right)\}, \gamma\right)$$

where

1. S is a finite state space in which the process evolution takes place.
2. A is the union of all available finite action sets which control the state dynamics.
3. $p(./.,.)$ denotes the state transition probability functions.
4. $r(.,.)$ gives the reward function defined on state transitions.
5. γ is the discount factor.

Let π^t be a decision rule at epoch t. The immediate expected reward column vector is defined by

$$r\left(\pi^t\right) := \left[r\left(s, \pi^t\right)\right]_{s\in S}$$

where, for each $s\in S$,

$$r\left(s, \pi^{\mathrm{t}}\right) := \sum_{a\in S} r\left(s, a\right)\pi^t_{s,a}$$

2.2 Total Discounted Expected Reward over Infinite Horizon

In this model, we consider a Markov decision problem of maximizing the performance of the system in the discounted return criterion over the infinite planning horizon. The decision-maker aims to determine the optimal balance between immediate reward and future one.

This subsection presents the discounted return criterion over infinite horizon to solve MDP as described in [5–10].

Definition 1 For arbitrary policy $\pi\in\Pi$ and starting state $s\in S$, the expected total γ-discounted reward, given discount factor $\gamma\in[0,1)$, is denoted by $\mathcal{V}_\gamma\left(s, \pi\right)$ and defined by

$$\mathcal{V}_\gamma\left(s, \pi\right) := \sum_{t=1}^{\infty}\gamma^{t-1}\mathbb{E}^s_\pi\left[r\left(X_t, A_t\right)\right] \tag{1}$$

where $r(.,.)$ are the reward functions of current state and action chosen.

Since $\gamma\in[0,1)$, the swapping of sum and expectation signs in (1) is ensured by the general theory of power series. Hence, we can write

$$\mathcal{V}_\gamma\left(s, \pi\right) := \mathbb{E}^s_\pi\left[\sum_{t=1}^{\infty}\gamma^{t-1}r\left(X_t, A_t\right)\right]$$

as a total discounted expected reward.

Definition 2 The γ-optimal value function is given by

$$\mathcal{V}^*_\gamma(s) := \sup_{\pi\in\Pi}\mathcal{V}_\gamma\left(s, \pi\right), \quad s \in S \tag{2}$$

and a policy π^* is γ-optimal if $\mathcal{V}^*_\gamma(s) := \mathcal{V}_\gamma$ (s, π^*), for all $s\in S$.

The existence of optimal stationary policies has been widely studied and established in the literature of the total expected discounted criterion for MDPs with finite state space and action spaces (see [5–10]). In addition, it is well known that the system performance V is the unique solution of the Bellman optimality equation

$$V_s = \max_{a \in A(s)} \left\{ r(s, a) + \gamma \sum_{s' \in S} p\left(s'/s, a\right) V_{s'} \right\}, \quad s \in S \tag{3}$$

In this context, the value iteration algorithm (VIA) and its variants are popular dynamic programming tools for solving such equation. Thus, we will give main schemes of VIA in the following subsection.

2.3 The Main Schemes of Value Iteration Algorithm

The VIA is well understood and very simple to implement. Hence, it is widely used to compute the optimal value vector utility and a corresponding optimal policy. In this subsection, we describe main schemes of value iteration which play an amazing rule in dynamic programming for solving Markov decision problems with discounted rewards.

Without loss of generality, we assume that $S = \{1, 2, \ldots, |S|\}$. All the following Schemes (5), (6), and (7) substitute updated values of components of V^{n+1} into the recursive formula (4).

- **Pre-Jacobi (PJ)**

$$V_s^{n+1} = \max_{a \in A(s)} \left\{ r(s, a) + \gamma \sum_{s'=1}^{|S|} p\left(s'/s, a\right) V_{s'}^n \right\} \tag{4}$$

- **Jacobi (J)**

$$V_s^{n+1} = \max_{a \in A(s)} \left\{ \left[r(s, a) + \gamma \sum_{s' \neq s} p\left(s'/s, a\right) V_{s'}^n \right] / \left[1 - p(s/s, a)\right] \right\} \tag{5}$$

- **Pre-Gauss-Seidel (PGS)**

$$V_s^{n+1} = \max_{a \in A(s)} \left\{ r(s, a) + \gamma \sum_{s'=1}^{s-1} p\left(s'/s, a\right) V_{s'}^{n+1} + \gamma \sum_{s'=s}^{|S|} p\left(s'/s, a\right) V_{s'}^n \right\} \tag{6}$$

- **Gauss-Seidel (GS)**

$$V_s^{n+1} = \max_{a\in A(s)} \left\{ \left[r(s,a) + \gamma \sum_{s'=1}^{s-1} p\left(s'/s,a\right) V_{s'}^{n+1} + \gamma \sum_{s'=s+1}^{|S|} p\left(s'/s,a\right) V_{s'}^{n} \right] / \left[1 - p(s/s,a)\right] \right\} \tag{7}$$

Our purpose is to improve the PJ value iteration scheme by the addition of new test to identify nonoptimal actions at each iteration.

Algorithm 1 : PJ Value Iteration Algorithm

Input: MDP
Output: The optimal value vector, An optimal policy
Initialize: $V_s^0 = 0$, $\forall s \in S$, n = 0. **Specify** $\varepsilon > 0$;
repeat
| **for** $s \in S$ **do**
| | $V_s^{n+1} = max_{a \in A(s)}\{r(s,a) + \gamma \sum_{s' \in S} p(s'/s,a) V_{s'}^{n}\}$;
| $n \leftarrow n+1$;
until: $\|V_s^{n+1} - V_s^{n+1}\|_\infty < \frac{1-\gamma}{2\gamma}\varepsilon$;
for $s \in S$ **do**
| $\pi^* \in argmax_{a \in A(s)}\{r(s,a) + \gamma \sum_{s' \in S} p(s'/s,a) V_{s'}^{n}\}$;
return V^n, π^*

3 Results and Discussion

3.1 New MDP

This subsection presents a transformation which allows converting the original MDP with any rewards into a new MDP with nonnegative rewards.

Let

$$M := \max_{s\in S, a\in A(s)} |r(s,a)| \tag{8}$$

and

$$\tilde{r}(s,a) := M + r(s,a), s \in S, a \in A(s) \tag{9}$$

By (8) and (9), it is obvious to show that: $\tilde{r}(s,a) \geq 0$ for all $s \in S$, $a \in A(s)$.

Let also

$$J_{\gamma}(s,\pi) := \mathbb{E}_{\pi}^{s}\left[\sum_{t=1}^{\infty}\gamma^{t-1}\tilde{r}(X_t, A_t)\right], \quad \forall s \in S, \quad \forall \pi \in \Pi \tag{10}$$

and

$$J_{\gamma}^{*}(s) := \sup_{\pi\in\Pi} J_{\gamma}(s,\pi), \quad s \in S$$

The following theorem plays an important role to link between original MDP and transformed one.

Theorem 1 All optimal policies for the original MDP are optimal for the new MDP and reciprocally. Moreover,

$$\mathcal{V}_{\gamma}^{*}(s) = J_{\gamma}^{*}(s) - \frac{M}{1-\gamma}, \quad \forall s \in S \tag{11}$$

Proof We use the same manner established in [4] to prove this theorem. □

3.2 A Novel Test for Suboptimality

Throughout this subsection, all rewards are nonnegative. Next, consider the following Bellman reward operator.

$$(\Psi V)_s = \max_{a\in A(s)}\left\{\tilde{r}(s,a) + \gamma\sum_{s'\in S} p\left(s'/s,a\right)V_{s'}\right\}, \quad s \in S$$

The following lemma provides lower and upper bounds on the optimal value vector.

Theorem 1 Let (V^n) $n \geq 0$ the sequence of vectors generated by Algorithm 1 is a monotone increasing contraction mapping in $\mathbb{R}^{|S|}$, with respect to the supremum norm, with contraction factor γ. In addition

$$V_s^n \leq J_{\gamma}^{*}(s) \leq V_s^n + 2M\frac{\gamma^n}{1-\gamma}, \quad s \in S, \quad n \geq 0 \tag{12}$$

Proof The monotonicity and contraction mapping properties follow from [5].

Hence, for all $n \geq 0$,

$$V_s^n \leq J_{\gamma}^{*}(s), \quad s \in S, \quad n \geq 0 \tag{13}$$

From the same reference we have

$$V_s^n - 2M\frac{\gamma^n}{1-\gamma} \le J_\gamma^*(s) \le V_s^n + 2M\frac{\gamma^n}{1-\gamma}, \quad s \in S, \quad n \ge 0 \tag{14}$$

Combining (13) and (14), we conclude that assertion (12) holds. □

Theorem 2 (New Test of Suboptimality) Let (V^n) $n \ge 0$ the sequence of vectors generated by Algorithm 1. If

$$r\left(s, a'\right) + \gamma \sum_{s' \in S} p\left(s'/s, a'\right) V_{s'}^n + 2M\frac{\gamma^{n+1}}{1-\gamma} \;< V_s^n \tag{15}$$

at some stage n, then any Markovian policy which uses action $a^{'}$ in state s is suboptimal.

Proof Suppose that the action $a^{'}$ is optimal in state s and at epoch n. We have

$$J_\gamma^*(s) = r\left(s, a'\right) + \gamma \sum_{s' \in S} p\left(s'/s, a'\right) J_\gamma^*(s) \tag{16}$$

From (14), (15), and (16), it follows.

$$J_\gamma^*(s) < V_s^n \tag{17}$$

Combining (13) and (17), we obtain: $J_\gamma^*(s) < J_\gamma^*(s)$. This completes the proof (Fig. 1). □

3.3 *New Iterative Algorithm*

Now, we give reminders about some famous tests that have played a crucial role in action elimination approach notably in Markov decision problems. Applying

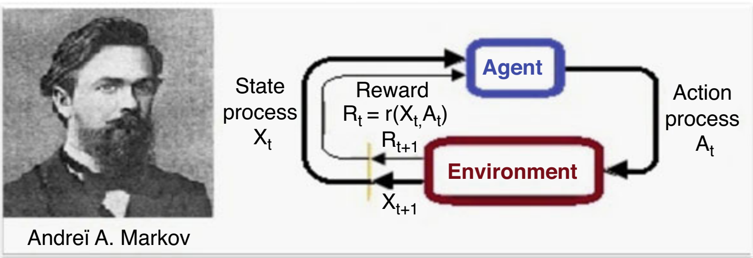


Fig. 1 Markov decision process model

Fig. 2 Computational complexity

MacQueen's bounds [1, 2] and Porteus bound [3] for the PJ algorithm leads to the following tests to eliminate action a in $A(s)$ permanently:

- **MacQueen Test:**

$$M^n(s,a) := Q^n(s,a) - V_s^n - \gamma \frac{a_n - b_n}{1-\gamma} < 0 \tag{18}$$

- **Porteus Test:**

$$P^n(s,a) := Q^n(s,a) - V_s^n - \gamma^2 \frac{a_{n-1} - b_{n-1}}{1-\gamma} < \ 0 \tag{19}$$

where

$$a_n = \min_s \left(V_s^n - V_s^{n-1}\right), \quad b_n = \max_s \left(V_s^n - V_s^{n-1}\right)$$

$$Q^n(s,a) = r(s,a) + \gamma \sum_{s' \in S} p\left(s'/s, a'\right) V_{s'}^n.$$

The proposed test in Theorem 1 may be integrated into the dynamic programming routine for solving Markov decision problems. Combining this test and the PJ value iteration algorithm, we provide the following algorithm as an amelioration of the standard VIA (Fig. 2).

Algorithm 2 : Improved VIA

Input: MDP
Output: The optimal value vector, An optimal policy
Initialize: $V_S^0 = 0$, n = 0, $\forall s \in S$. **Specify** $\varepsilon > 0$
$M = max_{s,a}|r(s,a)|$; $\tilde{r} = M + r(s,a)$, $\forall s$Î S, $\forall a$Î A(s);
repeat
| **for** $s \in S$ **do**
| | $Q^n(s,a) = r(s,a) + \gamma \Sigma_{s'}, eS\, p(s'/s,a) V_{s'}^n$;
| | $S^n(s,a) = Q^n(s,a) + 2M \frac{\gamma^{n+1}}{1-\gamma} - V_{s'}^n$;
| | **if** $S^n(s,a) < 0$ **then**
| | | $A(s) \leftarrow A(s) - \{a\}$;
| | $V_s^{n+1} = max_{aeA(s)} Q^n(s,a)$;
| $n \leftarrow n+1$
until: $\| V_s^{n+1} - V_s^{n+1} \|_\infty < \frac{1-\gamma}{2\gamma} \varepsilon$;
for $s \in S$ **do**
| $\pi^* \in argmax_{aeA(s)} Q^n(s,a)$;
| $W^n = V^n - \frac{M}{1-\gamma}$;
return W^n, π^*

Example 1 Consider an MDP with the following data:

$$S = \{1, 2\}; \quad A(1) = A(2) = \{a_1, a_2, a_3\}; \quad \gamma = 0.5.$$

Transition probabilities:

$$p\,(1/1, a_1) = 1; \quad p\,(1/1, a_2) = 0.6; \quad p\,(1/1, a_3) = 1; \quad p\,(2/1, a_2) = 0.4$$
$$p\,(1/2, a_1) = 0.5; \quad p\,(1/2, a_2) = 01; \quad p\,(2/2, a_3) = 1; \quad p\,(1/2, a_1) = 0.5$$

Rewards:

$$r\,(1, a_1) = -7; \quad r\,(1, a_2) = 9; \quad r\,(1, a_3) = -3;$$
$$r\,(2, a_1) = 12; \quad r\,(2, a_2) = -4; \quad r\,(2, a_3) = 5$$

Applying Algorithm 2, we obtain numerical approximations of the test function $S^n(.,.)$ in the following table (Table 1).

The following graphic representation gives the number of eliminated actions at each iteration by varying the discount factor inspiring from Example 1 (Fig. 3).

It is well observed that the rate of elimination of nonoptimal actions decreases when the discount factor γ increases.

Table 1 Semmouri and Jourhmane test

n	$S^n(1,a_1)$	$S^n(1,a_2)$	$S^n(1,a_3)$	$S^n(2,a_1)$	$S^n(2,a_2)$	$S^n(2,a_3)$
1	6.5	23.1	10.5	23.25	6.5	17
2	–5.05	11.58	–1.05	11.587	–5.2	5.375
3	–10.84	5.791	–6.84	5.791	–10.997	–0.418
4	–13.735	2.895	–9.735	2.895	–13.893	–3.314
5	–15.183	1.447	–11.183	1.447	–15.341	–4.762
…	…	…	…	…	…	…
23	–16.631	5.52E–06	–12.631	5.52E–06	–16.789	–6.210
24	–16.631	2.76E–06	–12.631	2.76E–06	–16.789	–6.210

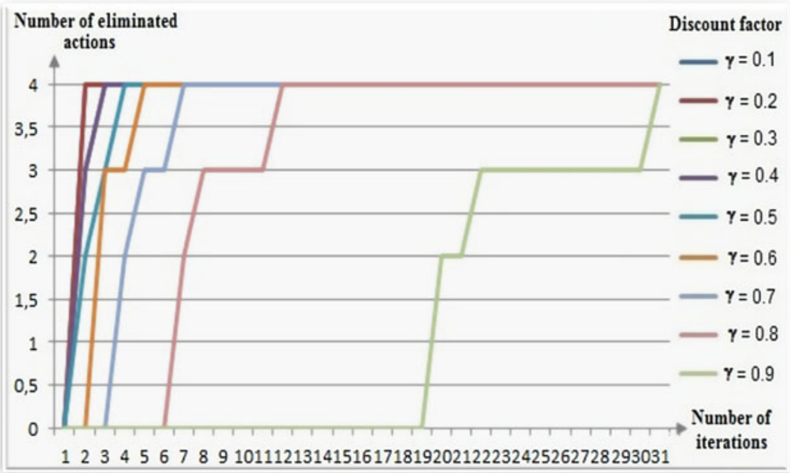


Fig. 3 Variations of the discount factor and suboptimality

Table 2 Macqueen test

n	$M^n(1,a_1)$	$M^n(1,a_2)$	$M^n(1,a_3)$	$M^n(2,a_1)$	$M^n(2,a_2)$	$M^n(2,a_3)$
1	–14.5	2.1	–10.5	2.25	–7.5	–4
2	–14.2	2.43	–10.2	2.43	–8.1	–3.77
3	–15.34	1.28	–11.34	1.28	–9.29	–4.92
4	–15.98	0.64	–11.98	0.64	–9.93	–5.56
5	–16.30	0.32	–12.30	0.32	–10.25	–5.88
…	…	…	…	…	…	…
23	–16.63	1.23E–06	–12.63	1.23E–06	–10.57	–6.21
24	–16.63	6.15E–07	–12.63	6.15E–07	–10.57	–6.21

Using MacQueen and Porteus tests, we obtain the following tables of test functions for $\gamma = 0.5$.

By applying the MacQueen test, the elimination of nonoptimal actions takes place from the first iteration in Table 2.

Table 3 Porteus test

n	$P^n(1,a_1)$	$P^n(1,a_2)$	$P^n(1,a_3)$	$P^n(2,a_1)$	$P^n(2,a_2)$	$P^n(2,a_3)$
1	//	//	//	//	//	//
2	–11.5	5.1	–7.5	5.25	–11.5	–1
3	–15.55	1.08	–11.55	1.08	–15.7	–5.12
4	–15.41	1.21	–11.41	1.21	–15.57	–4.99
5	–15.98	0.64	–11.98	0.64	–16.14	–5.56
…	…	…	…	…	…	…
23	–16.63	1.23E–06	–12.63	1.23E–06	–16.78	–6.21
24	–16.63	6.15E–07	–12.63	6.15E–07	–16.78	–6.21

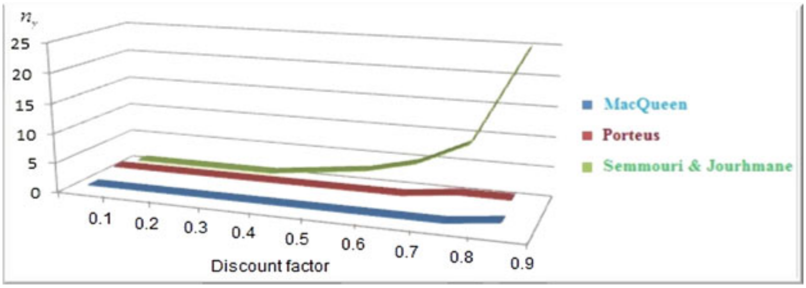


Fig. 4 Comparison of tests

The application of the Porteus test makes it possible to identify the optimal actions from the second iteration in Table 3.

Let $n_\gamma(T) := \inf \{n \in N/T^n(s,a) < 0,\ \ s \in S,\ \ a \in A(s)\}$ where $T\{M,P,S\}$. The following graphical representation summarizes the comparison between the three proposed tests when the discount factor γ varies (Fig. 4).

4 Conclusion

In this chapter, we developed a successive recursion algorithm as combination of the PJ value iteration algorithm and a new test to reduce computational efforts.

The main idea is to exclude from the computation process actions that cannot maximize the discounted gain over the infinite planning horizon. It is well known that the eliminated actions do not have any effect on the number of iteration required to reach the optimal value function.

Our object has not been to obtain the best result in action elimination literature, but to provide an idea of the merits of the various tests and improvements. This we state that our work is distinguished from those of MacQueen and Porteus in terms of computational complexity. In the previous tests, we must store the vector V^{n-1} and calculate V^n at iteration n. However, our test does not require the storage of the vector V^{n-1} in the computer memory. Therefore, the power and the efficiency of our result appear very important when the size of the state space is very large. Despite of this progress, we knowledge that MacQueen and Porteus tests are very faster for small sizes.

Acknowledgments The authors would like to thank all the members of the organizing committee of the ICDSIS 2020 Conference for their help and encouraging. Next, we are grateful to the referees for bringing to their attention, careful reading of the manuscript, valuable suggestions, and of a number of helpful remarks.

References

1. MacQueen, J. (1966). A modified dynamic programming method for Markovian decision problems. *Journal of Mathematical Analysis and Applications, 14*, 38–43.
2. MacQueen, J. (1967). A test for suboptimal actions in Markovian decision problems. *Operations Research, 15*, 559–561.
3. Porteus, E. L. (1971). Some bounds for discounted sequential decision processes. *Management Science, 18*, 7–11.
4. Semmouri, A., & Jourhmane, M. (2019). Markov decision processes with discounted cost: The action elimination procedures. In *IEEE, 2019 International Conference of Computer Science and Renewable Energies (ICCSRE)* (pp. 1–6).
5. Puterman, M. L. (1994). *Markov decision processes: Discrete stochastic dynamic programming*. Wiley.
6. Filar, J., & Vrieze, K. (1997). *Competitive Markov decision processes*. Springer.
7. Hu, Q., & Yue, K. W. (2008). *Markov decision processes with their applications*. Springer.
8. White, D. J. (1993). *Markov decision processes*. Wiley.
9. Sigaud, O., & Buffet, O. (2010). *Markov decision processes in artificial intelligence*. Wiley.
10. Bäuerle, N., & Rieder, U. (2011). *Markov decision processes with applications to finance* (1st ed.). Springer.

Modeling the Investment Attractiveness of Agricultural Enterprises in Different Regions

Nadiia Davydenko, Alina Buriak, Zoia Titenko, and Nadiia Mrachkovska

Abstract The purpose of the chapter is to develop methodological approaches to improving the assessment of investment attractiveness of agricultural enterprises, taking into account industry specificities.

During the research, scientific methods were used: modeling—to determine the impact of economic, financial, and production potential of agricultural enterprises on their investment attractiveness; economic and statistical—to evaluate the dynamics of capital investment; method of groupings—to determine the impact of capital investments on the results of enterprises; analysis and synthesis—to find out the reasons that cause changes in the volume of capital investments; tabular and graphical—to represent the results of the study; abstract-logical—in the implementation of theoretical and methodological generalizations.

The results of the study, i.e., the investment attractiveness of agricultural enterprises in different regions, were analyzed using an integral indicator that takes into account the impact of production, economic, and financial factors. There are four groups of regions, which correspond to different types of investment attractiveness.

The practical significance of the obtained results is to improve the methodology for determining the investment attractiveness of agricultural enterprises in terms of determining the integral indicator of investment attractiveness using economic and statistical methods.

Keywords Investments · Attractiveness · Efficiency · Agriculture · Integral indicator

N. Davydenko · A. Buriak · Z. Titenko (✉) · N. Mrachkovska
Finance Department, National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_21

1 Introduction

Agriculture occupies a leading position in the development of the economy of Ukraine, so it is important to increase the efficiency of agricultural enterprises. Investments play a significant role in resolving this issue. In today's economic environment, the issue of attracting internal and external investments, as well as their rational use, is of particular relevance. The current state of market transformations in Ukraine requires such conditions that would guarantee economic security to investors and ensure innovative development of the domestic economy.

One of the main factors that hamper the work of agricultural enterprises is the lack of sufficient resources. The effective functioning of an agricultural product requires considerable resources to use the problems produced in production, the raw materials used and used in production. It is worth noting that action is being taken to improve this issue.

For the efficient functioning of agricultural enterprises and the production of competitive products, significant investment resources are needed, which will allow solving the question of updating the material and technical base. The combination of factors such as investment expectations and financial risks has a decisive impact on investment decision-making. It should be understood that the influence of external and internal factors must be taken into account. An investor cannot resist the influence of external factors, but it is quite possible to minimize the investment risks directly related to the investment object.

Studies of theoretical and practical aspects of investment attractiveness are constantly in the focus of scholars. In particular, approaches to the interpretation of the concept of investment attractiveness were reflected in the works of I.O. Blanca [1], O.I. Kremenia [2], V.V. Gotra [3]. Research and systematization of factors that shape investment attractiveness is devoted to the work of O.E. Kuzmina [4], A.M. Zorova [5], O.V. Tovsteniuk [6]. Methodical approaches to assessing investment attractiveness are reflected in the works of such domestic and foreign authors as R. Bauer [7], N. Bandelj [8], S.I. Basalay [9], V. Babenko [10], R. Rajan [11], N.Y. Zakharova [12], S.V. Yukhimchuk [13]. Most scientists have researched the theoretical aspects of investment activity, but the question of increasing the level of investment attractiveness of the agrarian sector requires in-depth research, given the specifics of this field.

2 Results and Discussion

The transition to an innovative model of development of the agricultural sector of the economy depends directly on the investment—their volume, structure, and directions of application. In the context of economic globalization, investments serve as the foundation for ensuring the high competitiveness of Ukrainian producers, and they are an important prerequisite for successful integration of Ukraine into the

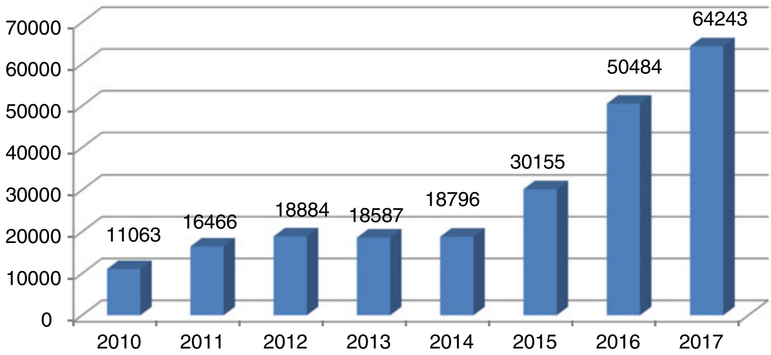


Fig. 1 Dynamics of capital investments in agriculture of Ukraine, million. (Source: Created by the author)

world economic space, achieving a high standard of living and food security of the state.

Given the intensification of market competition between enterprises for new markets for finished products, the assessment of the investment attractiveness of an individual entity is an important issue. On the one hand, it is the basis for the development of its investment policy, and on the other, the opportunity to identify deficiencies in the activity of the company, to provide for measures to eliminate them, and to improve the ability to attract investment resources. And this helps investors to decide in their investment decision.

The dynamics of capital investment in the agriculture of Ukraine indicates a rapid increase in their volume, especially over the last 2 years. It should be noted that over the past 5 years, capital investments have increased 3.5 times and amounted to UAH 64,243 million, correspondingly the share of capital investments in their total amount has also increased. In recent years, a positive trend has been observed, in particular, if in 2013 their share was 7.44% in total and in 2017, 14.33%, i.e., almost doubled (Fig. 1).

In the conditions of economic crisis in the agrarian sector of the economy of Ukraine, the level of investment activity decreases, as evidenced by the index of capital investments, which reflects the change of capital investments. It should be noted that, despite the increase in the volume of capital investments, the heterogeneity of the capital investment index is traced. In particular, there is a rapid increase in the index in 2015—160.4%, but in 2017 it decreased significantly—127.3%. For comparison, it is worth noting that the index of capital investments in food production in 2017 amounted to only 88.9% (Fig. 2).

Stimulating investment processes in agriculture is one of the priorities that requires solving a number of problems at all levels of management. The investment strategy of Ukraine's development should be systematically aimed at agricultural production, the productive functioning of which will ensure the country's food

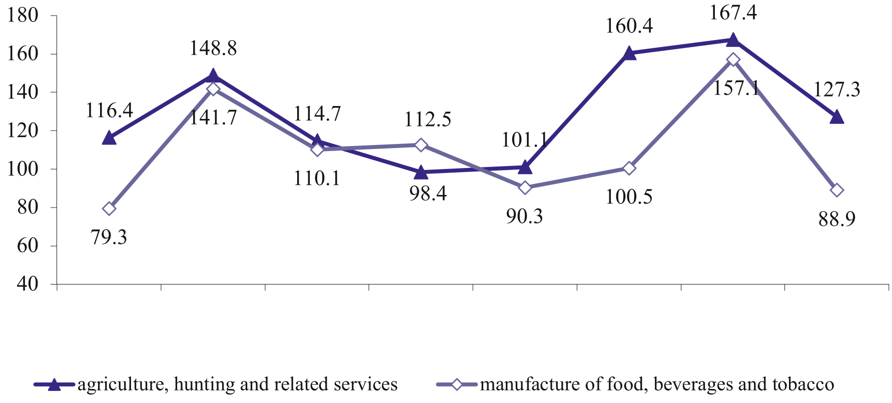


Fig. 2 Capital investment indices in Ukraine, %. (Source: Created by the author)

security, public health, efficient development of processing industries, machinery for agriculture, and other.

The sectoral structure of funds spent in agriculture is characterized by considerable unevenness. For a long time, agriculture has been a priority area for agriculture, with its share in total investment in the agricultural sector increasing from 71% in 2013 to 85% in 2017. The amount of invested funds in the development of plant growing has more than quadrupled and reached 54,237.7 million UAH.

The situation regarding investing in livestock development is somewhat different. Despite the fact that over the past 5 years, capital investments in livestock have almost doubled, but in terms of structure their share has decreased from 21.5% in 2013 to 11% in 2017. This situation is explained by the low attractiveness and long payback period of certain types of livestock products. Much of the livestock premises are physically and morally worn out, and livestock numbers and industry output sharply reduced.

The dynamics of investment in the livestock sector differ significantly from region to region. Thus, in 2017, most of the funds were spent on livestock development in Kyiv—UAH 1112.9 million, in Poltava—UAH 352.7 million, and Chernihiv oblasts—UAH 207.7 million, insignificant investment attraction indicators observed in the Trans Carpathian region for the development of animal husbandry in 2017 was allocated only 1.9 million UAH.

It is worth noting that the main source of financing capital investments in the agricultural sector for a long time remains the own funds of enterprises, due to which in 2017, 70% of the total volume was utilized.

Assessment of investment attractiveness is important as investments directly affect the end result of agricultural enterprises.

The distribution of the regions of Ukraine by the volume of capital investments per 1 ha of agricultural land confirms this fact, as with the increase of capital investments the economic efficiency of activity of the enterprises increases. With the growth of investment volumes, there is an increase in economic indicators

Table 1 Grouping of regions of Ukraine by volume of capital investment per 1 ha of agricultural land, 2015–2017

	Group of oblasts by investment volume per hectare of agricultural land, UAH			
Indicators	Up to 2000	From 2100–2600	More than 2600	Average
Number of areas	7	9	8	24
Capital investment per 1 ha of agricultural land, UAH	1624.78	2334.31	3328.41	2313.10
Gross production per 1 ha of agricultural land, UAH	4785.72	7318.62	9763.13	7008.68
Including plant growing	4208.16	5642.43	6699.25	5397.48
Animal husbandry	577.56	1676.19	3063.89	1611.20
Labor productivity, thousand UAH	258.27	281.88	305.86	282.66
Net profit per 1 ha of farmland land, UAH	3275.87	4437.11	4989.47	4298.79

Source*:* Calculated by the author

So, in the regions with more than 2600 UAH of capital investments per 1 ha of agricultural land, the gross production per 1 ha of agricultural land. The land is 9763.13 UAH, which is 40% more than the average in Ukraine. Net income per hectare of agricultural land in this group is the highest—UAH 4989.47, which is 16% more than the average (Table 1).

It should be noted that the investment attractiveness of agrarian enterprises is influenced by external and internal factors, which should be considered as the driving forces that ensure the creation of competitive advantages for a particular enterprise in the struggle for scarce investment resources [14, p. 381].

The environmental factors that have a significant impact on the investment attractiveness of enterprises, our beliefs include economic, political, scientific, technical, and environmental. However, it should be understood that, despite the significant role of external factors and the appropriateness of taking them into account when assessing the investment attractiveness of agricultural enterprises, their impact should not be overestimated. According to practice, financial stability and investment attractiveness of a particular enterprise play a significant role for investors, despite the characteristics of external factors.

Therefore, considerable attention should be paid to internal factors in the research process. These include financial, economic, production, technical, organizational and information.

With regard to the evaluation and justification of the investment attractiveness of the economic entity in the economic literature, depending on the objectives of the study, there are mainly two general areas: analysis of the financial position of the enterprise and its activities in the past period; the investor's own financial planning, or, in other words, financial design (carried out both for brand new investment projects and for existing enterprises, if the issue of investing in them is resolved) [15, p. 796].

In order to determine the investment attractiveness of an agricultural enterprise on the basis of a rating, it is necessary to use a comprehensive approach and take into account a certain set of indicators. During the development and streamlining of rating constraints, the author analyzed three groups of indicators that combined to characterize investment attractiveness, taking into account industry specificities.

It is advisable to carry out the assessment of the investment attractiveness of the regions according to the following algorithm: definition of a system of indicators for the assessment of the investment attractiveness of the regions in separate directions; calculation of certain indicators; standardization of average indicators; calculation of integral indices for each area and determination of the integral rating of the investment attractiveness of the region; construction of regions ranking by integral indexes and overall rating of investment attractiveness; the grouping of regions by the integral indices of each direction and the overall integral index of investment attractiveness; analysis of results (analysis of the actual state of investment attractiveness of regions); development of investment strategy for regional development.

The first step in assessing investment attractiveness is to identify the primary indicators, which are grouped into three groups, which characterize the following aspects of activity: resource support and production activity; financial condition and economic efficiency.

The assessment of resource provision and production activity was carried out using the following indicators: monetary valuation of land; the area of agricultural land per worker; production costs per 1 ha; gross production per 1 ha; productivity; energy capacity per 100 ha of acreage. Characterization of the financial condition was carried out using a number of financial indicators, in particular the coefficient of autonomy; ratio of own and borrowed funds; financial dependency ratio; debt capital concentration ratio; financial sustainability ratio; ratio of current assets to own funds; the coefficient of maneuverability of own capital; financial stability ratio. We have determined the economic indicators that characterize the efficiency of agricultural enterprises: profit per 1 ha; profit per 1 employee; profitability of operating activities; return on sales and return on assets.

Thus, the proposed list of indicators, in our opinion, allows to take into account such important components of the investment attractiveness of an agricultural enterprise, such as land supply, resource efficiency and business activity, quality of labor, financial stability, and efficiency of economic activity.

When generalizing the index, it is necessary to ensure the information unidirectionality of the indicators, that is, to identify stimulants and stimulators. It should be noted that the relationship between the generic indicator and the stimulus indicator is direct and that of the stimulus indicator is inverted.

Preliminary rationing is required to compare these indicators. The main task of normalization is to bring the indicators to one basis (dimensionless values), provided that the ratio between them is maintained [16]:

$$R_j = \sum_{i=1}^{n} \frac{x_{\max\ i} - x_{ij}}{x_{\max\ i} - x_{\min\ i}} + \sum_{i=1}^{n} \frac{x_{ij} - x_{\min\ i}}{x_{\max\ i} - x_{\min\ i}} \tag{1}$$

where

Rj is the sum of the ratings of a specific region for each of the indicators that characterize a particular aspect of investment attractiveness.
x_{ij} is the value of the *i*-th indicator of the *j*-th region.
$x_{\max\ i}$ the maximum value of the *i*-th indicator.
$x_{\min\ i}$ is the minimum value of the *i*-th indicator.

The first part of the formula is used to estimate stimulant performance, and the second part is used to evaluate stimulant performance.

The next stage of the study is to determine the arithmetic mean of the sum of integral indices by region for all indicators that characterize a particular group:

$$R_{cpj} = \frac{R_j}{n} \tag{2}$$

where

R_{cpj} is the arithmetic mean of the integral indices of a particular region for all indicators of a particular group.
n the number of indicators that were calculated for each group.

According to the results of the calculations, the integral rating is defined as the arithmetic mean of the sum of the ratings of a specific region for all groups by the formula:

$$I_j = \frac{\sum_{i=1}^{m} R_{cpi}}{m} \tag{3}$$

where

I_j is the arithmetic average of ratings of a specific region for all groups.
m is the number of directions for which the calculation was made.

To divide regions by rating, use the grouping method at regular intervals:

$$h = (x_{\max} - x_{\min})\,/m \tag{4}$$

where

$x_{\max}, x_{\min}$ is the highest and lowest value of the trait in the aggregate.
m is the number of groups.

Expert values of the selected indicators were established by the analysis of agricultural enterprises in different regions. In the course of the study, four groups

Fig. 3 Map of investment attractiveness of agricultural enterprises of regions of Ukraine. (Source: Created by the author)

of regions were identified, which corresponded to different types of investment attractiveness, in particular, group 1—investment attractive regions; group 4—risky.

In the distribution of investments by region of the country, there is excessive differentiation. Investments are mostly concentrated in economically developed regions where such investments can be paid off quickly. Research shows that agricultural enterprises of Kyiv, Odesa, Cherkasy, Kirovograd, and Vinnytsia regions are attractive for investment. At the same time, agricultural enterprises of Chernihiv, Ivano-Frankivsk, Zhytomyr, and Rivne regions showed a low level of their investment attractiveness (Fig. 3). This situation does not contribute to the even socioeconomic development of the country and increases the further growth of regional disparities.

The regional peculiarities of investing in the agriculture of Ukraine affect not only the equipment of agricultural enterprises by the means of logistics but also the results of their economic activity and investment attractiveness. More favorable natural and climatic conditions, higher level of investment support, and more powerful production potential contribute to the successful development of all branches of agriculture. Therefore, finding opportunities to attract investment for the development of the agricultural sector, creating favorable conditions for domestic and foreign investors are key in solving the problem of the consequences of the financial crisis, accelerating the pace of investment in agriculture [13].

To determine the real investment attractiveness of agricultural enterprises requires a comprehensive assessment taking into account the production and financial potential, as well as the potential of business activity and production efficiency. We find that this approach fully characterizes the investment attractiveness of enterprises.

At the same time, it should be noted that a necessary and important prerequisite for improving the investment attractiveness of agricultural enterprises and strengthening their financial sustainability is to resolve the land issue. Without the land market, the investment attractiveness of agricultural enterprises will be incomplete and remain low, which will not contribute to the flow of money to the investment market of the industry.

3 Conclusion

This chapter provides an integrated assessment of the investment attractiveness of agricultural enterprises in different regions of Ukraine, taking into account production and financial potential, as well as production efficiency. This approach can be quite feasible in the context of both an inflationary and a stable economy.

The proposed calculation method takes into account a number of indicators that characterize the internal environment of agricultural enterprises, providing a systematic approach to the study of investment attractiveness, which allows to identify both strengths and weaknesses, as well as to develop a strategy for their development.

Stimulating investment processes in agriculture is one of the priorities that requires solving a number of problems at all levels of management. The investment strategy of Ukraine's development should be systematically aimed at agricultural production, the productive functioning of which will ensure the country's food security, public health, efficient development of processing industries, machinery for agriculture, and other.

Conflict of Interest The authors declare that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancy have been completely observed by the authors.

References

1. Blank, I. A. (2001). *Investment management: [training course]*. K.: Elga-N, Nika-Center.
2. Kremen, O. I. (2014). Fundamentals of assessment of investment attractiveness of the region. *Scientific Bulletin of the State Higher Education Institute "Ukrainian Academy of Banking of NBU", 4*, 62–66.
3. Gotra, V. V. (2015). The current state, factors of providing and directions of activation of investment activity in agriculture. *Actual Problems of Economy, 6*, 168.
4. Kuzmin, O., Yastrubskyy, M., & Krevs, V. (2017). State of activities of higher education and prospects of investment and innovation. *Modern Issues of Economy and Law, 1–2*, C.118–C.127.

5. Zorova, A. M., & Chemerinska, K. P. (2015). Investment attractiveness of agricultural enterprises of Ukraine as a prerequisite for increasing foreign investment. *Scientific Bulletin of the Kherson State University, 12*(1), 38–41.
6. Tovsteniuk, O. V. (2013). Conceptual principles of diagnostics of investment attractiveness of enterprises. *Economy and State, 6*, 89–92.
7. Bauer, R., Koedijk, K., & Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. *Journal of Banking and Finance, 29*, 1751–1767.
8. Bandelj, N. (2002). *Creation of foreign direct investment in Central and Eastern Europe*. Princeton University.
9. Basalay, S. I., & Chorus, L. I. Accounting as a tool for increasing the investment attractiveness of agro-industrial corporations in a transition economy [Electronic resource]. http://knigi1.dissers.ru/books/library2/2782-4.php
10. Babenko, V. (2019). Formalization of the model of management of the technological innovations. In *CEUR Workshop Proceedings* (Vol. 2393, pp. 595–602). http://ceur-ws.org/Vol-2393/. URL: http://ceur-ws.org/Vol-2393/paper_431.pdf
11. Rajan, R. (2004). *Measures to attract FDI-investment promotion, incentives and policy intervention*.
12. Zakharova, N. Y. (2011). Problems of estimation of investment attractiveness of agrarian enterprises. *Finance, Banks, Investments, 3*, 76–80.
13. Yurchik, I. B., & Matsenko, A. A. (2016). Foreign investments in Ukraine: Problems and prospects. *Young Scientist, 5*(32), 204–207.
14. Tovsteniuk, O. V. (2012). Investment attractiveness as a diagnostic object. *Bulletin of the National University "Lviv Polytechnic", 727*, 381–386.
15. Skripnik, Y. V. (2016). Investment attractiveness of an agricultural enterprise. *Global and National Problems of Economy, 11*, 794–799.
16. Resolution of the CM of Ukraine of October 21. (2015). *On approval of the procedure and methodology for monitoring and evaluation of the effectiveness of implementation of the state regional policy*.

Prognostication of Financial Providing of Innovative Activities of Enterprises

Nadiia Davydenko, Anatolii Ivanko, Yuliia Nehoda, and Zoia Titenko

Abstract In order to ensure the effective development of enterprises and the economic growth of the country as a whole, it is important to introduce scientific and technological innovations or technologies. This chapter determines the relevance of innovative activity and discusses various aspects of innovation financing in Ukraine. The distribution of funds for innovative activity at Ukrainian enterprises is presented and analyzed. The sources of financing of innovative activity and their change during 2010–2017 are presented. In the course of the research, scientific methods were used: economic and statistical—to estimate the dynamics of the volume of expenditures by the directions of innovative activity in Ukraine and to analyze the sources of financial support of innovative activity; economic and mathematical modeling—to construct the forecast of financing of innovative activity; analysis and synthesis—to find out the reasons that cause changes in the financing of innovative activity; tabular and graphical—to represent the results of the study; abstract–logical—in the implementation of theoretical and methodological generalizations.

Results of the research—the dynamics of the sources of financing of innovative activity is analyzed and the forecast of volumes of financing of innovative activity till 2025 is made, taking into account the dynamics for 2010–2017 and the planned indicators of GDP growth in Ukraine.

Keywords Cost-effectiveness · Innovation · Forecasting · Financing

N. Davydenko · A. Ivanko · Y. Nehoda · Z. Titenko (✉)
Finance Department, National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine

Department of Agrarian Economics, Separated Subdivision of National University of Life and Environmental Sciences of Ukraine “Nizhyn Agrotechnical Institute”, Nizhyn, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_22

1 Introduction

In order to increase the competitiveness of domestic goods and services, at both the national and international markets, the issue of accelerating innovation processes at Ukrainian enterprises is becoming increasingly relevant. Innovation development greatly contributes to a more efficient use of enterprise production resources and attracts capital investment, which in turn contributes to the efficiency of economic entities and the economic growth of the country as a whole. In today’s context, it is impossible to resolve this issue without adequate financial support. This is due to the fact that in the current conditions, enterprises are experiencing a lack of their own financial resources, along with high interest rates on bank loans in Ukraine and the difficulty of attracting additional investments.

Increasing the innovation activity of economic entities and accelerating the production of innovative products are a condition of economic stability of each state. Innovation is a key driver for improving production efficiency, which is one of the components of the successful operation of the enterprise, so the intensification of the dynamics of economic growth of the economy of Ukraine implies innovation of entrepreneurial activity.

Theoretical and methodological aspects of innovative development of enterprises and their financial support are considered in the works of famous foreign and domestic scientists, in particular Boronos V.G. [1], Buturlakina V.A. [2], Babenko V. [3], Varnali Z.S. [4], Matelyukh N.P. [5], Michael E. [6], Tkalenko N.V. [7], and others. However, in spite of the thorough work of this problem, some aspects of financial support for innovative activity of enterprises in Ukraine require additional research.

2 Materials and Methods

For most businesses, the economic impact of innovations in the production process is directly dependent on their financial support. The ability of business entities to accumulate financial resources for the implementation of their own innovative projects directly affects their commercial success.

Most of the scientists in the analysis of financing of innovative activity in Ukraine highlight the following main problems: limited possibilities of attraction of state resources; deficit of own resources of economic entities; underdevelopment of institutional investors (banks, insurance companies, investment funds and companies, stock market, etc.); underdevelopment of innovative infrastructure; lack of an effective legal framework for attracting funds from foreign investors, including venture financing. No investment bank for long-term lending has been created yet in Ukraine, and no decision is being taken to establish a State Bank for Reconstruction and Development [8].

The conducted research shows that the development of the innovation sphere depends directly on its financial support. We believe that the urgency of the problems associated with the formation and improvement of the system of financial support for innovation activity is determined by the exceptional importance of innovation processes in the economic and social development of the country, as well as the need to create effective forms of financial relations that mediate each stage of innovation activity.

3 Results and Discussion

In today's challenges, the innovative path of development is an integral or even decisive component of the economic growth of any country in the world. The experience of the leading countries [9, p. 160] shows that only with the widespread introduction of innovative developments can it be possible to fill both Ukrainian and world markets with competitive products of domestic production. In this regard, the process of innovative substitution is urgently needed, since the development of the national innovation system depends to a large extent on the state's ability to mobilize its internal innovation potential in a timely manner.

Innovation is one of the most important factors of economic development, and the development of innovations has become the main condition for ensuring the competitiveness of the economy. In Ukraine, there is a growing understanding of the exceptional role of science and innovation in the national economy in the context of European integration benchmarks. However, despite the general recognition of the high importance of innovation processes for the national economy, the current state of their financial support system, as one of the key factors of innovation activity, indicates the inability to create the necessary conditions for realizing the existing innovation potential. The lack of adequate financial support, concentration, and efficient use of available financial resources hinders the process of activation of innovation activity and reduces the impact on the effectiveness of economic innovation [10, p. 67].

Innovative businesses are facing problems related to intellectual property rights. In most countries of the world, the state insures the risks of new innovative companies, but does not claim income from their activities, instead the state receives income through taxes paid by enterprises created on the basis of innovation. In Ukraine, however, the tax burden on innovative enterprises increases with their profitability, which is negative and complicates innovation.

An analysis of the dynamics of the main indicators of innovation activity in Ukraine shows that the number of enterprises engaged in innovation decreased by 51.2% during 2010–2017, from 1462 in 2010 to 759 enterprises in 2017; however, despite their decrease in the share of enterprises engaged in innovation activities, the total number of enterprises increased by 2.4 percentage points in 2016 compared to 2010 (Fig. 1). The most innovative enterprises in the regions are Kharkiv, Ternopil, Mykolayiv, Cherkasy, Kirovograd, Ivano-Frankivsk, Zaporizhia, Sumy, and Kyiv.

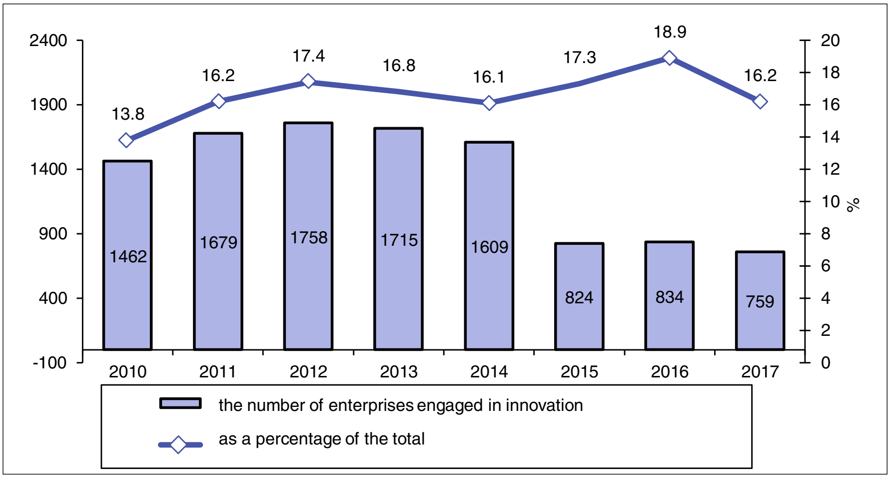


Fig. 1 Dynamics of innovation enterprises in Ukraine. (Source: Built by the author)

Innovative enterprise development involves not only product and technological innovations, which due to lack of resources have a number of problems in their implementation, but also management, marketing, and organizational innovations. It should be understood that in the conditions of today, this type of innovation can be a prerequisite for sustainable development of enterprises, as well as to accelerate the build-up of components of innovation potential that will ensure the innovative development of the enterprise.

The strengthening of innovative development of economic processes in the world practice is achieved through the commercialization of R&D, which is ensured by the stimulation of all participants of innovative activity. State funding plays a decisive role in this combination of direct and indirect innovation incentives.

Innovative enterprise development is not possible without sufficient financial support. Nowadays, the key is the problem of attracting investment resources in innovative development and their rational use, and the further economic development of enterprises, and the country as a whole depends on the solution of this problem.

Studies have shown that during the analyzed period in Ukraine, expenditures for innovation activities increased and in 2017 amounted to UAH 9117.5 million. It is worth noting that the largest share in the period covered by the purchase of machinery and equipment, respectively, their share in 2017 was 64.7%. The share of other expenditures for the years 2010–2016 in the total amount decreased substantially from 23% in 2010 to 11.3% in 2017, as a positive point is the fact that the dynamics of the growth of expenditures on domestic R&D in the total volume from 10 to 16% in 2017. Research has found that the cost of acquiring other external knowledge during the analyzed period has also decreased. Expenditure on external research in 2010–2016 remained almost unchanged (Fig. 2).

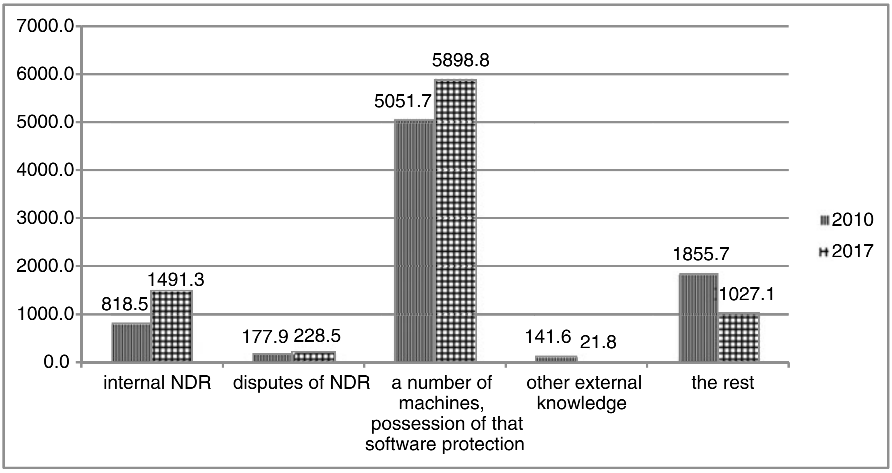


Fig. 2 Distribution of total expenditures by innovation activity in Ukraine. (Source: Built by the author)

The success of an innovation activity is largely determined by the forms of its organization and the methods of financial support. As new scientific developments and technologies become fundamental components of national security, developed countries find diverse opportunities to support and develop innovation. At the same time, a variety of methods of financing innovation activities and a range of measures for indirect support of innovation are being extended.

According to Art. 18 of the Law of Ukraine "On innovation activity," sources of financial support for innovation activity are [11]: state budget funds, local budget funds, own funds of specialized state and municipal innovative financial and credit institutions; own or borrowed funds of the subjects of innovative activity, funds (investments) of any natural and legal persons, other sources not prohibited by law.

In developed countries, financing for innovative activities comes from both public and private sources. However, in Ukraine, a completely different trend is being followed; It is worth noting that the share of own funds in the structure increased from year to year and in 2017 amounted to 84.5% (Table 1).

On the one hand, the use of own funds to finance innovation is characterized by stability, simplicity, and speed of their involvement, the ability to make flexible and prompt investment decisions, minimizing the cost of the project by the amount of interest on loans, which ensures high mobility of circulation of money and prevents bank risk. Time of their use. However, the constant lack of own funds and the high level of risk inherent in innovation activity do not always guarantee domestic enterprises high rates of development through self-financing of innovative measures.

Research shows that lending is one of the main forms of financing innovative activity in Ukraine. Compared to own funds, the share of credit resources is insignificant; however, it is the largest in the structure of borrowed funds. Thus, during 2010–2017, the share of credit remained virtually unchanged and in 2017

Table 1 Sources of financial support for innovation activity[a]

	2010		2015		2017	
	UAH million	As a percentage of the total	UAH million	As a percentage of the total	UAH million	As a percentage of the total
Total	**8045.5**	**100.0**	**13,813.7**	**100.0**	**9117.5**	**100.0**
Including at the expense of money						
Own	4775.2	59.3	13,427.0	97.2	7704.1	84.5
State budget	87.0	1.1	55.1	0.4	227.3	2.5
Local budgets	5.7	0.1	38.4	0.3	95.6	1
Extrabudgetary funds	0.9	0	1.4	0	0.3	0
Domestic investors	31.0	0.4	74.3	0.6	273.1	3
Foreign investors	2411.4	30	58.6	0.4	107.8	1.2
Credits	626.1	7.8	113.7	0.8	594.5	6.5
Other sources	108.1	1.3	45.1	0.3	114.9	1.3

Source: Built by the author

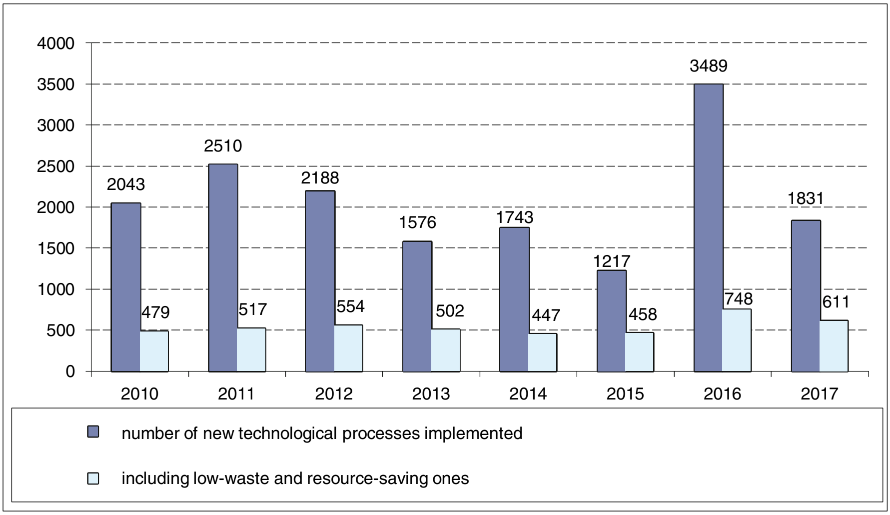


Fig. 3 Number of innovations implemented at industrial enterprises of Ukraine during 2010–2017. (*Source: Built by the author)

was 6.5%. It is worth noting also positive developments; so in 2017, the share of credit resources in the structure of sources of financing began to grow and reached the level of 2013, which indicates the activation of financial institutions and in the innovative processes of the Ukrainian economy.

However, according to the analysis, there are factors that have a negative impact on the development of enterprises on an innovative basis, in particular, a high degree of deterioration of fixed assets, dependence of financial results of activity on market conditions, large resource and energy intensity of the domestic industry, inefficient use of fuel and energy resources, weak link of domestic science with production.

Analyzing the industrial enterprises of Ukraine that have introduced innovations, it should be noted that during the analyzed period, their number decreased by almost 45%. Studies have found that in 2017, 459 enterprises were implementing innovative processes, of which 198 were introducing low-waste and resource-saving technologies, while the rest of enterprises were engaged in the introduction of innovative products.

Analyzing the introduction of innovations in Ukraine, it should be noted that in 2016 the number of new technological processes introduced increased by 3489. In 2017, their number decreased substantially and amounted to only 1831 technological processes. Resource-saving technologies in 2017 amounted to 611, which compared to 2010 increased by 27.5% (Fig. 3).

The conducted researches show that the volume of financing of innovative activity has considerable fluctuations by years and is described by the equation of the fourth parabola (Fig. 4).

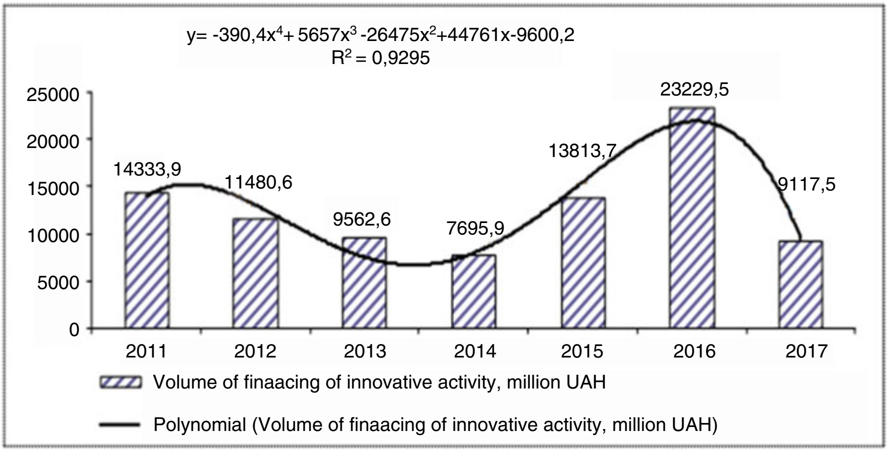


Fig. 4 Volume of financing of innovation activity. (Source: Built by the author)

In developed countries, innovation is a key factor in socioeconomic development. Recent research has found that in these countries, between 50 and 95% of GDP growth is derived from innovation, while up to 25% of the working population are engaged in innovative business. At the same time, when in Ukraine GDP growth based on innovation is less than 1% [10, p. 21].

Considering the experience of developed countries of the world, we believe that the amount of financing for innovation activities should depend on GDP. We consider the appropriate average share of financing to be determined by the moving average method. The projected value of the share of costs for innovation is 0.64%. However, the maximum value was 1.1 in 2011 and 0.97 in 2016. Therefore, the optimal forecast can be 1.035% of GDP (average of two values). Considering that the average GDP growth should be 3.3% in 2020, 3.8 in 2021, and 4.1 in 2022 [12, p. 21].

Taking into account the trends of the previous years and the planned indicators of growth of the gross domestic product, the financing of innovation activity in 2025 will average UAH 30,463 million, with the optimistic variant forecasted to reach UAH 47,772 million (Table 2).

The Strategy of Innovative Development of Ukraine for the period 2010–2020 defines in the conditions of globalization challenges that in order to ensure innovative restructuring of the economy and the impact of innovations on the economic system, it is necessary to increase the share of technologically high production up to 35–40%, the share of knowledge-intensive production to 25–30%, and the share of enterprises implementing innovations 3–3.5 times [13]. We understand that in order to achieve the above indicators, it is necessary to increase costs, so the importance of the efficiency of functioning of the financial mechanism in the formation of innovative type of economy of Ukraine is increasing. On the other hand, Ukraine's natural resources are one of the largest in the world. Having

Table 2 Forecast of financing of innovative activity, million UAH [a]

Year	GDP	Volume of financing of innovative activity	
		The middle option	Optimistic
2019	3,665,467	24,192	37,938
2020	3,786,428	24,990	39,190
2021	3,930,312	25,940	40,679
2022	4,091,455	27,004	42,347
2023	4,259,204	28,111	44,083
2024	4,433,832	29,263	45,890
2025	4,615,619	30,463	47,772

Source: Built by the author

identified the innovative way of development of the national economy, there is an urgent need for their rational use in the national economy. Given the available natural resource potential, having created the appropriate conditions for investment support for innovation, scientific progress and its commercialization can have a financial multiplier for investors.

4 Conclusion

Analysis of the current state of innovative development of the Ukrainian economy has made it possible to ascertain that there is an imbalance between the objective laws of social development and the conditions for obtaining and implementing innovations. This conclusion is supported by a number of factors: directly related to the innovation process, in particular the lack of economic conditions, the suboptimal structure of innovation and its financing; the inconsistency of some components of financial support for innovative economic development; lack of regulatory framework for financial support and the like.

Despite the presence of a number of concepts and programs for the development of science and innovation in Ukraine, as well as the periodic discussion of problems of innovation and science and technology at the parliamentary level, the adopted recommendations are largely not implemented, and financial, credit, tax, customs, and other levers of security development of innovative activities do not work. This is the main reason for slowing down the transfer of research results from scientific institutions to real sector enterprises. There are virtually no mechanisms in place in developed countries to support the transfer of research results.

Conflict of Interest The author declares that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancy have been completely observed by the authors.

References

1. Boronos, V. G., & Rybina, L. O. (2016). Problems of financial regulation of innovation activity in Ukraine. *Chernihiv Scientific Journal, 1*(7), 108–114.
2. Buturlakina, T. O., & Zavatko, V. V. (2015). Problems of functioning of venture capital as a source of financing of innovative activity in Ukraine and ways of improvement (on the example of the USA experience). *Global and National Problems of Economy, 4*, 791–794.
3. Ramazanov, S., Antoshkina, L., Babenko, V., & Akhmedov, R. (2019). Integrated model of stochastic dynamics for control of a socio-ecological-oriented innovation economy. *Periodicals of Engineering and Natural Sciences, 7*(2), 763–773. https://doi.org/10.21533/pen.v7i2.557
4. Varnaliy, Z. S., & Garmashova, O. P. (2013). *Competitiveness of the national economy: Problems and priorities of innovation provision: Monograph*. K.: Knowledge of Ukraine.
5. Matelukh, N. P., & Demyanchuk, G. V. (2014). Financing and promoting the commercialization of innovation in Ukraine: Problems and solutions. *Science and Innovation, 3*, 15–23.
6. Porter, M. E., & Stern, S. (2002). National innovation capacity (Chapter 2.2). In Porter, Sachs, Cornelius, McArthur, & Schwab (Eds.), *The global competitiveness report 2001–2002/E. Michael*. Oxford University Press.
7. Tkalenko, N. V. (2012). *Actual problems of implementation of an innovation-oriented model of national economy development: Monograph*.
8. Fominakh, V. I., Makhova, N. I., & Samokhlib, O. I. (2017). Problems of financing of innovative activity of enterprises. *Young Scientist, 10*(50), 1089–1092.
9. Marchenko, O. V. (2008). Determination of necessary and sufficient conditions for innovative development of the enterprise. *Actual Problems of Economics, 6*, 158–164.
10. Ptashchenko, L. O., & Sasko, V. V. (2014). Problems of financial support of innovative activity in Ukraine Economy and region. 6, 67–74.
11. *The Law of Ukraine "On Innovative Activity" No. 40-IV of 04.07.2002 on amendments and extra*. [Electronic resource]. http://portal.rada.gov.ua/
12. *Cabinet of Ministers of Ukraine Resolution of the Cabinet of Ministers of Ukraine On approval of the Forecast of economic and social development of Ukraine for 2020–2022*. No. 555 of May 15, 2019.
13. *Strategies for innovative development of Ukraine for 2010–2020 in the face of globalization challenges* [Electronic resource]. https://zakon.rada.gov.ua/laws/show/2632-17
14. Tarasyuk, M. V. (2017). *The current state of implementation of innovation policy of Ukraine and its financial support* [Electronic resource]. http://www.economy.in.ua/pdf/1_2017/6.pdf

Modern Methodological Approaches to Assessment of Social Responsibility of Pharmaceutical Companies

Yuliya Bratishko, Olga Posylkina, Olga Gladkova, and Maksym Bezpartochnyi

Abstract This chapter addresses topical issues of social development of pharmaceutical companies in the context of introduction of quality management systems. It proves the need of development of the method of assessment of social responsibility of pharmaceutical business that will be consistent with economic situation at any given point in time, and with due regard to the industry specifics. It further defines the essence and importance of social responsibility of pharmaceutical business. The suggested methods of assessment of social responsibility of pharmaceutical business are based on application of methods and techniques of mathematical modeling in economics and the latest information technologies. Introduction of the suggested assessment method would promote identification and assessment of problems in the sphere of social responsibility, laying down the groundwork for the creation of the efficient quality management, and developing the efficient management system in the context of socially responsible business.

Keywords Social responsibility · Social responsibility of pharmaceutical business · Assessment · Assessment results · Development · Pharmaceutical factory

1 Introduction

Social responsibility of business is one of the five basic prerequisites for the successful future business formulated in the course of a global study by IBM [1].

According to ISO 26000, "Social Responsibility Guidance Standard," social responsibility means responsibility for decisions affecting the society and envi-

Y. Bratishko · O. Posylkina · O. Gladkova
Department of Management and Economics of Enterprise, National University of Pharmacy, Kharkiv, Ukraine

M. Bezpartochnyi (✉)
Department of Economics and Marketing, National Aerospace University N. Zhukovsky "Kharkiv Aviation Institute", Kharkiv, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_23

ronment, a standard that encourages companies to act in a transparent and ethical manner contributing to sustainable development, including public health and well-being, consistent with applicable laws and international standards, integrated in and practiced by the organization in its relationships [2].

From this perspective, the meaning of “social responsibility of company” today expands far beyond the boundaries of philanthropy and charity and implies implementation of long-term social projects, investment in human capital development and environmental projects, introduction or resource and energy saving technologies. Corporate CEOs are becoming increasingly aware that socially responsible activities are not only about the additional expenditures but rather about considerable benefits resulting from the increased loyalty of customers, business partners and corporate staff, as well as enhanced investment attractiveness, establishment of long-term partnership with stakeholders, etc.

2 Literature Review

Many scholars address the problem of assessment and introduction of socially responsible management in the businesses. Over the years, the problems of SR and its assessment were studied by H. Bowen [3], A. Carroll [4, 5], K. Davis [6, 7], P.E. Druker [8], M. Friedman [9, 10], N. Haines, S. Rossi Alice, M. Schwartz [11], and many other scholars. In Ukraine, assessment and management of socially responsible business and sustainable social and economic development of enterprises were addressed in the studies of M.O. Kyzym, A.M. Kolot A.M [12], O.F. Novikova [13], N.M. Ushakova, etc. In pharmaceutical studies, certain aspects of assessment and management of sustainable social and economic development of the Ukrainian pharmaceutical companies and their social responsibility were addressed in the studies of A.A. Kotvitska [14], N.O. Tkachenko [15], B.P. Gromovyk [16], and other scholars. However, despite considerable scientific advancement, issues of social responsibility and sustainable development of pharmaceutical companies remain underexplored, especially when it comes to assessment of social responsibility with the help of modern economic and mathematical methods and information technologies, which proves the need to continue studies of this scientific problem.

Considering the specifics of pharmaceutical industry, the SRPhB shall be construed to mean the PhC activities aiming to provide quality, efficient, safe, and affordable medicines to the public; creating conditions for the development of professional competencies of employees in accordance with applicable social norms; as well as creation of conditions for efficient social benefits and protection of personnel; promoting the environmentally clean pharmaceutical production and facilitating the environmental improvements in the areas of presence of the PhC production facilities; participation in regional social programs for the development of communities, implementation of its own social projects, etc.

3 Research

In order to study contemporary tendencies in the context of social responsibility of companies and their sustainable growth, we analyzed the Reputation Institute studies involving 55,000 respondents. Top-10 companies in terms of corporate social responsibility in 2010–2018 are presented in Table 1.

It should be noted that medicine and pharmacy traditionally are socially conscious and responsible industries, and the same conclusion follows from the Industry Global CSR Index [17]. However, not a single pharmaceutical company made in to the Top-10 Global CSR RepTrak until 2018. Over the 9 years of studies, Novo Nordisk was the first pharmaceutical company to take No. 5 position in 2018.

Meanwhile, according to 2016 Global Pharma RepTrak® [17], perception of socially responsible behavior of pharmaceutical companies (PhC) has been growing steadily worldwide. For example, reputation of PhC in the UK, USA, Australia, Brazil, Germany, and Russia has grown over the last few years, reaching the index of 70 and higher. The leaders in terms of social responsibility are Novo Nordisk, Abbott Laboratories, Roche, Merck, Sanofi, Allergan, AstraZeneca, Eli Lilly, AbbVie, GSK, Novartis, Bristol-Myers Squibb, and Pfizer—pharmaceutical companies with reputation index varying from 65.9 to 68.7 (2016).

According to our studies, 63% of the Ukrainian PhCs build their governance pattern with the focus on economic effectiveness while acting in accordance with applicable laws, which is enough only for basic level of social responsibility (SR). And only 11% PhCs covered by our studies mastered the art of implementing the personnel, environmental, cultural, and social programs, aiming to strengthen their business reputation, image, and corporate culture. Curiously, the Ukrainian pharmaceutical companies demonstrate higher social responsibility internationally rather than domestically.

As far as smaller pharmaceutical companies are concerned, slightly over a third (34%) have social programs. As regards medium-sized companies, over 50% of employers (55%) have social programs. The index goes even higher (up to 62%) in companies with the staff exceeding 200 persons. Small- and medium-sized pharmaceutical companies are normally more inclined to charity, while the big companies are more focused on systemic projects that contribute to the development of social infrastructure and to the improvement of environmental situation. However, according to our studies, most of the Ukrainian pharmaceutical companies have SR programs of a rather spontaneous variety. Specifically, for 16% of Ukrainian PhC, social programs are just non-recurring, random projects. However, we also observed that the bigger and financially stronger companies tend to be more committed to development of a clear social responsibility strategy. About 60% of big PhC approach to this matter in a systemic manner, so unlike the small- and medium-sized companies (30–40%). In 47% of cases, the company itself covers the expenditures on social projects. Another 25% of companies rely on their employees' help whenever possible (the most frequently observed format of charity projects); 17% of companies implement their social projects through charitable

Table 1 Rating dynamics of the most socially responsible companies worldwide [17]

Company rating	Company/year								
	2018	2017	2016	2015	2014	2013	2012	2011	2010
1	Google	LEGO Group	Google	Google	Google	Microsoft	Microsoft	Google	Google
2	The Walt Disney Company	Microsoft	Microsoft	BMW AG	Microsoft	The Walt Disney Company	Google	Apple Inc.	Vodafone Group plc
3	LEGO Group	Google	The Walt Disney Company	The Walt Disney Company	The Walt Disney Company	Google	The Walt Disney Company	The Walt Disney Company	Microsoft
4	Natura	The Walt Disney Company	BMW AG	Microsoft	BMW AG	BMW AG	BMW AG	Microsoft	Apple Inc.
5	Novo Nordisk	BMW Group	LEGO	Daimler	Apple Inc.	Daimler	Apple Inc.	Daimler	BMW AG
6	Microsoft	Intel Corporation	Daimler	LEGO	LEGO	SONY	Daimler	SONY	SONY
7	Robert Bosch	Robert Bosch	Apple Inc.	Apple Inc.	Volkswagen	Intel Corporation	Volkswagen	LEGO	IKEA
8	Canon	Cisco Systems	Rolls-Royce Aerospace	Intel Corporation	Intel Corporation	Volkswagen	SONY	BMW AG	Volkswagen
9	Michelin	Rolls-Royce Aerospace	Rolex	Rolls-Royce Aerospace	Rolex	Apple Inc.	Colgate-Palmolive	Volkswagen	United Parcel Service, Inc.
10	IKEA	Colgate-Palmolive	Intel Corporation	Rolex	Daimler	Nestle	LEGO	Intel Corporation	Intel Corporation

organizations, public authorities, and local self-government bodies. And about 10% of companies appear to be aware that one may be useful without even making monetary donations, resorting rather to volunteering services or, for example, by providing the information support.

Therefore, our study helped uncover certain drawbacks in SR management in PhC. This calls for improvement of the existing theoretical framework for management of social responsibility of pharmaceutical businesses (SRPhB) and development of industry-specific methods for the assessment of social responsibility of companies.

4 Methodology and Result

According to the studies, practical assessment of social responsibility mostly focuses on identification of the PhC ratings, which is truly important and encourages the companies to develop their own SR systems. However, no matter how important these ratings may be, they would not help assess the SR development potential, their results are fairly subjective and limited to information that is only available to the experts whose opinions shape up the ratings. To promote better transparency, objectivity, and assessment of the SR level, and to provide tools to the PhC for the efficient control of the socially responsible development reserves, we suggest methodological approaches to the SRPhB assessment based on the economic and mathematical approaches and information technologies. What makes the suggested SRPhB assessment methodology unique is that it takes into account specifics of the PhC business and international standards regulating the performance of pharmaceutical companies. Figure 1 shows the SRPhB assessment system suggested by the authors.

Objectivity of the results of the SRPhC assessment largely depends on validity of the selected indicators (key values) on the basis of which the assessment proceeds. For the assessment of the SRPhB, we substantiated the requirements to indicators which will be used in calculations: (1) suitability; (2) correct assessment of the status of the object; (3) precision; (4) reliability; (5) completeness and entirety; (6) uniqueness; (7) simplicity yet substantivity; (8) the indicators may be quantitative; (9) the indicators shall be consistent with requirements of the assessment and provide the required correlation with indicators of production and economic activities of the PhC; and (10) comparability. Taking into account these requirements, we have formed a totality of local indicators characterizing the SRPhB by each specific component.

Local indicators of the SRPhB assessment were selected in Stage 1 by way of the expert-based method. The experts were represented by CEOs of the PhC, executive officers and employees of the HR, strategic planning and marketing services/departments, and scholars. Total experts count is 130 persons. Stage 2 involved checking the level of correlation between the local indicators of the SRPhB assessment in order to exclude highly interrelated indicators. If the pair

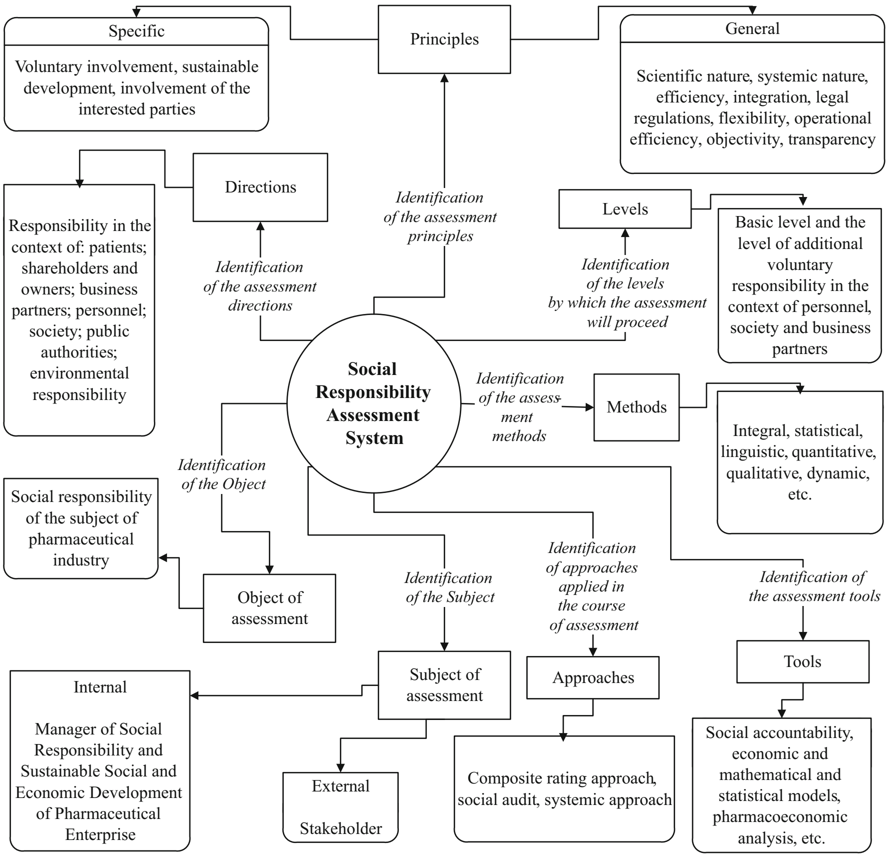


Fig. 1 Assessment of social responsibility of pharmaceutical business

correlation ratio exceeds 0.8, according to the Chaddock scale, the values are extremely interrelated, and it would be unreasonable to use both of them at a time, i.e., one indicator shall be removed from the totality. Out of any such pair, we would pick the indicator that had the least or no values of high correlation with other indicators comprising the totality following these calculations, we have selected 46 local indicators for the SRPhB assessment (Fig. 2).

Our SRPhB assessment method is based on the integral approach and taxonomic analysis. Therefore, SRPhB assessment may proceed as follows:

$$I_{SRPhB} = f\,(k_{YakLZ}; k_{NT}; k_{EK}; k_{STr}; k_{AKc}; k_{Bz};) \tag{1}$$

where I_{SRPhB} is the integral value of the SRPhB; K_{YakLZ} is the composite index of responsibility for the timely provision of quality, safe and affordable medicines (LZ) to the public; K_{NT} is composite index of responsibility for policy-making

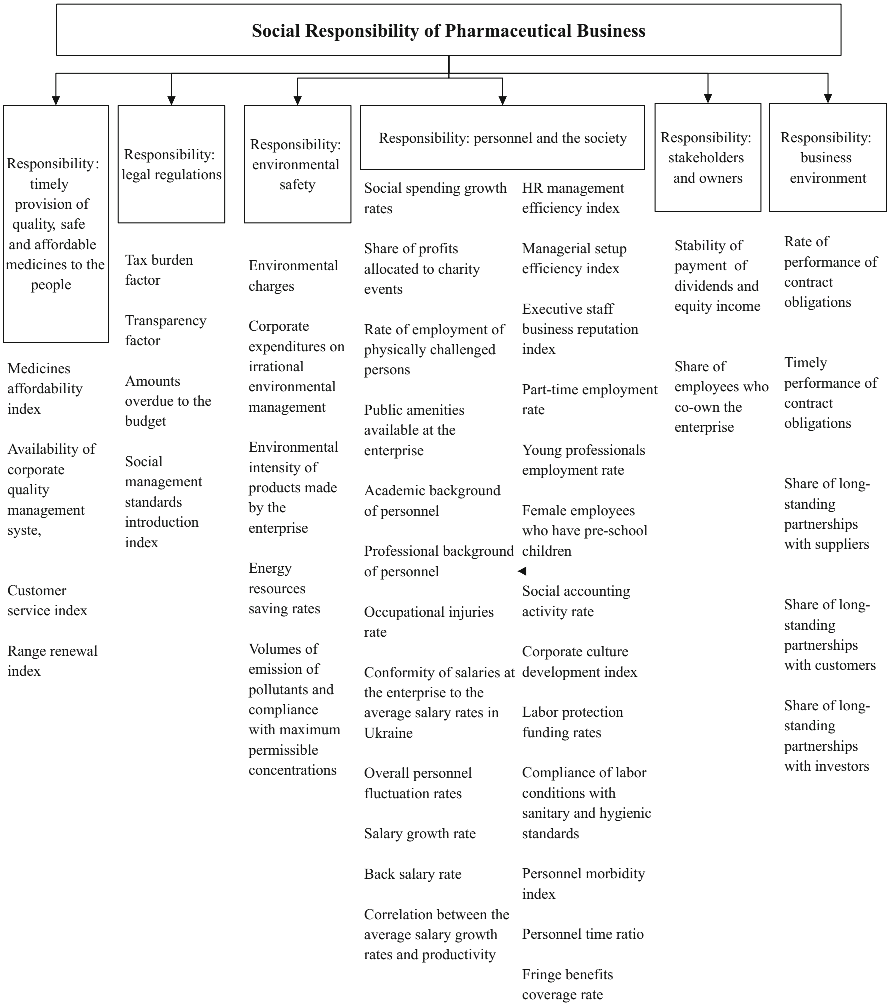


Fig. 2 Suggested system of indicators of the SRPhB assessment

and promotion of vocational and higher education; K_{Ek} is the composite index of responsibility for environmental safety; K_{STr} is the composite index of responsibility before the personnel and the public; K_{Akc} is the composite index of responsibility before shareholders and owners; K_{Bz} is the composite index of responsibility before business environment.

The abovementioned composite indices covering every SRPhB component are calculated on the basis of local indices set forth in Fig. 2. The suggested SRPhB assessment methodology has been tested on a number of pharmaceutical companies. The estimates are set forth in Table 2. The same SRPhB assessment methodology is complemented by the assessment of the synergic effect resulting from balanced

Table 2 2018 SRPhB assessment results for the studied Ukrainian PhC

Pharmaceutical enterprise	Integral SRPhB index	Synergetic factor	Cluster number	Rating (within a group, considering the synergetic factor)
Manufacturers				
PAT Farmak	0.78	1.17	1	4
Arterium Corporation	0.85	2.08	1	3
PAT NPC Borshchahivskiy Ch&Ph	0.63	–	2	7
FF Darnytsya	0.74	1.45	1	5
TOV FK Zdorovya	0.64	1.15	2	6
PAT Kyivskiy Vitaminny Zavod	0.53	–	2	10
TOV Micropharm	0.27	–	4	12
TOV Ternopharm	0.26	–	4	13
PrAT Lekhim-Kharkiv	0.48	–	3	11
PAT Pharmstandard-Biolik	0.49	–	3	9
Corporation “Yuriya-Pharm”	0.63	–	2	8
TOV Takeda Ukraine	0.86	1.18	1	1
TOV Teva Ukraine	0.87	1.61	1	2
Distributors				
TOV BaDM	0.62	1.22	2	2
JV Optima-Pharm Ltd.	0.43	–	3	4
Pharmacy chains				
“Med Service Group”	0.71	–	2	3
“Gamma-55”	0.61	–	2	5
“Apteka Nyzkykh Tsin”	0.72	2.16	1	1
“Leda”	0.51	–	2	7
“9-1-1”	0.54	–	2	6
“Apteka Dobroho Dnya”	0.66	–	2	4

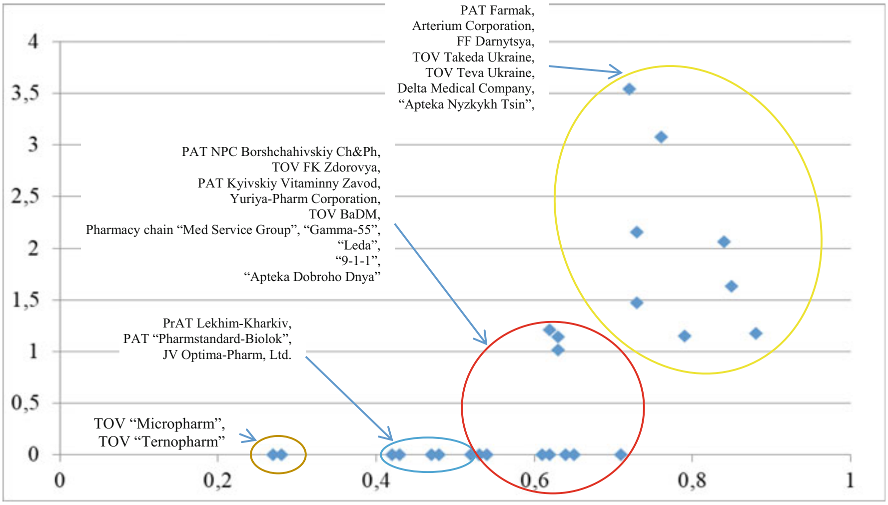


Fig. 3 Pharmaceutical companies clustered by the SRPhB level

management of every social responsibility component of PhC and is estimated with the help of the index that is calculated in accordance with the method described in our earlier studies [18].

Based on the calculations, we operated Statistica 12.0 and Grafikus to conduct cluster analysis of pharmaceutical companies by the level of SRPhB. Based on the results of the analysis, the studied PhC have been grouped into four clusters (Fig. 3). Each PhC cluster shall have an individual SRPhB strategy with implementation procedures outlined. Introduction of the suggested SRPhB assessment methodology will help identify bottlenecks in SRPhB governance, define the resource pool which the PhC will be able to apply for the promotion of its socially responsible activities, develop a package of measures to promote the effective SR management, contribute to the establishment of partnership with stakeholders, introduce principles of social responsibility in pharmaceutical industry, and promote the practice of open rating of the national PhC by the SRPhB, which will be an important step toward enhancement of investment attractiveness of the PhC and boosting the consumer loyalty.

Our study then proceeded to the next stage to establish correlation between the SRPhB and the sustainable development index of pharmaceutical companies. For purposes herein, the sustainable development means the advancement that is based on the balance between meeting the contemporary needs of the public and protecting the interests of future generations, including their need for higher quality of life, factoring in such aspects as health, safety, and clean environment. We have suggested the index system for the assessment of sustainable social and economic development of pharmaceutical companies in our earlier works [9, 22]. Under this approach, the sustainable social and economic development of pharmaceutical companies can be assessed as follows:

$$Y_{CCEP} = f\left(I_{EP}; I_{CP};\right), \tag{2}$$

where Y_{CCEP} is the summarizing taxonomic indicator of sustainable social and economic development of pharmaceutical companies; I_{EP} is the integral indicator of economic development of the PhC; I_{CP} is the integral indicator of social development of the PhC.

In turn, formalized calculation of the integral indicator of economic development (I_{EP}) and integral indicator of social development (I_{EP}) of the PhC can be described in formulations (3) and (4):

$$I_{EP} = f\left(Bm_K; F_K; M_K; Y_K; Ii_K;\right), \tag{3}$$

where Bm_K is the composite index of the production and technological component of economic development of the PhC; F_K is the composite index of financial component of economic development of the PhC; M_K is the composite index of marketing component of economic development of the PhC; Y_K is the composite index of management component of economic development of the PhC; Ii_K is the composite index of innovative investment component of economic development of the PhC.

$$I_{CP} = f\left(K_K; C3_k; Mm_k; Ti_K\right), \tag{4}$$

where K_K is the composite index of the HR component of social development of the PhC; Сз_K is the composite index of the social welfare and security component of the PhC personnel; O_K is the composite index of the PhC labor management component; Мт_K is the composite index of the motivational component in the PhC development; Ti_K is the composite index of the creative and intellectual component in the PhC development.

Table 3 summarizes the results of assessment of the level of development of the studied pharmaceutical companies.

Seeking to provide PhC with a package of effective tools to diagnose the reserves for improvement of socially responsible management, we suggest the Grafikus-based 3D graphic model. Figures 4, 5, and 6 are examples of how to build the abovementioned models for the studied PhC distributed in various clusters by the level of their social responsibility. On the graphic models, one may notice that the yellower area of the surface representing the development of a pharmaceutical enterprise stands for the higher level of development thereof, while the greater radius of the surface stands for the poorer development of the PhC. The study of social and economic performance of the Ukrainian pharmaceutical companies defines five levels of development (sustainable, high, medium, below medium, and low). For each level of the PhC development, we suggested key tasks for the improvement of sustainable development management system and developed the operational guidelines for the application of the social and economic control tools.

Table 3 2018 Indices of sustainable social and economic development of pharmaceutical companies in Ukraine

Pharmaceutical enterprise	Summarizing indicator of sustainable social and economic development (on a scale of 0–1)	Synergetic factor (if < 1—no synergetic effect is present)	Level of social and economic development	Cluster No.	Rating among the studied PhC (within the group)
Manufacturers					
PAT Farmak	0.69	No synergetic effect	Medium	3	5
Arterium Corporation	0.74	2.06	High	2	3
PAT NPC Borshchahivskiy Ch&Ph	0.64	No synergetic effect	Medium	3	7
FF Darnytsya	0.71	1.47	High	2	4
TOV FK Zdorovya	0.63	1.14	Medium	3	6
PAT Kyivskiy Vitaminny Zavod	0.54	No synergetic effect	Medium	3	10
TOV Micropharm	0.28	No synergetic effect	Low	5	12
TOV Ternopharm	0.48	No synergetic effect	Below medium	4	11
PrAT Lekhim-Kharkiv	0.56	No synergetic effect	Medium	4	9
PAT Pharmstandard-Biolik	0.61	No synergetic effect	Medium	3	8
Corporation “Yuriya-Pharm”	0.78	1.18	High	2	1
TOV Takeda Ukraine	0.73	1.63	High	2	2
TOV Teva Ukraine	0.67	1.63	High	2	2
Distributors					
TOV Arthur-K	0.63	No synergetic effect	Medium	4	3
TOV BaDM	0.63	1.21	Medium	3	2
Pharmacy Chains					
“Med Service Group”	0.71	No synergetic effect	Medium	3	3
“Gamma-55”	0.62	No synergetic effect	Medium	3	5
“Apteka Nyzkykh Tsin”	0.73	2.16	High	2	1
“Leda”	0.52	No synergetic effect	Medium	3	7
“9-1-1”	0.53	No synergetic effect	Medium	3	6
“Apteka Dobroho Dnya”	0.65	No synergetic effect	Medium	3	4

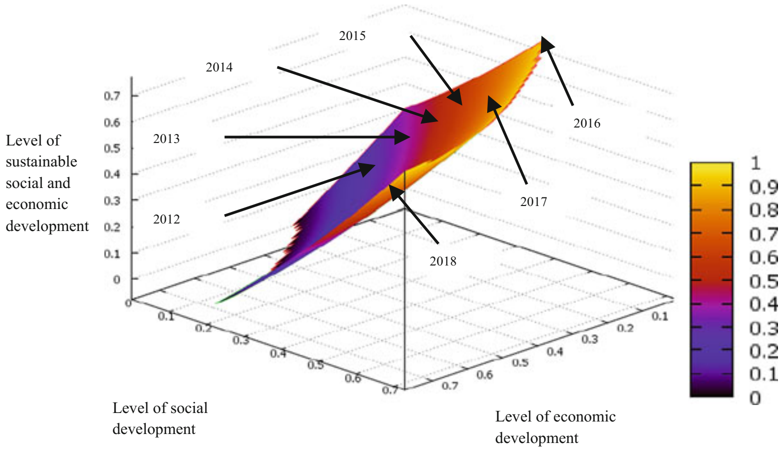


Fig. 4 Graphic model of PAT Farmak development

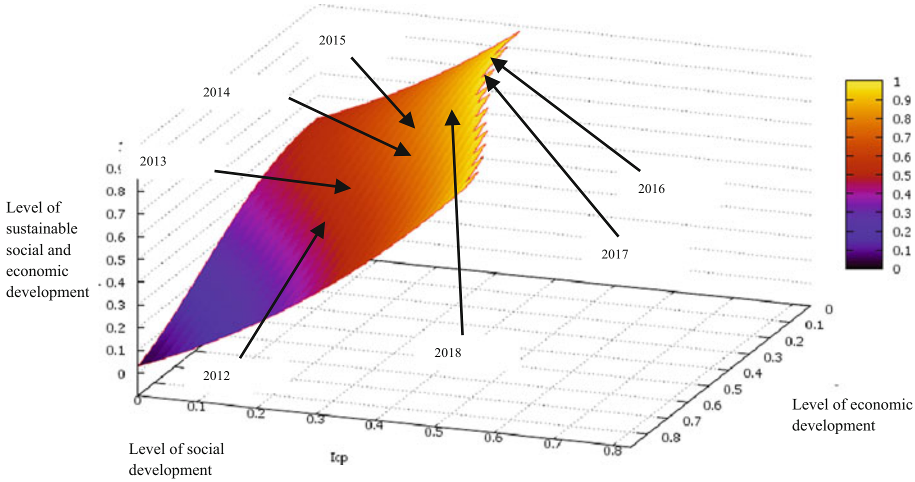


Fig. 5 Graphic model of Arterium Corporation development

In order to complete the tasks at hand, the PhC with sustainable and high level of social and economic development are recommended to focus on strengthening of the social responsibility system and on control of market forces while making their management arrangements; pharmaceutical companies with medium level of development should also activate the economic tools to control their interaction. When the level of development is below medium or low, one should apply all components of social and economic controls.

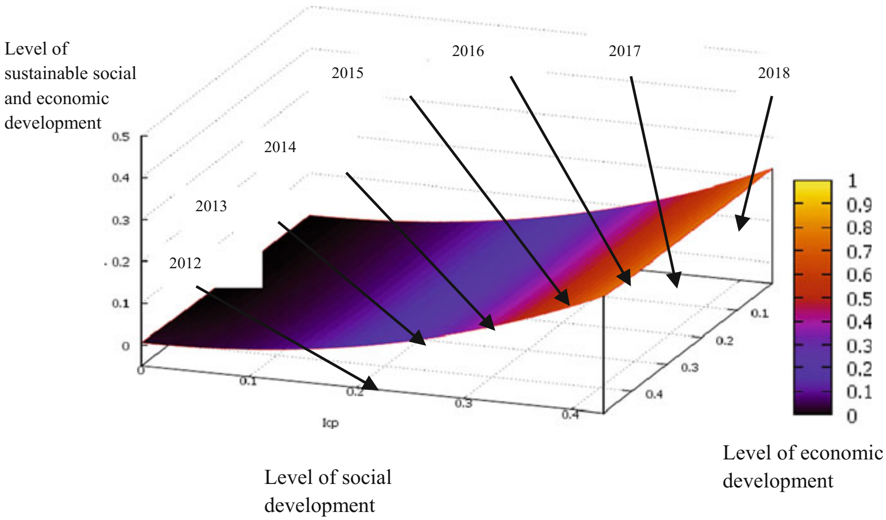


Fig. 6 Graphic model of TOV Micropharm development

5 Conclusions

Ratings of the most socially responsible international companies have been studied in real-time mode. Social responsibility management is proven to be topical for the promotion of sustainable social and economic development and attainment of strategic competitive edge by the companies.

Even though medicine and pharmacy are socially responsible industries, only one pharmaceutical company made it to top-10 of the most socially responsible companies worldwide over the last 10 years.

It has been established that 63% of the Ukrainian pharmaceutical companies covered by this study maintain their social responsibility only at the basic level. Notably enough, the Ukrainian pharmaceutical companies appear to be more socially responsible worldwide than domestically.

This study has defined the essence and the components of socially responsible pharmaceutical business (SRPhB).

We have suggested methodological approaches to the SRPhB assessment based on the economic and mathematical methods and information technologies. The suggested methodology was used for the assessment of social responsibility of more than 20 Ukrainian pharmaceutical companies with application of taxonomic analysis and clustering of the said PhC by the level of their social responsibility.

It has been established that the SR level and the level of sustainable social and economic development correlate, however, subject to balanced management.

For diagnosing of the reserves for improvement of socially responsible management, we have suggested the 3D graphic model based on the Grafikus service.

References

1. *Global study by IBM*. https://www.ibm.com/ua-en
2. *International Standard of Social Responsibility ISO 26000*. http://bcdst.kiev.ua/index.php/news/1-latest-news/56%2D%2Diso-26000
3. Bowen, H. (1953). *Social responsibilities of the businessman* (p. 276). Harper & Row.
4. Carroll, A. (1974). A three-dimensional conceptual model of corporate performance. *Academy of Management Review, 4*(4), 497–505.
5. Carroll, A. *Corporate social responsibility: Evolution of a definitional construct*. http://bas.sagepub.com/cgi/content/abstract/38/3/268
6. Davis, K. (1967). Understanding the social responsibility puzzle: What does the businessman owe to society. *Business Horizons, 10*(4), 46–47.
7. Davis, K. (1974). The meaning and scope of social responsibility. In J. W. McGuire (Ed.), *Contemporary management: Issues and viewpoints* (pp. 629–633). Prentice Hall.
8. Druker, P. (1984). The new meaning of corporate social responsibility. *California Management Review, 26*(2), 53–63.
9. Friedman, M. (1962). *Capitalism and freedom* (p. 202). University of Chicago Press.
10. Friedman, M. *The social responsibility of business is to increase its profits*. http://www.colorado.edu/studentgroups/libertarians/issues/friedman-soc-resp-business.html
11. Schwartz, M., & Carroll, A. (2003). Corporate social responsibility: A three-domain approach. *Business Ethics Quarterly, 13*(4), 503–530.
12. Grishnova, O. (2010). Social responsibility of business: Understanding the essence, contemporary status, problems and prospects of development in Ukraine. In A. Kolot, O. Grishnova, & O. Gerasymenko (Eds.), *Socialization of relationship in the sphere of labor in the context of sustainable development: Monograph* (pp. 189–210). KNEU.
13. Deich, M. (2012). Social responsibility in formation of conditions for sustainable development. In O. Novikova, O. Amosha, & V. Antonyuk (Eds.), *Sustainable development of an industrial region: Social aspects: Monograph* (pp. 223–229). IEP NAS of Ukraine.
14. Cherkashina, A., & Kotvitska, A. (2014). Study of contemporary status of social responsibility of the biggest pharmacy chains in the City of Kharkiv. *Governance, Economics and Quality Assurance in Pharmaceutical Industry, 2*, 36–42. http://nbuv.gov.ua/UJRN/Uezyaf_2014_2_8
15. Tkachenko, N. (2013). Analysis of definition "social responsibility" and its integrated construction in pharmaceutical. *Zaporozhskiy Meditsinskiy Zhurnal, 3*, 125–128.
16. Tkachenko, N., & Gromovyk, B. (2017). Conceptual analysis of social responsibility of a person and its integration in pharmaceutical practice. *Actual Problems of Pharmaceutical and Medical Science and Practice, 10*(3), 342–348. http://nbuv.gov.ua/UJRN/apfimntp_2017_10_3_19
17. *Industry Global CSR Index*. http://www.reputationinstitute.com
18. Bratishko, Y. (2014). Assessment of social responsibility of pharmaceutical business. In *Problems of social and economic development of business* (Vol. 1, pp. 205–209). Publish.house "BREEZE".
19. Posylkina, O., Bratishko, Y., & Kubasova, G. (2015). Formation of the system of indices of assessment of social responsibility of pharmaceutical businesses. *Clinical Pharmacy, 19*(2), 4–8. http://nbuv.gov.ua/UJRN/KlPh_2015_19_2_3
20. Social Responsibility and Organizational Ethics. (2001). *Encyclopedia of business and finance* (Vol. 1, 2nd ed., p. 814). Macmillan Reference.
21. Bratishko, Y. (2015). Cluster analysis of pharmaceutical enterprises in Ukraine by the level of sustainable social and economic development. In *Actual problems of modern economy development: Collection of scientific articles* (pp. 88–93). Thorpe Bowker.

Secure Mobile Agent Using Identity Combined Key Signature-Based Schnorr Protocol

Sanae Hanaoui, Jalal Laassiri, and Yousra Berguig

Abstract Mobile agent technology is a software paradigm which can migrate from one station to another, in order to execute a code or to collect information via network, and only recently that it has seen a significant degree of use in different domain applications as e-commerce. Because of its intelligence combined with its distributed architecture, including its capability to delegate, communicate, and negotiate tasks among other agents, it has gained widespread recognition. Yet because of its strong mobility the mobile agent faces some security issues as verifying the integrity and authenticity of information carried by the agent or its code while its migration over the network. In this chapter, we propose a protocol for transaction security by adopting a new identity combined key-based Schnorr signature inspired from the secure ID-based signature scheme and the Schnorr signature, applied to the agent technology, thus to validate its authentication as a digital transaction as an alternative for certificate-based proxy signatures to grant its data integrity, authentication, and non-repudiation in digital communications.

Keywords Mobile agent · Cryptography · Identity-based cryptosystem · Authenticated encryption · Secure transaction · Schnorr signature · Distributed systems · Intelligent systems

1 Introduction

In the computer science and network, the information integrity and authenticity check are a prime necessity [1]. Therefore, it is one of security requirement we must guarantee in the mobile agent technology. In our work, we will consider the agent validated as authentic (unmodified) by any other parties while its migration from a platform to another one in order to execute a distributed task assigned by a

S. Hanaoui (✉) · J. Laassiri · Y. Berguig
Laboratory of Computer Science, Systems and Optimization (ISO Lab), Department of Computing, Ibn Tofail University, Kenitra, Morocco
e-mail: sanae.hanaoui@uit.ac.ma; LAASSIRI@uit.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_24

user via validation of transaction. However, mobile agents are susceptible to several attacks, particularly by malicious hosts due to its ability of executing mobile code in a remote host. There has been a massive research to solve the agent security lacks. One of the reasonable and practical approaches to solve these issues is to provide software-based mechanism to prevent any kind of vulnerability. Nevertheless, it is very difficult to opt for implementing any kind of secure function in mobile agent since all the code and data of mobile agent are exposed to the remote host [2].

The coming years will see a tremendous growth in the digital content transactions through e-commerce. Therefore, it is estimated that mobile agent will be widely used for electronic commerce as an important key technology; in which we could delegate the agent the task of signing a message in a remote server on behalf of a customer without exposing his/her private key, it can be used not only to search for special products or services, but also to make a contract with a remote server [2]. As a result to emphasis the enforcement of its security, we will propose a multiple identity-based signature-based Schnorr to guarantee its authenticity. To guarantee authentication: by giving the recipient host a reason to believe that the agent was created by a known sender; non-repudiation: by assuring that the sender cannot deny having sent the agent; and integrity: to make sure that the agent was not altered in transit.

Our signing protocol is based on the identity-based Schnorr scheme. In our proposed protocol, we define a set of designated cosigners in a directory, selected by the system authority. And based on the original signer request, the system authority selects a N random number of disturbed registered cosigners' identities from the suggested set based on their availability and work load who can contribute in signing with the original signer, then communicate their set of identities to the Private key Generator, called the PKG, to generate a new combined secret key for the original signer associated with its corresponding public key, in our scheme, the public key of the original signer is embedded into the signature so that any verifier can verify that the right person only has signed the agent.

2 Related Work

In the literature, we find few researches about the multi-agent signature, either new contributions or improvements of old propositions.

2.1 Mobile Agent Security Using Signature Schemes

The interest of using the mobile agent technology is growing in the recent years; therefore, it becomes one of the more dynamic search areas specially because of its promising feathers which can empower the ecommerce, thence to prevent any drawbacks of this technology, many researchers are working on solving security issues especially the authentication problem because of its strong mobility.

Kotzanikolaou [3] had proposed a cryptographic solution for secure autonomous agents which is based on dynamically generated message-flexible multi-signatures. Their scheme does not compromise the advantages of the mobile paradigm. Bamasak and Zhang [4] proposed a method that allows the delegation of signature power to one or more entities that jointly play the role of a proxy signer especially in agent-based paradigm in which proxy signers are mobile agents that are executed in remote untrustworthy hosts. This feature is attractive for signature delegation, in addition they provided confidentiality protection to the proxy key, the method protects against repudiation of signature delegation by the original signer, repudiation of proxy signature generation by the proxy signer, and repudiation of receipt of the proxy signature by the signature recipient. Lin et al. [5] have proposed an improvement model based on blind signature and proxy host that would protect service hosts as well guarantee the fairness of online transactions. Their environment offers the promise of tracking malicious mobile agents once a service host is attacked. It also protects the information collected from a mobile agent from being manipulated. Saxena and Soh [6] had described a protocol using bilinear pairings that enables trust relationships to be formed between agent platforms in an ad hoc manner without actively involving any trusted third party, which can be used to authenticate agents before granting execution privileges. The main idea of their approach is the concept of "one-way" chaining. Tsaur [7] had developed a proxy signature scheme and a proxy authenticated encryption scheme for protecting mobile agents against malicious agent hosts using the ECC-based self-certified public key cryptosystem. Their introduced proxy signature scheme can protect users' private keys stored in smart cards and provide the fairness of contracts signed by agents. It is constructed using the ECC, and it integrates the identity-based public key cryptosystem with the self-certified public key cryptosystem. While Idrissi and Souidi [8] had introduced an approach based on cryptographic mechanisms, which involves an Identity-Based Key Agreement Protocol to get a session key and ensure authentication, an Advanced Standard Encryption (AES) for the confidentiality of data exchanged, as well as a Binary Serialization to get an easy and persistent portability of the agent across the network. Also Shi et al. [9] had conducted a concrete identity-based undetachable digital signature scheme with provable security in order to enable mobile agents signing securely on potentially malicious hosts in electronic commerce or in any other applications. Thus, in their scheme, mobile agents do not need to carry the private key when they generate digital signatures on behalf of the original signer; therefore, the private key will not be compromised. Their used encrypted function is combined with the original signer's requirement, so misuse of the signing algorithm can be prevented. Moreover, they adapted identity-based scheme, so the verification of the signatures generated by mobile agents does not require either verification of the entire certificate path or communication with the certification authority.

2.2 The Protocol Construction

In this section, we put forward our main construction of our disturbed ID-based combined key agent protocol, based on Schnorr signature. The signing protocol uses a cyclic group G of prime order p, a generator of a multiplicative subgroup Z_p^* g of G, and a collision-resistant cryptographic hash function H. The description of the construction is as follows: we adopted the four algorithms of the identity-based signature scheme, and we will add the fifth algorithm which will combine the keys. As a result, our protocol will consist of the following five algorithms. We inspired from the identity-based Scheme [10] and the Schnorr signature Scheme [11, 12]. The flow of functionalities of these algorithms of our protocol is described in Fig 1.

Setup: Given security parameters $\tau, \lambda \in Z(\tau > \lambda)$ as input, PKG runs this algorithm to generate system parameters.

Extract: Given a user's identity (ID), PKG runs this algorithm to generate initial private key.

Combine: Given the output of the extract algorithm, the PKG runs this algorithm to combine the partial initial shares $\{CSigner_j\}_j \in CoSigners$ from cosigners in the set $CoSigners$ to generate $S \leftarrow (Skey, Pkey)$.

Signature Generation: To generate a signature on a message m in our case, it would be the agent using the combined key received from the PKG after running the combine algorithm. This algorithm is run by the original signer.

Verification: Given the system parameters and a signature tuple; any verifier can check the validity of the signature using this algorithm.

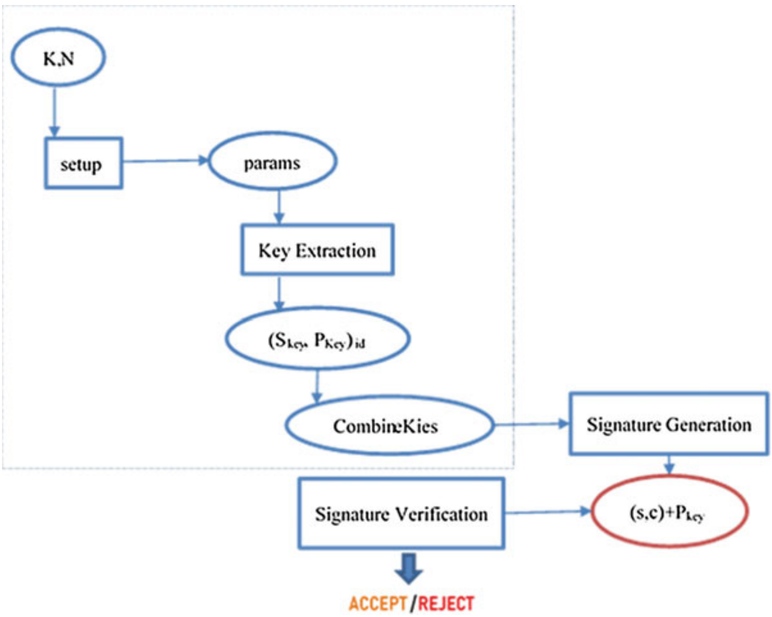


Fig. 1 Principle of the proposed protocol

3 Agent Security Application

We apply the ID-based combined key Schnorr signature to mobile agent situation. First, we elaborate the security problems for the mobile agent technology.

3.1 Security Problem

The complex computing which can be performed online with a non-permanent connection is likely requested by the distributed architectures in order to ensure the availability of information as well to enhance its efficiency. Hence, one of the technologies that designed to fulfill this process is the mobile agent technology, as with total independence of the environment where it is located/created, it has the ability to perform all necessary computation with efficiency. Yet, the autonomously migration between different host platforms of the agent might threaten its safety. Therefore, it is essential to guarantee its correct conduct without any risks of tempering the agent; on the other hand, it is crucial to certify that there will be no threat in execution of the agent in the destination host.

In this chapter, we are concerned by assuring the authenticity, integrity as well the non-repudiation of the mobile agent. Owing to the complexity of the dynamic interactions between the hosting platforms and the mobile agents makes it in reality a complex task to anticipate the behaviors of both the agents and the hosting platforms. In order to affirm that the claimed initiator of the agent is the same as the real initiator, we opted to apply our proposed ID-based combined key Schnorr signature described in the section above.

We model our proposed scheme on the following assumption:

– The mobile agent must be transferred with its accompanying signature. Each receiving platform H_i has to verify the signature before it is executed. Execution should only be possible if the verification process succeeds or other security policies of the platform are satisfied in case of adopting one.

3.2 The Identity-Combined Key Protocol for Agent Signature

See Table 1.

3.2.1 Initialization

The entities in the system are: a system authority (SA), users (U_i), hosts (H_i), and mobile agents (MA) generated by specific user or users in the original platform

Table 1 Notations used in the proposed protocol

Symbol	Description
SA	System authority
U_i	Users in the system suspectable signers
MA	Mobile agents generated by specific user or users in the original platform
H_i	Hosts
CS_i	Cosigners: Distrusted helper signers
I_i	The identity information of a user in the system
$I_{cs\,i}$	The identity information of a cosigner
$@(cs)$	The network address of a cosigner
X	The combined secret key for the original signer
Y	The corresponding public key of the secret key

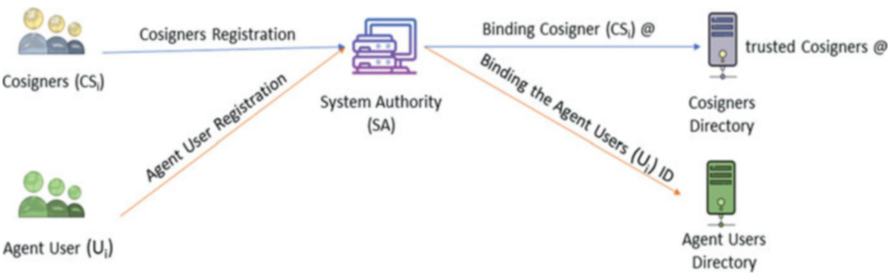


Fig. 2 System initialization

(original signer), cosigners (CS_i). We assume that SA is responsible for key generation, cosigners, and user registration (see Fig. 2).

The agent user U_i performs the following steps to register to SA:

Step 1. U_i executes the following tasks, respectively:

1. Selects an identity information, denoted by I_i.
2. Submits its identity information I_i to SA.

The cosigner (CS_i) performs the following steps to register to SA:

Step 1. CS_i executes the following tasks, respectively.

1. Selects an identity information, denoted by I_i.
2. Submits its identity information I_i as well as its network address denoted by @ to SA.

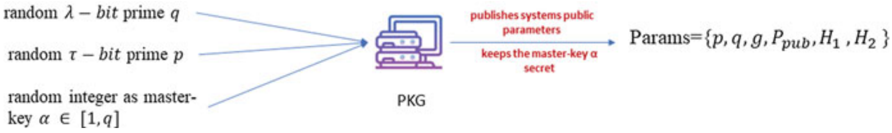


Fig. 3 System setup

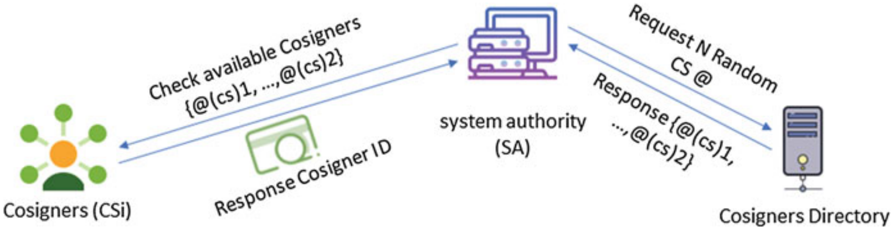


Fig. 4 Step 2 phase (1)

Fig. 5 Step 3 and Step 4 schema

3.2.2 The System Setup Phase

Private key generator PKG creates the systems public parameters in this phase, and then it releases these parameters to the system authority in communication. PKG randomly generates its master key and keeps it secret (see Fig. 3).

Step 1. The PKG generates public parameters: params $= \{p, q, g, P_{pub}, H_1, H_2\}$, $\alpha \leftarrow Generate(\tau, \gamma)$.
Step 2. The PKG publishes params and keeps as secret its master key.

3.2.3 The Key Generation Phases

In this phase, the agent user U_i requests the key generation from SA to obtain (public key, private key) (see Figs. 4 and 5).

Step 1. U_i requests to sign the agent with the corresponding ID in the instance t from the SA:

1. U_i : Submits $(AgentId, I_i)_t \rightarrow SA$.
2. SA: $L \leftarrow Verify(I)$. The SA verifies the existence of the agent user in the system, if $L \leftarrow Accept$, the SA continues the rest of the steps.
3. SA: $A \leftarrow$Archive ($SignReq_i$, $AgentId$, I_i, t). The SA saves the archive of the signing Request Operation for the period of time t.

Step 2. SA initiates a secure communication canal with the third trusted party the PKG and executes the following tasks for U_i.

1. SA: $\{CS_1, .., CS_N\} \leftarrow CsAddress(N)_t$. Requests N random network address of (CS)@ from the cosigner directory.
2. SA: $I_{csi} = \{I(cs)_1, \ldots, I(cs)_N\} \leftarrow CSGet(@(cs)_1, \ldots, @(cs)_N)$. SA Check available trusted cosigners in the network $\{@(cs)_1, \ldots, @(cs)_N\}$.
3. SA: $RCosigner \leftarrow RandomCS\,(I_{csi}, M)$.
4. SA submits system ID's {1_i,R _ Cosigner}to the PKG.

Step 3. Extract phase of the scheme. The PKG executes the following tasks for U_i and Random CS_i, respectively:

1. Computes the (private key, public key) with its corresponding identity where.

```
for i,  j in { 1 i,  R _  Cosigner }
x i =  αP j,   P j= H1 ( i j)  .
    S i ←  appendSet( x i)
```

Step 4. Combined keys phase of the scheme. The PKG executes the following steps for each x_i of both the original signer U_i as well as the random cosigners set.

1. Combines public keys and private keys: $Y \leftarrow g^X$ and $X \leftarrow \prod_{i=0}^{l} x_i mod p$. For the ($U_i$, $\sum R\_Cosigner_j$) where.

```
for x in S i
X,Y ← Combine( x)
```

2. PGK sends to the SA the couple $S \leftarrow (X, Y)$, where X : *secretkey*, Y : *publickey*.

Step 5. The SA shares to the original signer U_i the couple $S \leftarrow (X, Y)$.

3.2.4 Singing

The original signer U_i uses couple $S \leftarrow (X, Y)$. He signs the agent using the sign algorithm:

1. $U_i : (s, c) \leftarrow SignAgent(S, Agent)$.

3.2.5 Verification

At the arrival of the agent to destination host H_i (hosting platform) from the native platform, it receives in reality the encryption of the mobile agent serialized object along with the signature of the clear serialized object. Then the platform begins by decrypting the first part to obtain the plain format of the serialized object, which is later signed using our combined ID-based Schnorr signature scheme and compared with the second part of the received data. If both the signatures match, then the integrity of the mobile agent received is proven. By calling the Verify algorithm of the scheme:

2. H_i : ACCEPT/REJECT $\leftarrow$ $Verify(s, c)$.

4 Test and Application

The following depicts the computational complexity of the implemented signature protocol: we use the notations described in Tables 2, 3, and 4.

Comparing the complexity and running time of the signing protocol, we notice that timing result is very promising for the use of the proposed approach in securing agent application.

Table 2 Notations used for computational cost

Extract and key combining	ntmul+ texp+nth
Signature generation	tmul+texp+th+tadd
Verification	tmul+2texp+th

Table 3 The computational costs in the proposed protocol

Operation	512-bit (ns)	1024-bit (ns)	JAVA methods
SHA-256	38,730	58, 92	MessageDigest.getInstance (“SHA-256”)
Modular	16,956	34,932	Mod(BigInteger mod)
Multiplication	30,981	70,762	Multiply(BigInteger val)
Addition	0,563	6,5	Add(BigInteger val)
Exponentiation	4,155,056	9,155,056	modPow(BigInteger pow, BigInteger mod)

Table 4 Operation timing

	Setup and extract (ns)	Singing (ns)	Verification (ns)
Our Schnorr signature protocol	20,059,077	100	170

5 Conclusion

In this chapter, we proposed an efficient distributed ID-combined key-based Schnorr signature protocol to emphasize the agent security, which satisfies all the security requirements needed (provably secure with the hardness assumption to the difficulty of solving DLP as well as due to one-way property of the hash function $H(.)$), so in the above, we modeled our proposed protocol based on some assigned assumptions, and we simulated our protocol for agent signature, thus we noticed that timing result is very promising for the use of the proposed approach in securing agent application. As a perspective, we will further study our protocol, simulate it, and compare its performance with other different authentication protocols based on signature.

References

1. Bellare, M., Canetti, R., & Krawczyk, H. (1996). Keying hash functions for message authentication. In *BT—Proc. 16th Annu. Int. Cryptology Conf. Adv.* (pp. 1–15).
2. Steinfeld, R., & Hawkes, P. (2010). Information security and privacy. In *15th Australas. Conf. ACISP 2010, January 2002* (pp. 1–403).
3. Kotzanikolaou, P. (2001). *Dynamic multi-signatures for secure autonomous agents* (pp. 587–591).
4. Bamasak, O., & Zhang, N. (2004). *2004 ACM Symposium on applied computing. A secure method for signature delegation to mobile agents* (pp. 813–818).
5. Lin, M. H., Chang, C. C., & Chen, Y. R. (2004). A fair and secure mobile agent environment based on blind signature and proxy host. *Computers & Security, 23*(3), 199–212.
6. Saxena, A., & Soh, B. (2005). An authentication protocol for mobile agents using bilinear pairings. *Computer Engineering*, 1–11.
7. Tsaur, W. J. (2012). Secure communication for electronic business applications in mobile agent networks. *Expert Systems with Applications, 39*(1), 1046–1054.
8. H. Idrissi and E. M. Souidi, "Mobile agent security using ID-based agreement protocol and binary mobile agent security using ID-based agreement protocol and binary serialization, April 2016, 2015.
9. Shi, Y., Han, J., Li, J., Xiong, G., & Zhao, Q. (2018). Identity-based undetachable digital signature for mobile agents in electronic commerce. *Soft Computing, 22*(20), 6921–6935.
10. Provisions, D., & Resistance, E. (1998). Israel. *Earthquake, 196*, 1–45.
11. Schnorr, C. P. (1991). Efficient signature generation by smart cards. *Journal of Cryptology, 4*(3), 161–174.
12. Adams, C., et al. (2011). SYN cookie defense. *Encyclopedia of Cryptography and Security*, 1271–1273.

Model Tools for Diagnosing the Stability and Survivability of Economic Systems

Maksym Bezpartochnyi, Daniil Revenko, Halyna Dolha, and Nataliia Trushkina

Abstract In the process of conducting complex studies of the functioning of socio-economic systems of different levels and scales, scientists pay the greatest attention to the questions of the stability of its structure, since its viability and further development depend on the state of the structure of the system. The stability of the socio-economic system is the main feature of its sustainable and efficient dynamic development, and development itself implies a permanent transformation of the system and its modification, while the stability of the structure of the system implies the stability of its structural elements and the preservation of connections between them. The dualism of development while preserving the structure of the socio-economic system forms the basic idea of invariance. The invariance of the property of the structure of the system depends on the period considered and the dynamics of changes in the external environment. In the short term, maintaining the structure of the system is one of the basic conditions for its sustainable development, while in the long term, the structure of the socio-economic system must meet the requirements of the environment. If the structure of the socio-economic system is inefficient and unable to sustain the socio-economic systems to changes in the environment, this situation leads to crisis conditions and, in some cases, the death of the system as a whole.

Keywords Sustainability environment system · Structured approach · Modeling structure · Graphs · Economic systems

M. Bezpartochnyi (✉) · D. Revenko
Department of Economics and Marketing, National Aerospace University N. Zhukovsky "Kharkiv Aviation Institute", Kharkiv, Ukraine
e-mail: m.bezpartochnyi@khai.edu; d.revenko@khai.edu

H. Dolha
Department of Management and Tourism, Chernivtsi Institute of Trade and Economics of Kyiv National University of Trade and Economics, Chernivtsi, Ukraine

N. Trushkina
Department of Regulatory Policy and Entrepreneurship Development, Institute of Industrial Economics of NAS of Ukraine, Kyiv, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_25

1 Introduction

The impact of the crisis on the functioning of the economies of the countries and the question of ensuring the stability of enterprises appear rather sharply in the studies of domestic and foreign authors and sufficiently represented in the scientific literature, and therefore has a very ambiguous formalization. With the expansion of scientific research on the stability of socio-economic systems as complex and dynamic, the issues of treating and investigating the conditions of stability of socio-economic systems from the standpoint of systems theory have become most widespread, but with the question of modeling their structure and developing new methods of ensuring the stability of these structures, and however, their viability has received little attention. From the standpoint of deterministic and stochastic approaches, the stability of the socio-economic system depends on its dynamic properties and, accordingly, may change. A structural approach to modeling the stability of the socio-economic system answers the questions about changing the structure of the socio-economic system and the impact of these changes on its stability. This approach, unlike deterministic and stochastic, makes it possible to analyze changes in the system itself, and not only the disturbing influences of the environment and the overall development of the system. Therefore, the urgent task is to study the stability of the system, as well as to analyze the structural changes of the system itself. In practice, changes in the structure of the system, even minor ones, can lead to qualitative and quantitative changes in behavior and, consequently, in the stability of the system. The structure should be understood as a stable structure of the system, which is constant under various external actions on the system. System structure is a conservative characteristic of the state of the system, which consists of a specific order of the set of elements or parts of the system that interact with each other to realize the specific functions of the system. That is, structure is a means of realizing the goal of system development. Structural is also a characteristic of heterogeneous objects that seek to have a fixed or variable structure, that is, to be internally organized. Under the organization should be understood the dynamic characteristics of the structure of the system, which shapes its purposeful functioning. Among all the definitions, one of the key ones is the structural element—this is a single unit of the structure that performs one of the functions of the system.

2 Formation of a Graph

In general, the simplified structure of the socio-economic system S can be written as such a tuple [1, 2]:

$$S = \langle V, E, R \rangle \,, \tag{1}$$

where $V = \{v_i\}$ is a set of system elements; E is a set of weights of elements or ties; $R = \{r_k\}$ is a set of functions by structural elements of the system.

The weights of the vertices can be expressed by a number or function that characterizes the connections between elements of the socio-economic system. In practice, when modeling the structure of the socio-economic system, the mathematical apparatus of graph theory is more often used, which allows not only to form the structure of the system, rebuild it, or optimize it but also to study the system and evaluate its characteristics, in particular, stability as one of the most important ones. A graph is called the set of vertices and edges [3]. From the point of view of graph theory, the simplest structure of the socio-economic system is the set of elements of the system (V) with the connections between these elements given to them (the undirected graph). A visual method of representing the structure of the socio-economic system, from the standpoint of graph theory, is to display it in the form of a diagram indicating the vertices (elements of the system), as well as combining them with each other lines—edges that form vertices in edges of pairs. Representing the structure of the socio-economic system in the form of a graph makes it possible to visualize mathematical dependencies in the structures of the system.

A non-empty set V, which reflects the structure of the socio-economic system, can be represented as a set of elements of the system $\{v_1, v_2, v_3, \ldots, v_n\}$. And the connections between elements of a system can be represented as the set of all its two-element subsets $V^{(2)}$ in the form of:

$$V^{(2)} = \left\{ \begin{array}{lll} \{v_1, v_2\}, \{v_1, v_3\}, \ldots, & \{v_1, v_n\} & \\ \qquad\qquad\; \{v_2, v_3\}, \ldots, & \{v_2, v_n\} & \\ & \{v_{n-1}, v_n\} & \end{array} \right\}. \tag{2}$$

If in the structure of the system between its elements v_a and v_b there is no connection, in expression (2) they are replaced by values 0.

In a set of edges of a graph (E) we will consider the set of connections between elements, that is, the set of disordered pairs of different elements of the system.

The compact form of the analytical expression of the structure of the socio-economic system according to graph theory is the adjacency matrix, it is a symmetric square matrix $A = |a_{i,j}|$ exponent n, in which the element $a_{i,j}$ equals one if there is an edge in the graph $\{v_i, v_j\}$, namely v_i and v_j are adjacent, in other cases where there are no such edges, in the adjacency matrix the element value is 0. The adjacency matrix has the form:

$$A = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & v_2 & v_3 & \ldots & v_n \end{array} \\ \begin{bmatrix} 0 & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & 0 & a_{23} & \ldots & a_{2n} \\ a_{31} & a_{32} & 0 & \ldots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \ldots & 0 \end{bmatrix} \end{array}. \tag{3}$$

The main statistical characteristics of a graph formed on the basis of the structure of the system are the number of vertices (elements of the system) $n = |V|$, which determines the order of the graph and the number of edges $m = |E|$. A graph with the number n of vertices and m edges is called (n,m)-a graph.

The study of the stability of the structure of the socio-economic system requires determining the center of the graph formed by the structure of this system. The notion of the center of the graph is closely related to the concept of eccentricity of vertices (the distance between two vertices of the graph $e(v) = \max\ d(v_a, v_b)$), the maximum eccentricity is the diameter of the graph ($d(G) = \max\ e(v)$), and the minimum is the radius of the graph ($r(G) = \min\ e(v)$). The link is valid for any graph $r(G) \leq d(G)$. Determining the radius of a graph answers the question about the central vertex: if the condition is satisfied $e(v) = r(G)$, then the vertex is central. This vertex or group of vertices forms the global stability of the system. Disruption of connections between center elements (vertices of the graph) can lead to system failure, so management actions to maintain the stability of the system must be aimed at ensuring the structural integrity of the center of the system.

3 Structure Identification

Identifying the structure of the socio-economic system in graph form, indicating the causal connections between the elements of the system, extends information about the structure of the system, which uses oriented graphs (digraphs). Each prime graph G can be matched by an oriented graph if each of the edges is given one of two possible orientations. We call a digger a pair $\langle V, A \rangle$, if V is a nonempty set of elements of the system that are, respectively, vertices, A is a set of ordered pairs of different elements V, called arcs. Formation of knowledge about the direction or direction of connections in the studied graph of the structure of the system is important for the problems of analyzing and synthesizing the stability of the socio-economic system. The graphs have parallel arcs with common starting and ending vertices. The connection of the arcs at which the end of each arc is the beginning for a new arc is an oriented route (branch or frame) of the system. In the case where the end of the last arc coincides with the initial vertex of the first arc, the oriented route is called a closed or contour structure of the system. If the graph shows the redistribution of resources in the system, then it is called an oriented cycle. In other cases, an oriented graph formed by different arcs (if state is not closed) indicates that the socio-economic system is open.

Representing the structure of a socio-economic system on the basis of a graph allows us to look at the connections between the elements of the system. Forming strong connections between subsystems, system components, and system elements creates the stability of the system itself, so the task of diagnosing system structure connections is important. The main types of linked graphs are:

- A graph is called strongly connected if any two vertices formed by the elements of the structure of the system are mutually reachable (vertex v_a reachable from the vertex v_b, if there is a way $(v_a, \ldots, v_b)$), if the last vertex is reachable to the first, they are mutually attainable.
- Provided that at any pair of vertices of the digraph at least one is achievable, the digraph is called one-sided.

From the definitions it becomes clear that, with a strongly connected graph, the system provides the maximum branching of the structure, which is reflected in its stability; in disturbances in the external or internal environment that may cause some connections or system elements to break, it will be able to compensate for these losses by reallocating resources through other elements and system connections.

Analysis of the structure of the system involves the study of the statistical and dynamic characteristics of the system, decomposing it into separate subsystems and elements, and investigating the system connections between them. Analysis of the structure of the system gives prerequisites for drawing conclusions about the structural stability of the system, and it also allows to carry out topological decomposition of the structure of the system, to identify weak and strong connections between elements and subsystems, to allocate input and output nodes for resources and information of socio-economic system, to determine characteristics of the most important elements of the system, and to explore the maximum and minimum paths in the system graph.

4 Structure Analysis

One of the tasks of diagnosing the stability of the structure of the socio-economic system is to study the changes in the structure of the system over time. This task can be solved by plotting the system structures as a graph at these times and then comparing these graphs. The construction of a model of socio-economic system by deterministic methods allows to determine the stability characteristic of the system at a given time and to study its change in dynamics, and the spatial structure of the socio-economic system in the form of a graph, in turn, allows to determine the structural causes of stability change, so the construction graphs in dynamics and their analysis help to diagnose changes in the structure of system elements and the connections between them that influenced these changes. This task is complicated by comparing the images of graphs obtained at different intervals. In these circumstances, it is sufficient to apply the isomorphism of graphs. In our case, isomorphic graphs are considered to be a one-to-one correspondence between the set of vertices of two graphs in G_t and G_{t+z} structure of socio-economic system in the intervals t and $t+z$, which preserve the connection of adjacency, and we will call the graphs isomorphic, that is $G_t = G_{t+z}$. If graphs G_t and G_{t+z}, formed by a single structure of the socio-economic system, are not isomorphic and the system exhibits a change in the stability characteristic, then the elements and connections between them that have changed over time are investigated. The isomorphism of graphs can be easily detected on the basis of their invariance, that is, the characteristics of the graph G, which assume the same value for any graph isomorphic G. The analysis of graphs for isomorphism should begin by comparing the values of a certain set of invariants, going from simple to more complex. If the quantitative characteristics of the invariants do not coincide and the graphs are not isomorphic, they proceed to the next analysis—a deeper study of the change in the stability of the system.

One of the disadvantages of graphical representation of the structure of socio-economic system is the inability to study the dynamic properties of the structure of the system and its elements, that is, the simultaneous determination of the spatial structure of the system and its behavior over time. To remedy this shortcoming, we propose a method for dynamically displaying the change in structure and the connections between its elements based on a cause–effect model of cost-output. System-dynamic approach allows to analyze the behavior of the system depending on the organization of its structure. Based on this method, changes in the stability of the socio-economic system are analyzed on the basis of a formalized cause-and-effect diagram of changes in the connections between elements of the socio-economic system. The cause-and-effect graph shows the elements of the system structure, and the connections between them are displayed as cause–effect connections (dynamic changes) between each pair of system elements. In turn, these connections can be both positive and negative. The essence of the proposed method is based on the "cost–issue" model proposed by V. Leontiev [4]. The essence of this method is to analyze the costs of the structure of the reproduction in terms of individual sectors according to the main objective of the socio-economic system—to remain viable, while maintaining the sustainability of development, that is, to produce, redistribute, and reproduce resources and benefits to meet their needs and continued economic growth. In practice, the production objective of any socio-economic system can be attributed to two groups:

(a) The produced resources (products) are used for the current needs of the system (intermediate resources).

(b) The resources (products) consumed by the system itself that go to the accumulation (i.e., the increase of capital), called final products.

The graph of the system is formed on the basis of a symmetric matrix, each element of which shows an increase in the distribution of products of a particular industry (for the socio-economic systems of microlevels can be represented by a separate unit of the enterprise), which goes to the production consumption in different industries (subdivisions of the enterprise) of this system. Each of the indices of the matrix Δx_{ij}, this is, the difference between the volume of streams x_{ij}^{t} and x_{ij}^{t-1} from the field i to the field j at times t and $t-1$ for the purpose of production use. For the formation of the matrix, the indices of the difference in the volume of flows at times were chosen t and $t-1$, since these indicators characterize the stability of the socio-economic system. If the sum of all elements of the generated matrix $\sum_{i=1}^{n} \Delta x_{ij} > 0$ (n is the number of elements in the matrix), it says about the stability of the system. However, if the sum of all indicators of the matrix is negative, it indicates an unstable state of the system. Consideration of the first quadrant of the cost-output model reveals the internal structure of the socio-economic system: if the flow increase in the system is positive, then the system produces resources for final consumption. The formed matrix is an adjacency matrix for the construction of a multigraph and has the form

$$\Delta X = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & v_2 & v_3 & \ldots & v_n \end{array} \\ \begin{bmatrix} \Delta x_{11} & \Delta x_{12} & \Delta x_{13} & \ldots & \Delta x_{1n} \\ \Delta x_{21} & \Delta x_{22} & \Delta x_{23} & \ldots & \Delta x_{2n} \\ \Delta x_{31} & \Delta x_{32} & \Delta x_{33} & \ldots & \Delta x_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Delta x_{n1} & \Delta x_{n2} & \Delta x_{n3} & \ldots & \Delta x_{nn} \end{bmatrix} \end{array}. \tag{4}$$

Considering the properties of this matrix and the multigraph that it forms, the formed multigraph of this matrix is an oriented graph in which there are loops and parallel arcs between each two elements of the system. The availability of loops indicates consumption in the resource-producing industry, so the diagonal of the matrix matters. The presence of parallel arcs between each two elements of the matrix is explained by the fact that there are products that are manufactured in the industry i and consumed in the industry j; conversely, products are produced in the area j and consumed in the industry i. Each element Δx_{ij} is the weight of the arc and affects the multigraph. All elements Δx_{ij}, which are negative, impair the stability of the system, and conversely, if the elements Δx_{ij} have positive values, they form the stability of the system.

If we combine all the incident vertices with the edges with the greatest weights $\Delta x_{ij}/\sum_{i=1}^{n} \Delta x_{ij}$, we get a "framework of the system," which shows the combination of the largest industries with each other. These industries shape the accumulation of capital and the growth of resources in the system, that is, form the resilience of the system.

Using the method of researching the stability of the system on the basis of the model "cost–output" allows to identify the most important elements of the socio-economic system that affect its sustainability. The formation of a "system frame" in the study of its stability will allow to develop managerial influences on individual elements of the system in order to increase the global stability of the system.

The last stage of the study assesses the structural stability of the socio-economic system. Because resilience is a dynamic characteristic and V. Leontiev's work [4] shows that relative coefficients that reflect connections between industries are constant, comparing changes in structure over time is an ineffective task in analyzing structural resilience. Quantitatively, the system structure stability indicator reflects the degree of distribution of industrial-economic connections at the vertices of the center of the graph of the structure of the socio-economic system. The significant difference in the distribution of production connections between the center of the graph and its periphery (other vertices of the graph) plays a decisive role in the stability of the structure of the socio-economic system. The uneven distribution of production connections is affected by the branched structure of the system, which has a small number of production connections with the center of the structure, and the large peripheral trees, i.e., the higher the vertices of the graph to its center, the more stable the structure of the system. Significant differentiation between the center and the periphery of the structure creates economic instability and leads to inefficient allocation of resources. Therefore, it is suggested to evaluate the stability of the structure of the socio-economic system, represented as an oriented graph, as the ratio of the sum of the weights of the connections of the central vertices to the total sum of all the weights in the oriented graph

$$\mathrm{SS} = \frac{\sum_{i=1}^{b} v_C + \sum_{j=1}^{d} e_C}{\sum_{i=1}^{n} v + \sum_{j=1}^{m} e}, \tag{5}$$

where v_C is the weights of the vertices of the center of the graph; e_C is the weights of the arcs coming from the vertices of the center of the graph of the structure; b is the number of vertices of the center of the graph; d is the number of arcs coming out of the vertices of the center of the graph; v is the weights of vertices of graph of structure; e is the weights of arcs of graph structure; n is the number of vertices of the graph of the structure; e is the number of arcs of the structure graph.

The stability of the structure of the system can take the following values:

$0 < \mathrm{SS} < 0.3$—low stability of the structure of the system, the center does not provide basic production connections, the system can leave the state of stability if there is a slight destabilization within an individual industry or due to changes in production connections;

$0.3 < \mathrm{SS} < 0.6$—zone of sufficient stability of the structure of the system—the center of the structure graph accounts for about half of all production connections of the system structure, which forms a "producing star";

$0.5 < \mathrm{SS} < 1$—area of considerable stability of the structure of the system—the center of the structure accounts for a large number of production connections, the structure of the system is highly interconnected.

It should also be noted that structural resilience is formed in the very top of the graph, that is, in the productive sectors of the national economy, which have technological, economic, social, legal, and environmental opportunities for economic growth, full satisfaction of the needs of other industries in the formation of relevant services and intermediate or the final product.

Significant technical, technological, economic, social, regional, or other differentiation of industries in the national economy reduce its structural stability.

5 Identifying Structural Connections

Under the influence of dynamic processes at different levels of economic systems changes in the context and, accordingly, the structural connections of the socio-economic system with other systems and subjects of economic activity. A change in structural connections can be positive or negative for the socio-economic system. Under the influence of the different nature of crises, the intensity of structural changes in socio-economic systems is increasing.

The term "structural connections" means industrial connections between industries and sectors of the economy in the structure of the national economy. Structural connections can also be viewed from the point of view of the enterprise, when structural connections are promoted by production and commodity connections, as well as economic connections in rendering services to other enterprises, institutions, or organizations in order to ensure the smooth operation of the enterprise. Structural connections are ultimately an indicator of economic growth, because they are the means of exchange, reproduction, and redistribution of goods and services, as well as the process of value creation. Changes in the structural connections lead to changes in the structure of the supersystem of which the connection consists. When considering the effects of changes in structural ties and their effects on the stability of the socio-economic system, the following reasons that cause them should be highlighted:

– Changes in final consumption
– Changes in intermediate production
– Changes in labor productivity

An important task in the process of studying the sustainability of the functioning of the national socio-economic system is to analyze the structural connections between sectors and sectors of the national economy. The results of the system stability analysis depend on the quality of the identified structural connections of the system. The previous section has solved the problem of identifying the structure of the national economy and studying its stability. But considering the microlevel (level of sector, industry, market, enterprise group, enterprise), external connections act as production flows between industry and other industries, or enterprise and other enterprises and organizations that act as suppliers and customers of products of the enterprise. Given the strength of these connections, the industry or enterprise effectively performs its functions, and when these connections weaken or change dramatically or lose their socio-economic systems, they lose the opportunity for stable and balanced development, which leads to a loss of overall system stability. Therefore, in addition to the methodology of internal stability modeling, for micro-economic systems, it is advisable to assess the stability of the structural connections of socio-economic systems in which the sectors of the national economy or individual enterprises act.

Structural link changes occur under the influence of general economic distortions, structural crises change in the crisis period, one system element replaces another, and the link load is redistributed between system elements or other systems. In a collapsing environment, the system is transformed and, accordingly, in the case of the system's viability, it forms a new structure and new structural bonds. At the microlevel, the change in structural connections is always linked to economic processes: seasonal fluctuations, the stage of the economic cycle, processes of restructuring and reorganization.

As shown by the analysis of the stability of macroeconomic systems in qualitative and quantitative parameters of their development, the characteristic feature of the indicators is dynamic laminarity. Sustainability analysis at the sectoral level made it possible to distinguish the main characteristic of the microlevel—the non-stationary dynamics of the industry indicators, and accordingly the indicators of the functioning and development of enterprises belonging to these industries. Non-stationarity can be caused by structural changes of the following types:

– Change of the basic tendency of development of indicators, incline change of slope of time series
– One-time abrupt change of values (shift)
– A combination of the first two types of structural change

First of all, these structural changes can be caused by the following main reasons:

– Crisis phenomena
– Development of innovative processes and scientific and technological progress
– Change of physical capital, its intensive updating or absence of process of reproduction (deterioration of fixed assets)
– Change of consumer preferences, level of solvent demand from consumers

The task of assessing changes in structural connections at the level of sectoral and entrepreneurial socio-economic systems, as well as assessing the resilience of systems in the context of these changes, remains to be solved.

6 Method of Assessing Stability

The structure of the developed method for assessing the stability of the structural connections of the micro-economic systems of the microlevel is described below. Socio-economic systems can be sectoral systems, business structures, market structures.

The First Stage In the first stage, the information base of the research object is formed. An information source of statistical information is the annual table of expenditures—output in consumer prices.

The Second Stage In the second stage, the information base and sampling are distributed according to the area under study. That is, a sample for structural link analysis consists of two vectors of data showing the input production flows and outputs of an industry under study with all other sectors of the national economy.

The Third Stage At this stage, a vector of quantitative assessment of the structural connections of the sectoral socio-economic system with other sectoral systems is formed. The values of the vector are calculated as the sum of input and output production flows of resources, the industry is being explored with other industries. The received vector of values at the moment t is a quantitative expression of structural connections. If we express through the vector the structure of connections using graph theory, then we obtain a weighted graph in the center of which will be the branch under study, and radial arcs will be emitted from it e_j (the weight of the arc is the value of the vector) to every other branch j, with which the investigated industry has a structural connection, that is, provided $e_j > 0$. Also, in the third stage, several vectors of structural bonds are formed during the selected periods, which need to be investigated.

The Fourth Stage In the fourth stage, a model for assessing the stability of structural bonds is formed. It is advisable to evaluate structural link strengths using summary indices. Structural bonding resilience assessment should take into account the positive and negative changes in structural bonding between the industry (enterprise) being investigated and other microlevel entities. The construction of consolidated indexes for assessing the change in the structure involves taking into account the change of all elements of the system for each pair of periods to be compared. Also, it should be noted that when studying the socio-economic systems of the microlevel and their structural connections, the task of estimating the intensity of change of the structure and its internal and external connections must be solved.

To evaluate the change in external connections between industry and other industries, or between the research enterprise and other suppliers and contracting entities, the following factor is proposed, which provides for the calculation of the difference between a pair of sums that take into account the force and mass of the shift in structural dynamics between two periods.

$$S_z = \sum_{p=1}^{R} \left(\left(\frac{e_{1j}}{e_{0j}} \right)_p^+ - 1 \right) \frac{\left| e_{1j} - e_{0j} \right|}{\sum_{j=1}^{u} \left| e_{1j} - e_{0j} \right|} - \sum_{c=1}^{G} \left(1 - \left(\frac{e_{1j}}{e_{0j}} \right)_c^- \right) \frac{\left| e_{1j} - e_{0j} \right|}{\sum_{j=1}^{u} \left| e_{1j} - e_{0j} \right|} \qquad (6)$$

where

e_{1j} and e_{0j} are values of the total turnover between the investigated industry and the combined industry (j) current and previous years, respectively;

u is the number of industries that are linked by production connections to the industry under study;

$\left(\frac{e_{1j}}{e_{0j}}\right)^+$ and $\left(\frac{e_{1j}}{e_{0j}}\right)^-$ are connections that reflect the positive (>100%) and negative (<100%) dynamics of the shear force between two periods;

R and G are the number of connections related to the positive and negative changes in the dynamics of the shear force, respectively.

The main component in Eq. (6) is corconnections e_{1j}/e_{0j}, which shows the force of the shift in the structural connections between the studied industry and other related industries. As proved in Sect. 3, the sustainability of the socio-economic system can be ensured if the total income in it is positively increased. In other words, if the corresponding e_{1j}/e_{0j} will be more than 1, it will talk about the stability of the corresponding link in the system.

It should be noted that the main condition for the sustainable development of industry and business systems is the combination of changing structural connections with the positive dynamics and economic growth of the industry. The ratio in each of the sums of the modules in Eq. (6) shows the proportion of the corresponding link between industries in the aggregate of all connections between the industry, which is studied with all industries, that is, the expression of the mass of the structural shift. The shift mass calculation takes into account the change in the proportion of production connections in the structure of the set of all connections. The sign of the difference between the two sums in Eq. (6) shows how much the sum of the positive normalized displacements in the structural bonds outweighs the sum of the negative normalized displacements in the studied area. Thus, the coefficient developed takes into account the various parameters of structural displacement estimation in the bond estimation. This coefficient also allows us to estimate the direction of change in the structural connections of the system, that is, to answer questions whether the changes in the structure were positive or negative, or whether the structure of the connections was not change.

The Fifth Stage In the fifth stage, structural linkage stability is assessed based on the setting of boundaries for the change in the structural linkage coefficient of the socio-economic system. Also at this stage, conclusions are drawn from the results of the study.

A change in structural linkages can be considered effective and sustainable when these changes lead to the target state of the system, which results in positive economic growth of the system itself, with minimal financial, institutional, and time costs. According to the results of the calculation of the coefficient of stability of structural bonds, it can take different values, if the value of the coefficient is greater than zero, it will talk about the stability of structural bonds and, conversely, if the value of the coefficient is less than zero, it will talk about the unstable state of industry system connections. Table 1 shows the possible values of the structural coupling coefficient and their interpretation.

Depending on the level of stability and the depth of change in structural connections, transformational changes affect the development of society and the state of the economy:

– Changes in structural ties at the macroeconomic and sectoral levels have an impact on the environment of the functioning of the society, employment, income, which leads to changes in society as a whole.

Table 1 The interpretation of the values of the stability of structural connections

The value of the coefficient S_z	Structural bond characterization of the system
$S_z < 0$	The state of the structural connections of the system can be characterized as not stable. There is a gradual decrease in production flows through most of the structural connections between the industry and other industries. Or, if there is no significant positive increase in production flows for some structural connections, there is a significant decrease in production flows for all other structural connections
$S_z = 0$	A special case is the state of the system in which the balance between the positive and negative production flows across all structural connections. Or, there is almost no change in the structural connections of the systems
$S_z > 0$	Stable state of structural connections of the system with other systems, when the production flows for almost all structural connections are positive, with no significant negative phenomena in the individual structural connections or in the absence of the latter
$S_z > 1$	Adequate state of the structural connections of the system. In this state, the system can undergo major transformations associated with a radical change in structural connections. This situation may also be related to the "ball effect" when there is a significant increase in production flows against the background of their significant decrease in previous periods

– With changing structural connections, requirements for labor resources, their professions and qualifications are changing, as well as demand for labor resources in some sectors of the economy, changes affect the education sector.
– Changes in structural connections affect key economic indicators of economic development: economic growth, total consumer spending, physical capital value, foreign investment.
– Based on the developed method of assessing the stability of structural bonds, it is possible to distinguish its strengths and main disadvantages when using it. The developed index has its disadvantages, such as:

  Structural bond strength index values can be significant, meaning that there are no clearly defined limits on the coefficient change, which makes it impossible to determine the proportions of structural shifts.

  The developed method allows to estimate the change in structural connections, while the reasons and factors that influence these changes remain unknown.

– The results of the evaluation are not prognostic in nature but are instantaneous.
– The model is based on limited data, namely statistics on the cost-output method are published once a year, which makes it impossible to analyze changes in the structure of the economy and structural connections on a quarterly or monthly basis.

Among the positive aspects of the developed method and model are the following:

- Simulation results are easy to interpret.
- The model comprehensively takes into account the strength and mass of displacement in structural connections.
- The method is universal in that it can be used not only for the analysis of changes in industrial connections but also for enterprises of different sizes and sectors of the economy.

7 Conclusions

Thus, the above made it possible to determine that using a structural approach to model the socio-economic system accomplishes many challenges related to the diagnosis and management of the stability of such systems. Also, the use of a structural approach allows us to formalize and structure knowledge about the phenomena and processes that occur in the system and to predict them in the future. In combination with deterministic and stochastic approaches, the proposed approach allows to analyze the positive and negative changes in the external environment, which can affect the state and behavior of the socio-economic system, as well as to study the directions of development of the socio-economic system and to form on its basis the support system for management solutions.

References

1. Sergeeva, L. N., Bakurova, A. V., Vorontsov, V. V., & Zulfugarova, S. O. (2009). *Modeling the structure of viable socio-economic system*. CPU.
2. Sergeeva, L. N., & Bakurova, A. V. (2008). The concept of structural complexity of the socio-economic system. In *Management models in a market economy* (pp. 425–433).
3. Voronin, A. A., & Mishin, S. P. (2003). *Optimal hierarchical structures*. ICS RAS.
4. Leontiev, V. (1953). *Studies in the structure of the American economy*. Oxford University Press.

Features of Use of Simulation Modeling When Managing the Production and Economic Systems

Svitlana Sharova, Olena Androsova, Olesia Bezpartochna, and Victoriia Tsypko

Abstract One of the conditions for successful functioning of the modern company is the use of production systems (cs). Mainly their task is to maximize the reduction of time delays at various stages of production and the ability to adapt to changing internal and external conditions that affect production. For design, analysis and research of the armed forces use a virtual modeling. This is due to the fact that the experiments on the real facilities require large financial and labor costs. Simulation modeling enables to describe the behavior entirely using dynamic models. However, the increase in size of the studied the system increases the complexity of its modeling and general understanding of the internal interactions of various processes deteriorates, making changes requires a lot time-consuming. While the reuse of models or parts, in the case of modernization of production and analyze its current state, there is time-consuming task. The solution to these problems is the possibility of submitting complex systems using decomposition, separating system separate processes as a set of library components and automated supporting relevant data about the simulated objects. In addition to effective model building required its interaction with real objects to provide timely information about changes various parameters that will allow to predict in advance the possible consequences application of those or other actions.

Keywords Instrument making · AnyLogic · Simulation modeling · Production system · Computer experiment · Model · Complex system · Modeling system

S. Sharova · O. Androsova
Department of Finance, Banking and Insurance, Zaporizhzhya National Technical University, Zaporizhia, Ukraine

O. Bezpartochna (✉)
University of Finance, Business and Entrepreneurship, Sofia, Bulgaria

V. Tsypko
Department of Tourism, National Transport University, Kyiv, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_26

1 Introduction

The development of computer technology became the main guarantee for the emergence of a new scientific direction in the study of complex processes of functioning and development of economic systems—simulation modeling. When managing the production and economic systems (PES), it is very often necessary to deal with random factors: market demand, equipment failure, need for financial resources, etc. In order to evaluate the effects of PES under such conditions, it is sometimes possible to conduct an artificial experiment, during which, instead of conducting complex tests with real objects, special experiments are performed on mathematical models.

Such experiments are called simulation modeling, and their basis is the simulation model. Simulation models are a special class of mathematical models and differ from the analytical ones in that PES play a major role in their implementation. The power of computer facilities with modern software makes it possible to organize effectively the dialog mode of work within the framework of imitation systems.

The idea behind the simulation method is that instead of describing analytically the relationship between input and output states and metrics, an algorithm is created that maps the sequence of processes in the middle of the study object and then simulates the behavior of the object on a PC. Simulation modeling is a common variety of analog modeling, implemented with the help of a set of mathematical tools, especially imitating software products and technologies, which by means of analog processes allows us to conduct a purposeful study of the structure and functions of a real process in a dialog mode, as well as to optimize some of it parameters.

2 Use of Simulation Modeling

A simulation system is a set of simulation models that study internal and external functional descriptions of the phenomena and processes of PES. A simulation model is a calculable procedure that formally describes the object of the study and simulates its behavior. In its construction, there is no need to simplify the description of phenomenon or process, sometimes throwing away even significant particles in order to introduce it into a model suitable for the use of certain known mathematical methods. Simulation modeling is characterized by the imitation of the elementary phenomena that form the study process basis, with the preservation of their logical structure, the sequence of occurrence in time, and the nature and composition of information about the process state.

The model in this form is logical-mathematical (algorithmic). In relation to classification signs, the simulation models are divided into: static and dynamic; determined and stochastic; discrete and continuous. Every class of tasks proposes certain requirements to the simulation model. For example, in static simulation, the calculation is repeated several times for different conditions of the experiment, that is, the behavior of the object in a definite short period of time. In dynamic simulation, the system behavior is designed during the protracted period of time. In stochastic simulation, random variables with known distribution law are introduced into the model, and they are absent in the deterministic one. The order of the simulation model construction and its research is similar to the analytical models. However, for simulation models, there are a number of features, so it is advisable to re-describe the basic stages of imitation.

1. Definition of the system is the establishment of permissible functions of specific parameters, limitations, and quantitative estimates of the research system effectiveness.
2. Formulation of model is a transition from the real description of the system to the certain logical (abstract) chart.
3. Formation of information base is the collection of normative and statistical data and their submission in an appropriate form.
4. Model translation is the description of the model in a language available to the software system.
5. Adequacy assessment means that the model-based results should be fairly accurate with respect to reality.
6. Strategic planning is planning of an experiment that should provide the information you need.
7. Tactical planning is the definition of how each series of simulations is to be performed, as provided in the experiment plan.
8. Experimentation is the process of performing a simulation to obtain the desired results and conduct a sensitivity analysis.
9. Interpretation is a construction of conclusions from data that is got by an imitation.
10. Implementation is the practical use of the model and the results of the simulation.

As you can see, special attention is paid to the stages of planning experiments with the model, because simulation on a personal computer (PC) is an experiment, analysis, and finding the optimal solutions of algorithmic models (simulation models belong to this class of models) using these or other methods of experimental optimization on a PC. The only difference between a simulation experiment and a real object experiment is that the simulation experiment is performed on a model of the real system and not with the system itself.

3 Simulation of Economic Processes

Simulation of economic processes is mainly used in two cases:

- For managing a complex business process in which a managed economic object simulation model is used as a tool in the outline of an adequate computer-generated management system.
- When conducting experiments with discrete–continuous models of complex economic objects to obtain and investigate their dynamics in risk situations.

It is possible to distinguish the following typical tasks that are solved by simulation tools when managing economic processes:

- Modeling of logistics processes for determination of time and cost parameters.
- Managing the process of the investment project implementation at different stages of its life cycle, taking into account the possible risks and tactics of allocating financial resources.
- Forecasting the financial results of PES for a specific period of time, etc. The following list of tasks is incomplete and covers only those examples of models that are the fully described in the literature.

A simulation system that creates models to solve these tasks should have the following properties:

- Possibility of application of simulation programs together with special economic–mathematical models and methods based on management theory.
- Be characterized by instrumental methods of conducting structural analysis of a complex economic process.
- Ability to act in dialog mode.

To simulate a process or phenomenon on a PC, it is necessary to transform their mathematical model into a special modeling algorithm, according to which information will be generated in the PC, and describe the elementary phenomena of the research process, taking into account their relationships and mutual influences. Some of the processed information is printed and used to determine the functional characteristics of the process to be obtained as a result of the simulation. The central component of the modeling algorithm, in fact, will be a simulation model—a formalized scheme of the process. A formalized process scheme is a formal description of the procedure for the operation of a complex object of study and allows to calculate the corresponding numerical values of the initial characteristics ϖ for arbitrarily specified values of the input factors of the model (variables, x; deterministic, a; random, y). Other models (Fig. 1) are external mathematical support for the imitation process. The input models allow you to specify certain values of input factors. Static models of deterministic inputs are elementary arrays of constant values that correspond to certain model parameters. Dynamic input models change the values of deterministic factors over time according to the law a (t).

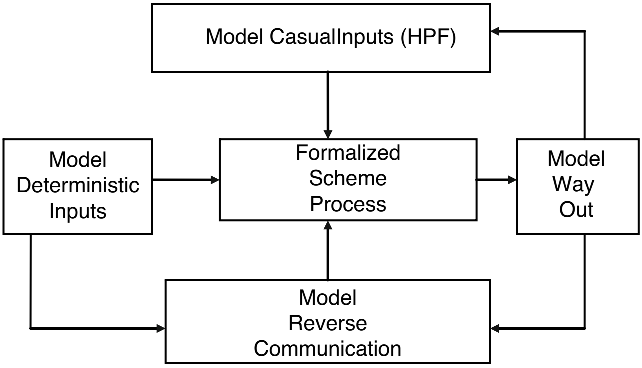


Fig. 1 Structure of a modeling algorithm of an optimization model with random factors

Random input models (random number generators) mimic the input of a random variable object with given distribution laws p (y).

The dynamic models of random inputs take into account that the laws of distribution of random variables are functions of time, that is, for each period of time, the form or characteristic of the law of distribution (mathematical expectation, variance, etc.) will be peculiar only to them. Due to the fact that the result obtained by playing a single implementation due to random factors cannot characterize the research process as a whole, a large number of such implementations have to be analyzed. Because only in this case, by the law of large numbers, the obtained estimates become statistically stable, then they can be accepted with certain accuracy for the estimates of the values sought.

The output model provides the accumulation, processing, and analysis of a set of random results. To do this, it is organized to perform multiple calculation of the values of the original characteristics at a constant value of factors a, x and different values of factors y (according to a given distribution law)—"cycle y." Therefore, the output model contains programs for tactical planning of the experiment on the PC—determining the procedure for each series of calculations that correspond to the specific values of a, x.

In addition, the outlined model solves the problem of processing the random values of the original characteristics, thereby eliminating the effects of random factors and entering the input of the feedback model. That is, the output model performs the procedure of reducing the stochastic problem to the deterministic one using the method of "result averaging." The feedback model enables, based on the analysis of the simulation results obtained, to change the value of the control variables and thus to implement the strategic planning function of the simulation experiment. Using the methods of optimal experiment planning theory, one of the functions of the feedback model is to present the simulation results analytically, to determine the response function equation.

In optimization, the output model based on the values of the output characteristics ω calculates the value of the objective function Z (ω) and, using the appropriate numerical optimization method, changes the values of the control variables to select the values that are the most favorable from the view point of the objective function. The procedure of the formalized scheme development consists of the structural decomposition of the object on the model, the choice of a mathematical scheme for a formalized description of the operation of each module, the formation of input and output information base for each module, and the development of a block diagram of the model control to display the interaction of individual modules. In the process of restructuring, the complex object is divided into relatively autonomous parts—modules. Further the links between them are fixed. It is expedient to structure the object in modeling so that the solution of a complex problem is reduced to a set of many simpler modules with the help of mathematical tools. The resulting structure scheme of the object can be further adjusted in terms of experience or convenience of information support of the algorithm. Further, for each constructed module, the choice of the mathematical description method, on which the corresponding model of operation will be constructed, is made. The next step is to formulate the necessary support according to the available methods of mathematical description. The integration of modules into a single model is performed on the basis of the operation models and information and procedural models described at the stage of the problem statement. This is solved by constructing a guided block diagram of a model that organizes the sequence of tasks to be solved. After building a managed block diagram, the content of the individual modules is detailed, specifying not only what should be done but also how it should be executed. More detailed and unambiguous instructions are given in relation to implementation of those or other procedures. Thus, the constructed formalized scheme contains the guided block diagram of the process, a description of each module, the rules for transferring one to another module, and the final list of desired values and functional characteristics. The formalized process diagram is the basis for further formalization of the simulation model and the construction of software for solution on PC.

4 Use Monte Carlo Method for Managing the Production and Economic Systems

The ability to model random variables and processes can in some way be used to simulate some real economic phenomena and production situations. In this case, the observation of a small number of occurrences of random variables is unlikely to be of benefit to us, and the observation of a large number of them makes it possible to draw correct conclusions about their average characteristics. This approach underlies the Monte Carlo method, which uses different boundary ratios of probability theory—the laws of large numbers and boundary theorems. The Monte Carlo statistical test is a simulation in the absence of any rules of behavior. The Monte Carlo sampling is the basic principle of simulation of systems containing stochastic or probabilistic elements.

According to the Monte Carlo method, the analyst can simulate and investigate the operation of a large number of complex systems or the behavior of the system relatively to the time horizon from a distant perspective. Without proving, we formulate a theorem that forms the fundamental basis of using the Monte Carlo method to calculate the mathematical expectation of a random variable on the basis of independent tests. In order for the arithmetic mean of independent tests of a random variable ξ to coincide with probability 1 before its mathematical expectation, it is necessary and sufficient that this mathematical expectation exists. The Monte Carlo method is a numerical method that makes it possible to model future values of a variable using its behavior over a time horizon. In various problems occurring in the construction and operation of complex economic systems, the values can be used which meanings are determined randomly. The examples of such quantities are: random moments of time during which orders are received for the firm; loading of production sites or other elements of PES; external actions (requirements or changes in laws, penalties, etc.); payment for a loan; receipts from the customer, etc.

The appropriate variables may be a number, a set of numbers, a vector, or a function. One variant of the Monte Carlo method for numerically solving problems with random variables is the statistical test method, which is to model random events. The Monte Carlo method is based on statistical tests and, by its nature, is extreme and can be used to solve completely deterministic problems such as inverting matrices, solving differential equations in partial derivatives, finding extrema, and numerical integration. In Monte Carlo calculations, statistical results are obtained by retesting. The probability that these results differ from the true ones by no more than the specified value is a function of the number of tests.

The basis of calculations according to the method is a random selection of numbers from a given probability distribution. For practical calculations, these numbers are taken from tables or obtained by certain operations, the results of which are pseudorandom numbers with the same properties as the numbers obtained by random sampling. There are many computational algorithms that make it possible to obtain some sequences of pseudorandom numbers. The use of the Monte Carlo method can have a significant effect in modeling the development of processes, the ordinary observation of which is undesirable or impossible, and other mathematical methods of application to these processes are either undeveloped or unsuitable due to the large number of assumptions that can lead to serious errors or incorrect conclusions. Therefore, it is necessary not only to observe the development of processes in the desired directions but also to evaluate the hypotheses regarding undesirable situations.

There are various methods of testing statistical hypotheses, among which the following criteria are often used in practice: 2 χ agreement, Cramer von Mises, Kolmogorov–Smirnova. The 2 χ criterion is preferred if the sample volumes N for which analysis is performed are large enough ($N > 100$). However, it is sometimes difficult or completely impossible to find 100 identical processes that develop with different inputs to analyze economic situations.

Kramer von Mises criterion produces good results with small sample sizes (*N* criterion). The Monte Carlo simulation process can be divided into the following stages:

1. Determining the stochastic nature of the input variable. This makes it possible to select the distribution that is required to perform the simulation. Most Monte Carlo modeling software has a probability distribution menu. They also have the ability to build a probability distribution based on the analyst's own conclusions, since modern computers have random number generators that allow you to get even numbers from 0 to 1.
2. Simulate the movement of input variables by repeatedly generating random numbers that are adjusted to such a calculation that they have the same probability distribution as the main variable. It means converting computer-generated random numbers into random variables for modeling. Adjusted random variables are input variables.
3. Modeling is the combination of input variables according to a system logic that describes how the input variables are linked and how the output variables are formed. Based on multiple random number generation, we get the future values of the variable we are looking for.
4. The multiple reiteration of the process makes it possible to find the average of the values obtained. This average is the future (expected) value of the simulated variable. Further, to determine the present value of the simulated variable for the future value is discounted at the appropriate discount rate.
5. Applying scattering techniques or variance reduction techniques to improve the accuracy of simulation results.

References

1. Moulder, J., & Elmagrabi, S. (1981). *Operations research* (Vol. 1, p. 712). Mir.
2. Reisberg, B. A. (1992). *Entrepreneurship and risk* (p. 62). Knowledge.
3. Rosen, V. V. (2002). *Mathematical models of decision making in economics* (p. 288). KD "University".
4. Takha, H. (1985). *Introduction to operations research* (p. 496). Mir.
5. Ferster, E., & Renz, B. (1983). *Methods of correlation and regression analysis* (p. 302). Finance and Statistics.
6. Fishburn, P. (1978). *The theory of utility for decision making* (p. 352). Science.
7. Khristianovsky, V. V., Polshkov, Y. N., & Shcherbina, V. P. (1999). *Economic risk and methods of its measurement* (p. 250). Donetsk State University.
8. Shannon, R. E. (1978). *Imitation modeling of systems: science and art* (p. 420). Mir.
9. Gordon, G. A. (1961). General purpose systems simulation program. In *Proceedings of the December 12–14, 1961, Eastern joint computer conference: computers—Key to total systems control* (pp. 87–104).

10. Schriber, T. (1987). The nature and role of simulation in the design of manufacturing system. In *Simulation in computer integrated manufacturing* (pp. 5–18). The Society for Computer Simulation.
11. Schulze, T., Lorenz, P., & Hinz, V. (2000). *Simulation und visualisierung*. SCS.
12. Karpuhin, I. N. (2013). *Research and implementation of systems of discrete event simulation based on graph models* (p. 321).
13. Podkritov, D. I. (2012). *Development of fault-tolerant multi-agent means of simulation modeling of systems with discrete events* (p. 114). Novosibirsk.
14. Nutt, G. J. (1972). *The formulation and application of evaluation nets. Ph.D. dissertation*. University of Washington.
15. Engelfriet, J., & Rozenberg, G. (1998). Elementary net systems. In *Lectures on Petri nets I: Basic models, ser. Lecture notes in computer science* (Vol. 1491, pp. 12–121). Springer.
16. Rzecki, K. (2013). Using the evaluation nets modeling tool concept as an enhancement of the Petri Net tool. In *Proceedings of the Fedcsis* (pp. 1007–1012).
17. Efremov, A. A. (2006). *E-network reliability modeling series-parallel technical systems with recovery* (Vol. 7, pp. 97–101). TPU News.
18. Kazimir, V. V., & Gray, A. A. (2012). Models of verification of plans accident elimination at coal mines. In *Mathematical machines and systems* (Vol. 1, pp. 129–138). IPMS of NASU.
19. Harmon, P., & Wolf, C. *Business process modeling survey*. http://www.bptrends.com/bpt/wpcontent/surveys/Process_Modeling_Survey-Dec_11_FINAL.pdf
20. Borshchev, A.V. *Simulation modeling: State of the region for 2015, trends and forecast*. http://simulation.su/uploads/files/default/2015-immod-14-22.pdf
21. Malikov, R. F. (2013). *Workshop on simulation modeling complex systems in AnyLogic 6* (p. 296). Publishing House of the BGPU.

Technological Competitiveness Formation Policy, Economic Security, and Growth in the Context of Ukraine

Taras Vasyltsiv, Ruslan Lupak, Vitalii Boiko, and Marta Kunytska-Iliash

Abstract Conceptual provisions of technological competitiveness are generalized, and the advantages of its identification as a strategic goal of the state policy aimed at guaranteeing safety and economic growth are substantiated. Casual interrelations of technological development, competitiveness, and economic growth are identified. An integral index of innovation and technological development is calculated according to the actual values of the indicators of innovation and technological development of Ukraine, using the method of principal components, which is compared to the integral index of economic security of the country. The obtained results are compared to Ukraine's rating positions in the international technological competitiveness ratings. Casual interrelations and algorithms for technological development, competitiveness, and economic growth in the form of a set of successive steps within the framework of the state technological policy are identified.

Keywords Technological competitiveness · Security and economic growth · Innovation and technological policy

T. Vasyltsiv
State Enterprise "Dolishniy Institute of Regional Research of NAS of Ukraine", Lviv, Ukraine

R. Lupak (✉)
Lviv University of Trade and Economics, Lviv, Ukraine

V. Boiko
Lviv Polytechnic National University, Lviv, Ukraine

M. Kunytska-Iliash
Stepan Gzhytskyi National University of Veterinary Medicine and Biotechnologies Lviv, Lviv, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_27

1 Introduction

The concept of technological development was first applied in economic science relatively recently. To a large extent, this was due to the emergence of classification of technological processes, awareness of the role of technologies used, their structure and measures of novelty in the context of socio-economic development, economic growth of states and regions. At the same time, the role, features and relations, factors and consequences of stimulation of innovation activities, formation and efficient use of intellectual capital in the context of economic competitiveness, and ensuring economic growth of countries have been the subject of scientific research of the majority of scientific schools for over a century. Initially, it was about research activities, creation and implementation of innovations, new developments, intellectual creative activity, which results are implemented in inventions, industrial prototypes, utility models, work-improvement suggestion, etc. Moreover, in almost all stages of the evolution of scientific idea, in studies of the leading scientific schools, and even in the practice of the most undemocratic and non-liberal systems of the state structure, direct correlation and impacts have been identified between the implementation of R&D results and the increase of production efficiency, and, consequently, competitiveness and economic development.

Studies resulted in the assertion that technological competitiveness is a leading component of competitiveness of the national economy, which characterizes the modernity and progressiveness of the technologies used, the extent of their spread in economy and society, quality of the system of production, attraction and implementation, availability and accessibility of resource support (primarily, investment and financial, intellectual and personnel, technical and technological, educational, research, and information), effectiveness of the use of results of innovation, and technological activities.

Vertical and typological correlations and interdependences have been worked out as well: the results of the policy of ensuring technological competitiveness of economy are manifested and can be evaluated at the macroeconomic level; however, they require necessary innovation and technological changes and shifts at all levels of competitiveness management, in particular, of the basic ones: product, enterprise, industry, territory, and economy as a whole.

Basic elements and systemic characteristics of the state policy of technological competitiveness management are the interests of the state, objects, and subjects of control, their respective functions, methods and means of management, sources of financial and investment support. Despite achievement of general goal of the state policy, ensuring technological competitiveness of the economy, it is important to ensure the implementation of a system of regulatory functions, which include economic security, social and economic efficiency, spatial and structural development, promotion of innovation, ensuring competition, financial and resource, innovation and technological integration, allocation of centers for research and innovation activity, and distribution of results as prerequisite elements of the system of economic growth of the country.

2 Technological Competitiveness: Methodological Discourse

Interrelations between population size and growth rate, and demand for technology and technological development were simulated in M. Kremer's models [1]. In particular, the assumption provided for growth of economy is determined by endogenously accumulated knowledge; technological progress is a growing function of population size (i.e., with an increase in the population, more people make scientific discoveries and implement them, correspondingly, knowledge is increasingly accumulated, which ensures technological progress); technological progress and GDP growth lead to an increase in the population at a faster pace than the growth of production output. The empirical form of the model was reflected as in the following formula [2]:

$$Y = AL^{\alpha}T^{1-\alpha}, \tag{1}$$

where Y is an aggregate production; A is a level of use of technologies; L is population size; T is a factor of land resources; α is a level of factors contribution.

Surprisingly, but these ideas, as well as the model determined and empirically proven, have not only lost their relevance but also become more important (although being controversial and requiring search for additional factors of influence) in the current conditions of growing migration activity of the population, its depopulation (especially in the countries Europe with a high level of technological development), growing differentiations in living standard of the population, etc. Studies that addressed human capital as an integral feature of individual knowledge, abilities, qualifications and skills, became a separate "stream" in the field of R&D relationships, development of more advanced technologies, and economic growth. It is obvious that the development and commercialization of advanced technologies are closely correlated with the availability of trained, well-educated personnel in the country, characterized by readiness and ability to implement the corresponding tasks.

Sectors of education and science, as well as manufacturing business, which creates its own research and testing sites for the development of innovations, in particular technological ones, are responsible for the presence of such intellectual and personnel resources. The role of subjects of financial and credit and investment sectors of economy is also very important here. According to G. Becker, the formation of intelligent staffing for technological development requires investment in human capital in the following areas: (1) training and education, (2) skill enhancement, (3) preservation and improvement of health, (4) migration, and (5) information search [3]. According to the author, in the modern world, the composition and structure of these investments are changing and supplemented, also in the direction of corporate communication and leadership, as well as in the use (moreover, a combination of physical and artificial intelligence) of information and communication technologies; the ability to forecast and predict the situation, to work ahead of the market situation and social behavior becomes increasingly important [4].

R. Solow was one of the first to consider human capital as a factor of production along with physical capital and labor [5]. Later, in the model of Mankiw [6], Romer [7], and other scholars, human capital is used as a production factor, and the process of its accumulation is fully in line with physical capital. As a result, the production function of R. Harrod acquired the following form [8]:

$$Y_t = K_t^{\alpha} H_t^{\beta} [A_t L_t]^{1-\alpha-\beta}, \tag{2}$$

where Y is production output; K is physical capital; H is human capital; A is technological development; L is labor factor; α, β are the levels of factors contribution; t is a moment of time.

In this model, however, withdrawal of physical capital outside of human capital as well as failure to consider the aforementioned investments in human capital raises a question of appropriateness.

R. Lucas's studies take into account the educational sector, in particular, as a segment of economic system, which ensures the formation of human capital, taking into account a series of circumstances and factors, the most important among which are: (1) technical parameter as the basis of productivity of educational system, (2) share of costs for education in the overall labor time reserve, (3) average current level of human capital. As a result, the following model is proposed:

$$Y = bK^{\alpha} [(1-u)\,H]^{1-\alpha} h^{e}, \tag{3}$$

where Y is production output; b is productivity parameter of end product sector; K is physical capital; H is an overall volume of human capital; $1 - u$ is a share of working hours of each individual, and, correspondingly, society as a whole in the sector of end product; h is an external effect of average level of education upon production of end product; e is end production elasticity coefficient according to the average level of human capital; α is a level of factors contribution [9].

In the framework of the proposed model, R. Lucas simulates possible directions and parameters of the state's influence on human capital and, accordingly, on the volume of economic growth [10]. Two behavioral variants are considered. The first one involves active subsidization of educational sector by the state and, thus, obtaining a positive external effect. Almost similar effect can be achieved in the other way: through stimulation or subsidization of savings. Due to this, the state has alternative options in terms of stimulating the impact of education on human capital and, ultimately, on ensuring economic growth. However, it should be noted at once that an entire set of additional factors and conditions that are not foreseen in this model, but whose influence may be critically significant, should be taken into account. Although, the tendencies characterized have somewhat different properties in the environment of a perfect market and in economies with not yet fully formed market institutions, establishments, rules of conduct, which is important in the context of understanding these features in the formation of the future state policy of ensuring technological competitiveness, determining its most qualitative and effective mechanisms, tools and means of regulatory influence.

It should be noted that the results of the study, which analyzes impact on economic growth of not so much factors of human capital as innovations, have comparatively higher theoretical and applied importance, because, first, the existing human capital still needs to be rationally and effectively used and, second, the implemented innovations are in fact the final stage in the policy of state support for technological growth. Here, scholars mostly use the following theses: there is a connection between the increment rates of production and the scope of implementation of technological advance; the increment rates depend on the behavioral factors and the quality of the state policy; technological advance includes expansion of the types of manufactured products, each of which is identified with a certain new technology; the main source of R&D financing is additional profit of the manufacturer, and the latter finances the research in order to receive such a profit. The relevant dependence is characterized by the following description:

$$Y = K^{e} A L^{1-e} N^{1-e}, \tag{4}$$

where Y is production output; K is capital; A is technological advance; e is production elasticity; L is a scope of labor; N is a number of available types of manufactured goods (correspondingly, the technologies).

Scientists highlighted the rights to intellectual property items as the relevant assets and a monopoly resource (though for a limited period of time) of the owner. This factor was taken into account in the models of economic growth under the pretext of the patent price as a sum of the flow of future discounted profit.

Innovative and technological development as a basis for the formation of technological competitiveness of a state is characterized by a number of components, levels and indicators. That is why, for the purpose of its analysis, it is necessary to use integral approaches that would describe its state and dynamics, its changes in a complex, and also provide a possibility to give a comparative characteristics of the innovative and technological development in the economy in general and according to the types of economic activity, economic levels, and stages of the reproduction processes.

In order to investigate the level of the innovative and technological development as a system, it is more appropriate to use the methodology suggested by the authors, which eradicates most of the detected drawbacks and is characterized by an extended structure of indicators, the use of multiplicative (nonlinear) form of the integral index, norming the indicators and their threshold values. This provides adequate diagnostics with the possibility to compare with the integral threshold (optimal) values.

The system of the indicators of the innovative and technological development in Ukraine was formed using the structural and functional model of the innovative and technological development (ITD):

$$\mathrm{ITD} = f\begin{pmatrix} \mathrm{LE}\uparrow \\ \mathrm{FIA}\uparrow \\ \mathrm{SS}\uparrow \\ \mathrm{IE}\uparrow \\ \mathrm{IAE}\uparrow \\ \mathrm{IPE}\uparrow \\ \mathrm{RIP}\uparrow \end{pmatrix}, \tag{5}$$

where LE is level of expenses on R&D; FIA is level of financing innovative activity; SS is number of professional engaged in R&D; IE is number of innovatively active industries; IAE is number of industries that have implemented innovations; IPE is number of industries that have sold innovative goods; RIP is amount of innovative industrial goods sold.

According to the next stage of the determined by us academic and methodological approach to the systemic evaluation of the level of innovative and technological development in Ukraine, dynamic rows of indicators can be calculated after norming (according to the formula (6)), which enables us to identify positive and negative tendencies of the innovative and technological development in Ukraine:

$$z_i = \begin{cases} y_i/y_{\max}, & \text{if } y_i \in S; \\ y_{\min}/y_i, & \text{if } y_i \in D. \end{cases} \tag{6}$$

where z_i is normed values of i- indicators of the innovative and technological development; y_i is actual values of indicators; $y_{\max}$, $y_{\min}$ are maximum and minimum of the actual values of indicators; S is multiplicity of the values of indicators which are stimulators; D is multiplicity of values of indicators which are de-stimulators.

The important information basis to implement controlled influence on the state of the innovative and technological development in Ukraine is defining the major coefficients with the help of the method of the main components and appropriate software, which enables us to determine the contribution of each indicator to the integral index and is a necessary basis for developing priority impact measures.

Using this approach enables us to determine the dynamics of the integral index of the innovative and technological development in Ukraine taking into account the nonlinear character of economic processes according to the formula:

$$\mathrm{ITD} = \prod_{i=1}^{n} z_{it}^{a_i}, \quad \prod_{i=1}^{n} a_i = 1, \quad a_i \geq 0, \tag{7}$$

where z_{it} is normed values of i-indicators of the innovative and technological development during t-time period; a_i is major coefficients of i-indicators of the innovative and technological development; n is number of indicators of the innovative and technological development.

The important aspect of the methodology of the analysis is using the method of main components to calculate major coefficients of indicators and also to calculate normed values of indicators during the analyzed time period and to determine integral coefficients of the indicator of the innovative and technological development in our country.

Integral values of the innovative and technological development in Ukraine have been obtained and compared with the integral values of the economic security of the country and integral values of the investment and innovative security, which enables us to generalize conclusions and determine relations and impacts between the innovative and technological development, investment and innovative security and economic security of the country.

3 Indicators and State of Innovation and Technological Development, Its Influence on Economic Security of Ukraine: Modern Empirics

Table 1 shows calculated standardized values of indicators, their weighing coefficients, and the integral index of innovation and technological development of Ukraine in 2005 and 2010–2017. First, summarizing the analysis, we obtained the basis for the conclusion about higher level of influence (on the integral value of the state of innovation and technological development of the country) of such indicators as share of enterprises that implemented innovations, among total number of industrial enterprises (weighing coefficient made up to 0.179), share of enterprises engaged in innovation activities, among total number of industrial enterprises (0.176), the proportion of sales of innovative products in the total volume of sales of industrial products (0.176), and the level of spending on R&D in GDP of the country (0.17). Exactly these factors have the greatest impact on the increase in innovation and technological activities in Ukraine and, accordingly, are required to be taken into account, and need special influence in the formation and implementation of the state policy in the field of technological competitiveness.

Second, in 2005–2017 there was a decline in the integral index of innovation and technological development in Ukraine. If in 2005 the value of this indicator was 0.166, then by 2017 it decreased to 0.116 (by 0.5). Moreover, the fact that in all years of the period under study, values of the integral index of innovation and technological development of the Ukrainian economy was characterized as low ones and was negative as well.

This was indicated and led to a decrease in the corresponding index in Fig. 1. Although, during the period under study, the situation with the state of economic security of Ukraine was unstable, yet its values were much higher, being in the range of 0.44–0.49.

On the other hand, this is an indication of the fact that the status of innovation and technological development in the country is a significant destructive factor in the

Table 1 Standardized values of indicators and integral index of innovation and technological development of Ukraine in 2005, 2010–2017 (Source: [11])

Indices	Standardized values according to the years									Weighing coefficients
	2005	2010	2011	2012	2013	2014	2015	2016	2017	
1. Level of spending on R&D, % of GDP	0.234	0.166	0.146	0.150	0.154	0.132	0.124	0.112	0.015	0.170
2. Level of innovation activity financing, % of GDP	0.374	0.211	0.311	0.231	0.189	0.140	0.197	0.189	0.139	0.138
3. Share of professionals, performing R&D work, persons per 1000 of persons employed, %	0.034	0.029	0.028	0.027	0.025	0.033	0.033	0.033	0.021	0.065
4. Share of enterprises, engaged in innovation activities, among total number of industrial enterprises, %	0.148	0.172	0.203	0.218	0.210	0.201	0.215	0.230	0.144	0.176
5. Share of enterprises that implemented innovations, among total number of industrial enterprises, %	0.124	0.176	0.197	0.209	0.198	0.186	0.233	0.262	0.149	0.179
6. Share of enterprises, selling innovation products, among total number of industrial enterprises, %	0.185	0.165	0.183	0.187	0.184	0.164	0.217	0.225	0.137	0.097
7. Proportion of sales of innovative products in the total volume of sales of industrial products, %	0.163	0.095	0.095	0.083	0.083	0.063	0.035	0.028	0.012	0.176
Integral values of innovation and technological development	**0.166**	**0.141**	**0.154**	**0.148**	**0.141**	**0.125**	**0.127**	**0.122**	**0.116**	–

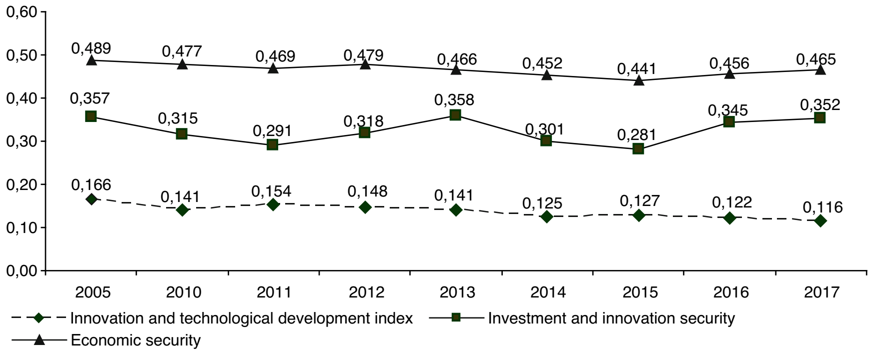


Fig. 1 Integral indices of innovation and technological development, investment, and innovation and economic security of Ukraine in 2005, 2010–2017. (Source: [11])

context of formation of economic security of the state. Moreover, in the information presented in the figure, higher values of the integral index of investment and innovation security of the country can be observed in comparison with the values of the index of innovation and technological development, which is an additional argument for the presence of significant problems in the field of development and implementation of modern technologies with the consequence of insufficient technological competitiveness of the national economy of Ukraine.

4 Consequences of the Status and Policy of Technological and Innovation Development for the International Competitiveness of Ukraine

These findings are confirmed by perhaps the lowest position of the Ukrainian economy in the Consolidated Innovation Index of the European Innovation Scoreboard (Fig. 2). In particular, the index of Ukraine made up to 22.9, which was the lowest value among the European countries that were analyzed. Difference of the index for Ukraine from the average index for the EU Member States was 4.5 times.

Ukraine (along with Romania, Macedonia, and Bulgaria) has fallen into a group of outsiders, called "modest innovators." Therefore, low innovation and technological activity, and in the end, the technological competitiveness of the Ukrainian economy are confirmed by recognized international ratings.

This resulted in the deterioration (preservation of low) of rating positions of Ukraine in the reputable international competitiveness indices (Fig. 3). Thus, according to the Global Competitiveness Index, Ukraine's economy in 2018 was ranked 81st place out of 137 economies in the world. Value of the analyzed index for Ukraine under the World Competitiveness Ranking was even worse: 59th position out of 63 countries.

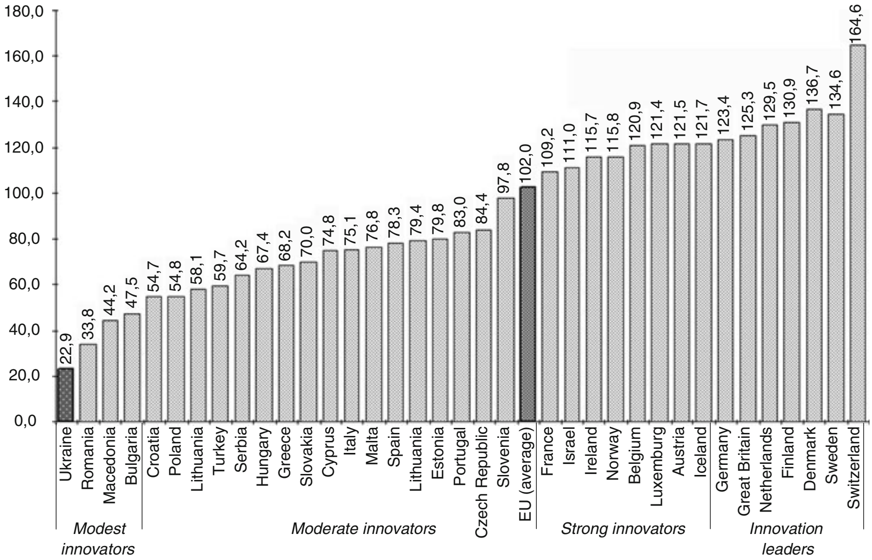


Fig. 2 Consolidated Innovation Index of the European Innovation Scoreboard at the beginning of 2017. (Source: [12])

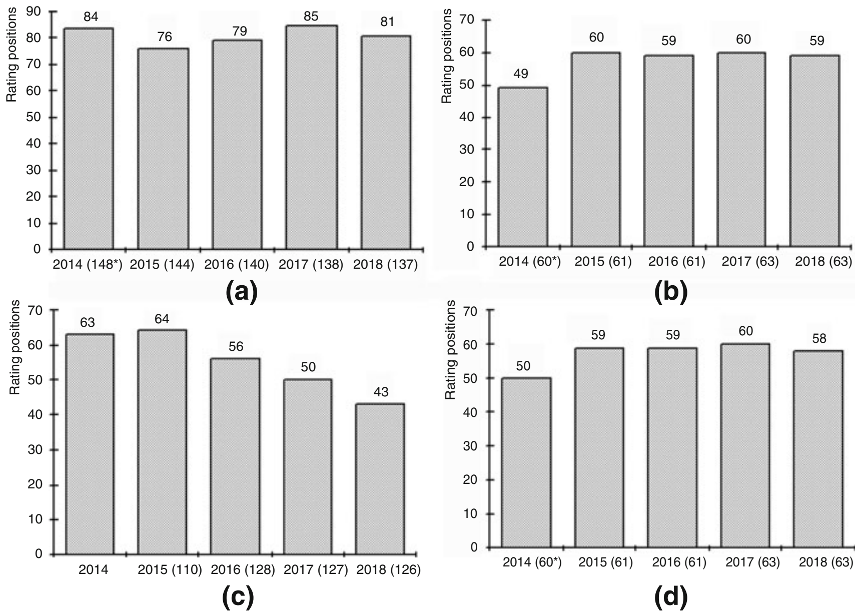


Fig. 3 Indices of rating assessment of competitiveness of Ukraine in 2014–2018. (**a**) Global Competitiveness Index. (**b**) World Competitiveness Ranking. (**c**) Global Innovation Index. (**d**) Digital Competitiveness Ranking. (Source: [13–15])

Positions of Ukraine in the Digital Competitiveness Ranking remain low: the 58th place out of 63 countries of the world in 2018, which is an evidence of underutilization of the potential of information technologies sector of the state in terms of ensuring the technological competitiveness of its economy.

It should be noted, though, that in 2014–2018, Ukraine improved its position in the Global Innovation Index. If in 2014 Ukraine occupied 63rd rating position out of 111 countries, then in 2018 it was 43rd out of 126 economies of the world. However, it is worth saying that these changes were not achieved directly due to technological innovations, but due to improvement of co-innovation (non-technological) factors.

5 System of Vectors and Conditions for Strengthening the Technological Competitiveness of the Country and Approaches to Security and Economic Growth

Despite limited scope of innovation and technological activity, low efficiency in the field of development and commercialization of modern technologies and insufficient technological competitiveness of the national economy, sufficient potential of scientific and technological and innovation development is maintained in Ukraine, which implementation is capable of accelerating the structural and technological modernization of the national economy and increasing the international competitiveness of the country. Ukraine is among the eight countries possessing necessary scientific and technological potential for the creation of aerospace technology, the ten worldwide leading shipbuilding countries, the nine largest exporters of weapons, and the leader in Europe in the field of IT outsourcing. Use of this potential and other opportunities requires activation of a well-balanced and effective state policy aimed at stimulating the increase in the scope of research and development and practical applicability of their results, increasing the amount of financing of scientific and technical and innovation activities, improving the intellectual and personnel provision of innovation activities, raising the level of innovation activity of the real sector of economy enterprises, improvement of interregional and intersectoral asymmetry of innovation development [16].

Institutional prerequisites for strengthening technological competitiveness of the Ukrainian economy lies in the enforcement of consistency and efficiency of statutes and regulations governing innovation and technological activities, increase in the efficiency of the state strategic planning of the national economy development on an innovative basis, structural reorganization and improvement of financing of research and development and innovation activity, distribution of expenses on performance of scientific and research and development works, allocation of investments on priorities of innovation development of economy, improvement of the system of intellectual property rights protection and competitive environment in the domestic market in general and in the domestic market of innovations and technologies in particular [17].

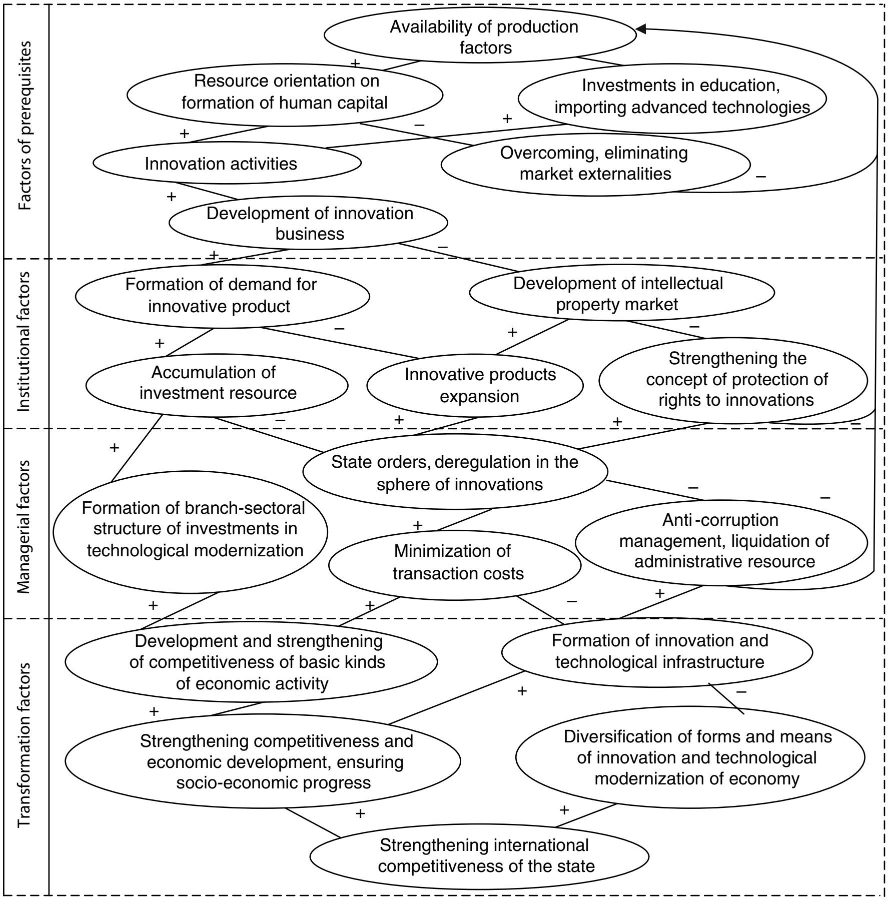


Fig. 4 Casual interrelation algorithms for technological development, competitiveness of the economy and economic growth. (Source: own)

Unambiguous empirical relationships between the factors described above and economic growth make it clear that the state, observing the “rules of the game” in the economy and thus fulfilling one of the most important functions, undoubtedly, contributes to a long-term economic growth. The degree of quality of the government apparatus, absence of corruption, administrative hindrances and barriers, and the ability to fully comply with all statutory regulations and provisions have a clear direct relationship with the competitiveness of the national economy and further economic growth. Summarizing the above facts, the basic casual interrelations and interdependencies in the field of the state policy of ensuring the technological competitiveness of economy are shown in Fig. 4.

It seems appropriate to point out that their identification and understanding, vision of their determinant factors, taking into account during formation of the state policy provisions in the field under consideration, allow for ensuring higher systemic quality and efficiency of the state regulation of technological competitiveness of economy.

Figure 4 presents to a greater extent the conditions and factors that ensure the growth of innovation activity and increase in the efficiency of the processes of creation and implementation of advanced technologies. Instead, some tendencies and interrelations between the successes of state regulation in the field of ensuring technological competitiveness and social and economic development of the country seem to have gained little attention.

6 Conclusions and Vision of Systemic Aspects of the State Policy

Casual interrelation algorithms of the technological development, competitiveness, and economic growth are a set of consistent steps within the framework of the state technological policy that enhance action of factors in the following groups: (1) prerequisites for technological competitiveness, (2) institutional environment, (3) management system, (4) transformation into the final result, and eventually provide an effect in the form of the country's international competitiveness. In the process of ensuring technological competitiveness of economy, a number of systemic economic (activation of innovation activity of market participants and strengthening of their competitiveness, capitalization of economy and efficient use of investments, provision of structural modernization of the national economy, diversification of activities, increase of GRP and GNP, modernization of technical and technological base of the real sector, input intensity reduction, productivity and efficiency growth, improvement of the system of national producer and its products protection, development of domestic market, import substitution, increase of government revenues, improvement of intersectoral innovation and technological cooperation), and inclusive (improvement of living standards, development of human capital, growth of scientific and educational potential of society, promotion of a more rational distribution of well-being and services and income of population, equalization of social imbalances, strengthening of national uniqueness and identity, counteracting labor migration, expansion of social opportunities of the state, development of consumer market and improvement of the quality of goods and services, lowering of the "burden" on physical capital and, consequently, preservation of raw materials resources) effects are achieved, which need to be taken into account during planning and control of the state technological policy efficiency.

References

1. Kremer, M. (1993). Population growth and technological change: One million B. C. to 1990. *Quarterly Journal of Economics, 108*, 681–716.
2. Kremer, M., Rao, G., & Schilbach, F. (2018). *Behavioral development economics*. Harvard University.
3. Becker, G. (1964). *Human capital*. Oxford University Press.
4. Becker, G. S., Kominers, S. D., Murphy, K. M., & Spenkuch, J. L. (2018). A theory of intergenerational mobility. *Journal of Political Economy, 126*, 7–25.
5. Solow, R. (1970). *Growth theory: An exposition*. Oxford University Press.
6. Mankiw, N. (1991). *Principles of microeconomics*. South-Western Cengage Learning.
7. Romer, P. (1990). Endogenous technological change. *Journal of Political Economy, 98*(5), 71–102.
8. Harrod, R. (1948). *Towards a dynamic economics: Some recent developments of economic theory and their application to policy*. Macmillan.
9. Lucas, R. (1988). On the mechanics of economic development. *Journal of Monetary Economics, 22*, 3–42.
10. Lucas, R. (2015). Human capital and growth. *American Economic Review, 105*(5), 85–88.
11. National Statistics Office of Ukraine. (2018). *Ukrainian statistical yearbook 2018*. National Statistics Office of Ukraine.
12. European Innovation Scoreboard. (2017). https://ec.europa.eu/docsroom/documents/24829
13. World Economic Forum. (2019). https://www.weforum.org
14. International Institute for Management Development. (2019). https://www.imd.org
15. World Intellectual Property Organization. (2019). https://www.wipo.int
16. Vasyltsiv, T., & Lupak, R. (2016). Strategic priorities of state structural policy of import substitution in Ukraine. *Actual Problems of Economics, 1*(175), 52–60.
17. Panukhnyk, O., Popadynets, N., & Fedotova, Y. (2019). Analysis and modeling of factor determinants of food provision at consumer market of Ukraine. *Global Journal of Environmental Science and Management, 5*(SI), 215–226.

Disruptive Innovation in Mining Industry 4.0

Sara Qassimi and El Hassan Abdelwahed

Abstract The mining industry is facing global challenges with undergoing significant market-changing demand and competitiveness. It has become imperative for mining companies to keep up with the real-time visibility on production quality and control, cycle times, machine status, and other important operational variables. In order to reach smart manufacturing, mining companies must seize the opportunity of Industry 4.0 to leverage the advancement of information technology. This chapter reviews the current research studies about the smart manufacturing in Mining Industry 4.0 that stands on the intersection of the emerging information technologies (IT 4.0), mining industry, and innovation. The review discusses and analyzes a plethora of innovative technologies that assist miners in their roles such as the Internet of Things, cyber-physical systems, digital twins, and so forth. These disruptive technologies address the issue of cyber-physical integration (CPI). They are the pillars of smart manufacturing in the Mining Industry 4.0. The review provides insights about the next horizon of the use of disruptive technologies in the mining industry toward data-driven smart manufacturing.

Keywords Innovation · Disruptive technologies · Cyber-physical integration · Mining Industry 4.0 · Smart manufacturing

1 Introduction

A seismic shift in the mining industry is occurring with the undergoing significant market challenges such as the global changing demand and competitiveness, market volatility, searching of reserves in new locations, continued declining

S. Qassimi (✉)
L2IS Laboratory, FSTG, Cadi Ayyad University, Marrakesh, Morocco
e-mail: sara.qassimi@uca.ac.ma

El H. Abdelwahed
LISI Laboratory, FSSM, Cadi Ayyad University, Marrakesh, Morocco
e-mail: abdelwahed@uca.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_28

commodity prices, etc. [1]. Mining factories have to be equipped with real-time visibility on production quality and control, cycle times, machine status, and other important operational variables to enable manager's instantaneous visibility for optimum decision-making [2]. Thus, it is achieved by filling the gap between the company/enterprise-level system (ERP, CRM, etc.) and lower level mining systems (shop floor systems). In this regard, mining companies must seize the opportunity of Industry 4.0 to leverage digital tools and technologies to fragment the real-time flow of information between the shop floor systems and the enterprise-level systems to deliver optimum operations and reach new levels of performance across the mining value chain.

The Mine 4.0 "Digital Mining" is the implementation of Industry 4.0 technologies that incorporate the digital transformation using information technologies (IT), while the smart manufacturing in Mining 4.0 stands on the intersection of the emerging information technologies (IT 4.0), mining industry, and innovation (see Fig. 1). It leverages the innovation in the mining industry using IT 4.0, the so-called disruptive technologies, such as the Internet of Things (IoT), big data, cloud, digital twin (DT), cyber-physical system (CPS), and advanced analytics, and so forth. These disruptive technologies connect the mining value chain from sensor to board management system, therefore from shop floor systems to enterprise-level systems to market. Once data collected from sensors is consolidated on the cloud-based big data platform, powerful analytics and machine learning algorithms can be launched to provide real-time operational analysis visualized on smart devices and operation boardrooms and desktops.

The current overview of the current digital situation of factories [3] claims that one of the most typical challenges is how to apply these new IT 4.0 to cope with the trend of smart manufacturing to achieve therein smart production operation and management. Thus, different strategies have occurred like the cyber-physical system-based manufacturing, Industrial Internet, Made in China 2025 [4]. Even with their different backgrounds, they have a common purpose of the cyber-physical integration (CPI) to achieve smart manufacturing [5]. The CPI aims to realize the interconnection and interoperability between physical and cyber worlds for moving digital factories forward toward smart manufacturing [3].

Therefore, the objective of this chapter is to review and analyze the academic progress that addresses the issue of CPI in manufacturing, especially in the mining industry. The purpose of reviewing is to answer the research questions:

– What is the fourth industrial revolution Industry 4.0?
– What are the features of Mining 4.0?
– What are the current research efforts to achieve smart manufacturing?
– What are the promising incomes of using disruptive technologies (IT 4.0) in the mining industry?

The rest of this chapter is organized as follows. Section 2 presents the industrial revolutions, the principles of Industry 4.0, and the features of smart manufacturing. Section 3 illustrates and discusses an overview of the pillars of Mining 4.0 and provides insights related to what is being researched corresponding to each disruptive technology in Industry 4.0. Section 4 presents a discussion of the next horizon of disruptive technologies in the mining industry in terms of promising incomes and points out future recommendations and perspectives. The conclusion of this chapter is delineated in Sect. 5.

2 Industrial Revolutions

Humankind has perfected its industry throughout history during four industrial revolutions [6]. The industry has evolved over the past centuries as follows (see Fig. 2). The first industrial revolution "Industry 1.0" started in the eighteenth century with the invention of mechanical production and manufacturing process using water power and the steam engine (e.g., steam locomotive). At the beginning of the twentieth century, the era of industrialization using oil and electricity gave rise to the second industrial revolution "Industry 2.0". The third industrial revolution "Industry 3.0" started in 1960 with the electronics and IT-based manufacturing to automate production, the so-called digital revolution. The fourth industrial revolution has occurred since the middle of the last century. The "Industry 4.0," a term coined by Klaus Schwab [7], blurs the lines between the digital domains and offline reality

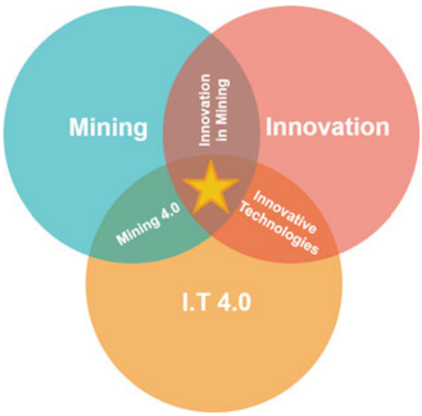


Fig. 1 Disruptive innovation in mining industry using IT 4.0

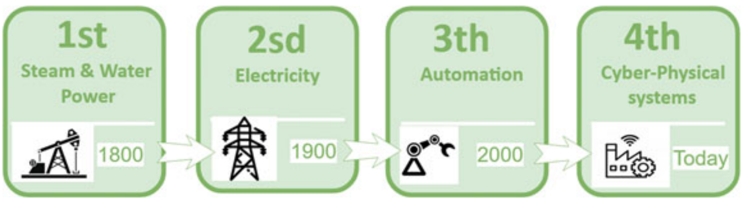


Fig. 2 The four industrial revolutions

with the use of innovative technologies. Industry 4.0 today's transformation does not represent a prolongation of Industry 3.0 but rather the arrival of innovative industrial technologies. The term Industry 4.0 is usually used as a synonym for the digitization of traditional manufacturing industries, which is commonly known as intelligent manufacturing standing in the intersection of manufacturing and artificial intelligence (AI) [8]. However, it goes further toward smart manufacturing which evolves with the emergence of the new generation of information technologies (IT 4.0), called disruptive technologies, fusing the physical and digital worlds. Conventional factories are implementing the new paradigm of Industry 4.0 for an ever-increasing focus on smart manufacturing transformation to achieve a higher level of operational productivity and efficiency [9].

3 Pillars of Mining 4.0: Disruptive Technologies

The disruptive technologies-enabled mining factories hold the promise of smart manufacturing, such as increasing their digitization, virtualization, production efficiency, data integration, computational capacities, and remote and autonomous operations to achieve therein their operational improvements in value creation. The McKinsey Digital report [10] claimed that leaders across industries are considering data as the core driver.

In the following, we define the disruptive technologies having a significant impact on manufacturing for rising innovation in the industrial domain and so for the mining industry. Following McKinsey's report [10], the disruptive technologies are clustered into four main pillars of the Mining Industry 4.0 (see Fig. 3).

3.1 Data, Computational Power, and Connectivity

The IoT technologies enable to connect all physical industrial equipments and objects by sensing their environment and communicating autonomously among each other and therein produce a huge amount of structured, semi-structured, and

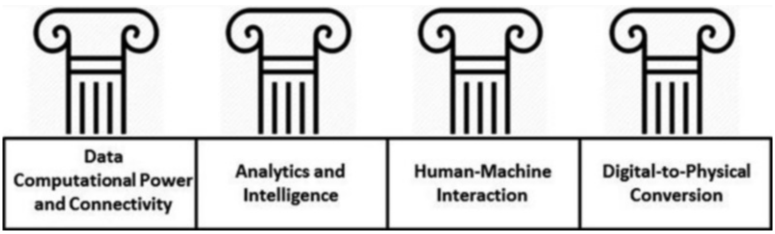


Fig. 3 Pillars of Mining 4.0 using disruptive technologies

unstructured data stored on big data platform and analytics software. The foundation for gathering and processing data is given by cloud computing providing services through a network of remote servers.

Physical Data Integration is the first stage of smart and cognitive manufacturing processing. The use of a flexible information model will integrate heterogeneous and distributed product information for IoT-based applications to provide further smart interactions [11]. To enable large-scale deployment of industrial services and extended applications, researchers focus on including semantic integration of massive data [12]. Big data is stimulated by emerging technologies that help to create smart factories where a huge amount of data is in real-time collected and analyzed [13]. Its main objective is to achieve inexpensive and fault-free processes with high quality and performance [14]. However, the collected real-time huge amount of data has to be analyzed and processed to provide meaningful value utilized for manufacturing decisions. Data science and data analytics are techniques that should be employed to analyze massive generated data coming from IoT applications and other existing information technology systems [15].

Internet of Things (IoT) enables the conversion of production resources into smart manufacturing object (SMO). These SMOs can sense their environments, interconnect, and interact with each other caring out automatic and adaptive manufacturing logic [16]. They are embedded with electronic sensors and other digital devices creating a network to collect and exchange data [17]. The IoT-enabled industry offers the connectivity of physical objects, services, and devices that enable an object-to-object communication and data sharing. Since its creation in the 1980s, the Radio Frequency IDentification (RFID) has been widely used to automatically identify and track tags attached to objects applied in various industries [18]. In the twenty-first century, the wireless communication—Wireless Sensor Networks (WSNs)—has become the most important infrastructure of the implementation of IoT [19, 20]. It is a system composed of sensors and radio frequency transceivers, power sources, and microcontrollers [21]. Thus, various software and hardware are used in the WSNs, such as the Internet Protocol version 6 (IPv6), high-speed and low-cost communication via WiFi and Wimax, also a local communication through RFID and an external communication using mobile platforms. IoT implementation in industrial services is empowered by the integration of WSN and RFID where WSN enables cooperative capabilities of multi-topologies communication. IoT researches focus on innovative technologies including interoperability, governance, efficiency, security, autonomous behavior, and reliability to achieve ambient intelligence and ubiquitous computing.

Cloud computing technologies offer high performance and low cost [22]. Virtualization technology provides cloud computing with various services to the user including software, hardware, platforms, and other IT infrastructure, resource sharing, dynamic allocation, flexible extension, and other advantages [23]. The service-oriented technologies, like service-oriented architecture (SOA) and web service, enable to achieve on-demand data integration, especially in big manufacturing data [24]. For instance, the cloud computing-related technologies Hadoop and MapReduce are used to store and process the collected data [25, 26]. Storing the large volume of data in a cloud computing center enables fast computation of uploaded data, significant systems' orchestration, and service orientation in manufacturing by considering the sharing of services and collaborative service component [27]. Cloud manufacturing [28], uses cloud computing, is constructed by a highly distributed network of resource called "smart networked manufacturing model." Its aim is to higher on-demand market productivity through a global cooperation and knowledge-intensive innovation.

3.2 Analytics and Intelligence

The smart manufacturing refers to a real-time collection, transmission, and analysis of data along with model-based simulation and optimization that positively impact all aspects of manufacturing [29]. The use of machine learning and data mining in manufacturing aims to reduce the cost efficiency and improve the quality of productivity. Advances in artificial intelligence, machine learning, and data analytics have taken place in the industrial area over the last few years.

Big Data Analytics (BDA) in manufacturing aims to change the perception of machines, processes, and their operations. Analytics can identify patterns in the data, model the behaviors of equipment, and predict failures and prescript on improving equipment availability and performance. For example, IBM's cognitive system Watson [30] produces accurate predictions and insights synthesized from vast amounts of unstructured data. It answers complex questions, provides real-time alerts, utilizes cognitive capabilities to enrich the information around manufacturing processes, to advise a user on exactly how to fix and resolve failures before they occur. Database management systems (DBMSs), business intelligence and analytics software, and big data leaders (e.g., Apache Hadoop data processing platform [25]) enable to improve data analysis by collecting a large amount of data provided by IoT devices. The BDA integration in the mining industry will enable a high performance and good quality production [2]. The prominent technologies of cyber-physical system (CPS) and digital twin (DT) are used in the core of CPI where CPS processes of a large amount of data and performs real-time operations [31]. And DT uses data analytics and machine learning algorithms to afford a cost-effective, reliable, and smart maintenance strategy [32].

The Big Data Analytics (BDA), IoT, and cloud computing-enabled industries refer to the three backbone concepts of the process of smart manufacturing (see Table 1). They have tremendous modern advanced manufacturing system effects on future mining industry. Their common aim is to enable optimized manufacturing resources and to offer smart decision-making in manufacturing systems [36].

Table 1 BDA-, IoT-, and cloud-based Industries 4.0 for smart manufacturing process

Concept	Research characteristics and technologies	References
BDA	Big data processing and analytics	
	Semantic integration of massive data	[12]
	Advanced decision-making models	
	Flexible information model for IoT-based applications	[11]
	Data science and data analytics	[15]
	Real-time huge amount of data collection and analysis	[13]
IoT	Real-time data collection	[33]
	RFID-enabled identification and tracking technologies	[18]
	Wireless Sensor Networks	[19, 20]
	Wireless cyber-physical system	[34]
Cloud	High performance and low cost	[22]
	Hadoop and MapReduce	[25, 26]
	SOA and Web Service	[24]
	Service-orientation in manufacturing	[27]
	Cloud manufacturing	[28]
	Cloud-based cyber-physical system	[35]

3.3 Human–Machine Interaction

Personal devices, touch interfaces, and virtual and augmented reality devices are ubiquitous and increasingly in use in the consumer world. The human–machine interaction has increased the familiarity with such devices, which will ease the implementation of natural feature interaction in the manufacturing environment [10]. For example, the German start-up Ubimax [37] offers solutions fully optimized for wearable computing and augmented reality.

Augmented and Virtual Reality (AR and VR) are the two most frequently mentioned technologies that have been applied to support the realization of Industry 4.0 [38]. They are expected to significantly grow in the next years [39]. The term Industrial Augmented Reality (IAR) is the application of AR to support industrial process. IAR has proven to be a suitable tool for manufacturing strategies [40]. It supports the operators by providing context-aware assistance, quality control, material management, and maintenance, direct or indirect visualization of real-world physical interaction, and so forth. For example, IAR pipeline provides advanced capabilities to scan and track the visual elements and features of the pipeline and display AR interactive information. It requires AR techniques of sensor and image fusion techniques to determine the position and orientation of the observer and the real/virtual object, such as techniques based on sensors (e.g., Inertial Measuring Unit (IMU), GPS, ultrasounds) and mage-based techniques using markers.

3.4 *Digital-to-Physical Conversion*

The actual intelligent manufacturing has been advanced by the use of disruptive technologies toward smart manufacturing essentially grounded on CPI (see Fig. 4). Two main technologies CPS and DT are merging simultaneously and provide greater efficacy and smartness in the manufacturing systems.

Cyber-Physical System (CPS) is the core foundation of Industry 4.0 [41]. It has been identified to be a major area of current research by the US National Science Foundation [42]. CPSs are blurring the boundaries of the real and virtual worlds (physical and cyber computational elements) where they are intertwining, interacting, integrating, and coordinating. CPS provides dynamic monitoring and control of the physical entities through real-time sensing, information feedback, and other services [43]. The large-scale cyber-physical platforms are built upon using virtual concepts and AI algorithms to address the challenges of heterogeneous data, distribution systems, and lack of communication [44]. In this regard, the model-based system applicable to CPS was designed using MATLAB and Simulink [45], Modelica comprising network communication, and real-time task scheduling [46]. Moreover, new extensions of CPS have emerged like, the Wireless Cyber-Physical System (WCPS), open-source simulation environment for wireless control systems, integrate Simulink and TOSSIM wireless sensor [34]; and the cloud-based cyber-physical system (CBCPS) is proposed as a new manufacturing paradigm by combining Internet of Things (IoT), cloud computing, Big Data Analytics, and CPS [35].

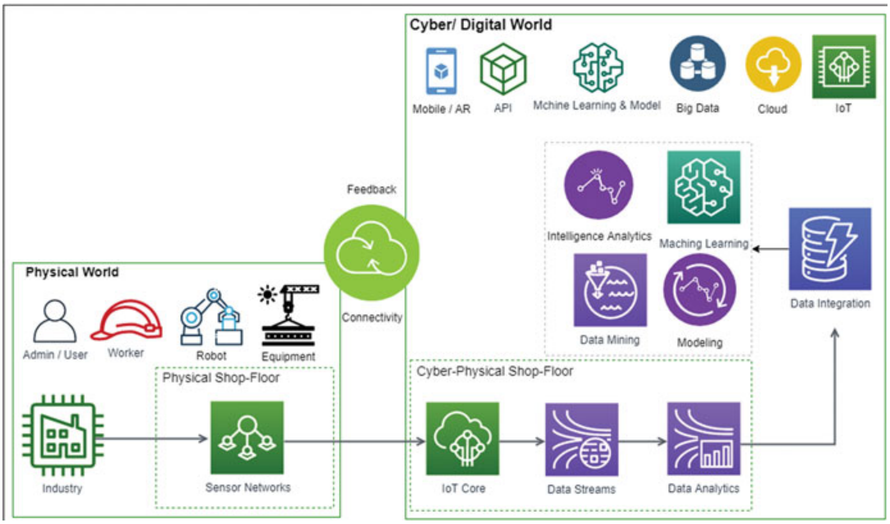


Fig. 4 Mapping between physical and cyber/digital worlds

Digital Twin (DT) creates virtual models of physical objects or process aiming to simulate their behaviors in the physical world to provide feedback. It is characterized by the seamless integration between the cyber and physical spaces [47]. It aims to create virtual models by realistically reproducing the physical appearance like twins (e.g., properties, structure and behaviors of the physical asset, object, process, or system) based on the cumulative real-time captured data and then to simulate their behaviors, spatiotemporal status, functions to predict and control their future states through feedback [48–50]. As shown in Fig. 4, the IoT Core acts as the entry point collecting real-time streaming data and then processing it in the data analysis phase (e.g., search for anomalies, real-time aggregation, correlation, filtering, sampling, etc.) before the storing and recording phase. Once the data starts flowing into the stream processing and analysis engine, the data is stored in cloud-based big data platform that analyzes and processes data through advanced analytics, machine learning, and deep learning and provides efficient decisions and recommendations to be executed in the physical part. Consequently, the DT enables manufacturing making more accurate decisions, monitoring, predictions, and production [51].

4 Discussion: Next Horizon of Mining Industry 4.0

The mining companies need to push for the next horizon of operational effectiveness of their production. The paradigm of Industry 4.0 aims to create smart factories that have manufacturing systems able to monitor physical processes in order to make smart real-time decisions [52]. Creating new operating business models and fostering digital culture are the new trends adopted by competitive industrial companies. A plethora of innovative technologies are created and deployed to assist miners in their roles such as the use of artificial intelligence, autonomous robots, drones, DTs, AR and VR, 3D and 4D printing, and so forth. They are in the core of new generation intelligent manufacturing by shifting from knowledge-based intelligent manufacturing to data-driven smart manufacturing [53]. The disruptive technologies of the Mining Industry 4.0 will offer ways of leveraging data to unlock its value potential. For instance, big data platforms will store and record relevant information, then advanced analytics will transform it into outcomes helping in decision-making and predictive maintenance, and later the 3D printing converts the digital data modeling-based construction into tangible workpiece. The aim of using these disruptive technologies is to achieve enhanced, connected, collaborative, and autonomous mining of the future. The McKinsey Digital report [10] gave an example from the oil and gas industry where companies lose 99% of their data through information leakage; consequently, none of the collected data are used to drive decision-making. Therefore, it is important to avoid information leakages by effectively managing the information along the digital thread, from the capturing data to reaching operational makers. The key to leveraging the potential of the Mining Industry 4.0 is by collecting and sharing information across the digital thread. This chapter has mentioned many research and development activities that are concerned about carrying out predictive maintenance-based data-driven digital model. The data-driven and knowledge-enabled smart manufacturing have emerged with CPI where DT has drawn considerable interest in the industry for cost-effective and smart maintenance. The DT, data-driven digital replica of the offshore asset, provides the computational model updated based on the collected and processed data coming from the different sensors installed in the offshore asset [32]. The digital model of the DT aims to predict the future condition of the physical asset [50]. An example of the offshore asset is the pipeline system. The digital twin of the pipeline system is modelled based on different field sensor data (e.g., motion sensor/accelerometer, wave radar, pressure and temperature sensors, etc.), where the IoT-based systems transmit the data to store in the cloud for real-time data aggregation and analytics to find the insight between the measured sensor data and provide actionable decisions for predictive maintenance capabilities using machine learning and thus save operating cost [32]. The smart mining industry will enable the monitoring and maintenance of the physical field based on actual sensor data from the physical field using disruptive technologies of Industry 4.0. Indeed, the use of these disruptive technologies will enable a significant shift to allow manufacturing sites to, for example, react faster, flexibility, and automation on production, and become self-optimizing, covering volume, and variety of production. The convergence of business IT and manufacturing IT systems will create relevant and efficient value through leveraging data using emergent technologies toward reflecting the Mining Industry 4.0 levers. Using the disruptive technologies of Industry 4.0 as the path to the future of manufacturing will allow mining companies to do more than just upgrading and optimizing physical assets to increase their operational effectiveness and maintain competitiveness in the global manufacturing market.

5 Conclusion

The disruptive technologies construct the pillars of smart manufacturing in the mining industry. The review describes and analyzes the core of new generation intelligent manufacturing that uses these emerging technologies (IT 4.0) to shift from knowledge-based intelligent manufacturing to data-driven smart manufacturing. It first presented a brief overview of the industrial revolution and a description of the principles of Industry 4.0 and features of smart manufacturing. Then, the article reviews the academic progress that addresses the issue of cyber-physical integration in manufacturing, especially in the mining industry. It gives the analysis of research studies that carry out a large number of theoretical researches and valuable techniques on digital manufacturing toward smart manufacturing. The specific data analysis corresponds to each research question by providing specific insights related to what is being researched. Finally, the final section concluded with future insights, recommendations, and perspectives of the promising incomes of using disruptive technologies (IT 4.0) in mining industry though data-driven predictive maintenance.

References

1. Ghodrati, B., Hoseinie, S., & Garmabaki, A. (2015). Reliability considerations in automated mining systems. *International Journal of Mining, Reclamation and Environment, 29*(15), 404–418.
2. Sishi, M. N., & Telukdarie, A. (2017). Implementation of Industry 4.0 technologies in the mining industry: A case study. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Singapore (pp. 201–205). https://doi.org/10.1109/IEEM.2017.8289880.
3. Cheng, Y., Zhang, Y., Ji, P., et al. (2018). *The International Journal of Advanced Manufacturing Technology, 97*, 1209. https://doi.org/10.1007/s00170-018-2001-2.
4. Tao, F., & Qinglin, Q. (2017). New IT driven service-oriented smart manufacturing: framework and characteristics. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. https://doi.org/10.1109/TSMC.2017.2723764.
5. Wang, L., & Shih, A. (2016). Challenges in smart manufacturing. *Journal of Manufacturing Systems, 40*(SI), 1.
6. Preuveneers, D., & Ilie-Zudor, E. (2017). The intelligent industry of the future: A survey on emerging trends, research challenges and opportunities in Industry 4.0. *Journal of Ambient Intelligence and Smart Environments, 9*, 287–298. https://doi.org/10.3233/AIS-170432.
7. Schwab, K. (2019). The Fourth Industrial Revolution: What It Means and How to Respond. Retrieved from https://www.foreignaffairs.com/articles/2015-12-12/fourth-industrial-revolution (Last accessed October 5, 2019).
8. Kusiak, A. (2018). Smart manufacturing. *International Journal of Production Research, 56*(1–2), 508–517.
9. Peruzzini, M., Grandi, F., & Pellicciari, M. (2017). Benchmarking of tools for user experience analysis in industry 4.0. *Procedia Manufacturing, 11*, 806–813.
10. Industry 4.0. (2015). How to navigate digitalization of the manufacturing sector. *McKinsey Digital 2015*.
11. Cai, H., Xu, L., Xu, B., Xie, C., Qin, S., & Jiang, L. (2014). IoT-based configurable information service platform for product lifecycle management. *IEEE Transactions on Industrial Informatics, 10*(2), 1558–1567.
12. Sabar, M., Jayaweera, P., & Edirisuriya, E. (2016). SAIF-refactored efficiency interpolation in the HL7 specifications development paradigm. *Journal of Industrial Information Integration, 4*, 35–41.
13. Lee, J., Bagheri, B. & Kao, H. (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing Letters, 3*(2015), 18–23.
14. Shrouf, F., Ordieres, J., & Miragliotta, G. (2014). *Smart factories in industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm* (pp. 697–701).
15. Cheng, Y., Chen, K., Sun, H., Zhang, Y. & Tao, F. (2018). Data and knowledge mining with big data towards smart production. *Journal of Industrial Information Integration, 9*, 1–13. https://doi.org/10.1016/j.jii.2017.08.001.
   Cheng, Y., Chen, K., Sun, H., Zhang, Y. & Tao, F. (2017). Data and knowledge mining with big data towards smart production. *Journal of Industrial Information Integration*.
16. Zhong, R. Y., Dai, Q. Y., Qu, T., Hu, G. J., & Huang, G. Q. (2013). RFID-enabled real-time manufacturing execution system for mass-customization production. *Robotics and Computer-Integrated Manufacturing, 29*(2), 283–292.
17. Xia, F., Yang, L. T., Wang, L., & Vinel, A. (2012). Internet of Things. *International Journal of Communication Systems, 25*(9), 1101–1102.
18. Zhai, C., Zou, Z., Chen, Q., Xu, L., Zheng, L., & Tenhunen, H. (2016). Delay-aware and reliability-aware contention-free MF-TDMA protocol for automated RFID monitoring in industrial IoT. Journal *of Industrial Information Integration, 3*, 8–19.
19. Finogeev, A. G., & Finogeev. A. A. (2017). Information attacks and security in wireless sensor networks of industrial SCADA systems. *Journal of Industrial Information Integration, 5*, 6–16.

20. Bag, G., Pang, Z., Johansson, M., Min, X., & Zhu. S. (2016). Engineering friendly tool to estimate battery life of a wireless sensor node. *Journal of Industrial Information Integration, 4*, 8–14.
21. Li, S., & Xu, L. (2017). *Securing the Internet of Things*. Syngress: Elsevier.
22. Mitra, A., Kundu, A., Chattopadhyay, M. & Chattopadhyay, S. (2017). A cost-efficient one time password-based authentication in cloud environment using equal length cellular automata. *Journal of Industrial Information Integration, 5*, 17–25.
23. Zhou, K., Taigang, L., & Lifeng, Z. (2015). Industry 4.0: Towards future industrial opportunities and challenges. In *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, Zhangjiajie (pp. 2147–2152).
24. Xiang, F., Jiang, G. Z., Xu, L. L., & Wang, N. X. (2016). The case-library method for service composition and optimal selection of big manufacturing data in cloud manufacturing system. *The International Journal of Advanced Manufacturing Technology, 84*(1–4), 59–70.
25. Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010). The hadoop distributed file system. In *Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)* (MSST '10) (pp. 1–10). Washington, DC, USA: IEEE Computer Society.
26. Tao, F., Cheng, Y., Xu, L., Zhang, L., & Li, BH. (2014). CCIoT-CMfg: cloud computing and Internet of Things based cloud manufacturing service system. *IEEE Transactions on Industrial Informatics 10*(2), 1435–1442.
27. Moghaddam, M., & Nof, S. Y. (2017). Collaborative service-component integration in cloud manufacturing. *International Journal of Production Research*, Published online 13 September 2017.
28. Mourtzis, D., & Vlachou, E. (2016). Cloud-based cyber-physical systems and quality of services. *The TQM Journal, 28*(5), 704–733.
29. Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. *Journal of Manufacturing Systems, 48*(Part C), 157–169.
30. IBM Web Site. (2019). Watson, Available at: https://www.ibm.com/watson (Last accessed October 5, 2019).
31. Edward, L. (2015). The past, present and the future of cyber-physical systems: a focus on models. *Sensors, 15*(3), 4837–4869.
32. Bhowmik, S. (2019). *Digital Twin of Subsea Pipelines: Conceptual Design Integrating IoT, Machine Learning and Data Analytics*. https://doi.org/10.4043/29455-MS
33. Zhong, R. Y., Peng, Y., Xue, F., Fang, J., Zou, W., Luo, H., et al. (2017). Prefabricated construction enabled by the Internet-of-Things. *Automation in Construction, 76*, 59–70.
34. WCPS. (2019). Wireless cyber-physical simulator, available at: http://wsn.cse.wustl.edu/index.php/WCPS:_Wireless_Cyber-Physical_Simulator (Last accessed October 5, 2019).
35. Mourtzis, D., & Vlachou, E. (2016). Cloud-based cyber physical systems and quality of services. *TQM Journal, 28*. https://doi.org/10.1108/TQM-10-2015-0133.
36. Liu, Y. K., & Xu, X. (2017). Industry 4.0 and cloud manufacturing: A comparative analysis. *Journal of Manufacturing Science and Engineering, 139*(3), 034701-1–8.
37. Ubimax Web Site. (2019). Innovative Solutions. https://www.ubimax.com/en/solutions/ (Last accessed October 5, 2019).
38. Liao, Y., Deschamps, F., Loures, E. D. F. R., & Ramos, L. F. P. (2017). Past, present and future of Industry 4.0-a systematic literature review and research agenda proposal. *International Journal of Production Research, 55*(12), 3609–3629.
39. Goldman Sachs Global Investment Research Technical Report: Virtual and Augmented Reality—Understanding the Race for the Next Computing Platform, Dec. 2017, Available: http://www.goldmansachs.com/our-thinking/pages/technology-driving-innovation-folder/virtual-and-augmented-reality/report.pdf. (Last accessed October 5, 2019).
40. Fraga-Lamas, P., Fernández-Caramés, T. M., Blanco-Novoa, Ó., & Vilar-Montesinos, M. A. (2018). A review on industrial augmented reality systems for the Industry 4.0 shipyard. *IEEE Access, 6*, 13358–13375.
41. Xu, L. D., Xu, E. L., & Li, L. (2018). Industry 4.0: State of the art and future trends. *International Journal of Production Research, 56*, 1–22.

42. Cyber-Physical Systems NIST. (2019). Available at: https://www.nist.gov/el/cyber-physical-systems (Last accessed October 5, 2019).
43. Liu, Y., Peng, Y., Wang, B., Yao, S., & Liu, Z. (2017). Review on cyber–physical systems. *IEEE/CAA Journal of Automatica Sinica, 4*(1), 27–40.
44. Giordano, A., Spezzano, G. & Vinci, A. (2014). Rainbow: an intelligent platform for large-scale networked cyber-physical systems. In *Proceedings of the 5th International Workshop on Networks of Cooperating Objects of Smart Cities (UBICITEC)*, Berlin, April 14.
45. MathWorks. (2019). Model-based design of cyber-physical systems in MATLAB and Simulink, available at: www.mathworks.com/discovery/cyber-physical-systems.html (Last accessed October 5, 2019).
46. Henriksson, D., & Elmqvist, H. (2011). Cyber-physical systems modeling and simulation with Modelica. In *Proceedings 8th Modelica Conference*, Dresden, June 20–22.
47. Tao, F., Zhang, H., Liu, A., & Nee, A. Y. (2019). Digital twin in industry: State-of-the-Art. *IEEE Transactions on Industrial Informatics, 15*, 2405–2415.
48. Schleich, B., Anwer, N., Mathieu, L., & Wartzack, S. (2017). Shaping the digital twin for design and production engineering. *CIRP Annals, 66*(1), 141–144.
49. Tao, F., & Zhang, M. (2017). Digital twin shop-floor: a new shop-floor paradigm towards smart manufacturing. *IEEE Access, 5*, 20418–20427.
50. Liu, Z., Meyendorf, N., & Mrad, N. (2018). The role of data fusion in predictive maintenance using digital twin. *AIP Conference Proceedings, 1949*, 020023.
51. Qi, Q., & Tao, F. (2018). Digital twin and big data towards smart manufacturing and Industry 4.0: 360 degree comparison. *IEEE Access, 6*, 3585–3593.
52. Wang, S., Wan, J., Zhang, D., Li, D., & Zhang C. (2016). Towards smart factory for Industry 4.0: A self-organized multi-agent system with big data based feedback and coordination. *Computer Networks, 101*, 158–168.
53. Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent manufacturing in the context of Industry 4.0: a review. *Engineering, 3*(5), 616–630.

Transport Factor in the Model of International Trade

Mykhaylo Voynarenko and Anatoliy Kholodenko

Abstract *Research goals and objectives*: to introduce transport factor into the model of international trade and to clarify related transformations.

Subject of research: model of international trade with transport factor.

Research methods used: economic and mathematical modeling, analytical methods, comparative analysis, computer simulation.

Results of the research: In the classical models, any world price deviation from the internal equilibrium price indicates expediency of foreign trade. In the same models with transport factor, there is some "dead zone" for international trade when world price deviation from the internal equilibrium price does not exceed the corresponding transportation costs.

In this chapter, it is established that each country trades some products with another country at the most favorable export conditions for itself, or at the most favorable import conditions for itself, or does not trade. Equality of marginal transportation costs to established transport rates at the optimal transport activity of each country with transportation of each product in each direction is also discussed in the chapter.

Keywords International trade · Transport · Equilibrium price · Demand · Supply

1 Introduction

International trade gains the increasing importance due to the globalization of economic development, at the same time the international trade system complexity grows.

M. Voynarenko
Khmelnytsky National University, Khmelnytskyi, Ukraine

A. Kholodenko (✉)
Odesa National Maritime University, Odesa, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_29

In classical models of international trade [1–3], the transport factor usually is not considered—for simplification or because of its alleged insignificance. Two-country model of trade and product innovation with fluctuations of innovation activities is considered in [4]. Two-country dynamic model of international trade and endogenous growth is studied in [5]. In [6], it is showed that international trade increases the conflict of interest between owners and middle managers within firms, and these eventually lead to decentralized corporate hierarchies. The effects of tariff changes on world steady states and transition dynamics in international trade are studied in [7]. In [8], optimal monetary policy in a two-country new Keynesian model with international trade in intermediate inputs is examined. In [9], a new simulated method of moments estimator to estimate elasticity of international trade from disaggregate price and trade-flow data is developed. A model of international trade based on new trade theory with financial constraints and non-homothetic preferences is proposed in [10]. Paper [11] identifies a new terms-of-trade externality that is exercised through tariff setting. An empirical implementation of a general-equilibrium model of international trade with heterogeneous manufacturing firms is presented in [12].

A game model of international trade with transport costs and duopolistic interaction between two monopolists located in two different countries is investigated in [13]. In [14], a simple two-country model of international trade is developed in which the transportation cost between countries is endogenously determined by pricing and investment decisions of cross-border transport infrastructure.

In fact, transport plays a significant role in trade volumes and prices, in particular, as it will be shown, causes a variety of equilibrium prices in the economic system.

So, the purpose of this chapter is the transport introduction in the international trade model and clarification of related transformations.

2 Results

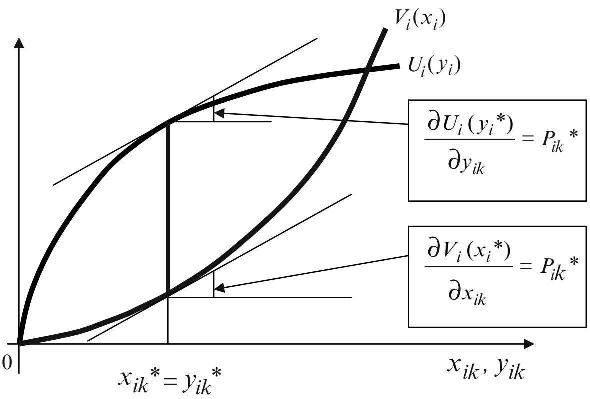
Let m products and n country participants in the economic system are considered. Each country i, $i = 1, \ldots, n$ is characterized by the continuous increasing functions $V_i(x_i)$ of costs for production in volume $x_i = \{x_{ik}\}_{k=1,\ldots,m}$, and $U_i(y_i)$ is the benefit (expressed in the commensurably with costs) from product consumption in volume $y_i = \{y_{ik}\}_{k=1,\ldots,m}$. The convexity down of functions $V_i(x_i)$ and convexity up (concavity) of functions $U_i(y_i)$, $i = 1, \ldots, n$ are traditionally supposed.

In the absence of international trade, the model of each country i is as follows:

$$U_i\,(y_i) - V_i\,(x_i) \to \max,\, y_i \le x_i,\, y_i \ge 0,\, x_i \ge 0. \tag{1}$$

Optimality conditions, certainly, are $y_i* = x_i*$, $\frac{\partial V_i(x_i*)}{\partial x_{ik}} = \frac{\partial U_i(y_i*)}{\partial y_{ik}} = P_{ik}*$, $k = 1, \ldots, m$ (Fig. 1), where P_{ik*} is the internal equilibrium prices on product k in the country i.

Fig. 1 Finding of the optimum in the closed national economy model



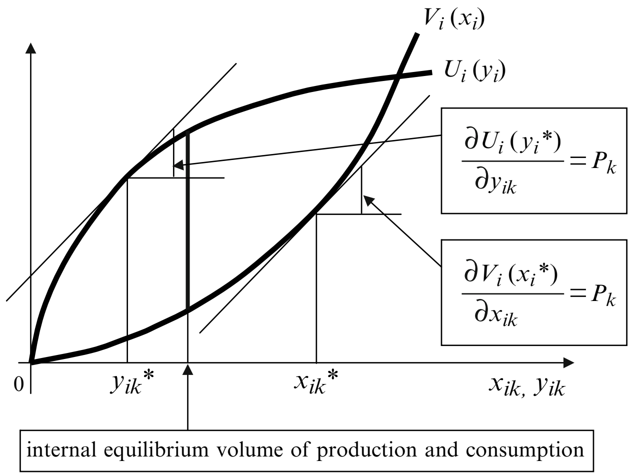


Fig. 2 Finding of the optimum in the open model of national economy

With the implementation of international trade (still without transport factor), the model of each country i turns the world prices $P = \{P_k\}_{k=1,\ldots,m}$.

$$U_i(y_i) - V_i(x_i) + (P, x_i - y_i) \to \max, \; y_i \geq 0, x_i \geq 0. \tag{2}$$

Here the optimum consumption volume y_{i*} in each country i can already be less than the own production volume x_{i*} or even to exceed it due to foreign trade at the world prices P, uniform for whole economic system (Fig. 2).

It is interesting that in Fig. 2 the world prices are higher internal equilibrium ones; therefore, own production x_{i*} increases, however the internal consumption y_{i*} decreases (it has recently been observed with the world prices on gasoline and its domestic production and consumption). Conversely, in the presence in the world market of cheap qualitative products, its own production is phased down, but consumption grows. To prevent such effects, there should be corresponding quotas and duties (which in this chapter are not considered).

The world prices P_* in the model (2) are equilibrium if the balance condition in the whole economic system is fulfilled for the optimum plans of the certain countries production and consumption corresponding to them:

$$\sum_{i=1}^{n} x_i * (P*) = \sum_{i=1}^{n} y_i * (P*) \tag{3}$$

Model (2)–(3) is rather convenient for analysis and finding of the equilibrium prices; however, it does not consider which countries have mutual trade at the established prices, and what transportation costs and restrictions at the same time arise.

Now we will introduce these factors. We will designate z_{ijk} the desired (from the country i point of view) trade volume of the country i with the country j of product k (positive z_{ijk} corresponds to export from the country i, negative one to import to it), $i = 1, \ldots, n, j = 1, \ldots, n, k = 1, \ldots, m$. It is clear, $z_{ijk} = 0$, $i = 1, \ldots, n$, $k = 1, \ldots, m$ (the country does not trade with itself). Respectively the prices are differentiated: $P_{ijk} = P_{jik}$ is the product price k in trade between the countries i and j, $i = 1, \ldots, n, j = 1, \ldots, n, k = 1, \ldots, m.$

We will also designate the transportation volume as t^i_{sjk} of the country i of product k between the countries s and j, the corresponding continuous convex down growing function of transportation costs of the country i as $W^i_{sjk} = \left(t^i_{sjk}\right)$, $i = 1, \ldots, n$, $s = 1, \ldots, n$, $j = 1, \ldots, n$, $k = 1, \ldots, m$; the transport tariff for product unit k from the country i to the country j as F_{ijk}; the share of transportation costs of product unit k from the country i to the country j, which is paid by the country i as α_{ijk} (from here $\alpha_{ijk} + \alpha_{jik} = 1$), $i = 1, \ldots, n, j = 1, \ldots, n, k = 1, \ldots, m.$

In case of the system of prices P_{ijk}, transport tariffs F_{ijk} and transportation conditions α_{ijk} of each country i determines volume of production x_{i*}, consumption y_{i*}, trade with another country z_{ijk*}, and transport activity $t^i_{sjk}*$, $i = 1, \ldots, n$, $s = 1, \ldots, n, j = 1, \ldots, n, k = 1, \ldots, m$, proceeding from the problem solution:

$$U_i\left(y_i\right) - V_i\left(x_i\right) + \sum_{j=1}^{n}\sum_{k=1}^{m} P_{ijk} z_{ijk} + \sum_{s=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\left(F_{sjk} t^i_{sjk} - W^i_{sjk}\left(t^i_{sjk}\right)\right) -$$

$$-\sum_{j=1}^{n}\sum_{k=1}^{m} \alpha_{ijk} F_{ijk}\left|z_{ijk}\right| \to \max, \tag{4}$$

$$\sum_{j=1}^{n} z_{ijk} = x_{ik} - y_{ik}, k = 1, \ldots, m \tag{5}$$

$$y_{ik} \geq 0, x_{ik} \geq 0, t^i_{sjk} \geq 0, s = 1, \ldots, n, j = 1, \ldots, n, k = 1, \ldots, m \tag{6}$$

We will consider the system of prices P_{ijk*}, transport tariffs F_{ijk*}, and transportation conditions α_{ijk*} to be equilibrium if for the optimum plans, corresponding to them, the following conditions are satisfied:

$$z_{ijk}* = -z_{jik}*, i = 1, \ldots, n, j = 1, \ldots, n, k = 1, \ldots, m \tag{7}$$

$$\sum_{i=1}^{n} t^i_{sjk}* = \left|z_{sjk}*\right|, s = 1, \ldots, n, j = 1, \ldots, n, k = 1, \ldots, m \tag{8}$$

i.e., the plans of the countries for mutual trade, made independently from each other, coincide (7), and these trade volumes are completely provided with transport of different countries (8).

In this variety of prices P_{ijk*}, transport tariffs F_{ijk*}, and transportation conditions α_{ijk*}, each country i behaves as follows. First, it defines the most attractive to itself conditions of export

$$E_{ik}* = \max_j \left\{ P_{ijk} * -\alpha_{ijk} * F_{ijk}* \right\} \tag{9}$$

and import

$$I_{ik}* = \min_j \left\{ P_{ijk} * +\alpha_{ijk} * F_{ijk}* \right\} \tag{10}$$

on each product $k = 1, \ldots, m$. Then it compares them with the optimum internal prices P_{ik*} of the corresponding products $k = 1, \ldots, m$ (Fig. 3).

Unlike Fig. 2 where any world price deviation from internal equilibrium one caused the foreign trade expediency, the certain "dead zone" for international trade appears here. If the internal equilibrium price P_{ik*}, though it deviates from the other countries prices P_{ijk*} offered to the country $i, j = 1, \ldots, n, j \neq i$, gets to this zone $[E_{ik*}; I_{ik*}]$, trade is unprofitable because of transportation costs.

If $P_{ik\ *} < E_{ik*}$, the export is favorable, and to that just very country (or those countries), from conditions of which E_{ik*} (the best conditions among the offered) is determined. If $P_{ik\ *} > I_{ik*}$, the import is favorable, and from that just very country (countries) from the conditions of which I_{ik*} is determined.

By the way, in Fig. 3 such arrangement $E_{ik\ *} \leq I_{ik*}$ has always to be carried out, otherwise the country i would have the opportunity to be engaged in usual speculation, buying up and reselling product k with benefit for itself and losses for the countries with such careless trade offers.

Necessary conditions of the equilibrium state in the international trade model taking into account transport (4)–(8) are as follows:

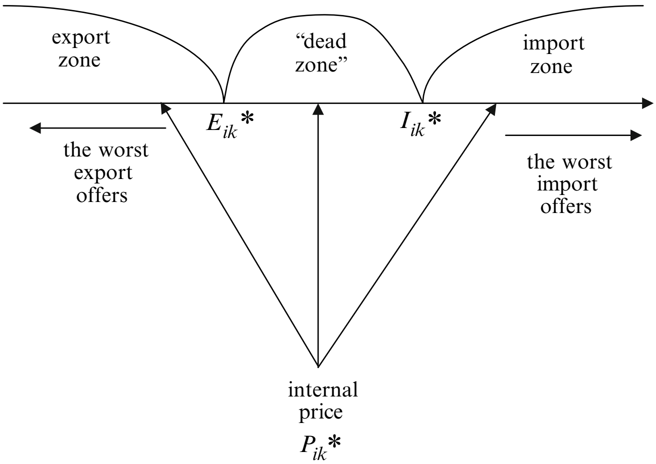


Fig. 3 Influence of the transport factor on international trade

$$z_{ijk} * \left(E_{ik} * - \left(P_{ijk} - a_{ijk} * F_{ijk} *\right)\right) \left(I_{ik} * - \left(P_{ijk} * + a_{ijk} * F_{ijk} *\right)\right) = 0$$

$$i = 1, \ldots, n, j = 1, \ldots, n, \quad k = 1, \ldots, m \tag{11}$$

$$\frac{dW^i_{sjk}\left(t^i_{sjk}*\right)}{dt^i_{sjk}} = F_{sjk}*, s = 1, \ldots, n, j = 1, \ldots, n, k = 1, \ldots, m \tag{12}$$

The condition (11) means that the country i carries out the product trade k with the country j on the most favorable to itself export conditions (if the second multiplier is equal to zero, then z_{ijk*} is positive), or on the most favorable to itself import conditions (if the third multiplier is equal to zero, then z_{ijk*} is negative), or does not carry out it at all (z_{ijk*} is equal to zero).

That is, at all varieties of prices P_{ijk*}, transport tariffs F_{ijk*}, and transportation conditions α_{ijk*}, each exporter of the given product k levels the export conditions for all real (not potential) country buyers (in case of the inequality, it would be expedient to change the export structure to advantage of more favorable country; therefore, the state would not be optimum for the export country). However, different exporters' prices for the same product can differ due to various transportation conditions (geographical position, communications network development, etc.). At the same time, even "expensive" suppliers find their ("relatives") buyers, whose demand does not switch due to economy on transportation costs.

Similarly, each country buyer levels final import prices (purchase prices plus transportation costs) for all its "active" country suppliers. Otherwise, the system state would not be in equilibrium—the given buyer would like to buy more from the countries more favorable to him due to reduction of other purchases. But in different country importers, these import prices can be different—depending on advantages of their geographical position and transport network.

Note that the prices P_{ijk*} of the "passive" countries, which are suppliers of the given country i (i.e., those countries j for which $P_{ijk\,*} + \alpha_{ijk\,*} F_{ijk\,*} > I_{ik*}$, $z_{ijk\,*} = 0$), can be still increased just as the prices of potential (not actual) countries, which are buyers of the given country i (the countries j for which $P_{ijk\,*} - \alpha_{ijk\,*} F_{ijk\,*} < E_{ik*}$, $z_{ijk\,*} = 0$) can be reduced—equilibrium in economic system will not be broken due to it. So, the equilibrium prices set in the international trade model are not the only one.

The condition (12) means equality of marginal transportation costs to the established transport tariff in the point of optimum transport activities of each country i for transportation of each product k on each direction sj.

Note that owing to the triangle inequality (the sum of any two side lengths is more than the length of the third one) from the point of view of transportation costs of the resale, trade through the third countries does not make sense; however, the transit is possible.

3 Conclusions

The introduction of the transport factor to the international trade model leads to the equilibrium price differentiation and the decrease in their general level owing to demand reduction in the world market due to transportation costs and supply reduction because of transport capacity limitation.

Transport development expands possibilities of supplying countries and consumers choice and increases the general foreign trade activity and the equilibrium prices; besides, transport service export acts as independent branch sector of economy.

In the classical models, any world price deviation from the internal equilibrium price indicates expediency of foreign trade. In the same models with transport factor, there is some "dead zone" for international trade when world price deviation from the internal equilibrium price does not exceed the corresponding transportation costs.

It is established that each country trades some products with another country at the most favorable export conditions for itself, or at the most favorable import conditions for itself, or does not trade. That is, beyond all diversity of prices, transport rates, and conditions of carriage, each exporter of the product aligns the conditions of its exports to all its real (not potential) country importers (in the case of inequality, it would be advisable to change the export structure in favor of more profitable country, therefore, the state would not be optimal for the exporting country). However, prices for the same product of different exporters may differ due to various transport conditions (geographical situation, the development of communication networks, etc.). Herewith, even "expensive" suppliers find their ("close") buyers, demand of them are not diverted due to savings on transport costs. Similarly, each country buyer aligns its final prices of imports (purchase price plus transport costs) to all of its "active" country suppliers. Otherwise, the state of the system would not be an equilibrium—this buyer wanted to buy more in a better country for itself by reducing the remaining purchases. But prices of different country importers can vary depending on their geographical advantages and the state of the transport network.

Equality of marginal transportation costs to established transport rates at the optimal transport activity of each country with transportation of each product in each direction is also discussed in this chapter. Note that because of the triangle inequality (the sum of the lengths of any two sides is more than the length of the third side), resale—trade through the third countries—is meaningless in terms of transport costs; but transit is possible. Hence, the introduction of transport factor to the model of international trade leads to differentiation of the equilibrium prices and a decrease of their overall level because of reduction in demand in the world market due to transportation costs and reduction of supply due to limited transport capacity. The development of transport enhances choice of country suppliers and consumers, increases the overall foreign trade activity and, consequently, the equilibrium prices.

References

1. Krugman, P. R., & Obstfeld, M. (1994). *International economics. Theory and policy*. Harper Collins College Publishers.
2. Lindert, P. H. (1986). *International economics*. IRWIN.
3. Salvatore, D. (1995). *International economics*. Prentice Hall Int.
4. Sushko, I., Gardini, L., & Matsuyama, K. (2018). Coupled chaotic fluctuations in a model of international trade and innovation: Some preliminary results. *Communications in Nonlinear Science and Numerical Simulation, 58*, 287–302.
5. Doi, J., Nishimura, K., & Shimomura, K. (2007). A two-country dynamic model of international trade and endogenous growth: Multiple balanced growth paths and stability. *Journal of Mathematical Economics, 43*(3–4), 390–419.
6. Marin, D., & Verdier, T. (2014). Corporate hierarchies and international trade: Theory and evidence. *Journal of International Economics, 94*(2), 295–310.
7. Alvarez, F. (2017). Capital accumulation and international trade. *Journal of Monetary Economics, 91*, 1–18.
8. Gong, L., Wang, C., & Zou, H. (2016). Optimal monetary policy with international trade in intermediate inputs. *Journal of International Money and Finance, 65*, 140–165.
9. Simonovska, I., & Waugh, M. E. (2014). The elasticity of trade: Estimates and evidence. *Journal of International Economics, 92*(1), 34–50.
10. Goksel, T. (2012). Financial constraints and international trade patterns. *Economic Modelling, 29*(6), 2222–2225.
11. Zissimos, B. (2009). Optimum tariffs and retaliation: How country numbers matter. *Journal of International Economics, 78*(2), 276–286.
12. Balistreri, E. J., Hillberry, R. H., & Rutherford, T. F. (2011). Structural estimation and solution of international trade models with heterogeneous firms. *Journal of International Economics, 83*(2), 95–108.

13. Batabyal, A. A., & Nijkamp, P. (2016). A game model of international trade, transport costs, invasive species, and protocol compliance. *Transportation Research Part D: Transport and Environment, 46*, 267–272.
14. Mun, S., & Nakagawa, S. (2010). Pricing and investment of cross-border transport infrastructure. *Regional Science and Urban Economics, 40*(4), 228–240.

Assessing Land Use/Land Cover Change Using Multitemporal Landsat Data in Agadir City (Morocco)

Ijjou Idoumskine, Ali Aydda, Abdelkrim Ezaidi, and Omar F. Althuwaynee

Abstract The aim of this study is to highlight Land Use/Land Cover (LULC) sever changes dynamics in Agadir city of Morocco during the period 1986–2019. For this purpose, Landsat imageries (5 TM, 7 ETM+, and 8 OLI) acquired in 1986, 1996, 2003, 2014, and 2019 years, were classified using Support Vector Machine (SVM) algorithm to produce LULC maps. First, spectral indices such as the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Built-up (NDBI) were used to discriminate the four LULC classes (built-up areas, bare lands, vegetation, and water bodies) of the study area. Moreover, change detection was applied on classified maps to characterize LULC dynamics class by class. The results reveal a spatial expansion over the 33 years, with high overall accuracy and kappa coefficient values. In detail, vegetation and bare lands have been decreased by 95.58% and 72.06%, respectively, while the built-up areas have been increased by an amount of 66.86%. Overall, the findings of this study could assist planners and decision-makers to guide, in a good manner, the sustainable land development of the city.

Keywords NDVI · NDBI · Supervised classification · Change detection · Landsat imagery

1 Introduction

It was recognized that Land Use/Land Cover (LULC) changes lead to severe environmental problems (climate change, ecosystem degradation, etc.). In light of this fact, whatever the causing act (natural or anthropic), it is essential to assess

I. Idoumskine (✉) · A. Aydda · A. Ezaidi
Department of Geology, Faculty of Sciences, Ibn Zohr University, Agadir, Morocco

O. F. Althuwaynee
Department of Energy and Mineral Resources Engineering, Sejong University, Seoul, Republic of Korea

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_30

and analyze LULC changes to understand its processes and to ensure a sustainable management (guide urbanization expansion, conserve ecosystem, etc.).

Practically, several studies have shown the strength of remote sensing and Geographical Information System (GIS) techniques for mapping production, detecting, and analyzing LULC changes and its environmental impacts [1–3]. Xiuwan [1] used multi-temporal Landsat TM to map and analyze land cover (LC) changes in the west coastal zone of Korea (Ansan City). Fichera et al. [2] used aerial photos and multitemporal Landsat TM scenes to map and analyze LC changes in the area of Avellino (south of Italy). Ahmed and Akter [3] used Landsat multi-spectral surface reflectance data to detect LC changes in southwest Bengal delta (Bangladesh). Bouiadjra et al. [4] used two scenes of Landsat imagery to assess vegetation cover change in Tessala mountains (Algeria).

Object-based classification is the most commonly used method to map LULC, and post-classification change detection technique is the most one used for analyzing surface change. Moreover, spectral indices of Normalized Difference Vegetative Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Difference Built-up Index (NDBI) are widely used to extract LULC.

In Morocco, where the overexploitation of natural resources associated with intense urbanization is prevailing, the rates of LULC dynamics are changing fast over the last decade. El Garouani et al. [5] used two Landsat (TM and ETM+) scenes to map land cover change and soil erosion within the Oued Tlata basin (northeast of Morocco). Mohajane et al. [6] trait the issue of vegetation changes in the Azrou Forest, Middle Atlas of Morocco, between 1987 and 2017 and the effects of LULC change on land surface temperature using multi-temporal Landsat satellite imagery. These studies and others have authenticated the LULC conversion throughout the kingdom.

In Agadir city, one of the largest seaside resorts in Morocco, as the process of urbanization becomes more intense in parallel with the start of the socio-economic development and the globalization of the city, the demand of land for various urban activities increases (in particular for secondary and tertiary activities). Hence, it is important to evaluate the magnitude and spatial extent of these changes within the city to ensure best future planning and management of land.

The present study anticipated to integrate NDVI and NDBI with object-based classification and post-classification change detection method in order to assess the spatiotemporal LULC changes in Agadir city during the period between 1986 and 2019.

2 Materials and Method

2.1 *Study Area*

Agadir city is the capital of Souss-Massa region and one of 13 administrative districts, which occupies the southern Moroccan Atlantic coast, with an approximate area of 112 km^2 [7] (Fig. 1).

Demographically, the district of Agadir had the highest growth rate in Morocco recorded at 3.7% during the period 1982–1994 [7]; the rural exodus was one of the main reasons for the population growth. Currently, the city experienced again an exceptional population growth, of about 421,844 inhabitants, according to the 2014 census [7]. The economy of the city prevails by tertiary activities (more than 50% of the territory), industry, trade, and fishery.

Geographically, Agadir city is marked by a semiarid climate with 340 sunshine days per year. The average annual temperature ranges from 14 °C to 16 °C in January and 19 °C to 25 °C in July. August is the hottest month (average temperature equal to 22.6 °C), while January is the coldest month (average temperature is 13.9 °C) and the annual precipitation is about 250 mm [8].

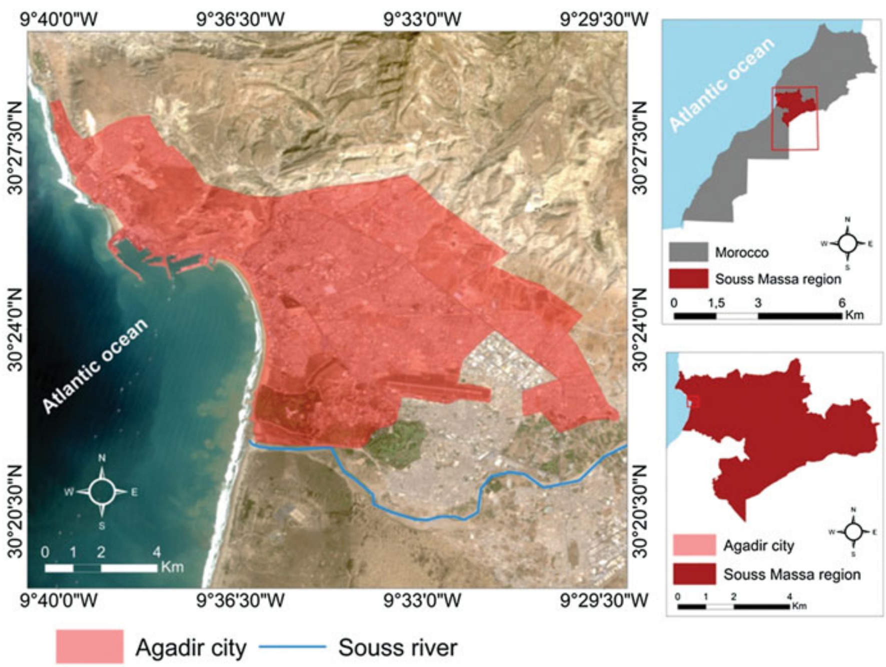


Fig. 1 Location map of the study area

Table 1 Characterizations of the used Landsat data

Acquisition dates	Satellite (sensor)	Bands number	Wavelength (µm)	Spatial resolution (m)
11/02/1986 07/02/1996	Landsat 5 (TM)	Band 1 blue Band 2 green Band 3 red Band 4 NIR Band 5 SWIR1 Band 6 thermal Band 7 SWIR2	0.45–0.52 0.52–0.60 0.63–0.69 0.76–0.90 1.55–1.75 10.40–12.50 2.08–2.35	30 30 30 30 30 120 30
11/02/2003	Landsat 7 (ETM+)	Band 1 blue Band 2 green Band 3 red Band 4 NIR Band 5 SWIR1 Band 6 thermal Band 7 SWIR2 Band 8 pan	0.45–0.52 0.52–0.60 0.63–0.69 0.77–0.90 1.55–1.75 10.40–12.50 2.09–2.35 0.52–0.90	30 30 30 30 30 60 30 15
08/02/2014 22/02/2019	Landsat 8 (OLI)	Band 1 coastal Band 2 blue Band 3 green Band 4 red Band 5 NIR Band 6 SWIR1 Band 7 SWIR2 Band 8 pan Band 9 cirrus	0.43–0.45 0.45–0.51 0.53–0.59 0.63–0.67 0.85–0.88 1.57–1.65 2.11–2.29 0.50–0.68 1.36–1.38	30 30 30 30 30 30 30 15 30

2.2 Dataset

Satellite Data In this study, remote sensing imagery data were used from Landsat 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper Plus (ETM+), and Landsat 8 Operational Land Imager (OLI). All data provided free of charge from the United States Geological Survey website (USGS) [9]. The dates of all images were chosen, in all, to be as closely as possible and in the same climate and vegetation season, and to achieve a homogeneous selection of acquisition dates and conditions of none or little cloudiness. Table 1 shows all information about the used Landsat imagery data.

Auxiliary Data Other geospatial data were used to represent the study area and to validate LULC classification maps. These data include ESRI shapefile administrative boundary of Agadir city, satellite images archive from Google Earth imagery and GP points collected during a field survey using Global Positioning System (GPS) receivers of Mobile Topographer 9.3.2 Apk for Android.

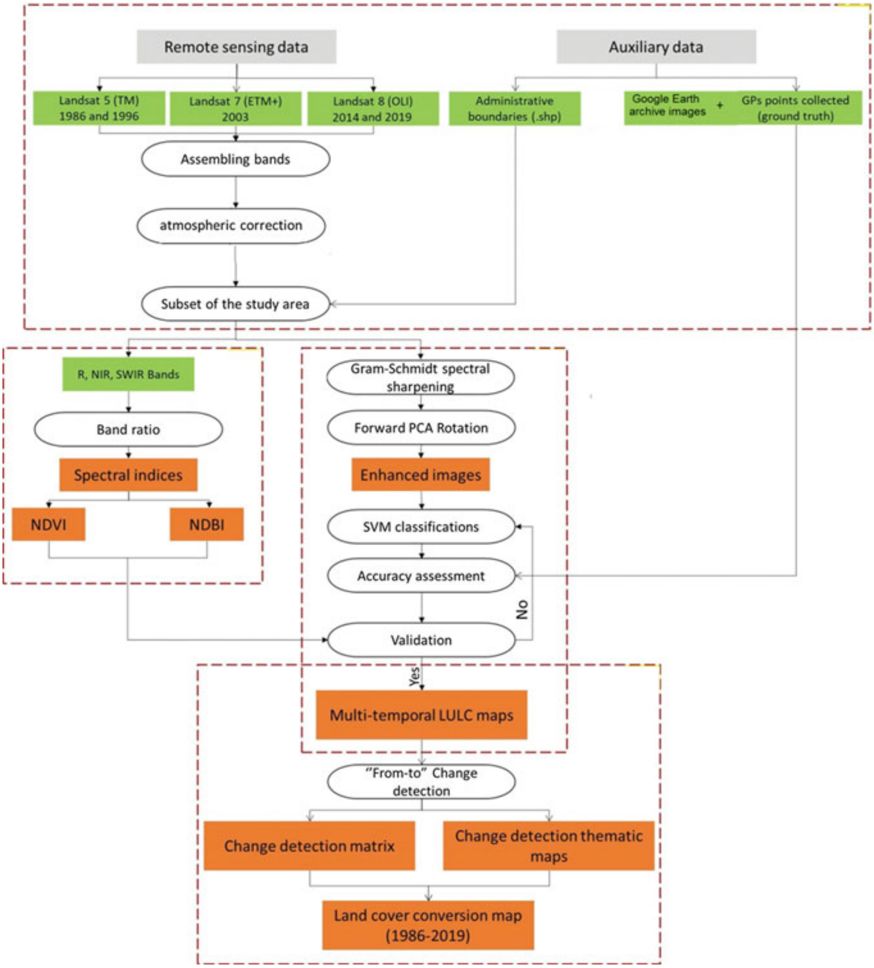


Fig. 2 Flowchart of the study

2.3 *Methodology*

The methodology followed in this study includes database preparation, indices extraction (NDVI and NDBI), classification and accuracy assessment, and change detection analysis. The flowchart used is summarized in Fig. 2.

Database Preparation All Landsat images used in this study were acquired in level 1T (L1T), which are already radiometrically calibrated and orthorectified [9].

First, the digital number (DN) value of each combined bands of each scene was converted to reflectance value by using ENVI 5.3 software. Then, the derived bands were atmospherically corrected by using Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) atmospheric correction model. Then, the corrected bands were clipped and subset to occupy the study area using the geo-referenced outline boundary of Agadir city.

Indices Extraction The use of index-based biophysical parameters to indicate relative abundance of features of interest are widely used in image processing. NDVI (Eq. 1) and NDBI (Eq. 2) were used for improving classification results and assisting land change detection task. However, these indices consist of spectral bands; the Red, Shortwave Infrared (SWIR) and the Near-Infrared (NIR) intervals of the electromagnetic spectrum.

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{red}) / (\mathrm{NIR} + \mathrm{red}) \tag{1}$$

$$\mathrm{NDBI} = (\mathrm{SWIR}–\mathrm{NIR}) / (\mathrm{SWIR} + \mathrm{NIR}) \tag{2}$$

Supervised Classification Implemented to classify the total of surface cover using training samples, collected from knowing points (representative classes) based on supervised machine learning. In this study, Support Vector Machine (SVM) algorithm used due to its ability to handle small training data sets [10] and often produces higher classification accuracy than the traditional methods [10, 11]. The representative LULC classes in our study area are built-up, vegetation, bare land, and water (Table 2).

Several spectral improvements were performed before starting supervised classification. Gram-Schmidt spectral sharpening was applied on prepared data of Landsat 7 ETM+ and Landsat 8 OLI in order to adjust their spatial resolution by using their own panchromatic band. Furthermore, for each scene, the principal component analysis (PCA) was applied to compress the redundant data of the used multibands and only the three first components which have maximum spectral information are used for classification.

Post-classification Change Detection All classified maps have been converted into polygon features (shapefile vector format) in order to evaluate and detect LULC change. The change detection was achieved by using vector spatial analysis (Intersect and Union) tools of ArcGIS 10.3 software.

Table 2 Land cover classes

LULC classes	Description
Urban	Residential, commercial, industrial, transportation, and facilities
Vegetation	Urban vegetation including parks and gardens
Bare land	Open spaces without vegetation or urban cover
Water	Rivers, lakes, or swimming pools

3 Results and Discussion

3.1 NDVI and NDBI Indices

Vegetation and built-up areas of each scene were extracted by using NDVI and NDBI indices (Fig. 3). Commonly, the NDVI values varies between −1 and 1; the values from −1 to 0 represent deep water, barren rocks, sand, or snow while the values between 0 and 1 indicate high to moderate density vegetation [12]. Moreover, the positive values of NDBI indicate built-up areas while negative values represent water bodies [13].

3.2 LULC Classification and Accuracy Assessment

LULC maps produced by SVM classifications were evaluated by using the error matrix, which consists of producer and user accuracies developed to compare randomly chosen test pixels with used ones for classification. For each scene, we have repeated the classification step many time until the best results of accuracy are obtained.

The final version of classified maps is shown in Fig. 4. Due to the heterogenic and the variability of urban cover of the study area; especially in its middle part, we could not obtain higher overall accuracies than the ones presented, even after repeating the classifications many times.

The kappa coefficients and overall accuracies achieved for final LULC maps are given in Table 3. Generally, the overall accuracy of all classified maps is higher than 80%, which means we have achieved the best classification for all used Landsat scenes.

3.3 LULC Change Detection

The statistical analysis of LULC change for the period 1986–2019 in Agadir city is shown in Figs. 5 and 6.

Effectively, during the 33 years, we remark an important growth of built-up class by 98.21% (24.78 km^2). Meanwhile, the bare land was reduced by 1.73 km^2 (6.84%) as compared to 8.11 km^2 (24.95%), 64.09 km^2 (26.68%), and 1.36 km^2 (4.75%) during the interval periods 1996–2003, 2003–2014, and 2014–2019, respectively, with a total amount of 26.84 km^2 (47.18%) for the entire period of the study. The vegetation cover had increased by 27.01 km^2 (422.85%) during the period 1986–1996, then had decreased by 7.98 km^2 (−23.95%), 10.06 km^2 (−39.56%), and 1.78 km^2 (−16.50%), respectively, during the interval periods 1996–2003, 2003–

2014, and 2014–2019. The water bodies have shown a relatively lower change rate during the period 1986–2019 (-0.89 km^2).

These results indicate that in 1986 as well as in 1996, the largest land cover class was bare land. After 2003, a strong growth of the urban land in line with population livelihood and economic activities (industrial and commercial) shows up and spread dramatically and affecting bare land and vegetation cover.

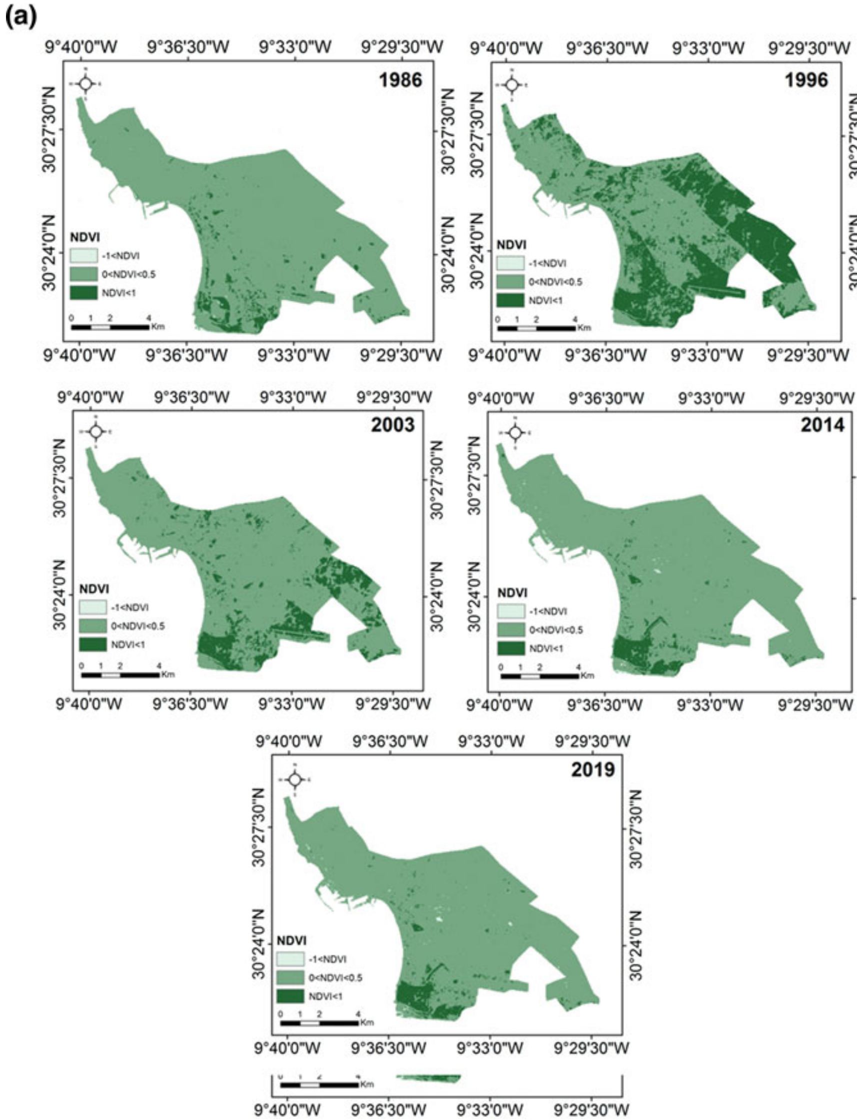


Fig. 3 Spatiotemporal distribution of NDVI (**a**) and NDBI (**b**)

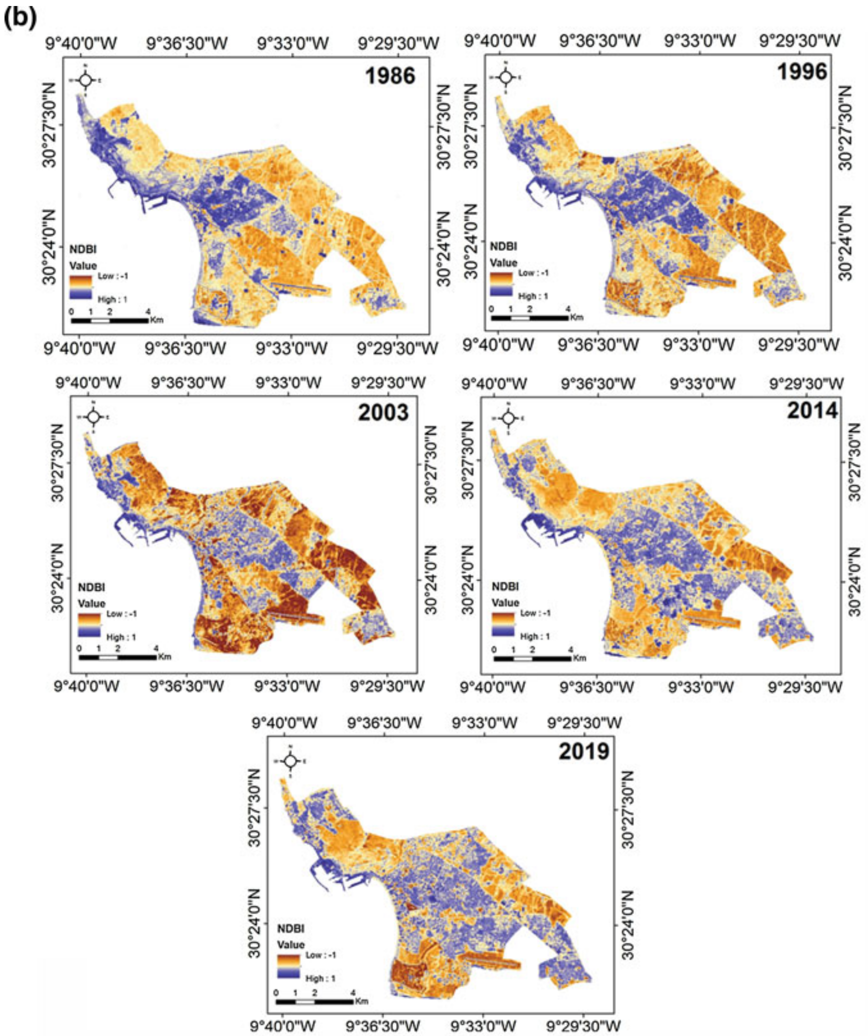


Fig. 3 (continued)

Consequently, we highlighted two clearly recognizable scenarios; (1) the urban area has increased drastically and rapidly, and (2) bare land and vegetation cover have gradually decreased.

Furthermore, LULC change map, for the interval period 1986–2019, indicates that the bare land was the main contributor in shaping urban area of the city followed by vegetation cover (Figs. 7 and 8).

Table 3 The accuracy assessment of classified LULC maps

	Landsat 5 (TM)				Landsat 7 (ETM+)		Landsat 8 (OLI)			
	1986 classification		1996 classification		2003 classification		2014 classification		2019 classification	
LULC classes	Producer (%)	User (%)	Producer (%)	User (%)	Producer (%)	User (%)	Producer (%)	User (%)	Producer (%)	User (%)
Bare land	73.42	97.31	97.85	94.79	96.41	99.77	32.62	97.80	65.79	98.78
Built-up	100.00	57.21	95.94	99.68	99.66	99.91	98.74	89.76	94.14	95.63
Vegetation	72.73	58.18	99.44	98.89	98.57	97.43	86.31	95.39	88.40	98.28
Water	100.00	98.94	0.00	0.00	100.00	99.77	11.11	20.00	37.50	21.82
Overall accuracy	85.45%		89.85%		98.97%		93.56%		81.83%	
Kappa	0.77		0.85		0.97		0.88		0.73	

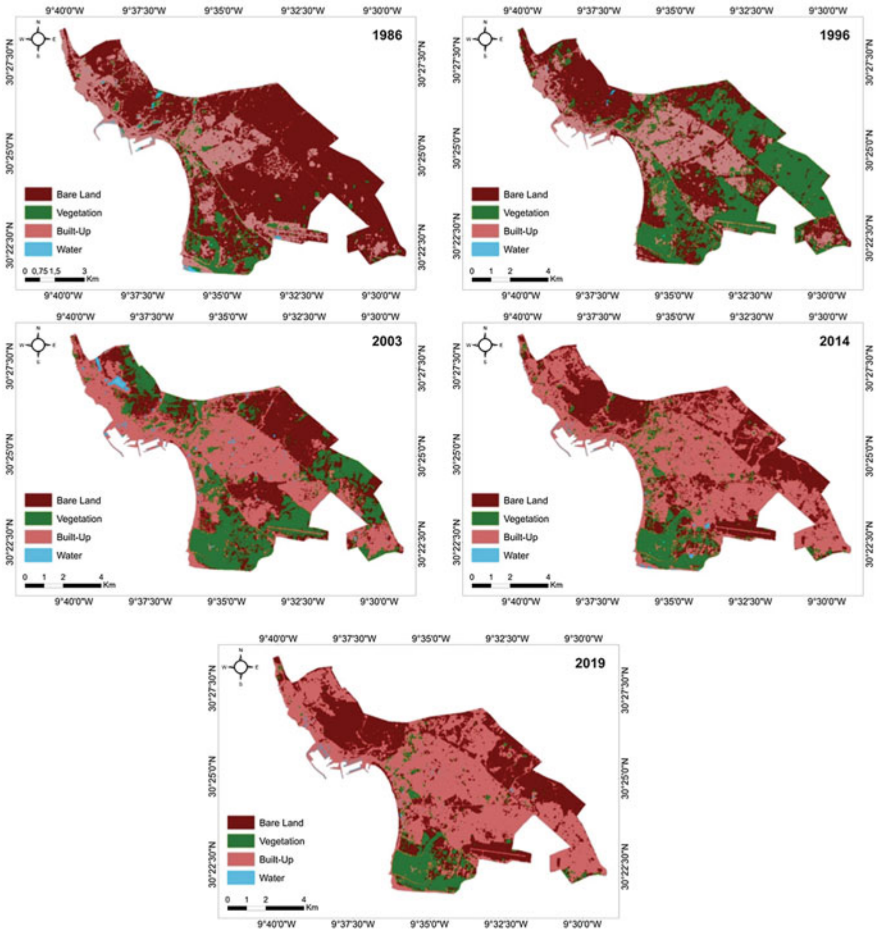


Fig. 4 LULC maps produced by classification process

Commonly, the LULC sever changes dynamics in Agadir city are caused by the urbanization activities and the construction of new urban areas in the early of 1960s after the 1960 earthquake. Moreover, the increase in area for urbanization purposes at the expense of vegetation and bare land, which have most suffered the effects of the expansion of the built-up areas, had many other reasons, such as the demographic crisis caused by the rural exodus of residents in surrounding rural areas and the economic development of the city during the last decades.

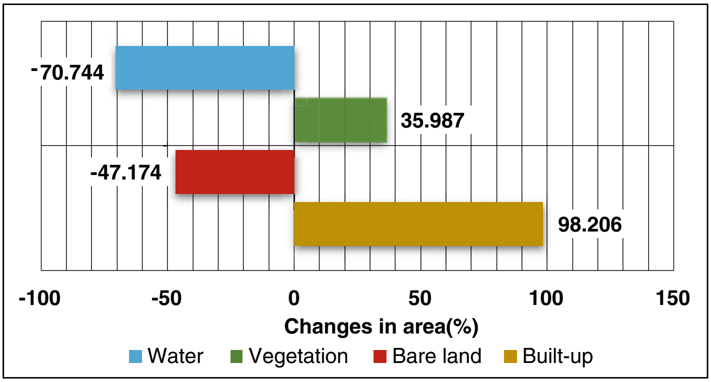


Fig. 5 LULC change area during the interval period 1986–2019

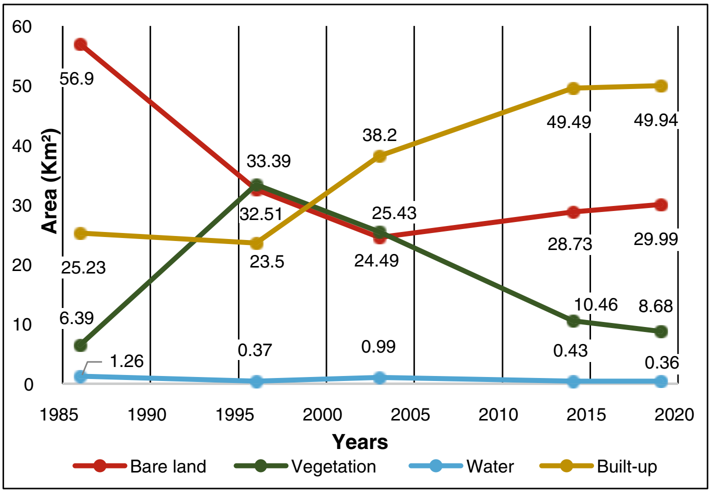


Fig. 6 LULC change scenario during the period 1986–2019

4 Conclusion

The multitemporal Landsat data, for Agadir city in Morocco for the period of 1986–2019, witnessed touchable variation in LULC spatial representation; therefore, it was processed to assess the change dynamics along that period. Several processing steps were conducted to achieve the main goal of the study, including spectral reflectance indices (NDVI and NDBI) extraction, SVM supervised classification, and post-classification comparison.

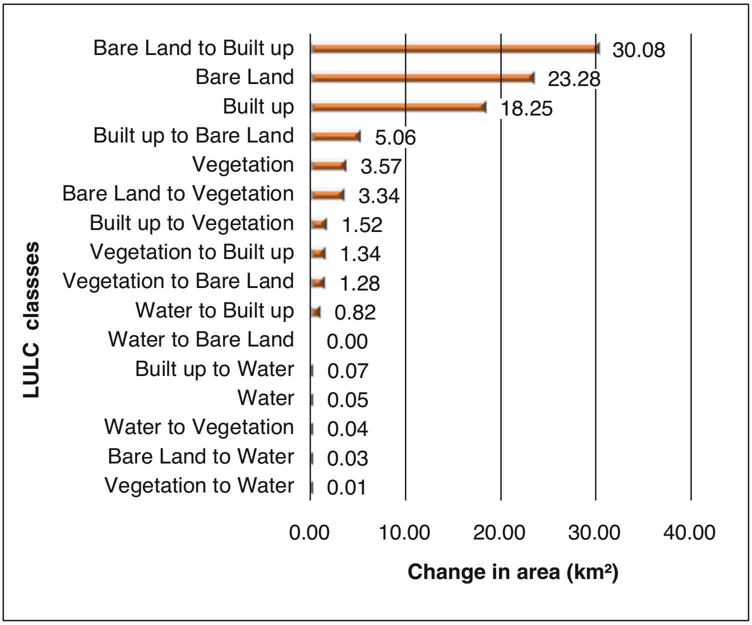


Fig. 7 Percentage of LULC classes conversion for the period 1986–2019

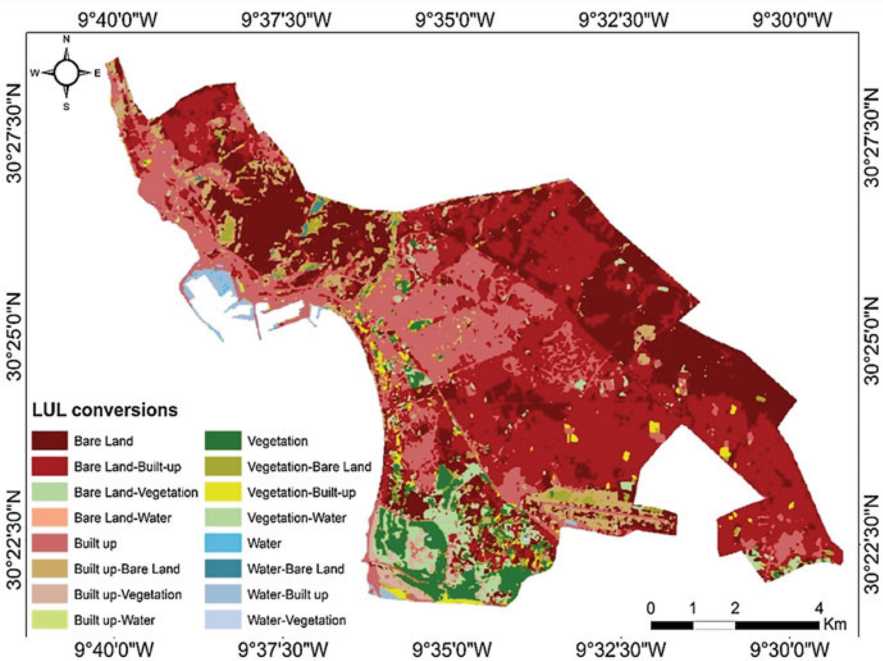


Fig. 8 Land cover conversion from 1986 to 2019

All the classified maps show a high statistical performance in terms of accuracy. The statistical results indicate an intense decreasing and rapidly changing trend in city over the last 33 years. It was found that the urban area has been increased, vegetative area and bare land have been decreased, and water bodies remained constant. The change detection map reveals that the bare land has been the main contributor in shaping urban area of the city followed by vegetation cover.

Eventually, after a successful application of the conducted study, where obtained results, can be an effective tool for land management and ecosystem conservation in the study area for next decade especially if were considered by city council and decision makers.

References

1. Xiuwan, C. (2002). Using remote sensing and GIS to analyse land cover change and its impacts on regional sustainable development. *International Journal of Remote Sensing, 23*(1), 107–124.
2. Fichera, C. R., Modica, G., & Pollino, M. (2012). Land cover classification and change-detection analysis using multi-temporal remote sensed imagery and landscape metrics. *European Journal of Remote Sensing, 45*(1), 1–18.
3. Ahmed, K. R., & Akter, S. (2017). Analysis of landcover change in Southwest Bengal delta due to floods by NDVI, NDWI and K-means cluster with Landsat multi-spectral surface reflectance satellite data. *Remote Sensing Applications: Society and Environment, 8*, 168–181.
4. Bouiadjra, S. E. B., Zerey, W. E., & Benabdeli, K. (2011). Étude diachronique des changements du couvert végétal dans un écosystème montagneux par télédétection spatiale: cas des monts du Tessala (Algérie occidentale). *Physio-Géo. Géographie Physique et Environnement, 5*, 211–225.
5. El Garouani, A., Chen, H., Lewis, L., Tribak, A., & Abharour, M. (2008). Cartographie de l'utilisation du sol et de l'érosion nette à partir d'images satellitaires et du sig idrisi au nord-est du Maroc. *Revue Télédétection, 8*(3), 193–201.
6. Mohajane, M., Essahlaoui, A., Oudija, F., El Hafyani, M., Hmaidi, A. E., El Ouali, A., Randazzo, G., & Teodoro, A. C. (2018). Land use/land cover (LULC) using landsat data series (MSS, TM, ETM+ and OLI) in Azrou Forest, in the Central Middle Atlas of Morocco. *Environments, 5*(12), 131.
7. HCP. *Recensement Général de la Population et de l'Habitat 2014 (RGPH)*. https://rgph2014.hcp.ma
8. *Climate charts Homepage*. https://www.climate-charts.com/Locations/m/FM60250.html
9. *UGSG Homepage*. https://www.usgs.gov/Landsat
10. Mantero, P., Moser, G., Member, S., Serpico, S. B., & Member, S. (2005). Partially supervised classification of remote sensing images through SVM-based probability density estimation. *IEEE Transactions on Geoscience and Remote Sensing, 43*(3), 559–570.
11. Aydda, A., Althuwaynee, O. F., Algouti, A., & Algouti, A. (2019). Evolution of sand encroachment using supervised classification of Landsat data during the period 1987-2011 in a part of Laâyoune-Tarfaya basin of Morocco. *Geocarto International, 34*(13), 1514–1529.
12. Jackson, R. D., & Huete, A. R. (1991). Interpreting vegetation indices. *Preventive Veterinary Medicine, 11*(3–4), 185–200.
13. Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing, 24*(3), 583–594.

Supply Chain Risk Management: Experimentation of the Ishikawa Diagram in Upstream Logistics

El Mehdi Gaou, Salma Gaou, Salah-Eddine Krit, and Noura Metawa

Abstract Nowadays organizations seeking to apply the concept of logistics cooperated, in another way cooperating all functions into a single string, starting with the acquisition of materials until after sales service. And in a competitive environment, all organizations are also sought to create a good image in the minds of customers. Therefore, these have to work on the quality, delay, and cost. In other words, the company must present to customers a cheaper product with good quality and determined delay. Many of studies have been done aimed to eliminate all possible risks in front of the supply chain. Our approach specifically concerns detailed analysis of dysfunctional of various potential hazards that may arise, by determining all risks through a mapping that can determine the severity of risks based on probability percentage. The results are presented in the form of a Ishikawa diagram (also called fishbone diagrams, herringbone diagrams, cause-and-effect diagrams, or Ishikawa), which is a diagram that shows the causes of an event. In light of the results of this analysis, many are problems may face, such as with raw materials, medium, method, measurement, equipment and means through an Ishikawa diagram. To overcome these problems, we have to establish all the recommendations to develop an action plan that we allow to avoid these risks.

Keywords Risk management · SCRM · Risk · Upstream logistics · Ishikawa diagram

El M. Gaou (✉) · S. Gaou · S.-E. Krit
Laboratory Engineering Sciences and Energies, Ibn Zohr University, Agadir, Morocco
e-mail: s.gaou@uiz.ac.ma; s.krit@uiz.ac.ma

N. Metawa
Faculty of Commerce, Mansoura University, Mansoura, Egypt
e-mail: n_metawe@mans.edu.eg

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_31

1 Introduction

The year 2019 was punctuated by a series of disasters. Hurricane Dorian in the Bahamasand some eastern coasts of the USA and their consequences funerals causing wind damage, are plainly illustrate the natural dimension of these disasters. The recall of millions of vehicles around the world by Volvo cars because of a defective component spotted in the engine of some of its models is another case of disaster; this time it is an industrial one.

In addition to natural events that have marked history, the development of human activities through the industrial era was accompanied by the appearance of new dangers. We can list a large number of disasters like mining, train derailment, and plane crash.

We can cite the following also as industrial disasters: the Three Mile Islands nuclear accident or the Chernobyl nuclear accident, the explosion of the factory in Bhopal, disaster factory of SEVESO, etc. [1], and also other disasters in different areas such as financial scandals revealed in the begin of 2000, such as those of Enron and WorldCom, or Lehman Brothers in 2008.

Taking a negative view of the possible consequences of an event on a system (natural system, company, project, etc.), the candying or science of danger proposes to study the sources, characteristics, and consequences of these events. In this context, a number of definitions have been proposed [2]:

- Incident: disruption of a component, unit or subsystem.
- Accident: disruption that affects a system as a whole like a factory, a business, or an industry.
- Risk: a hazard whose occurrence deprives a system of a resource and prevents it from reaching her goals.
- Crisis: surprising event, limiting the time allowed to develop a replica and that threatens their goals.

In addition to a negative aspect, disaster management refers to the urgent need where the intervention takes place.

Nevertheless, not all events give rise to disasters, and it belongs to the different stakeholders to understand their context and environment in order to implement prevention and learning means.

2 State of the Art

2.1 Risk: A Global View

Arabic, Latin, Italian, Spanish, etc. semantically it is difficult to establish precisely the origin of the term "RISK," the word "risk" as we know it today would have appeared in 1557 in a treatise of Henri Estienne [3] as the word feminine. It finally became a masculine word in the seventeenth century ("your money runs high risk," in The Impromptu of Versailles by [4]).

2.2 The Concept

"If something goes wrong, it will happen like that" is the sense of fatality that can characterize the occurrence of an event according to Murphy's Law. This remark makes clear the variety of "risky" situations that extend from the major chemical, nuclear, organizational risks in the company but also to financial and technological risks, as well as the risks associated with the use of the Internet.

This variety of sources, the perception of consequences, even the domain and cultures, and a multitude of definitions have been proposed. In addition to academic literature [5], contributions from organizations such as ISO, AFNOR, DGA, CIRANO, COSO, OGC [6, 7] were carried out.

Of this set of definitions, ISO/IEC Guide 73 can be considered as a commonly used reference. According to him: "Risk can be defined as the combination of the probability of an event and its consequences" [8]. Therefore, a risk combines a probability (for the hazard characteristic) and its consequences. From an outside point of view, Fig. 1 illustrates the general cause–consequence model of risk [5].

However, in some particular areas, the model presented in Fig. 1 can be complemented by notions revolving around risk. For example, in the field of management project, the causes as well as the risk factors can be specified. Indeed, these elements identified are a condition of the internal or external environment of the project whose existence is likely to influence the occurrence of an event and thus to modify the measurement of characteristics of the risk [5, 9].

From a quantitative point of view, the risk is very frequently characterized by the formula:

Criticality (risk) $=$ probability $\times$ impact. The risk can thus be presented in a matrix of risk (see Fig. 2) in two dimensions: probability and impact. It is then possible to lead a comparative multi-risk analysis in the same plan.

The areas determined in the risk matrix can help us not only prioritize risks but also choose management strategies, for example:

Fig. 1 General cause–consequence model

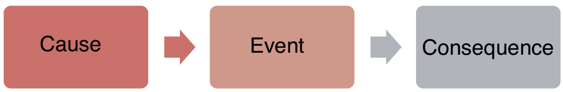


Fig. 2 Probability and impact matrix

- Avoid for risks located in the red zone
- Reduce or transfer: for the risks located n the yellow zone
- Accept: for risks located in the green zone

2.3 From the Notion of Risk to Supply Chain Risk Management

After studying the notion of risk previously in this paragraph will be presented the elements of review of the literature on risk in supply chain management, chain risks logistics, and finally supply chain risk management (SCRM).

Lambert et al. [10] define SCM as "the integration of key business processes from end user through original suppliers that provides products, services, and information that add value for customers and other stakeholders."

These processes do not include only traditional logistics activities such as warehousing, management, inventory, transport, but also non-traditional activities at the logistics such as purchasing, production support, packaging, administration sales, and sales order processing [11].

In addition, the SCM implies the integration, coordination, cooperation, and collaboration between the organizations throughout the supply chain. That means, according to Gimenez and Ventura [12], that the SCM requires integration both internally (intra-organizational) and externally (inter-organizational).

A first definition of risk in the field of supply chain management has been given by March and Shapira [13]. They define risk as "a variation in the distribution of possible supply chain outcomes, their likelihood, and their subjective values."

This definition can be found in this chapter that highlights an important criterion: the probability of occurrence of the risk. If this one is too strong, then the risk is no longer a risk but a certain event to come; if she is too weak, it will only be a chimerical and unfounded fear that managers will not have to look to manage.

Mason-Jones and Towill [14], Zsidisin et al. [15], and Juttner [16] sought to identify and understand the sources of supply chain risk.

Five categories of sources of risk are generally highlighted: the environment, demand, supply, process, and control. The vulnerability of the chain logistics can also be a factor aggravating the risk; it can be defined as "an exposure to serious disturbance arising from supply chain risks and affecting the supply." "It's a good way to serve the end customer market" [14].

Kraljic's [17] work on risk in the context of logistics/procurement shows that the risks exist because of the complexity of the supply market, characterized by the following elements: the shortage of suppliers, the renewal of products and technologies, barriers to entry, logistics costs, complexity, and market conditions of the suppliers (monopoly or oligopoly).

However, two terms are close to the "supply risk chain" and the "supply chain risk," the first is very operational and concerns supplies, deliveries, orders, and highly short-term management flow term. The second is more strategic and transversal and concerns management, implementation place, and organization of flows between the partners of a supply chain; its effects will have consequences on supply risk.

After defining the risk and specifying what constituted the risks associated with the chain logistics, we could present the definition of supply chain risk management (SCRM), proposed by Juttner [16]: "the identification and management of risks for the supply chain, through a coordinated approach to supply chain members, to reduce supply chain vulnerability as a whole."

In the rest of this communication, we will consider that the SCRM is constituted by risk management (and not just management, which implies both strategic and operational dimensions, a horizon of long-term and short-term appreciation term), a risk that may modify or even prevent all or part of the effective and efficient flow of information, materials, and products, or between the actors of a chain internal logistics to an organization, i.e., between the actors of a global logistics chain (between the supplier of the company and the customer of the company).

3 Contribution

3.1 Interest of Our Proposal

It is obvious that the notion of risk in organizations is omnipresent at all levels and concerns all the components of an organization as well as its stakeholders, in particular.

In fact, our study will focus solely on the risks related to the logistics chain and particularly in the upstream phase of the production process, and our study aims to bring a look and analysis.

3.2 Study of Ishikawa Diagram

The Ishikawa diagram will be applied for our example to analyze the root problems that impact the upstream of supply chain in an industrial organization. This simple and synthetic tool makes it possible to diagnose the possible causes of these problems as well as the effects of these risks.

4 Results and Discussion

In this part, we will focus on only a few identified risks in the Ishikawa diagram. Our analysis will focus particularly on the risks related to human and raw material methods, the last point will be the subject of a proposal for a model making it possible to anticipate the supply in an industrial unit and to anticipate any stop of the production in the upstream phase of the logistics chain.

4.1 Risk-Related to Methods

An error can cause disruptions that influence the entire supply chain. For example, an error in parameter setting ranges can create difference in the consumption of the material and at the same time an handicap for suppliers. In addition, at the planning level, any increase in the rate of production without the provision of the necessary information to the supplier can lead to unforeseen breaks in the production cycle.

4.2 Labor Force Risk

In a competitive context, for-profit organizations seek to convey a good image in the marketplace. As a result, they must work on the three performance axes: Quality/Cost/Delay.

To achieve these three reports, that is to say, to ensure good quality by optimizing the costs in a very short time, it is necessary to motivate and supervise the staff in their activities and tasks.

Labor plays a very important role in carrying out one or more tasks in an efficient and manageable way, because of good coordination, health and safety awareness, compliance with regulations and quality charters and specific training on positions, the availability of the hands of works, etc. All these elements are able to master risks related to the workforce.

4.3 Risks Related to Raw Materials

Currently, organizations are seeking to apply the concept of cooperative logistics, in other words by cooperating all the functions in a single chain, starting with the acquisition of the materials until the after-sales service.

Adopting a good raw material purchasing strategy is a very important element in meeting the needs of stakeholders. Otherwise, it could lead to malfunctions even stop production, the correlation of the raw material-production equation is therefore a priority.

In the same way, the organization uses the Just-In-Time (JIT) method to have the least amount of money locked in and applies the First-In-First-Out (FIFO) method to avoid the risk of oxidation of raw materials.

4.3.1 Proposal for a Supply Plan Improvement Model in an Industrial Production Structure

According to the JIT principle, our study will be based on a daily service rate that allows visibility on the receipt of the quantities requested while respecting the specific day in the schedule shared with the supplier of the raw materials.

The daily service rate or supplier service rate is an indicator that makes it possible to calculate an index in relation to the number of supply orders delivered under the conditions requested compared with the total order number.

From a database provided to us which makes it possible to calculate the service performance index rendered by the supplier as follows:

– Determination of the quantity to be delivered:

$$= \text{Need–Quantity delivered}$$

Calculate the Percentage Difference

$$= \text{IF (actual receipt date} > \text{expected date; } -100\%\text{; still to be delivered/need)}$$

At this level, if a delivery is received after the scheduled date, it will be considered as a delivery delay (deviation of -100%) despite if the quantity is the same as the expressed need.

On the other hand, if the calculated deviation is less than a tolerance interval of -12.5%, it will be considered a partial delivery.

Automation of the Remark Displayed to Users

$$= \text{IF}\left(\left(\text{Spread} = -100\%\text{; ``Delayed Delivery''; IF }\left(\text{Quantity delivered} < \text{Requirement}-\right.\right.\right.$$
$$\left.\left.\left.(\text{Requirement} * 0.125)\text{ , ``Partial Delivery,'' ``OK''}\right)\right)\right)$$

The remark displayed makes it possible to provide the following table users with the actual situation on the quality of service by the supplier.

The normal situation is presented by a simple comment in the form "OK." On the other hand, if the difference is -100%, it is possible to post a comment as "late delivery," in addition if the delivery is less than the quantity taking into account a tolerance index of -12.5%, the message will be displayed as "partial delivery."

Determination of the Lines where the Delay Exists

At this level, a partial delivery or a delivery delay will be represented by a -1 index which will allow us to know the number of items delivered without respecting the rules drawn by the supply service by the application of the formula following:

$$= \text{SI }\left(\text{Note} <> \text{`` ok''; } -1\text{; } 0\right)$$

Référence	Besoin	Quantité livrée	Date d'arrivée	RAL	Ecart en %	Remarque	Retard	Profondeur de retard
0062208000414	3,50 t	3,43	21/04/2014	-0,07	-2%	OK	0	0
0310015000108	4,26 t	4,02	21/04/2014	-0,243488	-6%	OK	0	0
0243020000414	4,94 t	4,95	21/04/2014	0,01	0%	OK	0	0
1150017000414	42,36 t	41,52	21/04/2014	-0,837	-2%	OK	0	0
0351015000446	6,92 t	8,47	21/04/2014	1,55176	22%	OK	0	0
0255030000103	25,00 t	16	21/04/2014	-9	-36%	Livraison Partiell	0	0
0595012000203	3,19 t	2,95	21/04/2014	-0,242	-8%	OK	0	0
0638012000414	5,00 t	4,6	23/04/2014	-0,3968	-100%	Retard de livrais	-1	2
0820009500203	11,37 t	13,56	17/04/2014	2,1853536	19%	OK	0	0
1200009500414	13,10 t	18,96	18/04/2014	5,86	45%	OK	0	0
0770025000119	13,44 t	11,52	19/04/2014	-1,9168	-14%	OK	0	0
1500012500414	36,04 t	34,2	20/04/2014	-1,836	-5%	OK	0	0
1060012500414	10,92 t	11,72	21/04/2014	0,8	7%	OK	0	0
0062706500414	3,50	Décaler S19						
0426006500203	3,37 t	3,95	22/04/2014	0,5764	17%	OK	0	0
0239015000203	5,42 t	7,81	22/04/2014	2,39128	44%	OK	0	0
0540015000414	7,45 t	8,39	22/04/2014	0,94	13%	OK	0	0
0556009000203	16,47 t	17,94	22/04/2014	1,46532	9%	OK	0	0
0433020000108	4,95 t	4,41	22/04/2014	-0,544725	-11%	OK	0	0
0512020000203	16,18 t	19,15	25/04/2014	2,9668	-100%	Retard de livrais	-1	3
0459015000203	5,61 t	7,27	17/04/2014	1,66	30%	OK	0	0
1490012000414	13,30 t	12,44	17/04/2014	-0,864	-6%	OK	0	0
1500006500414	10,41 t	11,6	17/04/2014	1,1909	11%	OK	0	0
1160009500203	9,16 t	9,34	22/04/2014	0,1782	2%	OK	0	0
0960015000108	16,00 t	14,92	24/04/2014	-1,08	-100%	Retard de livrais	-1	2
0124015000414	,62 t	0,98	22/04/2014	0,356	57%	OK	0	0
0320009500414	1,25 t	2,97	28/04/2014	1,722	-100%	Retard de livrais	-1	5
0104020000108	2,12 t	Décaler S19						
0306014000446	4,65 t	3,97	24/04/2014	-0,684	-100%	Retard de livrais	-1	1
0270015000414	8,10 t	10,34	23/04/2014	2,24	28%	OK	0	0

Fig. 3 Board: daily service rate

Calculation of the Delay Depth

This formula calculates the number of days of delay until the actual receipt of the late references by comparing the expected date of receipt with the current date using the Excel function TODAY ().

Find below an example of our realized action (Fig. 3):

The purpose of this file is to have a general summary to know the overall situation (Fig. 4):

The green bar represents the percentage of items delivered without delay, and the red bar signifies the change in the lag depth of each week.

5 Conclusion

This experiment is the first step of my proposal since by conceptualizing the principle, it will allow us to rely on this analysis to think about the implementation of an application to better detect and control risks in the upstream phase of the chain logistics.

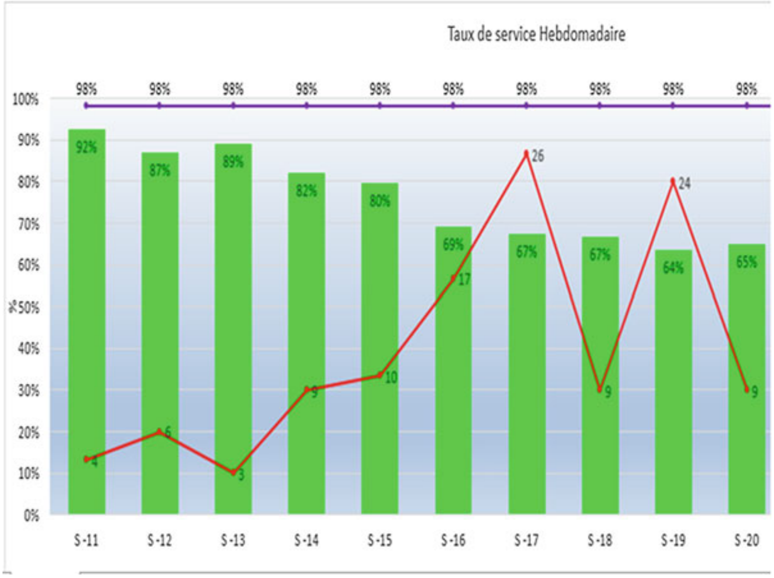


Fig. 4 Graph summary of daily service rate

References

1. Lagadec, P. (1981). *The major technological risk*. Pergamon Press.
2. Wybo, J. L. (1998). *Introduction aux cindyniques*. Ed. Eska.
3. Estienne, H. (1578). *Two dialogues of the new French language Italianized and otherwise.*
4. Lannoy, A. (2008). *Control of risks and dependability: Historical benchmarks and Methods* (Risk and Danger Science Series, Summary Notes Series and research).
5. Gourc, D. (2006). Towards a general model of risk for steering and conducting goods and services activities. *Habilitation to Lead Research, Toulouse, France: National Polytechnic Institute of Toulouse Transports Focus, 5*(2), 68–69.
6. ISO 31000. (2009). *International standards for business, risk management - principles and guidelines*. Guidelines to Quality in Project Management.
7. X50-117 AFNOR. (2003). *Risk management of a project*.
8. ISO Guide 73. (2002). *Risk management Vocabulary: Guidelines for use in standards*.
9. Ravalison, R. B. (2006). *Staging of information system projects!: Contributions to risk control in information system projects*. Doctoral thesis, National Polytechnic Institute of Toulouse, February 1.
10. Lambert, D., Cooper, M., & Pagh, J. (1998). *Supply chain management: Implementing issues and research opportunities*.
11. Tracey, M., Lim, J.-S., & Vonderembse, M. A. (2005). The impact of supply-chain management capabilities on business performance. *Supply Chain Management, 10*(3/4), 179–191.
12. Gimenez, C., & Ventura, E. (2005). Logistics-production, logistics-marketing and external integration: Their impact on performance. *International Journal of Operations and Production Management*.
13. Shapira, Z. (1987). Managerial perspectives on risk and risk taking. *Management Science, 33*(11), 1404–1418.

14. Mason-Jones, R., & Towill, D. R. (1998). *Shrinking the supply chain uncertainty cycle* (pp. 17–22). Logistics Systems Dynamics Group, Cardiff University.
15. Zsidisin, G. A., Panelli, A., & Upton, R. (1999). Purchasing organization involvement in risk assessments, contingency plans and risk management: An exploratory study. *Supply Chain Management; An International Journal, 5*(4), 187–197.
16. Juttner, U. (2005). Supply chain risk management. *International Journal of Logistics*.
17. Kraljic, P. (1983). Purchasing must become supply management. *Harvard Business Review, 61*(5), 109–117.

Model for Identifying the Most Important Student's General Competencies Based on Expert Data Processing

Pavlo Hryhoruk, Svitlana Grygoruk, Nila Khrushch, and Iryna Romanets

Abstract *Research goals and objectives*: Developing model to process expert data in order to identify the most important general competences for including them into the academic programs.

Subject of research: Identifying the list of most importance general competences by the processing of expert data for creation of academic programs for training specialists in economic cybernetics and applied mathematics.

Research methods: Comparative analysis, business analytics, expert data processing, methods for calculation of eigenvalues and eigenvectors.

Results of the research: The formulation of the normative content of the training specialists in the academic program should take place in terms of competences and learning outcomes. Stakeholders should be involved in the development of academic programs. This will increase the quality of the training specialists and will allow taking into account the needs of the regional labor market. Processing of expert data on the survey results of representatives of the Union of Industrialists and Entrepreneurs of the Khmelnytskyi region allowed to identify the most important general competences for the training of specialists in economic cybernetics and applied mathematics. These results were used in the design of relevant academic programs.

Keywords Model for identifying competences · Competence-based education · General competences · Academic program · Expert data processing · Eigenvalues · Eigenvectors · Measure of competence importance

P. Hryhoruk (✉) · S. Grygoruk · N. Khrushch
Khmelnytskyi National University, Khmelnytskyi, Ukraine

I. Romanets
Union of Industrialists and Entrepreneurs of Khmelnytskyi Region, Khmelnytskyi, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_32

1 Introduction

The autonomy of higher education institutions (HEIs), defined by the Law of Ukraine “On Higher Education” [1], determines the formulation of normative content of training of specialists in terms of competences and learning outcomes. This defines the content and structure of the academic program, which is a single set of educational components planned and organized to achieve the identified learning outcomes and to shape needed competences [2]. The development of academic programs should be based on the competence-based approach with taking into account the requirements to the specialists, proposed by the European Commission’s International Project “Tuning Educational Structures in Europe” (TUNING) [3, 4]. On the other hand, such training should take into account the needs of the regional labor market, the requirements of which must be reflected in the shaping of basic competencies of graduates.

Therefore, HEIs should work closely with all stakeholders to determine the content of learning, ensuring the quality of educational services that meet the market challenges and the needs and expectations of students, training of competitive professionals able to meet targets.

The purpose of this chapter is to develop model for processing expert data in order to identify the most important general competences for including them into the academic programs.

This chapter is organized as follows: Sect. 2 describes related works on competence-based education; Sect. 3 provides materials about approaches for expert data processing; Sect. 4 contains results of identifying most important general competences based on expert data processing; the last section consists of conclusions and direction of further studies.

2 Literature Review

The desire of Ukraine to integrate not only EU economic organizations but also in the educational area imposes significant commitments. Ukraine’s accession to the Bologna Declaration in May 2005, changes in public policy in the field of higher education, approval of the National Qualifications Framework [5], Laws of Ukraine “On Higher Education” [1], “On Education” [2] confirm improvement in this area.

A main concept of the Bologna process is the quality of higher education. A key innovative idea of modern education is the competence-based approach. As stated in the key document of Yerevan Ministerial Conference in 2015 (Yerevan Communiqué) [6], the main mission of the European Higher Education Area (EHEA) is to improve the quality of all components of educational process that should facilitate a closer link between learning and research at all levels of education, provide incentives for institutions, teachers, and students, to intensify activities aimed on developing creativity, innovation, and entrepreneurship. Academic programs for training specialists should be aimed on the formation and development of those competencies of students that will allow the best way to satisfy both their personal ambitions and the needs of society. They should maintain a transparent description of learning outcomes and student workload, flexible learning trajectories, and appropriate methods to assess learning outcomes. In addition, students, as full members of the academic community, as well as other stakeholders, must be actively involved in the development of curricula and in quality assurance procedures. Key competences for lifelong learning are identified and substantiated in [7].

The issues of quality assurance of the training of specialists according to the new paradigm of the organization of the educational process are crucial in the EHEA and, accordingly, reflected in the program documents of the conference of European ministers responsible for higher education like communique and declarations.

The European higher education quality assurance system is outlined in the "Standards and Guidelines for Quality Assurance in the European Higher Education Area" (ESG 2015) [8]. This document sets out a list of European standards for higher education quality assurance.

A normative document that regulates the content of the professional training of bachelors and masters in a separate specialty is the academic program. Harmonization of academic programs with the national qualifications framework involves a process of alignment, which establishes rules for identifying key competences and developing program's learning outcomes.

In accordance with the Law of Ukraine "On Higher Education" [1], the necessary component of the academic program is a list of competences. Despite the significant results obtained in the execution and implementation of the projects TUNING, CoRe 2, AHELO [3, 9, 10], the mentioned topics continue to attract the attention of a wide range of scholars. In particular, the article [11] presents the results of a survey of representatives of IT sphere on the necessary IT competences for both entry-level specialists and IT professionals. Specific IT competences are defined through answers to questionnaires. The presented results identify those competences that are considered to be critical, useful, or not relevant to academic programs related to IT area. The authors find that professional competences like professional ethics, oral and written communication, and problem solving are more general and can be applied to almost all directions in the field of information technology. On this basis, the importance of forming written / oral communication skills, professional ethics, teamwork, problem solving, and time management is emphasized, competency in the development of software. Instead, the importance of professional competencies is determined by the field of expertise. In particular, programing skills can be critical to programmers, but are less important for network administrators or IT managers.

Mambo G Mupepi [12] justifies the need to involve communities of practice to identify a set of competences when designing educational activities. The author has shown that such communities can also be important partners in establishing the required level of specialists for various industries.

The authors of [13] consider the impact of various factors on competency-based education (CBE) in higher education, in particular, changes in demographics, public and employer demands, shift of technologies and many others. The article also study the barriers and challenges faced by higher education in the implementation of this approach, as well as the areas and prospects for its development in the future.

In [14], features of implementing the BCE for adult students have been studied. In particular, it is argued that such students have high motivation, are ready for learning, and are rich in life experience, and they often just need the opportunity to test the own knowledge and skills already acquired in their lifetime. In addition, considered approach offers flexible educational trajectories that adults often need. Among the problems authors noted, adult students concern about the ability to succeed and opportunities for continuing education.

The issues of determining a list of competences and their assessment in the designing of academic programs were also not overlooked. The papers [15, 16] present approaches to use mathematical methods in the analysis of the educational context based on the identification and classification of competences and learning outcomes. They are based on courses' organizational documents like syllabi and assessment procedures.

Journal [17] contains a set of articles reflecting the issues related to the academic programs and techniques aimed at developing student's competences. In particular, the effect of influence of tuning project and other related projects on the change in the learning paradigm is described. Also in these studies was presented an overview of innovative changes in the field of engineering education in China in recent years, the approach to establish relationship between assessment at the separate courses level and levels of the program as a whole, and many other issues related to the CBE. Analysis of the implementation of the EU best practices in the educational process of Ukraine's HEIs and the identification of key competences for appropriate academic programs is presented in studies [18–20].

3 An Approach for Processing of Expert Data

Studied objects are often considered as components of weakly structured systems, which cannot be unambiguously described and formalized. An important role in the evaluation of such objects is heuristic techniques including expert methods. One of their most significant advantages is the ability to obtain unique information that cannot be obtained by a statistic way.

The result of an expert assessment of the investigated objects using a quantitative measure is the matrix expert scoring:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & & & \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}, \tag{1}$$

where x_{ij} is scoring for ith object obtained from jth expert, $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$, where m is the number of assessed objects, n is the number of experts.

The purpose of further expert assessment processing is to obtain generalized data and new information contained in the hidden form. Depending on the purposes of expert evaluation in processing the results of the survey, the following main tasks arise:

- Calculation of a generalized assessment of objects and the design of their generalized ranking.
- Identifying the measure of experts' qualifications.

When matrix of the X like (1) is obtained by the quantitative approach, generalized evaluation of objects provided that the experts have the same qualifications, can be calculated as a sample mean:

$$x_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}, \tag{2}$$

where x_i is a generalized evaluation of ith object, $i = 1, 2, \ldots, m$.

On the basis of the obtained average marks, there is a ranking of objects in accordance with the ordering of these estimates in descending order (or growth, depending on the meaning of the assessment).

If it is desirable or necessary to take into account expert qualification (competence), objectivity, and knowledge, we use weighting coefficients q_j for each jth expert, which reflect the relative level of competence of each expert. Group assessment x_i for the ith object is calculated by the formula:

$$x_i = \sum_{j=1}^{n} q_j x_{ij}. \tag{3}$$

The calculation of expert's qualification coefficients and the generalized assessment of the objects taking into account the competence of experts based on the posteriori information presented in the form of a matrix X like (1), may be carried out using the following recurrent formulae:

$$x_i^{(t)} = \sum_{j=1}^{n} x_{ij} q_j^{(t-1)}, \quad i = 1, 2, \ldots, m, \tag{4}$$

$$\lambda^{(t)} = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} x_j^{(t)}, \tag{5}$$

$$q_j^{(t)} = \frac{1}{\lambda^{(t)}} \sum_{i=1}^{m} x_{ij} x_i^{(t)}, \quad j = 1, 2, \ldots, n, \tag{6}$$

where $t = 1, 2, \ldots$. Initial values of expert's qualification coefficients considered as equal: $q_j^{(0)} = \frac{1}{n}, \quad j = 1, 2, \ldots, n$.

After the transformation, Eqs. (4) and (6) are written in the matrix form as follows:

$$E^{(t)} = \frac{1}{\lambda^{(t-1)}} B E^{(t-1)}, \tag{7}$$

$$Q^{(t)} = \frac{1}{\lambda^{(t)}} C Q^{(t-1)}, \tag{8}$$

where $E^{(t)}$ is a vector of generalized objects' assessments on the tth step; $Q^{(t)}$ is a vector of expert's qualification coefficients on the tth step; $B = [XX^T]$, $C = [X^T X]$, X is a matrix of initial expert assessments of objects.

For $t \to \infty$, formulas (7) and (8) will take the form (with a certain modification):

$$BE = \lambda_B E, \tag{9}$$

$$CQ = \lambda_C Q. \tag{10}$$

That is, the components of the vectors E and Q can be calculated as the eigenvectors of the matrices B and C that correspond to the first (largest modulo) eigenvalues of the appropriate matrices.

This approach will be used to evaluate the importance of general competences based on expert data.

4 Processing of Expert Assessment of the Importance of Competences

In order to take into account, the needs of the regional labor market during the formation of the of academic program content for training specialists in the Khmelnytskyi National University, an expert survey of the stakeholders was conducted. Experts were members of the Union of Industrialists and Entrepreneurs of the Khmelnytskyi Region. The survey was conducted in 2016 within the framework of the implementation of the TEMPUS-JPGR ALIGN project "Achieving and checking the alignment between the academic programs and qualifications frameworks" [21] and concerned two academic programs for the training of specialists in economic cybernetics and applied mathematics.

They have in common that they are focused on the training of specialists in the field of mathematical modeling of socio-economic systems and the use of information and communication technologies. The choice of these programs was also due to the fact that, first, the project provided for a pilot assessment of the quality of these programs by the expert commission of the EU partners, and second, Khmelnytskyi National University was the only one in the Khmelnytskyi Region, which is preparing bachelors and masters in these academic programs. The purpose of the survey was to identify the most important competences that should be developed at the graduates of these programs. A questionnaire was developed for the survey. Respondents were asked to select five of the most important general competences from the list of ones identified within TUNING project. Experts should provide them with appropriate scores from 1 to 5, where the value of 5 corresponds to the most important competence and the value of 1 to the least important one. The remaining competences that were not selected by experts were set to zero.

Thirty-two respondents were involved in the survey (by the number of questionnaires that were returned for processing). According to the survey results, the matrix of expert assessments was formed.

Expert data processing showed the following results. It turned out that nine competences from the proposed set of ones were not selected by any of the experts and did not receive any score respectively. They were excluded for further analysis. In particular, among such competences were:

- Ability to motivate people and move toward common goals.
- Ability to design and manage projects.
- Ability to work in an international context.

In our view, with regard to the absence of the first two competences, one can conclude that, since primary positions of graduate are not guiding, stakeholders did not attach importance to the formation of these competences during studying in higher educational institutions. Probably, they consider that these competences may be successfully formed during professional activity. Insufficient attention to the ability to work in an international context can be explained by the lack of stable commercial and industrial relationships between entrepreneurs of Khmelnytskyi region and foreign partners.

Figure 1 shows a diagram that illustrates the frequency of choice of the remaining 22 competencies by experts (without taking into account the score that experts assigned to competence).

The analysis of the figure shows that experts often chose the following competences:

- Ability to search for, process, and analyze information from a variety of sources (21 times).
- Ability to identify, pose and solve problems (19 times).
- Ability for abstract thinking, analysis, and synthesis (16 times).
- Capacity to generate new ideas (creativity) (14 times).
- Ability to undertake research at an appropriate level (13 times).

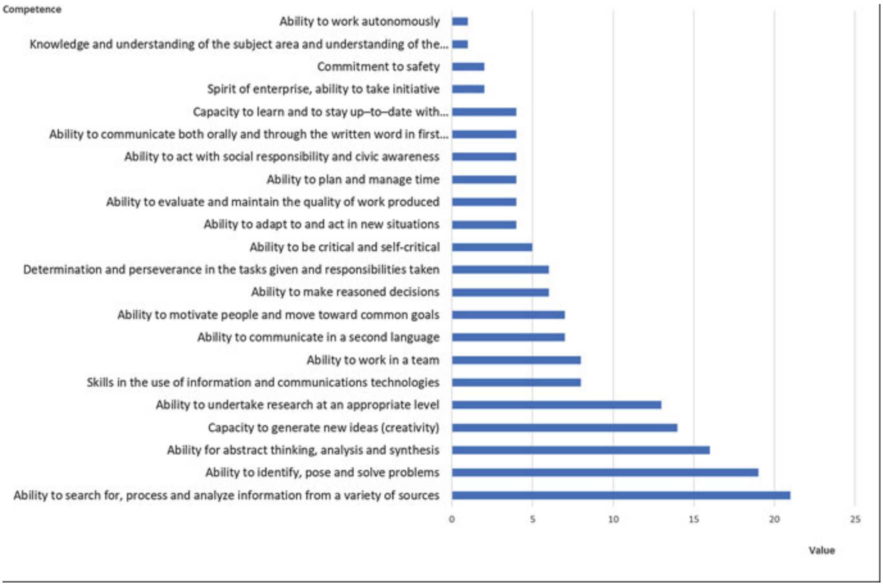


Fig. 1 Frequency of competencies selected by experts

It can be concluded that the stakeholders give main advantage to the competences related to ability to solve problems and make decisions, generate new ideas, work with different sources of information, ability to think critically, analyze, synthesize, undertake research. At the same time, the competence associated with the formation of skills in the use of information technology is only in the sixth place. Probably, experts believe that such skills should be formed at the pre-university stage, and students already possess them sufficiently.

Now consider the importance of competences by the value of the total amount of points. Figure 2 shows the corresponding diagram.

The analysis of the results shows that competences “Capacity to generate new ideas (creativity)” and “Ability to identify, pose and solve problems” were exchanged places in the list of the most important competencies, but they also ranked among the five most important ones. For the rest of the competencies if there are changes, then they are minor. Consequently, these results confirm the previous conclusion about the list of most important general competences.

Let us calculate the assessments (measures) of the importance of competences, considering the degree of experts’ qualifications and knowledge in the field of study are the same. To do this, consider the matrix $B = [XX^T]$, where X is the matrix of the initial expert scores for competences. In the next step, we calculate the first eigenvalue of matrix B, the largest one in absolute value, and the normalized eigenvector corresponding to this value. The components of this vector reflect the importance of relevant competences. Arranging the components of the vector in descending order of values, we obtain the following results regarding the importance of competences, that are shown in Fig. 3.

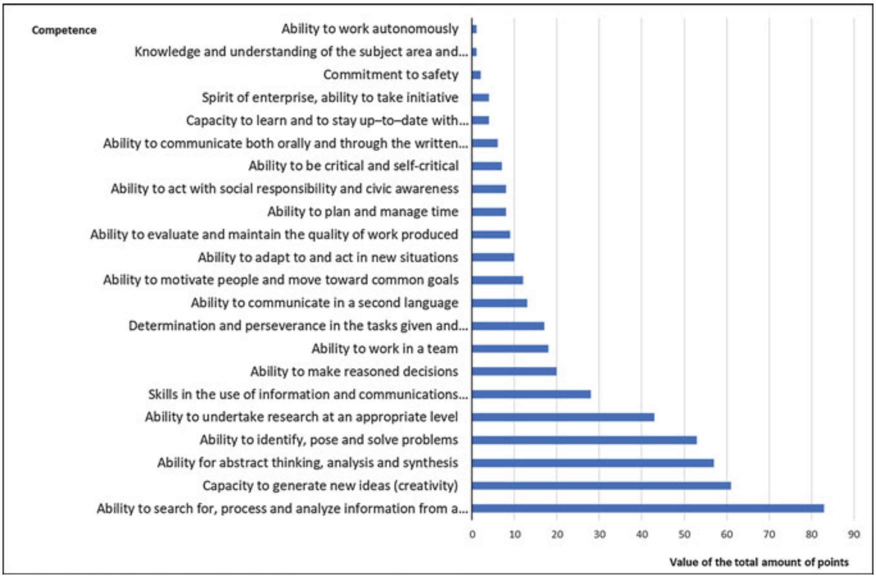


Fig. 2 Importance of competences by the value of the total amount of points

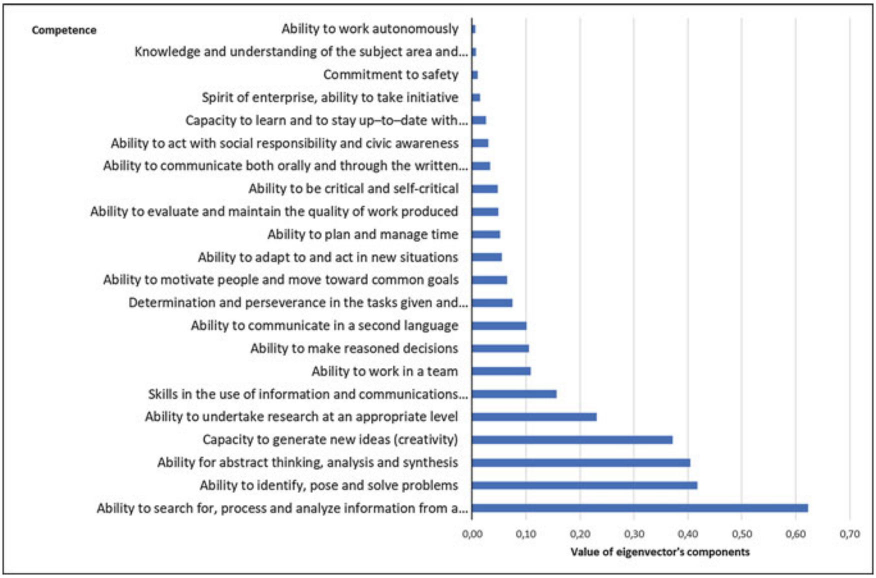


Fig. 3 The importance of competences by the value of the eigenvector's components

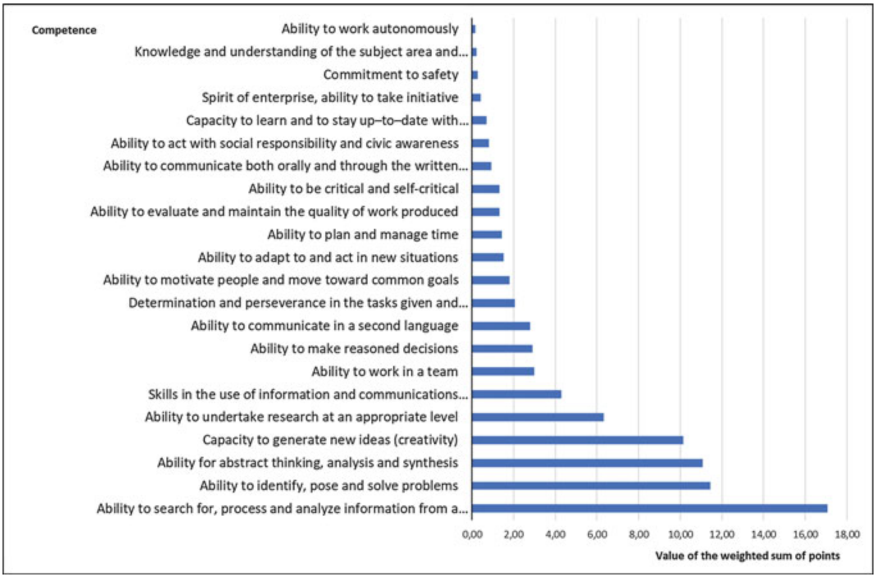


Fig. 4 Importance of competences by the value of the weighted sum of points

As it follows from the analysis of Fig. 3, the list of the first five competences ordered on the significance corresponds to the result obtained in the analysis of the frequency of the choice of competencies by experts (see Fig. 1). We note, as in the previous case, for other competencies, there are also minor changes in the assessment of their importance.

Let us calculate generalized estimates of general competences using the formula (3), taking into account the expert's qualification coefficients. These coefficients are obtained as components of the eigenvector, which corresponds to the first eigenvalue of the matrix $C = [X^T X]$, that is most in absolute value, where X is the matrix of the initial expert scores of the general competences. The importance of general competences in terms of the weighted sum of points calculated by formula (3) is shown in Fig. 4.

We note that the order of general competences is fully consistent with the one obtained in the preceding case.

Thus, it can be argued that according to the opinion of experts of the Union of Industrialists and Entrepreneurs of the Khmelnytskyi region, the training specialists in economic cybernetics and applied mathematics should primarily focus on the formation of the following competences:

- Ability to search for, process, and analyze information from a variety of sources
- Ability to identify, pose, and solve problems
- Ability for abstract thinking, analysis, and synthesis
- Capacity to generate new ideas (creativity)
- Ability to undertake research at an appropriate level

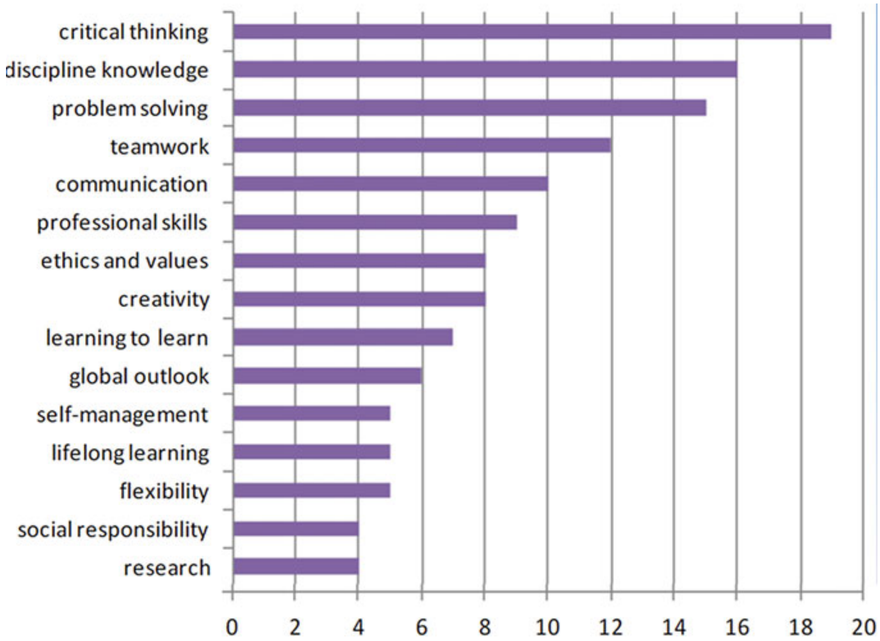


Fig. 5 Types of learning outcomes (fragment). (Source: [22, p. 55])

These results are consistent with those obtained by processing expert data on the selection of learning outcomes within the AHELO project (Fig. 5) [22].

The results of identifying the most important general competences were reflected in the academic programs, which are used in the educational process at the Khmelnytskyi National University for the training specialists in economic cybernetics and applied mathematics.

Also, these results were taken into account by increasing the workload for the courses "Fundamentals of Research" and "Decision-Making Systems" for the curricula of the bachelor's level, as well as the implementation of a mandatory course "Methodology and Organization of Research" for training masters, in which students should compose paper on the results of their research.

5 Conclusions and Outlook

The main scientific result of the study is the identification of a list of the most important general competences for the training specialists on academic programs of economic cybernetics and applied mathematics at the Khmelnytskyi National University. The result is obtained through the expert data processing. Members of the Union of Industrialists and Entrepreneurs of the Khmelnitsky region were experts. The used mathematical tools may be applied to solve similar problems in designing academic programs for other fields, processing of the survey data obtained from other groups of experts, in particular, students, teachers, etc. The directions of further studies are conducting an expert survey of stakeholders with the further data processing to identify the most important professional competences, taking into account the content of the new higher education standards of Ukraine.

References

1. *Law of Ukraine "on higher education"*. http://osvita.ua/legislation/law/2235/
2. *Law of Ukraine "on education"*. https://zakon.rada.gov.ua/laws/show/2145-19
3. *A tuning guide to formulating degree programme profiles*. http://www.core-project.eu/documents/Tuning_Guide_Publicada_CoRe.pdf
4. *Tuning Educational Structures in Europe*. http://www.unideusto.org/tuningeu/
5. *National qualifications framework*. http://zakon5.rada.gov.ua/laws/show/1341-2011-%D0%BF
6. *Yerevan communiqué*. http://www.ipq.org.ua/upload/files/files/03_Novyny/2015.05.15_Ministerial_Conference/Yerevan%20Communique.pdf
7. European Commission. (2005). *Common European principles for teacher competencies and qualifications*. European Commission. http://www.pef.uni-lj.si/bologna/dokumenti/eu-common-principles.pdf
8. *Standards and guidelines for quality assurance in the european higher education area (ESG)*. https://enqa.eu/wp-content/uploads/2015/11/ESG_2015.pdf
9. *AHELO reports and documents*. http://www.oecd.org/education/skills-beyond-school/ahelodocuments.htm
10. CoRe Projects. *Competences in education and recognition*. Guide. http://www.core-project.eu/core%202/guide.aspx
11. Kovacs, P. J., Davis, G. A., Caputo, D. J., & Turchek, J. C. (2005). Identifying competencies for the it workforce: A quantitative study. *Issues in Information Systems, VI*(1), 339–345. http://iacis.org/iis/2005/Kovacs_Davis_Caputo_Turchek.pdf
12. Mupepi, M. G. (2017). https://www.igi-global.com/chapter/using-communities-of-practice-to-identify-competencies/167901). Using communities of practice to identify competencies. In M. Benke, F. Hurst, S. Krauss, F. Mili, & D. Shulman (Eds.), *Handbook of research on competency-based education in university settings* (pp. 157–167). IGI Global. https://doi.org/10.4018/978-1-5225-0932-5
13. Irvine, C. K. S., & Kevan, J. M. (2017). https://www.igi-global.com/chapter/using-communities-of-practice-to-identify-competencies/167893). Competency-based education in higher education. In M. Benke, F. Hurst, S. Krauss, F. Mili, & D. Shulman (Eds.), *Handbook of research on competency-based education in university settings* (pp. 1–27). IGI Global. https://doi.org/10.4018/978-1-5225-0932-5
14. Navarre, M., Wozniak, K., Marienau, C., Wilbur, G., Tolliver, D. E., & Meyer, P. (2017). https://www.igi-global.com/chapter/using-communities-of-practice-to-identify-competencies/167905). Learning, adults, and competency-based education. In M. Benke, F. Hurst, S. Krauss, F. Mili, & D. Shulman (Eds.), *Handbook of research on competency-based education in university settings* (pp. 210–231). IGI Global. https://doi.org/10.4018/978-1-5225-0932-5
15. Gonçalves, M. J. A., Rocha, Á., Cota, M. P., & Pimenta, P. C. C. (2017). Model for identifying competencies and learning outcomes (MICRA). In M. M. Pinheiro & D. Simões (Eds.), *Handbook of research on engaging digital natives in higher education settings* (pp. 154–169). IGI Global. https://doi.org/10.4018/978-1-5225-0039-1.ch007

16. Sefiani, N., Boumane, A., Campagne, J.-P., & Bouami, D. (2012). Process of identifying competencies based on a functional approach. *International Journal of Engineering Science and Technology, 4*(1), 265–275.
17. *Tuning Journal for Higher Education 6*(*1*) (2018). http://www.tuningjournal.org/index.php/tuningjournal/
18. Pukhovska, L. P. *Formation of "competence" ideas in European pedagogical education*. https://www.narodnaosvita.kiev.ua/?page_id=1736
19. Sobolieva, S. (2018). Formation of social competence of the future specialists in economics in the process of teaching social and humanitarian disciplines. *Pedagogical Sciences: Theory, History, Innovative Technologies, 2*(76), 197–206. https://doi.org/10.24139/2312-5993/2018.02/197-206
20. Tolochko, S. V. (2018). Structured system analysis of the experience of Ukraine in determine key competencies ScienceRise. *Pedagogical Education, 5*, 36–42. https://doi.org/10.15587/2519-4984.2018.139110
21. *ALIGN: Achieving and checking the alignment between academic programmes and qualification frameworks*. http://align.brusov.am/
22. OESD. (2013). *Assessment of higher education learning outcomes*. Feasibility Study Report. Vol. 3—Further Insights.

Innovative Approaches to Determining the Monopolitisation Level of Regional Primary Residential Real Estate Markets of Ukraine

Olena Pavlova, Kostiantyn Pavlov, Olena Panukhnyk, Leonid Ilyin, Oksana Apostolyuk, and Nazariy Popadynets

Abstract In general terms, real estate market is a constituent of national economy, embodying a lion's share of the world's total wealth, and its stable functioning is one of the most complex market processes occurring in any state. Ukraine is currently experiencing the stage of forming a market system including the market of primary residential real estate. Due to constant demand for residential real estate, rather insufficient solvency of the population, and the disproportionality of certain regional primary residential real estate markets (RPRREM) development, the question of regional real estate market competitiveness and their monopolisation level is extremely significant today. For the purpose of this in-depth study, a large number of statistical, administrative, governmental, and scientific sources elaborated by leading scientists in Ukraine and abroad were processed in order to determine the state of primary real estate market in the regions of the country. A number of measuring instruments for the monopolisation level (concentration) of regional primary residential real estate markets (RPRREM) were employed: the market concentration factor; the Herfindahl–Hirschman index; dispersion of market shares; entropy of market shares; the Gini index. The results of the study indicate that methodological approaches were generalised by the authors, as well as specific tools and methods of scientific knowledge, which, in turn, were introduced into a unified model of research with further definition of the monopolisation level of regional primary residential real estate markets (RPRREM) of Ukraine.

O. Pavlova · K. Pavlov · L. Ilyin
Lesya Ukrainka Eastern European National University, Lutsk, Ukraine
e-mail: pavlova.olena@eenu.edu.ua

O. Panukhnyk
Ternopil Ivan Puluj National Technical University, Ternopil, Ukraine

O. Apostolyuk
Ternopil National Economic University, Ternopil, Ukraine

N. Popadynets (✉)
SI "Institute of Regional Research Named After M. I. Dolishniy of the NAS of Ukraine", Lviv, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_33

Keywords Competition · Competitive relations · Market monopolisation · Market concentration · Real estate market · Housing · Constructor · Developer · Residential object

1 Introduction

Real estate markets conditioned by transformation processes in Ukraine are characterised by institutional asymmetry and structural incompleteness, and their crisis economic situation entailed by lower economy efficiency indicators, the financial and credit system's imperfection, general decline in production, shortage of real incomes, and population's social stratification reinforcement naturally slowing down their development.

When investigating regional primary residential real estate market, its integral functioning system, and the subordination of operating non-liquid commodity markets, as well as financial markets should be adhered to. Meanwhile, the inherent feature of residential real estate is the material form of existence, the space immobility, and gradual transfer of market value over time. Besides, as a result of practical realisation of housing objects' purchase and sale operations, the market contributes to the development of other types of economic activities, creating a unique environment. Taking into account the mentioned conditions, it is possible to state that the primary market of residential real estate is one of the basic forming constituents of national economy.

The concept of competition and competitiveness has been researched by such scholars as Ohlin B [1]., Hekscher E [2]., Samuelson P [3]., Smith A [4]., Babenko V [5]., Melnyk M [6]., Suhina E [7]., Koshkalda I [8]., Vasyltsiv T [9]., Mikulikova, Z [10]. The study focused on "compulsory nature" of products, industry, and region as a production factor conditioned by competitive advantages. In fact, we believe that the competitive advantages of production factors insufficiently reflect the potential of an object. A certain region (regional market) with a high level of enormous potential as a research object can be classified as advantageous. However, Mann R [11]. believed that competitiveness of real estate is a cluster of different characteristics and distinguished them from similar objects according to the satisfaction level of an individual consumer, the cost of its purchase, operation, and allocatedness by the consumer among options offered by sellers, due to significantly higher economic and technical peculiarities [11].

Consequently, the monopolisation level of regional residential real estate markets (RRREM) and competition within a certain RRREM is an important condition for their functioning and development, as it promotes the implementation of the most efficient production methods and offers more flexible prices and new conceptual solutions [12].

2 Materials and Methods

The concentration of sellers reflects the relative scalability and number of construction developer companies working in the industry. Their smaller number testifies to a higher concentration level. If the number of construction developer organisations in the markets coincides and if these organisations diverge considerably, the level of concentration is higher. Although, it is of foremost importance to determine what serves as an indicator of the construction developer organisation's size and market boundaries. In order to characterise the concentration indicators of the largest construction developer organisations on a certain RRREM, a scalability indicator is used, also called the threshold market share.

To analyse the state of the competitive environment, the coefficient of market concentration $CR_{\underline{n}}$ is calculated as the sum of market shares of the largest firms operating on the market. It characterises the percentage share of several largest enterprises in the total market volume. This indicator is calculated by the ratio:

$$\mathrm{CR}_n = \sum_{i=1}^{n} S_i, i = 1, 2, \ldots, n, \tag{1}$$

where n is the number of largest firms in the market for which the indicator is calculated; Si is the share of the ith firm in the market (in the industry).

If the concentration index approximates to a value of 100%, RRREM can be characterised by a high degree of monopolisation. The Linda index is used to determine the number of firms and those of them holding leading positions in RPRREM. With this aim, the index is calculated in several stages: first, for two largest CDOs, then for three, etc. until the continuity of functions is disturbed:

$$\mathrm{IL}_2 = \frac{S_1}{S_2} x 100\%, \tag{2}$$

$$\text{If, } S_1 = 50\%,\ S2 = 25\%, \text{ then } \mathrm{IL}_2 = 200\%, \tag{3}$$

For three largest CDO the Linda Index is calculated by the formula:

$$\mathrm{IL}_3 = \frac{1}{2}\left[\frac{S_1}{(S_2 + S_3)/2} + \frac{(S_1 + S_2)/2}{S_3}\right] x 100\%, \tag{4}$$

For four largest CDO the Linda Index is calculated by the formula:

$$\mathrm{IL}_4 = \frac{1}{3}\left[\frac{S_1}{(S_2 + S_3 + S_4)/3} + \frac{(S_1 + S_2)/2}{(S_3 + S_4)/2} + \frac{(S_1 + S_2 + S_3)/3}{S_4}\right] x 100\%, \tag{5}$$

The Linda index reduction by addition of a CDO means that the core is not yet formed. Calculation of this indicator is carried out as a percentage sum of market share squares of all entities in the market in its total volume:

$$\mathrm{HHI} = \sum_{i=1}^{n} S_i^2, \tag{6}$$

where *Si* is the fraction of *i*th input in the industry, in per cent; *n*—the number of CDO in the industry, *n* often taken as 50. Market shares can therefore be expressed in fractions or percentages. In the first case, the HHI will gain a value from 0 to 1, in the second case, from 0 to 10,000.

One of the main shortcomings of the HHI index is that the accuracy of its calculation requires a complete analytical basic framework for all market participants, setting which is currently complicated. The HHI index as a concentration level indicator is directly related to the Lerner's monopoly power indicator. This feature is widely used in economic research [13]. In course of microeconomics, the index characterising monopoly power is considered as a certain value, the price of which exceeds the marginal expenditure:

$$L = P - \frac{MC}{P} = -\frac{1}{e_D}, \tag{7}$$

where P is price per item, MS marginal costs associated with the production of an additional production item, e_D is elasticity of demand for this company's price. The larger the gap between R and MS, the greater is the market monopolisation degree. The value of L ranges between 0 and 1. In perfect competition, the Lerner index is 0.

The value of the Lerner index can be directly related to HII index for the oligopolistic market, assuming that it is described through the Cournot model [13–15]. In this case, the Lerner index for an individual enterprise will be calculated according to the formula (monopoly power index):

$$L_i = \frac{S_i}{e_D} \tag{8}$$

in which s_i is a CDO's market share, e_D is market demand's elasticity index [12].

In this case, the average index for the industry (provided that the shares of enterprises in the market serve as scales) will be:

$$L = -\frac{\mathrm{HHI}}{e_D}, \tag{9}$$

There is also a dependence of the Lerner index on the concentration level to be mentioned, taking into account the consistency of enterprises' pricing policy:

$$\text{for CDO,} \qquad L_i = \frac{b}{e_D} - (1 - b)\,, \frac{k_i}{e_D}, \tag{10}$$

$$\text{for a construction industry,} \qquad L = -\frac{b}{e_D} - \frac{(I - b)\,\text{HHI}}{e_D}, \tag{11}$$

where b stands for an indicator of the CDO's pricing policy consistency (the degree of conspiracy), which acquires the value from 0, corresponding to the interaction of the CDO according to the Cournot curve, to 1, in consonance with the cartel agreement. The higher the pricing policy consistence indicator, the lower is the dependence of the Lerner index on the CDO by its market share, and for the construction industry—on the vendors concentration level. In international practice, the Gini coefficient and the Lorentz curve are used to determine the level of monopoly power enterprises. The market share dispersion index evaluates the degree of each CDO market share's deviation from the average market share. Market shares dispersion is calculated in concordance with the formula [16]:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(S_i - \frac{1}{n}\right)^2, \tag{12}$$

where S_i is ith CDO's share, n is the total number of CDOs on the market.

The smaller the market shares dispersion index, the more homogeneous is the size of CDO and the economic entities share on the market; the lower is concentration level. Conversely, the greater the dispersion size, the more unequal is the market, the weaker is the competition and the stronger is the power of large CDO operating on this market [13]. The value of the HHI index is related to the dispersion of the CDO shares on the market by the equation as follows:

$$\text{HHI} = n\sigma^2 + \frac{1}{n}, \tag{13}$$

In order to determine the disproportion level of market shares distribution among its participants, the variation coefficient was also used:

$$\nu = \frac{\sigma}{\overline{S}} \cdot 100\%, \tag{14}$$

where $\overline{S} = \frac{1}{n}$ is the average market share.

Another indicator of the uneven distribution level of market shares is the entropy index. The entropy index shows the mean logarithm value of the number inverse to the market share, weighted by the market shares of firms:

$$E = \sum_{i=1}^{n} S_i \text{ In } \left(\frac{1}{S_i}\right), \tag{15}$$

The entropy coefficient is an indicator reciprocal of concentration. If the value of the entropy index ranges from 0 to 0.5, this means that the market is monopolised or close to single monopolisation. A quantitative interpretation of the Lorentz curve is the Gini G coefficient, reflecting the distribution of the total amount of the citizens' income (firms) among their individual groups. Its value may fluctuate within the range of 0–1. If the income is evenly divided, there is an approximation of the coefficient in the direction of 0. The higher the level of the value of the indicator, that is, its approximation to 1, the more distorted are distributed incomes in the market within the society.

The Gini coefficient helps to determine the average income difference between two recipients. Similarly, provided that the Gini coefficient is equal to 0.2, it will mean that the average difference in the income of the recipients belonging to this figure will constitute 40% relative to the average income of the figure. The Gini coefficient is the ratio of the segment A area generated by the Lorentz curve and the line of even distribution to the triangle A area + B below the line of even distribution: $G = A/(A + B)$.

$$B = \frac{h}{3}\left(S_0 + 4S_1 + 2S_2 + 4S_3 + \cdots + 4S_{n-1} + S_n\right), \tag{16}$$

here S_i is ith value of the Lorentz curve ordinate.

The number of n values must be even. The area of segment A is calculated by the ratio.

$$A = \frac{1}{2} - B \tag{17}$$

3 Results and Discussion

Assessment of the monopolisation level of regional primary residential real estate markets in Ukraine.

According to our proposed model of Ukraine's territory regionalisation with distinction between six regional markets of primary residential real estate, as reflected in Table 1: Western, Eastern, Northern, Southern, Central, and Capital. Each of these markets possesses a certain degree of independence and has its own construction concentration peculiarities of primary housing stock and the intensity of CDO competition.

Analysis of the competitive situation that occurs between the CDOs as of January 1, 2019, within the framework of regional markets for primary housing construction, was carried out due to proposed methodological approaches to determine the monopolisation degree of regional primary residential real estate markets.

Thus, Western Regional Real Estate Market (RPRREM) must be considered first, characterised by the highest level of primary housing construction intensity. The statistics on developer companies and the number of construction sites for this

Table 1 Geographic and demographic features of regional primary residential real estate markets of Ukraine, as of January 1, 2019 (indicators of the temporarily occupied territory of the Autonomous Republic of Crimea, the city of Sevastopol, Donetska and Luhanska oblasts as of January 1, 2014)

Area sq. m	Per cent from the country's total area	Population, '000	Per cent from the country's total
Western region: Volynska, Rivnenska, Lvivska, Ternopilska, Khmelnytska, Zakarpatska, Ivano-Frankivska, Chernivetska			
131,277	21.7	10,640	25.8
Cetral region: Kyivska, Zhytoyrska, Vinnytska, Cherkaska, Kirovogradska			
129,964	21.6	6780.2	15.8
Northern region: Chernihivska, Sumska, Poltavska, Kharkivska			
115,862	19.2	6290.9	14.7
Southern region: Odesska, Mykolayivska, Khersonska, AR Crimea			
112,450	18.6	4601.6	10.8
Eastern region: Dnipropetrovska, Zaporizka, Donetska, Luhanska			
112,295	18.6	11,444	26.1
Capital region: KYIV			
839	0.1	2916.2	6.8
Total throughout Ukraine			
603,766	100	42,672.9	100

Sources: developed by the author based on data [17]

region in 2018 are given in Table 2. Based on these data, degree of concentration and competition intensity in primary residential real estate market was assessed. The total number of operating construction companies in the region is 128, the total number of units under construction is 362.

This is how the market concentration ratio CRn is calculated for three, four, five, and seven largest developers. According to what we have:

$$\mathrm{CR}_3 = 9.1\%; \mathrm{CR}_4 = 11.9\%; \mathrm{CR}_5 = 14.4\%; \mathrm{CR}_7 = 19.3\ \%\ .$$

The value of $\mathrm{CR}_4 = 11.9\%$ means that the four largest companies, "RIEL", "BC Melnyk", "Dita", and "Ternopilbud", comprise 11.9% of the market. The five largest companies possess 14.4% of the market, seven companies 19.3% of the Western RPREM's. Since all calculated coefficients do not exceed 40%, according to V.G. Shepherd's classification. This proves the absence of monopoly and a high level of competition in primary residential real estate market of Western RPRREM.

The Linda index was calculated to assess the relationship between largest enterprises which constitute the "core" of primary real estate market. This index facilitates to determine which CDOs hold the dominant position on Western RPRREM. According to the obtained results, the meanings are as given below:

$$\mathrm{IL}_2 = 109.1\%; \mathrm{IL}_3 = 71.8\%; \mathrm{IL}_4 = 79.3\%$$

Table 2 Calculation of primary housing market concentration for Western RPRREM

No п/п	Company	Number of objects	Si	HHI	Dispersion indicator	Entropy index	*Sn*	*Fn*
1	"Frankivskyi Dim"	1	0.0028	0.000008	0.0000	0.0163	0.003	0.008
2	LLC BC "Victoriya"	1	0.0028	0.000008	0.0000	0.0163	0.006	0.016
3	LLC "Rivnenskyi MZK"	1	0.0028	0.000008	0.0000	0.0163	0.008	0.023
4	JE "Lutsk-tekhnobud"	1	0.0028	0.000008	0.0000	0.0163	0.011	0.031
5	SB "Group"	1	0.0028	0.000008	0.0000	0.0163	0.014	0.039
6	"Fine City Development" Company	1	0.0028	0.000008	0.0000	0.0163	0.017	0.047
7	IDK "FD Group"	1	0.0028	0.000008	0.0000	0.0163	0.019	0.055
8	HCC "Ridnyi DIM"	1	0.0028	0.000008	0.0000	0.0163	0.022	0.063
9	GC "RANG"	1	0.0028	0.000008	0.0000	0.0163	0.025	0.070
...	...	...	...	...	...	...	...	...
109	"City of Dreams"	5	0.0138	0.000191	0.0000	0.0591	0.594	0.852
110	"Zelenyi Dvir"	5	0.0138	0.000191	0.0000	0.0591	0.608	0.859
111	"Halytska Budivelna Gildiia"	5	0.0138	0.000191	0.0000	0.0591	0.622	0.867
112	"Hazda"	5	0.0138	0.000191	0.0000	0.0591	0.635	0.875
113	"SK Group"	6	0.0166	0.000275	0.0001	0.0680	0.652	0.883
114	"Listyng"	6	0.0166	0.000275	0.0001	0.0680	0.669	0.891
115	"Creator Bud"	6	0.0166	0.000275	0.0001	0.0680	0.685	0.898
116	"Vambud"	6	0.0166	0.000275	0.0001	0.0680	0.702	0.906
117	"Blago"	6	0.0166	0.000275	0.0001	0.0680	0.718	0.914
118	"HalZhytlobud"	7	0.0193	0.000374	0.0001	0.0763	0.738	0.922
119	"Budivelnyi Alliance Group"	8	0.0221	0.000488	0.0002	0.0842	0.760	0.930
120	"Bever-Alliance"	8	0.0221	0.000488	0.0002	0.0842	0.782	0.938
121	BO "Lutsksatechmontazh" No 536	9	0.0249	0.000618	0.0003	0.0919	0.807	0.945

(continued)

Table 2 (continued)

No п/п	Company	Number of objects	Si	HHI	Dispersion indicator	Entropy index	*Sn*	*Fn*
122	"Yarovytsia"	9	0.0249	0.000618	0.0003	0.0919	0.831	0.953
123	"Integral-Bud"	9	0.0249	0.000618	0.0003	0.0919	0.856	0.961
124	"Halytskyy Dvir"	9	0.0249	0.000618	0.0003	0.0919	0.881	0.969
125	"Ternopilbud"	10	0.0276	0.000763	0.0004	0.0991	0.909	0.977
126	"Dita"	10	0.0276	0.000763	0.0004	0.0991	0.936	0.984
127	BC "Melnyk"	11	0.0304	0.000923	0.0005	0.1062	0.967	0.992
128	"RIEL"	12	0.0331	0.001099	0.0006	0.1129	1.000	1.000
	Total	362	1.0000	0.0137	0.0059	4.5488	37.547	64.500

Sources: developed by the author based on data [17]

The index growth on the third step means that the "core" of primary residential real estate market in Western region is formed by two largest companies: "RIEL" and "BC Melnyk". To assess the distribution of "market power" among all market actors, the Herfindahl–Hirschman index HHI was calculated, which resulted in:

$$\mathrm{HHI} = 1.9$$

Followed by the calculations, we concluded that Western RPRREM, which is investigated, is a market for perfect competition. This is due to the large number of CDOs, most of which build one or two primary housing stock.

To estimate uneven distribution of market shares, the entropy coefficient and the dispersion index of market shares were calculated. Accordingly, the value of the dispersion index was obtained:

$$\sigma^2 = 0.00005$$

Hence, it is inferred that primary Western RPRREM is low-dispersed. In fact, a severe competition is observed, and there is not a single dominating firm. The coefficient of variation for this market is:

$$V = 87\%$$

This explicates the heterogeneity of real estate market with both large CDO leaders and small ones managing one or two housing projects. However, most CDOs carry provides services to one or two objects. The number of such CDOs is 78 (60.9%) from the total 128 CDOs on the market. Then the entropy of the CDO market shares was calculated, according to which:

$$E = 4.55$$

The high level of entropy indicates a high level of competition in Western RPRREM. The Gini coefficient was calculated with the use of previously applied methodological approaches:

$$G = 0.42$$

This value of Gini coefficient indicates a fairly significant differentiation of market shares possessed by different CDOs and a correspondingly significant differentiation in their revenues. In fact, we conclude that the average difference in market shares is 84% of this share's value.

Central RPRREM was studied next, which is, like Western RPRREM, characterised by a high degree of construction intensity. To the Central RPRREM we assigned: Kyivska, Zhytomyrska, Vinnitska, Cherkaska, and Kirovogradska oblasts. The statistical data on the CDOs and the number of objects under construction for this RPRREM in 2018 are given in Table 3. Using the above indicators, the intensity of competition within Central RPRREM was calculated. The total number of operating CDOs in Central RPRREM is 92, the total number of objects under construction is 171. After that, the CRn market concentration ratio was calculated for three, four, five, and seven largest CDOs. According to this:

$$\mathrm{CR}_3 = 11.7\%;\ \mathrm{CR}_4 = 14.6\%;\ \mathrm{CR}_5 = 17.5\%;\ \mathrm{CR}_7 = 22.2\%.$$

The value of $\mathrm{CR}_3 = 11.7\%$ means that the market share of 11.7% is possessed by the three largest companies—"Profgroup", "Arial Investment Development", and "Nadiia Building Company". The five largest companies occupy 17.5% of the market, seven companies—22.2% of Central RPRREM. Since all calculated coefficients do not exceed the value of 40%, according to the V.G. Shepherd's classification this is proves the absence of monopoly and a high level of competition in Central Regional Primary Real Estate Market.

The Linda index was calculated to estimate the relations between the largest CDOs comprising the "core" of primary residential real estate market. This index explicates which CDOs hold the dominant position in primary housing market. As a result:

$$\mathrm{IL}_3 = 133.3\%;\ \mathrm{IL}_3 = 75.0\%;\ \mathrm{IL}_4 = 92.1\%.$$

The growth of index on the third step means that the "core" of the Central RPRREM is formed by two largest companies: "Profgroup" and "Arial Investment Development". For the estimation of "market power" distribution among all market participants, the Herfindahl–Hirschman index was calculated, which resulted in:

$$\mathrm{HHI} = 162.$$

Table 3 Calculation of primary housing market concentration for Central RPRREM

No п/п	Company	Number of objects	Si	HHI	Dispersion index	Entropy index	*Sn*	*Fn*
1	PE "RESPECT"	1	0.0058	0.000034	0.0000	0.0301	0.006	0.011
2	HC "Kyivskyi kvartal"	1	0.0058	0.000034	0.0000	0.0301	0.012	0.022
3	HC "Campa"	1	0.0058	0.000034	0.0000	0.0301	0.018	0.033
4	"Finansovyi budivelnyi holding"	1	0.0058	0.000034	0.0000	0.0301	0.023	0.043
5	"Ukrzhytloinvest	1	0.0058	0.000034	0.0000	0.0301	0.029	0.054
6	"TrestBoryspilsilbud"	1	0.0058	0.000034	0.0000	0.0301	0.035	0.065
7	"TEOS Developer Group"	1	0.0058	0.000034	0.0000	0.0301	0.041	0.076
8	"Sfera zhytlobud"	1	0.0058	0.000034	0.0000	0.0301	0.047	0.087
9	"Strakhovskyi Group"	1	0.0058	0.000034	0.0000	0.0301	0.053	0.098
10	"Serviszemproekt"	1	0.0058	0.000034	0.0000	0.0301	0.058	0.109
...	...	...	...	...	...	...	...	...
75	"R-Building"	3	0.0175	0.000308	0.0000	0.0709	0.585	0.804
76	"Novyi Dim"	3	0.0175	0.000308	0.0000	0.0709	0.602	0.815
77	"Druzi Development"	3	0.0175	0.000308	0.0000	0.0709	0.637	0.837
78	"Dovira Development"	3	0.0175	0.000308	0.0000	0.0709	0.655	0.848
79	"Global Development"	3	0.0175	0.000308	0.0000	0.0709	0.673	0.859
80	"Budregioninvest"	3	0.0175	0.000308	0.0000	0.0709	0.690	0.870
81	BC "Forum "	3	0.0175	0.000308	0.0000	0.0709	0.708	0.880
82	"BD Holding"	3	0.0175	0.000308	0.0000	0.0709	0.725	0.891
83	"Alliance Novobud"	3	0.0175	0.000308	0.0000	0.0709	0.743	0.902
84	"In-build"	3	0.0175	0.000308	0.0000	0.0709	0.760	0.913
85	"Evrodim.com"	3	0.0175	0.000308	0.0000	0.0709	0.778	0.924
86	"Atlant"	4	0.0234	0.000547	0.0002	0.0878	0.801	0.935
87	"AVM Development Group"	4	0.0234	0.000547	0.0002	0.0878	0.825	0.946
88	"NovaBudova"	5	0.0292	0.000855	0.0003	0.1033	0.854	0.957
89	"Vinnytsiabud"	5	0.0292	0.000855	0.0003	0.1033	0.883	0.967
90	"Building Company Nadiia"	6	0.0351	0.001231	0.0006	0.1175	0.918	0.978
91	"Arial Investment Development"	6	0.0351	0.001231	0.0006	0.1175	0.953	0.989
92	"ProfGroup"	8	0.0468	0.002189	0.0013	0.1433	1.000	1.000
	TOTAL	171	1.0000	0.016176	0.005306	4.33075	31.82	46.50

Sources: developed by the author based on data [17]

This value is much greater than the value of the HHI for Western RPRREM. On the other hand, according to the provided classification, Central RPRREM is characterised by perfect competition.

The entropy coefficient and the dispersion index of market shares is to be calculated. According to this the index of dispersion is obtained:

$$\sigma^2 = 0.00006$$

Consequently, Central RPRREM is marked by low dispersion. In other words, there is a tough competition and the lack of dominance of one or two CDOs. The coefficient of variation for this market is:

$$V = 70\%$$

This testifies to the low level of heterogeneity of Central RPRREM. Although there are both small and large CDOs, the majority of the CDO manage one or two objects. There are 72 small CDO of this (78.3%) within 92 BOD on the market.

The next step was to determine the entropy value of CDO market shares. According to the calculations the following was retrieved:

$$E = 4.33$$

The high level of entropy indicates high competition on the market. The value of Gini coefficient was calculated accordingly:

$$G = 0.32$$

Such value of the Gini coefficient explicates a smaller differentiation of market shares, held by different CDOs than that of Western RPRREM. The market is fairly homogeneous, CDO market shares differ by 64% of this share.

Based on the methodological and practical approach described above, we conducted a study on the level of competition between the construction market for primary housing and for other regions of Ukraine: Northern RPRREM; Southern RPRREM; Eastern RPRREM; Capital RPRREM. The grouped and systematised results of our studies are given in Table 4.

4 Conclusions

The analysis of the table indicates that there is a situation close to perfect competition for three RPRREMs (Western, Central, and Capital). In each of these markets, there is a large number of CDOs (50–100), and there is no apparent predominance of one or two separate CDOs. The CR_5 concentration index does not

Table 4 Monopolisation level characteristics of regional primary residential real estate

RPRREM	CR_5, %	Number of firms in the "core"	HHI	σ^2	V	E	G	Market type
Western	14.4	2	2	0.00005	0.87	4.55	0.42	Perfect competition
Central	17.5	2	162	0.00006	0.70	4.33	0.32	Perfect competition
Capital	39.8	3	485	0.00056	1.23	3.47	0.50	Perfect competition
Southern	55.9	2	1142	0.00254	1.66	2.85	0.53	Compet. with Monopoly El.
Northern	58.3	2	890	0.00198	1.02	2.72	0.49	Compet. with Monopoly El.
Eastern	48.5	3	725	0.00094	0.55	2.75	0.29	Compet. with Monopoly El.

exceed 40%. The Herfindahl–Hirschman index does not exceed 500. The entropy index exceeds 3 which indicates a high level of competition in these markets.

On another three RPRREMs (Southern, Northern, and Eastern regions), a market situation is observed as "competition with the elements of a monopoly". The number of builders is smaller (20–30), and, at the same time, there are several CDOs—outstanding leaders in terms of the number of objects under construction. The CR_5 concentration index ranges within 50%–60%. The Herfindahl–Hirschman index is 700–1100. The entropy index exceeds 2, which indicates a fairly high level of competition on the RPRREM. The value of Gini coefficient for Central and Eastern RPRREMs is the lowest indicating the high homogeneity of these markets. The value of Gini coefficient for the Capital, Northern, and Southern RPRREM is significantly higher, indicating the presence of leading CDO building up a large number of residential properties in primary housing stock.

References

1. Ohlin, B. (1967). *Interregional and international trade*. Harvard University Press.
2. Heckscher, E. (2006). *The impact of foreign trade on income distribution*. TEIS. (in Russian).
3. Samuelson, P. (1989). *Economics* (13th ed.). McGraw Hill.
4. Smith, A. (2007). In S. M. Soares (Ed.), *An nquiry into the nature and causes of the wealth of nations*. MetaLibri.
5. Babenko, V. (2019). Formalization of the model of management of the technological innovations. *CEUR Workshop Proceedings, 2393*, 595–602.
6. Melnyk, M., Korcelli-Olejniczak, E., Chorna, N., & Popadynets, N. (2018). Development of regional IT clusters in Ukraine: Institutional and investment dimensions. *Economic Annals-XXI, 173*(9–10), 19–25.
7. Ulytsky, O. A., Suhina, E. N., Tkach, V. A., & Kamushkov, A. S. (2019). Economic mechanisms of coal mining waste management at realization of public-private partnerships. *Ugol, 4*, 39–41. (in Russian).
8. Kalashnikova, T., Koshkalda, I., & Trehub, O. (2019). Mathematical methods of data processing in formation and evaluation of georal structure in agricultural enterprises. *Global. J. Environ. Sci. Manage, 5*, 87–95.
9. Vasyltsiv, T., & Lupak, R. (2016). Justification of strategic priorities of state structural policy of import substitution in Ukraine. *Actual Problems of Economics, 1*(175), 52–60.
10. Mikulikova, Z. (2018). The mathematical model of the operation problem presented on a particular model. *Perspectives - journal on economic issues, 1*(4), 51–63.
11. Mann, R. V. (2007). Features of the diagnosis of the state and prospects for the development of the real estate market. *Economy & The State, 9*, 15–17. (in Ukrainian).
12. Strishenets, O. M., & Pavlov, K. V. (2016). Features of competitive relations on regional real estate markets. *Scientific Bulletin of Uzhgorod University. Series "Economics", 1*(2), 35–38. (in Ukrainian).
13. Pavlov, K. (2018). Methodological approaches to determining the monopolisation level of primary residential real estate regional markets in Ukraine. *Scientific Journal "Economics, Entrepreneurship, Management", 1*(5), 63–72.
14. Drapikovskii, O. I., & Ivanova, I. B. (Eds.). (2012). *Trends in the real estate market in Ukraine: Realities and forecasts. 2007–2013*. Art Economics. (in Ukrainian).

15. Panukhnyk, O., Popadynets, N., & Fedotova, Y. (2019). Analysis and modeling of factor determinants of food provision at consumer market of Ukraine. *Global Journal of Environmental Science and Management, 5*(SI), 215–226.
16. Minyaylo, O. I., & Buryan, A. M. (2016). The current state of competition in the primary real estate market of Kyiv. *Business Inform, 4*, 126–130. (in Ukrainian).
17. *Official site of "State service of statistics of Ukraine"*. http://www.ukrstat.gov.ua (in Ukrainian).

The Impact of Investment Activity Parameters on Uneven Development of the Regions of Western Ukraine

Andriy Pilko, Volodymyr Stefinin, Svitlana Danylo, and Nazariy Popadynets

Abstract The research of the dynamics of impact of changes in investment activity parameters in Ukrainian regions and the level of their development asymmetry is relevant and in demand in current development stage. This chapter suggests the approach to evaluation of the impact of investment activity parameters on the level of development asymmetry of the regions of Western Ukraine. The dynamics of regional asymmetry in the period under research is analyzed based on the previously conducted estimation of the level of development asymmetry of Western Ukrainian regions by the parameters of sustainable development and levels of development harmonization and social tension. It was possible to obtain scientifically valid results due to the application of the methods of information analysis, indicative analysis, correlation-regression analysis, and estimation of the parameters of dynamic econometric models. Scientific novelty of the chapter lies in improvement of existing approaches to estimation, analysis, and prognosis of the impact of investment activity parameters on the regional asymmetry rates by the parameters of sustainable development and levels of harmonization and social tension.

Keywords Region · Development · Misbalances · Asymmetry · Level of sustainable development harmonization · Social tension · Model · Investment

A. Pilko · V. Stefinin
Vasyl Stefanyk Precarpathian National University, Ivano-Frankivsk, Ukraine

S. Danylo
Uzhhorod Institute of Trade and Economics of Kyiv National University of Trade and Economics, Uzhhorod, Ukraine

N. Popadynets (✉)
SI "Institute of Regional Research Named after M. I. Dolishniy of the NAS of Ukraine", Lviv, Ukraine

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_34

1 Introduction

In order to efficiently manage the investment component of socio-ecological-economic development, it is necessary to develop a set of instruments to estimate and analyze the level of socio-ecological-economic development, its unevenness, misbalances, and safety as well as to analyze the system of relations between the parameters of socio-ecological-economic development and parameters that characterize investment climate of the region and efficiency of its use.

Expansion of the problem of estimation and analysis of the level of regional development asymmetry to the problematics stipulates estimation of the level of socio-ecological-economic development of regions under research and of regions' development misbalances same as identification of factors that define the patterns of regions' development and impact the level of development misbalances.

As Pena [1] rightfully mentions, currently the development tendencies of Ukrainian regions stipulate the deepening of all types of regional misbalances in economic space. Such processes typically result in growing level of social tension, increasing labor migrations, deteriorating demographic situation, and finally reduction of development capacity of territorial systems and regions of Ukraine, where the misbalances of socio-ecological-economic development are substantial in a long-term period and of systemic nature. In other words, high rates of regional socio-ecological-economic development asymmetry observed in several reporting periods can be the basis for forming of preconditions for the emergence of economically backward and depressive territorial systems, systemic inefficient use of local resources, and exhaustion and degradation of some regions' capacities.

Domestic and foreign researchers use the concept of "asymmetry" in the process of forming of efficient mechanisms to overcome spatial disparities of regional development and to estimate their impact on the efficiency of the use of resources. According to Krasnonosov and Yermolenko [2], asymmetry helps adequate characterization of current development conditions of Ukrainian regions. They understand asymmetric as "territorially misbalanced development of the country, increasing gap between the economically strong and weakly developed regions with peculiar discrepancy of regional parameters."

We understand "asymmetry of regional development" in the framework of our research as the discrepancies in the level and dynamics of social, economic, and ecological development of some regions, which come as the result of the impact of various endogenous and exogenous factors with regard to the regions with their peculiar features.

It is obvious that positive aspects of the impact of asymmetry in some cases reduce to improvement of the use efficiency of local resources and forming of relevant competitive advantages of some regions. Meanwhile, as noted by Tyshchenko [3], the negative consequences of regional spatial asymmetry can exceed dozens of times the few positive economic and social effects caused by different levels of provision of territorial systems with main types of resources.

In our opinion, it is important to go beyond the objective estimation of the level of sustainable development of the regions under research in the process of forming of efficient mechanisms to manage regional socio-ecological-economic development and to determine and to analyze the level of harmonization of such development as well. We will consider the level of harmonization of a region's sustainable development as the balance that can be achieved only in case of creation of relevant conditions to align the goals of economic, ecological, and social development.

Currently, there are many scientific works devoted to the analysis of the impact of unbalanced development of regions on the nature of changes taking place in the relevant space. In particular, Doctor [4] argues that economic and political asymmetry as well as institutional, social, and economic deficits prevent the deepening of regionalism and international integration.

Petrakos [5] analyzes the spatial structure of Southeastern Europe with the view to estimate regional misbalances. The results of conducted analysis show that Southeastern Europe faces the increasing regional asymmetry, growing role of mega-cities and much lower rates of socio-economic development of peripheral regions.

There is also the range of approaches to estimation of regional asymmetries and development misbalances. For example, they are calculated by the Herfindahl coefficient and quintile coefficient [6], regional asymmetry coefficient (AS) [7], decile coefficient [8], standard variation level [9], and Euclidean distance [10]. The level of misbalances is advisable to be calculated by the methods of ranging using the Lorenz curve and Gini coefficient, Theil index, scaling methods, average deviation, and standard variation as outlined in refs. [11–15].

It is worth mentioning that possible options of solving the tasks of estimation and analysis of misbalances and asymmetries of regional socio-ecological-economic development are actively examined by both domestic and foreign researchers. Western methods of estimation of development misbalances in some groups of countries and regions usually stipulate: development of rankings (complex estimations) of international and regional development, use of the methodics of research of income inequality (Lorenz curve, Gini coefficient), and use of indicators that show the dispersion, standard deviation, variation coefficient, etc. [16–18].

Maslyhina [19] provides classification of parameters and methods of quantitative evaluation of spatial development misbalances. In particular, she outlines the parameters of the scope of inequality (Williamson's coefficient of variation, Klotsvog-Mahomedov coefficient, Gini Index, Hoover Index, Theil Index, Atkinson Index, Kolm Index, Moran's and Geary's Spatial Correlation Indices, etc.), parameters and methods that characterize the structure of inequalities (Local Spatial Autocorrelation Index (Getis-Ord Index), asymmetry and kurtosis coefficients, cluster analysis, Theil and Atkinson Indices), as well as methods showing the dynamics of inequalities (dispersion parameters, Gini Index, Hoover Index, Theil Index, Atkinson Index, Kolm Index, decile coefficient, convergence analysis, etc.).

In Ukraine, the processes of estimation of regional development are regulated by the Resolution of the Cabinet of Ministers of Ukraine "On Approval of Estimation of Interregional and Intraregional Differentiation of the Region's Socio-Economic

Development" [20], which outlines the procedures of estimation and analysis of socio-economic situation, level, and quality of life in order to find interregional and intraregional asymmetries and inequalities of socio-economic development.

Moreover, some issues of estimation of the level, misbalances, and asymmetries of regional development are regulated by 2020 State Regional Development Strategy of Ukraine 2020 [21] and the Law of Ukraine "On the Foundations of State Regional Policy" [22].

The issues of estimation of investment component of socio-ecological-economic development of a region and its modeling are covered in ref. [23]. Hasnah [24] suggests the scheme of analysis of the impact of key economic growth factors on the GDP rates of regions in Malaysia based on econometric methods and models. Capital investment and foreign investment volumes and number of employed were selected among the factors of economic growth [24]. The suggested scheme of analysis contributes to quantitative estimation and ranging by impact of economic growth factors on the result indicators of the regions' economy and inequalities of regional development.

Sîrbu [25] estimates the impact of foreign investment on aligning of development indicators of Romanian regions.

This chapter aims to show the results of conducted research directed at development of theoretical basis and scientific methodical foundations of increasing the efficiency of management of regional socio-ecological-economic development based on economic-mathematical modeling of the processes of asymmetry estimation and analysis and estimation of lag effect of impact of regional development investment component on its asymmetry.

2 Materials and Methods

The suggested [10] scheme of estimation of the level and degree of harmonization of Western Ukrainian regions' sustainable development lies in improvements in terms of accounting of possible multicollinearity and modification considering the available statistical information and based on the approach that stipulates the calculation of sustainable development rate outlined in ref. [26]. Sustainable development parameter I_{sd} is calculated based on the rates of economic (I_{ec}), ecological (I_e), and social development (I_s). In order to determine the weight of each of these parameters, we suggest calculation of actual shares of pairwise correlation coefficients based on estimation of the nature of interrelations between the relevant economic, social, and ecological parameters and the parameter of social tension calculated for each of regions under research for 2000–2017.

Similar to [10, 26], the level of sustainable development harmonization is considered as the angle between the vector I_{sd}:

$$I_{\text{sd}} = \frac{I_{\text{ec}}}{3} + \frac{I_e}{3} + \frac{I_s}{3}, \tag{1}$$

And "ideal" vector equidistant from each of coordinates I_{ec}, I_e, I_s:

$$1 = \frac{I_{ec}}{3} + \frac{I_e}{3} + \frac{I_s}{3}, \tag{2}$$

The angle is measured in degrees and determined by the ratio:

$$\alpha = \frac{I_{ec} + I_e + I_s}{\sqrt{3} * \sqrt{I_{ec}{}^2 + I_e{}^2 + I_s{}^2}}, \tag{3}$$

Equidistance I_{sd} from each of coordinates I_{ec}, I_e, I_s corresponds to the most harmonious socio-economic development. Approaching of this rate to one of the coordinates indicates actual priorities of the region's development by the relevant dimension with comparatively lacking investment in development by two other directions [26]. Index I_{sd} and the level of harmonization of sustainable development is calculated through its components I_{ec}, I_e, I_s. In particular, in order to calculate the value of economic development parameter (I_{ec}), the following formula is suggested:

$$I_{ec} = k_{stii} * I_{ii} + k_{strt} * I_{rt} + k_{stki} * I_{ki} + k_{strz} * I_{rz} + k_{stek} * I_{ek} + k_{stGRP} * I_{GRP}, \tag{4}$$

where I_{ii}, I_{rz} are the values of economic dimension parameters, including the volumes of foreign direct investment, companies' retail turnover, capital investment, average monthly wages, net exports, and Gross Regional Product per capita.

k_{stii}, k_{strt}, k_{stki}, k_{strz}, k_{stek}, k_{stGRP} are actual shares of pairwise correlation coefficients based on estimation of the nature of interrelations between the relevant economic parameters and social tension parameter.

In order to calculate the value of ecological development parameter (I_e):

$$I_e = k_{stsdz} * I_{sdz} + k_{stpdz} * I_{pdz} + k_{stuv} * I_{uv} + k_{stvu} * I_{vu} + k_{stvo} * I_{vo}, \tag{5}$$

where I_{sdz}, I_{pdz} are the values of ecological dimension parameters, namely the volumes of air pollutants emissions from stationary polluting sources, volumes of air pollutants emissions from mobile polluting sources, wastes generation and management, cost of protection, and rational use of natural resources.k_{stsdz}, k_{stpdz}, k_{stuv}, k_{stvu}, k_{stvo} are the actual shares of pairwise correlation coefficients based on estimation of the nature of interrelations between the relevant ecological parameters and social tension parameter.

The value of social development parameter (I_s) was calculated according to the following formula:

$$I_s = k_{stpz} * I_{pz} + k_{stll} * I_{ll} + k_{stzz} * I_{zz} + k_{stdz} * I_{dz} + k_{stsr} * I_{sr} + k_{stps} * I_{ps}, \tag{6}$$

where I_{pz}, I_{ll}, I_{zz}, I_{dz} are the values of social dimension parameters, namely the number of outpatient clinics, number of hospital beds per 10 thousand persons, number of general education and pre-school education establishments, rate of average monthly total resources per one household, total amount of assigned subsidies.

k_{stpz}, k_{stll}, k_{stzz}, k_{stdz}, k_{stsr}, k_{stps} are the actual shares of pairwise correlation coefficients based on the estimation of the nature of interrelations between the relevant social parameters and social tension parameter [10].

3 Results and Discussion

The value of social tension parameter (I_{st}) was calculated based on the method of Povstyn [27]. In order to calculate the parameter of social tension, the information across the following parameters was used: dropout rate, number of people suffering injuries, number of discharged persons, number of registered crimes, volumes of air pollutants emissions from mobile polluting sources, volumes of air pollutants emissions from stationary polluting sources, number of deaths per 1000 persons of actual population, number of persons with mental health and behavior disorders, number of entities registered in EDRPOU, fixed investment rates, number of arriving persons, number of employed, average monthly wages, amount of housing space available for population, number of health facilities, number of births per 1000 persons of actual population, average monthly total resources per one household.

In order to determine the degree of development asymmetry of the examined regions, we used the Williamson's weighted coefficient of variation in our research [6].

Conducted calculations brought us to conclusions that the acceptable situation in terms of the level of sustainable development and its harmonization was observed in Ivano-Frankivsk, Lviv, Transcarpathian, Chernivtsi, and Ternopil regions depending on the report period (Tables 1 and 2). "Ideal" vector could be observed only in Transcarpathian region in 2000, when the level of sustainable development harmonization reached its maximum value. In time, the level of sustainable development deteriorates along with its harmonization level, which can be explained by both systemic determined and stochastic impact of endogenous and exogenous factors.

The ranging of the impact of factors on the level of economic, social, ecological, and sustainable development of Western Ukrainian regions [10] shows that capital investment is one of the determining rates that directly or indirectly influences the major directions of regional development.

With this regard, we have analyzed the impact of capital investment on the rates of asymmetry of regional socio-ecological-economic development by the parameters of sustainable development, level of sustainable development harmonization, and social tension rate.

Calculated rates of regional asymmetry of Western Ukrainian regions by the Williamson's weighted coefficient of variation by the parameters of sustainable

Table 1 Level of sustainable development of Western Ukrainian regions in 2000–2017

Year	Ivano-Frankivsk	Lviv	Transcarpathian	Chernivtsi	Ternopil
2000	0.960289197	0.954137542	0.894341548	0.666708251	0.980512206
2001	0.957419358	0.883027081	0.945166503	0.626825825	0.836784075
2002	0.858659327	0.719969015	0.809497596	0.597003656	0.765589707
2003	0.807784758	0.785434344	0.746528087	0.551406506	0.805041321
2004	0.845944232	0.745127055	0.650187864	0.506693411	0.857505606
2005	0.486893314	0.610448044	0.550120163	0.425140438	0.76646922
2006	0.387110608	0.631517475	0.285846391	0.357233025	0.62814557
2007	0.346443282	0.570333864	0.245957302	0.185672142	0.452815119
2008	0.293934581	0.429238861	0.366234926	0.239635155	0.402546797
2009	0.493809913	0.512094391	0.365716397	0.23740418	0.377056705
2010	0.358357921	0.442894531	0.195381301	0.455551841	0.317030196
2011	0.136834766	0.285031801	0.124548287	0.424356542	0.206832534
2012	0.045412465	0.0684898	0.161585498	0.354972273	0.218628256
2013	0.055536105	0.00207165	0.155987188	0.333333333	0.090408783
2014	0.25948236	0.179671184	0.405205863	0.366517255	0.107803415
2015	0.439431947	0.530264419	0.45407231	0.449895842	0.18900308
2016	0.419827431	0.484775992	0.48499664	0.431612412	0.141371814
2017	0.480532206	0.491104462	0.49423618	0.455385664	0.204195835

Table 2 Level of sustainable development harmonization of the regions of Western Ukraine in 2000–2017

Year	Ivano-Frankivsk	Lviv	Transcarpathian	Chernivtsi	Ternopil
2000	0.058415384	0.067872484	0	0.6153915	0.02810026
2001	0.007892475	0.109566212	0.083781888	0.61528111	0.165625722
2002	0.07733212	0.310226927	0.018631415	0.61550889	0.210115728
2003	0.143920234	0.12951701	0.070627494	0.61550466	0.013099467
2004	0.050662567	0.105704055	0.080227051	0.61514675	0.08294542
2005	0.500868697	0.139746115	0.086234291	0.61505878	0.212353013
2006	0.576147233	0.178101207	0.110853906	0.6159485	0.094829489
2007	0.581366845	0.17380721	0.615891699	0.81632525	0.196137928
2008	0.255221466	0.364545158	0.037106174	0.615459	0.047535289
2009	0.439589702	0.397971725	0.246499531	0.61650205	0.047157002
2010	0.442988771	0.43087748	0.081762399	0.44891725	0.211944711
2011	0.60186673	0.499856344	0.090419316	0.591875	0.376969546
2012	0.755483422	0.627447078	0.349812273	0.84545386	0.35457038
2013	0.955316618	0.955316618	0.955316618	0.95531662	0.955316618
2014	0.768031934	0.338239863	0.313274705	0.77273781	0.618796832
2015	0.700673633	0.465783175	0.303246528	0.46879612	0.561950615
2016	0.77451354	0.649422524	0.522388822	0.69118592	0.60223431
2017	0.796549532	0.632144237	0.543795556	0.71446223	0.65323665

Table 3 Level of regional asymmetry of Western Ukrainian regions by the Williamson's weighted coefficient of variation by the parameters of sustainable development

Year	Ivano-Frankivsk	Lviv	Transcarpathian	Chernivtsi	Ternopil
2000	0.132920885	0.181855355	0.124958665	0.107322	0.119552298
2001	0.131937203	0.180279575	0.124337382	0.106646	0.118719636
2002	0.149746482	0.204400513	0.141476849	0.121157	0.134803875
2003	0.130990606	0.178651406	0.123803936	0.10596	0.11774645
2004	0.166814779	0.227434662	0.157818539	0.134936	0.149760362
2005	0.201317467	0.274343405	0.190548995	0.162814	0.180444577
2006	0.224978687	0.306458396	0.213048144	0.181927	0.201315776
2007	0.354342097	0.482465752	0.335745682	0.286597	0.316515432
2008	0.240931157	0.327829255	0.228407502	0.194872	0.214765485
2009	0.245514714	0.333796565	0.232954119	0.198643	0.218441242
2010	0.133811565	0.18183606	0.12705559	0.108299	0.118833253
2011	0.390284191	0.53001869	0.371087016	0.315958	0.345945401
2012	0.684977838	0.929426454	0.652075402	0.554776	0.606057082
2013	0.876339943	1.188301667	0.834964547	0.710071	0.773781501
2014	0.416712596	0.564737887	0.397390296	0.337854	0.367220958
2015	0.279281255	0.378374803	0.266569119	0.226576	0.245677342
2016	0.31673278	0.42884129	0.30228981	0.256964	0.27809529
2017	0.29663547	0.44325697	0.33648795	0.284653	0.25647894

development, level of sustainable development harmonization and social tension rate are outlined in Tables 3, 4, and 5.

We hypothesize that the impact of capital investment volumes on the rates of regional socio-ecological-economic development asymmetry is not only of immediate but also lag nature. The hypothesis was verified by the results of correlation-regression analysis between the actual values of dependent variable (asymmetry level) and lag values of factor variable (volumes of capital investment), as well as by the results of application of Alt-Tinbergen's method to determine the maximum lag length.

In order to build the distribution-lag models and to calculate the predicted values of asymmetry under the impact of capital investment for 2018–2020 for the regions of Western Ukraine, we used the following conventional values:

Y_t—level of regional asymmetry

x_t—volumes of capital investment in the current time period

$x_{t-1}, x_{t-2}, x_{t-3}$ $\mathrm{x_t}$— volumes of capital investment in time period $t-1$ $t-2$, $t-3$ (lag variables)

The dynamic econometric models of analysis and prognosis of asymmetry level estimated by the method of Almon by the parameters of sustainable development and the predicted values of asymmetry rate for 2018–2020 are outlined below:

Ivano-Frankivsk region:

Table 4 Level of regional asymmetry of Western Ukrainian regions by the Williamson's weighted coefficient of variation by the parameters of the level of harmonization

Year	Ivano-Frankivsk	Lviv	Transcarpathian	Chernivtsi	Ternopil
2000	1.475004034	2.018022846	1.386648422	1.190934419	1.326654747
2001	1.073822144	1.4672753	1.011968046	0.867976943	0.966245843
2002	0.837194673	1.142751523	0.790961245	0.677355871	0.753654338
2003	1.08480266	1.47950701	1.025286036	0.877512217	0.975120783
2004	1.125296319	1.5342249	1.064609635	0.910248514	1.010250918
2005	0.658141641	0.896876071	0.622937642	0.532265892	0.58990455
2006	0.719367037	0.979897568	0.681219248	0.581709014	0.643705122
2007	0.517042399	0.703995522	0.489907223	0.418190879	0.46184718
2008	0.80033086	1.088991034	0.758729485	0.647329758	0.713413107
2009	0.538084352	0.731568002	0.51055582	0.435357504	0.478748553
2010	0.45514954	0.618501093	0.432169622	0.368370398	0.404201988
2011	0.429390708	0.583126619	0.408269975	0.347616864	0.380609167
2012	0.340691876	0.462274871	0.324326977	0.275932439	0.301438546
2013	0.165786045	0.224802984	0.157958645	0.134331201	0.146384033
2014	0.351260447	0.476035724	0.334973059	0.284788194	0.309542354
2015	0.256132369	0.347012314	0.244473909	0.207795858	0.225313796
2016	0.128289297	0.173697675	0.122439323	0.104080746	0.112639586
2017	0.154697853	0.219864317	0.143267981	0.119846254	0.136477893

Table 5 Level of regional asymmetry of Western Ukrainian regions by the Williamson's weighted coefficient of variation by social tension rate

Year	Ivano-Frankivsk	Lviv	Transcarpathian	Chernivtsi	Ternopil
2000	0.574812015	0.78642753	0.540379657	0.46411	0.517000002
2001	0.328405634	0.448734902	0.309488875	0.265452	0.295505713
2002	0.545213643	0.744204116	0.515104641	0.44112	0.490808937
2003	0.743002729	1.01334352	0.70223862	0.601025	0.667879449
2004	−1.89350596	−2.58159913	−1.79139011	−1.53165	−1.69992215
2005	−0.90840252	−1.237916635	−0.85981207	−0.73466	−0.81421802
2006	−0.49677901	−0.676695652	−0.47043499	−0.40172	−0.44452856
2007	−0.52317449	−0.712344866	−0.49571749	−0.42315	−0.46732466
2008	−0.35353619	−0.481048227	−0.3351593	−0.28595	−0.31514136
2009	−0.39674875	−0.53941113	−0.37645098	−0.321	−0.35299835
2010	−0.34533613	−0.469276043	−0.32790055	−0.27949	−0.30668064
2011	−0.37092522	−0.503728574	−0.35268027	−0.30029	−0.32878573
2012	−0.27622781	−0.374805458	−0.2629594	−0.22372	−0.2444018
2013	−0.2826467	−0.383263998	−0.26930186	−0.22902	−0.24956844
2014	−0.25062681	−0.339654848	−0.23900564	−0.2032	−0.22086065
2015	−0.40391674	−0.547232992	−0.38553153	−0.32769	−0.35531633
2016	−0.40202959	−0.544329155	−0.38369709	−0.32617	−0.35298694
2017	−0.38301564	−0.523614779	−0.37685412	−0.31425	−0.33147856

$$Y_t = 0.044 + 0.189x_t + 0.086x_{t-1} \qquad R^2 = 0.753$$
$$Y_{2018} = 0.526 \qquad Y_{2019} = 0.583 \qquad Y_{2020} = 0.603$$

Lviv region:

$$Y_t = 0.154 + 0.53x_t - 0., 145x_{t-1} \qquad R^2 = 0.723$$
$$Y_{2018} = 0.852 \qquad Y_{2019} = 0.892 \qquad Y_{2020} = 0.885$$

Transcarpathian region:

$$Y_t = -0.008 + 0.31x_t + 0.11x_{t-1} \qquad R^2 = 0.725$$
$$Y_{2018} = 0.355 \qquad Y_{2019} = 0.412 \qquad Y_{2020} = 0.442$$

Chernivtsi region:

$$Y_t = 0.12 + 0.047x_t + 0.115x_{t-1} \qquad R^2 = 0.88$$
$$Y_{2018} = 0.228 \qquad Y_{2019} = 0.229 \qquad Y_{2020} = 0.25$$

Ternopil region:

$$Y_t = -0.074 + 0.15x_t + 0.26x_{t-1} \qquad R^2 = 0.827$$
$$Y_{2018} = 0.323 \qquad Y_{2019} = 0.343 \qquad Y_{2020} = 0.404$$

The calculated predicted values of the level of asymmetry under the impact of capital investment show that asymmetry will grow in Ivano-Frankivsk, Transcarpathian, Chernivtsi, and Ternopil regions by the parameters of sustainable development, i.e., the tendency toward the deterioration of the development of regions is expected. Only in Lviv region, the level of asymmetry will remain around its current level.

The dynamic econometric models of analysis and prognosis of asymmetry level estimated by the method of Almon by the parameters of harmonization level and the predicted values of asymmetry rate for 2018–2020 are outlined below:

Ivano-Frankivsk region:

$$Y_t = 1.13 - 0.021x_t - 0.16x_{t-1} - 0.13x_{t-2} - 0.126x_{t-3} \qquad R^2 = 0.839$$
$$Y_{2018} = 0.136 \qquad Y_{2019} = 0.067 \qquad Y_{2020} = -0.05$$

Lviv region:

$$Y_t = 1.3 - 0.3x_t - 0.05x_{t-1} - 0.07x_{t-2} - 0.36x_{t-3} \qquad R^2 = 0.77$$
$$Y_{2018} = 0.283 \qquad Y_{2019} = 0.122 \qquad Y_{2020} = 0.066$$

Transcarpathian region:

$$Y_t = 1.13 - 0.57x_t - 0.17x_{t-1} - 0.018x_{t-2} - 0.125x_{t-3} \qquad R^2 = 0.753$$
$$Y_{2018} = 0.393 \qquad Y_{2019} = 0.278 \qquad Y_{2020} = 0.22$$

Chernivtsi region:

$$Y_t = 0.574 - 0.41x_t - 0.027x_{t-1} + 0.14x_{t-2} + 0.09x_{t-3} \qquad R^2 = 0.801$$
$$Y_{2018} = 0.444 \qquad Y_{2019} = 0.389 \qquad Y_{2020} = 0.349$$

Ternopil region:

$$Y_t = 1.25 - 0.51x_t - 0.07x_{t-1} - 0.014x_{t-2} - 0.337x_{t-3} \qquad R^2 = 0.814$$
$$Y_{2018} = 0.409 \qquad Y_{2019}0.302 \qquad Y_{2020} = 0.195$$

We can observe that the level of asymmetry of regional development falls under the impact of the change of capital investment volumes by the parameter of harmonization level, i.e., the tendency is positive from the viewpoint of investment influence on the rate of harmonization of regional socio-economic-ecological development in all regions of Western Ukraine. In other words, current tendencies of fixed investments will promote the harmonization of development of Western Ukrainian regions. However, based on the results of prognosis of asymmetry level by sustainable development parameters, we can conclude that general development tendency does not contribute to achievement by the regions under research of the development parameters close to the sustainable development parameters, i.e., the misbalances tend to grow.

The dynamic econometric models of analysis and prognosis of asymmetry level estimated by the method of S. Almon by the parameter of social tension and the predicted values of asymmetry rate for 2018–2020 are outlined below:

Ivano-Frankivsk region:

$$Y_t = -0.138 + 0.026x_t - 0.22x_{t-1} \qquad R^2 = 0.828$$
$$Y_{2018} = -0.393 \qquad Y_{2019} = -0.502 \qquad Y_{2020} = -0.516$$

Lviv region:

$$Y_t = -0.368 + 0.112x_t - 0.21x_{t-1} \qquad R^2 = 0.808$$
$$Y_{2018} = -0.457 \qquad Y_{2019} = -0.513 \qquad Y_{2020} - 0.545$$

Transcarpathian region:

$$Y_t = 0.25 + 0.27x_t - 1.016x_{t-1} \qquad R^2 = 0.733$$
$$Y_{2018} = -0.388 \qquad = -0.341 \qquad Y_{2020} = -0.518$$

Chernivtsi region:

$$Y_t = 0.2 + 0.044x_t - 0.62x_{t-1} \qquad R^2 = 0.839$$
$$Y_{2018} = -0.189 \qquad Y_{2019} = -0.152 \qquad Y_{2020} = -0.238$$

Ternopil region:

$$Y_t = 0.24 + 0.12x_t - 0.75x_{t-1} \qquad R^2 = 0.825$$
$$Y_{2018} = -0.354 \qquad Y_{2019} = -0.342 \qquad Y_{2020} = -0.418$$

The calculated predicted values of the level of asymmetry under the impact of capital investment show that asymmetry will diminish in Ivano-Frankivsk and Lviv regions by the parameter of social tension. In Transcarpathian, Chernivtsi, and Ternopil regions, the steady tendency of the change of the level of regional development asymmetry by the parameter of social tension is absent, which is caused by the change of the volumes of capital investment into the economy of abovementioned regions.

4 Conclusions

The conducted analysis of the results of application of existing methodics to estimate the level of misbalances and asymmetry of socio-ecological-economic development of a region as well as the predicted values of the level of asymmetry of regional development under the impact of the change of capital investment volumes calculated based on the developed dynamic econometric models and statistical data of the regions of Western Ukraine show that the asymmetries of regional development by the parameters of sustainable development are expected to grow under the current tendencies toward the practical realization of investment policy, i.e., the misbalances will aggravate in future. However, improvement of the situation is expected by the parameters of harmonization of sustainable development and the results of prognosis of asymmetry level by the parameter of social tension show the positive trend to the reduction of development of misbalances in Ivano-Franskivska and Lviv regions, which is the positive phenomenon, and the absence of the tendency toward the change of development of misbalances by the parameter in the rest regions of Western Ukraine. Existing unequivocal positive tendencies toward the reduction of development misbalances in two industrial-agricultural regions of Western Ukraine (Ivano-Frankivsk and Lviv regions) and the absence of negative tendencies of the parameter in the rest of regions under research, which are agricultural-industrial by their nature, testify to the fact that growing volumes of capital investment create preconditions in the long-term perspective to conduct systemic transformations directed at reduction of misbalances of regional development by the parameters of sustainable development, as far as the level of development asymmetry by the parameter of development harmonization is being secured currently.

References

1. Pena, T. V. (2013). Regional inequalities of socio-economic development and directions of their reduction and overcoming. *Collection of scientific papers of VNAU. Series: Economic Sciences, 2*(77), 67–81. (in Ukrainian).
2. Krasnonosova, O. M., & Yermolenko, O. O. (2013). Theoretical aspects of forming of spatial asymmetry of labour resources location in a region's economy. *Business Infrom, 1*, 52–55. (in Ukrainian).
3. Tyshchenko, O. P. (2011). Regional misbalances: Nature, methodology, definition and estimation. *Economic Theory, 4*, 55–64. (in Ukrainian).
4. Doctor, M. (2012). Prospects for deepening Mercosur integration: Economic asymmetry and institutional deficits. *Review of International Political Economy, 20*(3), 515–540.
5. Petrakos Dimitris, G. (2004). Economou spatial asymmetry in southeastern Europe. *The Journal of Economic Asymmetries, 1*(1), 127–149.
6. Vahovych, I. M., & Tabalova, O. Y. (2012). *Regional asymmetries of sustainable development of Ukraine: Diagnostics and measuring mechanisms*. Volyn Polihraf. (in Ukrainian).
7. Chuzhykov, V. I. *Territorial asymmetry in Europe: Modeling for Ukraine*. http://www.univ.rzeszow.pl/pliki/Zeszyt8/22_czuzikow_ukraina.pdf (in Ukrainian).
8. Lavrovskyi, B. L. (1999). Measuring the regional asymmetry on the example of Russia. *Economic Issues, 3*, 45–52. (in Russian).
9. Naumenko, Z. H. (2015). Methodics of estimation of regional misbalanced development. *Economy: Realities of Time. Scientific Journal, 4*(20), 33–40. (in Ukrainian).
10. Pilko, A. D., & Harda, T. P. (2018). Models of estimation and analysis of regional development. *Development Economy, 2*(86), 24–35. (in Ukrainian).
11. Melnyk, M., Korcelli-Olejniczak, E., Chorna, N., & Popadynets, N. (2018). Development of regional IT clusters in Ukraine: Institutional and investment dimensions. *Economic Annals-XXI, 173*(9–10), 19–25.
12. Ramazanov, S., Antoshkina, L., Babenko, V., & Akhmedov, R. (2019). Integrated model of stochastic dynamics for control of a socio-ecological-oriented innovation economy. *Periodicals of Engineering and Natural Sciences, 7*(2), 763–773.
13. Panukhnyk, O., Popadynets, N., & Fedotova, Y. (2019). Analysis and modeling of factor determinants of food provision at consumer market of Ukraine. *Global Journal of Environmental Science and Management, 5*(SI), 215–226.
14. Ulytsky, O. A., Suhina, E. N., Tkach, V. A., & Kamushkov, A. S. (2019). Economic mechanisms of coal mining waste management at realization of public-private partnerships. *Ugol, 4*, 39–41. (in Russian).
15. Vasyltsiv, T., Lupak, R., & Osadchuk, Y. (2017). Assessment of the level of penetration of import dependence in the context of the import substitution policy in Ukraine. *Economic Annals-XXI, 167*(9–10), 13–17.
16. Vdovichen, A. A., & Vdovichena, O. H. (2015). Approaches to methodics of evaluation of regions' misbalanced development. *Regional Economy, 1*(57), 6–17. (in Ukrainian).
17. Kalashnikova, T., Koshkalda, I., & Trehub, O. (2019). Mathematical methods of data processing in formation and evaluation of sectoral structure in agricultural enterprises. *Global. J. Environ. Sci. Manage, 5*(SI), 87–95.
18. Vasyltsiv, T., & Lupak, R. (2016). Justification of strategic priorities of state structural policy of import substitution in Ukraine. *Actual Problems of Economics, 1*(175), 52–60.
19. Maslyhina, V. Y. *Quantitative estimation of economic and social spatial inequalities in Volga Federal District*. https://naukovedenie.ru/PDF/22evn413.pdf (in Ukrainian).
20. On approval of estimation of interregional and intraregional differentiation of the region's socio-economic development, resolution of the cabinet of ministers of Ukraine as of 20 May 2009 No 476. http://zakon2.rada.gov.ua/laws/show/476-2009- (in Ukrainian).
21. On approval of 2020 state regional development strategy: Resolution of the cabinet of ministers of Ukraine as of 6 August 2014 No 385. https://zakon.rada.gov.ua/laws/show/385-2014-%D0%BF (in Ukrainian).

22. Law of Ukraine “on foundations of state regional policy” as of 5 February 2015 No 156-VIII. https://zakon2.rada.gov.ua/laws/show/156-19, last accessed 2019/08/08 (in Ukrainian).
23. Pilko, A. D., & Kis, V. V. (2016). Setting and solving the task of estimation and analysis of investment component of region’s development safety. *Economy Problems, 3*, 300–306. (in Ukrainian).
24. Hasnah, A., Er, A. C., Ahmad, A. R., Lyndon, N., & Ahmad, S. (2013). An analysis of the impact of foreign investment on regional disparities: A case of Malaysia. *Asian and Social Science, 9*(14), 7–17.
25. Sîrbu, R.-M. (2014). European Union strategy and foreign direct investments impact on Romania’s regional development. *Procedia - Social and Behavioral Sciences, 124*, 442–450.
26. Zhurovskyi, M. Z. *Sustainable development of Ukrainian regions*. NTUU “KPI”, 2009. (in Ukrainian).
27. Povstyn, O. V. (2010). Methods of estimation of social tension in society. *Economic Space, 35*, 169–180. (in Ukrainian).

Drivers of Urban Growth in the Rural Areas of Kolkata City: An Approach to Smart City

Sushobhan Majumdar, Uday Chatterjee, and Uttara Singh

Abstract Urban expansion and urban growth are prominent indicators of the degree and nature of urbanization of big cities across the world. Kolkata, one of the major cities of eastern India, has not only undergone rapid urban growth since 1980s but also been witness to large influx of refugees from the neighboring states. However, majority of peripheral area being unplanned impede the process of its transformation into a smart city. The present chapter aims at finding the drivers of urban growth in the rural tracts of Kolkata metropolitan area. For the purpose of analysis of LULC, Landsat satellite imagery has been used to measure urban growth and change in land use. Erdas Imagine and Principal Component Analysis (PCA) SPSS software have been used for classification of satellite images and statistical analysis, respectively. It was observed that the population density, tertiary sector, distance from roads, proximity to Kolkata city, good connectivity, etc. were the major drivers of urban growth. This study will assist the policy makers in their decision-making process and empower them in formulating the strategy for transforming Kolkata into a smart city. It will also be crucial for sustainable planning and future modeling of the area.

Keywords Urbanization · Urban growth · Principal component analysis · Smart cities · Future modeling

1 Introduction

The growth in world population is concentrated mainly in the urban areas, as most of the demographic growth is mostly centered here. Demographic growth in the

S. Majumdar (✉)
Department of Geography, Jadavpur University, Kolkata, West Bengal, India

U. Chatterjee
Bhatter College, Vidyasagar University, Dantan, West Bengal, India

U. Singh
CMP Degree College, University of Allahabad, Prayagraj, Uttar Pradesh, India

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_35

urban area occurs due to infrastructural development and increase in employment facilities. According to the United Nations, almost half of the world's population is residing in urban areas or proximity to cities [1]. Among the Southeast Asian countries, urban growth is relatively high in the developed countries, which is mainly due to the rural to urban migration [2–8]. In the developing countries like India, urban growth impacts a country's economic growth as most of the people in India live in cities or near the city areas, and it is estimated that India's GDP will increase by 30% by 2030 [9]. For this reason, most of the policy makers are trying to focus on controlling the growth of urban areas [10–20]. The urbanization in India is very rapid because of the expansion and infrastructural development of the urban areas. According to [21], it is a major challenge to minimize the problem of urbanization in India. Kolkata is one of the cities in the eastern parts of India, which has experienced a demographic growth rate of 10.30%.

For the qualitative measurement of the land use and land cover change, LULC classification needs to be done. Detailed description of the land use and land cover change and its proper assessment is very much needed for the proper planning and management of the area. With the change in land use and land cover, it is very much necessary to identify the drivers or the major forces of urban growth in the rural areas of the city. To find the drivers of urban change is major objectives of this study.

Kolkata city is now increasing at an unprecedented rate for which lands are also converted very rapidly. Drivers played an important role behind the growth of the city as major drivers are pull factors of urbanization. In this chapter, an attempt has been made to find out the major drivers behind the growth of Kolkata city as these are the key factors behind the development of smart city.

2 Data and Methods

2.1 Study Area

Kolkata and its surroundings regulated by Kolkata Metropolitan Development Authority (KMDA) is known as the Kolkata Metropolitan Area (KMA). It is one of the largest metropolitan areas in the country, which is located in the eastern part of the country. KMDA was introduced by the Calcutta Metropolitan Planning Organization (CMPO) by an act in 1965. In 1966, a development plan was formed which was basically a perspective plan funded by World Bank. After its establishment, total area of KMDA was only 1380 sq. km. After 2012, few surrounding gram panchayats have been added with Kolkata Municipal Corporation (KMC). Now total area of KMA is 1841.47 sq. km consisting of different constituent units like municipal corporation, municipality, census town, out growths, and rural mouzas. KMA covers nearly six districts in West Bengal. Those are Kolkata, 24-

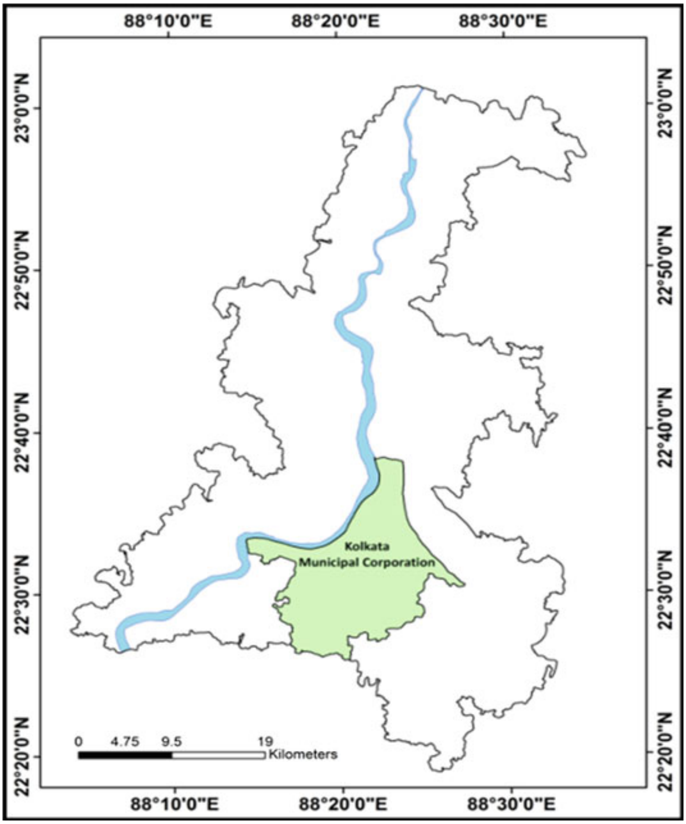


Fig. 1 Location of Kolkata metropolitan area

Parganas (South), 24-Parganas (North), Howrah, Hooghly, and Nadia, respectively. All types of planning in KMA are under the jurisdiction of KMDA (Figs. 1 and 2).

2.2 *Materials*

For the identification of land use and land cover change, supervised classification has been done using Landsat satellite images obtained from the United States Geological Survey. At the time of LULC classification, thermal images have been excluded. During the supervised classification, maximum likelihood algorithm has been used. At the time of classification, training samples have been collected from all over the areas. Nearly 1000 training sites have been chosen for the classification of the images. The area has been classified into four categories, i.e., built-up area, vegetation, water body, and fallow land, as these are major features in this area.

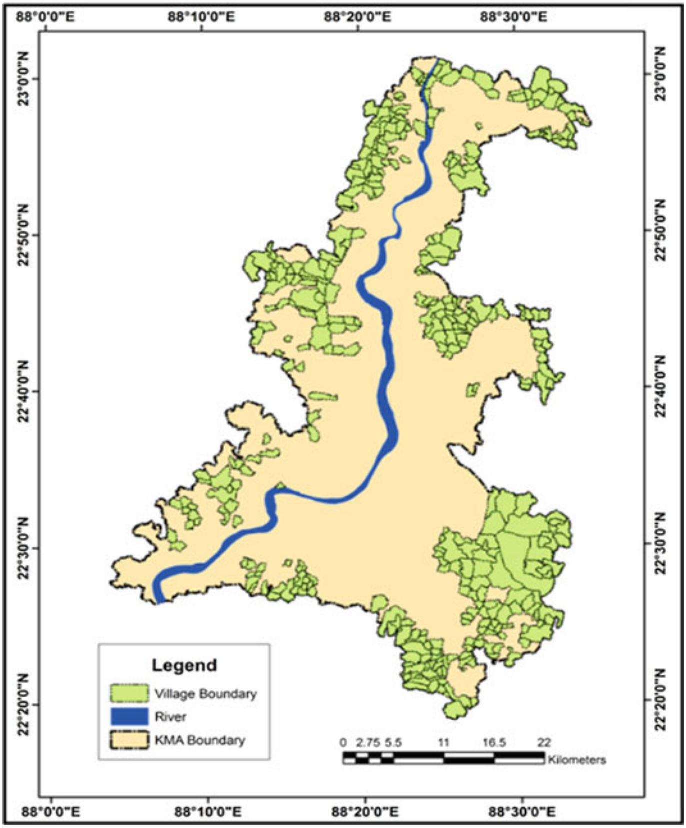


Fig. 2 Location of Kolkata rural units

To find out the drivers of urban growth, different criteria have been used. For the analysis purposes, those indicators have been categorized into three categories, i.e., socio-economic, land use, and infrastructural criteria (Table 1). To identify the forces among the rural areas, census data of the year 2011 has been used. For the identification of the driving forces Categorical Principal Component Analysis (CAT PCA) has been used using SPSS software. Before applying it, the variables have been categorized into three categories (Table 2).

3 Discussion and Analysis

Kolkata is one of the cities in the eastern parts of India, which have experienced rapid growth of population from the past few decades, which is mainly because of

Table 1 Present composition of KMA

Name of constituent units	No. of units	Area (sq. km)	Population (million)
Municipal corporations	155	364.53	18.14
Municipalities	39	667.28	73.13
Census towns	3	272.32	57.99
Out growths	6	4.68	0.17
Rural areas	371	532.66	9.48
Total KMA	574	1841.47	158.92

Table 2 Criteria for determining driving forces

Sl. No.	Name of the criteria		
Socio-economic criteria		Land use criteria	Infrastructural criteria
1.	Population (total)	Non-agricultural land (percentage)	Distance from nearest CBD (km)
2.	Population density (persons/sq km)	Vegetation (percentage)	Distance from railway track (km)
3.	Literacy (percentage of total population)	Agricultural land (percentage)	Availability of bus services (km)
4.	Workers: Male (percentage of total workers)	Waste land (percentage)	Distance to higher education (km)
5.	Workers: Female (percentage of total workers)		Distance to primary health center (km)
6.	Percentage of primary workers		Nearest to market (km)
7.	Percentage of secondary workers		Piped water supply
8.	Percentage of tertiary workers		Drainage facility
			Toilet facility

Table 3 Land use and land cover (%) of KMA

Year	1980	1990	2000	2010	2019
Built-up area	6.50	16.44	25.93	30.17	33.60
Vegetation	66.96	60.15	44.89	41.98	40.37
Water bodies	12.14	9.21	5.15	4.78	3.80
Fallow land	14.40	13.79	23.64	23.07	22.23

the influx of refugees from the neighboring areas (which includes migration from Bangladesh, other states of India, and other districts of West Bengal) (Table 3).

By analyzing the trends of land use and land cover characteristics from 1980s in and around Kolkata city, it has been found that percentage of built-up area has been increased with time, i.e., 6.50–33.60 between 1980 and 2019. In 1980, the percentage of vegetation cover is nearly 67% where percentage of vegetation in 2019 is only 40%. Because of the urban growth or the growth in built-up

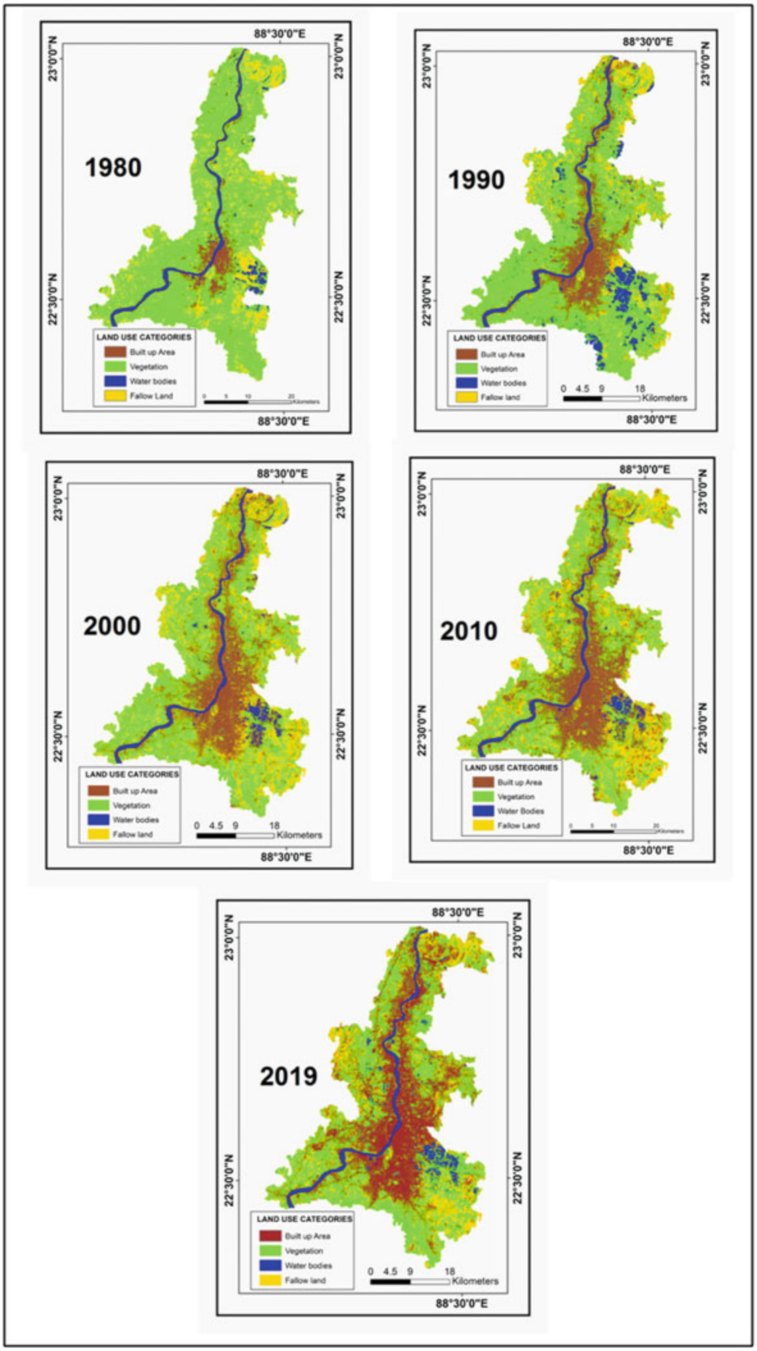


Fig. 3 Land use and land cover classification of KMA from 1980 to 2019

area, percentages of water bodies, wetland, and vegetation areas are declining very rapidly which creates adverse impacts on the local land resources. By field verification, it has been found that the change in land use and land cover is high in the peripheral areas of the city, which are basically rural mouzas of several gram panchayats. Those areas are under the jurisdiction of local panchayat institutions. There is no proper planning of these areas, for this reason the growth of these areas is totally unplanned. The growth of the population is also high in these areas. So, an attempt has been made to find out the drivers of urban growth in these areas that will help the planners to make Kolkata a smart city (Fig. 3).

After the principal component analysis, regarding the driving forces of Kolkata city, it has been found that among the socio-economic criteria population, population density, literacy, and male workers are the major criteria which attract the people to move into these areas. Population, population density, literacy, and male workers occupy nearly 90% of the total variance, as these are the major factors. These four criteria are the major driving forces that rapidly change rural areas to semi-urban areas or urban areas (Tables 4 and 5).

Among the land use criteria, it has been found that non-agricultural land and vegetation are the major criteria, which occupies nearly 99% of the total variance. Among the rural areas around Kolkata city, the rate of demographic growth is high where the percentage of non-agricultural land and vegetation is high as vegetation area can easily be converted into urban areas. For this reason during land use and land cover classification, it has been found that most of the vegetation areas have been converted into urban areas within few years of time (Table 6).

Among the infrastructural criteria, distance from the nearest CBD, distance from the railway tracks, and availability of bus service are the major indicators among the rural areas of KMA. From the field verification, it has been found that there is a tendency among the inhabitants of this area to settle down in a place, which is nearer to the state headquarters or district headquarters. In case of the distance from the nearest railway tracks it has been found that the inhabitants prefer to stay near the railway tracks because of good connectivity from the other areas. Except this, these areas act as a storehouse of different fresh vegetables, fruits, flowers, milk, fishes, etc. From these areas, those products used to supply to Kolkata for the daily needs of the city people. In the rural areas around Kolkata city except railway, bus service is most powerful criteria, which affects urban growth. By these routes, people use to go to Kolkata for their daily needs.

4 Recommendation

From the above discussion regarding the urban growth of Kolkata city, which is needed to make Kolkata a smart city, it has been found that there is no master plan in Kolkata. There is only a draft development plan by KMDA, which was planned a few decades ago. This plan is unable to address the problems, which are an impediment in making Kolkata a smart city. Kolkata is one of the old metropolitan

Table 4 Major driving forces among the socio-economic criteria

Component	Initial eigen values			Extraction sums of squared loadings		
	Total	Percentage of variance	Cumulative percentage	Total	Percentage of variance	Cumulative percentage
Population	2.263	28.30	28.301	2.264	28.30	28.3
Population density	1.894	23.71	52.009	1.897	23.71	52.0
Literacy	1.227	15.46	67.473	1.237	15.46	67.4
Workers: Male	1.016	12.57	80.042	1.006	12.57	80.0
Workers: Female	0.785	9.44	89.486			
Primary workers	0.531	7.01	96.494			
Secondary workers	0.260	3.50	100.00			
Tertiary workers	0.000	0.000	100.00			

Table 5 Major driving forces among the land use criteria

Component	Initial eigen values			Extraction sums of squared loadings		
	Total	Percentage of variance	Cumulative percentage	Total	Percentage of variance	Cumulative percentage
Non-agricultural land	2.831	70.773	70.773	2.831	70.772	70.772
Vegetation	1.144	28.597	99.370	1.144	28.597	99.370
Agricultural land	0.025	0.629	99.998			
Waste land	0.000	0.002	100.000			

cities in India, but growth of population is still high in this area. By analyzing these criteria, it has been found that among high population growth is one of the factors, which adversely impacts the natural ecosystem of the city. The density of population is also very high in the city areas as compared to the rural areas. Lack of open spaces and low price of land attract people to move outside the city areas, and urban growth takes place. Government officials make strict plan to control the demographic growth of the area. By analyzing the land use criteria, it has been found that planners should formulate a strict plan to stop the illegal conversion of the land. After analyzing the infrastructural criteria, it has been found that the growth pattern of Kolkata is still centralized. The decision makers should focus on the decentralized growth of the city, which will result in reducing the population pressure on the city.

5 Conclusion

From the above analysis, it has been found that Kolkata city is witnessing rapid peripheral expansion. For this reason, most of the land like vegetation, wetlands, and water bodies have been converted into the urban area or built-up area within the

Table 6 Major driving forces among the infrastructural criteria

Component	Initial eigen values			Extraction sums of squared loadings		
	Total	Percentage of variance	Cumulative percentage	Total	Percentage of variance	Cumulative percentage
Distance from nearest CBD	1.671	18.571	18.571	1.671	18.570	18.570
Distance from railway track	1.307	14.526	33.097	1.307	14.525	33.095
Availability of bus services	1.026	11.398	44.493	1.026	11.397	44.493
Distance to higher education	0.987	10.969	55.461			
Distance to primary health center	0.915	10.167	65.627			
Nearest to market	0.859	9.549	75.175			
Piped water supply	0.791	8.789	83.963			
Drainage facility	0.761	8.459	92.423			
Toilet facility	0.682	7.577	100.000			

few years of time. By analyzing the major driving forces of Kolkata city, it has been found that among the three different criteria, population, population density, literacy, male workers, vegetation, non-agricultural land, distance from the CBD, distance from the railway track, and availability of the bus services are the major driving forces which attract the people to move into these areas. To transform Kolkata into a smart city, policy makers should regulate or control these driving forces. It will lead to sustainable development of the city. The methods, which have been used in this chapter, will give a new perspective for developing smart cities. The methodology used can subsequently be applied to various small-sized or medium-sized cities.

References

1. United Nations. (2008). *World urbanization prospects: The 2007 revision*. United Nations Department of Economic and Social Affairs/Population Division.
2. Barredo, J. I., Demicheli, L., Lavalle, C., Kasanko, M., & Mccormick, N. (2004). Modelling future urban scenarios in developing countries: An application case study in Lagos. *Nigeria. Environment and Planning B: Planning and Design, 32*, 65–84.
3. Tian, G., Jiang, J., Yang, Z., & Zhang, Y. (2011). The urban growth, size distribution and spatio-temporal dynamic pattern of the Yangtze River Delta megalopolitan region, China. *Ecological Modelling, 222*(3), 865–878.
4. Yue, W., Liu, Y., & Fan, P. (2013). Measuring urban sprawl and its drivers in large Chinese cities: The case of Hangzhou. *Land Use Policy, 31*, 358–370.

5. Shalaby, A., & Tateishi, R. (2007). Remote sensing and GIS for mapping and monitoring land-cover and land-use changes in the Northwestern coastal zone of Egypt. *Applied Geography, 27*, 28–41.
6. Pilon, G., Howarth, J., & Bullock, A. (1988). An enhanced classification approach to change detection in semi-arid environments. *Photogrammetric Engineering and Remote Sensing, 54*, 1709–1716.
7. Matthew, L., & Wu, J. G. (2002). A gradient analysis of urban landscape pattern: A case study from the Phoenix metropolitan region, Arizona, USA. *Landscape Ecology, 17*, 327–339. https://doi.org/10.1023/A:1020512723753
8. Sui, D. Z., & Zeng, H. (2000). Modeling the dynamics of landscape structure in Asia'semerging Desakota regions: A case study in Shenzhen. *Landscape and Urban Planning, 758*, 1–16.
9. Planning Commission. (2000). *Report of the steering committee on urbanisation* (p. 232). Planning Commission, Government of India. Retrieved from http://planningcommission.gov.in/aboutus/committee/strgrp12/strrep_urban0401.pdf
10. Duranton, G. (2009). Are cities engines of growth and prosperity for developing countries? In M. Spence, P. C. Annez, & R. M. Buckley (Eds.), *Urbanization and growth* (pp. 67–114). World Bank Publications.
11. Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., & Bai, X. (2008). Global change and the ecology of cities. *Science, 319*, 756–760.
12. Spence, M., Annez, P. C., & Buckley, R. M. (2009). *Urbanization and growth*. World Bank Publications.
13. Barnes, K. B., Morgan, J. M., III, Roberge, M. C., & Lowe, S. (2001). *Sprawl development: Its patterns, consequences, and measurement*. A White Paper. Towson University. http://chesapeake.towson.edu/landscape/urbansprawl/download/Sprawl_white_paper.pdf/
14. Hedblom, M., & Soderstrom, B. (2008). Woodlands across Swedish urban gradients: Status, structure and management implications. *Landscape and Urban Planning, 84*, 62–73.
15. Herold, M., Goldstein, N. C., & Clarke, K. C. (2003). The spatiotemporal form of urban growth: Measurement, analysis and modeling. *Remote Sensing of Environment, 86*, 286–302.
16. Lassila, K. D. (1999). The new suburbanites: How American plants and animals are threatened by the sprawl. *The Amicus Journal, 21*, 16–22.
17. Litman, T. (2007). *Evaluating criticism of smart growth*. Victoria Transport Policy Institute.
18. Muniz, I., Calatayud, D., & Garcıa, M. A. (2007). Sprawl causes and effects of urban dispersion. In F. Indovina (Ed.), *The low density city* (pp. 307–347). Diputacio´ de Barcelona.
19. Weng, Q. (2001). A remote sensing-GIS evaluation of urban expansion and its impact on surface temperature in the Zhujiang Delta. *China International Journal of Remote Sensing, 22*(10), 1999–2014.
20. Iqbal, M. F., & Khan, I. A. (2014). Spatiotemporal land use land cover change analysis and erosion risk mapping of Azad Jammu and Kashmir, Pakistan. *Egyptian Journal of Remote Sensing and Space Sciences, 17*, 209–229.
21. Swerts, E., Pumain, D., & Denis, E. (2014). The future of India's urbanization. *Futures, 56*, 43–52. https://doi.org/10.1016/j.futures.2013.10.008

A New Cellular Automaton Model for Traffic Flow Based on Biham, Middleton, and Levine (BML) Model

Salma El Bakkal, Abdallah Lakhouili, and El Hassan Essoufi

Abstract Traffic light systems in cities are a very important topic to study due to many challenges such as traffic congestion, traffic accidents, and gridlock. In this chapter, we propose a new model for traffic flow in city networks based on cellular automata and modifying the basic ideas of the most known Biham, Middleton, and Levine (BML) model. This model combines the simplicity of the BML model with the characteristic of the actual urban traffic networks in order to improve and develop new results. Our model indicates that the waiting line of the intersections has a great influence on the waiting time and the average velocity of the vehicles. The experiments show that the proposed algorithm reduces the necessary time to have a free flow state from a jammed state and keeping the traffic fluency in intersections.

Keywords Traffic light systems · Cellular automata · (BML) model · Waiting time

1 Introduction

Nowadays, traffic congestion is one of the most challenging problems related to transportation network in countries. Also, managing traffic in a congested network needs a clear study of traffic flow parameters. Therefore, for almost half of a century, a lot of traffic flow theories and models have been developed to manage urban congestion. Researches in this subject started in 1955 when Lighthill, Whitham, and Richards proposed a model based on the analogy of vehicles in traffic flow

S. El Bakkal (✉) · A. Lakhouili
Laboratory Informatics, Imaging and Modeling of Complex Systems (LIIMCS) FST of Settat, Hassan 1st University, Settat, Morocco
e-mail: s.elbakkal@uhp.ac.ma; abdallah.lakhouili@uhp.ac.ma

El H. Essoufi
Mathematics, Computer Science and Engineering Sciences FST of Settat, Hassan 1st University, Settat, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_36

and the particles in the fluid, since then, the mathematical description of traffic flow has resulted two different concepts for modeling vehicular traffic. Macroscopic models focus on the complete road flow taking into account different parameters such as driver behavior, distance headways, vehicles locations... etc.. Otherwise the microscopic approaches focused on the mobility of each vehicle, the individual behavior and the interactions of each vehicle. The later models are more adapted to the study of urban traffic flow.

In recent years, most of the microscopic models developed by using the language of cellular automata (CA) [1]. Cremer and Ludwig [2] are the first who proposed the first model of cellular automaton applied to the road traffic in 1986, and the Biham, Middleton, and Levine model [3] is the first classical model applied to the urban traffic. This later describes the principal factors affecting phase transitions and identifies the different states of urban traffic flow [4]. A lot of researchers have proposed many extensions of the BML model. Fukui in 1996 [5] considered the average velocity in the BML and introduced the individual high-speed of vehicles by using the running speed of the traffic flow. Nagatani in 1995 [6] focused to reduce the numbers of gridlocks in the BML and improve the running status of the traffic by using the cloverleaf junction. Cuesta (1993) [7] and Nagatani (1995) [6] were the first to propose switch rules in the BML model. Ding et al. (2011) [8] have explored the mean field theory in the BML model, and Hu et al. [4] were the first to propose a modified BML to predict the urban traffic jams in real time.

This chapter proposes a novel automaton cellular model in order to improve the results contributed by the BML model by modifying the main rules of this later and keeping the properties of the jamming transition and the characteristics of the traffic jam. The outline of this chapter is organized as follows: Sect. 2 presents a background about the BML model, in Sect. 3 we present a detailed description of our proposal approach, in Sect. 4 we discuss the experimental results, and the conclusion and the prospects for future work are reported in the last section.

2 Backgrounds

The BML is a two-dimensional cellular automaton (CA) model with a square lattice; each cell represents an intersection of two streets (eastward street and northward street) and the space between these intersections is ignored. The cell can be in one of two states: empty or occupied by a car moving to the east or the north. In order to enable movement of the vehicles, every odd time steps the eastward vehicles attempt to advance if the cell in front is empty, whereas the northward vehicles attempt to move every even time step. In this model, the initial number of northward vehicles is equal to that of the eastward vehicles and the density of each direction is preserved because row changes are not allowed and the periodic boundary conditions are applied that means every car that leave the lattice reenters on the opposite side (see Fig. 1).

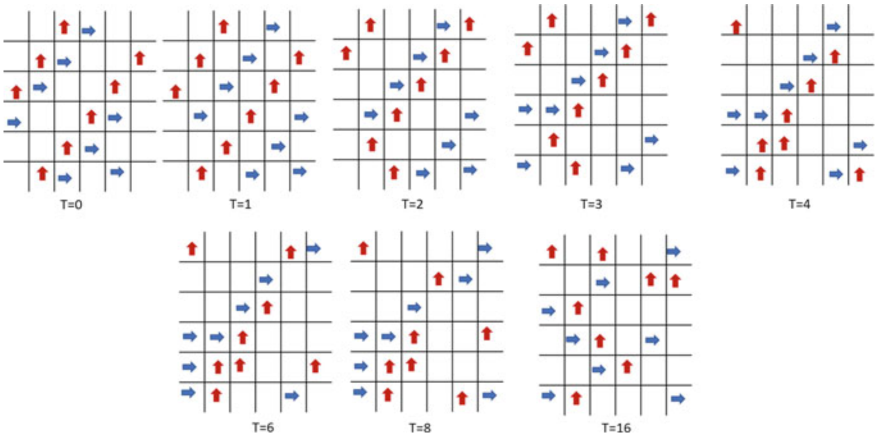


Fig. 1 Example of the BML model application

The update process of this model present two phases: the jammed phase begins to form when we start with a high density, and the cars are blocked by others cars and the free flow occurred when all cars can move at each time step with maximum velocity. In Fig. 2 we present a simple illustration of the transition phase from the free flow state to the full-jammed state.

The results of this model show that if we start with a low density, all cars can move at each time step with maximum velocity; otherwise, if we start with a high density a global jam begins to form until no cars can move as shown in Fig. 3.

We have implemented the BML model and studied its behavior via computer simulation, and we discovered some limitations:

- In the BML model, the vehicles are distributed in the eastward and northward directions with equal numbers which occurred at the diagonal form of the jammed state.
- In the BML model, a vehicle moves forward if the cell in front is empty and the vehicles behind stay in their actual position which is not consistent with the rules of the traffic flow.
- In the BML model, an eastward vehicle (resp, Northward vehicle) can move only at odd time step (resp, even time step) even if the cell in front is empty which can increase the waiting time of vehicles.
- İn the BML model, we cannot describe the transition phase from the jammed state to free flow state due to the periodic boundary conditions.

In 1996, Janos Torok and Janos Kertesz proposed an updated version of the BML mode known by the Green Wave Model (GWM) [9]. The main difference between these two models is that in the GWM if two cars are neighbors (no empty space between them), they will stay together, and nothing could separate them as shown in Fig. 4. But in the BML model, a car moves only if the cell in front is empty.

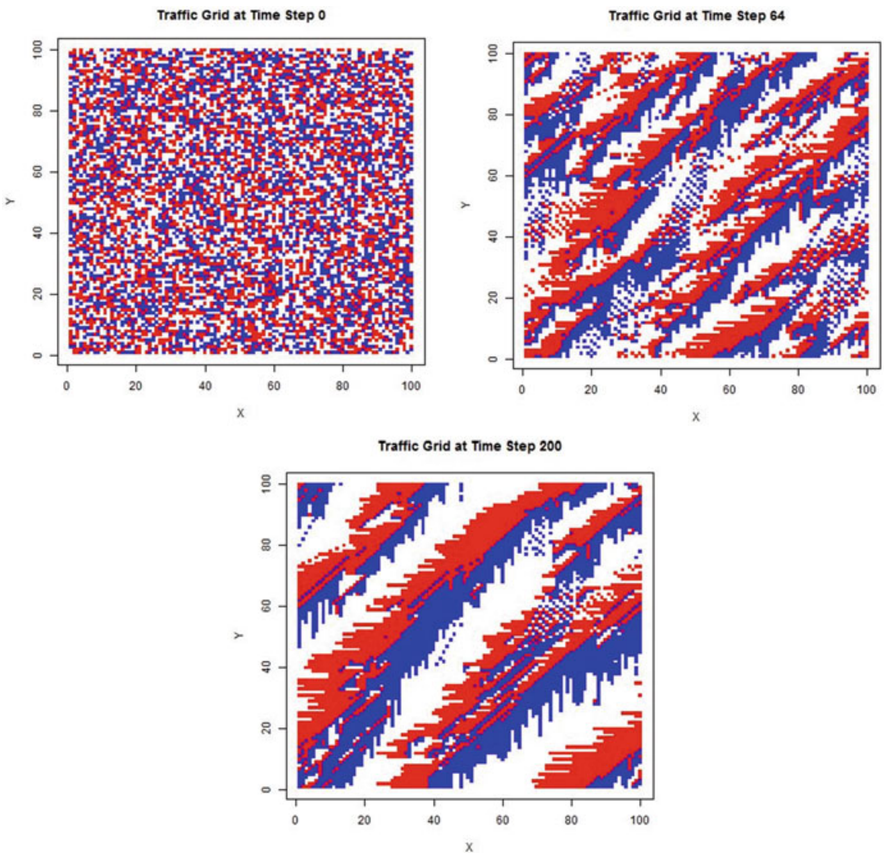


Fig. 2 The transition phase from free flow state to full-jammed state. See [10]

The results of the Green Wave Model show that the probability to have free flow state depends on the density and the size of the system. Also, if we start with a high density, we will have only jammed states, but if the densities are low, only flow states exist as shown in Fig. 5.

3 Proposal Approach

In this part, a new cellular automaton is presented to improve the results provided by the BML model. Our model is a two-dimensional automaton cellular defined with a square lattice with linear size L where the cells present an intersection between a north street and an east street, and the space between the intersections is ignored. Each cell may contain an eastward car, a northward car or is empty. The dynamic

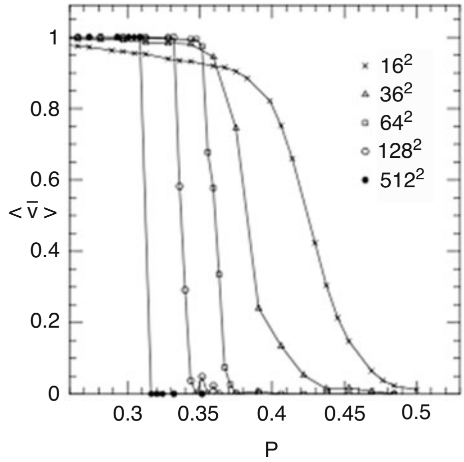


Fig. 3 The ensemble average velocity <*v*> as a function of the concentration *p* for five different system sizes. As the system size increases, the transition becomes sharper, and the ensemble average velocity changes rapidly from <*v*> = 1 to <*v*> = 0. See [3]

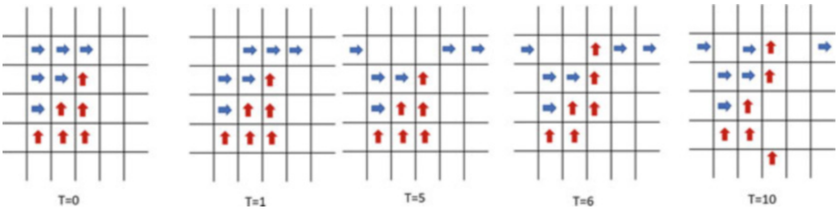


Fig. 4 Example of the Green Wave Model application

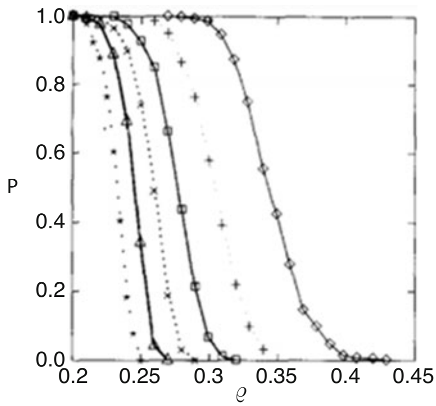


Fig. 5 The probability that a configuration goes to free flow limiting state (*P*) versus the density (¢). The different symbols represent different system sizes: $L = 32$, $L = 64$; $L = 128$; $L = 256$, $L = 512$; $L = 1024$. See [9]

of cars is controlled by a traffic light system where each direction has two states: green state means that the cars are allowed to move and read state means that the cars cannot move in this step of time.

The traffic light state is described with a Boolean variable such as follows:

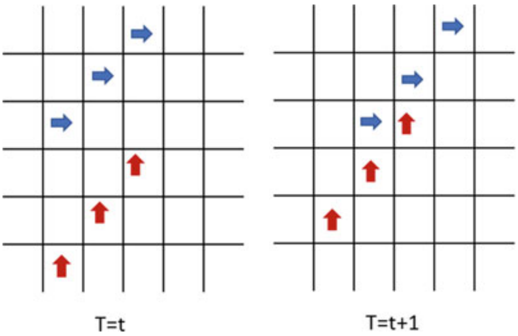


Fig. 6 Example of rule 1

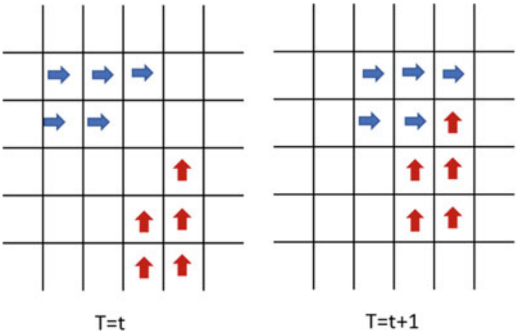


Fig. 7 Example of rule 2

$$TLSi,\ j = \begin{cases} 1, & \text{if green} \\ 0, & \text{else} \end{cases} \tag{1}$$

The state of the cells is presented in this chapter with a Boolean number that is equal to 1 if it is occupied and to 0 otherwise.

$$Ci,\ j = \begin{cases} 1, & \text{if occupied} \\ 0, & \text{else} \end{cases} \tag{2}$$

The step of time T is defined as the necessary time for a vehicle to pass to the next cell because the space between the intersections is ignored.

The dynamics of this model is controlled by a traffic light system with the below rules:

- Rule 1: every car can move if the cell in front is empty whatever the value of the step of time as shown in Fig. 6.
- Rule 2: if a car moves, all the cars behind move in the same step of time (see Fig. 7).
- Rule 3: if we have a northward and an eastward vehicle attempt to move to the same cell, we calculate the waiting line of each direction (waiting line of the

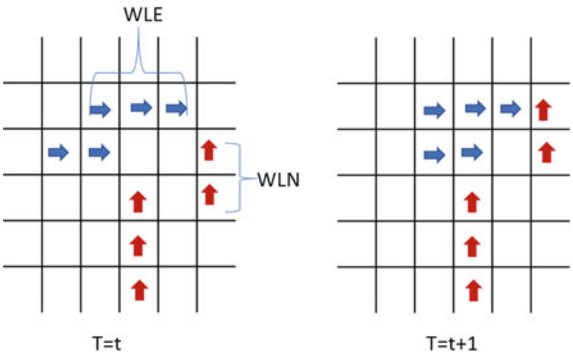


Fig. 8 Example of rule 3

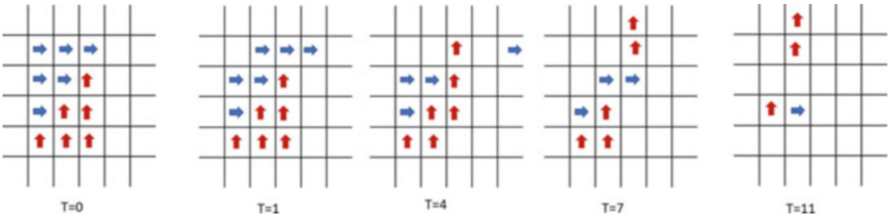


Fig. 9 Example

eastward WLE (3) and the waiting line of the northward WLN (4)) and which has the smallest value has to move the next step of time (see Fig. 8).

$$\text{WLE}i,\, j = \sum_{k=0}^{i} Ck,\, j \tag{3}$$

$$\text{WLN}i,\, j = \sum_{k=j}^{L} Ci,\, k \tag{4}$$

In order to study the transition phase from the jammed state to the free flow state, we suppose in this chapter that the periodic boundary conditions are not applied which means every car leave the lattice, the total density decreases.

Considering the example shown in Fig. 9, this example presents a cellular automaton with $L = 6$, $p = 0.33$. Starting with this example, we need 11 steps of time to have a free flow state; we suppose that the step of time is equal to 20 s, so 220 s is the required time to move from the jammed state to the free flow state.

4 Results and Discussions

Most researchers focus on the transition phase from the free flow state to jammed state of the BML model. But in this chapter, we focus to study the necessary time to have a free flow state starting with a full-jammed state for that we evaluate our proposed model in different size of systems with different densities as shown in Fig. 10.

Figure 11 presents a comparison of the performance of our system using our proposed approach and BML model. The results show that our approach is more performant than the BML model in terms of the time required to have the free flow state. Note that in this example, density $p = 0.8$.

In order to show the robustness of our model, we studied the average waiting time of the two models. The results of this study show that, with our model, vehicles have more chance to leave the lattice early than with the BML model as shown in Fig. 12. Note that in this example density $p = 0.5$.

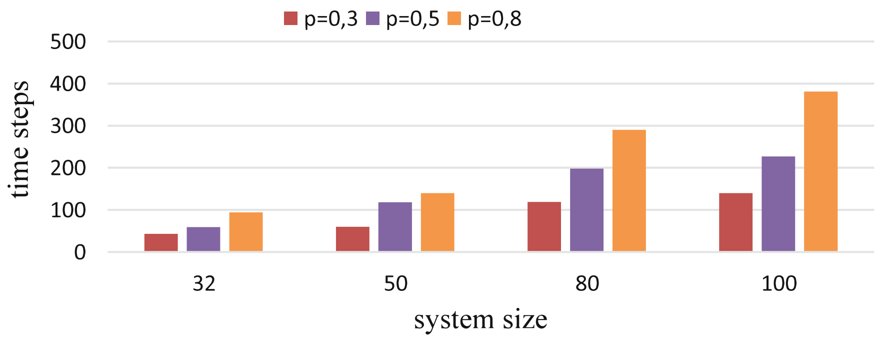


Fig. 10 The corresponding histogram of the results

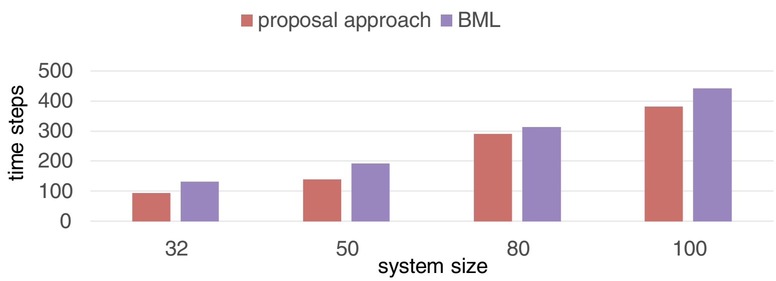


Fig. 11 Comparison between the proposed approach and the BML model

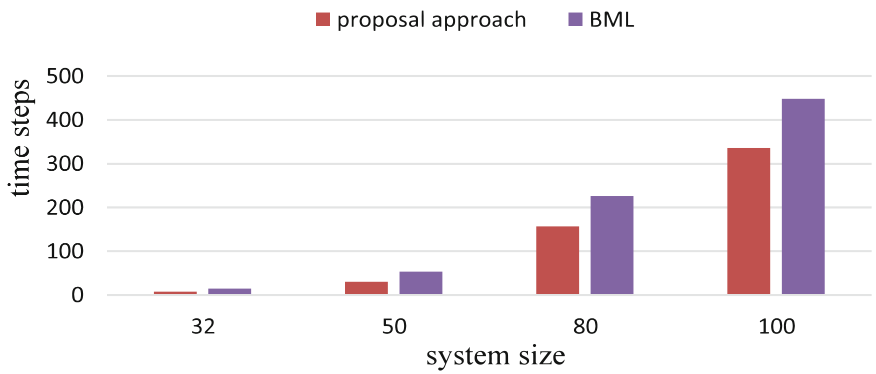


Fig. 12 The histogram of the average waiting time

5 Conclusion

We have presented in this chapter a new cellular automaton based on the BML model which describes the traffic flow in two dimensions and using the waiting line as a parameter to manage the traffic light system. The main advantage of the proposed method is that it keeps the traffic fluency in intersections. In comparison to the BML model, the experimental results show that our approach has the ability to reduce the waiting time of the vehicles and the time required to move from the jammed state to the free flow state.

İn the future work, we will use another parameter to manage the system such as the free space between the intersections, we may also use a multi-agent system to control the communication between the flows.

References

1. Brockfeld, E., et al. (2001). Optimizing traffic lights in a cellular automaton model for city traffic. *Physical Review E, 64*, 056132.
2. Cremer, M., & Ludwig, J. (1986). A fast simulation model for traffic flow on the basis of Boolean operations. *Mathematics and Computers in Simulation, 28*, 297.
3. Biham, O., Middleton, A. A., & Levine, D. (1992). Self-organization and a dynamical transition in traffic flow models. *Physical Review A, 46*(10), R6124–R6127.
4. Hu, W., Yan, L., Wang, H., Du, B., & Tao, D. (2017). Real-time traffic jams prediction inspired by Biham, Middleton and Levine (BML) model. *Information Sciences, 381*, 209–228.
5. Fukui, M., & Ishibashi, Y. (1996). Traffic flow in 1d cellular automaton model including cars moving with high speed. *Journal of the Physical Society of Japan, 65*(6), 1868–1870.
6. Nagatani, T. (1995). Self-organization in 2D traffic flow model with jam-avoiding drive. *Journal of the Physical Society of Japan, 64*(4), 1421–1430.
7. Cuesta, J. A., Martínez, F. C., Molera, J. M., & Sánchez, A. (1993). Phase transitions in two-dimensional traffic-flow models. *Physical Review E, 48*(6), 4175–4178.

8. Ding, Z.-J., Jiang, R., Huang, W., & Wang, B.-H. (2011). Effect of randomization in the bihammiddletonlevine traffic flow model. *Journal of Statistical Mechanics, P06017*.
9. Török, J., & Kertész, J. (1996). The green wave model of two-dimensional traffic: Transitions in the flow properties and in the geometry of the traffic jam. *Physica A: Statistical Mechanics and Its Applications, 231*(4), 515–533.
10. Davies, J. *Biham-Middleton-Levine Traffic Model*. https://www.jasondavies.com/bml/

A Multicriteria Decision Model for Optimizing Costs and Performances for a Cloud User

Youcef Bezza, Ouassila Hioual, and Ouided Hioual

Abstract Cloud computing is a recent technology in the field of computing. This paradigm is gaining a great place in everyday life by providing services on demand for the benefit of clients. Optimizing the characteristics of cloud providers has become indispensable so combinatorial optimization is a necessary mechanism combining several techniques of discrete mathematics and computer science to solve real-life difficulties. Given the number of cloud providers that are very large and the variety of criteria for choosing them, decision making has become increasingly difficult and requires a robust multicriteria system. In this chapter, we propose a model of a multicriteria decision support system for the optimization of costs and performances in cloud computing. The proposed model uses a modified neural network with two hidden layers: the first one uses the ELECTRE technique and the second a modified Tabu search algorithm; this modification consists of using the weighted sum instead of the objective functions. The input for our modified neural network is cloud performance.

Keywords MCDA · Cloud computing · Combinatorial optimization · Artificial neural networks

1 Introduction

How quickly things change, cloud computing has changed from an uncertain and confusing concept to a plan that large and small organizations are beginning to adopt as part of their overall computing plan. Cloud computing is a process of providing a group of shared computing resources that contain applications, computing, storage,

Y. Bezza (✉)
ICOSI Laboratory Khenchela, Abbes Laghrour of Khenchela, Khenchela, Algeria

O. Hioual · O. Hioual
LIRE Laboratory Constantine, Abbes Laghrour of Khenchela, Khenchela, Algeria

Abbes Laghrour of Khenchela, Khenchela, Algeria

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_37

development, and deployment platforms as well as business processes [1]. Cloud computing is currently commonly used to designate the delivery of software, infrastructure, and storage services over the Internet.

Customers of the cloud can profit from other organizations providing services related to their data, software, and other computing needs on their behalf, without the need to own or run the usual physical hardware and software themselves. Given the number of cloud providers that are very large and the variety of criteria for choosing them, decision-making has become increasingly difficult. Multicriteria decision-making is a famous branch of decision-making. It is a branch of a general class of operation research models which deal with decision and problems under the presence of a number of decision criteria. This major class of models is very often called MCDM [2]. In the presence of multiple criteria, a unique optimal decision for the problem does not exist but rather many or even infinitely many decisions are suitable [3]. Multicriteria decision aid (MCDA) is an advanced field of operations research, which provides decision-makers (DMs) and analyzes a wide range of methodologies, which are well suited to the complexity of decision problems [4].

One of the important things in our world is the search for optimal states. In our case, we search to optimize cost and system performances. Optimization is a very important subject especially in multicriteria problems. An optimization problem is a situation which requires deciding for one choice from a set of possible alternatives in order to reach a predefined/required benefit at minimal costs. Optimization is the procedure of resolving an optimization problem, i.e., finding suitable solutions for it [5]. The need for cloud services is increasing day by day as the number of providers, and cost is a very important factor in choosing the right provider. The criteria are multiple and contradictory which makes the decision-making very difficult, for that we propose a multicriteria decision support model that optimizes costs and performances of cloud providers and helps the customer to choose the best and the optimal providers. Our model is a combination of optimization methods and MCDA algorithms. The structure and running are inspired from the neural network, and we use a modified neural network which is inspired from the basic multilayer neural network with two hidden layers. The first one is presented with ELECTER method, and the second is presented with a modified Tabu search algorithm using the weighted sum.

In this chapter, we present a multicriteria decision support system for the optimization of costs and performance in cloud computing. We start with a general introduction, and then, we present some related works. In the third section, we detail and explain our architecture and the functioning of our system and illustrate it with a case study. Finally, we conclude the chapter with a conclusion and give some perspectives.

2 Related Work

In the present section, we focus on some work for optimizing cloud computing services selection. Seyedali Mirjalili and Andrew Lewis [6] proposed a novel nature-inspired meta-heuristic optimization algorithm, called Whale Optimization Algorithm (WOA). The algorithmic rule is inspired by the bubble-net searching strategy. WOA is tested with 29 mathematical optimization problems and six structural design problems. WOA was very competitive with meta-heuristic optimizers and superior over conventional techniques.

In 2014, Heba Kurdi and et al. [7] developed a novel COMbinatorial optimization algorithm for cloud service COMposition (COM2) that can efficiently use multiple clouds. The proposed algorithm, COM2, guarantees that the cloud with the maximum number of services will always be selected before other clouds, which increases the possibility of fulfilling service requests with minimal overhead. The results of the study suggest that the COM2 algorithm retains a low number of combined clouds without compromising the number of examined services, which significantly affects execution time.

Yuhui Shi et al. [8] suggested a new multi-objective brain storm optimization algorithm in which the clustering plan is applied in the objective space instead of in the solution space in the original brain storm optimization algorithm for solving single objective optimization problems. Two versions of multi-objective brain storm optimization algorithmic rule with completely different characteristics of divergent operation were verified to check the utility and effectiveness of the steered algorithmic rule. Experimental results show that the planned multi-objective brain storm optimization algorithmic rule could be a terribly promising algorithmic rule.

In [9], the authors proposed a modified Self-adaptive Multi-objective Brain Storm Optimization (SMOBSO). Instead of the k-means clump of the normal algorithmic rule, the algorithmic rule adopts the straightforward clump operation to extend the looking speed. At the same time, the open likelihood is introduced to avoid the algorithmic program stable gear into native optimum. Associate in nursing adjustive mutation methodology is employed to grant an uneven distribution on solutions. The proposed algorithm is tested on five benchmark functions; and therefore, the simulation results showed that the changed algorithmic rule increase the variety additionally because of the convergence with success.

Xin-She Yang et al. [10] proposed a framework for self-tuning algorithms so that an algorithm to be tuned can be used to tune the algorithm itself. It is found that totally different (completely different) parameters might have different sensitivities and thus require different degrees of tuning. Parameters with high sensitivities need fine-tuning to attain optimality. Parameter tuning is the process of tuning an algorithm that is used to find the best parameter settings so that an algorithm can perform the best for a given set of problems.

3 Proposed Model

Cloud computing is characterized by consumers who use cloud services as needed, consume shared resources as a service that can rapidly and elastically go down or scale up as needed, pay only for what is used and access services over a networked infrastructure [11]. Cloud Service Providers (CSPs) (e.g., Google, Microsoft, Amazon) are vendors who lease to their customers cloud computing resources and services that are dynamically utilized based on customer's demand according to a certain business model [12]. With the advancement of IT, the number of cloud providers is increasing rapidly. So, it is so hard to make a decision about which cloud provider a consumer can choose. Because of this, we propose a model who help consumers to choose the best provider who meets the needs and optimizes costs and performance.

3.1 Architecture of Our System

Figure 1 illustrates the general architecture of our system which is composed of a neural network and a set of cloud computing providers.

Our model is composed of a multilayer neuron network but not a basic or standard one, it means it is modified and adapted to our context. The input of our neuron is the cloud providers. Each provider has parameters, and these later represent the criteria of making a choice. The first hidden layer works with the ELECTRE (Elimination and Choice Translating algorithm) method which is used to rank proposed providers and extract the best ones. The second hidden layer is

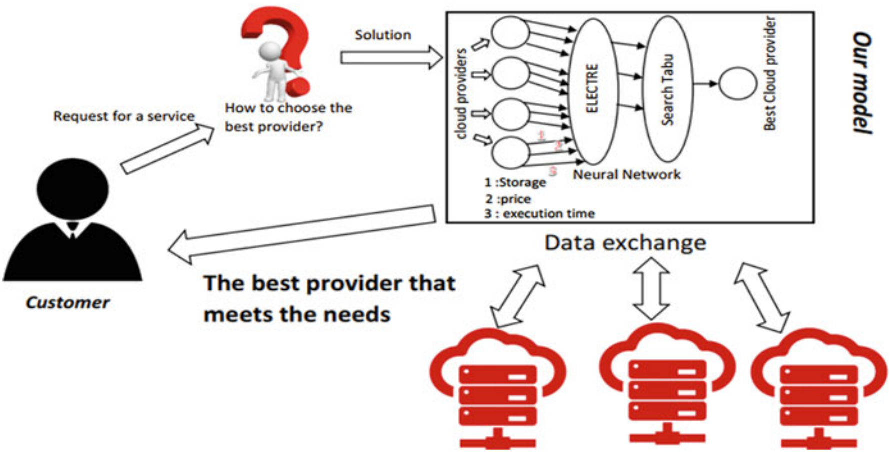


Fig. 1 Architecture of our model

represented by Tabu search algorithm; this algorithm explores the best providers that are transferred from the first layer and gives as result the optimal provider that meets the customer needs. Thus, this second layer is used for optimizing the results of the first hidden one.

3.2 Functionality of the Proposed Model

In our system, the input is data extracted from the data center of each candidate provider; these data are storage space offered to the end user, price, and execution time. These three criteria are transferred to the first hidden layer in order to work with in the ELECTRE method. Then, when we obtain the rank of the best cloud providers, we sent the best providers to the second hidden layer which work with the algorithm of "Tabu search" (cf. Fig. 2). After running the Tabu search algorithm, we can get the best optimized cloud provider.

3.3 Background of ELECTRE and Tabu Research Algorithm

ELECTRE ELECTRE is a multicriteria decision-making procedure that can be useful when a set of alternatives must be classified according to a set of criteria reflecting the decision maker's preferences [13]. To start with, ELECTRE needs the evaluation of two indices, the concordance index and the discordance index, defined for each pair of alternatives.

ELECTRE is composed of two main steps. The first step is the preparation of the Decision Matrix (cf. Table 1), where *vij* represents the value of alternative *i* with respect to criteria *j*.

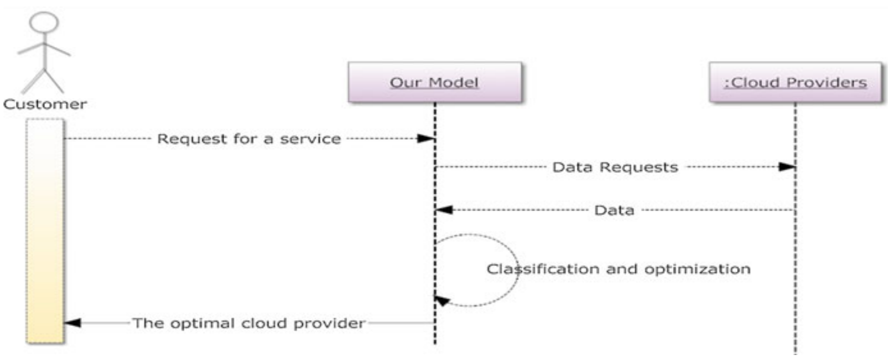


Fig. 2 Sequence diagram

Table 1 Decision matrix

Alternatives		Criteria			
		C1	C2	...	Cn
	a1	V11	V12	...	V1n
	a2	V21	V22	...	V2n
	...	...	...	...	...
	an	Vn1	Vn2	...	Vnn

The second step is the computation of concordance and discordance matrix.

Computation of Concordance Matrix Concordance matrix is generated by adding the values of weights of concordance set elements. The strength of the hypothesis that alternative Ai is at least as good as alternative Aj is measured by the concordance index between the pair of alternatives Ai and Aj and is calculated by (1) [14]:

$$C\,(a,b)=\frac{\forall j: gj\,(a)\geq gj\,(b)\sum kj}{k}\ \text{with}\ k=\sum_{j=1}^{n}kj \tag{1}$$

Computation of Discordance Matrix The discordance index D (a, b) is defined by (2) and (3):

$$D\,(a,b)=0\ \text{if}\forall j,\, gj\,(a)\geq gj\,(b) \tag{2}$$

If not

$$D\,(a,b)=\frac{1}{\delta}\,\text{MAX}\ j\,[gj\,(b)-gj\,(a)] \tag{3}$$

where δ is the maximum difference between the same criterion for two actions given.

The overclassing relation for ELECTRE is constructed by comparing the concordance and unconformity indices with thresholds of concordance c and unconformity d, a outclasses b, if:

$$\text{aSb} \iff C\,(a,b)\geq c\ \text{et}\ D\,(a,b)\leq d.$$

3.3.1 Tabu Search

Basic Ingredients of Tabu Search

A chief way to exploit memory in Tabu search is to classify a subset of the moves in a neighborhood as forbidden (or tabu) [15].

A neighborhood is constructed to identify adjacent solutions that can be reached from the current solution [16].

The classification depends on the history of the search, and particularly on the regency or frequency that certain move or solution components, called attributes, have participated in generating past solutions [15].

A tabu list records forbidden moves, which are referred to as tabu moves [17].

Tabu restrictions are subject to an important exception. When a tabu move has a sufficiently attractive evaluation where it would result in a solution better than any visited so far, then its tabu classification may be overridden. A condition that allows such an override to occur is called an aspiration criterion [15].

Basic Tabu Search Algorithm

Step 1: Choose an initial solution i in S. Set i* = i and k=0.
Step 2: Set k=k+1 and generate a subset V* of solution in N(i,k) such that either one of the Tabu conditions is violated or at least one of the aspiration conditions holds.
Step 3: Choose a best j in V* and set i=j.
Step 4: If f(i) < f(i*) then set i* = i.
Step 5: Update Tabu and aspiration conditions.
Step 6: If a stopping condition is met then stop. Else go to Step 2.

Tabu Search Stopping Conditions. Some immediate stopping conditions could be the following [18]:

1. N(i, K+1) = 0. (no feasible solution in the neighborhood of solution i).
2. K is larger than the maximum number of iterations allowed.
3. The number of iterations since the last improvement of i* is larger than a specified number.
4. Evidence can be given than an optimum solution has been obtained.

Modified Tabu Search Algorithm Because of the criteria which are heterogeneous, we must calculate their global performances. For that, we will use the weighted sum. The weighted sum tries to calculate for each action (ai) its global performance given by *V* (ai). This overall performance is calculated using the expression (4):

$$\mathrm{V}\,(\mathrm{ai}) = \sum_{j=1}^{n} wj * eij \tag{4}$$

where *wj* represents the coefficient of importance of the criterion *j* and *eij*, the evaluation of the action ai with regard to the criterion *j*.

So, the modified tabu search algorithm is as follows:

Step 1: Choose an initial solution i in S. Set i* = i and k=0.

Step 2: weighted sum function.
Step 3: Set k=k+1 and generate a subset V* of solution in N(i,k) such that either one of the Tabu conditions is violated or at least one of the aspiration conditions holds.
Step 4: Choose a best j in V* and set i=j.
Step 5: If f(i) < f(i*) then set i* = i.
Step 6: Update Tabu and aspiration conditions.
Step 7: If a stopping condition is met then stop. Else go to Step 3.

4 Case Study

This section examines a scenario of a customer to illustrate, first, the steps or procedures that must be followed according to our model, and second, the interaction between the different components of the proposed architecture. We assume that a customer search for a provider to buy some storage space, he has a list of some cloud provider, but which provider meets the needs?

Initially, we have the following performance table of cloud providers (Table 2):

The importance of each criterion in decision-making is translated by a weight *kj* such that:

4.1 Computation of Concordance Matrix

We calculate the computation of the concordance index as follows:

$$C\text{ (Dropbox, Box)} = \frac{0+3+2}{8} = \frac{5}{8} = 0.625$$

$$C\text{ (Microsoft One Drive, Box)} = \frac{3+3+2}{8} = \frac{8}{8} = 1$$

Table 2 Performance table

	Storage	Price	Execution time
Dropbox	2 Go	7€/1 To	0.25 ns
Box	10 Go	8€/100 Go	0.80 ns
Google drive	15 Go	9.99€/1 To	0.73 ns
Microsoft one drive	5 Go	7€/1 To	0.25 ns
Amazon	5 Go	4€/1 To	0.73 ns
Apple Icloud	5 Go	9.99 €/2 To	0.78 ns
Hubic	25 Go	5€/1 To	0.83 ns

4.2 Computation of Discordance Matrix

The discordance index is calculated as follows:

$$\delta = 25\text{–}2 = 23$$

$$D\,(\text{Dropbox, Box}) = \frac{8}{23} = 0.34$$

The limited concordance = 1
The limited discordance = 0.26
So, we have:
Google Drive S Box which mean Google Drive overclass Box
Microsoft One Drive S Dropbox; Amazon Cloud Drive S Apple Icloud.
After analyzing the two matrixes, we obtained: the four best providers are: Google Drive, Microsoft One Drive, Amazon, Hubic.

4.3 Modified Tabu Search

Now, we have the four best providers, we will extract the best one using the modified Tabu search algorithm which represents the second hidden layer of our modified neural network. The progress of the algorithm is as shown in Fig. 3.

S:{Google Drive, Microsoft One Drive, Amazon, Hubic}, S is the set of initial solutions.

we assume that google drive is s0 and we set s*=s0,s* The best solution so far.
s:new solution from the neighborhood of s*.
f(s): objective function.
f(s*): the value of the best solution.
l: tabu list, initially is empty.

Stop condition is: the number of element in tabu list is the number of elements of initial list −1.

The supposed weights for each criterion of the weighted sum are in Table 3. We choose a solution from the neighborhood s of s0, s=hubic.

We calculate the object function of each provider:
f(google drive)=15*3+9.99*3+0.73*2=351.50
f(hubic)=25*3+5*3+0.78*2=481.56
f(google drive)<f(hubic).

So, we put hubic in tabu list, and we choose another solution from the initial solution list.

l:{hubic}.
s:Amazone.
f(Amazone)=43.07

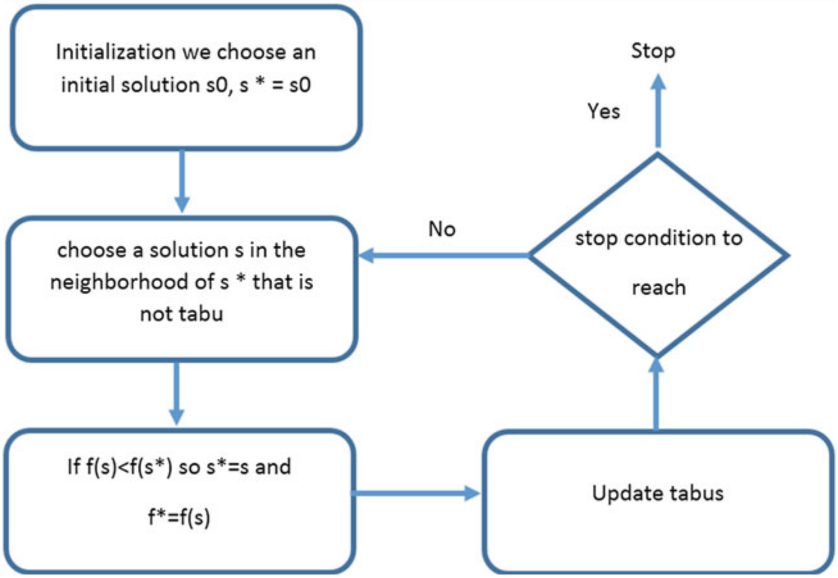


Fig. 3 Operation of the algorithm

Table 3 Performances weight table

Criteria	Storage	Price	Execution time
Weight (kj)	3	3	2

f(google drive)>f(amazon), we put google drive in the tabu list and s*=Amazon.

We choose another solution from the initial list: s=Microsoft one drive.

f(Microsoft one drive)=132,5.f(Amazon)<f(Microsoft one drive). So, we put Microsoft on the tabu list.

Finally, the stop condition is checked and Amazon is the best provider. The result of our modified neural network is Amazon.

5 Conclusion

In this chapter, we presented a new multicriteria decision support model for costs and performances optimization in cloud computing using the modified Tabu search algorithm and a modified neural network with two hidden layers. The first layer uses the ELECTRE method and the second uses the modified Tabu search algorithm. The obtained results are reliable. In the future, we will improve our model using other optimization algorithms and multicriteria analysis methods with several criteria of choice.

References

1. Hurwitz, J., Kaufman, M., & Halper, F. (2012). *Cloud services for Dum-mies®*. IBM Limited Edition, Wiley.
2. Pohekar, S. D., Ramachandran, M. (2003). *Application of multi-criteria decision making tosustainable energy planning—A review*.
3. Wieceka, M., Ehrgott, M., Fadel, G., et al. (2008). Multiple criteria decision making for engineering. *Omega, 36*, 337–339.
4. Zopounidis, C., & Doumpos, M. (2002). Multi-criteria decision aid in financial decision making: Methodologies and literature review. *Journal of Multicriteria Decision Analysis, 11*, 167–186.
5. Weise, T. (2011). *Global optimization algorithms–theory and application* (3rd ed.).
6. Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in Engineering Software, 95*, 51–67.
7. Kurdi, H., Al-Anazi, A., Campbell, C., & Al Faries, A. (2015). A combinatorial optimization algorithm for multiple cloud service composition. *Computers and Electrical Engineering, 42*, 107–113.
8. Shi, Y., Xue, J., & Wu, Y. (2013). Multi-objective optimization based on brain storm optimization algorithm. *International Journal of Swarm Intelligence Research, 4*(3), 1–21.
9. Guo, X., Wu, Y., Xie, L., Cheng, S., & Xin, J. (2015). *An adaptive brain storm optimization algorithm for multiobjective optimization problems* (pp. 365–372). Springer. Y. Tan et al. (Eds.): Part I, LNCS 9140. https://doi.org/10.1007/978-3-319-20466-639
10. Yang, X.-S., Deb, S., Loomes, M., & Karamanoglu, M. (2013). A framework for self-tuning optimization algorithm. *Neural Computing and Applications, 23*(7–8), 2051–2057.
11. Carroll, M., van der Merwe, A., & Kotzé, P. (2011). *Secure cloud computing: Benefits, risks and controls*. IEEE. https://doi.org/10.1109/ISSA.2011.6027519
12. Prodan, R., & Ostermann, S. (2009). A survey and taxonomy of infrastructure as a service and web hosting cloud providers. In *10th IEEE/ACM internation conference on grid computing*.
13. Aiello, G., Enea, M., & Galante, G. (2006). *A multi-objective approach to facility layout problem by genetic search algorithm and Electre method*. Elsevier. https://doi.org/10.1016/j.rcim.2005
14. Sahaaya Arul Mary, S. A., & Suganya, G. (2016). Multi-criteria decision making using ELECTRE. *Circuits and Systems, 7*, 1008–1020. https://doi.org/10.4236/cs.2016.76085
15. Glover, F., Kelly, J. P., & Laguna, M. (1995). Genetic algorithms and tabu search: Hybrids for optimization. *Computers and Operations Research., 22*(1), 111–134.
16. Reeves, C. R. (1993). *Modern heuristic techniques for combinatorial problems*. Wiley.
17. Hillier, F. S., & Lieberman, G. J. (2005). *Introduction to operations research* (8th ed.). McGraw-Hill.
18. Hertz, A., Taillard, E., & Werra, D. *A tutorial on tabu search*. http://www.cs.colostate.edu/~whitley/CS640/hertz92tutorial.pdf

Improving Throughput for Mobile Nodes

Aboulwafa Mohamed and Enneya Nourddine

Abstract With traffic increase in a wireless network beyond its capacity, the quality of service (QoS) degrades. In this chapter, we review the impact of mobility on throughput, in the case of infrastructure networks using IEEE 802.11 Wi-Fi standard. The study was made by simulating different scenarios using Network Simulator-3 (NS-3). First, we compare the throughput between static and mobile nodes, both connected to a QoS station. We found that the throughput may vary depending on the simulation scenario. Second, we propose a modification to EDCA parameters by adding a new access category reserved for mobile nodes regardless of their original access category. The tests show that with the proposed solution the mobile nodes can have a better throughput.

Keywords EDCA · Throughput · Quality of service · Mobility

1 Introduction

The IEEE 802.11 protocol has become the most dominant and is vastly used in Wireless Local Area Networks (WLANs) [1]. IEEE 802.11 WLANs are easily deployed with low cost in zones like airports, office, hotels, and residence homes. With an increasing popularity of multimedia applications (VoIP and streaming) that requires a guaranteed QoS. QoS stations (STAs) allow a better use of the network by providing prioritized QoS services which classify all the traffic to multiple Access Categories (ACs). The protocol includes the 802.11e standard that proposes a new function for the MAC layer, known as the hybrid coordination function (HCF). This function uses a contention channel access method based on EDCA. EDCA is designed to provide prioritized QoS and improve the distributed coordination function (DCF) belonging to the original 802.11 standard.

A. Mohamed (✉) · E. Nourddine
Informatics, Systems and Optimization Laboratory, Faculty of Science, Ibn Tofail University, Kénitra, Morocco
e-mail: enneya@uit.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_38

It is important to study the impact of mobility on different protocols and scenarios as the performance of different network protocols can be affected in different ways. In this chapter, we chose to study the impact of mobility on throughput in IEEE 802.11n protocol, and we choose random walk mobility model because it is close to a normal usage of wireless devices around the office or in a public space.

The aim of this study is to compare the throughput between static and mobile nodes in the case of classic EDCA parameters and in the case of our new access category.

This chapter is organized as follows: Sect. 2 describes some related work, Section 3 describes basic theory of WLAN IEEE 802.11e and EDCA MAC layer. In Sect. 4, we evaluate the impact of mobility on throughput via simulations, and present the results of our new model. Finally, Section 5 concludes the chapter.

2 Related Work

There are quite a few attempts to study the impact of mobility on the performance of Wi-Fi networks. In this chapter, for brevity, we refer to only a selected set of literature.

In [1], the authors tested the mobility impact on throughput and found that under medium and high mobility, the real-time traffic is penalized.

In [2], they conclude that under mobility conditions, a new concept for admission control is required that constantly monitors the channel conditions and reevaluates admission control decisions.

In [3], the simulations show that the velocity of the nodes impacts the number of packets that are successfully received by the destination node.

The authors in [4] have showed that at the time of movement nodes, the throughput scaling changes completely.

In [5] where the authors made a real-life experiment, they found that mobility is the most dominant cause of link failure for links with a long lifetime.

3 Brief Description of 802.11e EDCA

3.1 Distributed Coordination Function (DCF)

The fundamental access method of the IEEE 802.11 MAC is a DCF known as carrier sense multiple access with collision avoidance (CSMA/CA). The DCF must be implemented in all STAs. For an STA to transmit, it have to sense the medium to determine if another STA is transmitting. If the medium is not determined to be busy, the transmission may proceed.

3.2 IEEE 802.11e Standard

The IEEE 802.11e standard introduces the enhanced distributed channel access (EDCA). While backward compatible with distributed coordination function (DCF), EDCA enhances DCF by providing a distributed access method that can support service differentiation among different classes of traffic. The DCF is supposed to provide a channel access with equal probabilities to all stations contending for the channel access in a distributed manner. However, equal access probabilities are not desirable among stations with different priority frames. The EDCA is designed to provide differentiated, distributed channel accesses for frames with eight different priorities (from 0 to 7) by enhancing the DCF (Table.1).

To access the channel, an STA shall determine that the medium is idle through the use of the carrier sense function for the interval specified. The time interval between frames is called the inter-frame space (IFS) [6] (Fig. 1).

Table 1 UP-to-AC mappings

Priority	UP (same as 802.1D user priority)	802.1D designation	AC	Designation (informative)
Lowest	1	BK	AC_BK	Background
↓	2	–	AC_BK	Background
	0	BE	AC_BE	Best effort
	3	EE	AC_BE	Best effort
	4	CL	AC_VI	Video
	5	VI	AC_VI	Video
	6	VO	AC_VO	Voice
Highest	7	NC	AC_VO	Voice

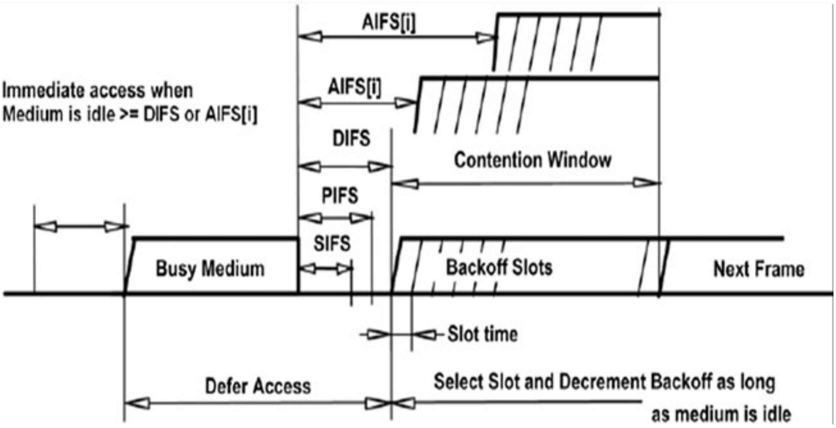


Fig. 1 Some IFS relationships

The Inter-Frame Spaces are:

SIFS: Short Inter-Frame Space
PIFS: PCF Inter-Frame Space
DIFS: DCF Inter-Frame Space
AIFS: Arbitration Inter-Frame Space (used by the QoS facility).

3.3 EDCA

The EDCA mechanism provides differentiated, distributed access to the wireless medium for STAs using eight different user priorities (Ups). The EDCA mechanism defines four access categories (ACs) that provide support for the delivery of traffic with UPs at the STAs. The AC is derived from the UPs as shown in Table 1 [7].

In order to achieve traffic differentiation, EDCA assigns higher priority ACs with smaller CWmin, CWmax, and AIFS to influence the successful transmission probability (statistically) in favor of high-priority ACs. The smaller the parameter values (such as AIFS, CWmin, and CWmax), the greater the probability of gaining access to the medium. Each AC within a station behaves like an individual virtual station: it contends for access to the medium and independently starts its backoff procedure after detecting the channel being idle for at least an AIFS period. The backoff procedure of each AC is the same as that of DCF. When a collision occurs among different ACs within the same station, the higher priority AC is granted the opportunity to transmit, while the lower priority AC suffers from a virtual collision, similar to a real collision outside the station. Figure 3 [8] shows the four access categories inside a QoS station (Fig. 2).

The preferred parameter values of each AC and the standard recommends are shown in Table 2 [9].

Besides the parameters stated above, there is another parameter called transmission opportunity (TXOP limit) which represents the maximum allowed duration for each acquired transmission opportunity, a station may transmit multiple frames within the assigned TXOP limit.

3.4 Quality of Service Metrics

Quality of service (QoS) is defined as the measure of performance for a transmission system that reflects its transmission quality and service availability. Service availability is a crucial element of QoS. Before QoS can be successfully implemented, the network infrastructure must be highly available [10]. The network transmission quality is determined by:

Delay: (or latency) is the amount of time it takes for a packet to reach the receiving end point after being transmitted from the sending end point. This time

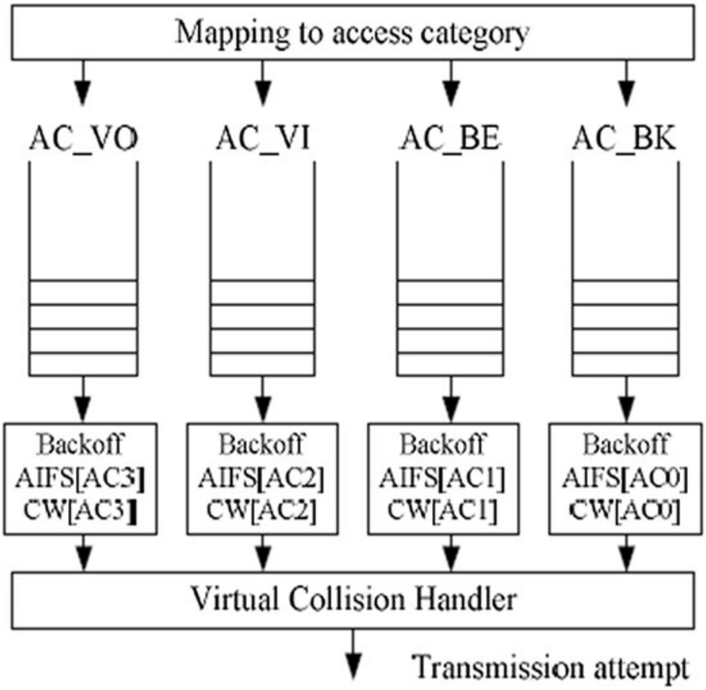


Fig. 2 Four access categories in IEEE 802.11e

Table 2 802.11e EDCA parameter set

Priority	AC	Designation	AIFSN	CWmin	CWmax
3	AC_VO	Voice	2	7	15
2	AC_VI	Video	2	15	31
1	AC_BE	Best effort	3	31	1023
0	AC_BK	Background	7	31	1023

period is called the end-to-end delay and can be divided into fixed and variable network delay.

– Fixed network delay—Includes encoding and decoding time (for voice and video), and the finite amount of time required for the electrical or optical pulses to traverse the media en route to their destination.
– Variable network delay—Generally refers to network conditions, such as queuing and congestion that can affect the overall time required for transit.

Jitter: (or delay-variance) is the difference in the end-to-end latency between packets.

Loss: (or packet loss) is a comparative measure of packets successfully transmitted and received to the total number that were transmitted. Loss is expressed as the percentage of packets that were dropped.

4 Results and Discussion

Our study of the impact of mobility on throughput showed that there is degradation in throughput in mobile nodes, we tried to remedy to this by adding a new access category that will be used exclusively by mobile nodes. To do so we modified the EDCA parameters in NS-3 to add a new access category which we called "AC_MO."

To test our new model, the parameters of the new AC will be kept the same as the parameters of access category Voice (AC_VO) except for Txoplimit.

4.1 Simulations

In this section, we use NS-3 simulator to evaluate the impact of mobility on throughput in IEEE 802.11, we choose 802.11n as the PHY layer.

The mobility of the nodes is according to the In the random walk mobility model, in which nodes move one movement unit with random speed in a direction chosen randomly between 0° and 360°, and then pick another direction [11].

Simulation Set 1

To compare the throughput of static and mobile nodes, we have considered two scenarios, namely scenario 1 and scenario 2. In each scenario, 10 stations (STAs) are connected to the same access point (AP).

Scenario 1 consists of all the STAs given fixed positions and connected to the access point with an access category of type Voice (AC_VO).

Scenario 2 consists of all STAs being mobile according to the random walk mobility model, and connected to the access point with an access category of type Voice (AC_VO).

We compare the throughput for each station by simulating both scenarios, the results represent the mean value of 10 simulations for the same scenario. The results are shown in Fig. 3.

Simulation Set 2

To compare the throughput of nodes connected with different Access Categories, we have a total of 20 nodes connected to the same access point (AP); all the nodes are mobile using the random walk mobility model, 10 nodes are connected to the AP using the access category of type voice (AC_VO), and the remaining nodes connected to the AP using the new access category of type mobile (AC_MO).

We compare the throughputs between the two access categories, and the results represent the mean value of 10 simulations and are shown in Fig. 4.

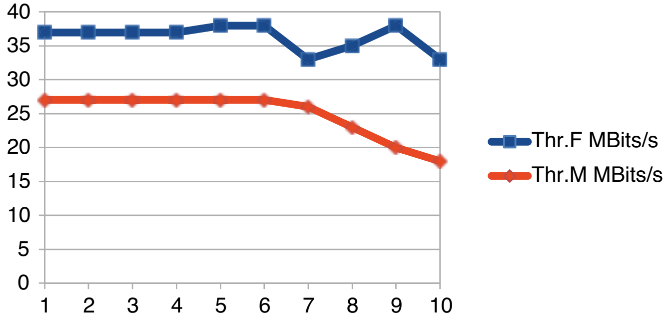


Fig. 3 Throughput of nodes connected with AC_VO and AC_MO in MBits/s

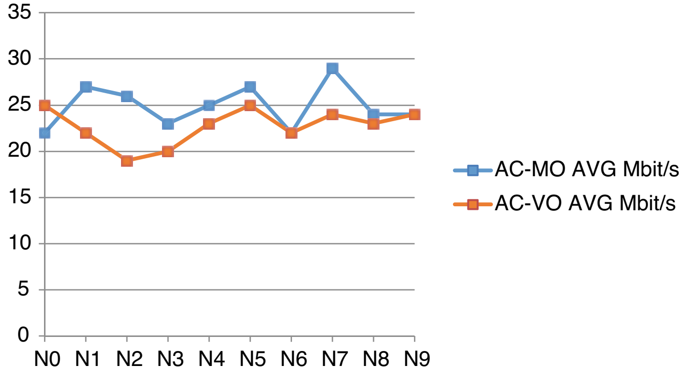


Fig. 4 Throughput of fixed and mobile nodes in Mbit/s

4.2 Discussion of Results

The results in simulation 1 show that while using the same access category (AC_VO), the difference in throughput between the static and mobile nodes is considerable (in the order of 10Mbits/s) which lead us to think that mobility of a node can have an impact on its throughput.

Based on the results of simulation 1, we decided to add a new access category that will be dedicated for mobile nodes.

To test the impact of the new AC on mobile nodes, we used the setup described in simulation set 2 in which 10 nodes are connected to the AP with AC_VO and the other 10 nodes connected to the AP with AC_MO. The results show that difference in throughput between the two sets of nodes was reduced, and the mobility does not impact the throughput as much as it did before we introduced the mobile access category (AC_VO).

5 Conclusion and Future Work

In this chapter, we have evaluated the impact of mobility on throughput, we observed that the mobile nodes had a lower throughput compared to the static nodes, and we tested our proposed model which consists of adding a new access category reserved only for mobile nodes regardless of their traffic category.

The tests of the new access category show the impact of mobility on throughput was reduced considerably.

The purpose of this study was purely explorative, to test our theory about the impact of mobility and if the proposed model could help with throughput fairness. Now that we have tested the model, the next step is to integrate it fully with the 802.11e standard.

References

1. Dridi, K., Javaid, N., Daachi, B., & Djouani, K. (2009). IEEE 802.11 e-EDCF evaluation through MAC-layer metrics over QoS-aware mobility constraints. In *Proceedings of the 7th international conference on advances in mobile computing and multimedia* (pp. 211–217). ACM.
2. Banchs, A., Azcorra, A., García, C., & Cuevas, R. (2005). Applications and challenges of the 802.11e EDCA mechanism: An experimental study. *IEEE Network, 19*(4), 52–58.
3. Zogkou, M., Sgora, A., Chatzimisios, P., & Vergados, D. D. (2014). EDCA mechanism and mobility support evaluation in IEEE 802.11 s WMNs. In *Ultra modern telecommunications and control systems and workshops (ICUMT), 2014 6th international congress on IEEE* (pp. 204–209).
4. Grossglauser, M., & Tse, D. (2001). *Mobility increases the capacity of ad hoc wireless networks* (pp. 1360–1369). IEEE INFOCOM.
5. Lenders, V., Wagner, J., Heimlicher, S., May, M., & Plattner, B. (2008). An empirical study of the impact of mobility on link failures in an 802.11 ad hoc network. *IEEE Wireless Communications, 15*(6), 16–21.
6. Dridi, K., et al. (2009). IEEE 802.11e-EDCF evaluation through MAC-layer metrics over QoS-aware mobility constraints. In *MoMM '09: Proceedings of the 7th international conference on advances in mobile computing and multimedia* (pp. 211–217).
7. IEEE Std 802.11. (2012). *Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*.
8. Cai, L. X., Ling, X., Shen, X., Mark, J. W., & Cai, L. (2009). Supporting voice and video applications over IEEE 802.11n WLANs. *ACM Wireless Networks (WINET), 15*(4), 443–454.
9. Arif, T. Y., & Sari, R. F.. (2010). *Performance comparison of video traffic over WLAN IEEE 802.11e and IEEE 802.11n*. UBICOMM.
10. Mohapatra, P., Li, J., & Gui, C. (2003). QoS in mobile ad hoc networks. *IEEE Wireless Communications, 10*(3), 44–53.
11. Zonoozi, M. M., & Dassanayake, P. (1997). User mobility modeling and characterization of mobility patterns. *IEEE Journal on Selected Areas in Communications, 15*(7), 1239–1252.

Connected Objects, an Asset to Improve Customer Experience

Badr Machkour and Ahmed Abriane

Abstract "We didn't do anything wrong, but somehow, we lost" that is how Nokia CEO Stephen Elop ended his speech saying during the press conference to announce the mobile phone branch of NOKIA to be sold to Microsoft. Nokia was the king of the mobile phone, but it missed learning and the change led to the collapse of their empire with the iPhone as well as the strategy of smartphones under Windows. Indeed, learning agility is not an option. Digital transformation and technological innovations today bring fabulous opportunities to meet the challenges and without digital adoption the tomorrow's company may not only evolve but disappear for the benefit of more agile and responsive competitors.

In 2010, we already produced, in 2 days, as many data as humanity in all its history. In addition, the best use of these data allows any company, whatever its sector of activity and whatever the context related to the use of the object, to improve all of its procedures. As a result, all this data, when aggregated and analysed, offer the opportunity to enhance the customer experience while providing them with tailor-made and fully individualized experiences. The management of the customer experience has become a major focus of strategic business consideration. It is an integral part of the company's offer and can therefore be the source of its difference with its competitors. In the same way, the connected object optimizes the customer experience, helping to make it more fluid, easy and responsive. It also offers an unprecedented experience by providing the opportunity for brands to materialize their services in different customer universes.

This publication aims to analyse the real impact of connected objects, or any other implemented innovation, on the realization of savings thanks to their capacities to optimize the processes as well as their influence the customer experience considered as an essential lever in the success of a business nowadays. Thus, we will propose a strategic model to make recommendations on the opportunity of the connected objects in the context of the management of the customer experience from

B. Machkour (✉) · A. Abriane
Faculty of Law, Economics and Social Sciences of Agadir, Laboratory of Studies and Research in Economics and Management, Agadir, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_39

three axes, namely the optimization of uses and more fluid experience, participation in the strategy, measuring the customer experience.

Keywords Customer experience · Satisfaction · Connected objects · Digitalization · Digital transformation · Innovation

1 The Optimization of Uses and More Fluid Experience

Use is naturally the main focus of influence on the customer experience. That said, it is important to understand the uses of new technologies and how they impact the client's different worlds.

Traditionally, a strategy that focuses on the customer experience requires understanding the customer, knowing what they want and measuring their expectations. "The company that successfully applies a value strategy has the gift of empathy. She puts herself in the client's shoes to identify his needs and measure their relative importance without forgetting any of them. It intimately understands the reasons for and conditions under which the customer uses his product or service. She masters the customer experience" [1]. Customer empathy and understanding are essential to a successful customer experience. In this logic, everything that is important for them must be placed at the heart of the company's projects.

To do this, it is best to ask the customer about his expectations. However, it is sometimes necessary for the company, based on the knowledge of its customers, to decide for them what is important and interesting. This is particularly the case when implementing innovative offers.

"If I had asked my customers what they wanted, they would have answered me: a faster hors". This famous sentence by Henry Ford clearly shows that this approach is not new but that it is particularly intense and is all the more relevant in a context where technologies are constantly emerging. Are customers able to understand what new technologies such as connected objects, artificial intelligence or virtual reality could bring them? It is likely that the answer is negative for the majority of them. In this case, the company must take risks, make a bet and offer its customers something new that can create a need, a positive experience.

Determining the right way to provide a positive customer experience or improve an existing one by using innovative technologies may seem complicated or confusing. Fortunately, there are tools that provide answers, leads and a methodological framework.

1.1 Customer Experience Dimensions

To try to model and understand what connected objects can offer the customer as an experience, it is possible to use a concept proposed by Pierre Daems[1] as well as by Batat and Frochot [2]. It is a model that focuses on several dimensions of the customer experience. Each of these dimensions represents a way of influencing the proposed experience. These dimensions can be used in a very advanced way depending on the sector of activity or type of experience studied. However, these dimensions can be used simply to directly and efficiently classify, study and verify easily the contributions to the experience that connected objects or other innovations bring. Indeed, a non-exhaustive list of dimensions accompanied by a description of the value that the connected objects and their data could bring:

- *Accessibility*: This dimension makes the company's offer more accessible. It can be a lever acting on temporal access or on physical and geographical access. This dimension can be materialized by an object allowing access to the offer without moving, or it can be an artificial intelligence interacting with the customer late at night.
- *Reliability*: This dimension concerns the impacts related to the client's reinsurance on what he will experience through a demonstration of the company's mastery. For example, they may be connected objects made available to provide a reliable waiting time based on real-time data.
- *Conviviality*: This dimension could rather be highlighted in the context of a human relationship, because it particularly plays on the customer's emotions, in all its place in all the levers to be activated by innovations related to connected objects, since it is possible to significantly improve already automated interactions by offering, for example, an adaptation based on video and audio analyses of emotions, but also by equipping intelligent interfaces with humorous or empathetic capacities.
- *Clarity*: This dimension influences the company's transparency towards its customer. It can be very important in certain sectors or contexts. It is possible to imagine objects to provide information to the customer on regulatory indicators.
- *Prevenance*: This dimension concerns the levers related to the anticipation of the client's needs. The analysis of the data generated by the connected industrial objects allows preventive maintenance to be carried out. Similarly, it is possible to predict customer needs by equipping objects to order their consumables according to their use, for example, or to offer a service based on the detection of a cycle, a habit.
- *Responsiveness*: This dimension acts on the time-related lever in the customer experience. It is about providing an immediate experience, in real or near real time. Many objects exploit this improvement by offering buttons connected to their customers in order to access services more quickly, if not immediately.

[1] P. Daems, Aube Conseil, Customer Experience Dimensions, National Bank of Canada, 2014.

- *Ease*: This is the dimension most exploited by connected objects. The characteristic of this dimension is to optimize processes and make the customer experience simpler, more intuitive, effortless. Connected objects offer the possibility to fluidify the experience in many fields and contexts such as home automation, health, energy, entertainment, by removing the frictions related to the complexity of using other interfaces. Artificial intelligence can also contribute significantly to the experience by supporting complex interactions.
- *Esteem*: This dimension corresponds to the contributions that connected objects, the analysis of their data and the intelligence that exploits them can make to the customer experience by personalizing it, making it truly unique and on measure. These objects can, for example, make it possible to physically greet a customer by name.
- *Credit*: This is a dimension that influences the client's ability to have an incomplete or uncertain experience. In some contexts, the advice and skills of companies are important to ensure a positive customer experience. This lever is traditionally linked to the values of the company's employees. When the human is not necessarily available or operating, artificial intelligence offers the ability to respond to this.

These dimensions are applicable at each of the key moments when the customer interacts with the company. These moments have a strong impact on the customer experience. Once these key moments of the customer experience have been mapped, it is therefore possible to ask questions to guide thinking and determine the opportunities of new technologies that could be offered to customers. For example:

> What dimension could be activated at this moment of truth to bring a positive experience or improve the customer experience?
>
> Does this new technology activate a dimension that can improve or make the customer experience positive at this key moment?
>
> How to use this technology to propose an improvement on such a dimension of the experience at this key moment?

Even without necessarily applying this analysis grid, some companies have long understood the value of these dimensions and their positive impact on the experience at all times during their customers' careers. We can use these dimensions to assess how the customer experience is changed.

In addition, other brands use connected objects to enhance the customer experience, such as Disney and its Magic Band or some insurance companies. These objects can influence one or more key moments in the customer's journey. The graph in Fig. 1 illustrates how these objects improve the experience through the prism of customer experience dimensions. This analysis, which uses an empirical classification, shows that the vast majority of life moments proposed by connected objects highlight ease or reactivity. This is therefore the primary use that brands make of these technologies in the customer experience: to meet an expectation of ease or simplicity, reactivity or immediacy.

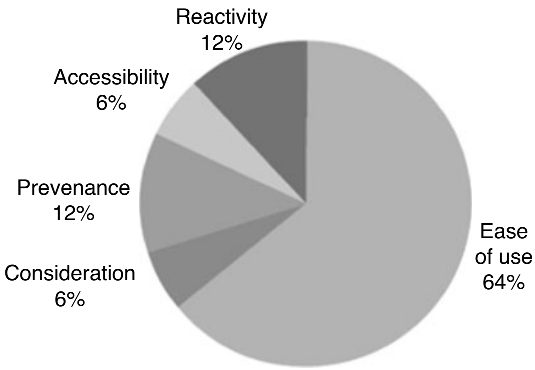


Fig. 1 The proportions of the activated dimensions

Connected objects, such as the Magic Band, designed to be fully integrated into the entire customer experience, make it possible to activate a panel of various sizes even if those of ease and reactivity remain largely represented.

Among these dimensions, thoughtfulness stands out because the object makes it easier to identify the customer and therefore to personalize his experience by anticipating his needs. It can be noted that being identified through an object used only in this context may be better perceived by customers than if the identification and localization were done by the smartphone, which is more personal.

The application of the other dimensions of the grid is for the moment less present in this evaluation of connected objects, but reliability could be highlighted by a connected object: for example, the taxi company “Uber” could have proposed the real display of waiting time via the object which, thus making the information more reliable, would reassure the customer. User-friendliness, clarity and competence seem to be further removed from the customer experience that can be provided by technology such as connected objects. However, they can, thanks to the data they generate, provide other technologies, such as artificial intelligence, with the ability to activate them. In addition, new objects and new ways of using their data are emerging from day to day that will certainly cover all dimensions and possibly others that are still missing from the practical sheet.

Connected objects therefore present many possibilities, and because it is a new technology, they are the subject of high expectations on the part of customers. However, we must be vigilant about the risks associated with this novelty and the expectations that could be imagined. Indeed, we can think that an object such as a light bulb, a button or a bracelet, even if it has communication capabilities to offer services, is ultimately only a simple gadget, as an article in La Tribune suggests: “Owning an activity tracker, a connected pair of jeans or a refrigerator that sends an alert the day before the yogurt expiry date is still perceived as part of the gadget”.[2] The connected object would in this case be “an object for which the promise of

[2] S. Rolland, “Pourquoi les objets connectés ne décollent pas (encore),” La Tribune, 11 Septembre 2015.

satisfaction is most often based on novelty, accessory, futility and/or strangeness: all characteristics likely to naturally attract the consumer's attention and trigger an impulse purchase". It would be a "purchase amplifier" but would have a "very short life cycle". "Anglo-Saxon professionals sometimes use the expression 'bells and whistles', which is literally untranslatable, but which refers to all the options and other gadgets added to a commodity to seduce a wider target group".[3]

In this definition of gadget, the notions of futility and impulse purchase but also of life cycle will shed light on the real opportunity of these objects to meet a need for companies and customers. Le Monde already mentioned this point of vigilance in 1966: "an object that does not aim at any aesthetic research, that claims no service, that is useless or whose function is so futile that one can guess that its creation was not dictated by a need". It is easy to find examples on the Internet of connected objects that you might find futile or even useless. Objects that do not really meet a need. For example, we can mention a connected oven that can be remotely activated when you are away but which therefore requires you to leave the food at room temperature for hours. Or a connected object that sends an alert when a smoke detector rings and is about five times more expensive than a real smoke detector.

In addition, the smartphone is a very complete connected object, it is even an Internet connection hub for many other connected objects. It is an object that the customer often wears. It has significant computational capabilities and an interface that allows for intuitive and complex interactions. It can therefore easily perform functions that connected objects could offer and relegate them, therefore, to the rank of gadgets because their usefulness would immediately be questioned; even if these connected objects did contribute to positive dimensions of the customer experience.

It is therefore important to ask the question of the substitution of the object for the mobile phone when selecting the project: can the smartphone provide the same experience? This is the question that Chronodrive asked itself for Izy, for example, to which their supplier Hiku answered: "The Hiku 1D linear scanner is a dedicated hardware component that scans more than 50 times per second, making instant scans very accurate. With a mobile application only, every time you want to scan an item you have to find your phone, unlock it, find your application, launch it, tap 'add', choose the scan function, wait for the camera to load, point it to the barcode and then wait for the results. With Hiku you press a button and you are done".[4] In this example, we realize that the object has a real interest in relation to the mobile phone.

2 Participation in the Strategy

The connected objects and the data they generate make it possible to differentiate the offer offered to the customer via the service or via customer experience

[3] Lehu, J. M. (2012), L'Encyclopédie du marketing, commentée et illustrée, Eyrolles.

[4] https://support.hiku.us/hc/fr.

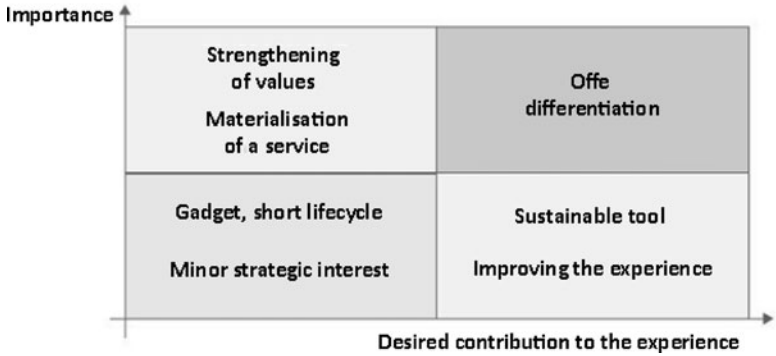


Fig. 2 The strategic opportunities of the connected object

management. Indeed, the extent of this differentiation will be more or less important depending on the place that the project holds in the company's strategy (Fig. 2).

The more important the implementation of connected objects and the intelligent processing of their data will be in the company's strategy, the more they will bring a real gain to the customer experience and help differentiate the company's offer from that of its competitors.

However, a connected object can be offered for strategic reasons for the company without significant contributions for the customer. The purpose will be to support a marketing or communication policy.

On the other hand, if the connected object has a low impact on the customer experience and does not contribute to the company's various strategic axes, then it will only be a gadget with no real interest for the customer and the company.

2.1 Marketing and Communication Levers

The connected object is perceived as an innovative component of the customer experience in a wide variety of sectors. The sectors of activity concerned may be transport, distribution, entertainment or insurance. It is interesting to note that while connected objects can meet local objectives dedicated to customer experience management, they also contribute to the company's overall objectives and vision.

For some, these technologies can be used to promote and make more visible their premium service and experience, which are the hallmarks of their brand promise. The connected object can thus become a representation, a physical embodiment of a service, a strategic opportunity to strengthen its image.

2.2 Image Enhancement

Brand image is an extremely important aspect in the customer's experience with the company, since it is one of the first links to which they will be attached. Often the first impression that the company will give will be made via the image it conveys. This is why brands compete with each other in imagination to make this impact as positive as possible.

To achieve this objective, brands naturally have a large number of marketing tools at their disposal. One of them is to communicate about integrating new technologies. This allows the brand to emphasize its innovative character in the era of time. This effect can be assimilated to the cultural dimension of consumption.

This principle, further developed by the CCT (Customer Culture Theory),[5] shows that consumers live their experience in a personal way. Indeed, ideological, socio-cultural and symbolic aspects influence his experience with companies or with the product he consumes.

Typical examples of the influence of such factors include eco-responsibility, manufacturing in foreign countries, social considerations, but also new technologies, science and innovation, which are cultural factors in their own right.

2.3 Communication Oriented Towards New Technologies

The experience expected by the client is also based on the memorization of past experiences and is now also increasingly influenced by the stories of other clients. With globalization and the relocation of production workshops, consumers are offered a wider range of products, and can often benefit from better prices. Let's take a minute to imagine the number of artisans, manufacturers, existing brands and products created every day around the world: a phenomenally huge number. Lost by the diversity of similar offers on offer, how can we differentiate between this and that product today? Future generations of customers are no longer interested in it: today, what matters is the customer experience and the company's brand image. Not satisfying a customer costs money when his satisfaction pays off. It is important to remember that setting up interactions with the customer to improve the perception of his buying experience does not generate high costs, but on the contrary these interactions are synonymous with ROI (return on investment). This shopping experience is part of the customer's memory, whether good or bad, and is not without consequences on his consideration and future behaviour towards the brand concerned. Not all experiences are memorable, but only those that go beyond initial expectations are memorable. However, an effective customer satisfaction strategy is

[5] W. Batat, I. Frochot, Experiential Marketing: how to design and stimulate the customer experience, Dunod, 2014.

not to seek this overcoming over the entire customer relationship but to target it to specific contact points in the customer journey.

The improvement of the customer journey can be defined around three main axes. First of all, it is relevant to develop a strong affinity with the brand through multi-channel interactions before, during and after the purchase act, on all the customer's experience factors (fluidity, reactivity, quality, etc.). It is also essential to identify and analyse all the contact points in order to choose them and integrate them as closely as possible into the company's strategy. The proposed route must be consistent and coherent with the company's image, values and products. Finally, it is essential to understand and identify the profitability of this approach by measuring the quality of customer satisfaction and its financial impact. This monitoring will also make it possible to identify potential improvements for defined customer segments and in a given competitive environment.

With the advent of Web 2.0, consumers can freely exchange their opinions: on blogs, through comments, forums but also on social networks. The brand, or rather the company, can no longer control its reputation: it evolves independently of its goodwill. Thus, the brand reputation is built by customers but also by non-customers: this is a horizontal communication. The multiplicity of digital exchanges between consumers, and the nature of their content, today create a real ecosystem of recommendations that is decisive for the company's reputation and sustainability.

This ecosystem is particularly important when the offer is experiential in nature, given the difficulty of pre-testing it. In addition, product offers also increasingly tend to transform the act of buying into a shopping experience. "With digital, the customer's opinion, his voice, are sacred". The difficulties of controlling these interactions are not without importance and consequence for companies. Potential customers will consider with interest the opinions of other users and will thus create an expectation or reluctance towards the company.

Notices left on the Web influence consumer satisfaction by changing the criteria and waiting threshold, both upwards and downwards. However, as stated above, satisfaction is "the consumer's response to the evaluation of the perceived gap between his or her previous expectations and the current performance of the product as perceived after consumption" [3]. This is why it is very important to seriously consider the opinions left after each user experience and to establish a strategy. It is in the company's best interest to feel responsible for these interactions and to manage them. Indeed, general dissatisfaction could lead to behaviour that is detrimental to the company, such as negative word of mouth or infidelity. On the contrary, general satisfaction would ensure better sales and a profitable ROI.

Digital interaction in response to good and bad opinions would be an essential part of the company's strategy. First of all, it is relevant to accept, listen and acknowledge the defects of the service or product. Through this transparency and responsiveness, the customer will feel respected and listened to. Thereafter, it is important to prevent this type of crisis. For example, H&M removed and apologized as soon as possible on social networks following the release in early January 2018 of the "Coolest Monkey" sports sweatshirt worn by a young black skinned model. Finally, it is necessary to ensure regular monitoring of this monitoring and to provide information on answers, thanks to compensation to customers in order to establish a so-called vertical interaction: a two-way interaction between the company and its customers.

Thus, artificial intelligence, virtual reality, and more recently the blockchain, but also connected objects, have become trophies to be displayed in a window, covered by the media in a multi-channel way. The customer therefore obtains information about the brand through social networks, word of mouth, reputation. The experience that the customer expects is built from these marketing and communication actions.

However, the experience expected by the client is also based on the memory of past experiences and is increasingly influenced by the comments of other clients. It is therefore necessary for the company to provide its customers with an experience that respects its vision and values. These may be more or less known to customers, but influence their perception.

When Amazon, a major player in online commerce, renowned for its excellent logistics and delivery management, makes an announcement about the project to launch an express delivery service using autonomous drones,[6] the company operates an excellent communication channel. Indeed, this innovative and technological object in the brand's colours would ideally be aligned with the values and image that the company conveys. On the other hand, for a company whose logistics and delivery are not the main strengths, which can even be sources of dissatisfaction, it would not be appropriate to follow the same strategy. The effect could be the opposite and the customer could reject the brand experience.

It is important for the company to integrate digital or digital aspects so that they are in harmony with its strategy, but also with its values because they are also the values of its customers: "For the customer experience to be memorable, engaging, robust and fascinating [4]", "it must contain strong attributes that are in line with the company's values" [5].

2.4 Differentiation Through Customer Experience

The concept of customer experience management has become widespread in companies over the past few years. An accelerating factor in this evolution is the increasing presence of technologies in the interactions between companies and their customers. Thus, the notion of customer experience is often mentioned when mentioning the Web, social networks, and more generally all the channels more or less related to new technologies, as well as the way to tune them and orchestrate them to elicit a reaction and create a link with the customer.

All sectors of activity take this concern into account and the task is complex. If there was a time when customers could only exchange with their salespeople

[6] https://amazon-presse.fr/Nos-communiqu-s/Nos-communiqu-s/Communiqu-/amazon/fr/2019-06-18/.

through a very limited number of channels, by going to the store or writing a letter, they can now engage in a conversation through a variety of channels: web, e-mail, SMS, social media, phone, chat, etc. Customers are mastering these new means of communication, which are multiplying and contributing to their experience with the brand.

The resulting technologies and communication methods have evolved very quickly and had to be integrated very quickly. The management of the experience was initially based on the physical point of sale and the telephone. Then, the Internet revolutionized the concept with the appearance of e-commerce sites. The development of the smartphone and mobile application concept in 2007 has created a new upheaval. The customer was able to access the Internet on the move, in a more ergonomic and practical way. This has changed the uses, the possibilities of contact points, and has democratized other channels such as social networks and features such as geolocation or payment via the Internet.

Connected objects are a source of questioning. Indeed, they offer the possibility of more intuitive interactions with objects and the client's direct environment, they strengthen the link thanks to the particular connection they offer. This is a new perspective of more spontaneous and positive exchanges between the brand and the customer. Thus, the more the connected objects help to differentiate the company's offer from that of its competitors in terms of experience, the more they will represent a strategic opportunity for the company. The axis of differentiation can be so strong that it can imply a totally new offer for the customer and give the company a significant competitive advantage.

For brands, retailers and producers, it is interesting to note that the role of connected objects in the customer experience and their place in the company's strategy could be considered in a much more significant way. Indeed, mastering the object and its data also implies controlling the potential for creating services and the associated customer experience. This is the case for resellers, for example, who know their customers but also the products they sell and can thus combine this knowledge to offer new services and a new experience to their customers. Above all, this would make it possible to anticipate an intermediation of its services by household equipment manufacturers, since the connection of objects somehow brings the customer closer to the reseller but also to the manufacturer.

Therefore, mastering the information generated by the algorithms that analyse the data of connected objects is currently of critical strategic importance, particularly because it will provide the best customer experience and thus have a competitive advantage that allows for privileged positioning. In this context, it is therefore an important issue for distributors of connected products, but also for manufacturers, in order to position themselves in the value chain. This is true for all sectors: energy, insurance, etc.

This is pointed out by Porter and Heppelmann [6] in specifying the opportunities that connected objects present for manufacturers: "Connected and intelligent products enable companies to maintain deep and solid relationships that can reduce the need for partners' distribution channels. Companies can also remotely diagnose product performance or failure problems, sometimes maintaining, reducing dependence on partner services. By reducing the role of intermediaries, companies can potentially capture more revenue and boost their margins".[7] For insurers, for example, achieving one of their major objectives, which is prevention, requires a strategy based on the experience provided by connected objects. This strategy will have to be developed over the long term, since it is based on a necessary change in the uses and habits of their policyholders. It is therefore easy to understand why they need the most complete and sustainable vision possible of the customer, and that they must rely on objects to increase their access to information.

This desire to differentiate competitors through the customer experience can be a major component of the strategy of some companies, which use new technologies to achieve these objectives. This methodology can be applied in a quasi-obsessive way by some successful companies. They advocate innovation at the service of the customer.

This is, among other things, the case of Chronodrive, which advocates a breakthrough in the way it does its shopping to offer its customers an optimal experience by following a strategy of differentiation through customer experience that allows it to create an offer that Chronodrive was alone to offer by launching the concept of a "drive" based on technological innovations related to mobile applications. It is in this spirit that the brand was also the first to offer a new way of shopping with a connected object. The Chronodrive brand was once again the only one to offer this experience, until its competitors adapted, since the concept was then adopted by other brands.

This connected object is therefore part of a strategic approach based on a break in the customer experience. The direct objective is to continue to differentiate the brand's offer for household purchases. The disruptive capacity of this offer could be significant. Customer use will determine this in the future. This new experience is in any case remarkable enough for direct competitors in the sector to also launch their objects about a year later.

3 Customer Experience Measurement

The elements put into perspective above show the strategic importance of customer experience management, and that the way customers live their experience can influence their perception of the brand. The company must therefore ensure that it masters the experience it wishes to provide to its customers so that their perception is as positive and memorable as possible.

[7] Porter M., Heppelmann J., "How Smart, Connected Products Are Transforming Competition", Harvard Business Revue, November 2014.

The strategy must take this into account. Good customer experience management is pragmatic: its objective is to ensure that two perspectives are aligned as closely as possible, as Elisabeth Lefranc demonstrates in her thesis[8]:

– The positive experience it wishes to offer.
– The customer's perception from his own experience.

This notion is all the more important today because the customer experience is highly visible for companies. However, in recent years, studies have shown that the company's and customers' perspectives are not at all aligned: where the majority of companies think they know their customers and provide them with a quality experience that differentiates them, very few customers think the same thing in return. A study carried out by IBM and EConsultancy in 2015 even provides surprising data: a large panel of companies surveyed believe they have a very good knowledge of their customers since 81% of them ensure that they have a holistic or near holistic vision. However, when customers are asked about the ability of these same brands to understand them, the majority of them answer "no".

This implies that there is therefore a gap between what companies want to provide to their customers and what customers express in terms of their experience.

3.1 From Service Quality to a Successful Customer Experience

There are tools on which to base an understanding of the mechanisms that govern the company's proposal and the client's experience. The work of Parasuram et al. [7] introduced different components into the notion of service: the service that is specified and delivered by companies is also expected and perceived by customers.

The company's objective is to understand how the service provided by the company is different from the service expected by the customer. As part of managing the customer experience offered by the company, in order to ensure that it is well lived, it is also necessary to bring the perspectives of the customer and the company closer together, no longer through the provision of service, but through the customer experience.

This means trying to find out what the customer's real experience was. The results of this lived experience must be recovered from the field. The analysis of this data will improve the company's communication and the implementation of the customer experience.[9]

It is important to know and understand the client's experience. This measure is indeed decisive in successful customer experience management, as it ensures that the experience desired by the company and the customer's experience are aligned.

---

[8] Le Management de l'expérience client : au-delà des enquêtes satisfaction, la mesure de l'expérience vécue, Elisabeth Lefranc, 2013.

[9] Le Management de l'expérience client : au-delà des enquêtes satisfaction, la mesure de l'expérience vécue, Elisabeth Lefranc, 2013.

The key to a positive experience therefore seems rather obvious: it is enough to evaluate it. However, this data is not easily retrievable.

In order to collect this valuable data, the company must focus on identifying their sources. To do this, it has two main areas of recovery: on the one hand, information corresponding to the lived experience that can be directly derived from an explicit feedback from its customers, and on the other hand, information that can come from data generated directly or indirectly by its customers.

To retrieve direct information about the customer experience, companies can use quantitative sources. These sources are suitable for measuring service quality, including customer satisfaction. However, satisfaction surveys may not fully capture the customer experience. Other measures make more sense to obtain a general statistical view of customer experience, such as the Net Promoter Score (NPS) or the Customer Effort Score (CES). The first shows the customer's attachment to the brand. The second is to determine the difficulties that customers have experienced during their exchanges with the brand. These two indicators are generally very closely related, as it is conceivable that clients who have made very little effort and who have had a smooth experience may offer a high NPS. However, these measures, which propose benchmarks and trends based on expectations, do not provide information on the reasons for these trends and what particularly influences the customer experience: emotions.

Feeling and emotion are important concepts to consider. They help to determine how memorable and positive the experience was for the client. This feeling is generated when the client lives his experience according to what he expected from that moment. If this axis of analysis is ignored, the company could find itself faced with confusing situations: the customer can obtain a service or product that meets his needs, which are optimized to make it easy to use, but yet he will have a bad feeling about his exchange with the company or product. Why? Why? Because of a difference between what he imagined about the experience he was going to have and what he really experienced.

Direct sources of information are interesting to understand how the client experienced it. Qualitative studies offer the possibility of clearly probing it and extracting information that is not visible in quantitative information. They are often unstructured and can be based on interviews or behavioural or even emotional analyses.

Unstructured data analysis technologies provide particularly interesting information for processing these sources. Especially since, from now on, these sources are no longer necessarily solicited by the company but are offered by the customer on the Internet (via forums, social networks) but also via connected objects.

Very efficient tools are already available thanks to advances in the processes of analysing large amounts of data. They make it possible to analyse the experience of customers through information not requested by the company. This information, generated directly or indirectly by customers, is a gold mine for companies that want to know their customers' experiences. This data can be retrieved in large quantities on social networks or via customer feedback directly to the company. Data can also be extracted indirectly generated by customer behaviour towards products or services, such as website visits, for example.

3.2 Sentiment and Context Analysis

Take the example of "Watson for Social Analytics", which provides the opportunity to determine the experience of clients. In 2014, Bob Iger, Disney's Chief Executive Officer, publicly reported results that show that 90% of users of the connected "Magic Band" bracelet rate the experience as excellent or very good.

This information remains quantitative, indicating a trend, but does not provide accurate information about the client's experience. It would obviously be tedious and costly to conduct a qualitative study and interview thousands of clients, and then conduct a count to determine their feelings about the experience.

To complete this analysis, the "Watson for Social Analytics" tool provides quantitative information on sentiment. The tool is based on unstructured data collected from forums and social networks. It is a source of information where customers' comments do not have a filter. The results are therefore not biased by the survey methodology. In this example, the analysis, based on nearly 30,000 sources mentioning the "Magic Band", shows that nearly half of the mentions on the Disney bracelet are positive while about 5% are negative (Fig. 3).

These powerful analytical tools provide valuable measurements to identify areas for potential improvements in the customer experience.

As Elisabeth Lefranc [5] points out, "an ultimate quest is to integrate measurement as a component of the desired experience". In other words, for the measure to be effective, this aspect should be treated as an essential part of the customer experience management strategy. The measurement should be integrated from the design of the experiment, in order to define as soon as possible how to determine it as precisely as possible.

While connected objects can be used to improve or offer new experiences, they also allow information to be retrieved in order to build a vision of the customer's

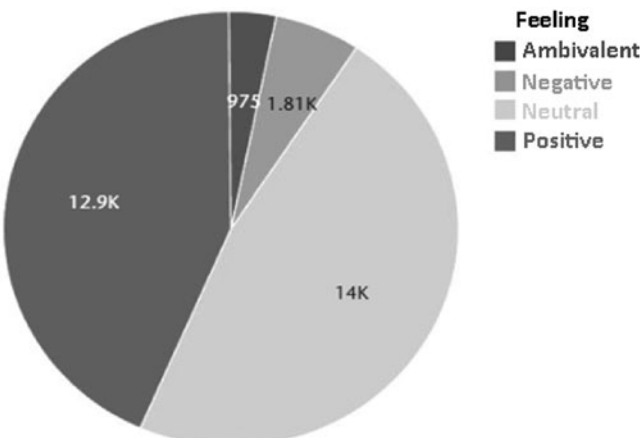


Fig. 3 Magic Band sentiment analysis. (Source: "Magic Band benchmark", [8])

experience. As we have seen before, it is tedious to measure the client's experience in a representative and reliable way. Objects offer the possibility of measuring with customers. However, this vision will remain partial if it is not associated with a return linked to feeling, or even emotion. We have previously seen, through solutions such as "IBM Watson for Social Analytics", that it is possible to retrieve information oriented on customers' feelings about a particular subject based on a semantic analysis of social networks. It would therefore be possible to apply analysis algorithms with the same objective to the data from the connected objects.

For example, there are currently telephone conversation analysis tools, such as the one offered by Cogito, that are able to determine emotions and feelings. These are tools that have been widely used in call centre management in recent years. This technology provides indicators on customer sentiment and brand relationship experience, and therefore on their feeling of the company. This even makes it possible, when combined with intelligent real-time processing, to help the operator by coaching him and giving him advice to guide the conversation if he starts talking too fast or if he is confronted with a frustrated or even angry person.

Voice is not the only analytical vector that connected objects could use to measure the customer experience: the image could also be used to determine emotions, feelings and therefore the customer experience. This is a possibility offered by, for example, connected cameras combined with new technologies related to image analysis.

3.3 Complex Analysis of Human States and Measurement of Emotions

Local shops are on the verge of a revolution, led by a company that won a Gartner award in the "cool vendors" section of the 2017 "Consumer Goods Manufacturing" report: Cloverleaf. This company is based on artificial intelligence based on the analysis of emotions from video streams from Affectiva, which makes it possible to categorize customers' feelings according to different categories such as joy, disappointment, disgust, dissatisfaction or surprise.[10]

This ability to analyse emotions has been integrated into shelves used in local shops. Thus, the shelf is able to detect the customer's emotions and determine how the customer reacts to a promotion or product launch, or to a new packaging. This is essential information for brands, which can therefore carry out very rapid A/B testing. The company ensures that this technology performs the scans anonymously, it collects general information such as customer reaction, age, gender. The concept goes even further, since the cameras embedded in the shelf are also able to detect if the customer is paying attention to the product. In this case, LCD panels, positioned

[10] https://venturebeat.com/2017/01/12/this-device-from-cloverleaf-and-affectiva-tracks-your-emotional-reaction-to-products-while-shopping.

at the locations traditionally used for price labels, display messages dynamically and adapt to customer reactions.

This solution offers local businesses a level of behaviour analysis comparable to that available on e-commerce platforms. These panels have already been tested in Procter & Gamble's innovation centre with a store equipped with shelves and a demonstration store that did not have shelves. The average increase in sales observed in the equipped store is 37%.[11]

The ability to detect and measure emotions in real time no longer requires human presence, but can be achieved at any time when the customer is experimenting with the brand. Analysis software is now able to detect as many characteristics on a face as a human could: gender, age, head and iris positioning, emotions, etc. This is also what Emotient, acquired by Apple in 2016, offers.

Connected objects could therefore offer, via behavioural and physiological measurements, information that could automate the measurement of the customer's experience and emotions. This would allow the different perspectives of the customer experience to be adjusted more quickly and systematically, in particular what the company wants and what the customer is experiencing, thanks in particular to artificial intelligence, which will have the ability to act in real time with the customer according to the emotions detected.

References

1. De Bodinat, H. (2007). *Les mystères de l'offre – Enquête sur une stratégie orpheline* (p. 56). Pearson Education France.
2. Batat, W., & Frochot, I. (2014). *Marketing expérientiel: Comment concevoir et stimuler l'expérience client*. Tendances Marketing, Dunod.
3. Tse, D. K., & Wilton, P. C. (1988, May). Models of consumer satisfaction formation: An extension. *Journal of Marketing Research, 25*, 204–212.
4. Gilmore, J. H., & Pine, B. I. I. (2002). Customer experience places: The new offering frontier. *Strategy & Leadership, 30*, 4–11.
5. LeFranc, E. (2013). *Le Management de l'expérience client: au-delà des enquêtes satisfaction, la mesure de l'expérience vécue*. CNAM.
6. Porter, M. E., & Heppelmann, J. E. (2014). How smart connected products are transforming competition. *Harvard Business Review, 92*, 64–88.
7. Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64*(Spring), 12–40.
8. IBM Watson Analytics. (2016). *Magic Band sentiment benchmark*.

[11] http://www.shelfpoint.com/who.html.

A Comparative Study of Existing Fuzzy Query Systems of Database

Mama Rachid and Machkour Mustapha

Abstract Relational Database Management Systems (RDBMS) have become, without a doubt, the core of any computer system. Besides, in many cases information is found to be naturally fuzzy or imprecise, that's why fuzzy query systems have become indispensable to represent and manage this information and especially facilitate interrogation to a non-expert user, but the problem is that Boolean queries do not allow the user to use vague and imprecise language terms in the qualification criteria of the searched data or to express preferences between these criteria, which is often a request legitimate end user. Nowadays, there are many proposals that allow users to make fuzzy queries on relational databases. In this chapter, we will briefly review the main attempts to find a perfect solution to this problem, highlighting their advantages, disadvantages and difficulties encountered.

Keywords Database management systems · Fuzzy query · Flexible queries · Fuzzy logic

1 Introduction

Data plays an important role in our daily lives, such as bank accounts, games, social networks, videos, etc. However Relational Database Management Systems (RDBMS) have become the core of any computer system. In many cases, the information is naturally fuzzy or inaccurate because the knowledge that humans have about the world is almost never perfect. Thus, the knowledge on which human reasoning is based, are almost always tainted with uncertainties and inaccuracies. In fact, these imperfections emanate from the very nature of man and the world. Traditional interrogation systems are unable to deal with uncertainty and vagueness.

M. Rachid (✉) · M. Mustapha
Information Systems and Vision Laboratory, Faculty of Sciences, Ibn Zohr University, Agadir, Morocco
e-mail: mama.rachid@edu.uiz.ac.ma; m.machkour@uiz.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_40

In addition, a relational database management system supports a structured query language (SQL) for data processing.

This language is based on standards based on Boolean interpretations that prevent database experts from processing fuzzy information. To illustrate this problem, consider a user who consults by the Internet a database of car rental offers. The user wants to rent a new car, cheaper and with reduced fuel consumption, this request can be expressed as:

```
SELECT * FROM tbl_cars
WHERE prod_year=" new" AND Price="cheap"
AND fuel_consumption="small"
```

with tbl_carsis a table that contains numeric data. The problem is that traditional interrogation systems are unable to handle these kinds of fuzzy terms such as "new," "cheap," and "small."

Several works have been proposed in the literature to introduce flexibility in database querying. Most of these works have used the fuzzy sets and fuzzy logic formalism to model linguistic terms such as ("new," "small") and to evaluate predicates with such terms. The main idea of this work is to extend the SQL language and add an additional layer of a classical DBMS to evaluate fuzzy predicates [1–3].

In this chapter, we present a comparative study of the most relevant fuzzy Query systems of database, along with the advantages and the drawbacks of each one, and finally a conclusion.

2 Background

The problem of the representation and processing of "imprecise" information has been widely studied by several authors [2, 4]. However, all the models published to give the solution to this problem have their advantages, disadvantages and their limitations. The problem is not trivial, it is necessary to modify the structure of the relations and, with these, the operations defined on them. To help store imprecise information and to consult it in a flexible way, this information requires the study of a multitude of cases that do not occur in the classical model.

The first models are mainly theoretical models of fuzzy relational databases, among them the model of Bukles-Petry [5] proposed by Buckles and Petry in 1980, This is the first model that uses the similarity relations in the relational model, he defined a fuzzy representation for relational databases, in which non-fuzzy databases are a special case of this model. The structure for representing imprecise information that has been defined in this model differs from ordinary relational databases in two important ways: n-tuplet components do not need to be unique values and a similarity relation is required for each set of domains in the database. A fuzzy relation R is defined as a subset of the Cartesian product $2^{D_1} \times 2^{D_2} \times \cdots \times 2^{D_n}$ where 2^{D_i} is any not null member of the domain base set D_i.the domains are either discrete scalars, or discrete numbers from a finite or infinite set.

The values of a particular tuple can be simple scalars or numbers (including nulls) or a finite set of scalars or numbers. .for example (Table person):

NAME	APTITUDE	AGE
{SMITH}	{AVERAGE,GOOD}	{21,22,23}

The resemblance relation that exists on each of the domains serves to represent and direct the imprecision. It establishes a measure of similarity s (x, y) between the different values of the domain on which it is defined. It is defined by the user and the resemblance values are between 0 and 1. (0: Completely different; 1: Completely similar). In a query, the user asks about tuples that satisfy a given condition for a given similarity threshold. For example:

PROJECT(***PERSONNE*** *:APTITUDE ,AGE) WITH THRES(APTITUDE*≥0.5*) ,*
THRES(AGE)≥0.75

The tuples of the relationship are grouped into equivalence class, according to the relations and the thresholds of similarities defined on them. The main disadvantages of this model are:

- it does not model well all the fuzzy aspects of information (for example fuzzy modifiers, fuzzy quantifiers ...).
- atomicity is not guaranteed in the representation of information.
- the integrity of the database is not guaranteed.
- the result having several interpretations.

However, we can mention some advantages such as the use of resemblance relationships is an appropriate and intuitive tool for representing imprecision, and the use of different thresholds for each of the attributes.

Umano et.al. [6] proposed in 1980 a model based on the theory of possibilities. It is one of the first models of fuzzy relational databases. He uses as value:

- Possibility distributions
- Undefined: $\pi_{A(x)}(d) = 0, \forall\ d \in D$, Unknown: $\pi_{A(x)}(d) = 1, \forall\ d \in D$ And Null = {1/Unknown,1/Undefined}with D is the discourse universe of $A(x)$ and $\pi_{A(x)}(d)$ is the measure of possibility.

The consultation in this module returns three subsets: the tuples that clearly satisfying the consultation, tuples that approximately satisfy the consultation and tuples that do not clearly satisfy the consultation. Among the advantages of this model is that it can assign a degree of belonging to each tuple of the relationship, as well as it can store possibilities distributions. But there are the limits like does not handle non-scalar data, do not support the similarity relationship, and also does not model well all the fuzzy aspects of the information (for example Fuzzy modifiers, Fuzzy quantifiers ...).

In 1984, Henri Prade and Claudette Testemale [7] proposed a model based on the distribution of the possibility introduced by Zadeh [8, 9] to represent and process partial, uncertain or fuzzy data and to take into account vague queries. It is an approach that generalizes the representations of Buckles-Perty and Umano-Fukami and describes an extended relational algebra. The value of the attributes and the vague predicates are represented by means of possibility distributions evaluated by [0,1]. The data structure is similar to that used in the Umano-Fukami model. It uses measures of possibility and necessity to satisfy the conditions established in the consultation. The allowed domains in this model are:

- Finished set of scalar. Example $D = \{\text{red, blond, brown}\}$
- Finished set of number. Example $D = \{21,22,23\}$
- Set of fuzzy numbers or fuzzy labels. Example $D = \{\text{small, medium, big}\}$

The possible values for these domains are:

- Precise values. Example 25
- Interval values. Example [30,34]
- Fuzzy values. Example "good," "about-10," "bad-to-very-bad"
- Null values "Unknown" and "does-not-apply"
- A distribution of possibility. Example {1/M, 0.6/D}

For example, the relation person, can correspond to a table such as:

Name	Age	Family-situation[a]
David	25	Unknown
Tom	[30,34]	U
Paul	Young	{1/W,1/D}
Jean	About-50	{1/M,0.6/D}

[a] *M* married, *U* unmarried, *D* divorced, *W* widow(er)

{1/M, 0.6/D} means that there is a possibility equal to 1 that the person is married, and possibility equal to 0.6 that he is divorced, and a zero possibility for the others.

Although this module has defined an acceptable generalization for the representation of uncertain and incomplete information. There are still some disadvantages like:

- It does not model well all the fuzzy aspects of the information (for example Fuzzy modifiers, Fuzzy quantifiers, Fuzzy group by ...)
- Do not support the similarity relationship
- Does not support multivalued attributes such as spoken-language (David) = English and Arabic
- Do not model values that are related to each other

In 1985, Maria ZAMANKOVA and Abraham KENDEL [10] proposed another fuzzy relational database model, this model is based on research in relational data and theories of fuzzy sets and the possibility. It allows to recover the information desirable by the application the rules of linguistics fuzzy terms of the query. Among the advantages of this module is that it takes into account individualization. A user can define specific functions or rules that can be added to the system vocabulary. For example, a definition of a fuzzy set AGE may differ from one user to another even though the Age data is the same in the database. This model consists of three parts:

- A database of values (VDB) that store the actual data values.
- An explanatory database (EDB) that stores definitions for fuzzy subsets and fuzzy relationships is one part that reflects a user's knowledge profile.
- A set of translation rules that are used to manipulate adjectives.

The allowed domains in this model are:

- Set of discrete scalars. Example color = {red, blond, brown}
- Set of discrete or continuous numbers
- The unit interval [0, 1]

And the possible values for these domains are:

- Simple scalars or number
- A possibility distribution
- A real number in the interval [0,1] which is the value of the membership function or distribution of possibility
- Null value

For example, the person relationship may represent as:

Name	Age	Hair color	Smart
David	25	0.8/black + 0.3/brown	0.5
Tom	30	0.6/red + 0.7/blond	0.4
Paul	82	1/black	0.9

Among the disadvantages of this module is that do not support fuzzy quantifiers, fuzzy grouping, and fuzzy join. Also, the dependency with the relational model, which is not treated by this model. In addition, it does not allow the user to specify the accuracy with which the conditions involved in a query are met.

The most generalized model is the Generalized model for fuzzy relational database (GEFRED) which was proposed in 1994 by Medina, Pons and Villa [11]. it constitutes an eclectic synthesis of the various published models to treat the problem of the representation and the treatment of the fuzzy information by means of the relational databases. it is based on the Generalized fuzzy Domain (D) and on the Generalized fuzzy Relationship (R), one of the main advantages of this model is that it consists of a general abstraction that makes it possible to treat different approaches, even those that may seem very disparate. The possible data in the GEFRED model can be consulted in [11].

Based on the theoretical GEFRED model and the resources of the classical relational model, Medina and .al have developed a module called Fuzzy Interface for RelationalSystems (FIRST) to extend the capabilities of a classic DBMS so that it can represent and manipulate imprecise information. It is based on the client-

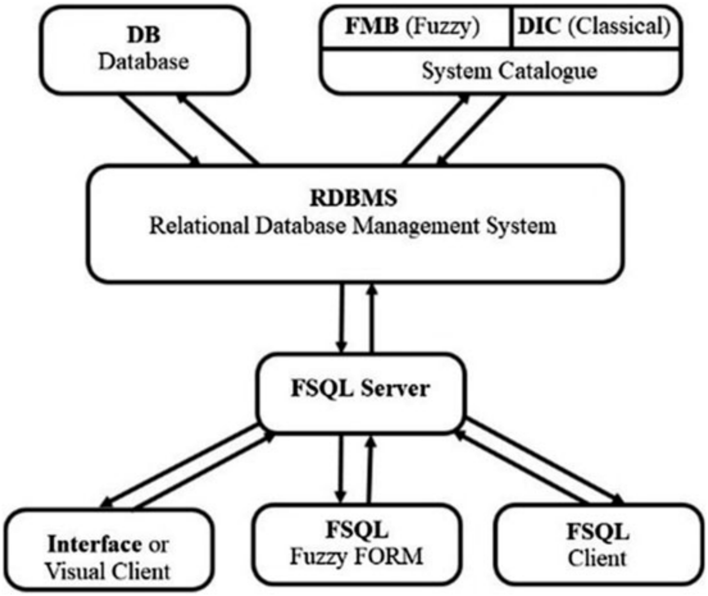


Fig. 1 FIRST architecture

server RDBMS architecture provided by Oracle. It adds new components (Fuzzy Meta Knowledge Base "FMB," FSQL Server, etc.) to the existing structure to handle imprecise information. Figure 1 shows the general architecture of this model.

This model uses a specific query language called Fuzzy SQL (FSQL),it is an extension of SQL to allow flexible queries. It already extends the existing commands in SQL, but it also incorporates novelties like fuzzy attributes, fuzzy constants, fuzzy comparators, fuzzy quantifiers ...

For example, if we consider a table person and we want to find young person (with a threshold of 0.4) who live in New York and have a salary greater than or equal to the trapezoidal distribution [100,300,500,800] .the FSQL query is written:

```
SELECT name ,CDEG(age),salary FROM person WHERE age FEQ $young 0.4
AND salary FGEQ  $[100,300,500,800] AND address = 'new york'
```

The FMB component deals with the storage of attributes that allow fuzzy processing and the information of each of them according to their type in a relational format .while the FSQL server's role is to extract the queries written with the FSQL language and translate them into SQL language using the information contained in the FMB. (see Fig. 2).

Although this model has several advantages in the representation and processing of information fuzzy, there are the weaknesses, such as:

- The problem concerns the choice of the type of the attribute (FTYPE1, FTYPE2, or FTYPE3), because an attribute can be in FTYPE1 cases and in other cases FTYPE2.

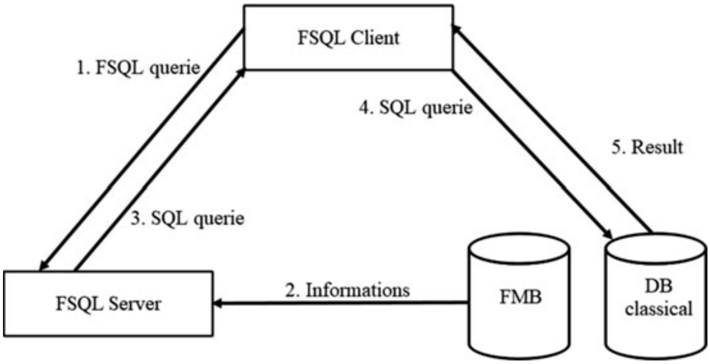


Fig. 2 FSQL Server operations

- The GIFRED model theoretically defines some features that have not yet been implemented in this module(Example Fuzzy group by).
- The approach uses a parser/translator to check and convert an FSQL query to SQL respecting the definitions of any fuzzy terms or operators stored in another database (FBM). This will slow down the query process.
- This module requires a good description of the different operations to be done at the database level and at the FBM level. This operation becomes heavier and trickier if the database becomes very large.
- The SQL language remains unusable by a non-expert user.
- This model does not allow the user to describe the fuzzy database (FDB) schema or manipulate its FDB.

Several approaches have been proposed to improve the FIRST model like that of José Galindo [12] which introduced the second version (FIRST 2) which contains new comparators, new fuzzy attributes, new fuzzy constants, and new feature in executionthresholds ... etc. and Martinez [13] who introduced an approach to extend non-scalar attribute management using ontology, he presents a new system (see Fig. 3) that combines fuzzy logic and ontology for get an answer as complete as possible.

Another fuzzy query language was proposed by Patrick Bosc and Olivier Pivert [14] in 1995 called SQLF for the purpose of remedying the problems posed by the SQL language in flexible queries. The structure of the SQL base block is kept in SQLF:

```
SELECT   [distinct ]  [n| t | n, t]  <attributes>
FROM <relations>  WHERE  <fuzzy condition>
```

The “FROM” clause does not undergo any change, the changes concern two points: the calibration of the result and the nature of the authorized conditions which may contain Boolean or gradual conditions, or both connected by connectors. In SQLF, a multi-relation block combines projection, restrictions, and algebraic or fuzzy joins:

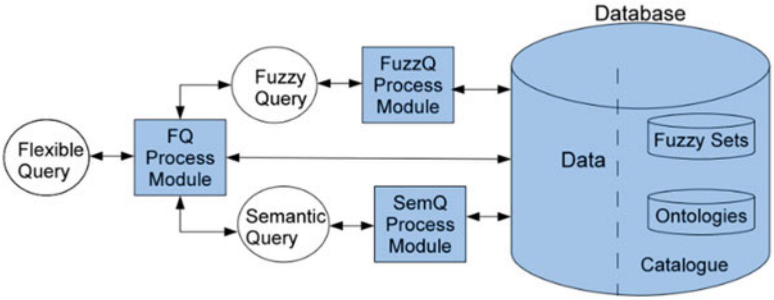


Fig. 3 System architecture proposed by Martinez

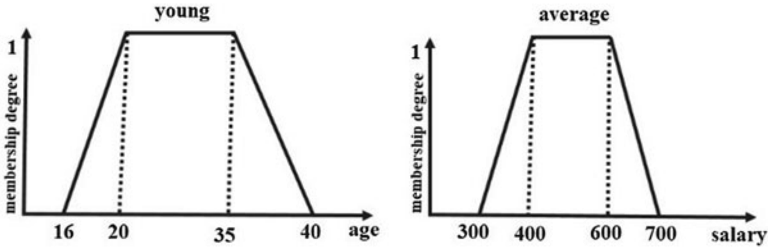


Fig. 4 Definition of fuzzy sets 'young' and 'average'

SELECT distinct R.A, S.B.
FROM R, S WHERE f_{c_R}*and* f_{c_S}*and (R.C* θ *S.D).*

where (R.C θ S.D)is the fuzzy join condition for example (R.C "roughlyequal to" S.D) and f_{c_R} (resp. f_{c_S}) it's a selection expression on R (resp S). The relational and set operations used in Boolean queries have been extended to take into account fuzzy predicates and return fuzzy relationships. An example of a fuzzy query addressed to the person table using the predicates shown in Fig. 4 would be:

```
SELECT name ,age,salary
FROM person  WHERE age = 'young' AND salary = ' average'
```

All valid queries in SQL are always valid in SQLF, it is an important point for optimizing the evaluation of the query. Among the disadvantages of SQLF is that do not manage non-scalar data, and do not support the similarity relationship.

SQLF has undergone several enhancements like the one made by Kacprzyk and Zadrozny [15], as part of their FQUERY package for Access, to increase the efficiency of the fuzzy query engine.

3 Results

We presented a brief study on proposed models. The tables (extract from [13]) below show a summary of our study (Tables 1 and 2).

Table 1 Comparison of most relevant characteristic in fuzzy query systems (part 1)

Model	Medina [11]	Bosc [14]	Zemankova [10]	Prade [7]	Martinez [13]	Umano [6]	Kacprzky [15]	Buckles [5]
Manage scalar data	.	.	.	.	.	.	.	.
Manage non-scalar data	.		.	.	.			.
Similarity relationship	.		.		.			.
Possibility distributions	.	.	.	.	.		.	
Degree in attributes level	.			.	.	.		.

Table 2 Comparison of most relevant characteristic in fuzzy query systems (part 2)

Model	Medina [11]	Bosc [14]	Zemankova [10]	Prade [7]	Martinez [13]	Umano [6]	Kacprzky [15]	Buckles [5]
Degree in tuple level	.	.	.	.	.	.	.	
Fuzzy modifiers		.	.					
Fuzzy quantifiers	.	.					.	
Fuzzy comparison operators	.	.	.	.	.	.	.	.
Fuzzy group by		.					.	
Fuzzy joins	.	.		.		.		
Nesting	.	.						
Store fuzzy data	.		.	.		.		.
Fuzzy queries	.	.	.	.	.	.	.	.
Extension SQL language	.	.			.	.	.	

We conclude that none of the proposals is complete, most of them give a partial version of the representation and processing of imprecise information and the implemented proposals also depend on the platform.

4 The Intended Model

The implementation of any system requires a detailed study. This study must consider the needs of users. The desired system must be flexible, able to provide the appropriate mechanisms for the representation, processing and retrieval of fuzzy information in all its forms, in addition it must be considerably collaborate with the commercial DBMS in an efficient way to obtain better performances. And, must be regardless of the platform.

Our system must be complete, having all the features and operators to recover and process fuzzy information. The previous systems are incomplete; the model that offers more functionality is the GEFRED model that has been proposed by Medina et al. [11].

Our system must be able to consider the fact that the user can be non-expert for example do not know the schema of the database. So, the system must have a friendly graphical interface to facilitate the tasks. For example, help the user to define his own linguistic terms. And allow users to compose their questions in natural language and receive the answer in natural language.

5 Conclusion

To sum up, even though many models have been proposed, either by using theories of possibility or fuzzy logic, the problem of implementing a flexible fuzzy query system of database is still persisting. Therefore, improvements need to be made to provide flexible and user-friendly interfaces to RDBMSs.

References

1. Zemankova-Leech, M., & Kandel, A. (1984). *Fuzzy relational databases—A key to expert systems*. Verlag TUV Rheinland GmbH. B. Simpson et al. Title of paper goes here if known. unpublished.
2. Galindo, J., Urrutia, A., & Piattini, M. (2006). *Fuzzy databases: Modeling, design and implementation*. Idea Group Publishing.
3. Mama, R., & Machkour, M. (2019). A study on fuzzy interrogation systems of database. In *International conference of computer science and renewable energies (ICCSRE)* (pp. 1–6).
4. Kacprzyk, J., Zadrozny, S., & De Tré, G. (2015). Fuzziness in database management systems. *Fuzzy Sets and Systems, 281*(C), 300–307.

5. Buckles, B. P., & Petry, F. E. (1982). A fuzzy representation of data for relational databases. *Fuzzy Sets and Systems, 7*(3), 213–226.
6. Umano, M., Hatono, I., & Tamura, H. (1995). Fuzzy database systems. In *Proceedings of the FUZZIEEE/IFES'95 workshop on fuzzy database systems and information retrieval* (pp. 53–36).
7. Prade, H., & Testemale, C. (1984). Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries. *Information Sciences, 34*(2), 115–143.
8. Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems, 1*(1), 3–28.
9. Zadeh, L. A. (1981). Test-score semantics for natural languages and meaning representation via PRUF. In B. B. Rieger (Ed.), *Empirical semantics* (Vol. 1, pp. 281–349). Bochum Brockmeyer.
10. Zemankova, M., & Kandel, A. (1985). Implementing imprecision in information systems. *Information Sciences, 37*(1–3), 107–141.
11. Medina, J. M., Pons, O., Vila, M. A., & GEFRED. (1994). A generalized model of fuzzy relational databases. *Information Sciences, 76*(1–2), 87–109.
12. Galindo, J. (2005). New characteristics in FSQL: A fuzzy SQL for fuzzy databases. *Computer Journal of WSEAS Transactions on Information Science and Applications, 2*(2), 161–169.
13. Martinez-Cruz, C., Noguera, J. M., & Vila, M. A. (2016). Flexible queries on relational databases using fuzzy logic and ontologies, information sciences. *Vol., 366*(20), 150–164.
14. Bosc, P., & Pivert, O. (1995). SQLf: A relational database language for fuzzy querying. *IEEE Transactions on Fuzzy Systems, 3*(1), 1–17.
15. Kacprzyk, J., & Zadrożny, S. (2001). SQLf and FQUERY for access. In *Proceedings of the IFSA World Congress and 20th NAFIPS international conference* (Vol. 4, pp. 2464–2469).

QOS-Aware IOT Services Composition: A Survey

Boucetti Rabah, Hemam Sofiane Mounine, and Hioual Ouassila

Abstract The IoT services composition is a complex task, where the necessary services choice and the handsets in the proper order as an integrated composite service is the main challenge in a highly constrained environment. The services composition has been the subject of research in several academic works where the authors try to find the best way to combine and compose the services in an optimal way. In this paper we examine the composition mechanism of IoT services, by focusing on QoS-aware approaches, by a comparative study of some proposed approaches in the literature.

Keywords Services composition · IoT · Internet of Things · Quality of service · QoS

1 Introduction

The Internet of Things is a new paradigm that involves the integration of the physical world into the virtual world, they can change our life way by giving to our objects the ability to interconnect, exchange data and self-cooperating to achieve automatically a common goal, by the perception of the environment using sensors and change their state using actuators, by exploiting the existing internet infrastructure. Today, there is more and more IoT to make our lives more comfortable and our activities more profitable [1, 2].

Due to the increasing deployment of these smart objects and their heterogeneities, the encapsulation of their services into web services has been adopted

B. Rabah (✉)
ICOSI Laboratory, University Abbés Laghrour-Khenchela, Khenchela, Algeria

H. S. Mounine
University Abbés Laghrour-Khenchela, Khenchela, Algeria

H. Ouassila
LIRE Laboratory of Constantine, University Abbés Laghrour-Khenchela, Khenchela, Algeria

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_41

to improve homogeneity and facilitate their management. The encapsulation of the heterogeneous smart devices as web services uses principally the SOA model. In the service-oriented architecture (SOA), Web services are the basic components of any processing where their principles are based on description, publication, and invocation. The weak coupling between the components in the SOA architecture offers the reusability and composability of these services [3].

The services composition consists of combining the functionalities of several atomic services into a single business process in order to meet the complex needs that a single service can not satisfy by complying with the constraints (in terms of QoS) imposed by the user [4].

The IoT infrastructure is highly dynamic and heterogeneous. Providers may suggest a set of heterogeneous intelligent objects in the form of web services that can be collected by a coordinator service in order to provide a composite service [5]. Among the challenges of smart objects, the composition of IoT services. Combining the services of certain sensors and actuators to achieve a composite service adapted to the user's demands in a constrained and unstable environment with the limited processing and energy resources is a real challenge, compared to web services deployed in computers rich in resources and computing power [6].

The user request includes the functional properties of the services that are needed (service behaviors) as well as the non-functional properties (quality of services), such as cost, response time, reliability, etc. So, it is appropriate to introduce a quality of service (QoS) strategy for selecting and composing IoT services.

In this sense, several research works are devoted to the composition of IoT services, many of them take into account the nonfunctional properties (QoS) in this process, trying to give the most optimal approach to satisfy the preferences of the user in terms of QoS.

In this paper, we examine the compositional mechanism of IoT services by focusing on QoS-sensitive approaches by proposing a survey of some existing approaches in the literature with a comparative study based on specific criteria.

The structure of this survey study is designed as follows: in Sect. 2, some basic concepts related to the composition of IoT services are introduced. In Sect. 3, we review related work. The comparative study of approaches QoS-aware IoT services composition is provided in Sect. 4. Section 5 presents a discussion. Finally, a conclusion in Sect. 6.

2 Context

2.1 Web Service vs IoT Service

Web services are the main elements in the service-oriented architecture (SOA), they are accessible software components, open and described, offering atomic services. They exchange information (SOAP: Simple Object Access Protocol) independently on each other, the only link is the contract stipulating the means of communication, it means that the realized architecture is decoupled (weakly coupled). Describing web services in a standard way (WSDL) in a directory (UDDI) allows their reuse and their compositions in order to construct service offered to many users [7].

In realty the direct application of web services standards on IoT services is not appropriate and it will cause problems. The IoT environment is strongly constrained, it seems appropriate to disqualify SOAP, especially because of its heaviness. IoT requires the brevity and lightness of the exchanges that can be made with the RESTful approach via HTTP, with more concise way [8]. In addition, IoT services are directly related to the unstable physical world, unlike traditional services which are more stable virtual entities [6]. Also, IoT applications are more specifically adapted to their uses and their objects, and can only be used by its designer, the roles of the designer/the consumer are sometimes played by the same person in the IoT [9].

2.2 IoT Services Composition

In general, IoT smart devices are heterogeneously and dynamically integrated to provide functionality as a result of user requests. Each IoT device is linked to one (more) atomic service(s) that provides the functionality of the device. In most cases, a single IoT service cannot satisfy the user's compound requests because some of the complex services are provided by a set of IoT devices. In this case, a composite service is required. A composite service adequately combines a set of atomic services into one to successfully meet a complex objective that provides added value [10].

Given the diversity of IoT service in the cloud providers, a composite task is not necessarily managed by a single provider, which increases the complexity of the composition scenario. Naturally, the user sends his request to the service providers to obtain an appropriate response; it happens that a solitary provider is not sufficient to satisfy this request. Consequently, the involvement of several providers would be necessary [5].

Since the IoT environment is very dynamic, the quality of service criteria of the IoT services that are candidates for the composition are constantly oscillating. Therefore, it must be determined whether the composite service meets the user-imposed service level requirements by aggregating the quality of service criteria for atomic services. The quality of service of the composite service is related to the models of the composition. There are four composition models: sequential, cycle, parallel and branch, as described in Fig. 1a–d respectively [11, 12]:

(a) The sequential model, the tasks are executed in a sequential order;

(b) The cycle model, at least one task must be run more than once;

(c) The parallel model, the parallel tasks executed simultaneously, move on to the next task until all these parallel tasks are completed;

(d) The branch template, it selects a single task from a set of optional options and moves on to the next step.

2.3 Quality of Service (QoS)

Service qualities are keys that allow to take into consideration a composite service among several recommended candidates after the aggregation of service levels of atomic services according to the composition model (Fig. 1). The cost, response time, reputation, reliability are generally the main attributes of the quality of service on the Internet. However, IoT information is closely related to the physical world, making geographic location of devices an important attribute for user satisfaction. The quality of service in the IoT can be modeled as follows: price, response time, reliability, reputation, and geographical location. Each attribute can be defined as follows [4].

- Cost: the cost that the user must pay for the acquisition of sensory information;
- Response Time: The time interval between submitting a user request and accepting the service response. It includes the query processing time;
- Reliability: the rate of successful transmission of sensory data to users over a period of time. The success rate of the information service offering may be affected by the environment;
- Reputation: It is considered as a global indicator of the user experience. Service reputation is defined as the average score of the multi-assessment;

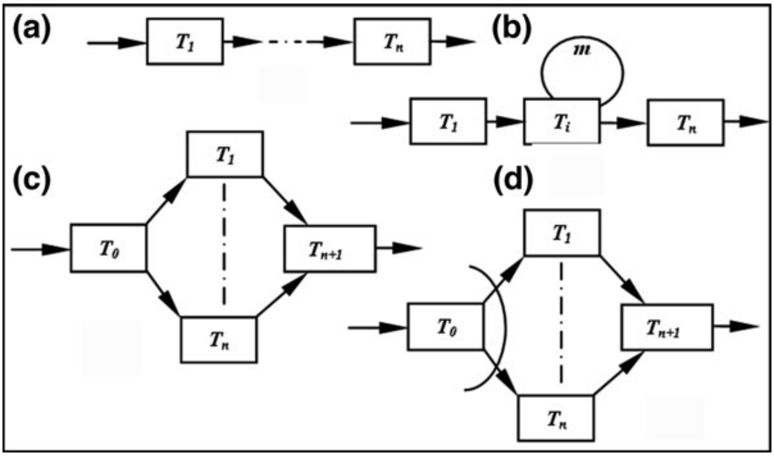


Fig. 1 Basic models for web service composition [11]

- Location: Sensory data is closely related to the geographic environment of the devices. There is always a distance between the detection devices and the destination specified by the user. The smaller geographical distance, the more precise the sensory data.

3 Related Work

According to our reviews, no analysis about IoT service composition sensitive to the QoS has been produced. In this section, we summarize the main and closest works related to the Iot service composition:

In [13], a comparative study of existing IoT service composition approaches in general has been described (QoS-sensitive or non-QoS-sensitive), and compared against each other based on the following criteria: Adaptivity, Autonomy and Independence, Decentralization and Distribution, Protection and Confidence Management and Optimization of Composition.

In [14], the authors propose a systematic study of service composition approaches in IoT. They focus on several IoT service composition approaches (published between 2012 and 2017) based on the Systematic Literature Review (SLR) method, in which they present a technical taxonomy with regard to functional and non-functional aspects of composition approaches. A comparison is made between the studied approaches based on certain technical aspects such as system correction factors in functional properties approaches, and QoS factors, algorithms and existing platforms in non-functional approaches. The authors discuss the advantages and disadvantages of each approach, with some tips for resolving weaknesses.

Other closely works such as [15–17] are realized, but in the field of web services. In [15], a systematic review of the literature is presented. The authors attempt to systematically classify and compare existing research methods and techniques on the composition of Web services that take into account the QoS (published between 2005 and 2015) based on the artificial intelligence.

In [16], a review about the web service composition approaches is proposed. The authors categorized these approaches into three categories: workflow-based, XML-based, and ontology-based. Then, a comparison between these approaches is made according to certain criteria, such as quality of service, scalability and correction.

The work [17] offers a comparison between semantic web service composition approaches. The approaches are categorized into two classes: Semantic Web Services composition approaches with QoS support, and Semantic Web Services composition approaches without QoS support. The systematic comparison between the two categories of approaches is based on six criteria: extensibility, non-determinism, dynamic aspect, adaptability, domain independence, accuracy, semantic capacity and QoS awareness. The objective of this study was to identify the best approach that can be used for the composition of semantic Web services.

4 QoS-Aware IoT Services Composition Approaches

In this section we will present some existing approaches in the literature relative to the composition of IoT services sensitive to the quality of service. After, we will make a comparative study of these approaches.

4.1 QoS-Aware Approaches

In this sub-section, we present some suggested approaches for the composition of IoT services sensitive to the quality of service:

In [18], the authors proposed a method based on the cooperation between two algorithms: the Genetic algorithm (GA) and the PSO (Particle Swarm Optimization) to solve the problems of composition of the IoT service controlled by the quality of service (QoS), the proposed approach provides the following elements: Improvement of the best local optimal priority strategy which guarantees the weighting of components in the local aptitude of a task and in the aptitude of the Composition Service. The relative importance of the quality of service attributes of a composite service is the same as the size of the potential candidate services. The local enhancement of an IoT service can properly guide algorithm search when a better candidate service is selected to accomplish the corresponding task.

In [19], the QoS-oriented service composition (QSC) is formulated as a Multi-Criteria Goal Programming model (MCGP) to discover the appropriate composite service that meets the constraints of the user's service quality. A multi-population genetic algorithm (MGA) is proposed to solve this model. The MCGP automatically assigns high-quality web services to the composite service and also finds candidate composite services by easing QoS constraints when no composite service can strictly satisfy the constraints of the service quality imposed the user.

In [20], the authors consider both energy consumption and quality of service, they have modeled the composition problem of the IoT service into a short-path optimization problem with two objectives (BSPO: Bi-objective Shortest Path Optimization): (1) Minimize the quality of service, including execution time, network latency and price of service. (2) Minimize the energy consumption of the composite service. They use the exact algorithm called pulze to solve the problem. The proposed model is scalable for large scale IoT environments and offers an optimal balance between QoS level and energy consumed.

In [14], the authors propose Clustering-based and QoS-aware services Composition Algorithm (CQCA) for the composition of services in an ambient intelligence environment (AmI). First, by using the K-means method, the candidate services will be partitioned into clusters, each cluster represents a QoS level, this partitioning allowed to determine the relevance of the candidate services for the composition process. For the sake of optimality and by using the characteristics of the resulting clusters, a new formulation of the utility function of the candidate services is proposed by the authors to eliminate the candidate services which are not promising in terms of quality of service which reduces the search space as well as the composition time Then, they determine the QoS level of the candidate services that satisfy the global QoS constraints by exploiting the lexicographic optimization method. Finally, from the selected candidate services, a search tree is constructed to determine the quasi-optimal compositions.

In [21], the authors introduce energy into workflow management in wireless sensor networks (WSN). First, they extend the traditional CoS (Quality of Service) model with the proposition of a method for calculating energy. Then, they adopt a quality of service constraint decomposition approach. By considering the QoS of the service and its energy, the global constraints will be decomposed into a set of local quality of service constraints for the atomic services. Finally, local services will be efficiently selected.

In [22], an energy-centric and QoS-aware IoT (QoS) service composition approach is proposed. The authors model and solve the problem as a multi-objective optimizations problem which allows to minimize power consumption in order to ensure high availability of composite services while satisfying the user QoS requirements. First, each sub-problem is linked to a quality of service attribute and processed by a lexicographic optimization method to preselect service based on the quality of service level. The preselected candidate services are then compared to select the best service using Pareto-related relative dominance relation concept based on its energy profile, QoS attributes and user preferences. The EQSA algorithm (energy-centered and QoS-aware services selection algorithm) is able to find solutions very close to the optimum (about 98%).

In [23], for an event service composition, the authors propose a QoS aggregation scheme to compute the global QoS vector. The degree of optimization of the event composition is calculated using a QoS utility function, taking into account the user constraints and the weighting vector. Finally, a genetic algorithm is developed to create optimal event service compositions. They were able to show that the genetic algorithm developed gives between 79% and 97% of the optimized results, exploiting the compromise between convergence time and degree of optimization.

In [12], the authors decompose the complex calculation model of the QoS into four basic models and assign to each model a calculation method. To suit the IoT characteristics, a back-Track (BT) algorithm is proposed to calculate the quality of service. After analysis and comparison with two algorithms (integer programming (IP) and the genetic algorithm (GA)), the results show that BT algorithm is better suited to the large scale IoT services.

In [24], a service quality monitoring and service composition system (SMACS) in environmental sensor networks is presented. Reliability and reactivity are the two QoS parameters estimated by SMACS using a naive Bayesian classifier. In addition, SMACS uses these parameters to get an optimal web services composition, which improves the overall quality of service of the entire environmental sensor network by carefully excluding the failed nodes when responding to user requests.

In [4], the authors propose an effective quality of service strategy to solve the problem of the IoT information services composition, they use multi-attribute decision making (MADM) [25] to evaluate each IoT service and select the elites in terms of QoS to be part of the composition. An improved genetic algorithm (GA) is used as a global optimization method. The optimal solution is the service composition scheme with the global maximum QoS value.

4.2 Comparative Study

In this sub-section, we will try to give a comparative study illustrated in “Table 1” between the approaches mentioned in section *A*. This comparison is based on the following criteria:

1. Method of resolution: it means the composition algorithms and methods used in the calculation and evaluation of QoS;
2. Scalability: it is the evaluation of the method (the algorithm) on the IoT environment on a large scale;
3. Optimality: to measure if the approach gives optimal results or not (optimal, quasi-optimal);
4. Approach: it means the approach used in the composition of IoT services, Non-heuristic, heuristic or meta-heuristic approach;
5. Multi-objective optimization: to analyze if multi-objective optimization is used in compositional procedures or not;
6. The QoS parameters considered: this is the enumeration of the quality of service parameters considered in the service composition method (algorithm).

5 Discussion

Non-heuristic approaches use the algorithms: pulse algorithm [20], linear programming (LP) [21], the backtracking algorithm (BT) [12] and the naive Bayesian Classifier [24]. The heuristics and meta-heuristics use the Genetic algorithm (GA) and Optimization by particle swarms (PSO) [18], the multi-population genetic algorithm (MGA) [19], the K-means method and Lexicographic Optimization [26], Lexicographic Optimization and Pareto Domination [22], the Genetic Algorithm (GA) [23], the Enhanced Genetic Algorithm [4].

The table also shows that most solutions give quasi-optimal results like [4, 18, 19, 23, 26], the [22] is very close to the optimal, these solutions are based on heuristic and meta-heuristic approaches. On the other hand, optimal compositions are found in [12, 20, 21, 24] where the authors used non-heuristic approaches.

Table 1 Comparative study between the approaches already proposed

	The resolution method	Scalability	optimality	Approach	Multi-objective optimization	The QoS parameters considered
[18]	Genetic algorithm (GA) + Optimization by particle swarms (PSO)	Not evaluated	Near to optimal	Meta-heuristic	Not supported	Cost (C), Availability (A), Time (T), Reliability (R).
[19]	Multi-population genetic algorithm (MGA)	Yes	Near to optimal	Meta-heuristic	Supported	Execution time (t), Reliability (r), Execution cost (c).
[20]	Pulse algorithm (bi-objective shortest path optimization (QoSU, EP))	Yes	Optimal	Non-heuristic	Supported	Execution time (t), Cost (C) + Energy consumption (EP).
[26]	Clustering-based and QoS-aware services Composition Algorithm (CQCA) (K-means method + Lexicographic optimization + Search tree)	Yes	Near to optimal	heuristic	Not supported	Response time (ReT), Throughput (Th), Availability (Avl), Reliability (Rel), Cost (C)
[21]	K-means + utility function + Linear Programming (LP)	Not evaluated	Optimal	Non-heuristic	Supported	Execution Time, Cost, Availability + Energy consumption
[22]	Energy-centered and QoS-aware Services selection Algorithm (EQSA) (Lexicographic optimization + Pareto dominance)	Yes	Very close to optimal (about 98%)	Meta-heuristic	Supported	Cost, Availability, Reliability + Energy consumption

(continued)

Table 1 (continued)

	The resolution method	Scalability	optimality	Approach	Multi-objective optimization	The QoS parameters considered
[23]	QoS utility function + genetic algorithm (GA)	Yes	Near to optimal	Meta-heuristic	Not supported	Latency (L), Price (P), Energy Consumption (E), Bandwidth Consumption (B), Availability (Ava), Completeness (C), Accuracy (Acc), Security (S)
[12]	Backtracking algorithm (BT)	Not evaluated	Optimal	Non-heuristic	Not supported	Response time (q_t), Reliability (q_r), Availability (q_a).
[4]	Improved genetic algorithm (Calculation of QoS performance based on Multi-attribute decision- making (MAMD)).	Yes	Near to optimal	Meta-heuristic	Not supported	Price, Responsetime, Reliability, Reputation, Geographic location
[24]	Service Monitoring And Composition System (SMACS) (Naïve Bayesian classifier)	Not evaluated	Optimal	Non heuristic	Not supported	Reliability, Responsiveness

Approaches in [4, 19, 20, 22, 23, 26] are scalable on large scale IoT environments. On the other hand, the approaches [12, 18, 21, 24] are not evaluated.

The composition of IoTQoS-aware services is modulated as a problem of multi-objective optimizations in [4, 19–22].

The QoS parameters considered in the composition approaches of IoT services cited are: Cost, Availability, Reliability, Run Time, Response Time, Rate, Latency, Bandwidth Consumption, Completeness, Accuracy, Security, and Responsiveness. In the approaches of [20–23] in addition to the QoS, introduce the energy consumption as a parameter in the optimization of the composition. Geographic Position is also considered in [4] with the other QoS parameters.

6 Conclusion

In this paper, we have described a comparative study of some approaches of the composition of IoT services that are sensitive to quality of service. We have made a comparison based on the algorithms used, which we noticed that the majority of these algorithms are heuristic or meta-heuristic, this is justified by the fact that the composition of web services taking into account the QoS is NP-hard and it must be resolved within an acceptable time.

It was found that the most important QoS attributes were response time, cost, availability, and reliability. Energy consumption and geographical position are two attributes that are important in the composition of IoT services, and this can also be justified by the need for energy optimization of connected objects that are also closely related to the physical world (location).

QoS-aware IoT service composition approaches have a large share in the field of service composition in general and we believe that it requires even more research in order to be able to optimally satisfy user preferences in terms of QoS.

References

1. White, G., Nallur, V., & Clarke, S. (2017). Quality of service approaches in IoT: A systematic mapping. *Journal of Systems and Software*, 186–203.
2. Jang, J., Jung, I. Y., & Park, J. H. (2017). *An effective handling of secure data stream in IoT*. Applied Soft Computing.
3. Issarny, V., Bouloukakis, G., Georgantas, N., & Billet, B. (2016). Revisiting service oriented architecture for the IoT: A middleware perspective. In *14th international conference on service oriented computing (ICSOC)*.
4. Yang, Z., & Li, D. (2014). IoT information service composition driven by user requirement. In *2014 IEEE 17th international conference on computational science and engineering* (pp. 1509–1513).
5. Baker, T., Asim, M., Tawfik, H., Aldawsari, B., & Buyya, R. (2017). An energy-aware service composition algorithm for multiple cloud-based IoT applications. *Journal of Network and Computer Applications, 89*, 96–108.

6. Li, K., & Jiang, L. (2012). The research of web services composition based on context in Internet of Things. In *IEEE international conference on computer science and automation engineering (CSAE)* (pp. 160–163).
7. Papazoglou, M. P., & Van Den Heuvel, W.-J. (2007). Oriented architectures: Approaches, technologies and research issues. *The VLDB Journal*, 389–415.
8. Guinard, D., Trifa, V., Karnouskos, S., Spiess, P., & Savio, D. (2010). Interacting with the SOA-based internet of things: Discovery. *Query, Selection, and On-Demand Provisioning of Web Services, IEEE Transactions on Services Computing, 3*(3), 223–235.
9. Karmouch, A., Roussel, G., Festor, O., Nicaud, C., Bonnin, J., & Ghamri-doudane, Y. (2013). *Architecture and application protocols for services choreography in the Internet of Things*. Doctoral thesis.
10. Deng, S., Xiang, Z., Yin, J., Taheri, J., & Zomaya, A. Y. (2018). Composition-driven IoT service provisioning in distributed edges. *IEEE Access, 6*, 54258–54269.
11. Zhao, X., Song, B., Huang, P., Wen, Z., Weng, J., & Fan, Y. (2012). An improved discrete immune optimization algorithm based on PSO for QoS-driven web service composition. *Applied Soft Computing, 12*(8), 2208–2216.
12. Ming, Z., & Yan, M. (2013). QoS-aware computational method for IoT composite service. *The Journal of China Universities of Posts and Telecommunications*, 35–39.
13. Aoudia, I., Benharzallah, S., Kahloul, L., & Kazar, O. (2017). *A comparative analysis of IoT service composition approaches*.
14. Asghari, P., Rahmani, A. M., & Javadi, H. H. S. (2018). Service composition approaches in IoT: A systematic review. *Journal of Network and Computer Applications, 120*, 61–77.
15. Jatoth, C., Gangadharan, G. R., & Buyya, R. (2017). Computational intelligence based QoS-aware web service composition: A systematic literature review. *IEEE Transactions on Services Computing, 10*(3), 475–492.
16. Tabatabaei, S. G. H. , & Ibrahim, S. (2008). *A review of web service composition approaches.*
17. Radziahand, M., & Furkh, Z. (2011). Comparative evaluation of semantic web service composition approaches. *Communications in Computer and Information Science*, 283–290.
18. Jin, L., Yuxi, C., Xu, C., Jianli, D., Kaushik, C., Qiping, H., & Shenling, W. (2013). *A cooperative evolution for QoS-driven IOT service composition*. Automatika.
19. Li, Q., Dou, R., Chen, F., & Nan, G. (2014). A QoS-oriented Web service composition approach based on multi-population genetic algorithm for Internet of things. *International Journal of Computational Intelligence Systems*, 26–34.
20. Osama, A., Ibrahim, M., & Tein-Yaw, C. (2018). *Bi-objective optimization for energy aware IoT service composition*. IEEE Access.
21. Tong, E., Chen, L., & Li, H. (2017). Energy-aware service selection and adaptation in wireless sensor networks with QoS guarantee. *IEEE Transactions on Services Computing, 13*, 829–842.
22. Khanouche, M., Amirat, Y., Chibani, A., Kerkar, M., & Yachir, A. (2016). Energy-centered and QoS-aware services selection for internet of things. *IEEE Transactions on Automation Science and Engineering*, 1–14.
23. Gao, F., Curry, E., Intizar, A., Bhiri, S., & Mileo, A. (2014). QoS-aware complex event service composition and optimization using genetic algorithms. In X. Franch, A. K. Ghose, G. A. Lewis, & S. Bhiri (Eds.), *Service-oriented computing. ICSOC*. Lecture Notes in Computer Science.
24. Cuong, N., Domínguez, F., Long, N. T., Touhafi, A., & Steenhaut, K. (2016). Service composition with quality of service management in environmental sensor networks. *International Journal of Ad Hoc and Ubiquitous Computing, 23*(3/4), 216–229.
25. Xu, J., & Wu, W. (2006). *Multiple attribute decision making theory and methods* (pp. 37–116). Tsinghua University Press.
26. Khanouche, M., Attal, F., Amirat, Y., Chibani, A., & Kerkar, M. (2019). Clustering-based and QoS-aware services composition algorithm for ambient intelligence. *Information Sciences, 482*, 419–439.

SDN-Based Approaches for Heterogeneity and Interoperability in Internet of Things: An Overview

Sihem Benkhaled, Mounir Hemam, and Moufida Maimour

Abstract The Internet of Things (IoT) and Software Defined Network (SDN) are two emerging technologies. The IoT aims to connect devices over the Internet. Networks on which these IoT devices operate will continue to be heterogeneous and the network becomes more complex. SDN paradigm provides efficient network management by decoupling the control plane and the data plane, it allows facing network heterogeneity and interoperability (networks-to-network communications) through a programmable manner, this paradigm allows the communication between various objects independently of their respective technologies and thus allows dealing with the problem of heterogeneity. The main objective of this chapter is to present an overview of SDN-based pertinent solutions that consider network level interoperability and heterogeneity in the IoT.

Keywords Internet of Things (IoT) · Interoperability · Heterogeneity · Software Defined Networks (SDN)

1 Introduction

Internet of Things (IoT) [1] is a novel emerging technology where any things from anywhere are connected together as well as to the Internet infrastructure at any time. IoT combines aspects and technologies coming from different disciplines. IoT devices such as laptops, smartphones, sensors and actuators are expected to exceed 45 billion by 2020. These devices may use various types of technologies and communication protocols and may produce large volumes of data. Hence, the need

S. Benkhaled (✉) · M. Hemam
Department of Computer Science, University of Khenchela, Khenchela, Algeria

M. Maimour
University of Lorraine, CRAN, UMR, 7039, Nancy, France
e-mail: moufida.maimour@univ-lorraine.fr

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_42

for installing new network access and core devices will increase and hence create several issues such as interoperability, heterogeneity, security, and scalability [2].

Particularly, in [2] the authors state that 40% of the potential benefits of IoT can be obtained with the interoperability between IoT systems. It is fascinating that this issue can be easily and dynamically resolved thanks to the paradigm of Software Defined Networking (SDN). It mainly allows hiding all the control and management operations by setting them inside a middleware layer, which alleviates the dependency from vendors.

In this chapter, we are interested in the use of Software defined networks technology to provide solutions to network level interoperability. Section 2 introduces the issue of network interoperability in the IoT and Sect. 3 presents the concept of SDN. Section 4 presents SDN-based IoT networks. Section 5 overviews research work where SDN-based technologies are used to overcome issues related to heterogeneity and interoperability in the IoT system.

2 Network Interoperability in IoT

The IEEE defines interoperability as the ability to interconnect and communicate different systems to form a cost effective and easy to implement network [3]. With respect to heterogeneity, interoperability can be seen from different perspectives, namely, device interoperability, network interoperability, syntactic interoperability, and semantic interoperability.

As networks on which IoT devices operate on will continue to be heterogeneous, network interoperability will continue to be more than required. Network level interoperability deals with mechanisms to enable seamless message exchange between systems through different networks (networks of networks) for end-to-end communication. In fact, IoT devices generally rely on various short ranged wireless communication and networking technologies [3]. To make systems interoperable, each system should be able to exchange messages with other systems through various types of networks. Due to the dynamic and heterogeneous network environment in IoT, the network interoperability level should handle issues such as addressing, management, routing, resource optimization, scalability, security, QoS, and mobility support [3].

In this level, protocol interoperability is the main focus. At the standardization level, the IETF has developed a set of standards for routing such as RPL, CORPL, and CARP and solutions for encapsulation including 6LowPAN, 6TiSCH, 6Lo, and Thread [3]. Another new solution to address interoperability in this level is software-based approaches such as Software Defined Networking (SDN) which hides all the control and management operations from the IoT devices by setting them inside a middleware layer, which alleviates the dependency from vendors [4].

3 Software Defined Networking (SDN)

SDN is a new networking paradigm that addresses the limitations of current network infrastructures. It brings flexibility allowing different objects connected to heterogeneous networks to communicate with each other. Simultaneous connections of various communication technologies can be handled.

Network management decisions such as routing, scheduling can be done at the SDN controller.

SDN architecture can be characterized by: (1) Decoupled data and control plans; (2) Routing decisions based on flow, rather than destination; (3) Control logic is moved to an external entity (SDN controller or network operating system); and (4) Network programmable by software applications that interact with data plane devices [4].

The separation of interests or concepts, introduced between the definition of network policies, their implementation in switching hardware, and the routing of traffic, is the key to the desired flexibility: by breaking the network control problem into manageable parts, SDN makes it easy to create and introduce new network abstractions, simplifying network management and making it easier to evolve.

SDN is composed of three layers: Data plan layer, Control layer, and application layer [5]. Th*e data plan layer* is composed of various networking equipments which form the underlying network that forward network traffic. It could be a set of network switches and routers in the data center. This layer would be the physical one over which network virtualization would be laid down through the control layer (where SDN controllers would sit and manage underlying physical network).

The control layer is an intermediate between the infrastructure (data plane) layer and the application layer. It is implemented in a logically centralized controller (or network operating system), simplifying policy enforcement and network configuration and evolution. It can manage physical devices with various features and functions via different interfaces from lower layer (southbound API).

The application layer is an open area to develop as much innovative application as possible by leveraging all the network information about network topology, network state, network statistics, etc. There may be several types of applications which can be developed like those related to network automation, network configuration and management, network monitoring, network troubleshooting, network policies and security. Such SDN applications can provide various end-to-end solutions for real world enterprise and data center networks.

Traditional networks are usually rigid, fixed and complex which is difficult to manage and configure. Heterogeneity, scalability, and interoperability have been major challenges. SDN is a promising paradigm to provide solutions to network interoperability and heterogeneity issues through virtualization [4].

SDN is applied to the IoT to facilitate networking applications (heterogeneity of protocols, QoS management, and security). Even though the devices have different protocols, the forwarding devices in the router are able to convert them in an understandable form by the receiver [4].

4 IoT Systems with Software Defined Networking (SD-IoT)

In recent years, the SD-IoT has become one of the main topics of IoT-related researches. It deals with several issues for deploying and managing numerous nodes of IoT infrastructure by adopting the idea of SDN [6].

The IoT aims to connect objects over the Internet and the SDN is expected to resolve traditional network issues like heterogeneity, interoperability, and scalability by allowing orchestration for network management and network interoperability by decoupling the control plane and the data plane [3].

Integrating IoT and SDN will increase network efficiency as it will make it possible for a network to respond to changes or events detected at the IoT application layer through network reconfiguration. It allows routing efficiency, high data transmission, network management and resource allocation for the IoT devices to meet the growing need of the user demands [7].

The general architecture for SDN-based IoT is shown in Fig. 1 redrawn from [1]. SD-IoT architecture deals with clear separation of concern between services provided in the control plane and the data plane. Control plane specifies the management of network traffic and data plane specifies the mechanisms to forward traffic to desired destination. SDN-based IoT management specifies how the applications on top of the Management Layer interact with the control plane and the coordination among them.

There are three layers in the SD-IoT architecture. The data plane includes sensors which sense and forward data in the network. The control plane that consists in the

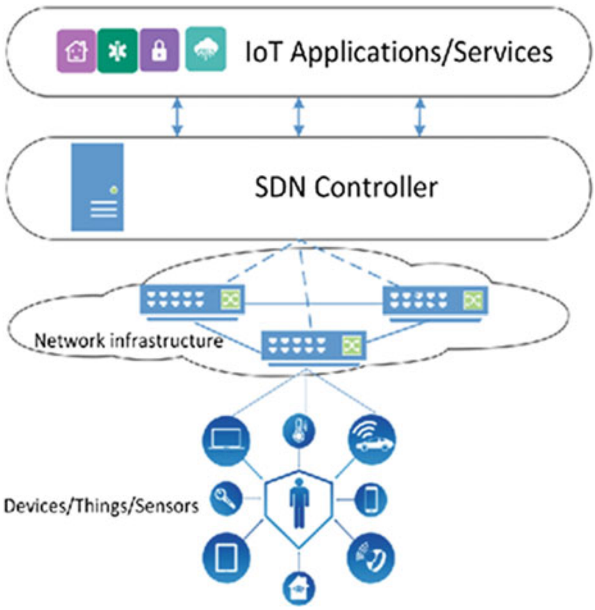


Fig. 1 IoT architecture based SDN technology [1]

controller which controls the whole network. The application plane includes diverse network applications such as routing.

IoT networks need to be based on a robust architecture to answer these requirements and inspired by SDN to make them more flexible.

The SDN controller provides programmability and flexible management for flow forwarding state in the data plane by having a global view of the network. SDN can facilitate high data transmission, spectral efficiency, resource allocation and network management for the IoT devices for fulfilling growing need of the customer demands (Fig. 1).

5 Overview of SDN–Based Solutions to Network Heterogeneity and Interoperability in IOT

SDN has been applied to IoT to facilitate networking applications while dealing with heterogeneity, network and device management, routing, and security issues. In order to have an interoperable network and overcome these issues, it is necessary to create and manage the network devices efficiently so an interoperable network is obtained with efficient data exchange [8].

This section gives a general overview of how applied SDN to IoT in order to solve heterogeneity and interoperability issues in the connectivity and network level. We focus on the pertinent approaches (summarized in Table 1*)* which can be potentially fulfilled by SDN technology.

In this context, Jie. Li, et al. proposed a general framework for IoT based on SDN with distributed controller and managing. Another conceptualization of generic IPv6-based IoT devices architecture based on SDN for smart cities is presented in [15]. Hung et al. [10] propose SDN-based management framework for the IoT devices, as the 6LoWPAN protocol gains momentum, particularly in view of the IoT new network management solutions are developed to cater for these devices.

Jia Qiang Liu, et al. proposed an SD-IoT architecture in [9], for the smart management of IoT devices. In [13], A. Desai et al. proposed an architecture for IoT integrated into SDN in which IoT devices are managed by OpenFlow based management devices. Wu et al. in [19], presents UbiFlow framework which provides the integration of the SDN and the IoT, proposed an efficient flow control and mobility management in urban multi-networks using SDN distributed controllers. In UbiFlow architecture, IoT network is partitioned into small network chunks/cluster. Each partition is controlled by a physically distributed SDN controller, the IoT devices in each partition may be connected to the different access point for different data requests. In [12] the authors propose SDN-based cloud platform approaches for IoT network connectivity. The work in [14] presents general SDN-based architectures to facilitate the scalability, heterogeneity, and interoperability among IoT devices or nodes.

Table 1 Proposed IoT approaches based on SDN for heterogeneous networks

Approaches	Goals	Proposition	Control plane decoupling	Advantage
Liu et al. [9]	Device management High level management capabilities Low latency and heterogeneity Mobility using fog computing	Edge computing enabling the IoT	Distributed ODL, Onix, and ONOS controllers SD fog gateways, OF-switches	Supports multiple identification and communication Technologies Multiple SD fog gateways ensure interoperability Centralization leading to security enhancement to some extent
Huang et al. [10]	Scalability Efficient interoperability and traffic Engineering	Framework for monitor IoT devices	Centralized SDN controllers	Architecture based on independent IoT system and efficient management IoT devices
Hakira et al. [11]	Networking, mobility, standardization, security, QoS	IoT architecture combining SDN with message-based publish/subscribe DDS middleware	Centralized SDN controller	Filtering and fusion mechanism for efficient traffic engineering
Desai et al. [12]	Heterogeneity, IoT device to cloud communication	OF-enabled management device	Distributed NOX, POX, ODL	Proposed an OF-enabled management device, which will make network simpler
Salman and Jain [13]	Single solution for multiple challenges: Scalability, heterogeneity, QoS, latency, reliability, security	Centralized SDN control network for IoT with decentralized data management Layered model: Works in application, control, network, and device layers Implements SD-gateways in the fog with specialized algorithms	Distributed SDN controller	Inter-controller communication. Intelligent fog nodes SDN controller uses management protocols (i.e., NetConf and Yang, OF-Config and extended OF) Use of unified application for communication

(continued)

Table 1 (continued)

Approaches	Goals	Proposition	Control plane decoupling	Advantage
Lin et al. [14]	Heterogeneity, interoperability, scalability, security, QoS	Centralized global view Heterogeneous devices with various data formats for information modeling	Centralized layered IoT controller	Adaptable network state Minimized latency and optimized interoperability and scalability Better performance and flow scheduling
Li. et al. [15]	Heterogeneity, interoperability, scalability, availability, security	IoT gateways and SDN switches. Distributed network OS	Distributed SDN controller	Distributed OS providing centralized control Global view of the underlying physical distributed network environment
Iyer et al. [16]	Heterogeneity, interoperability, latency, scalability, reliability, security	SDN gateway/router	Distributed network OS Distributed POX	Discovering IoT devices from different domains Real time evaluation for latency in IoT devices and sensors, using Raspberry Pi
Mann et al. [17]	Interoperability, device discovery, scalability, security, efficiency management, flexibility	An SDN-IoT architecture with NFV implementation	SDN controller and virtualized IoT gateways	Enhanced performance and management of hardware, software and virtual resources Device discovery with enhanced connectivity
Qin et al. [18]	Heterogeneity, scalability, mobility	Replacement of traditional gateway with SDN gateway	Distributed ONOS, ODL	Improved network efficiency and agility SDN-enabled gateway Intelligent routing protocols and caching techniques Architecture only

(continued)

Table 1 (continued)

Approaches	Goals	Proposition	Control plane decoupling	Advantage
Desai et al. [13]	Scalability, heterogeneity, agile and inexpensive	SD solution for IoT to forward store, and secure data	Distributed (multiple SDN controllers)	Maniples SD application modules to facilitate IoT network
Wu et al. [19]	Interoperability, device discovery, scalability, security, efficiency management, flexibility	An SDN-IoT architecture with NFV implementation	SDN controller and virtualized IoT gateways	Enhanced performance and management of hardware, software and virtual resources Device discovery with enhanced connectivity
Julia and Skarmeta [8]	– Efficient routing – Cost-effective deployment	Resolves CAPEX issues in IoT	Centralized SDN controller	Efficient inter-domain routing Less connected and deployed devices, hence cost-effective
Qin et al. [18]	Low cost IoT Enhanced scalability and interoperability	An SDN/NFV-enabled edge node, which orchestrates end-to-end SDN-IoT services	Distributed SDN controller IoT gateways OF-switches	ODL and OpenStack Nova/Havanna service controller GMPLS controlled optical network Multi domain network architecture

Salman and Jain [13] proposes architecture, with layered model, for IoT with decentralized data and centralized control. Authors also discuss IoT challenges like scalability, big data, heterogeneity, and security. The proposed four layered model consists of Application, Control, Network, and Device Layer.

The architecture uses unique identifiers in device layer that ensures interoperability, security, and quick address. Software Defined Gateways (SD-Gateways) [20], a virtualized abstraction of a common gateway supporting extended OpenFlow protocol to communicate with the SDN controllers.

Li et al. [16] discusses issues like interoperability from the perspective of devices, data, communication protocols, and re-usability of data generated from IoT devices. Moreover, authors suggest resource utilization, openness and interoperability by using a layered architecture which includes Device Layer (responsible for collecting data), Communication Layer (contains SDN-enabled switches and gateways), Computing Layer (having SDN Controller), and Service Layer (which provides services). The IoT devices communicate with the SDN gateway/router through sinks, like Raspberry Pi. The gateway/router then forwards the data to the SDN controller. The SDN controller manipulates the data as per the application requirements located at the service layer. This is done by programming the SDN controller. Limitations of this work include sink and sensing devices, which work independently while only responsible for aggregating and caching the data received from IoT devices. Its architecture lacks security mechanisms and routing algorithms both in SDN controller and IoT Gateway.

Ojo et al. [17] proposes a replacement of traditional IoT gateways, with specialized SDN-enable gateways, these gateways are capable of managing wired and wireless devices, and claims to be more flexible, efficient, and scalable. Authors also claim that the gateway can perform efficient traffic engineering with intelligent routing protocols and caching techniques across less constrained paths. However, the work is limited in defining intelligent routing algorithms, and performance evaluation or implementation in real time which is considered a future direction. Hakiri et al. [11] discusses five key network related challenges of IoT, such as current standardization efforts, mobility management, recurring distributed systems issues, communication protocols, and security and privacy. They outline an IoT architecture that combines SDN with message-based publish/subscribe Data Distribution Service (DDS) middleware to solve variety of issues like networking, mobility, standardization, and QoS (Quality of Service) support. In this framework, smart devices are linked with SDN-based IoT gateways to communicate with SDN forwarding devices. Furthermore, an SDN controller connects to the forwarding devices using southbound APIs allowing asynchronous, anonymous, and many-to-many communication semantics.

Martinez-Julia and Skarmeta [8] used SDN to allow different objects from different networks to communicate with each other using IPv6 and at the same time simplify the management and control operations of various objects types by adding an additional IoT controller over the SDN controller. Thus, even so the devices have different protocols; the forwarding devices in the router convert it in a form understandable by the receiver. This enables the communication of diverse devices in the network.

A middleware is designed and implemented by Qin et al. [18], which is composed of a layered IoT SDN controller to manage distributed, heterogeneous, and dynamic IoT multi-network. In their research, a central controller monitors the existing resources and schedules the data streaming according to the specific service requirement e.g., a minimum data rate, maximum tolerable delay or packet loss for each separate flow.

Finally, in the same [18] Multi-network information architecture (MINA) proposed with reflective self-observing that maintaining proactive plans approach to address this problem of heterogeneity for multi-network IoT environment by providing a centralized global view much like SDN for overlay structure and can perform analysis of the network state.

6 Discussion

In this chapter, we agree that the lack of interoperability is a consequence of several problems such as the heterogeneity in the objects, the communication protocols and the network routing protocols.

The different approaches, based on SDN, presented in this chapter, give certain advancement in research. It should be noted that the most SD-IoT approaches are neither implemented nor validated. So, we should contribute with an implemented approaches, this should address the transparent mobility of intelligent objects between different IoT architectures, research on routing mechanisms to overcome the disadvantages of traditional routing protocols and use the virtualization of network.

We can try to integrate this approach in a whole IoT architecture by considering the other interoperability levels like semantic data. Each layer of IoT interoperability model is the prerequisite for upper layers.

7 Conclusion and Perspectives

In this chapter, we have answered two questions, what are the different levels of IoT interoperability and how the SDN technology can face one of this levels that is network level.

We have presented the most pertinent SD-IoT approaches. Those approaches trend to focus more on communicating and networking layers by exploit the characteristics of SDN technology. All these presented approaches focus on designing solutions to solve the interoperability in network functions, only in one level without considering the other upper IoT levels.

Finally, we consider it is necessary to develop an interoperability approach which covers different layers in the same IoT architecture. Each layer of IoT interoperability model is the prerequisite for upper layers; not only the networking level but the semantic data level also it's a very important challenge.

References

1. Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer Networks, 54*(15), 2787–2805.
2. Sundmaeker, H., Guillemin, P., Friess, P., & Woelfflé, S. (2010). Vision and challenges for realising the internet of things. *Cluster European Projects on the Internet Internet Things, European Commission, 3*(3), 34–36.
3. Hahm, O., Baccelli, E., Petersen, H., & Tsiftes, N. (2016). Operating systems for low-end devices in the internet of things: A survey. *IEEE Internet of Things Journal, 3*(5), 720–734.
4. Ashton, K. (2009). The internet of things. *RFID Journal, 22*(7), 97–114.

5. Bizanis, N., & Kuipers, F. (2016). *SDN and virtualization solutions for the internet of things: A survey* (pp. 5591–5606). IEEE Access.
6. Kreutz, D., & Ramos, F. (2015). Software-defined networking: A comprehensive survey. In *Proceeding IEEE* (pp. 14–76).
7. Han, Z., & Ren, W. (2014). A novel wireless sensor networks structure based on the SDN. *International Journal of Distributed Sensor Networks, 10*(3), 1–8.
8. Julia, P. M., & Skarmeta, A. F. (2014). *Extending the internet of things to IPv6 with software defined networking*. White Pap.
9. Liu, J., Li, Y., Chen, M., Dong, W., & Jin, D. (2015). Software-defined internet of things for smart urban sensing. *IEEE Communications Magazine, 53*(9), 55–63.
10. Huang, H., Zhu, J., & Zhang, L. (2014). An SDN-based management framework for IoT devices. In *Proceeding 25th IET Irish signats systems cong. China. Ireland international conference on information and communication technology (ISSC, CIICT)* (pp. 175–179).
11. Hakiri, A., Berthou, P., Gokhale, A., & Abdellatif, S. (2015). Publish/subscribe-enabled software defined networking for efficient and scalable IoT communications. *IEEE Communications Magazine, 53*(9), 48–54.
12. Gharbaoui, M., Martini, B., Adami, D., Antichi, G., Giordano, S., & Castoldi, P. (2014). On virtualization-aware traffic engineering in OpenFlow Data Centers networks. In *Proceeding of the IEEE network operations and management symposium (NOMS)* (pp. 1–8).
13. Salman, T., & Jain, R. (2013). *Networking protocols for internet of things* (pp. 1–28).
14. Li, D., Shang, Y., He, W., & Chen, C. (2014). EXR: Greening data center network with software defined exclusive routing. *IEEE Transactions on Computers, 64*(9), 2534–2544.
15. Li, J., Altman, E., & Touati, C. (2015). A general SDN-based IoT framework with NVF implementation. *ZTE Communications, 13*, 342–345.
16. Iyer, A., Kumar, P., & Mann, V. (2014). Avalanche: Data center multicast using software defined networking. In *Proceeding of international conference on communication systems and networks (COMSNETS)* (pp. 1–8).
17. Mann, V., Vishnoi, A., Kannan, K., & Kalyanaraman, S. (2012). CrossRoads: Seamless VM mobility across data centers through software defined networking. In *Proceeding of network operations and management symposium (NOMS)* (pp. 88–96).
18. Qin, Z., Iannario, L., Giannelli, C., Bellavista, P., Denker, G., & Venkata, N. (2014). MINA: A reflective middleware for managing dynamic multinetwork environments. In *Proceeding of network operations and management symposium (NOMS)* (pp. 1–9). IEEE.
19. Wu, D., Arkhipov, D. I., Asmare, E., Qin, Z., & McCann, J. A. (2015). Ubi flow: Mobility management in urban scale software defined IoT. In *IEEE conference on computer communications (INFOCOM)* (pp. 208–216).
20. Diaz-Montes, J., Diaz-Granados, M., Zou, M., Tao, S., & Parashar, M. (2015). Supporting data-intensive workflows in software-defined federated multi-clouds. *IEEE Transactions on Cloud Computing*. https://doi.org/10.1109/TCC.2015.2481410

Parallel Collaborative Normalization of Security Events for Mobile Agent SIEM Systems

Nabil Moukafih, Ghizlane Orhanou, and Said El Hajji

Abstract Current Security Information and Event Management (SIEM) systems use traditional information retrieval methods that produce heavy processing power, and then allocate their local resources to normalize collected logs for advanced correlation. Recently, a Mobile Agent-based SIEM was introduced that uses mobile agents for event collection and normalization at the source device in near real time. In this work, we introduce parallelization to MA-SIEM by comparing two approaches of normalization, the first approach is the multi-thread approach that is used by current SIEM systems, and the second one is the multi-agent approach that is used by MA-SIEM. Our results showed a promising performance gain if we combine both normalization approaches.

Keywords SIEM · Normalization · Mobile agents · Multi-threading

1 Introduction

In today's computer network, many types of equipment are in place to ensure the security and availability of systems and data. The goal is to cope with the number and complexity of cyberattacks that are dramatically increasing each year. Verizon showed in their report that 76% of breaches were financially motivated [1]. This leads to tailored attacks that can compromise systems often in minutes or even seconds, which creates the need for a tool capable of detecting and stopping breaches in real-time.

SIEM systems are example of the tools that can receive and analyze events from different security devices (IDS, firewalls, etc.) in real time [2]. SIEM systems use log files as a major source of relevant information when trying to detect attacks executed against a network. A longstanding problem for SIEM tools is to normalize

N. Moukafih (✉) · G. Orhanou · S. El Hajji
Laboratory of Mathematics, Computing and Applications - Information Security, Faculty of Sciences, Mohammed V University in Rabat, Rabat, Morocco
e-mail: nabil_moukafih@um5.ac.ma; orhanou@fsr.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_43

log data before the analysis and correlation process, this problem comes from the following reasons: (1) Each security equipment generates log data in many format and structures, which adds the effort of transforming each generated log file into a common representation. (2) Sensors in networks produce a huge amount of log data. Gartner defines a network with a small deployment (300 event sources or fewer) can generate a sustained Event per Second (EPS) rate of 1500 EPS or less [3].

In this context, the normalization issue was addressed by proposing a new approach to event normalization, based on mobile agents, that externalizes this process and executes it on the source device. Compared to traditional SIEM systems, in which the normalization phase is performed locally using the server's resources, the Mobile Agent-based SIEM architecture (MA-SIEM) uses mobile agents for real-time event collection and normalization on the source device. MA-SIEM has the following advantages: (1) dedicated mobile agent normalizers that reduce the bandwidth of the network. (2) The externalization of event normalization improves the performance of the SIEM system by leaving its resources primarily for event correlation. More details are given in the paper [4].

This work presents the following main findings:

1. Highlighting the limitations or challenges of normalization in MA-SIEM and SIEM in a network containing a lot of log data to be normalized.
2. Introducing parallel normalization in MA-SIEM by comparing two approaches, the first is often used locally by SIEM systems and the second is based on a multi-agent system.
3. A hybrid model is introduced that shows a promising performance gain in normalization for MA-SIEM systems.

This chapter is organized as follows: Sect. 2 examines the related work to event normalization in current SIEM systems. The third part highlights the main difference between the problem of normalization in SIEM and MA-SIEM. A comparison of two normalization approaches is performed in Sect. 4 and a new hybrid approach is presented. In the last section, we give our conclusions.

2 Related Work

Research on event normalization in security systems is divided mainly into two categories. The first category includes works on developing a universal security log format. The goal is to facilitate analysis in SIEM systems and invite vendors to agree on a format and a syntax of security event records. For instance, authors in [5] proposed an extension of the Common Event Expression or CEE by introducing new structures. The new format OLF or Object Log Format was implemented later in the SAL (Security Analytics Lab) platform [6] and used to detect attacks using simple SQL queries. An overview of some log formats used by SIEM vendors can be found in [7].

SIEM vendors can be found in [7]. This chapter falls into the second category that attempts to improve the performance of SIEM systems. The goal is to optimize the normalization process within SIEM systems so that they can keep up with the constantly growing events of the network. In this context, a hierarchical knowledge base for event normalization was proposed in [8]. The aim of the work was to provide a hierarchy of sub-log formats which offer multiple fields to store event details instead of describing the event type in its entirety, the authors reached a normalization speed of 37,000 events/s. The work in [9] used a different normalization pattern called disruptor pattern which is supposed to be locking free. The authors found that the disruptor pattern is 20% better than using blocking queues for exchanging event data through the distributors. The scalability of the disruptor pattern was tested in [10] by the same team in which they compared it to some existing distribution frameworks.

Finally, the authors in [4] proposed a new approach of normalization which completely externalizes the collection and normalization and executes it on the source device. This approach could potentially free up SIEM resources to do better correlation and real-time analysis of the collected normalized logs, perhaps leading to better decision-making. In this chapter, we will continue the investigation on the benefits of using mobile agent in SIEM by introducing parallel normalization in MA-SIEM.

3 Collaborative Normalization

Normalization is the process of transforming all the formats of the collected events into a single format usable by the SIEM system, it uses a technique called parsing which analyzes and scans the content of the log message to extract using Regular Expressions. In the next subsection, we will highlight some of the challenges that SIEM and MA-SIEM systems face when normalizing events.

3.1 Traditional SIEM

One of the biggest challenges in analyzing generated event logs is the huge amount of log data produced by devices in corporate networks and governments. This is closely related to the device type, available resources, logging level, security policies, and the location of the device on the network. Therefore, we can have two identical servers with two very different EPS values.

Thus, organizations must follow a methodology for implementing SIEM systems taking into account the number of events generated per day in the network [11]. Common sense tells us that we should implement a SIEM capable of handling as many events as any network device could produce simultaneously as a result of a security incident as shown in Fig. 1. But as indicated in the SIEM Benchmarking

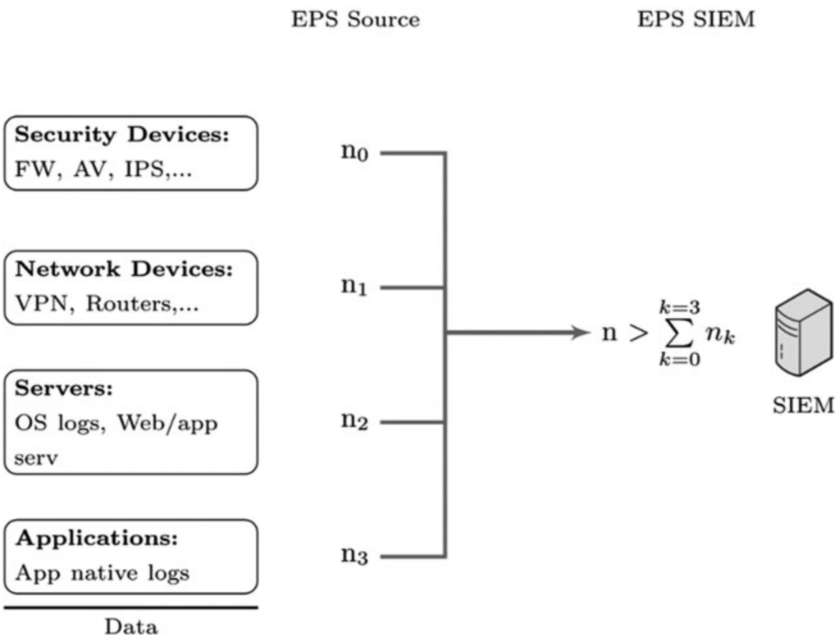


Fig. 1 Normalization in a traditional SIEM system

Report provided by SANS, this scenario is not practical, or necessary because there is no realistic scenario that would imply that all devices will send a maximum EPS [12]. In addition, many events at a time will create network bottlenecks and render SIEM collectors useless.

So, parallelization is one of the most effective methods for improving event normalization, because it helps SIEM systems handle security events generated on parallel. The goal of this section is to give a brief overview of the overall normalization process in SIEM systems and highlight where parallelization can be used.

As depicted in Fig. 2, the normalization workflow can be broken down into three steps:

1. *Sources*: one or more servers produce logs collected at a central location by the Log Receiver component.
2. In reception, the received log data is divided into separate events, which are then passed to the *Normalization* component.
3. In the *Normalization* component, several workers then process the logs and normalize them in a standard format. After normalization, all logs are forwarded for further processing.

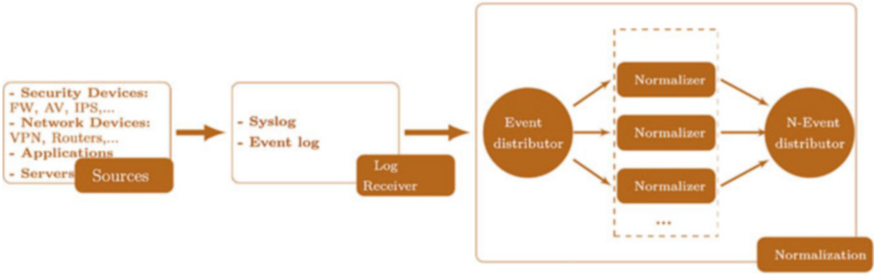


Fig. 2 Overview of the normalization process for event logs

It is clear that the event distributors are two points or bottlenecks that limit the overall performance in this workflow. We can usually solve this problem by adding a few extra workers to handle the receipt of the logs because there is no heavy processing to perform. Normalization, on the other hand, is very intensive in terms of processing and requires a lot of resources. It is therefore based on multi-threading. It also relies on the exchange of events through distributors.

Finally, the multi-threading approach is often run locally on the SIEM server and uses its own resources. The previous study in [4] suggests that the externalization of the normalization process will optimize the performance of SIEM systems and leave its resources for advanced analysis modules like correlation, leading to better decision-making.

3.2 MA-SIEM

The issue with MA-SIEM system is that it uses normalization agents with limited resources (single thread agents) and doesn't scale to the number of events generated on the network [4].

Interestingly, MA-SIEM's different philosophy of event normalization introduces also a slightly different problem for parallelization: If current SIEM servers should consider the number of events generated per day in the whole network in order to provide approximately the same or equivalent processing capacity, MA-SIEM's mobile agents must take into account the EPS value of only the source device and not the entire network as depicted in Fig. 3. That is, when a mobile agent visits a firewall to normalize its generated events, for example, it must have in its code the optimization techniques necessary to handle these events at an optimal speed. Thus, the mobile agent will not drop any event if the firewall log generation is too high.

The challenge for MA-SIEM systems is therefore when a normalizer agent is sent to normalize the events of a particular sensor, the processing capability of the agent must be close to that of the device ($\alpha \approx n_i$). Generally speaking, for a multi-agent

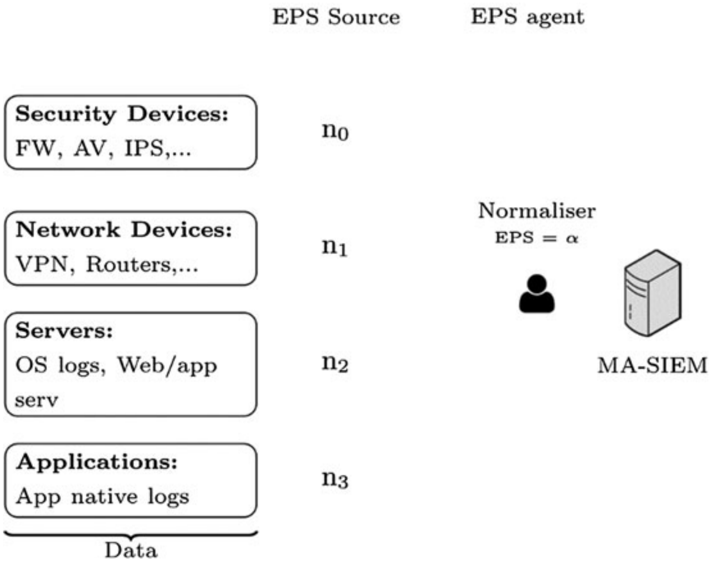


Fig. 3 Normalization in MA-SIEM system

system, agents can work together to solve a complex problem. This can be done in a three-stage activity:

- Problem decomposition: In this stage, the overall problem is divided hierarchically into smaller, simpler sub-problems until individual agents have the appropriate granularity to resolve these sub-problems.
- Sub-problem solution: When the individual agents solve the identified sub-problems, agent share information with other agents because this information may be useful for other agents.
- Solution synthesis: The goal of this phase is to combine all the individual/partial solutions (with the same hierarchy used for decomposition) for an overall solution.

More details of this can be found in [13]. Interestingly, we can use these notions to use a similar multi-agent architecture that performs parallel normalization. This architecture is introduced in the next subsection.

3.3 Externalizing Parallel Normalization in MA-SIEM

It is time to define the normalization models that will be studied in this article.

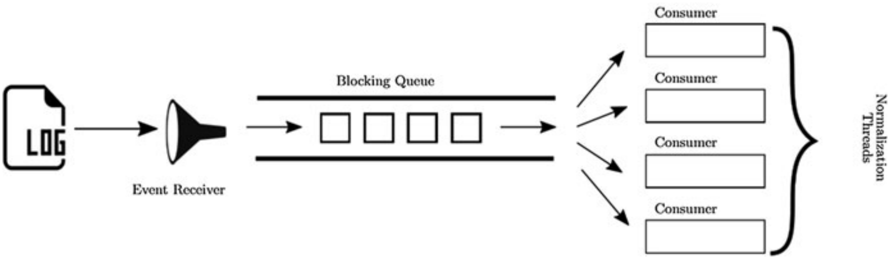


Fig. 4 Event processing approach using blocking queue

3.3.1 Multi-Threaded Blocking Queues

The first is blocking queues, a typical programming model that resolves event exchange through distributors and provides parallel processing, as shown in Fig. 4.

A blocking queue is an implementation of one of the classic models of parallel processing: the producer/consumer model. The pattern uses one or more "producers" that produce information or tasks and one or more "consumers" that consume the information. Both producers and consumers should be able to work simultaneously (hence, parallel processing).

The blocking queue uses a queue that blocks when a consumer thread tries to *dequeue* from it and the queue is empty, and blocks a producer thread if it tries to *enqueue* items and the queue is already full. The consumer thread is blocked until some other producer thread inserts an information/item into the queue, and vice versa, e.g., a producer thread is blocked until some other consumer thread makes space in the queue by consumer some information [14]. The difference now in this chapter, is that the normalization pattern will be on the source device instead of using it locally in the SIEM server.

3.3.2 Multi-Agent Systems

Using the concepts of problems decomposition, sub-problem solution and solution synthesis defined earlier. We propose the following multi-agent architecture as depicted in Fig. 5.

The distributer agent sends the generated events to the normalizer agents. Each normalizer agent will independently normalize the requested events using its knowledge base and send the result to a new distributer/writer agent for further processing.

The main difference between the multi-agent model (Fig. 5) and the blocking queue model (Fig. 4) is how the producer component receives information: All threads in the multi-thread model have a shared memory thanks to the queue between the producer and the consumer component, whereas in the multi-agent systems, each consumer agent is independent from one another and has a direct

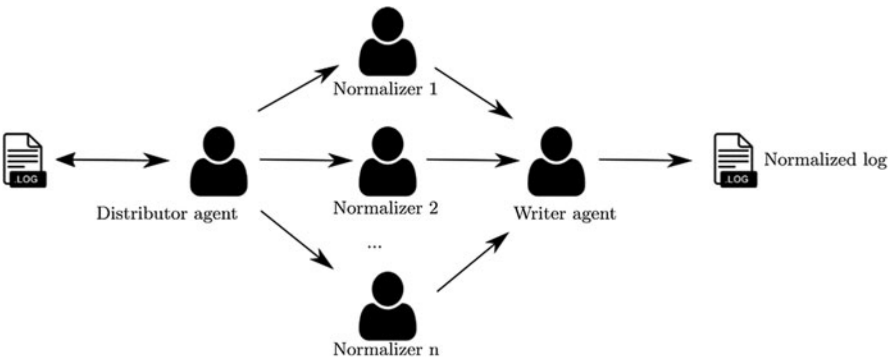


Fig. 5 Multi-agent system for parallel normalization

access to the consumer. The goal of the next section is to study the performance of the two normalization approaches in order to decide which approach will be used in MA-SIEM system.

4 Implementation and Results

Both models were implemented using Java Agent DEvelopment Framework or JADE. JADE is a software framework that simplifies the implementation of multi-agent systems through a middle-ware that complies with the FIPA specifications [15], it also offers a set of graphical tools that support the debugging and deployment phases. The framework is developed in Java.

The data used for implementation is from the "Scan of the Month" Honeynet Challenges: "Scan 31 - Discover how an OpenProxy is abused." The Honeynet Project aims to help security specialists sharpen their forensics and attack detection skills by sharing real attack data collected from honey nets from all over the world [16]. The HN31 contains over 450 k log lines collected from a specially configured Apache web server, specifically designed to be used as a honeypot open proxy server or ProxyPot. Table 1 provides more details regarding the logs contained within the HN34 challenge.

To make the task more complex, all log files were combined into a single file and shuffled randomly. The goal is to make the normalizer component perform a searching algorithm for the appropriate normalization rule for each log line. This is actually the reason why normalization is one of the most processing intensive tasks in SIEM systems. Details of the hardware and software used in the implementation are given below:

– *Hardware*: Dell R820 Server, CPU Intel Xeon E5-2620/2 GHz (6 cores), 16GB RAM. Running VirtualBox 6.0.14 (hosting debian 9.0).
– Software: JavaSE-12, JADE 4.5.0.

Table 1 Log files present in the HN31 challenge

Service	Log file	# of logs
HTTP sever	access_log	202,145
	error_log	696,442
	audit_log	3,643,818
	ssl-access_log	2313
	ssl-error_log	5236
	ssl-enging_log	9910
	ssl-request_log	2313
All	All	4,562,177

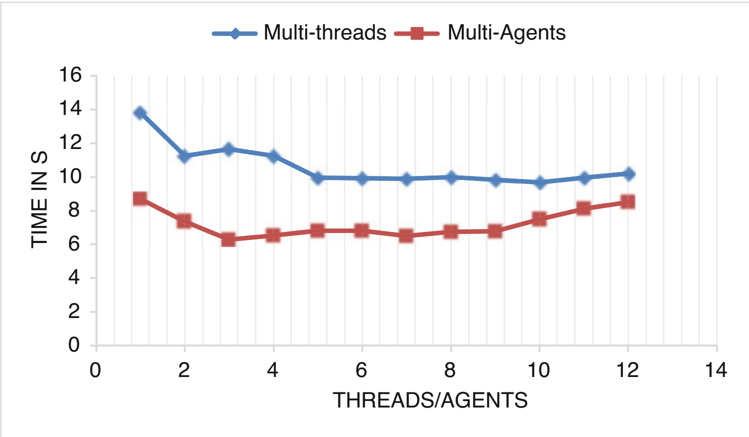


Fig. 6 Multi-threaded vs. Multi-Agent normalization

– Java Libraries: named-regexp, misc. add-on (jade.misc package).

Figure 6 shows the results of our implementation. Note that for the multi-threaded approach, we actually used two queues: The first queue is between the Event Distributor and the Normalizers, where the Event Distributor puts the events to normalize. When the Normalizers process the events, they put the results on the second queue so that N-Event Distributor retrieves the results and puts them on a single file. The goal is to avoid using multiple threads to write a single file which might lead to a resource contention situation.

As for the multi-agent approach, we have deployed single threaded agents to work in parallel in order to generate events. Unlike the multi-threaded approach, communication between the Producer Agent, Normalizers and the Writer Agents is direct and uses ACL messages (Agent Communication Language). ACL is a message format specified by the FIPA [15] international standard for agent interoperability to facilitate communication between JADE agents.

From the graph above we can see that the multi-threaded approach is faster than the multi-agent approach. Although this was initially expected, the results were

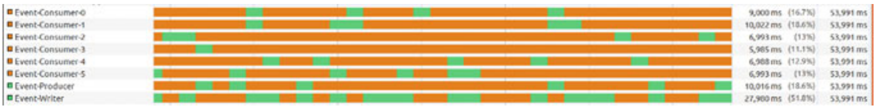


Fig. 7 Locking during normalization with multiple threads (green—thread running, red—thread blocking)

interesting, because they show an interesting scenario: The more threads the multi-threaded approach uses, the longer the model takes to normalize events. On the other hand, the multi-agent approach performs slightly better when we add more agents to the model. Note that each test was run ten times per agent/thread, so that a mean runtime could be calculated.

Ideally, the multi-threaded approach should show an improvement in performance when more threads are involved, but this is only true when the consumers/normalizers are "fast enough" to keep the queue basically empty, because then the producer/Event Producer can go full speed. However, if the normalizers run at maximum capacity and still cannot hold up with the queue, they will be then blocked instead of risking a memory overflow. This is clearly shown in Fig. 7 where most of the time normalizers (consumers) are blocked. As we mentioned before, this is because they have to perform a searching algorithm for the normalization rule before normalizing events.

The multi-agent system is a lot slower when compared to the multi-threaded approach, this is due to the use of ACL messages to communicate. More precisely, we noticed that Jade uses a special queue for messages to be delivered. When the size of this queue exceeds a predefined limit, the MessageManager component (Using Jade's terminology) begins to slow down the agents sending messages so that its internal Thread Pool can flush the queue. Unfortunately, this causes agents to become progressively very slow.

Although the previous argument is related to mobile agent platforms, we believe that it is still possible to achieve a lock-free implementation if we combine both approaches, or at least we can achieve a normalization model that doesn't get slower even if it inherits more resources after reaching maximum capacity. This can be possible if we use multi-agent normalizers that use multi-threads, the threads will help agents normalize events faster, and all agents will work parallel to handle all the source devices events. The promising performance gain can be achieved in the architecture in Fig. 8.

The architecture described in Fig. 8 proposes to use a combined approach which overcomes the performance limitation cause by ACL messages in multi-agent systems by using temporary clusters to store results instead of communicating directly. Also, each agent will use an optimal number of threads (in our case 4) to improve the overall performance.

Figure 9 shows that the combined approach normalized data in 6 seconds which is equivalent to the optimal speed of the blocking queue approach, however and as

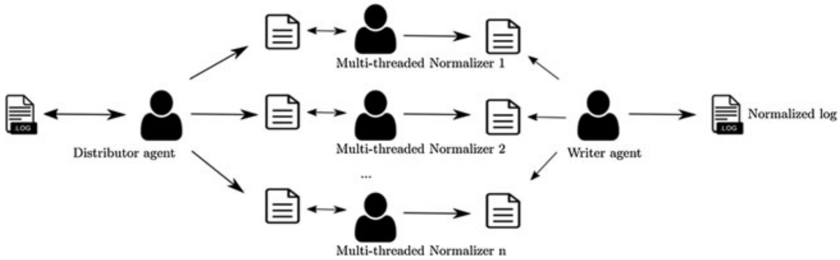


Fig. 8 Compromise approach that uses multi-agent and multi-threads

```
INFO: --------------------------------------
Agent container Container Firewall@localhost is ready.
--------------------------------------------
All 2 agents are there!
Agent Producer@localhost:1099/JADE terminated.
Creating cluster: tmp_Consumer1.csv
Creating cluster: tmp_Consumer0.csv
All 4 threads have finished.
 Stopping..
Consumer1 finished processing. Quitting..
Collector-agent Consumer1@localhost:1099/JADE terminated.
Deleting the file
Recieved Counter number 0
Consumer0 finished processing. Quitting..
All 4 threads have finished.
 Stopping..
Collector-agent Consumer0@localhost:1099/JADE terminated.
Deleting the file
Recieved Counter number 1
All 2 finished. Concatinating..
Creating cluster: access_log-norm0.csv
Completed in 6.282116654 s
Deleting..
Writer agent Writer@localhost:1099/JADE terminated.
```

Fig. 9 Implementation results of the combined approach

we have seen from previous results, the multi-agent approach doesn't proportionally reach bottlenecks when more agents are involved. Future work will concentrate on this.

5 Conclusion

The aim of this research is optimizing the normalization process within SIEM systems as it allows such tools to operate on structured and parsed formats. In this context, MA-SIEM was proposed in order to optimize current SIEM systems by executing normalization of events on the source device instead of executing it locally using the server resources. This is important, because this approach leaves SIEM systems dedicated for advanced and correlation which can lead to better decision-making.

In this chapter, we extended the work by introducing parallelization to MA-SIEM. To do this, we highlighted some of the challenges of normalization in current SIEM systems and presented new ones that should MA-SIEM systems consider. Next, we selected two approaches of parallelization in both cases and compared them: The multi-threaded normalization that's often executed on the central SIEM node, and the multi-agent SIEM that is deployed by MA-SIEM. Our results suggest that it is possible to achieve a lock-free implementation if both approaches are adequately combined. We are currently in the process of investigating this subject.

References

1. Verizonent. (2018). *2018 Data Breach Investigations Report* (p. 8). https://www.verizonenterprise.com
2. Miller, D. (2011). *Security information and event management (SIEM) implementation*. McGraw-Hill.
3. Kavanagh, K. M., & Bussa, T. (2017). *Magic quadrant for security information and event management*. Gartner, Technical Report.
4. Moukafih, N., Orhanou, G., & Elhajji, S. (2020). Mobile agent-based SIEM for event collection and normalization externalization. *Information and Computer Security, 28*(1), 15–34.
5. Sapegin, A., Jaeger, D., Azodi, A., Gawron, M., Cheng, F., & Meinel, C. (2013). Hierarchical object log format for normalisation of security events. In *2013 9th international conference on information assurance and security (IAS)*.
6. Azodi, A., Jaeger, D., Cheng, F., & Meinel, C. (2013). Pushing the limits in event normalisation to improve attack detection in IDS/SIEM systems. In *2013 international conference on advanced cloud and big data*. IEEE.
7. Moukafih, N., Sabir, S., Lakbabi, A., & Orhanou, G. (2017). SIEM selection criteria for an efficient contextual security. In *2017 international symposium on networks, computers and communications (ISNCC)*.
8. Jaeger, D., Azodi, A., Cheng, F., & Meinel, C. (2015). Normalizing security events with a hierarchical knowledge base. In *Information security theory and practice* (pp. 237–248).
9. Jaeger, D., Sapegin, A., Ussath, M., Cheng, F., & Meinel, C. (2015). Parallel and distributed normalization of security events for instant attack analysis. In *2015 IEEE 34th international performance computing and communications conference (IPCCC)* (pp. 1–8). IEEE.
10. Jaeger, D., Cheng, F., & Meinel, C. (2018). Accelerating event processing for security analytics on a distributed in-memory platform. In *2018 IEEE 16th international conference dependable, Autonomic and Secure Computing, 16th international conference on pervasive intelligence and computing, 4th international conference on big data intelligence and computing and cyber science and technology congress (DASC/PiCom/DataCom/CyberSciTech)*.
11. Brad, H. (2012). *Estimating log generation for security information event and log management*. SolarWinds, Technical Report.
12. Michael, B. J. (2009). *Benchmarking security information event management (SIEM)*. SANS Institute InfoSec Reading Room.
13. Wooldridge, M. J. (2009). *An introduction to multiagent systems* (2nd ed.). Wiley.
14. Martin, M. (2015). *Java concurrency essentials*.
15. Bellifemine, F., Caire, G., & Greenwood, D. (2007). *Developing multi-agent systems with JADE*. Wiley.
16. Honeynet Project. (2005). *Scan of the month 31*. http://old.honeynet.org/scans/scan31/

An Experimental Analysis of Reward Functions for Adaptive Traffic Signal Control System

Abu Rafe Md Jamil, Kishan Kumar Ganguly, and Naushin Nower

Abstract Adaptive traffic signal control (ATSC) can play an effective role to reduce traffic congestion which becomes a serious issue in any urban area. The ATSC aims to optimize the traffic flow by dynamically changing traffic signals based on real-time traffic data. Recently, deep reinforcement learning (DRL) has shown its potential in solving the timing of traffic lights to minimize the overall travel time of the vehicles. However, the performance of DRL significantly depends on the proper designing of the reward function. In this chapter, a rigorous experimental analysis among different single and multiple objective reward functions is performed. It has shown that single objective reward functions perform better than multiple objective reward functions. A new single objective reward function is designed that can improve the overall traffic flow. Based on the analysis, it can be concluded that delay, queue length, and travel time are suitable candidates for the reward function.

Keywords Traffic signal control · Deep reinforcement learning · Adaptive system · Reward functions

1 Introduction

Nowadays, traffic congestion is increasingly aggravating in almost every urban area, and it affects people's daily life, fuel consumption, and time and causes air pollution. According to the 2015 Urban Mobility Scoreboard, in the USA, three billion gallons of fuel are wasted for traffic congestion and travelers get stuck in their cars for almost seven billion extra hours (forty-two hours per traveler) [1]. Thus, traffic congestion becomes a serious problem in many cities because of the growing population, increasing urbanization, traffic mismanagement, inadequate traffic infrastructure, and inefficient traffic signal control. However, it is quite

A. R. Md Jamil (✉) · K. K. Ganguly · N. Nower
Institute of Information Technology, University of Dhaka, Dhaka, Bangladesh
e-mail: bsse0722@iit.du.ac.bd

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_44

impossible to stop urbanization, as a result, more public transportation is needed to facilitate the increasing population. Furthermore, improving traffic infrastructure is highly costly. The easier way to reduce cost is to manage the traffic signal efficiently. The reduction of 1% traffic congestion using efficient traffic signal can save billions in a year [2]. An Intelligent Transportation System (ITS) aims to improve the performance of the traffic light control systems by taking real-time traffic information as an input and adjusting the duration of traffic light dynamically.

Most of the existing traffic signal systems are controlled by a predefined fixed time plan without considering the real traffic scenario [3, 4]. In this approach, traffic signal timing is predefined based on historical data. However, traffic flow may change from time to time. Therefore, the fixed time traffic signal control cannot efficiently manage fluctuating traffic demands. Recently, adaptive traffic signal control (ATSC) attracts the researcher's attention because it can manage and control road traffic more intelligently by real-time traffic information (queue length of lanes, waiting time of the vehicles, etc.) and can adjust the traffic light accordingly [5]. However, setting the optimal traffic signal is a complex task, and the success of ATSC depends on the appropriate setting of traffic signal [6]. Reinforcement learning (RL) has been considered as a well-suited learning method for traffic signal control applications and can be used for setting optimal traffic signals.

RL [7] as a machine learning technique for ATSC has shown promising results in recent studies [8–10]. RL does not need to know the full knowledge of the environment. Rather, it interacts with the environment in a sampling fashion to gain knowledge about the environment. The learning agent observes the environment and takes an action. According to the action, the environment changes and the agent gets a scalar reward [11]. The agent learns a policy from taken action according to the observed environment so that it can get the maximum reward. However, traditional reinforcement learning is infeasible to use in traffic signal control because of the two key challenges: (1) how to represent the environment and (2) how to define the correlation between environment and action [12]. The state and action space grows exponentially in real-world traffic scenario, so it becomes infeasible to maintain a state–action pair in the tabular method. To deal with this problem, recent studies [10, 13] used deep reinforcement learning techniques, such as Deep Q-Learning (DQN) [14]. A deep neural network is used as a function approximation technique to predict action for the environment.

In the reinforcement learning technique, one of the most important tasks is to design the reward function [6]. It is the return value from the environment after taking an action. Whenever the agent takes the right action, then it gets a high reward in return. The agent's learned policy always tries to maximize the reward value. If the reward function is not designed properly, the policy can never converge to an optimal point. However, there is no precise rule to choose the correct reward function in RL [6]. Many different objectives can be considered to design a reward function in RL for ATSC. In this chapter, we proposed a reward function that is suitable for real-life traffic configurations. We also provide a comparative study and experimental justifications for choosing the appropriate reward function for different traffic scenarios based on its objective.

The rest of this chapter is organized as follows. Section 2 provides related work regarding reward designing in the area of adaptive traffic signal control. The details of the system design are described in Sect. 3. The experiment setup and result analysis are described in Sect. 4. The whole chapter is concluded in Sect. 5.

2 Related Work

In the literature, different types of reward functions are designed for adaptive traffic control system. In this section, we review some widely used existing reward functions in deep reinforcement learning for ATSC. The existing reward functions can be divided into two categories: (1) single objective and (2) multiple objectives.

2.1 Single Objective Reward Function

In this approach, the reward function is designed considering one traffic parameter. The approach seeks to optimize the parameter to get optimum output. Most of the existing works are single objective [13, 15–18]. The difference of queue lengths between opposite directions (north-south/south-north direction and east–west/west–east direction) is considered as a reward (Eq. 1) in [13].

$$r_t = |\max_{i=1,2}\{q_t^{e-w,i}, q_t^{w-e,i}\} - \max_{i=1,2}\{q_t^{s-n,i}, q_t^{n-s,i}\}| \tag{1}$$

where i indicates the lane number, e–w is east to west direction, w–e is west to east direction, n–s is north to south direction, s–n is south to north direction, and q_t is the queue length at time t.
With the reward function in Eq. 1, the learning process maximizes the difference of queue length between WE and NS. If the difference is high, the agent gets a high reward value, but it does not represent a very smooth traffic scenario. Therefore, the reward function gives wrong feedback to the agent.

The difference of cumulative delays (Eq. 2) is considered as a reward function in [15, 16]. Delay is defined in Eq. 3.

$$r_t = D_{t-1} - D_t \tag{2}$$

Here, D_{t-1} is the cumulative delay at time $t-1$ and D_t is the cumulative delay at time t.

$$D = 1 - \frac{average\ speed\ of\ vehicles\ in\ lanes}{maximum\ allowed\ lane\ speed} \tag{3}$$

With this reward function, the agent tends to maximize the speed ratio of vehicles, which can represent the smooth traffic flow in some cases.

The difference of cumulative waiting time of vehicles between before taking the action and after the action (Eq. 4) is considered as a reward function in [17, 18].

$$r_t = W_{t-1} - W_t \tag{4}$$

where W_{t-1} is the waiting time of vehicles at time $t-1$ and W_t is the waiting time at time t. With this reward function, the agent tries to minimize the waiting time for vehicles. However, only waiting time of vehicles cannot provide an appropriate measure for efficient traffic flow.

2.2 Multiple Objectives Reward Function

Multi-objective is an emerging research theme in RL for ATSC that has received some attention in recent years [6, 10, 12, 19]. This approach seeks to optimize multiple parameters in the network at a time. It is assumed that multiple parameters may lead to better solutions and it may require less convergence times.

Pol et al. [10] defined a reward function considering a number of teleports j, a number of action switches c, a number of emergency stops e, a sum of delay d, and a sum of wait time w as parameters.

$$r_t = -0.1c - 0.1\sum_{i=1}^{N} j_i - 0.2\sum_{i=1}^{N} e_i - 0.3\sum_{i=1}^{N} d_i - 0.3\sum_{i=1}^{N} w_i \tag{5}$$

The first three coefficients of the equation do not affect the deep reinforcement learning (DRL) process [20]. For this reason, the feedback returned from the reward function is misleaded.

Wei et al. [12] designed a reward function considering the sum of queue length L, the sum of delay D, the sum of updated waiting time W, the indicator of light switches C, the total number of vehicles passed N, and the total travel time T of the passed vehicles.

$$r_t = -0.25\sum_{i \in l} L_i - 0.25\sum_{i \in l} D_i - 0.25\sum_{i \in l} W_i - 5C + N + T \tag{6}$$

Here, the indicator of light switches C has a very high negative coefficient though it does not affect the DRL process according to [20]. Furthermore, the total travel time T is added within the equation. It indicates that, whenever the travel time increases, the agent will get a high reward. But, the higher value of travel time indicates bad traffic condition. These two factors mislead the feedback with this reward function.

Mannion et al. [6] experimentally reviewed three different RL algorithms with different reward functions. The reward functions were designed considering the difference of average waiting time, the difference of average queue length, and the unweighted sum of the two above reward functions. They concluded that different types of reward functions work better in different traffic scenarios. However, they considered only two parameters where a lot of other parameters can be considered in reward function.

In this chapter, we mainly focus on reward function in deep reinforcement learning for ATSC. We experimentally reviewed six reward functions (four single objective and two multi-objective functions). All of these reward functions are based on information that is easy to obtain in both simulation and real life. Two multi-objective reward functions are taken from [10] and [12]. Among the four single objective reward functions, two of them are taken from existing studies and one (queue length) is modified from [6]. We designed a single objective reward function considering the difference of travel times between action. With the reward function based on travel time, the agent always minimizes the travel time of vehicles. A minimum travel time always can be a good indicator of smooth traffic flow.

3 System Description

In this section, we formulate the traffic signal control as deep reinforcement learning problem by describing the states, action, and reward function.

The traffic signal control problem is considered as an agent-based system. Each agent is responsible for controlling the traffic light of a single intersection. The environment (traffic scenario) is represented as a state to an RL agent and the agent takes action (change the traffic light). The traffic scenario changes after the change of the traffic light, and the reward is calculated with a defined reward function. If the traffic scenario improves, the agent gets a high reward and gets low reward otherwise. The agent saves its experience in the experience replay memory [17].

In RL, defining the state is a crucial aspect for ATSC [6]. Different types of approaches have already proposed in the literature on how the state can be represented to the agent. Queue length [5] and waiting time [9] are considered as a state. Some studies [12] take queue length, waiting time, and the number of vehicles as a state. Some recent studies [15–19] tend to take an image snapshot of the road intersection as a state. We represented the state using three different metrics of traffic features in this research that are (1) queue length of all lanes, (2) waiting time of all vehicles in the lanes, and (3) the number of vehicles in each lane. There are, in total, eight incoming lanes in the intersection. For each lane, we take the aforementioned three metrics as the state for the RL agent; thus, we get twenty-four entries of numeric value as a state.

The action for an RL agent in ATSC is to change the present traffic phase or to stay on the present traffic phase by adjusting the traffic light signal. There are two different traffic phases in our simulation. In one phase, the vehicles in the east and west lanes can cross the intersection. The vehicles in the north and south lanes can cross in other traffic phase. Therefore, the agent's action set is $A = \{0,1\}$. At any time step t, the agent receives a state s_t and takes an action $a_t \in A$. $a_t = 0$ indicates to stay in the current traffic phase and $a_t = 1$ indicates to change the current traffic phase.

Reward function designing is one of the most important issues in RL for ATSC. It is a scalar value that an agent receives as a feedback after taking an action. The feedback determines the effectiveness of taken action. The traffic factors that can be considered to define reward functions are waiting time, delay, queue length, travel time, flow rate, green waves, accident avoidance, a number of emergency stops, speed restriction, fuel conservation, etc. But there is no hard and fast rule to set a reward function for all traffic conditions.

We experimented the traffic simulation training process with six different reward functions. Among them, four functions are single objective and two are multi-objective.

1. Difference of cumulative delays (Eq. 2) [15, 16]
2. Difference of waiting times (Eq. 4) [17, 18]
3. Difference of queue lengths (Eq. 7) (modified from [6])
4. Difference of travel times (Eq. 8) (proposed)
5. Considering five different factors (Equation 5) [10]
6. Considering six different factors (Eq. 6) [12]

$$r_t = Q_{t-1} - Q_t \tag{7}$$

$$r_t = T_{t-1} - T_t \tag{8}$$

where Q_{t-1} is the total queue length of the time step $t - 1$, Q_t is the total queue length of time step t, T_{t-1} is the travel of vehicles at time step $t - 1$, and T_t is the travel time of vehicles at time step t.

To the best of our knowledge, travel time is never used alone for reward function. Travel time is considered in reward function alongside the other five traffic features in [12]. The other reward functions are commonly used in the literature. We propose travel time as a new reward function which can alone represent the smoothness of traffic flow. Therefore, we designed a new reward function considering the difference of travel times between action.

To evaluate and measure the smoothness of traffic flow in a road intersection, the system needs to have some performance metrics. Some widely used performance metrics in RL for ATSC are road intersection's average waiting time, average delay, throughput of the intersection, and average queue length of vehicles [6]. We considered four different metrics to evaluate the performance of RL agent which are (1) average waiting time, (2) average delay, (3) average queue length, and (4) average number of vehicles passed (throughput).

4 Experiment and Results

In this section, we describe the experiment setup for traffic simulation and discuss the result.

4.1 Experiment Setting

We conduct simulation using a well-known open-source simulator Simulation of Urban MObility (SUMO) (Version 1.3.1) [21] for our experiment. The simulation road intersection is a four-way road intersection, and each road has two incoming lanes and two outgoing lanes. The length of each road is 300 m. The incoming lane's maximum speed can be 70 km/h and the outgoing lane's maximum speed can be 40 km/hour. Each vehicle in simulation has a length of 3 meters and the minimum gap between vehicles is 1 meter. There can be four possible routes: (1) W–E: going from west to east, (2) E–W: going from east to west, (3) N–S: going from north to south, and (4) S–N: going from south to north.

The RL agent parameters are summarized in Table 1. We used ϵ-greedy approach to solve the explore–exploit dilemma. We have used the same network structure for all the experiments. The neural network consists of five layers (an input layer, an output layer, and three hidden layers). The input layer consists of 24 neurons as the state is designed with 24 values. The hidden layers consist of 12, 8, and 4 neurons, respectively. The output layer is constructed with two neurons. The values in the output layer indicate the q-values for action (0,1). The sigmoid activation function is used in the hidden layers. The weights of the network are updated using RMSProp optimizer [22].

In our experiment, we used synthetic data with four different traffic conditions which is used in [12]. In the first configuration, simple changing traffic is used with a very high biased ratio. An equal steady traffic demand is used in configuration 2.

Table 1 Agent parameters

Agent parameters	Values
Minimum time interval of action	5
update period of target network	5
Learning rate	0.005
Memory size	1000
Sample size	300
Mini-batch size	20
Training epochs	500
ϵ for exploration	0.05
Discount factor	0.9

Table 2 Traffic configurations

Conf.	Routes	Arrival rate (cars/s)	Start time (s)	End time (s)
1	W-E	0.4	0	3600
	N-S	0.4	3601	7200
2	W-E	0.4	0	7200
	N-S	0.4	0	7200
3	W-E	0.2	0	7200
	N-S	0.4	0	7200
4	Conf. 1		0	7200
	Conf. 2		7201	14,400
	Conf. 3		14,401	21,600

Table 3 Performance metrics of reward functions in configuration 1

Reward functions	Performance metrics (Average values)			
	Delay	Queue length	Waiting time	Throughput
Reward 1	2.1231	9.5454	18.099	0.8050
Reward 2	2.6406	10.8271	2.376	0.8060
Reward 3	2.1442	9.4723	35.4075	0.8059
Reward 4	2.9665	11.6551	2.5889	0.8062
Reward 5	3.8722	12.719	12.7392	0.7961
Reward 6	3.2620	12.194	9.105	0.8050

Unequal but steady traffic demand is used in configuration 3. Configuration 4 is generated by combining the three different abovementioned traffic configurations. Table 2 shows the traffic arrival rates for four different types of configurations.

4.2 Performances on Synthetic Data

The performance metrics of the reward functions are described in this section. For metrics delay, queue length, and waiting time, the lower value is better. For throughput, the higher value is better.

The performance metrics of the experimented reward functions for configuration 1 are described in Table 3. Analyzing the value of the metrics, we can say that no reward function performs significantly better in this configuration. The configuration is designed in such a way that it has two different parts. In one part, there are vehicles for one direction (WE) and there are vehicles in the opposite direction (NS) in another part. The two parts are very biased in traffic flow. For this reason, each reward function provides better value for the metric that it is designed with. For example, the reward designed with cumulative delay provides a better value for delay.

Table 4 Performance metrics of reward functions in configuration 2

Reward functions	Performance metrics (Average values)			
	Delay	Queue length	Waiting time	Throughput
Reward 1	6.467	23.036	157.101	1.5824
Reward 2	6.7685	23.7356	32.9819	1.6114
Reward 3	5.4813	22.77	7.044	1.6159
Reward 4	7.15	24	21.9926	1.615
Reward 5	7.398	23.966	23.012	1.5
Reward 6	7.4423	23.929	25.1974	1.488

Table 5 Performance metrics of reward functions in configuration 3

Reward functions	Performance metrics (Average values)			
	Delay	Queue length	Waiting time	Throughput
Reward 1	4.9109	21.2095	4.2221	1.212
Reward 2	5.4322	21.8871	6.3648	1.21
Reward 3	5.7089	21.734	11.6407	1.1763
Reward 4	5.5587	21.8749	7.2938	1.212
Reward 5	5.7027	21.651	19.936	1.1306
Reward 6	5.7495	21.655	15.4048	1.16

Table 6 Performance metrics of reward functions in configuration 4

Reward functions	Performance metrics (Average values)			
	Delay	Queue length	Waiting time	Throughput
Reward 1	4.8492	18.3577	399.5098	1.192
Reward 2	5.2051	19.3276	11.398	1.205
Reward 3	4.7106	18.0505	398.8789	1.1631
Reward 4	4.3377	18.072	5.0608	1.202
Reward 5	5.87425	19.6665	17.3418	1.127
Reward 6	5.8723	19.6643	27.751	1.15

Table 4 describes the performance metrics for configuration 2. Considering the metric values, we can say that the reward function designed with queue length outperforms the other reward functions. Therefore, it can be concluded that the reward function designed with queue length performs better for equal traffic flow in all directions.

The metric values in Table 5 indicate that the reward designed with cumulative delay outperforms the other reward functions in configuration 3. We can summarize that the reward designed with cumulative delay performs better for unequal steady traffic flow.

The metric values in Table 6 show that the reward designed with travel time outperforms the others in two metrics (delay and waiting time). For the other two metrics (queue length and throughput), the rewards with queue lengths and waiting time provide better values, respectively. However, if we take a close look at the values, we can see that the highest values for the two metrics are not significantly better than reward 4 (travel time) values. Therefore, we can say that if the traffic flow fluctuates, the reward function designed with travel time performs better.

None of the multi-objective reward functions performed well in any configuration. All the traffic features considered in the multi-objective reward functions do not affect the traffic flow [6]. Furthermore, the features may have a correlation among them, which affects the performance of multi-objective reward functions.

5 Conclusion

In this chapter, we have designed a single objective reward function and evaluated its performance. The new reward function is experimentally analyzed alongside with mostly accepted five other reward functions. We have shown that our designed reward function performs better than other functions in certain traffic scenarios. Four different traffic configurations are used for the experiments. From the result analysis, it can conclude that delay, queue length, and travel times are the most prominent candidates for designing reward functions for different traffic scenarios. Consideration of the correlation among the traffic features while designing multi-objective reward functions can be a future research direction.

References

1. Urban Mobility Scorecard. (2015). The Texas A&M Transportation Institute and Inrix, Inc., USA 9.2015:10.
2. The Economist. (2014). The cost of traffic jams. https://www.economist.com/blogs/economist-explains/2014/11/economist-explains-1.
3. Miller, A. J. (1963). Settings for fixed-cycle traffic signals. *Journal of the Operational Research Society, 14*(4), 373–386.
4. Webster, F. V. (1958). Traffic signal settings. Road Research Technical Paper 39.
5. Baher, A., Pringle, R., & Karakoulas, G. J. (2003). Reinforcement learning for true adaptive traffic signal control. *Journal of Transportation Engineering, 129*(3), 278–285.
6. Mannion, P., Duggan, J., & Howley, E. (2016). An experimental review of reinforcement learning algorithms for adaptive traffic signal control. *Autonomic Road Transport Support Systems* (pp. 47–66). Cham: Birkhäuser.
7. Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning* (vol. 2(4)). Cambridge: MIT press.
8. Kuyer, L., Whiteson, S., Bakker, B., & Vlassis, N. (2008). Multiagent reinforcement learning for urban traffic control using coordination graphs. *Machine Learning and Knowledge Discovery in Databases*, 656–671.
9. Wiering, M. A. (2000). Multi-agent reinforcement learning for traffic light control. In *Machine Learning: Proceedings of the Seventeenth International Conference* (ICML'2000).
10. van der Pol, E., & Oliehoek, F. A. (2016). *Coordinated deep reinforcement learners for traffic light control*. NIPS.
11. Glorennec, P. Y. (2000). Reinforcement learning: An overview. In *Proceedings European Symposium on Intelligent Techniques* (ESIT-00), Aachen, Germany.

12. Wei, H., Zheng, G., Yao, H., & Li, Z. (2018). Intellilight: A reinforcement learning approach for intelligent traffic light control. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2496–2505). ACM.
13. Li, L., Lv, Y., & Wang, F.-Y. (2016). Traffic signal timing via deep reinforcement learning. *IEEE/CAA Journal of Automatica Sinica, 3*(3), 247–254.
14. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529533.
15. Genders, W., & Razavi, S. (2016). Using a deep reinforcement learning agent for traffic signal control. Preprint. arXiv:1611.01142.
16. Mousavi, S. S., Schukat, M., & Howley, E. (2017). Traffic light control using deep policy-gradient and value-function-based reinforcement learning. *IET Intelligent Transport Systems, 11*(7), 417–423.
17. Gao, J., et al. (2017). Adaptive traffic signal control: Deep reinforcement learning algorithm with experience replay and target network. Preprint. arXiv:1705.02755.
18. Liang, X., et al. (2018). Deep reinforcement learning for traffic light control in vehicular networks. Preprint. arXiv:1803.11115.
19. Coşkun, M., Baggag, A., & Chawla, S. (2018). Deep reinforcement learning for traffic light optimization. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE.
20. van Dijk, J. (2017). *Recurrent neural networks for reinforcement learning: An investigation of relevant design choices*.
21. Krajzewicz, D., Erdmann, J., Behrisch, M., & Bieker, L. (2012). Recent development and applications of SUMO-simulation of urban mobility. *International Journal On Advances in Systems and Measurements, 5*(3&4).
22. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. Preprint. arXiv:1412.6980.

Smart Contracts: Between the Attractiveness of a Registry Catalog Implementation and Convenience of a Jurisdictional Background

Mohamed Laarabi, Abdelilah Maach, and Abdelhakim Senhaji Hafid

Abstract Smart contracts are a common topic in scientific discussions held by computer scientists. Due to the dynamic nature of its applications, smart contracts must be considerably more adaptable, responsive and controllable. Yet the long-standing focus on security issues has affected negatively the development of smart contracts in the transfer of ownership of the real estate.

In this paper, we study the possibilities and the limitations of how to transfer properties on the block chain using a digital record, as a back-support in case of litigation. We propose storing a registry catalog to handle the property historical data and criminal record of the owners. The execution of the smart contract is enforced using a digital identity of the two parties. We design the components of what the registry should contain to integrate with block chain infrastructure and we also analyze and discuss the government endorsement needed for the success of the implementation.

Keywords Smart contract · Block chain · Registry · Ethereum · Digital identity · Fractional rights · Colored coins

1 Introduction

Currently, when someone needs to buy sell or rent a property, he or she must go through a centralized system. That system depends on all its structure on a third party that can safeguard the accuracy of the data, the safety of the operation and the history of all the transactions that have been made on that property. However,

M. Laarabi (✉) · A. Maach
Computer Engineering Department, Mohammadia School of Engineers, Mohammed V University, Rabat, Morocco
e-mail: maach@emi.ac.ma

A. S. Hafid
Montreal Block Chain Lab, University of Montreal, Montreal, Canada

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_45

this centralized system has shown flaws in the form of longer procedures and huge transaction fees.

In 1994, legal scholar Nick Szabo realized that contracts could be converted to computer code, or in other words, a smart contract. A smart contract is a code that enforces the terms of an agreement between untrusted parties and releases assets once the pre-defined actions have been accomplished.

Compared to traditional contracts, a smart contract does not have a third party resulting in lower transaction fees. This idea of not having third-party involvement did not come to light until the appearance of the block chain network, a system that can self-execute contracts, store and replicate transactions on a decentralized ledger and supervised by a network of computers. The smart contract can be developed in different block chain platforms such as NXT Bitcoin or Ethereum. Each one of these platforms highlights a diverse dimension of coding complexity. In this paper, we will concentrate on the smart contract that is supported by Ethereum because of its coding dialect (turning-completeness) that allows the creation of a more advanced and customized contract [1–5].

Today, the connection between smart contracts and governments is quite dim, that's what motivated us to give a contribution to this field of research. Several organizations are attempting to utilize Ethereum-based smart contracts for their activities. Ex.slock. is a German company that uses smart contracts for its leasing services. However smart contracts are far from being perfect for larger real estate transactions. We now face a critical urge for new radical thinking in order to provide a better solution. The research is primarily contributing with improved knowledge about smart contract technology and its potentials and challenges within the real estate industry. One conclusion from the study is that the government is capable of adapting to this new technology which is why the topic needs to be investigated deeper and to develop this relationship further for a successful implementation. There are still numerous unanswered inquiries, like how to write the code and by who? Is it developers or legal advocates? Should we give developers some jurisdictional background?

1.1 Purpose and Research Questions

This paper aims to look at the smart contract from a technical view and try to shed light on the issues that keep the administrations from embracing them on a larger scale.

In this study, we aim to answer the following research questions:

- RQ1: How to solve the problem of ownership? To answer this question a literature review is made on the prerequisites for storing all the legal documents and former owners of the property on the blockchain. The methodology used for the question is the systematic mapping study.

- RQ2: Designing contracts and writing them are issues that we are endeavoring to explain, which brings us to the question, what smart contract should contain to fulfill the legal agreement between the parties?

1.1.1 Contributions

This paper makes these contributions:

The design, implementation, and evaluation of a registry catalog, an application that uses smart contracts for automated real estate purchases.

A discussion of the legal use of code as law, if code is a contract then code is law.

The remainder of this paper is organized as follows:

In Sect. 2, we describe our dataset and give a brief overview of related work. In Sect. 3, we present and discuss our findings. The last segment of Sect. 3.2 will be devoted to smart contracts in relation to the registry record implementation process. In Sect. 3.3, we discuss and address threats to the validity of our results. Finally, we provide concluding remarks.

2 Related Work

Many researchers have explored the application of smart Contracts in the real estate domain. For example, Alex Mizrahi in his paper entitled, "A Blockchain-based Property Ownership Recording System," smart contracts have been proposed as a mechanism to prevent the ownership of the property from being stolen. Others have investigated the privacy implications of using smart contracts in different transactions, but very few mentions of the potential of a smart contract as a security solution. Furthermore, very few articles have addressed the performance of those contracts, as who will write them, how they will be taxed, or whether they can stand up in court in the case of litigation between the parties. Leaving a huge gap in research that prevents any practical use of smart contracts as a legal document in the near future [6, 7].

Study design:

Adopting systematic mapping, allows us to explore the previous work that has been done on smart contracts. We have divided our research into three categories: security, language design, and privacy.

We will then proceed to the discussion chapter, introducing our main contributions and limitations.

2.1 Security of the Ownership

Starting with the security of the ownership transfer, few articles depicted explicitly ownership transfer. We have found one article that gives a real solution using the bitcoin platform and treating the property like any other coin on the blockchain. Alex Mizrahi [7], suggests to attach the property with a colored bitcoin using the Bitcoin platform with each transaction containing the following:

- An input spending coins from registry's Bitcoin address.
- Data contained in OP_RETURN output.
- A genesis output which assigns colored coin to the property owner.

The advantages of doing so are reduced implementation cost as we can use a system that is already functional. The ability to transfer the colored coins and the bitcoin in one transaction makes it conceivable to do trustless trades.

However, the colored coin may lack effectiveness in the transfer of bigger property such as houses, buildings, or parcels of land because the seller can attach properties to many-colored coins and sell it to many other parties.

Transferring real estate via the Bitcoin platform seems conceivable but not be practical. Should any litigation or a transfer error occur in the transaction, the real estate will face a significant impact on its value and furthermore on the identity of the actual owner.

2.2 The Language Design of a Smart Contract

The language design of a smart contract using the systematic mapping methodology, most of the previous works discuss the option of standardizing the coding language of the smart contract. Clack et al. [8], address in their article how Contracts are written, how they are executed, how they are enforced, and how to ensure that the execution of a contract is faithful to the meaning of the legal documentation.

Following Grigg's Ricardian Contract triple, they define the smart contract as an agreement whose execution is both automatable and enforceable. Automatable by a computer, although some parts may require human input and control. Enforceable by either legal enforcement of rights and obligations or tamper-proof execution. Admitting that a legal smart contract is not automatically executed it's more automatable as it needs the intervention of the human, restricting the human intercession in the code writing. It is important to provide a framework that supports complex legal agreements. The focus was on the definition of a contract dividing it into operational semantics and denotational semantics while differentiating between smart contract codes that are fully automated and smart legal contracts that have a hint of human input. Moreover, they give a profound look at how to move from executable code to a common language that not only can be used in legal cases but can extend to the fields of finance and biology [9–13].

2.3 Privacy

Transactions that occur on the block chain are viewable by all the viewers which can be a problem for those who want their sensitive data to remain private. In his article about methods of implementing privacy on the block chain, Vitalik states that a solution to this problem involves multi-party computation [14], which requires that you trust a group of the participants involved in storing the information [8].

We need to comprehend that block chain, in general, does not solve the issues of privacy. On the contrary, block chain exists to make transactions progressively straightforward and transparent.

3 Discussion

In the first section, we will be talking about the role and the significance of a record implementation as a vital element for the success of a property transaction on the block chain. In the second section, we will be discussing the requirement and the content of a property smart contract. The last segment will be devoted to the smart contract concerning the registry record implementation process.

3.1 Storing Registry Catalog on the Ethereum Block chain

When they tried to change the recording system from physical folders to computer data in the late 1980s and mid-1990s in the United States of America, a huge number of records were lost. Moreover, keeping records internally may be about as secure as keeping it in a public ledger. Sadly, there are numerous cases of security breaching and owners changing either by fraud or human mistakes. Readily accessible records could potentially save time and money for both the government municipality in terms of minimizing salaries and the time of the transaction (Fig. 1).

3.1.1 Owner Identification

In terms of identifying the new owner, we can use a single storage location for all data. The idea is to choose the Key containing the "Demand-of-changing-ownership" sent by the registry to the buyer and, in the meantime, tracking personalized data access permissions [15].

We design the components of what the registry should contain to integrate with blockchain infrastructure (Sect. 3.3). We assume that information available in the blockchain is already trustworthy. Our design introduces keys of identity and

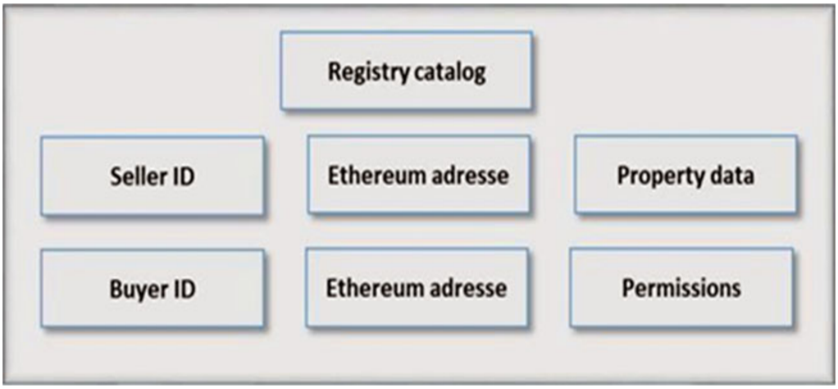


Fig. 1 Registry catalog showing data content for each operation

sufficiently able to incorporate with a SQLite [15]. Any new buyer can partake in the system. All one needs to do is coordinate the ID with the permissions access.

3.1.2 Digital Identity

As we have proposed that registries must be digitized giving these smart contracts a back-support and validation requirement. But even when those are digitized we would still be facing the challenge of verifying the owner's identity.

We have said that the purpose of the registry is to validate assets and tell who has rights to which asset. The smart contract is there to cut the middleman and to support the territory jurisdictional laws, yet the requirement for digital identity is vital for the complete success of the solution we are offering.

Canada is on the lead in this field, with the block chain-based national ID system named secure key. Dubai is additionally building their own identity platform [16]. Having an existing legal-data record for identities can shorten the time for the implementation of smart contracts. Users.

will no longer need to have an identity just for the block chain, and also having a digital identity verified by the state will give provide top-notch data [16].

3.2 Smart Contracts Design and Regulations

The discussion over who should compose the contract has been largely solved in the paper of Nick Szabo [1]. Is it a lawyer or a software engineer? Szabo proposes that the writer must be a legal scholar with some analytic and intelligent reasoning. We concur that the requirement for a legal advisor in the drafting process is essential. The client can go for a hybrid approach having it drafted by a lawyer and then turn it to an executable code by a programmer. Clearly, the draft will enhance over time. To make it work, the client must turn the fundamental components of a contract to a running code. However, some data cannot be acquired from the block chain such as historical information. Hence the registry implementation is some form of guarantee.

The smart contract is, after all, an agreement between parties, states Morris R. Cohen [17]. In the block chain, those parties are unknown to each other in most cases, so the code must be written to perfection and must be executed faithfully.

However, all articles could not solve the issue of whether a smart contract can be acknowledged in court? Whether the language of a contract can be admissible in all cases? To solve this problem a hybrid solution between previous works should be done. We need to select our parameters in order to make the smart contract practical. A language capable of supporting a complex code that contains legal clauses.

- A common definition of what's a smart contract and accepted by all the parties.
- The role of the government to make sure of the validity of the Contract and give legal legitimacy to it.
- The role of the registry catalog in order to identifier the historic and transactions that have been made on the property.

The language of a smart contract must fulfill all the legal obligations of the host country. We recommend that legal smart contracts should be written by lawyers and then turn those clauses into executable code, agreed to by all parties and supported by the local court.

3.3 Smart Contract and Registry Record Process

Contracts call different contracts by message calls. Each time a solidity contract calls an element of another contract by delivering a message call. Each call has a sender, a recipient, a payload, a value, and a measure of gas. Providing a solid local call technique to the address type that functions as pursues: [address.call.gas (gas).value (value) (data)] [18].

The subject that can be a wellspring of disarray is the concept of the non-operational legal interpretation of some parts of the contract in the message call. Contract functions can call other contract functions. But the interpretation of some parts can create a problem and therefore the hybrid solution emerges between using smart contract and a registry to back it up with the help of a legal identity record for individuals and institutions. The processing time can be slow sometimes as the blocks come with a list of transactions to be executed irrespective of their complexity [18].

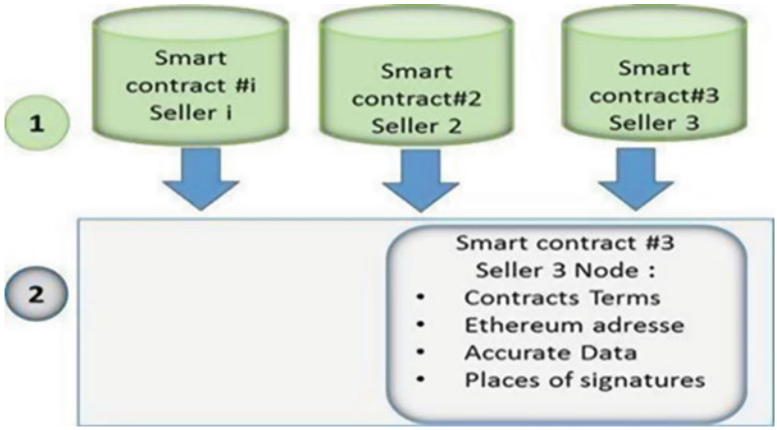


Fig. 2 Smart contract content

3.3.1 Phase 1: Smart Contract

Before the smart contract implementation on the Ethereum, smart contract must contain these components:

- Contracts Terms.
- Ethereum address.
- Accurate Data.
- Places of signature.

As we wrote in the first section, Smart Contract must be written by developers (off-chain phase) that have a solid jurisdictional background, therefore the integration must be flawless, as once the execution of the contract is made there can be no other corrections (Fig. 2).

3.3.2 Phase 2: Registry Workflow

Once the smart contract has been successfully implemented on the system, the RAR (Registry Administration Record) starts to examine independently the existence of the property and the ownership of historical records via public and private keys. It also stores the actual operation if it matches the execution conditions. The RAR will provide a solution for the identity problem since land parcels and buildings can be tied to many parties. The RAR will send a notice to all the parties that are associated with the transaction. Here we must think of developing decentralized blockchain-based identity platforms that allow all the physical participants to have a digital identity. This goal is a little far from reality at present but we can possibly use an existing validated identity system and connected it to the blockchain via hashes (Fig. 3).

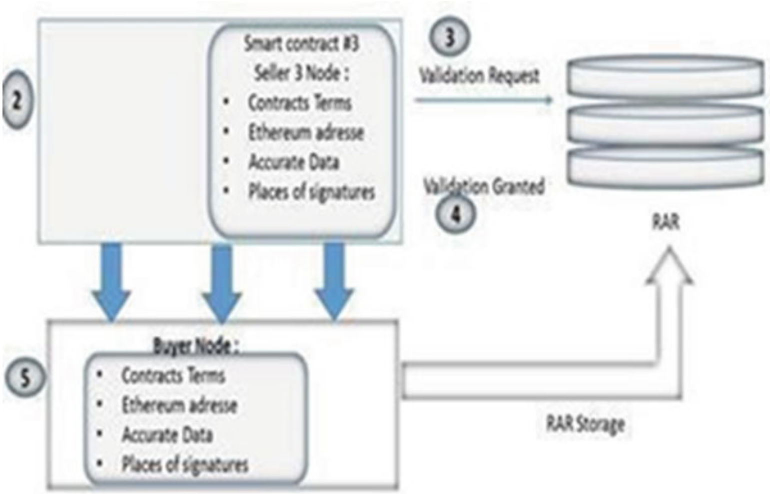


Fig. 3 Registry workflow

Fig. 4 Transaction pool price

3.3.3 Phase 3: Miner Node Accepts the Transaction

The Ethereum network contains two sorts of user's miner nodes and the non-miner nodes. The miners are the ones responsible for working the transaction in the block. Miners maintain a record of all the transactions before evaluating it (Fig. 4).

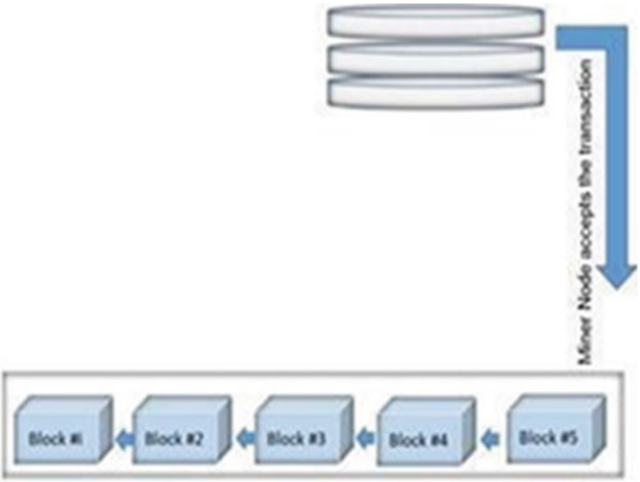


Fig. 5 Miners accepts the transactions after RAR verification

This specificity can create a huge problem, as some transactions will not be executed because of the price. In order to solve this problem, we suggest another application to the registry by making a local pool explicit for the transactions that are approved by the registry. We recommend also that instead of uploading a registry catalog in the form of images we can turn all the previous contracts into a text that will minimize the cost. For example, a page of texting can cost between \$5 and \$7. To absorb the cost, the government must take charge of data storing and include the price as a new tax on each completed transaction (Fig. 5).

3.3.4 Phase 4: Database Physical Record

Having an off-chain record means all the transactions on and off can be audited if needed without the requirement for a public key or reading permission protocol. In addition, we must put into our consideration that not all the people will run towards this new application. We will likely see a resistance to change. So to keep a physical record is to ensure a smoother change process. Also for added security for users, the existing on an off-chain record is vital in illuminating complex situations such as fractional occupancy in other terms the rights of the property that are divided between many parties whether they are the rights of a land, image, or mineral resources. Therefore, we ought to expect the utilization of a physical record until we can all come to common ground (Fig. 6).

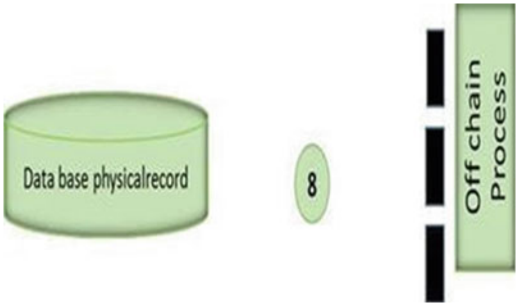


Fig. 6 Off-chain physical record

4 Conclusion

We have presented in this paper the implementation of a decentralized registry catalog. The execution of the smart contract is enforced using a digital identity of the two parties. We design the components of what the registry should contain to integrate with block chain infrastructure. We also explore the role of governments in maintaining trust in those contracts. The limitation we have faced during this exploratory study can be divided into three areas:

- **Digital identity:**
  Consider the possibility that somebody stole the key or the digital identity and tried to use it, therefore the need for a second signature is a must. The bank and registrar can act as a second signer that will provide at least a more secure transaction. Regrettably, using this as a solution will involve a third party which is not what block chain is built for [16].
- Smart contract language:
  The significant challenges are obvious, the need for a better template that fulfills the will of both parties in all transactions; even the complex ones such as fractional rights. Until now, there is not enough research in this area.
- **Government's endorsement:**
  In every new product or a radical change, the confidence provided by the state is crucial for the success of the project. Yet smart contracts confront a colossal opposition from many countries due to the need for a new manual that addresses specifically these types of products in terms of taxes, state rights, heritage, etc.

4.1 Further Work

We have identified several subjects for the future. First, we continue our process in order to provide a new template for smart contracts and make it more flexible by exploring the non-operational parameters of a contract. We are talking about the parts of the contracts that are difficult to put in an electronic representation [9].

Contracts may likely contain more sophisticated characters as we have seen previously. Storing registry data can have a radical change on the business by cutting so much red tape by digitalizing not only contracts but also administrations. Therefore more opportunities for a whole new market are possible. Clearly, we need to answer many more questions on how much a smart contract can really support without neglecting the cost.

The development of smart contract require The development of so many other areas, the block chain itself for example the language also, as most parameters in now templates have simple types; such as date, number; and value [9], beyond simple and complex, parameters can also be expressions, containing references or names.

The urge for trustful data of property is one the fundamental reasons we contend for hybrid approach between off-chain record and a block chain registry. Governments must manage the monetary system and find new ways of how to benefit from flow created by smart contracts.

References

1. Popper, N. (2015). Decoding the Enigma of Satoshi Nakamoto and the Birth of Bitcoin. *The New York Times*, 15 mai 2015 (ISSN 0362-4331, lire en ligne [archive], consulté le 27 février 2016).
2. Buterin, V. A next generation smart contract & decentralized application platform. (2014). On Stake. [Online]. Available: https://blog.ethereum.org/2014/07/05/stake/retreived 19/02/2019.
3. Christidis, K., & Devetsikiotis, M. (2016). Blockchains and smart contracts for the internet of things. In *IEEE access* (Vol. 4, pp. 2292–2303). https://doi.org/10.1109/ACCESS.2016.2566339
4. Wu, C., Zhang, L., Li, Z., Li, Q., & Zhang, Y. (2019). Exploring potential and feasibility of binary code sharing in mobile computing. *IEEE Transactions on Cloud Computing*.
5. Liu, L., Chang, Z., Guo, X., Mao, S., & Ristaniemi, T. (2018). Multiobjective optimization for computation offloading in fog computing. *IEEE Internet of Things Journal, 5*(1), 283–294.
6. Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic mapping studies in software engineering. In *EASE'08 Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering* (pp. 68–77).
7. Bentov, I., Gabizon, A., & Mizrahi, A. (2016, February). Cryptocurrencies without proof of work. In *International conference on financial cryptography and data security* (pp. 142–157). Springer.
8. Clack, C. D., Bakshi, V. A., & Braine, L. (2016). Smart contract templates: Foundations, design landscape and research directions. arXiv preprint arXiv:1608.00771.
9. J. R. Douceur, The Sybil attack, in peer-to-peer systems (lecture notes in computer science). (2002), : Springer, Mar. pp. 251–260. [Online]. Available: http://link.springer.com/chapter/10.1007/3-540-45748-8_24.
10. Grigg, I. (2004). The Ricardian contract. In *Proceedings of the First IEEE International Workshop on Electronic Contracting* (pp. 25–31). IEEE.
11. Jeong, S., & Ahn, B. (2021). A study of application platform for smart contract visualization based blockchain. *The Journal of Supercomputing*. https://doi.org/10.1007/s11227-021-03879-1
12. Grigg, I. (2015). *The sum of all chains — let's converge! Presentation for Coinscrum and Proof of Work*. http://fi-nancialcryptography.com/mt/archives/001556.html

13. Gotcu, M.-L. (2016). *Legal breakthrough for blockchain technology*. Tilburg Law School.
14. Wilkinson, S., Lowry, J., & Boshevski, T. (2014). Metadisk a blockchain-based decentralized file storage application. *Storj Labs Inc., Technical Report, hal*, 1–11.
15. Meier, P., Beinke, J., Fitte, C., Jan Schulte To Brinke, & Teuteberg, F. (2021). Generating design knowledge for blockchain-based access control to personal health records. *Information Systems and e-Business Management., 19*. https://doi.org/10.1007/s10257-020-00476-2
16. Graglia, J. M., & Mellon, C. (2018). Blockchain and property in 2018: At the end of the beginning. *Innovations: Technology, Governance, Globalization, 12*(1–2), 90–116.
17. Cohen, M. R. (1933). The basis of contract. *Harvard Law Review, 46*(4), 553–592. https://doi.org/10.2307/1331491
18. Rahman, A., Hassanain, E., & Hossain, M. S. (2017). Towards a secure mobile edge computing framework for hajj. *IEEE Access, 5*, 11768–11781.

Conceptual Denoising Auto-Encoder with Long Short-Term Memory for Remaining Useful Life Estimation in Predictive Maintenance

Paul Menounga Mbilong, Fatima-Zhara Belouadha, and Mohammed Issam Kabbaj

Abstract Most manufacturers using industrial machines or equipment that must operate 24 h a day are faced with the fear of a failure that would immobilize all their activity and result in huge maintenance costs. To anticipate these possible failures, conditioned based maintenance (CBM) or predictive maintenance (PDM) is presented as a solution; it consists of setting up maintenance policies that are based on the condition of the machine, unlike periodic maintenance. One of the sub-areas of predictive maintenance is the prediction of the remaining useful life of a machine (RUL) before it fails. Accurately estimating the RUL is a continuous challenge for every company in the industrial sector, so it is essential to find approaches that allow us to be very precise in our predictions. In this chapter, we propose a conceptual approach to have a good estimation of the RUL that goes through a noise reduction step with an auto-encoder model that will then be combined to a sequential LSTM Network.

Keywords Predictive maintenance · Deep learning · Long short-term memory · Remaining useful life

1 Introduction

First of all, we can define maintenance as all the actions and processes that allow a product to be maintained in a good working condition [1]. The PDM, on the other hand, is a process that is based on the conditions of a product at a given time before carrying out a maintenance action [1, 2]. CBM is opposed to periodic or preventive maintenance, which consists in carrying out maintenance tasks at regular intervals without really taking into account the condition of the machine. The two main problems that predictive maintenance proposes to solve are the

P. Menounga Mbilong (✉) · F.-Z. Belouadha · M. I. Kabbaj
Ecole Mohammadia d'Ingénieurs, Mohammed V University in Rabat, Morocco
e-mail: paulmbilong@research.emi.ac.ma; belouadha@emi.ac.ma; kabbaj@emi.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_46

diagnosis and prognosis of resources or equipment. Diagnosis deals with the tasks of detecting, isolating and identifying failures when they occur [2]. The second major issue, prognosis, deals with problems of predicting failures before they occur. The prognosis can be treated in two ways, the first way is to predict in advance the time remaining at a machine before one or more failures occur based on the present and past state of that machine, the second way is to predict the chances (probability) that the machine will fail at a specific time in the future. There are three possible approaches to deal prognosis and diagnosis; we have the analytical model approach, the knowledge-based approach and the data driven approach or also kwon as machine learning approach. The analytical model approach allows to give a physical interpretation of the state of degradation of an equipment by the use of analytical models such as algebraic or differential equations, to implement it requires a good knowledge of the physical processes involved on the system being studied, they are usually performed by experts in the field. After its implementation, it must be validated with a battery of data sets. Thus, using statistical techniques, it is possible to determine the degradation threshold of the studied system from the responses produced by the model. Its main disadvantage is that it can only be applied to a limited number of components, i.e., it is difficult to generalize to all types of components without a thorough knowledge of them [3].

The data-based approach, is also known as the artificial intelligence approach or Machine Learning approach, because most of the research in this area makes extensive use of tools from machine learning. With this one, the diagnosis and prognosis of the health of the equipment are made using the history of the machine data, which will be analyzed by algorithms to identify a degradation model of the machines concerned and thus be able to anticipate the future states or conditions of a resource. One of its negative points is that it depends on the ability of the data to reproduce the degradation of a system [1], in other words if the data collected are good enough (the data perfectly reflect the conditions of the machine), the resulting model will only be better and conversely the model obtained will be wrong.

The third approach is the knowledge-based approach, it is somewhat similar to the physical model-based approach in that it relies heavily or exclusively on the knowledge of the experts who implement the model. It is carried out in two ways, either by using the knowledge of experts (i.e., in uses a set of logical conditions) or by using fuzzy logic. A final approach, although obvious, is often also mentioned, namely the hybrid approach, which consists in combining previous approaches to take advantage of each other's advantages. Following the great advances in the field of machine learning and the interesting results shown by this discipline in recent years, especially in deep learning, the latter appears to be the ideal approach for our purpose. In this chapter, we use data-driven approach and we focused on prognosis, more specify we decided to work in RUL estimation. For any company in the industrial sector, having the ability to predict the occurrence of a failure on machine is a challenge, in other words knowing the remaining lifetime of a machine is crucial. To correctly estimate the RUL with good accuracy, it is important to use models that take into account the chronological aspect of the data. For this, in deep learning, the Recurrent Neural Network (RNN) has been proposed as a solution.

2 Related Works

In the literature, many papers deal with or attempt to answer the issue of estimating RUL useful life remaining on a machine. In [4], the author explains a two-phase method for estimating RULs. An offline phase, which is based on the search for variables that contain information on degradation behavior using an unsupervised variable selection method. Several health indicators (HI) are constructed from the selected variables, which represent degradation over time, and recorded in the offline database as models. In phase two called online, the method finds the offline HI most similar to the online HI, using the nearest *k*-neighbor classifier to use it as a RUL predictor. The method finally estimates the state of degradation using a discrete Bayesian filter.

In [5] authors, used LSTM for RUL estimation, in [6] and [7] authors used deep Convolutional Neural Network (CNN) based regression approach for estimating the RUL, in [8], which is a in predictive maintenance, the author presents an approach and solution to the challenge of the IEEE Prognostics and Health Management (PHM) 2008 conference. The solution uses an advanced recurrent neural network (RNN) architecture to estimate the remaining useful life of the system. This solution ranked second in the overall ranking of the competition.

In the chapter, we propose a conceptual model LSTM that can be coupled with a first step of noise reduction to get good accuracy in prediction of RUL, like [9], the authors use auto-encoder to denoising geophysical datasets.

The rest of the document is structured as follows. Section 3 describes data preparation and evaluation metrics use in RUL; Sect. 4 explains the architecture proposed, and Sect. 5 concludes the chapter.

3 Problem Setting

To be able to predict the RUL of machines, the idea is to install sensors on these machines first. Sensors are going to collect information about machines; these information can be temperature, pressure, noise etc. This information can be structured in columns. Imagine sensors that have generated n columns of data for $m$machines and then assume that each machine with a lifetime l, we can write $X_{l_j}^{m_i}$, denote data from machine m_i with lifetimev_j, where $X^m \in \mathbb{R}^{v_j \times n}$. This configuration is for the training set, in machine learning it is important to have a training set and a test set. In the test set we have k machines, from which we have taken a number of cycle t before they fail, we can write $P_{t_j}^{k_i}$, to denote test engine unit k_i with part of lifetime t_j, where $P^k \in \mathbb{R}^{t_j \times n}$. The goal is to get good estimation of the remaining useful life for all P^k given X^m. Before using the data, it is important to put in place mechanisms to put all the data on the same scale.

3.1 Data Preparation

This step consists in performing transformations on our data so that they can be easily integrated with our models.

When the data have different orders of magnitude, some algorithms take longer to find the ideal setting for the weights of the models. Normalization and standardization allow the data to be reduced to a very close scale. In this chapter, we use normalization; it will be applied to the columns, using the following formula,

$$X_{\text{scale}} = \frac{X_{\text{before}} - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where X_{before} represents the value of a column cell before normalization,

X_{min} is the minimum value in a column, $X_{\max}$ the maximum value in a column, X_{scale} the new value of the cell after normalization.

The data sets used in this chapter are not labeled in order to allow this it is useful to create labels that more or less capture reality, this is the purpose of the following section.

3.2 Target and Performance Evaluation Function

The data contained in the dataset that we used are not labeled, in order to be able to regress, it is necessary to label them by associating to each machine unit the remaining lifetime after each operating cycle, for this purpose, the easiest way to perform this task is to use the following formula,

$$\text{Rul}_{i\,(\text{unit}_k)} = \text{Cycle}_{\max(\text{unit}_k)} - \text{Cycle}_{i\,(\text{unit}_k)} \tag{2}$$

where $\text{Cycle}_{\max(\text{unit}_k)}$represents the maximum cycle of the unit (engine),

$\text{Cycle}_{i\,(\text{unit}_k)}$represents the cycle i of unit cycle k,
$\text{Rul}_{i\,(\text{unit}_k)}$ is the remaining useful life of the unit cycle k in cycle i

To measure the performance of our models, the evaluation score is computed, so that the function used strongly penalizes models that have overestimated the remaining useful life, it was provided during PHM'08, it is given by the following formula,

$$\text{Scores} = \begin{cases} \sum_{i=1}^{n} \left(e^{-\frac{p_i}{13}} - 1\right), & \text{when } p_i < 0 \\ \sum_{i=1}^{n} \left(e^{-\frac{p_i}{10}} - 1\right), & \text{when } p_i \geq 0 \end{cases} \tag{3}$$

where n is total number of engine unit in test set,

$$p_i = \mathrm{RUL}_{\mathrm{pred}_i} - \mathrm{RUL}_{\mathrm{truth}_i} \tag{4}$$

where $\mathrm{RUL}_{\mathrm{pred}_i}$ is model prediction for engine unit i, and $\mathrm{RUL}_{\mathrm{truth}_i}$ is the truth remaining life of engine unit$i$in test set.

4 Network Architecture Proposed

We proposed a model that can combine LSTM network to a pre-processing preliminary stage for reducing noise.

4.1 Steps to Follow

In predictive maintenance coupled to machine learning, the commons steps used by people are data acquisition and data manipulation. Data acquisition is a process that involves collecting and storing data from a physical resource for predictive maintenance purposes. The data collected are generally of two types. The first type concerns event data that must specify a situation and what has been done to correct it, and the second type concerns data on the state of health of resources, which are the most abundant, usually from sensors. Data for monitoring the state of health of resources are the most abundant and varied, they can come from vibrations, engine oil analysis, acoustic sensors, pressure sensors, temperature sensors, ultrasound, to name a few. After that, model can be fitted in a model for prediction, in this chapter, we propose a first step to reduce noise with auto-encoder; Fig. 1 summarizes the steps we used. First, we get the data from our sensors, then we normalize them, knowing that ours data have noise due to several factors such as calibration failure, we deleted it with auto-encoder model, in parallel we label our data by associating each line an integer corresponding to the number of cycles remaining before engine unit fault like Eq. (2). Let us note here that our test data will also undergo a noise reduction with auto-encoder. Therefore, we can make our predictions.

4.2 Denoising Data

Data manipulation is a step of data filtering and analysis, filtering is used to reduce noise and errors in the data. Indeed, event data can have many errors since they are generally entered manually by humans and as for resource status data, errors in these can be due to problems on sensors. Data analysis allows us to better understand and interpret the data, for this we can use several tools (such as algorithms, physical

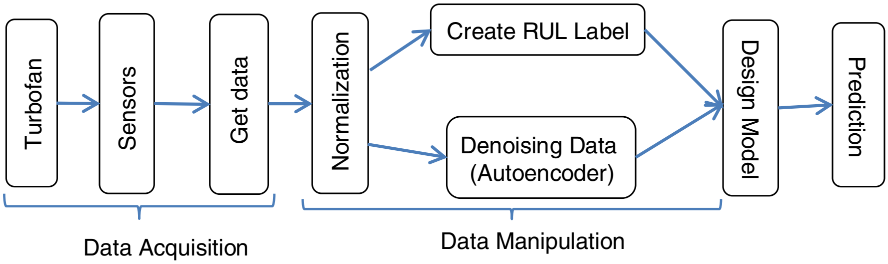


Fig. 1 Steps for building a model

models) and each tool used depends on the type of data in our possession. Given the variety of data, they can be grouped into three categories or types of data, the value type, the waveform type, the multidimensional type. The data of type, value type, waveform type, multidimensional type are data extracted at specific times as variables representing the state of a resource whose characteristic is respectively to be single value (example temperature), to be a time series (example acoustic emission), to be multidimensional (example infrared images). This step of manipulating data of any type is also called attribute engineering, it generally includes signal analysis, selection and extraction of relevant attributes [2, 3]. Signal analysis is a process of interpreting, generating and transforming a raw signal, for waveform data the most commonly used techniques can be grouped into three categories, time domain, frequency domain and time-frequency domain. Attribute selection, which is a step widely used in machine learning, consists in taking a sub-section of all input attributes in order to make the final model less costly to run, as for attribute extraction, it generally consists in combining certain attributes in order to produce new more relevant attributes. These different processes can be used successively or individually according to needs.

Most of the time, the sensors often have, more or fewer adjustment errors, also contact errors between the sensors and the target. All these small defects often add noise to the data. For reduce as much as possible this noise, we propose to use an auto-encoder model (Fig. 2). An auto-encoding neural network is an unsupervised learning algorithm that applies backpropagation by assigning equal target values to the inputs.

The auto-encoder tries to study for a given X a $f(X) \approx X$ function. The auto-encoder allows removing the noise in the data as shown in [10], authors used it to reduce noise and in speech enhancement. After the noise is removed and standardization and labeling operations have been carried out, we can apply an LSTM model to make our predictions.

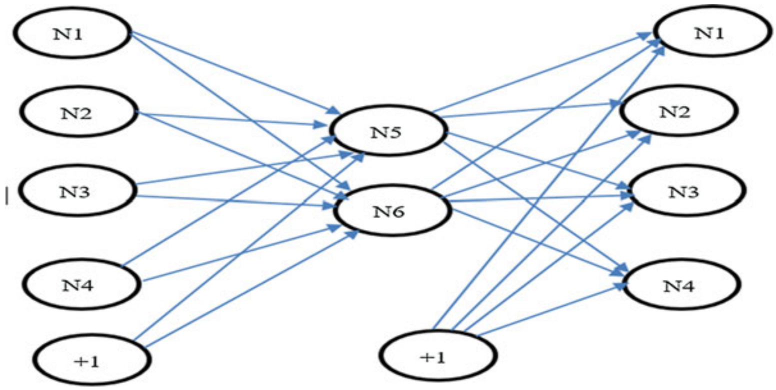


Fig. 2 Auto-encoder model

4.3 Design Model

Most of the methods developed for estimating the RUL do not take into account the presence of noise in the data and this can lead to prediction errors. Using the data generated by our auto-encoder, we can apply a simple sequential LSTM model, unlike the very complex model proposed in the literature. Long-term short-term memory are a special type of RNN, capable of learning long-term dependencies. LSTM are designed to explicitly memorize information over long periods. In this part, the LSTM model built must have the particularity of not being very complex because that is the goal indeed our hypothesis is that the current methods are complex because they completely omit the data filtering part.

5 Conclusion

In this chapter, we have proposed an approach for estimate remaining useful life of machinery, in two steps, first step we perform noise reduction with an auto-encoder model after that we applied a sequential LSTM model for prediction remaining useful life estimation. Most of the time, the auto-encoder is used for anomaly detection and dimension reduction tasks, in our approach we propose to work with the data generated by an auto-encoder instead of working directly with the dataset noise data for the training and test phases. For future work, we plan to put this approach into practice by applying it to real data to demonstrate its effectiveness.

References

1. Shin, J.-H., & Jun, H.-B. (2015). On condition based maintenance policy. *Journal of Computational Design and Engineering, 2*, 119–127.
2. Jardine, A., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*, 1483–1510.
3. Jahnke, P. (2015). *Machine learning approaches for failure type detection and predictive maintenance* (pp. 51–83). Technische Universitat Darmstadt.
4. Mosallam, A., Medjaher, K., & Zerhouni, N. (2016). Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. *Journal of Intelligent Manufacturing, 27*, 1037–1048.
5. Zheng, S., Ristovski, K., Farahat, A., & Gupta, C. (2017). Long short-term memory network for remaining useful life estimation. In *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)* (pp. 88–95).
6. Babu, G. S., Zhao, P., & Li, X.-L. (2016). Deep convolutional neural network based regression approach for estimation of remaining useful life. In *International Conference on Database Systems for Advanced Applications* (pp. 214–228).
7. Jayasinghe, L., Samarasinghe, T., Yuen, C., Low, J. C. N., & Ge, S. S. (2018). Temporal convolutional memory networks for remaining useful life estimation of industrial machinery. *arXiv Preprint* arXiv:1810.05644.
8. Heimes, F. O. (2008). Recurrent neural networks for remaining useful life estimation. In *2008 International Conference on Prognostics and Health Management* (pp. 1–6). IEEE.
9. Bhowick, D., Gupta, D. K., Maiti, S., & Shankar, U. (2019). Stacked autoencoders based machine learning for noise reduction and signal reconstruction in geophysical data. *arXiv Preprint* arXiv:1907.03278.
10. Lu, X., Tsao, Y., Matsuda, S., & Hori, C. (2013). Speech enhancement based on deep denoising autoencoder. In *Interspeech* (pp. 436–440).

Scar Tissue Evaluation in the Left Ventricular Endocardial Wall Using Pixel-Based Concept

Yashbir Singh, S. Deepa, Heenaben Patel, Joao Manuel R. S. Tavares, Salah-Eddine Krit, and Weichih Hu

Abstract Scar tissue is an identified cause for the development of malignant ventricular arrhythmias in patients of myocardial infarction, which ultimately leads to cardiac death, a fatal outcome. The scar formation is an irreversible process of the formation of dead muscle cells that is related to the blockage of a coronary artery. Here, we present a framework to find the possible scar tissue location on the left ventricular (LV) endocardium wall using computed tomography (CT) with delayed-enhancement images by implementing a pixel-based concept. We performed automated LV segmentation to find the LV endocardial wall, extracted the pixel value of the endocardial wall for each image in the sequence, performed morphological operations and marked the particular regions where is the more possible region of the scar tissue on the endocardial wall of LV. This proposed methodology will help in a better understanding of scar tissue remodeling and ways to improve LV function.

Keywords Scar tissue · LV endocardial wall · Morphological operations · Myocardial infarction · Cardiac remodeling

Y. Singh · S. Deepa · W. Hu (✉)
Department of Biomedical Engineering, Chung Yuan Christian University, Taoyuan City, Taiwan

H. Patel
West Virginia University, Morgantown, WV, USA

J. M. R. S. Tavares
Universidade do Porto, Porto, Portugal
e-mail: tavares@fe.up.pt

S.-E. Krit
Ibn Zohr University, Agadir, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_47

1 Introduction

The scar tissue development in patients who suffered from Myocardial Infarction can lead to fatal outcomes due to the development of ventricular arrhythmia. According to the WHO (World Health Organization) report, over a million people suffer from cardiovascular disease (CVD) each year and it is predicted that the number of CVD patients in the United States will increase by 2.5 times by the end of 2050 and other developed countries [1]. During the myocardial infarction, scar tissue formation on the myocardium wall leads to the rupture of the myocardium. The size and the location of the scar are very critical factors to determine the outcome of the patients who are suffered from the initial myocardial infarction as the left ventricular (LV) pump function also depends on infarct size. It is important to understand the process of the development of LV remodeling [2]. The number of infarcts cannot survive the mechanical loads and they get rupture shortly because of catastrophic and fatal complication issues. The few infarcts develop a thin myocardial wall, to increase the myocardial wall stresses. Though, it has been verified remarkably difficult to design therapies that can improve heart function or remodeling. There are numerous exciting new therapies that are still under development to assess the scar tissue because it needs a deep understanding of infarct scar formation, properties and how the scar structure affects not only cardiac mechanics as well as electrical conduction and reflex hemodynamic compensations in the clinical aspects [3]. Here, we represent an overview of published research articles before in the field of scar detection and quantification algorithms. We have shown all reports on how they were evaluated (Table 1).

Whenever we focus to the computational field, a lot of studies has been done using CMR modalities for scar detection on myocardium wall but not using CT modalities and in the case of algorithm implementation, they have used only standard deviation and average of the intensity value method which is very common for scar detection till now. Our motive is to use the CT modalities because CT scan works by acquiring several X-rays at many angles, these modalities are very quick, faster and widely available which is a very challenging task. Here, we designed an automated method for the detection of myocardium scar tissue using computed tomography (CT) with delayed-enhancement images by implementing a pixel-based concept.

2 Materials and Method

2.1 Data Acquisition

We acquired the cardiac CT images datasets of five Inpatients which are captured ten sets of timing frames at the same position with different contrast mediums, including a complete cardiac cycle by the Philips computerized tomography instrument.

Table 1 A brief summary of previously published scar detection articles

Publications	Subjects	*n*	Modality	LV/LA	Algorithm	Assessment
Kolipaka et al. [4] (2005)	Human	23	CMR	LV	SD	Percentage scar
Positano et al. [5] (2005)	Human	15	CMR	LV	Clustering	Percentage scar
Yan et al. [6] (2006)	Human	144	CMR	LV	SD	Percentage scar
Schmidt et al. [7] (2007)	Human	47	CMR	LV	SD	Infarct size
Hennemuth et al. [8] (2008)	Human	21	CMR	LV	EM fitting	Percentage scar
Oakes et al. [9] (2009)	Human	81	CMR	LA	SD	Percentage scar
Detsky et al. [10] (2009)	Human	15	CMR	LV	Clustering	Infarct size
Tao et al. [11] (2010)	Human	20	CMR	LV	Otsu thresholding	Dice
Knowles et al. [12] (2010)	Human	7	CMR	LA	MIP	Percentage scar
Lu et al. [13] (2012)	Human	10	CMR	LV	Graph-cuts	Infarct size
Rajchl et al. [14] (2014)	Human	50	CMR	LV	SD, FWHM	Percentage scar

LV left ventricular, *LA* left atrium, *n* number of datasets, *SD* standard deviation, *FWHM* full-width-at-half-maximum, *MIP* maximum intensity projection, *EM* expectation-maximization fitting

Every data set has 409 images with 512 × 512 dimensions. We used the delayed-enhancement cardiac CT images dataset. Dataset was in the DICOM (Digital Imaging and Communications in Medicine) format. We used the RadiAnt DICOM Viewer program to see the dataset. We extracted the desired dataset that exhibits the LV part (149 CT images) and few layers were avoided due to its insignificance. Data is provided by the National Institute of Hospital of Yang-Ming University. The informed consent procedure and the study was conducted as per the Institutional Review Board of National Yang-Ming University Hospital.

2.2 Automatic Segmentation of LV

We have used Segment CT software for automatic LV segmentation. It is applied on short-axis stacks, by the entering points in the reconstruction process and found LV. This software is optimized for automated description of chambers from CT images, helps in the finding of endocardium and epicardium. We considered only the desired dataset for the study that shows the LV part, performed automatic LV segmentation, found LV epicardial and endocardial wall.

2.3 Mathematical Calculation of Pixel Value

This step is involved with the selection of the LV myocardium wall and extracted the pixel value of the selected myocardium wall region. We performed SD & average of the pixel value for finding the possibilities of scar area in the desired dataset. This idea was taken from the literature which is if three SD values are more than the average intensity value of a healthy myocardium region that shows the possibilities of scar tissue [4, 7, 15]. We have focused to find out the pixel value of the LV myocardium wall region because every pixel present on the surface as HU provides the clues to precede this research work. From the literature, we get the idea HU allows a simple method to characterize certain tissue. In the CT scan image, HU is directly proportional to the degree of X-ray attenuation to every pixel to exhibit the image, represents the density of the tissue [16, 17]. Here in this step, we performed the automated localization and subsequent cropping of the selected myocardium wall that infers to LV of heart and calculated SD and the average value (Fig. 1).

2.4 Implementation of Region Growing Algorithm

This part has performed on MATLAB, R2018a platform. The seeded region growing algorithm for image segmentation, proposed by Adams and Bischof [18], is very easy to execute method. This method is very efficient, easy to apply on gray scale images, and can be extended to color images as well. The algorithm begins with a set of seed points in an individual region. The seed points compare to their neighbors on the basis of a similarity criterion and then the neighbor's pixel is computed by 4-connectivity where the pixels are connected horizontally and vertically. This is the simplest similarity criteria to calculate the difference between the intensity value of the image pixel and the corresponding region mean. In this way, we performed SD and average calculation of the region growing area which is LV myocardium (Fig. 2).

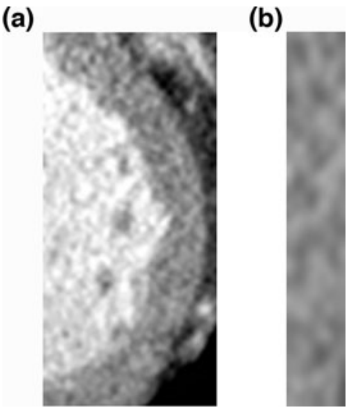


Fig. 1 (**a**) LV myocardium wall, (**b**) selected myocardium wall region

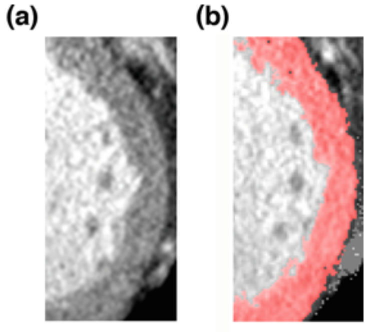


Fig. 2 (**a**) Original image, (**b**) region growing

2.5 Implementation of Morphological Operations

In this step, we tried to implement basic morphological operators which are erosion, dilation, opening and closing that helps to get some clues. Firstly, we performed erosion function which means "shrink" or "thin" of an image in a binary image. After erosion, we moved to the Dilation function which means "grow" or "thicken" an image in a binary image. Further, we performed Opening and closing functions which make breaks narrow isthmuses, eliminates thin protrusions, fuses narrow breaks, long thin gulfs, and eliminates small holes [19]. In the next step, we performed masking of the original image with a close image and we found the bold boundary color. Here, we set the criteria if two STD and one average value are higher than the pixel value that will be considered as a contrast area in the particular dataset. Finally, we were able to get a clear view of the contrast area.

3 Results

In this study, we investigated the method for the automatic prediction of scar tissue on the LV myocardium wall. According to literature, the presence of the scar tissue on the endocardium of the LV plays a significant role in determining function and remodeling. Here, we calculated the Standard Deviation (SD) and an average of selected myocardium wall pixel values of every slice, calculated average 58.7777 and SD 20.7579 of the whole dataset. This was the very initial step. After this, we performed region growing algorithm on the original dataset and calculated the average, SD and number of pixels present in the particular growing area of the dataset which is the myocardium wall of LV, performed standard deviation for the whole population which is found 59.89 and 21.2014. Here, our result is getting a match with the previous studies; it means we have performed the analysis in the right direction. In the further step, we performed the morphological algorithm of image processing such as erosion, dilation, closing, and opening. After region growing, we did erosion where the image shrinks (Fig. 3) and after that performed dilation where

Fig. 3 Erosion of images

Fig. 4 Dilation of the images

Fig. 5 Closing of the images

the images become thicker (Fig. 4) but the images were small holes in the dilated part, performed closing algorithm which removes the small holes (Fig. 5).

After this step, we did masking of the image with the original to closed image; provide the masked image where the area becomes a bit bold color (Fig. 6). Here we considered the threshold value of the mean of intensity which converts the image into binary form, we found black and white spots, and white spots represent the high intensity means more high contrast area (Fig. 7). Finally, we marked the point where is the more pixel value that shows the possible area of scar tissues. Various studies have exhibited a linear relationship between HU and the pixel value in CT images

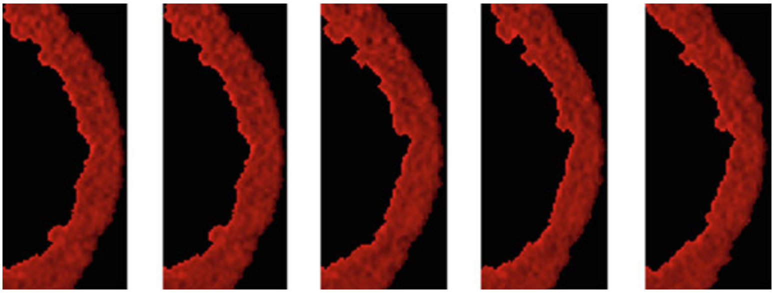


Fig. 6 Masking of the original images

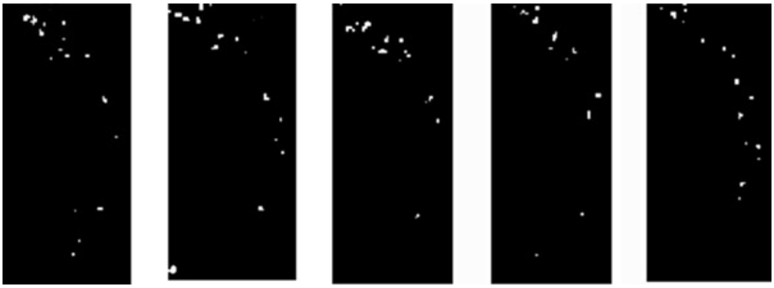


Fig. 7 Clear view of Contrast area

and suggested that pixel value can be used for estimation of tissues. Here, in this study, we have shown the connection of the pixel value to tissue recognition.

4 Discussion

Numerous clinical studies have exhibited that myocardial scars are very significant in the evaluation of the recovery of function after revascularization [20]. In the literature, CMR is the only imaging modality allowing for the identification of scar tissue till now. Scar appears whiter in comparison to normal myocardium because of high spatial resolution. CMR modality has become the most popular method for assessing the presence of myocardial scars [4–10]. In this research, we have delineated a novel automated method to find the scar on the endocardium wall from cardiac CT cardiac images using the pixel-based concept, to our knowledge, this is the first description of LV myocardial scar tissue identification from cardiac CT images. However, this work is a bit challenging due to the large variability of cardiac structures across patients. Various research groups have validated the potential of imaging data to provide meaningful insights into scar structure for the intention of modeling ventricular arrhythmia. The aim of this research is to provide a standardized methodology for evaluating scar tissue which will help to provide a better understanding of the structure and possible location of the scar.

5 Conclusion

Pixel-based concept defiantly plays a major role in the quantifying scar tissue on the LV endocardial wall. As we know, computed tomography (CT) with delayed-enhancement images is a very challenging imaging technique for tissue detection. This study helps to give a new direction and this proposed framework gives some standing and acceptability for future algorithms of scar detection techniques.

6 Limitations

This study included a very limited number of patients who had myocardial infarctions and the methodology of the present study requires future validation on a large scale study.

Acknowledgments The author acknowledges support from the National Institute of Hospital of Yang-Ming University and this study was also supported by the Department of Biomedical Engineering, Chung Yuan Christian University Taiwan.

References

1. WHO. (2013, March). *Cardiovascular diseases (CVDs), fact sheet number 317*. http://www.who.int/mediacentre/factsheets/fs317/en/index.html
2. Amigoni, M., et al. (2007). Mitral regurgitation in myocardial infarction complicated by heart failure, left ventricular dysfunction, or both: Prognostic significance and relation to ventricular size and function. *European Heart Journal, 28*(3), 326–333.
3. Ashikaga, H., et al. (2007). Magnetic resonance–based anatomical analysis of scar-related ventricular tachycardia: Implications for catheter ablation. *Circulation Research, 101*(9), 939–947.
4. Kolipaka, A., et al. (2005). Segmentation of non-viable myocardium in delayed enhancement magnetic resonance images. *The International Journal of Cardiovascular Imaging, 21*(2–3), 303–311.
5. Positano, V., et al. (2005). A fast and effective method to assess myocardial necrosis by means of contrast magnetic resonance imaging. *Journal of Cardiovascular Magnetic Resonance, 7*(2), 487–494.
6. Yan, A. T., et al. (2006). Characterization of microvascular dysfunction after acute myocardial infarction by cardiovascular magnetic resonance first-pass perfusion and late gadolinium enhancement imaging. *Journal of Cardiovascular Magnetic Resonance, 8*(6), 831–837.
7. Schmidt, A., et al. (2007). Clinical perspective. *Circulation, 115*(15), 2006–2014.

8. Hennemuth, A., et al. (2008). A comprehensive approach to the analysis of contrast enhanced cardiac MR images. *IEEE Transactions on Medical Imaging, 27*(11), 1592–1610.
9. Oakes, R. S., et al. (2009). Clinical perspective. *Circulation, 119*(13), 1758–1767.
10. Detsky, J. S., et al. (2009). Reproducible classification of infarct heterogeneity using fuzzy clustering on multicontrast delayed enhancement magnetic resonance images. *IEEE Transactions on Medical Imaging, 28*(10), 1606–1614.
11. Tao, Q., et al. (2010). Automated segmentation of myocardial scar in late enhancement MRI using combined intensity and spatial information. *Magnetic Resonance in Medicine, 64*(2), 586–594.
12. Knowles, B. R., et al. (2010). 3-D visualization of acute RF ablation lesions using MRI for the simultaneous determination of the patterns of necrosis and edema. *IEEE Transactions on Biomedical Engineering, 57*(6), 1467–1475.
13. Lu, Y., et al. (2012). Automated quantification of myocardial infarction using graph cuts on contrast delayed enhanced magnetic resonance images. *Quantitative Imaging in Medicine and Surgery* 2.2, 81.
14. Rajchl, M., et al. (2013). Interactive hierarchical-flow segmentation of scar tissue from late-enhancement cardiac MR images. *IEEE Transactions on Medical Imaging, 33*(1), 159–172.
15. Kim, R. J., et al. (1999). Relationship of MRI delayed contrast enhancement to irreversible injury, infarct age, and contractile function. *Circulation, 100*(19), 1992–2002.
16. White, S. C., & Pharoah, M. J. (2009). *Oral radiology: Principles and interpretation*. St. Louis, Mosby.
17. Mah, P., Reeves, T. E., & McDavid, W. D. (2010). Deriving Hounsfield units using grey levels in cone beam computed tomography. *Dentomaxillofacial Radiology, 39*(6), 323–335.
18. Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 16*(6), 641–647.
19. Eddins, S. L., Gonzalez, R. C., & Woods, R. E. (2004). *Digital image processing using matlab* (pp. 6–12). Princeton Hall Pearson Education Inc..
20. Kaandorp, T. A., Lamb, H. J., Van der Wall, E. E., De Roos, A., & Bax, J. J. (2005). Cardiovascular MR to assess myocardial viability in chronic ischemic LV dysfunction. *Heart, 91*, 1359–1365.

Comparison Study on Some Convolutional Neural Networks for Cerebral MRI Images Segmentation

Hicham Moujahid, Bouchaib Cherradi, and Lhoussain Bahatti

Abstract Manual segmentation of brain tumors from MRI images is very frustrating and time consuming for medical doctors and relies on accurate segmentation of regions of interests. Convolutional neural networks (CNN)–based segmentation has gained a huge amount of attention over the last few years due to its speed and automated aspect. As the CNN models are becoming more efficient for image analysis and processing, they increasingly defeat previous state-of-the-art classical machine learning algorithms. Through this study, we provide an overview of CNN-based segmentation models for quantitative brain MRI image segmentation. As this has become a fast-expanding field, we will not survey the entire existing landscape of methods, but we will focus on the three best outperforming algorithms according to evaluation parameters. First, we review the current conventional methods and deep learning architectures used for segmentation of brain lesions. Next, we perform deep performance comparison based on accuracy and loss function of some relevant selected CNN methods. Finally, a critical analysis of the current study is made to identify all pertinent issues and limitations to work on.

Keywords Convolutional neural network · Machine learning · Brain tumor segmentation · MRI images

H. Moujahid · L. Bahatti
SSDIA Laboratory, ENSET of Mohammedia, Hassan II University of Casablanca, Mohammedia, Morocco

B. Cherradi (✉)
SSDIA Laboratory, ENSET of Mohammedia, Hassan II University of Casablanca, Mohammedia, Morocco

STIE Team, CRMEF Casablanca-Settat, Provincial Section of El Jadida, El Jadida, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_48

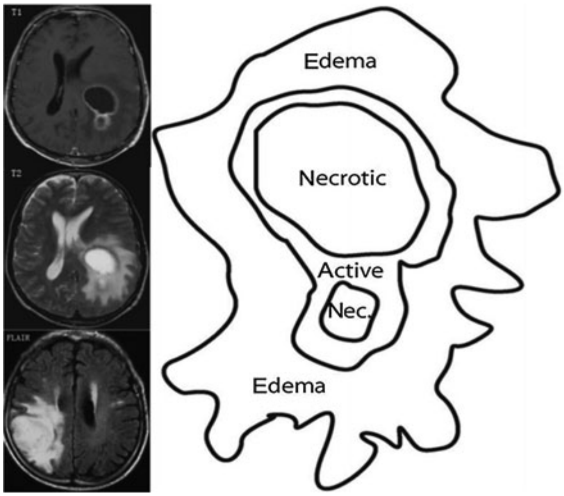


Fig. 1 Left: Three types of brain tumor MRI images with three imaging modalities: T1 with contrast, T2 and FLAIR; right: the three main components after segmentation [2]

1 Introduction

Malignant brain tumors are still the second biggest cause of death in the world for patients between 15 and 34 years old. Therefore, the early and accurate diagnosis is very important to get an efficient treatment. This task requires a deep analysis of MRI brain images to achieve an accurate segmentation by medical experts [1].

The goal of brain image segmentation is to separate the different cerebral tissues. The main tissues in normal MRI brain images contain white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). Pathological images are detected based on the examination of tumor tissues according to their density, shape, location, size, and edge. Additional types of tissues such as active cells, necrotic core, and edema can appear (Fig. 1). Based on the segmentation results, we can not only determine the type and the grade of the tumor but also the volume and the shape.

Manual method of brain tumor segmentation is still the most efficient approach and is usually used by radiology experts. The main inconvenience of this method is that it is linked to human mistakes and it is a laborious time-consuming task.

Automatic methods of brain medical image segmentation are still a challenging problem for scientists.

Recently, many researchers become interested in artificial neural networks (ANN), especially convolutional neural networks (CNN) to achieve medical image segmentation tasks [3]. CNN is a traditional ANN performing convolution where parameters like weights and biases can be learned over the network. It enables machines to solve complicated pattern recognition problems and extract objects from 2D or 3D images using the ability of CNN models to learn from large data sets.

Several visual recognition challenges are organized each year over the world in purpose to improve the performance of existing related models retrieving the best accuracy rate on the image classification and object extraction tasks [4].

Healthcare organisms produce huge amounts of medical data. In particular, radiology centers produce by prescriptions of neurologist doctors, and a lot of MRI examinations contain extremely valuable information which can serve for learning processes in CNN algorithms. This technique leads to big improvements in their performance surpassing traditional analysis methods [5].

In this chapter, we propose a comparative study of some CNN methods that present a high accuracy on brain tumor segmentation task, especially the models that are proposed by the teams obtained the first position on some famous challenges [6–8].

The rest of the chapter is organized as follows: In Sect. 2, we will introduce an artificial neural network concept and cover the structure of convolutional neural networks and its mean processes. In Sect. 3, we will make a detailed comparison for three best CNN models in well-known competitions organized for the purpose of brain tumor segmentation. In Sect. 4, we discuss the performance evaluation results for some relevant CNN models applied to brain tumor segmentation. Then, Sect. 5 concludes this chapter.

2 Methods

2.1 Artificial Neural Networks

Artificial neural network (ANN) is a system with a concept that is inspired from the biological brain which is the most complicated smart system that processes data received from the five senses in the body. The processing ability is obtained from synapses in the form of a connection between large amounts of biological neurons, forming a huge neural network able to process a lot of complicated operations [9].

2.2 Convolutional Neural Networks

In deep learning context, a convolutional neural network is a subclass of the ANN network, most commonly applied to analyze image and video contents. As its name indicates, a CNN is an ANN network containing at least one convolution layer.

CNN network that systematically has a convolution layers intermixed with a pooling layer and connected with and before a fully connected layer [9].

The idea of a CNN is relatively old. This model had been proved to work better for handwritten character recognition as early as 1998 [10]. However, due to the failure of these networks to process too much larger images, they become slowly

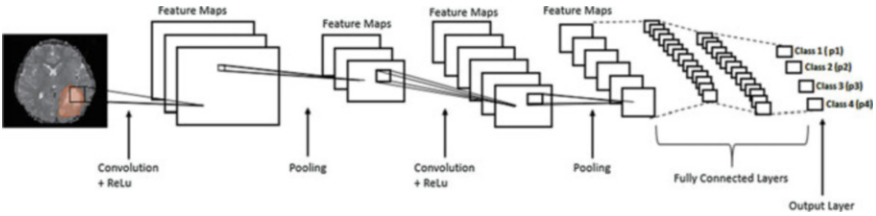


Fig. 2 Convolutional neural network architecture illustration

old-fashioned. This was because of memory and graphic processing constraints, and the lack of large training datasets. With the increase in computational power thanks to wide availability of GPUs, CPU, RAM, and the introduction of large-scale datasets like the Image-Net [11], it was possible to train complex models.

Some recent research papers focus on developing a CNN algorithm that classifies the extracted brain tumor from MRI to labels of specific brain tumor types, which is real exploitation of convolutional deep learning methods [12].

2.3 Building a CNN Blocks

CNNs have architectural constraints to reduce the processing complexity and ensure pattern detection disregarding the translational aspect of objects. A CNN consists of two principal constraints.

Shared weights: A convolution unit receives its input from multiple neurons from the previous layer, the input neurons share their weights creating proximity, every shared weight set is known as a kernel (convolution kernel) and is propagated through the network layers.

Local connectivity: Layer neurons are connected only to neurons in the next layer. This design reduces the number of connections between successive layers.

CNN network architecture should contain three types of layers as illustrated in Fig. 2: convolutional layers, pooling layers, and fully connected layer [5].

Pooling layer always comes after the convolution layer in a CNN network. Its main goal is to reduce computing complexity by minimizing the number of parameters.

Building a model requires initialization of its parameters before performing a convolution operation, then add convolution layers to the CNN classifier, then defining a pooling operation by choosing the adequate activation function. At the end adding fully connected layers to a CNN model is necessary.

2.4 *Methodology for Brain Tumor Segmentation with CNN*

There are some principle stages for MRI segmentation: pre-processing, features extraction, features selection, training, validation, testing, CNN classification, and post-processing.

2.4.1 Pre-Processing

During this step, dataset is prepared and regularized to feed the network. The result of this process can be saved under a specific extension and used again for training or testing step later. The mean sub-processes of this step are normalization, standardization, and whitening. The outputs of data preprocessing can be utilized to feed various neural network models [4].

2.4.2 Feature Extraction

This step consists of extracting and converting original raw data to useful features expected to be irredundant and useful. The given step is necessary for the future learning steps.

2.4.3 Feature Selection

This process consists of selecting top ranked features forming a subset of original features according to predefined criteria. This is an important process and is usually used to reduce dimensionality in deep learning. Recently, many researchers have created a lot of feature selection methods designed for different goals and have some advantages and inconvenient [9].

2.4.4 Training

Learning is achieved through this step by minimizing as much as possible the loss function over iterations. CNN-based models usually use the soft-max loss function or the sigmoid cross-entropy function [9].

CNN networks are generally trained using the back-propagation algorithms by using the chain concept to implement and execute the gradient descent (GD) algorithm. However, for datasets with a huge number of images, using GD is not recommended. A model called stochastic gradient descent (SGD) must be used to compute gradients according to individual image or sub-dataset rather than the whole dataset [10].

When training a network with a lot of parameters, a powerful regularization method is required to reduce over-fitting such as dropout method, especially to improve generalization for complex CNN networks.

In dropout method, neurons are dropped randomly with a probability p while training. However, a partial subset of neurons is trained in a single iteration of SGD based algorithm. When testing, all neurons must be utilized [13].

Batch normalization (BN) is another useful process that regularizes and improves generalization and speeds up network convergence. This technique is the best solution for the problem of internal covariate shift, where the distribution of inputs is changing continuously for all layers while training. This phenomenon usually slows down training and requires careful initialization [14].

Four important factors are used to estimate the performance of the network during training step:

True positives (TP): Images containing the tumor and the neural network judges matching the ground truth.

Negative positives (NP): Images not containing the tumor and the neural network identifies properly matching the ground truth.

False positives (FP): CNN classifies images as containing tumor but that actually do not contain any tumor in the ground truth.

False negatives (FN): CNN classifies images as not containing any tumor but are actually containing a tumor in the ground truth, which is a very important parameter in case of MRI tumor detection [9].

For quantitative analysis, we define the ground truth areas as circular regions (ROI). Based on the above definitions, we can compute the precision (P), recall (R) and F1-score as:

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{F1} = \frac{2 * P * R}{P + R} \tag{3}$$

F1 combines (Eq. 2) and (Eq. 3) to get a specific parameter; it can be interpreted as a weighted average of the P and R, and its value is between 0 and 1.

Precision tells us that quantity of the time we have been able to identify the object correctly, but the recall tells us how many objects are found in the dataset. In other words, the neural network has not been able to recognize all of the objects, but it is very sure of its judgment once it has identified one.

2.4.5 Validation and Testing

This step leads to determine the best network in terms of performance by calculating the error function using a minimized dataset. Usually, networks are trained with huge dataset to minimize error function.

In fact, this procedure produces some over-fitting to the validation set. The performance of the selected network should be approved by calculating its performance on another independent dataset which is the test set.

Test set is used to make unbiased evaluation of the model fit in the training dataset; however, test dataset must be independent of the training dataset, but that represents the same probability distribution as the training dataset. If a model fits to the training dataset as well as the test dataset, then we get a minimal over-fitting [1].

2.4.6 Classification

Consists of joining images to labels according to the calculated probability, the accuracy of this task defines the real performance and precision of the model.

2.4.7 Post-Processing

This step is the last step of generating over-segmented regions, like super-pixels, for obtaining the final segmentation result.

However, post-processing method is used for simplifying the network in order to improve the output without losing much in terms of prediction accuracy. At each hidden layer of the network, this method tries to find a small subset of features/nodes that can be used to reconstruct the entire layer approximately [9].

3 Comparison of Three CNN Models

Many CNN architectures have shown interesting results in recent publications. They differ in depth of the network and the number of used. Many organizations organize events to increase competitiveness of teams to improve their CNN models as much as possible. Multimodal Brain Tumor Segmentation Challenge (BRATS[1]) has the goal of evaluating the state-of-the-art methods for brain tumors segmentation by providing a 3D MRI dataset with ground truth tumor segmentation labels approved by experts of the field [8].

[1] https://www.med.upenn.edu/sbia/

3.1 Model 1: Ensembles of Multiple Models and Architectures (EMMA)

EMMA is a deep learning model which can be run with excellent performance. It gets the first position in the BRATS 2017 competition among more than 50 teams. The model wins the first position by obtaining the dice coefficient scores as below: 73.8 for enhancing tumor, 90.1 in the whole, and 79.7 in core. The next two models in position are ULC-TIG and MIC_DKFZ [6].

This algorithm is combining multiple configured and trained CNN models. The system won the competition by achieving the overall best performance in the testing phase, based on dice score.

Traditional CNN architectures if not regularized can over-fit noise in the training data, which leads to mistakes when they are used for generalization. Emma proceeds of averaging the bias infused by individual model configurations [6].

The first employed architecture is Deep-Medic. It is a fully 3D and multi-scale CNN, designed for processing efficiently 3D images. Two Deep-Medic models are used in this experiment. The first is the residual version previously presented in the BRATS 2016, and the other is a wider variant doubling the number of feature maps at each layer. The models are trained using a cross-entropy loss function conserving their original parameter.

This model includes three different 3D-FCN and two versions of 3D-Unet model with a few modifications.

3.2 Model 2: CNN-Based Segmentation of Medical Imaging Data

The proposed model is first ranked in BRATS 2015 and ISLES[2] challenges where results in BRATS 2015 are 0.61, 0.85, and 0.72 average dice for enhanced tumor, whole tumor, and tumor core, respectively.

3D segmentation is demanding more memory and processing; therefore, this CNN-based method with three-dimensional filters which is proposed consists of combining segmentation maps created at different points in the network. The work is similar to U-Net CNN architecture [7], with two modifications:

- Combining multiple segmentation maps created at different scales.
- Using element-wise summation to forward feature maps from one stage of the network to another.

The architecture consists of contracting and expanding stages found in other works using similar networks, where feature maps from the first stage are combined

[2] http://www.isles-challenge.org/

with feature maps from the second stage via long skip connections. The prevalent approach in combining feature maps in other works has been concatenation. In this experiment, concatenation is compared with element-wise summation.

Element-wise summation directly inserts local details found in the feature maps of the contracting stage to the feature maps of the expanding stage. In addition to this, all of the layers between the source and the destination of the skip connection can be viewed as one large residual block.

Furthermore, this method looked at combining segmentation maps created at different points in the network. The network uses PReLU activations.

Finally, summation was found to produce slightly worse results, while combining multiple segmentation maps was shown to speed up convergence without hurting the final performance. To tackle the issues of class imbalance, this is a challenging aspect of medical image segmentation [7].

3.3 Model 3: Auto Encoder Regularization Based

This algorithm is the most ranked one for BRATS 18 challenge named NVDLMED.

BRATS 2018 training dataset contains 285 cases, 210 of high grade and 75 of low grade, each with four 3D MRI modalities (T1, T1c, T2, and FLAIR) rigidly aligned.

The proposed approach follows encoder–decoder based CNN architecture with an asymmetrically larger encoder to extract image features and a smaller decoder to reconstruct the segmentation mask. To better cluster the features of the encoder endpoint, it follows the variational auto-encoder (VAE) approach [8].

The RESNET [15] blocks are used in the encoder part, but each block contains two convolutions including normalization (Group Normalization) and ReLU, followed by additive identity skip connection [8].

The decoder architecture is the same as the encoder, the difference by just one block per each special level. The special size of the end of the decoder is the same as the original size.

VAE consists of reducing the result of the encoder output to lower dimensional space and then reconstruct it following the same structure as the decoder.

The loss function is defined as follows:

$$L = L_{\text{dice}} + 0.1 * L_{\text{L2}} + 0.1 * L_{\text{KL}} \tag{4}$$

where

- L_{dice} is a soft dice loss applied to the decoder output to match the segmentation mask.
- L_{L2} is an L2 loss on the VAE branch output to match the input image.
- L_{KL} is standard VAE penalty term.

Table 1 Competition positions with dice score (%) obtained on different models

Method	Enhanced tumor	Whole tumor	Tumor core	Challenge
EMMA [6]	73.8	*90.1*	79.7	BRATS 2017
CNN-based Segmentation of Medical Imaging Data [7]	61	85	72	BRATS 2015
Hourglass CNN [16]	59	81	63	BRATS 2018
S3D-UNet: Separable 3D U-Net for Brain Tumor Segmentation [17]	73	88	80	BRATS 2018
NVDLMED [8]	*76*	88	*81.5*	BRATS 2018

4 Discussion

This chapter presents an overview of several methods for the automatic brain tumor segmentation task, using 3D convolutional neural networks. There is a large-scale of methods, but we propose to compare and analyze some methods including the three best ranked architecture implementations in BRATS challenges over the last 3 years according to dice coefficient.

CNN-based Segmentation of Medical Imaging Data gets the best performance (dice score) on BRATS 2015 challenge, showing the results of: 61% for enhancing tumor, 85% in the whole, and 72% in core. The model performance for EMMA in the validation and testing stage of BRATS 2017 dataset are on the top of the list with dice score: 73.8% for enhancing tumor, 90.1% in the whole, and 79.7% in core. NVDLMED gets the best performance on BRATS 2018 challenge, showing the results of: 76% for enhancing tumor, 88% in the whole, and 81.5% in core.

The other models in Table 1 are included just for comparison. The result shows clearly a relative amelioration over time on successive BRATS challenges.

All the methods are presenting a significant complexity calculation, which leads to the necessity of developing new challenging mechanisms to improve processing task and be able to train a huge dataset easily in order to obtain better precisions and reduce errors.

5 Conclusion and Perspectives

Brain tumor segmentation presents a very challenging task for researchers all over the world. Many laboratories propose each year new methods for automatic segmentation of brain tumors to surpass the limitations presented in the previous methods. However, the mean blocks that face those methods and need a lot of development are processing capacities due to the constraint of training neural networks with an extensive number of datasets to improve predictions and reduce errors, especially when classifying and segmenting the 3D MRI images of brain tumors.

Time consuming is another big limitation for the training stage that lies to the processing complexity too and needs to be minimized as much as possible over the future research.

References

1. Dubey, R. B., Hanmandlu, M., & Vasikarla, S. (2011). Evaluation of three methods for MRI brain tumor segmentation. In *2011 Eighth International Conference on Information Technology: New Generations*. https://doi.org/10.1109/itng.2011.92
2. Liu, J., Li, M., Wang, J., Wu, F., Liu, T., & Pan, Y. (2014). A survey of MRI-based brain tumor segmentation methods. *Tsinghua Science and Technology, 19*(6), 578–595. https://doi.org/10.1109/tst.2014.6961028
3. Parihar, A. S. (2017). A study on brain tumor segmentation using convolution neural network. In *2017 International Conference on Inventive Computing and Informatics (ICICI)*. https://doi.org/10.1109/icici.2017.8365336
4. Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging, 35*(5), 1240–1251. https://doi.org/10.1109/tmi.2016.2538465
5. Nwankpa, C., Ijomah, W.L., Gachagan, A., & Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. ArXiv, abs/1811.03378.
6. Kamnitsas, K. et al. (2018) Ensembles of multiple models and architectures for robust brain tumour segmentation. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., & Reyes, M. (Eds.), *Brain lesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries*. BrainLes 2017. Lecture Notes in Computer Science (vol. 10670). Cham: Springer. https://doi.org/10.1007/978-3-319-75238-9_38.
7. Kayalibay, B., Jensen, G., & Smagt, P.V. (2017). CNN-based segmentation of medical imaging data. ArXiv, abs/1701.03056.
8. Myronenko, A. (2019). 3D MRI brain tumor segmentation using autoencoder regularization. *Lecture Notes in Computer Science*, 311–320. https://doi.org/10.1007/978-3-030-11726-9_28
9. Kevin Zhou, S., Greenspan, H., & Shen, D (2018). *Deep learning for medical image analysis*. https://doi.org/10.4103/jpi.jpi_27_18.
10. Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324. https://doi.org/10.1109/5.726791
11. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision, 115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y
12. Buda, M., Saha, A., & Mazurowski, M. A. (2019). Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in Biology and Medicine, 109*, 218–225. https://doi.org/10.1016/j.compbiomed.2019.05.002
13. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. ArXiv, abs/1207.0580.
14. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. ArXiv, abs/1502.03167.
15. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. *Lecture Notes in Computer Science*, 630–645. https://doi.org/10.1007/978-3-319-46493-0_38

16. Benson, E., Pound, M. P., French, A. P., Jackson, A. S., & Pridmore, T. P. (2019). Deep hourglass for brain tumor segmentation. *Lecture Notes in Computer Science*, 419–428. https://doi.org/10.1007/978-3-030-11726-9_37
17. Chen, W., Liu, B., Peng, S., Sun, J., & Qiao, X. (2019). S3D-UNet: Separable 3D U-net for brain tumor segmentation. *Lecture Notes in Computer Science*, 358–368. https://doi.org/10.1007/978-3-030-11726-9_32

The Implementation of ERP Systems for Developing Human Resources in Moroccan Public Sector Organizations

Malak Bouhazzama and Said Mssassi

Abstract Implementation of enterprise resource planning (ERP) systems in the public sector is not quite the same as in the private area. The aim of this study is to explore the impact of ERP system on Human Resources Management in Public Sector Organizations (PSO) using a perspective of the Balanced Scorecard (BSC) performance measurement of Human resources information system (HRIS). On the first hand, this study presents evidence based on individual interview conducted with sample of 12 ministerial departments in Morocco which have implemented the system. On the other hand, 650 public employees have been selected to carry out the skills assessment in order to conduct training activities at the level of the HRIS. Empirical field shows that the implementation of ERP system in Public resulted in a positive impact on HRM and a multitude of recommendations suggest a strategy to develop HRM in PSO.

Keywords Enterprise resource planning · ERP implementation human resources information system · Public sector · Organizational performance · Balanced Scorecard · Morocco

1 Introduction

The ERP systems gained popularity as the software application that would provide profitability, efficiency and productivity. In this purpose, public sector learning from industry has started to implement this system to improve the services to citizens and to offer a better governance and transparency. In this context, commercials target three public sector government agencies including Federal, state, and local. This study is conducted to investigate the implementation of ERP on PSO using two different methods for credible results by using BSC wish could help evaluate the performance of the HRIS. Research questions are:

M. Bouhazzama (✉) · S. Mssassi
National school of Management, Tangier, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_49

RQ1: What are the impacts of ERP implementation toward the HRM in PSO?
RQ2: Is that performing well enough to build a HR strategy?

2 Literature Review

2.1 Enterprise Resource Planning

ERP system is one of the most widely used applications that could provide a significant change in the organization. O'Leary defines ERP systems as "computer-based systems designed to process an organization's transactions and facilitate integrated and real-time planning, production, and customer response." So the fundamental component of ERP framework incorporates all business capacities in the association into a solitary database framework to show an all encompassing perspective on the business. For, Kumar and Hillegersberg it is a "configurable information systems packages that link information and information-based processes within and across functional areas in an organization." This sophisticated system can include many functions as marketing and accounting software, supply chain management and human resource management which will be in the center of this study.

2.2 Impacts of ERP on Organization

ERP is now the prevailing form of computing in large organizations in the private and public sector. An effective ERP framework execution can abbreviate generation cycles, builds exactness of interest for materials the board and sourcing and prompts stock decrease as a result of material administration, and so on. Also it very well may be utilized as an essential instrument for re-designing.

Anyway different examinations have uncovered that not all ERP executions are fruitful. As per Gray A. Langenwalter, ERP usage disappointment rate is from 40 to 60%. Different investigations have been directed to examine the effects of ERP framework on authoritative execution to demonstrate the capacity of ERP, there are additionally a few investigations demonstrated the ERP framework execution doesn't really brings to upgrade organizational performance However the focused weight released by the procedure of globalization is driving execution of ERP extends in progressively enormous numbers. For example, an examination in Greece on 193 organizations uncovered that ERP framework makes the information gathered and prepared simpler and quicker. Be that as it may, there is no decrease of work force which is negative according to the HRM strategy.

2.3 PSO in Morocco

The Moroccan Government got involved in the implementation of the strategy of administrative reform. In effect, this program worked out with the help of the World Bank, with the European Union and with the Bank African of Development. Different measures have been taken by the Moroccan government to improve the quality and responsibility of several ministerial departments and offices in order to offer better performance while guaranteeing more prominent straightforwardness transparence. The difficulties of globalization have enhanced the power for a superior open division benefits so as to accomplish upper hands and Moroccan open area need to have a world-class execution and fathom the universal benchmarks. It is about a strategy of reform of the public administration which aims at lightening administrative structures, at simplifying procedures, at ameliorating performances and at raising the quality of benefits provided by means of:

- Work into a modern administration that can contribute to the country's competitiveness and sustainable development while ensuring the medium-term viability of its macro-economic framework;
- Ensure better and lower-cost services through the strengthening of the administration's intervention efficiency and the rationalization of its expenditure by setting up an innovative budget system based, on the one hand, on the accountability of managers in return for their commitment to achieving pre-defined goals and assessing their achievement in relation to the expenses incurred, and on the other hand, the development of performance-based control and Accountability;
- Develop and strengthen the process of decentralization necessary to establish a management of proximity to the public thing in line with the concerns of the populations;
- Promote government online tools to promote public access to budget information and administrative services;
- Reaffirm the strategic role of human resource management to make it the true vector of a modern, efficient, responsible and citizen administration as well as the determining factor in the dynamics of reforms.

2.4 Specificities of Human Resources Management in PSO

Human capital is a major investment for public organizations as the goods produced are of public interest or even essential. The agents who design, produce and deliver them must rely on state-of-the-art skills. The Moroccan Government got involved in the implementation of the strategy of administrative reform. In effect, this program worked out with the help of the World Bank, with the European Union and with the Bank African of Development. Different measures have been taken by the Moroccan government to improve the quality and responsibility of several ministerial.

In this purpose, forward-looking management of jobs and skills allows you to predict, decide on, invest in, track and assess the effects of human capital needs. The forward-looking management of jobs and skills guides all of the HRM's instruments, which are thus finalized by a results-based management perspective. It is no longer just a matter of counting the number of people paid, but of making the most of them from a win-win perspective where the organization achieves its results better while the employee develops its potential better during the course of its work. The approach is part of the new two-way moral contract that connects the employer and the employee: the former develops the second one he needs while the employee contributes to the organization from which he derives several types of rewards.

All in all, the forward-looking management of jobs and skills has the advantage of valuing all the jobs of an organization by inserting them into a common approach. This integration makes it possible to deploy HRM instruments such as mobility or promotion to the best of their ability. Nevertheless, we cannot speak about system competitive forward-looking management of jobs and skills without disposing of competitive tools. The valuation of these last would be then the keystone to diagnose the whole system as well as to disclose relation with the notion of performance within the public Moroccan administration. Basic tools which concern this study are:

- Job and skills review to classify jobs and post offices of departments.
- Skills Assessment is the analysis and the valuation of professional and personal competences, as well as aptitude and motivations of a person..
- Questionnaire valuation to recognize its performances and evolution of career.
- The Human Resources Information System (HRIS): As its name points it out, a Software package of Inserted Management (PGI) or ERP (Enterprise Resources Planning).

2.5 ERP and MHR

Investments of ERP systems have the purpose to develop specific capabilities and assets, in addition to managerial and technical competencies, to produce value and opportunities for differential long-term benefits. Also, ERP systems Implementation can basically create direct operational benefits, or it can indirectly expedite business growth by stimulating innovations and capacity building around the technology. However, the investigation by Skok and Legge in 2002 managed the issue of the key partners and distinguished four primary gatherings associated with ERP usage ventures: the executives, clients, engineers and experts. The study of Skok and Legge has examined the key stakeholders power to influence the outcome of ERP project and their strategies to gain support for the project (Table 1).

This examination by Skok and Legge gave a few bits of knowledge into the alternate points of view of the four significant partners associated with an ERP

Table 1 Results about involvement of stakeholders in the ERP implementation

Stakeholders	Factors	Results
Management	Incentives, top management and stakeholders, customers	Involvement and retaining
Consultants	Knowledge transfer, motivation, communication	Power of influence
Developers	Performance, skills shortage, communication	Designing the configurations of the system
Users	International dimensions, training, sharing culture	Performance

venture. The specific accentuation was made in the zones of staff maintenance, clashes in ERP ventures, overseeing advisors, and social and business process changes.

3 Research Methodology

3.1 Analyze the Research Variables

A universal model of performance measurement is not defined anywhere and varies in different organizations, but it is an integral part of management. We can call it a set of mechanisms and procedures to ensure quality and efficient management of the organization. In this case, the objective is not to produce universal laws, but rather to propose new valid and robust theoretical conceptualizations, rigorously developed. Abductive reasoning fits perfectly with our research in that we do not seek to test theoretical hypotheses regarding performance indicators of ERP systems or to develop universal laws on this subject, but rather our research approach aims to understand and explain how the successful implementation of ERP system can contribute efficiently at building a performing strategy of human Resources to come out with a PSO's HRM model that is effective, efficient and efficient in the light of data collected in the field. It turns out that we find ourselves in the position of interpreting observer. In this frame, we opted for the qualitative method by case study, because she will allow us to understand, to analyze and to discern anomaly at the level of ERP system in PSO via the indicators of performance through her exploratory nature.

On the first hand and drawing on the forward-looking scoreboard theory Robert S. Kaplan and David Norton who questioned the only financial indicators for performance measurement, the performance indicator repository will be based on four dimensions of a model of HRM: Financial aspect, Internal business process, Learning and growth and customer to quantify efficiency, efficiency, coherence and relevance of ERP system. Nevertheless, the dimension of customer will be replaced by stakeholders as seen in the literature review according to Skok and Legge.

Table 2 Interviews in diverse PSO

PSO	Number of interviews
Ministry of Economy and Finance	6
Ministry of Equipment and Transport	1
Ministry of Health	1
Ministry of Energy, mines, water and environment	1
Ministry of the Employment and Vocational Training	1
Ministry of Culture	1
Ministry of Public Service and Modernization of administration	1
Ministry of Housing, urbanism and City policy	1
Ministry of the Interior	1
High commissioner for water and forestry and the fight against desertification	1
Ministry of Energy, mines, water and environment	1
National Office of railways	1
National Land Conservation agency and mapping	1

The interviews which were to the number of sixteen were accomplished in the twelve ministries (Table 2).

On the other hand, as The survey research method is one of the most leading research methods used in information system research in order to determine dependant and independent variables of an environment without having any control on it, 650 survey questionnaires were distributed to following employees of ministry of finance through electronic mails. The above mentioned ministry was identified because in these PSO ERP systems (named as GISRH) is implemented, active and working. The target population includes all leaders or service managers those from the above population who access the GISRH.

4 Discussion

The results indicated that the implementation of ERP system in PSO has beneficial impact toward its performance which flag a good system implementation in the HRM policy. Results show the ERP provides information for decision-making, and strategic planification. Many leaders proposed to unify the language of PSO because union is strength and performance can only be achieved by adopting the same methodology while respecting the specificity of each direction and insist on having clear objectives and measuring their achievement. Unfortunately, despite the high costs of implementing an HRIS and the need for a long period of training for the familiarization of the tool, only 10 PSO have appropriated the tool. This reservation in relation to the HRIS is explained by the complexity of the qualitative module, which despite its existence in the system is generally not frequently used which requires an internal development of the tool. Even more, the resistance of technological change on the one hand and against modernization by the forward-looking management makes the tool difficult to use in public administration.

5 Conclusion

This investigation gives a superior comprehension and commitment in the writing with respect to the effect of ERP usage with regards to Moroccan open segment associations. This examination has figured out how to demonstrate that by applying the most recent framework in the open part association, it has gotten one of the proficient technique so as to upgrade the adequacy and productivity of the presentation of the open segment association. Thus, the aftereffects of this examination ought to urge the open part to introduce comparable techniques to accomplish comparative outcomes appeared by neighborhood experts in Morocco.

SysML Extension by AADL Specification for WSN Modeling

Hilmani Adil, Lasfar Youssef, Maizate Abderrahim, and Zakari Abdelouahed

Abstract In the area of software engineering and system architecture, the design and implementation process has taken a general methodological approach, encompassing all the activities necessary to design, evolve, and verify a system providing an economical and efficient solution to needs of a client while satisfying all the stakeholders. For a long time, system engineers have used modeling techniques, the best known of which are AADL (Analysis and Architecture Design Language) and SysML (System Modeling Language). AADL was born as a language specific to the field focused on avionics and later was revised to represent and support a more general category of embedded systems in real time. SysML is an extension of the Unified Modeling Language (UML) to support engineering and modeling of the system. Based on these two languages and these two tools, several approaches have been created and developed to meet the needs and requirements of the system and also to overcome the problems and limitations of these two languages. In this paper, we will study and examine the two AADL and SysML modeling languages in a thorough way in order to extend the SysML language by the AADL language specifications for the modeling and design of wireless sensor networks (WSNs).

Keywords Modeling process · WSN · Performance evaluation · SysML · AADL

1 Introduction

In recent years, wireless sensor networks (WSNs) have performed enormously in various fields including medicine, military, and environmental domains [1–3]. This

H. Adil (✉) · L. Youssef · M. Abderrahim
RITM-ESTC, CED-ENSEM, University Hassan II, Casablanca, Morocco

Z. Abdelouahed
LGIPM, University of Lorraine, Metz, France
e-mail: abdelouahed.zakari@univ-lorraine.fr

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_50

sensor network is an ad hoc network made up of different sensor nodes which are electronic components whose role is to capture and measure specific parameters in a monitoring space called a capture field.

These sensor nodes are small nodes that use limited energy resources and low communication and processing power to collect information in a given geographic area and transfer them to the well node (called Sink). These limits are very important in the study, analysis, and deployment of a network of wireless sensors to ensure network longevity and also to guarantee better performance and reliability in receiving data collected in a practical way and effective [4]. These sensor networks have another problem that must be taken into consideration and which is to model the behavioral and non-behavioral part of the latter in order to carry out the design and the deployment using the most modeling languages used to implement their functioning and behavior.

Model-based engineering (MBE) has emerged as a key set of technologies for designing complex systems. This technology has seen in recent times a tremendous progress in the realm of development and modeling of real-time systems among them wireless sensor networks (WSNs). This modeling of systems is generally relying on the use of one of the two most distinguished and most used languages currently: AADL and SysML.

AADL [5] defines several categories of components, divided into three main families of components:

- The software components describe the application elements of the architecture.
- The hardware components define the elements of the execution platform.
- The hybrid components, described through the system component, are used to structure the application architecture (in blocks).

In addition, AADL supports the prediction and early analysis of critical system qualities, such as performance, planning, and reliability.

SysML [6, 7] is a modeling language with a graphic syntax developed and standardized by Object Management Group (OMG). It has been designed to, inter alia, capture software interactions with physical entities and is widely used in systems engineering. Compared to UML 2, SysML adds support for systems engineering (for example, by requirements engineering and quantitative analysis of the physical aspects of the system), while removing many UML software centric constructs.

The main objective of this paper is to propose a profile-specific modeling for wireless sensor networks (WSNs) based on the use of a new design technique which is based on an extension of the SysML language by adding the critical specifications of the AADL language in order to carry out and implement the modeling of wireless sensor networks.

The rest of the document is organized as follows. Section 2 provides an overview of WSN modeling methods. In Sect. 3, we will explain the requirements of the SysML profiles for the modeling of WSN. Then, in Sect. 4 we will give a new contribution based on a proposed SysML profile. Finally, we conclude with a conclusion and a perspective in Sect. 5.

2 WSN Modeling Languages

In this section, we will study and examine the different modeling languages that are the most distinguished and used today in the design and implementation of complex and real-time systems. Among these languages are UML, MARTE, BIP, SysML, and AADL.

2.1 Current Modeling Languages

UML [8] (Unified Modeling Language) is a unified modeling language and object oriented by the OMG which is based on the use of graphs and diagrams to visualize, specify, build, and document systems covering different areas of application between the software systems, based on all standardized and object-oriented methods. UML offers 13 diagrams that are used for the description and design of a system. These diagrams are divided into two categories of diagrams which are:

- Structural diagrams: These diagrams play a crucial role in the description of the body part of the system based on the identification of objects with their behaviors and their operations, not forgetting the operations and methods attributed to them.
- Behavior diagrams: These diagrams describe the dynamic part of a system that is based on the output of the results to the user once the system interacts with external events.

MARTE [9, 10] (Modeling and Analysis of Real-Time and Embedded Systems) is designed to take over from its predecessor called profile for Scheduling, Performance and Time Specification (SPT); it is a recent OMG UML profile consisting of a platform, which describes the real-time parameters and aspects of systems in UML models. This language also relies on the modeling of systems. Considering the static part, defining the components, functions, and methods, and the dynamic part describing the interaction of the system with its entourage and its environment, producing real-time results. The main objective of MARTE is to create and propose a common and unique modeling language to describe and specify at the same time the software and hardware aspects of embedded systems in real time. This language consists of three major models:

- Application model: defines aspects and functionalities of the system.
- Resource model: describes the implementation and execution infrastructure of the behavioral part of the system based on non-functional properties and parameters.
- Allocation model: describes the interaction between applications and resources.

BIP [11] (Behavior, Interaction, Priority) is a language with a theoretical basis for the construction and analysis of embedded systems in real time from heterogeneous components. This language allows the implementation and behavioral description of real-time embedded architectures. The BIP models are obtained by the association and integration of three modeling parts:

- The lower part defines the behavioral aspect of a component or object as a set of transitions (that is, an extended state machine with data).
- The intermediate part describes the connection with the behavior part using the connectors which provide a description of the interaction between these two parts.
- The upper part consists of a group of properties and priority features used to describe the laws and rules of operation and planning of interactions.

Modeling using the BIP language is based on the use of connectors and priorities to develop and build a hierarchical architecture of composite components from atomic components. This modeling has created an open and clear vision that has made it possible to design the behavior and operation of components and the overall structure of a system.

2.2 Presentation of the SysML Language

SysML [12–14] is a modeling language developed by OMG, INCOSE, and AP233 and is oriented to the field of system engineering which is based on the graph to specify, analyze, describe, verify, and validate the feasibility of a system before its realization. SysML has been designed to use and rely on an open source license to develop its contribution and use. SysML is defined as an extension of a subset of UML using the profile mechanism defined by UML.

The SysML language is a way to form and consolidate in a single, common model all feature components, properties, constraints, and parameters to model and design the entire system. To carry out this modeling, the SysML language is based on diagrams that make it easier to approach complex systems and also to implement their operations and their realizations.

With modeling using the SysML language, a system can be modeled and also can be designed in several manners and in different ways in order to formalize the overall behavior of the system. This freedom of modeling can create problems when implementing the realization of the operation of this system. This language does not allow the realization and the execution of the behavior of the system which limits its performance.

2.3 Presentation of the AADL Language

AADL (Architecture Analysis and Design Language) [15, 16] is an SAE international standard established in 2004 that produces a new project to analyze system designs prior to development and that supports the system architecture description for embedded systems.

The objective of the AADL language is to perform an efficient and precise modeling that includes all functional and non-functional parts of the system without forgetting all these properties and functions such as complexity, schedulability, message transmission, and hardware description (processor, memory, bus, etc.). To carry out this modeling, each component of the system is described in AADL according to two aspects. The first, type, corresponds to the functional interface of the component. The second, the implementation, describes the content of the component (subcomponents, time properties, connections, etc.). Each component belongs to one of the ten categories predefined in AADL, classified into three families: hardware (Execution Platform), software (Application Software), and hybrid (Composite).

To model an architecture, the AADL components must be declared as subcomponents (instances) of other AADL components. At the highest level, a system component contains all instances of other components. Most AADL components can contain subcomponents that allow a hierarchical description of the system.

3 SysML Profiles Requirements for WSN Modeling

This paper proposes a new process of modeling wireless sensor networks that is based on the combination of two languages SYSML and AADL precisely on the notion of profile in UML/SYSM; this notion allows specializing a part of the AADL language by SYSML in order to produce a profile that will provide a complete modeling support for WSN designers.

3.1 The SysML-AADL Profile

The initial experiments with the SYSML and AADL modeling languages have revealed difficulties in handling these tools. Certainly, the syntax of these languages is very wealth, but the semantics and methodology to be adopted remain dependent on the application field and the context of use. Hence the idea of specializing and customizing a profile dedicated to sensor networks based on the complementarity of these two languages and the possibility offered by extensions with SYSML profile (which is also a UML profile). The ultimate goal is to create a specific profile for wireless sensor networks by integrating new features based on feedback from these two languages.

3.2 SysML-AADL Profile Requirements for WSN

Generally modeling is a very important operation when it is poorly mastered it can engender a model that does not exactly represent the system and which, in some cases, can cause defects in the behavior or execution of the system which in turn leads to a renunciation of the solution.

In our case, we proposed a new profile of modeling and analyzing SYSML-AADL that will be served as a support for the designers and developers of the WSN. Indeed, a WSN is based on four components:

- Nodes containing sensors.
- Wireless interconnection networks.
- A base station for grouping information and acting as an interface between the nodes and the user.
- Processing resources at the base station level to manage data correlation.

The SYSML-AADL profile must consider all the requirements related to these components (lifetime, scheduling, communication, etc.) in order to produce a complete solution which allows building a complete support for the modeling of the WSN. In this part, we will describe all these requirements which are divided into four categories:

- Requirements related to the structure of WSN.
- Requirements related to the behavior of WSN.
- Requirements related to energy constraints.
- Specific requirements.

3.2.1 Requirements Related to the Structure of WSN.

A wireless sensor network consists of a group of small wireless node networks; each node group is connected to a gateway. The gateways send the collected data to a base station via a transmission network.

In order to model this structure, our profile must support the following requirements:

- REQU STRUCT-1: The modeling of the sensor nodes and their characteristics (commercial name, operating systems supported ...) as well as the internal components that ensure the communication and the data processing (e.g., transducer, processor).
- REQU STRUCT-2: The modeling of communication links between two devices (between two sensor nodes for example) and their characteristics (error rate ...).
- REQU STRUCT-3: The modeling of sensors, gateways, and base stations and their characteristics.
- REQU STRUCT-4: The modeling of the transmission network specification.

3.2.2 Requirements Related to the Behavior of WSN

In addition to collecting and transmitting data, a wireless sensor network can also perform other tasks, such as changing behavior in the event of a failure.

- REQU BEHAVIOR-1: Modeling the process of data collection.
- REQU BEHAVIOR-2: The modeling of the maintenance process (periodic sending of messages to indicate the state of health of a sensor node).
- REQU BEHAVIOR-3: Modeling the configuration process and auto-reconfiguration.

3.2.3 Requirements Related to the Energy Constraints

Energy conservation is an important issue for WSN as the nodes contain a limited amount of energy, which cannot usually be renewed.

REQU ENERG-1: Our profile must define the modeling entities for the specification of the energy properties of the different elements of the model.

3.2.4 Requirements Related to the Specific Constraints

The scope of use of the WSN is very wide, which generates a specific constraint related either to the environment of use or to the mechanism of configuration of the network. However, our SYSML-AADL profile must be scalable and extensible to support any possible constraints.

4 Contribution: Proposal of a SysML Profile

4.1 Why a Specific Profile Definition?

As we saw in the previous section, SysML is an ideal language to serve the systems modeling as the WSNs without forgetting the AADL. We now have to make the link between these two languages. We set ourselves the objective of demonstrating the interest of such a combination by first developing a SysML profile for AADL. As we know the ideal would be to develop a DSL (Domain Specific Language), completely dedicated to the WSNs. This objective will interest us in a second time and we chose at first to start pragmatically by defining a SYSML profile. The SysML profile will solve a number of AADL language limitations:

- It will provide a development environment and graphic support for the expression of the constraints of the WSNs.
- It will provide SysML tools that are now starting to be used in industrial settings.

- It will make it possible to check the consistency of the operators defined by their implementation in OCL constraints.
- It will make the transition between the expression of needs and the analysis in an efficient way and with a minimum of information loss (seamless).

The expected progress of the project and the expected results:

The development of this profile is not the mean goal. It is part of a thesis aiming at the elaboration of a set (WSNs modeling and code generation as long as testing) of software tools for the modeling of the WSNs.

The need to develop a SysML profile requires analysis of the expression and taking into account the AADL elements. The idea is to make (starting with UML) the taxonomy of the types of association. First intra-model (relations, compositions, etc. in a class diagram for example), then inter-model (as dependencies), with in this case the mono levels of meta-modeling (M1) or multi-levels (M1/M2, for example the type conforms to).

4.2 Mapping Component Type and Component Implementation

In AADL, component types and component implementations describe the component, respectively, the visible interfaces and the internal structure. In our case, we will use the blocks for the modeling of the component's types and the components implementations of AADL. To differentiate between these last two, we will use two new stereotypes, "ComponentType" and "ComponentImpl," two SysML extensions.

The SysML blocks are extensions of the UML classes; they are chosen for the modeling of the components to describe the two structural and behavioral parts of the system, by adding only the block modeling; this will allow the use of other SysMLconstructs (ports and parts) for consistent modeling of AADL parts (subcomponents and ports).

AADL provides ten components of different categories. For each identifier in the AADL category, we reserve a stereotype with attributes representing the properties of the chosen category. The values associated with the AADL properties are automatically projected as values of the attributes of the stereotypes applied to the blocks. All category identifiers are generalized by "ComponentType" and "ComponentImpl." We apply two stereotypes for each block: a stereotype specifies whether it is ("ComponentType") or ("ComponentImpl"), and the other will specify its component categories.

In AADL, the "ComponentImpl" can refine the specification of the "CompenantType" by adding implementation details. For example, specifying the subcomponents, connections, modes, properties, and the flow.

A realization relationship in the AADL enables all the values of the "CompenantType" properties to be transferred to "ComponentImpl," and gives accessibility to all aspects and specifications of the “ComponentType” flows, from the “CompónantImpl.”

For our Model in UML/SysML, this is done by “AADL_Realization,” each relation of realization in our model must be stereotyped by “AADL_Realization.”

4.3 Application and Evaluation of the Profile

In this section, we conducted a case study based on wireless sensor network modeling using the profile proposed in the previous section. In this true study, we will evaluate the capability of this profile to meet the modeling requirements of a wireless sensor network (WSN) system consisting of two sensor nodes, a head cluster node and a base station using AADL/OSATE.

To start the study, first you need to create a “ComponentType” system type that allows to gather all the different components needed to structure the architecture of the sensor network. In our case, we will create two nodes, a CH and a BS. For example, specifying the subcomponents, connections, modes, devices, properties, and the flow. To accomplish this task, it is necessary to model the implementation part “ComponentImpl” of this “ComponentType” in order to refine and define its internal structure in terms of interactions between the different subcomponents (see Fig. 1).

After modeling the internal architecture of the system, it is important to define the functions and processes to be executed for each component of the system. To achieve this part, you need to create a new “ComponentType” type process that defines the function to perform for each component using “ComponentImpl” to refine the behavior of the component with more detail and accuracy. In our case,

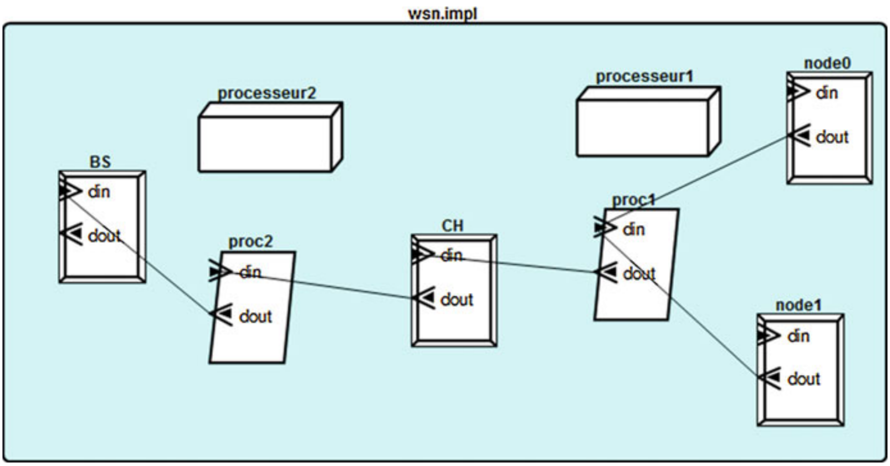


Fig. 1 Implementation of the proposed system WSN

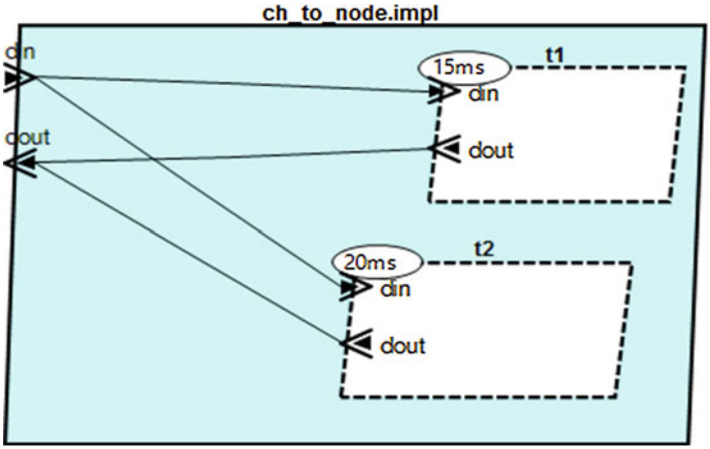


Fig. 2 (**a**) Implementation of CH processes, (**b**) implementation of BS processes

it is necessary to determine the tasks that will be executed by the two nodes, the CH and the base station without forgetting to define the technique of access control to the wireless communication medium which will allow to transmit several flows of traffic on a single channel in real time (using TDMA).

Once the processes to be performed by the WSN system have been completed, the tasks to be performed for each network component are obtained, as shown in Fig. 2.

According to the study carried out, we can conclude that the developed model based on the extension of SysML using the AADL specifications has gathered all the features and details to model the WSN system. As expected, many aspects of the SysML model required the use of AADL features because SysML proved to be insufficient and powerless.

5 Conclusion and Perspectives

The works presented in this article are "in progress." Our study of the state of the art makes it possible to show the pluridisciplinary and applied nature of the research that can be done within a framework of the modeling of the WSNs. The concrete development of an experimental platform will demonstrate the interest of the approach described and raise other questions both practical and theoretical on the development of hardware/software-based engineering models.

References

1. Kalantary, S., & Taghipour, S. (2014). A survey on architectures, protocols, applications, and management in wireless sensor networks. *Journal of Advanced Computer Science & Technology, 3*(1), 1–11.
2. Akyildiz, L. F., & Vuran, M. C. (2010). *Wireless sensor networks*. Wiley.
3. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). A survey on sensor networks. *IEEE Communications Magazine., 40*, 102–114.
4. Hilmani, A., & Maizate, A. (2017). A study of self-organization protocols in wireless sensor network. *Mediterranean Telecommunications Journal, 7*(2).
5. Behjati, R., Nejati, S., Briand, L., & Selic, B. (2011). An AADL-based SysML profile for architecture level systems engineering: Approach, metamodels, and experiments. https://www.simula.no/sites/default/files/publications/Simula.simula.380.pdf
6. SysML. http://www.sysml.org
7. Belloir, N., Bruel, J. -M., Hoang, N., & Pham, C. (2008). *Utilisation de SysML pour la modelisation des reseaux de capteurs sans fil, Conference sur les Langages et Modèles à Objets (LMO)*, Montreal, Canada, March 3–7, 2008 (pp. 171–186). http://www.cepadues.com/
8. Booch, G., Jacobson, I., & Rumbaugh, J (1996). *The unified modeling language*, Unix Review. ccs.neu.edu
9. André, C. (2007). Université Nice Sophia Antipolis, Le temps dans le profil UML MARTE. UNSA-CNRS. wwwi3s.unice.fr
10. OMG: UML profile for Modeling and Analysis of Real-Time and Embedded systems (MARTE), beta 1, August 2007, document ptc/07-08-04.
11. Chkouri, M. Y. (2010). *Modélisation des systèmes temps-réel embarqués en utilisant AADL pour la génération automatique d'applications formellement vérifiées*. hal.archives-ouvertes.fr
12. http://www.uml-sysml.org/sysml
13. Weilkiens, T. (2011). Systems engineering with SysML/UML: Modeling, analysis, design. Books.google.com
14. Feiler, P. H., Gluch, D. P., & Hudak, J. J. (2006). *The architecture analysis & design language (AADL): An introduction*. dtic.mil
15. Feiler, P. H., Lewis, B. A., & Vestal, S. (2006). The SAE Architecture Analysis & Design Language (AADL) a standard for engineering performance critical systems. In *Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control*. IEEE.
16. Hugues, J., & Singhoff, F. (2009). Développement de systèmes à l'aide d'AADL – Ocarina/Cheddar. In *Tutoriel présenté à l'école d'été temps réel*, Paris, Septembre 2009.

The Value of Simulations Characterizing Classes of Symbiosis: ABCs of Formulation Design

Khadija Basaid, Bouchra Chebli, James Furze, El Hassan Mayad, and Rachid Bouharroud

Abstract *Senecio glaucus* ssp. *coronopifolius* (Maire) and *Ridolfia segetum* (L.) are competitive-ruderal plants growing in Morocco. Their essential oils were tested for antifungal activity in vitro against *Botrytis cinerea*, agent of gray mold in plants, using volatile phase method (VF) and poisoned food method (PF). *S. glaucus* and *R. segetum* showed antifungal potential using VF method (68 and 80% inhibition at 640 μl/disc), and PF method (83 and 98% inhibition at 16 μl/ml), respectively. Accordingly, both essential oils are potential natural substitutes for chemical pesticides. Inhibition of fungal growth and simulations of synergistic interactions are facilitated by use of plant oils. Qualification of essential oils using preliminary fuzzy models specifies the groups of units, which create formulation design and prediction of inhibition. The latter can be expanded in simulations of synergy in variable scenarios.

Keywords Synergy · Simulation · Formulation · Essential oils · Fuzzy models

K. Basaid (✉) · B. Chebli
Laboratory of Mechanic Process Energy and Environment, National School of Applied Sciences, Ibn Zohr University, Agadir, Morocco
e-mail: b.chebli@uiz.ac.ma

J. Furze
Faculty of Sciences, Laboratory of Biotechnology and Valorization of Natural Resources, Ibn Zohr University, Agadir, Morocco

El H. Mayad
Laboratory of Mechanic Process Energy and Environment, National School of Applied Sciences, Ibn Zohr University, Agadir, Morocco

Faculty of Sciences, Laboratory of Biotechnology and Valorization of Natural Resources, Ibn Zohr University, Agadir, Morocco

R. Bouharroud
Integrated Crop Production Unit, Institut National de la Recherche Agronomique d'Agadir, Agadir, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_51

1 Introduction

Crop protection is constantly faced with challenges to satisfy customer demands, while promoting crop productivity and ecosystem maintenance. The search for new control agents is continuous, as crops are subject to attacks by many pathogens (fungi, bacteria, nematodes, insects, etc.). This work investigates plant essential oils for biological control of post-harvest fungi. There are two approaches to deal with this subject. The first one is a basic approach, in which we characterize plants with laboratory tests and use statistics for results interpretation. It is a limited approach in terms of mastery of interaction on different levels. Hence we complement it with a second one, a functional approach, where we apply mathematical theory and modeling to plant systems. This approach enables understanding of how the relationship works between different classes subject of the investigation, and within the classes, thus applying this knowledge for simulation of the synergy between different members of the ecosystem community. First, we discuss the use of plant essential oils for crop protection, then we give an overview of fuzzy modeling and the importance of its application in this context. Finally, we discuss the choice of plant species used, and we state objectives of the investigation.

The frequent applications of commercial fungicides can cause negative effects on the environment and food safety [1], particularly in the post-harvest period because of the short time between treatment and consumption [2]. Furthermore, it leads to the development of resistance in pathogens. The fungus, e.g., *Botrytis cinerea*, became resistant to specific fungicides in a short period of time [3]. This species belongs to *Botrytis*, a highly diverse genus, with more than 30 species that are necrotrophic pathogens decaying infected plant tissues [4]. The most common and the most important species of this genus is *Botrytis cinerea* Pers., an ubiquitous pathogen responsible for losses in over 200 crop species worldwide in pre- and post-harvest [5, 6].

The negative perception of synthetic fungicides used for control of *B. cinerea* drives attention towards natural alternatives, mainly among plant essential oils [7]. These phyto-compounds are biodegradable, cause minimal effects on non-target organisms, and delay the occurrence of resistance in pests [8]. Thus, inhibition of *B. cinerea* by essential oils made the subject of several studies [9–13], which provides motivation to investigate other plant species. Further, we wanted to expand the knowledge on these novel species by characterizing plant systems with mathematic precision, ultimately enabling with fuzzy logic methods.

For Sugeno and Yasukawa, fuzzy modeling enabled formation of a system model, with the help of a description language, which consists of fuzzy logic with fuzzy predicates, where, in a larger sense, fuzzy modeling was used to describe system behavior in a qualitative way, using a natural language. In a constricted sense, they referred the use of fuzzy modeling to describing systems with fuzzy quantities [14]. Since then, fuzzy logic has evolved. It is underpinned by set theory and, as such, is used to mathematically describe the quantification of complex design problems and form control strategies on which one bases a rule structure [15].

Seminal thinking about fuzzy modeling started with initial papers of Zadeh, as he described a human action qualitatively using fuzzy algorithms [16]. Later in 1973, he published his most remarkable work related to qualitative modeling, which established the foundation for fuzzy control. In this paper, he introduced the concept of linguistic variables and proposed to use fuzzy IF-THEN rules to formulate human knowledge [17]. Motivated by these ideas of "fuzzy algorithm" and "linguistic analysis," Mamdani applied fuzzy logic initially to control [18]. Mamdani and Assilian established the basic framework of fuzzy controller [19]. By the 1980s, Takagi, Sugeno, and Kang proposed another fuzzy system whose inputs and outputs are real-valued variables [20, 21].

Identification procedures and advanced knowledge investigation require the development of highly precise models [22]. Mathematical process models give a better insight and furthermore understanding of the complete system, some specific system behaviors or process properties through analyzing the model structures and parameters. Such knowledge gains from process models may be taken as cornerstones for future design decisions [23]. In the context of biological modeling, areas of study that need to be detailed are plant systems and process operations.

Plant species *Senecio glaucus* ssp. *coronopifolius* and *Ridolfia segetum* (L.) Moris are species of South-West Moroccan flora. They combine competitive and ruderal strategies with traits of stress tolerance. The chemical profile of these species has been described [24, 25]. There are no studies on biological activities of *S. glaucus* ssp. *coronopifolius*, and only few studies on biological properties of *R. segetum*, which reported the antioxidant, antibacterial, anti-inflammatory, and HIV-1-inhibiting activities of the essential oil [24, 26, 27]. No reports were made on antifungal activity of these oils or on simulation of their interaction with fungi. Application of mathematical theory and modeling enables simulation of the synergy between plants and different life forms. Furthermore, it will help in the design of a function that works in different scenarios, allowing prediction of inhibition with high accuracy.

Our objectives are to show antifungal potential of essential oils of *S. glaucus* ssp. *coronopifolius* and *R. segetum* against *B. cinerea* with laboratory experimentation. Further, we apply mathematical approaches towards simulation of synergy in different scenarios.

2 Materials and Methods

In this section, the laboratory tests are detailed. These include extraction of essential oils from plants, isolation of the fungus, and application of oils to it by means of two methods. Further, we introduce fuzzy inference systems, mainly the Mamdani system which we used in this work.

2.1 Experimental Work

Collection of Plants and Extraction of Essential Oils. The collection of *S. glaucus* ssp. *coronopifolius* and *R. segetum* plants from the region of Souss Massa Darâa (South-West of Morocco) was done at flowering stage. After drying, whole plants of *S. glaucus* ssp. *coronopifolius* (stems, leaves, flowers, and roots) and aerial parts (stems, leaves, flowers, and fruits) of *R. segetum* were used for extraction of essential oils. The method used was hydro-distillation via a Clevenger for 4 h. The average yield of essential oils was 0.08% for *S. glaucus* ssp. *coronopifolius* and 0.05% for *R. segetum.*

Fungal Isolation. Isolation of *B. cinerea* was performed from rotten beans infected by the fungus, by transferring small fragments of the rotten beans to sterile Petri dishes, filled with fresh potato dextrose medium (PDA) mixed with antibiotic, to prevent bacteria growth. Plates were put in incubation at 25 °C during 7 days. After series of purification, we attained a pure culture of the fungus, which was maintained on PDA at 4 °C.

Antifungal Activity Assay. Two methods were used to test antifungal activity of essential oils against the fungus. They are the poisoned food method (PF) [28] and volatile phase method (VF) [29], with some variations. The PF technique requires the mixture of essential oils, emulsified by Tween 80 with PDA, right before being emptied into Petri dishes (9 cm diameter). The concentrations tested ranged from 0.25 to 16 μl/ml. The controls were filled with PDA alone. Moreover, the tested fungi was inoculated using a 6 mm mycelial plug from a 7-day-old culture. Plates were incubated for 7 days at 25 ± 2 °C (Fig. 1a). In VF, Petri dishes (9 cm) filled with PDA and inoculated with the fungus were turned upside down, inverted lids had sterile Whatman No. 1 filter paper discs, which contained different amounts of essential oils (10, 20, 40, 80, 160, 320, and 640 μl/disc). The controls had sterilized paper filter discs impregnated with distilled water. Plates were incubated at 25 ± 2 °C during 7 days (Fig. 1b).

According to the two techniques described above, for each treatment, three replicate plates were inoculated. Fungus growth was determined by measuring the surface covered by the mycelium on petri dishes. Antifungal activity of essential oils is expressed by inhibition percentage of mycelial growth (*I%*) and evaluated according to the formula (1) of [30]:

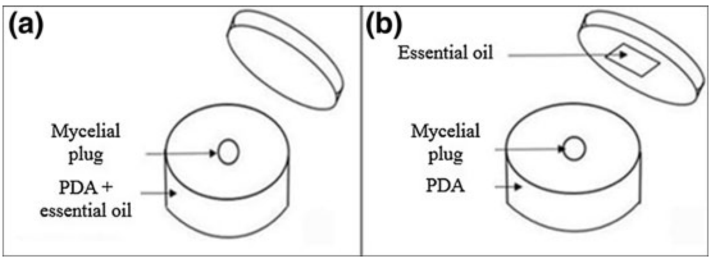


Fig. 1 Poisoned food and volatile phase method

$$I\% = (\mathrm{Dt} - \mathrm{Di}) * 100 \tag{1}$$

Dt represents mycelial growth diameter in control Petri plates and Di represents mycelial growth diameter in treated ones.

Statistical Analysis. Growth inhibition data were submitted to analysis of variance (ANOVA). Mean values were compared using Tukey's test (also called Tukey's Honest Significant Difference) ($P < 0.05$). The latter compares specific groups' means with each other, to bring out different ones. For each pair of means, we determine HSD value, using formula (2):

$$\mathrm{HSD} = (\mathrm{Mi} - \mathrm{Mj}) / \surd (\mathrm{MSw}/nh) \tag{2}$$

where:

- Mi and Mj are two different means, their subtraction is the difference between the pair of means.
- MSw represents the mean square within the pair of means.
- n stands for number in the group.

Step 1 of Tukey's test is to perform the ANOVA test. If we find a significant F, we move to step 2, where we choose two means from the ANOVA output. Step 3 is to calculate the HSD value for the Tukey test with the help of formula (2) showed above.

Step 4 is to find the HSD value among values of Tukey's critical value table. Step 5 is to compare the HSD value calculated in Step 3 with the critical value found in Step 4. The two means are considered significantly different, if the calculated value is bigger than the critical value. Further, we calculated the IC_{50} value (concentration required for inhibition of 50% of mycelial growth). We calculated log values of essential oil concentrations, then determined their probit with inhibition percentage of the test fungus, using the linear regression. Calculated IC_{50} value represented the average of three replications.

2.2 Fuzzy Inference Systems

Fuzzy systems have a knowledge base consisting of fuzzy IF-THEN rules. These are statements which are characterized by continuous membership functions [31]. In terms of inference process, there are two main types of Fuzzy Inference Systems (FIS): the Mamdani-type [18] and the TSK-type [20, 21]. In this investigation, we will focus on application of the first type.

Bayesian inference systems show the use of fuzzy logic to qualify modeling frameworks. In terms of use, the Mamdani FIS is widely used; it gives practical results with a more simple structure, and has an intuitive and interpretable nature in

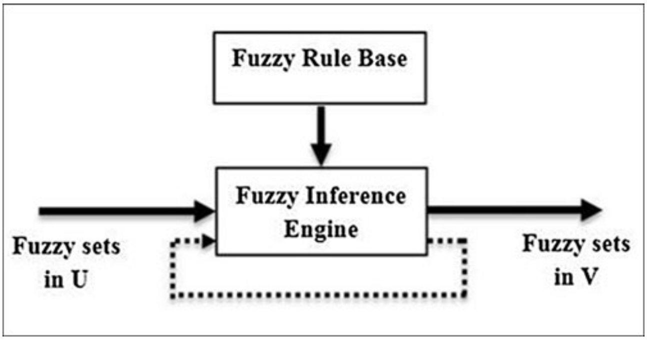


Fig. 2 Process of Mamdani system [31]

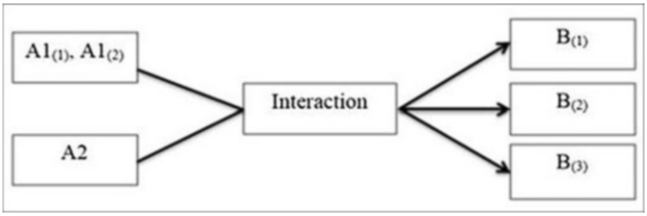


Fig. 3 Qualitative determination of inhibition

its rule base [32]. They can be defined with fuzzy inference and fuzzy consequence. A basic configuration of a Mamdani system is shown in Fig. 2.

2.2.1 Input space U C R^n, Output space V C R

Collection of IF-THEN rules (fuzzy rules) constitute the rule base. The inputs and outputs are fuzzy sets. The fuzzy inference engine uses fuzzy logic principles in combining fuzzy rules, to plot from fuzzy input sets (U C R^n) to fuzzy output sets (V C R) [31]. The same system has been used to calculate plant strategies [33–35]. If the dashed feedback line in Fig. 2 exists, the system becomes the so-called fuzzy dynamic system [31]. Application of a Mamdani system in this investigation is shown in Fig. 3.

This system has low-resolution input. A1(1) and A1(2) are essential oils of *S. glaucus* and *R. segetum*, respectively. A2 is the fungus *B. cinerea*. The interaction between input variables is inhibition. The latter varies from no inhibition B(1), to medium B(2) or maximum inhibition B(3) depending on oil doses. This system can be projected to various scenarios. Accordingly, candidates of input variables are different types of oils, different plants and may also include different temperatures or environments. The output are a fungus, different types of fungus or different types of life forms (insects, bacteria, nematodes, etc.). The operation describes the interaction.

3 Results

3.1 Antifungal Activity of Essential Oils

Table 1 presents the effects of *S. glaucus* ssp. *coronopifolius* and *R. segetum* essential oils concentrations on mycelial growth after 7-day incubation period of *B. cinerea* at 25 ± 2 °C, using the PF technique.

Inhibition percentage increases with concentration. Its maximum value reached 83% at 16 μl/ml for *S. glaucus* ssp. *coronopifolius*, and 98% at 16 μl/ml for *R. segetum*. IC_{50} values are 7.87 and 2.70 μl/ml, respectively (Figs. 4 and 5). Using the micro-atmosphere technique, the results of the effect of essential oils vapors on the growth of *B. cinerea* mycelium after incubation at 25 ± 2 °C for 7 days, are presented in Table 2.

Maximum inhibition achieved by *S. glaucus* ssp. *coronopifolius* is 68% at 640 μl/disc, whereas *R. segetum* provided 80% inhibition of the fungus at both concentrations 640 μl/disc and 320 μl/disc (Figs. 6 and 7).

The distribution of results obtained with volatile phase method (VF) is a Poisson distribution, whereas results obtained with poisoned food method (PF) have a sigmoid distribution. This difference in distribution reflects on difference in resolution. Thus PF method has low resolution and VF offers higher resolution. Hence, for the following section, we will use results of VF method for application of a high-resolution system.

3.2 Application of a Mamdani System

Figure 8 presents a higher resolution system based on the low-resolution system showed in Fig. 3. Given that chemical characterization of the oils used in this study has yet to be determined, we referred to references on similar oils to determine major

Table 1 Inhibition percentages of *B. cinerea* growth at various concentrations of the two essential oils using the PF technique

Concentration (μl/ml)	Inhibition (%)	
	S. glaucus	*R. segetum*
0	0.0 a	0.0 a
0.25	0.0 a	0.0 a
0.5	0.0 a	0.0 a
1	0.0 a	0.0 a
2	0.0 a	33 b
4	13 b	78 c
8	60 c	79 c
16	83 d	98 d

Values of the same column followed by the same letter are not different in a significant way according to Tukey's test ($p \leq 0.01$)

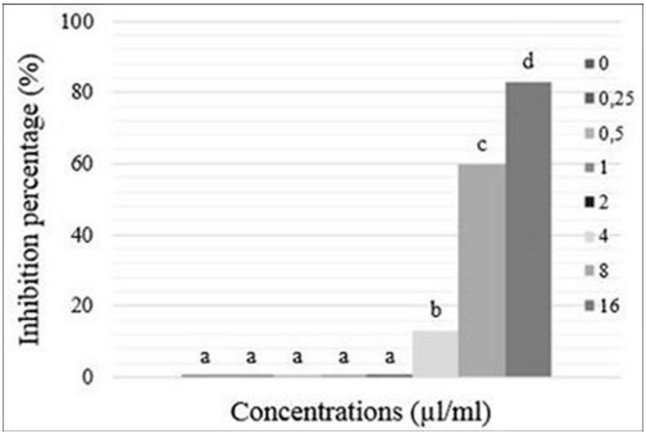


Fig. 4 Inhibition percentages of essential oil of *S. glaucus* ssp. *coronopifolius* against *B. cinerea* at different concentrations using poisoned food method, values assigned the same letter do not differ significantly, according to Tukey's test ($p \leq 0.01$)

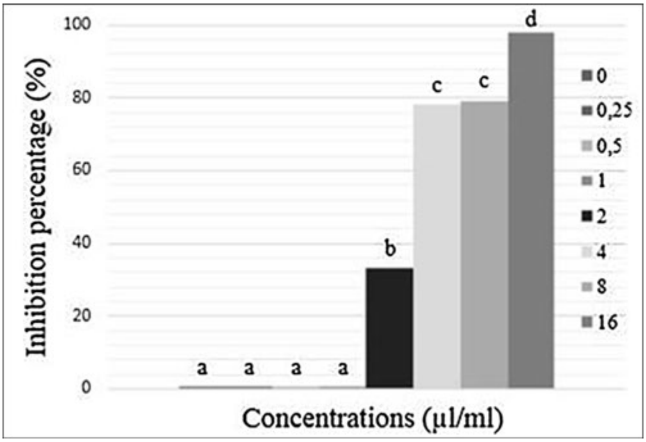


Fig. 5 Inhibition percentages of *R. segetum* essential oil against *B. cinerea* at different concentrations using poisoned food method, values assigned the same letter do not differ significantly, according to the test of Tukey ($p \leq 0.01$)

compounds of oils, and to attribute to each major compound a percentage in which it is present in oils [24–26, 36]. Letters from A to G represent major compounds of essential oils (myrcene, dehydrofukinone, α-phellandrene, *p*-cymene, terpinolene, dillapiol, and myristin) and H stands for *B. cinerea*. The quantity in oil attributed to each compound is the result of Eq. (3):

$$Q = P * \mathrm{OQ} \tag{3}$$

Table 2 Inhibition percentages of *B. cinerea* growth at various concentrations of the two essential oils using the VF technique

Concentration (µl/disc)	Inhibition (%)	
	S. glaucus	*R. segetum*
0	0.0 a	0.0 a
10	0.0 a	0.0 a
20	0.0 a	0.0 a
40	0.0 a	0.0 a
80	0.0 a	0.0 a
160	0.0 a	0.0 a
320	0.0 a	80 b
640	68 d	80 b

Values of the same column followed by the same letter are not different in a significant way according to Tukey ($p \leq 0.01$)

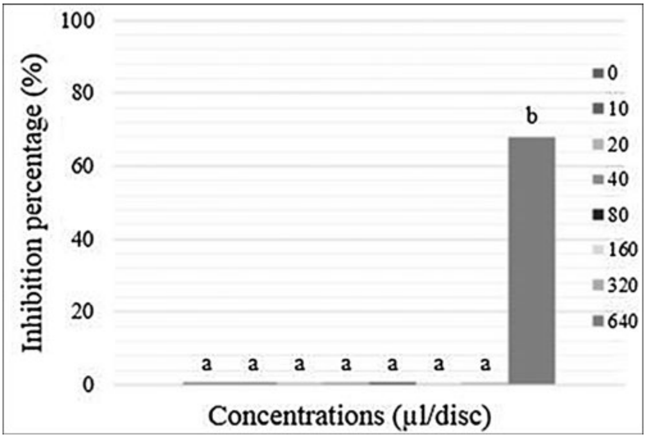


Fig. 6 Inhibition percentages of *S. glaucus* ssp. *coronopifolius* essential oil against *B. cinerea* at different concentrations using volatile phase method, values assigned the same letter do not differ significantly, according to the test of Tukey ($p \leq 0.01$)

where *Q* is the quantity in oil required for each compound, *P* is the percentage in which the compound is present in the oil, and OQ is the quantity in oil applied to achieve inhibition.

The percentages of major compounds in oils according to the literature are as follows:

P (myrcene) = 24% (maximum value) [25].
P (dehydrofukinone) = 21% (maximum value) [25].
P (α-phellandrene) = 58% (average of the range 53.0–63.3%) [24].
P (*p*-cymene) = 12% (average of the range 8.4–15.2%) [36].
P (terpinolene) = 10% (average of the range 11.9–8.6%) [24].
P (dillapiol) = 47% (maximum value) [26].

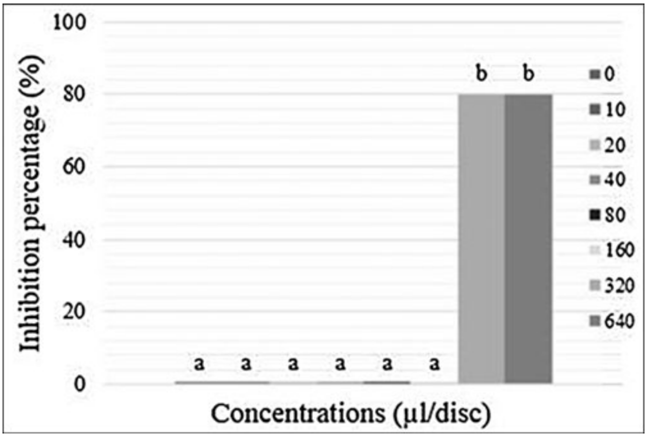


Fig. 7 Inhibition percentages of *R. segetum* essential oil against *B. cinerea* at different concentrations using volatile phase method, values assigned the same letter do not differ significantly, according to the test of Tukey ($p \leq 0.01$)

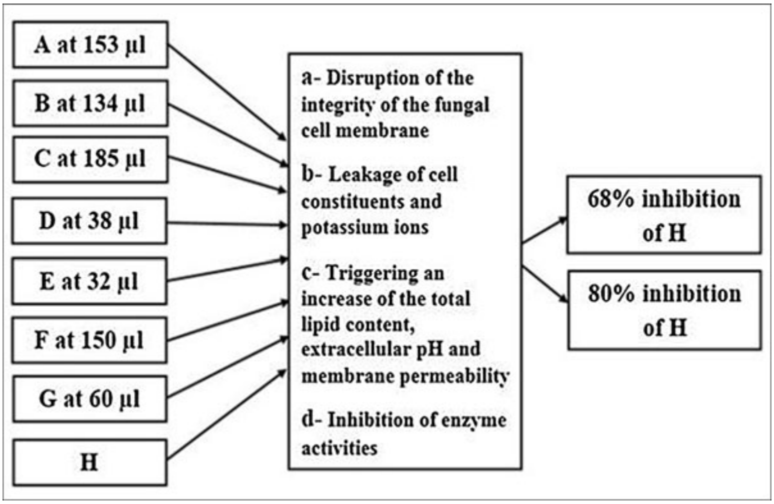


Fig. 8 Identification of groups participating in inhibition

P (myristin) = 19% (maximum value) [26].

For values of quantity in oil (OQ), we will use those found with VF method (high resolution): 640 µl for compounds myrcene and dehydrofukinone present in *S. glaucus* oil, and 320 µl for the rest of the compounds (α-phellandrene, *p*-cymene, terpinolene, dillapiol, and myristin) which are present in *R. segetum* oil. The input of this system is major compounds of oils (A to G). The operation is different

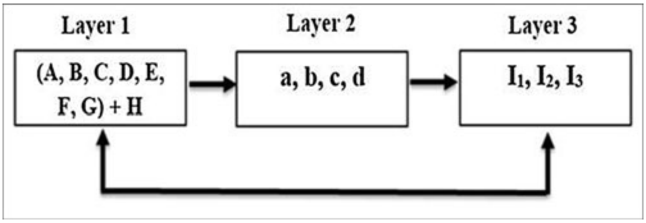


Fig. 9 Process of inhibition

mechanisms of action of these compounds on the fungus (H). The outcome is the inhibition percentage of the fungus.

For a more simplified system structure, we propose the process presented in Fig. 9.

The input of the system are major compounds of oils (A to G). The operation is their mechanisms of action (a to d). The output of the system is inhibition of the fungus: I1 is inhibition 1 (68% inhibition of H), I2 is inhibition 2 (80% inhibition of H), and I3 is no inhibition (0% inhibition of H). The arrow, which links layer 1 with layer 3, is indicative of a dynamic relation. The interaction between members of layer 1 creates feedback in the form of inhibition. Increase in layer 1 produces increase in layer 3, and a decrease in layer 3 has a feed forward to layer 1. This process hints at Eqs. (4–6), by which we can design a formulation that is effective on the fungus.

$$\mathrm{A} * 153 + \mathrm{B} * 134 + \mathrm{H} = 68\% \downarrow \mathrm{H} \tag{4}$$

$$\mathrm{C} * 185 + \mathrm{D} * 38 + \mathrm{E} * 32 + \mathrm{F} * 150 + \mathrm{G} * 60 + \mathrm{H} = 80\% \downarrow \mathrm{H} \tag{5}$$

$$(\mathrm{A} + \mathrm{B} + \mathrm{C}) * 0 + (\mathrm{D} + \mathrm{E} + \mathrm{F} + \mathrm{G}) * 0 + \mathrm{H} = \mathrm{H} \tag{6}$$

Equation (4) is built on the results obtained by *S. glaucus* essential oil. In the equation, the oil is replaced by its major compounds myrcene (A) and dehydrofukinone (B), multiplied by the quantities in which they can be used. With the addition of the fungus, the result for this equation is inhibition of the fungus by 68%.

Equation (5) is based on the results obtained by *R. segetum* essential oil. In the equation, the oil is replaced by its major compounds α-phellandrene (C), *p*-cymene (D), terpinolene (E), dillapiol (F), and myristin (G), multiplied by the quantities in which they can be used. With the addition of the fungus, the result for this equation is inhibition of the fungus by 80%.

Equation (6) combines major compounds of *S. glaucus* and *R. segetum* essential oils, all multiplied by 0. With the addition of the fungus, the result for this equation is the presence of the fungus, which indicates that in the absence of compounds of oils, there is no effect on the fungus.

4 Discussion

Oils of both species studied inhibit *B. cinerea* and are optional antifungal agents. Yet, *R. segetum* (L.) Moris provides greater inhibition of the fungus than *S. glaucus* ssp. *coronopifolius*. The methods used infer a difference in the chemical profiles of the oils. Major components of *S. glaucus* ssp. *coronopifolius* oil are: myrcene (24%) and dehydrofukinone (21%) which gives the oil its distinct odor [25]. These two compounds have not been reported as antifungal against *B. cinerea*. Yet, myrcene showed antifungal activity against *Fusarium oxysporum* [37]. A study by [38] showed that dehydrofukinone had great antifungal activity against *Pycnoporus sanguineus* and *Gloeophyllum trabeum*.

R. segetum produces two types of oils: those that contain monoterpene hydrocarbons as major compounds (α-phellandrene, *p*-cymene, and terpinolene) [24, 27, 36, 39–41], and those that additionally contain phenylpropanoids like myristicin and dillapiol, as major compounds or in considerable amounts [26, 42]. *P*-cymene has previously showed good antifungal activity against *B. cinerea* [43], and other fungi such as *Pythium ultimum, Fusarium acuminatum, Aspergillus niger, Fusarium solani, P. digitatum, R. solani, F. oxysporum, Verticillium dahliae, Alternaria mali, A. flavus, Ammophilus fumigatus,* and *Penicillium* sp. [37, 44–46]. Terpinolene showed antifungal action against *B. cinerea* [47], and other fungi such as *Aspergillus niger, A. flavus, A. fumigatus,* and *Penicillium* sp. [44]. α-phellandrene is reported to have great antifungal activity against *Penicillium cyclopium* [48]. Myristicin was shown to have antifungal action on *Aspergillus flavus* and *Aspergillus ochraceus* [49], and dillapiol showed great antifungal activity against *Colletotrichum acutatum, Botryodiplodia theobromae,* and *Clinipellis perniciosa* [50, 51].

Given the diversity of molecules present in both essential oils, antifungal activity seems to result from a combination of several modes of action, involving different cellular targets: For example, α-phellandrene inhibits the growth of *Penicillium cyclopium* by acting on membrane permeability. That leads to leakage of cell constituents, and increase of the total lipid content [48]. Yu et al. [46] suggested that *p*-cymene gains antifungal activity by inhibiting enzyme activities, as it was shown to have inhibitory activity on pectin methyl esterase isolated from *A. solani* and *P. infestans*, and cellulase isolated from *P. infestans* and *A. niger*. Dillapiol disrupts the integrity of cell membranes. Due to its lipophilic characteristic, it interferes with fatty-acid-chain constituents of the membrane [50].

Application of a basic Mamdani system is the initial step towards prediction of inhibition. To characterize the interaction going on between oils and fungi, we used a high-resolution Mamdani system. The system presented in Fig. 8 shows possible mechanisms of action of major compounds of oils tested in fungus inhibition. This system has applications in crop protection. Figure 9 presents a simplified structure based on Fig. 8, which gives a hint on the estimation of a recipe for a formulation (Eqs. 3–5), which can be applied against the fungus. This formulation would need to be further tested directly on plants infected with the fungus to ensure its effectiveness and lack of phytotoxicity on plants. Furthermore, plants in their ecosystem are not in seclusion, they are surrounded by other pathogenic life forms (insects, bacteria, nematodes, etc.) and are subject to their attacks, which opens up our minds to other scenarios where essential oils can have a key role in crop protection, thanks to their chemical diversity. Thus it is very interesting to test inhibitory effects of essential oil compounds on them to verify those ideas. I am currently exploring the effects of these oils on plant parasitic insects and nematodes, which enables the formation of broader functional scenarios that have ecosystem application.

Determination of rules for chemical profiles of oils and their outcome of inhibition on different life forms allows creation of formulas that possess several targets, which will promote plant growth, and also benefit plant ecology, as essential oils don't have effects on non-target organisms which include beneficial organisms that live in symbiosis with plants in the ecosystem.

The next step after qualifying is quantifying (measurement) with quantifiable modeling structure, otherwise known as quantified hybrid systems that use higher mathematic. Higher mathematic approaches rely on centralized means and unit variance. Further they make use of probability and distribution functions. Quantifiable modeling with defined variables and measured resolution produces accurate results, and they can be further expanded through other mathematic approaches with 100% certainty [52]. Application of higher mathematic method involves the formation of a framework for synergy, in which case we are considering different memberships (plants, fungi, nematodes, insects, bacteria, and environment). Figure 10 shows a low-resolution system that makes a framework for synergy in different scenarios between different life forms.

The input of this system is plants (P), environment (E), and different life forms (fungi (F), nematodes (N), insects (I), and bacteria (B)) present in plant ecosystem. The interaction between these members is antagonism, symbiosis, and

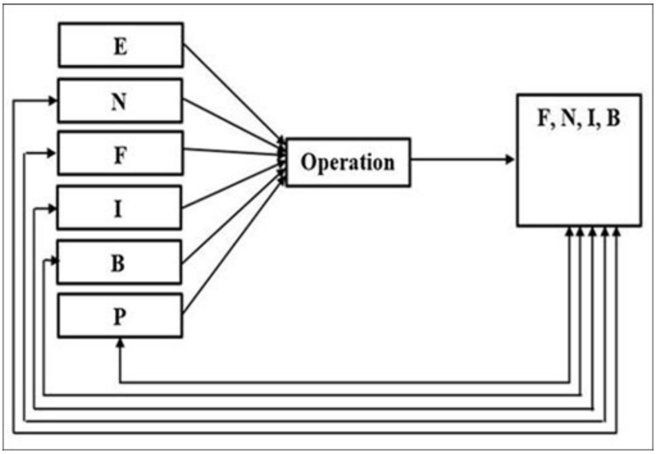


Fig. 10 Simulation of synergy in variable scenarios

commensalism. The output of the system is life forms in increase, decrease, or intact. This system is a dynamic system as the double arrows show feedback and feed forward happening between classes of interactions and between different organisms.

Measurement and specific resolution of memberships enable formation of quantifiable systems (Takagi-Sugeno-Kang systems) [20, 21]. The latter enable simulation of synergy between and within membership components. TSK systems can be expanded through further mathematic approaches with 100% certainty, creating a function for prediction of inhibition in variable scenarios. Thus, preliminary results presented in this article must be further expanded to determine the details for precise calculation of the interaction via Takugi-Sugeno-Kang model structure.

5 Conclusion

This study validates antifungal activity of essential oils of two Moroccan competitive-ruderal plants against the gray mold disease, in vitro. Examination with various concentrations of both oils exhibited promising prospects for their utilization against the fungi tested. However, further work is required to indicate the difference in the mechanism of action between the essential oils. This mechanism may be concentration dependent or involve additional metabolites. Simulation of the interaction between plants and different life forms guides the design of plant-based chemical formulations that can promote crop productivity, maintain ecosystem structure and thus maintain and increase diversity within the ecosystem.

References

1. Norman, C. (1988). EPA sets new policy on pesticides cancer risks. *Science, 242*, 366–368.
2. Chebli, B., Hmamouchi, M., Achouri, M., & Idrissi Hassani, L. M. (2004). Composition and *in vitro* fungitoxic activity of 19 essential oils against two post-harvest pathogens. *Journal of Essential Oil and Research, 16*, 507–511.
3. Elad, Y., Yunis, H., & Katan, T. (1992). Multiple resistance to benzimidazoles dicarboximides and diethofencarb in field isolates of *Botrytis cinerea* in Israel. *Plant Pathology, 41*, 41–46.
4. Fillinger, S., & Elad, Y. (2016). *Botrytis – The fungus, the pathogen and its management in agricultural systems*. Springer.
5. Williamson, B., Tudzynski, B., Tudzynski, P., & Van Kan, J. A. L. (2007). *Botrytis cinerea*: the cause of grey mould disease. *Molecular Plant Pathology, 8*, 561–580.
6. Chebli, B., Achouri, M., Idrissi Hassani, L. M., & Hmamouchi, M. (2003). Chemical composition and antifungal activity of essential oils of seven Moroccan labiatae against *Botrytis cinerea* Pers: Fr. *Journal of Ethnopharmacology, 89*, 165–169.
7. Sharma, N., & Tripathi, A. (2008). Effects of *Citrus sinensis* (L.) Osbeck epicarp essential oil on growth and morphogenesis of *Aspergillus niger* (L.) Van Tieghem. *Microbiological Research, 163*, 337–344.
8. Isman, M. B. (2000). Plant essential oils for pest and disease management. *Crop Protection, 19*, 603–608.

9. Cid-Pérez, T. S., Torres-Muñoz, J. V., Nevárez-Moorillón, G. V., Palou, E., & López-Malo, A. (2016). Chemical characterization and antifungal activity of *Poliomintha longiflora* Mexican oregano. *Journal of Essential Oil Research, 28*, 157–165.
10. Fraternale, D., Flamini, G., & Ricci, D. (2016). Essential oil composition of *Angelica archangelica* L. (Apiaceae) roots and its antifungal activity against plant pathogenic fungi. *Plant Biosystem, 150*, 558–563.
11. El Ouadi, Y., Manssouri, M., Bouyanzer, A., Majidi, L., Bendaif, H., Elmsellem, H., Shariati, M. A., Melhaoui, A., & Hammouti, B. (2017). Essential oil composition and antifungal activity of *Melissa officinalis* originating from North-East Morocco, against postharvest phytopathogenic fungi in apples. *Microbial Pathogenesis, 107*, 321–326.
12. Xueuan, R., Dandan, S., Zhuo, L., & Qingjun, K. (2018). Effect of mint oil against *Botrytis cinerea* on table grapes and its possible mechanism of action. *European Journal of Plant Pathology, 151*, 321–328.
13. Banani, H., Olivieri, L., Santoro, K., Garibaldi, A., Gullino, M. L., & Spadaro, D. (2018). Thyme and savory essential oil efficacy and induction of resistance against *Botrytis cinerea* through priming of defense responses in apple. *Food, 7*, 11.
14. Sugeno, M., & Yasukawa, T. (1993). A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems, 1*, 7–31.
15. Furze, J. N. (2014). *Global plant characterisation and distribution with evolution and climate*. PhD Thesis. https://eprints.uwe.ac.uk/secure/23121/
16. Zadeh, L. A. (1968). Fuzzy algorithms. *Information and Control, 12*, 94–102.
17. Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics, 3*, 28–44.
18. Mamdani, E. H. (1974). Application of fuzzy algorithms for control of simple dynamic plant. *Proceedings of the IEEE, 121*, 1585–1588.
19. Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies, 7*, 1–13.
20. Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics, 15*, 116–132.
21. Sugeno, M., & Kang, G. T. (1988). Structure identification of fuzzy models. *Fuzzy Sets and Systems, 28*, 15–33.
22. Lughofer, E. (2011). *Evolving fuzzy systems-methodologies, advanced concepts and applications*. Springer.
23. Lin, C. C., Segel, L. A., & Handelman, G. H. (1974). *Mathematics applied to deterministic problems in the natural sciences*. Macmillan Publishing.
24. Cabral, C., Poças, J., Gonçalves, M. J., Cavaleiro, C., Cruz, M. T., & Salgueiro, L. (2015). *Ridolfia segetum* (L.) Moris (Apiaceae) from Portugal: A source of safe antioxidant and anti-inflammatory essential oil. *Industrial Crops and Products, 65*, 56–61.
25. De Pooter, H. L., De Buyck, L. F., Schamp, N. M., Aboutabl, E., De Bruyn, A., & Husain, S. Z. (1986). The volatile fraction of *Senecio glaucus* subsp. *coronopifolius*. *Flavour and Fragrance Journal, 1*, 159–163.
26. Jabrane, A., Ben Jannet, H., Mastouri, M., Mighri, Z., & Casanova, J. (2010). Chemical composition and in vitro evaluation of antioxidant and antibacterial activities of the root oil of *Ridolfia segetum* (L.) Moris from Tunisia. *Natural Product Research, 24*, 491–499.
27. Bicchi, C., Rubiolo, P., Ballero, M., Sanna, C., Matteodo, M., Esposito, F., Zinzula, L., & Tramontano, E. (2009). HIV-1-inhibiting activity of the essential oil of *Ridolfia segetum* and *Oenanthe crocata*. *Planta Medica, 75*, 1331–1335.
28. Rhayour, K., Bouchikhi, T., Tantaoui-Elaraki, A., Sendide, K., & Remmal, A. (2003). The mechanism of bactericidal action of oregano and clove essential oils and of their phenolic major components on *Escherichia coli* and *Bacillus subtilis*. *Journal of Essential Oil Research, 15*, 356–362.
29. Soylu, E. M., Kurt, Ş., & Soylu, S. (2010). *In vitro* and *in vivo* antifungal activities of the essential oils of various plants against tomato grey mold disease agent *Botrytis cinerea*. *International Journal of Food Microbiology, 143*, 183–189.

30. Pandey, D. K., Tripathi, N. N., Tripathi, R. D., & Dixit, S. N. (1982). Fungitoxic and phytotoxic properties of the essential oil of *Hyptis suaveolens*. *Journal of Plant Diseases and Protection, 89*, 344–349.
31. Wang, L.-X. (1997). *A course in fuzzy systems and control*. PTR Prentice-Hall.
32. Mendel, J. M. (2001). *Uncertain rule-based fuzzy logic systems: introduction and new directions*. Prentice Hall.
33. Grime, J. P. (1979). *Plant strategies and vegetation processes*. Wiley.
34. Furze, J. N., Zhu, Q. M., Qiao, F., & Hill, J. (2013). Linking and implementation of fuzzy logic control to ordinate plant strategies. *International Journal of Modelling, Identification and Control, 19*, 333–342.
35. Furze, J. N., Zhu, Q., Hill, J., & Qiao, F. (2017). Biological modelling for sustainable ecosystems. In J. N. Furze, K. Swing, A. K. Gupta, R. H. McClatchey, & D. M. Reynolds (Eds.), *Mathematical advances towards sustainable environmental systems* (pp. 9–42). Springer.
36. Fleisher, Z., & Fleisher, A. (1996). Volatiles of leaves and flowers of *Ridolfia Segetum* L. Moris. Aromatic plants of the holy land and the Sinai. Part XII. *Journal of Essential Oil and Research, 8*, 189–191.
37. Sekine, T., Sugano, M., Majid, A., & Fujii, Y. (2007). Antifungal effects of volatile compounds from black zira (*Bunium persicum*) and other spices and herbs. *Journal of Chemical Ecology, 33*, 2123–2132.
38. Silva, D. T., Bianchini, N. H., Muniz, M. F. B., Heinzmann, B. M., & Labidi, J. (2016). Chemical composition and inhibitory effects of *Nectandra Grandiflora* leaves essential oil against wood decay fungi. *Drewno., 59*, 189–198.
39. Marongiu, B., Piras, A., Porcedda, S., Tuveri, E., & Maxia, A. (2007). Comparative analysis of the oil and supercritical CO_2 extract of *Ridolfia segetum* (L.) Moris. *Natural Product Research, 21*, 412–417.
40. Palá-Paúl, J., Velasco-Negueruela, A., Pérez-Alonso, M. J., & Ramos-Vázquez, P. (2002). Volatile constituents of *Ridolfia segetum* (L.) Moris gathered in southern Spain, Andalucia Province. *Journal of Essential Oil Research, 14*, 206–209.
41. Palá-Paúl, J., Velasco-Negueruela, A., Pérez-Alonso, M. J., & Vallejo, M. C. G. (2005). Volatile constituents of *Ridolfia segetum* (L.) Moris gathered in Central Spain: Castilla la Mancha Province. *Journal of Essential Oil and Research, 17*, 119–121.
42. Jannet, H. B., & Mighri, Z. (2007). Hydrodistillation kinetic and antibacterial effect studies of the flower essential oil from the Tunisian *Ridolfia segetum* (L.). *Journal of Essential Oil Research, 19*, 258–261.
43. Nabigol, A., & Morshedi, H. (2011). Evaluation of the antifungal activity of the Iranian thyme essential oils on the postharvest pathogens of strawberry fruits. *African Journal of Biotechnology, 10*, 9864–9869.
44. Hammer, K. A., Carson, C. F., & Riley, T. V. (2003). Antifungal activity of the components of *Melaleuca alternifolia* (tea tree) oil. *Journal of Applied Microbiology, 95*, 853–860.
45. Kordali, S., Cakir, A., Ozer, H., Cakmakci, R., Kesdek, M., & Mete, E. (2008). Antifungal, phytotoxic and insecticidal properties of essential oil isolated from Turkish *Origanum acutidens* and its three components, carvacrol, thymol and p-cymene. *Bioresource Technology, 99*, 8788–8795.
46. Marei, G. I. K., & Abdelgaleil, S. A. M. (2018). Antifungal potential and biochemical effects of monoterpenes and phenylpropenes on plant pathogenic fungi. *Plant Protection and Science, 54*, 9–16.
47. Yu, D., Wang, J., Shao, X., Xu, F., & Wang, H. (2015). Antifungal modes of action of tea tree oil and its two characteristic components against *Botrytis cinerea*. *Journal of Applied Microbiology, 119*, 1253–1262.
48. Zhang, J.-H., Sun, H.-L., Chen, S.-Y., Zeng, L., & Wang, T.-T. (2017). Anti-fungal activity, mechanism studies on α-Phellandrene and nonanal against *Penicillium cyclopium*. *Botanical Studies, 58*, 13.
49. Valente, V. M. M., Jham, G. N., Jardim, C. M., Dhingra, O. D., & Ghiviriga, I. (2015). Major antifungals in Nutmeg essential oil against *Aspergillus flavus* and *A. ochraceus*. *Journal of Food Research, 4*, 51–57.

50. De Almeida, R. R. P., Souto, R. N. P., Bastos, C. N., Da Silva, M. H. L., & Maia, J. G. S. (2009). Chemical variation in *Piper aduncum* and biological properties of its dillapiole-rich essential oil. *Chemistry and Biodiversity, 6*, 1427–1434.
51. Vizcaíno-Páez, S., Pineda, R., García, C., Gil, J., & Durango, D. (2016). Metabolism and antifungal activity of saffrole, dillapiole, and derivatives against *Botryodliplodia theobromae* and *Colletotrichum acutatum*. *Boletín Latinoam. y del Caribe Plantas Med. y Aromáticas, 15*, 1–17.
52. Furze, J. N., Zhu, Q., Qiao, F., & Hill, J. (2013). Mathematical methods to quantify and characterize the primary elements of trophic systems. *International Journal of Computer Applications in Technology, 47*, 314–325.

Propagation of the Electromagnetic Waves in Asymmetric Defective One-Dimensional Photonic Comb-Like Structure

Younes Errouas, Youssef Ben-ali, Abdelaziz Ouariach, Zakaria Tahri, and Driss Bria

Abstract In this work, we report the existence of pass bands (transmission bands) and large forbidden bands (gaps) for the comb-like structure composed of a periodicity of segments along a finite monomode waveguide with two different lateral branches of different lengths, which play the role of the resonators, grafted onto *N* equidistant sites (also called nodes). These gaps come not only from the periodicity of the system but also from the resonance states of the two grafted branches (eigen modes of the lateral branches). The width of these band gaps is very sensitive to the ratio between the lengths of the lateral branches located at the same site. In the other hand, the presence of geometrical defects in the lateral branches (resonators) situated in the j site of the comb-like structure may give rise to one or more localized states (defects modes) in the gaps. These states are very sensitive to the defects lengths and the lengths of the undisturbed lateral branches located in the same site. The result of this article shows that we can obtain two defects modes in the gaps with a high transmission coefficient and a good quality factor *Q*. Hence, this structure can be used in the electromagnetic telecommunications field in the design of new filters with performance.

Keywords Comb-like · Asymmetric · Photonic · Defects · Filters

Y. Errouas (✉) · A. Ouariach · D. Bria
Laboratory of Materials, Waves, Energy and Environment, Team of Acoustic, Photonic and Materials, Faculty of Sciences, University Mohamed First, Oujda, Morocco

Y. Ben-ali
Engineering Sciences Laboratory (LSI), Multidisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University, Taza Gare, Morocco

Z. Tahri
Laboratory of Materials, Waves, Energy and Environment, Team of Acoustic, Photonic and Materials, Faculty of Sciences, University Mohamed First, Oujda, Morocco

Laboratory of Applied Sciences, National School of Applied Sciences of Al-Hoceima, Al-Hoceima, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_52

1 Introduction

The initial research on high-transmission comb-like (HTC) filters, some of which has appeared in reports and patents [1, 2], was motivated by a need for filters that can be electrically tuned to block the transmission of any narrow spectral line within a band of wavelengths. Optical combs are also of tremendous interest for other applications, such as multi-wavelength coherent light wave communications, optical arbitrary waveform generation, generation of low-phase noise or agile ultra-broadband microwaves, and signal processing [3–7]. Here, we report their significant advances in photonic crystal-based filters made possible by the ability to quickly adjust the comb's timing and shape its power spectrum (transmission coefficient).

One-dimensional photonic comb-like structure also called star waveguides structure is a system composed of an infinite or finite segment (also called waveguide) along which stars of N' finite lateral branches (which play the role of the resonators) are grafted at N equidistant sites, N and N' being integers [8–11]. The important property of this structure is the presence of narrow pass bands (transmission bands) separated by large photonic forbidden bands (gaps) where the electromagnetic wave cannot propagate [12], analogs for the electronic gaps in the band structure of semiconductor crystal [13, 14]. Wide band gaps and narrow pass bands can be obtained by an appropriate choice of the parameters, in particular the ratio between the two characteristic lengths of the segment and resonator, but more especially by grafting several dangling lateral branches N' at every site N [10]. The band gaps exhibit extraordinary dispersion characteristics, which allow manipulating and controlling the flow of electromagnetic waves and can lead to many potential applications in field of photonics and electromagnetics. These band gaps are very wider than band gaps in the photonic crystals [15–17].

The physical characteristics of this comb-like structure are the periodicity, which is the distance between the two sites, the length of each grafted branch and the relative permittivity of the constituent materials. In this 1D comb-like waveguides structure, it is emphasized that the diameter of the waveguide should be much smaller than the wavelength in order to allow the propagation of single mode through the waveguides [8]. On the other hand, it has been shown that the presence of defect surface (defect localized between the first substrate and the structure) or defect segment (defect can move inside structure) in the comb-like structure may give rise to localized states (called defect modes) inside the band gaps. These electromagnetic states are very sensitive to different parameters of the structure [18, 19, 20].

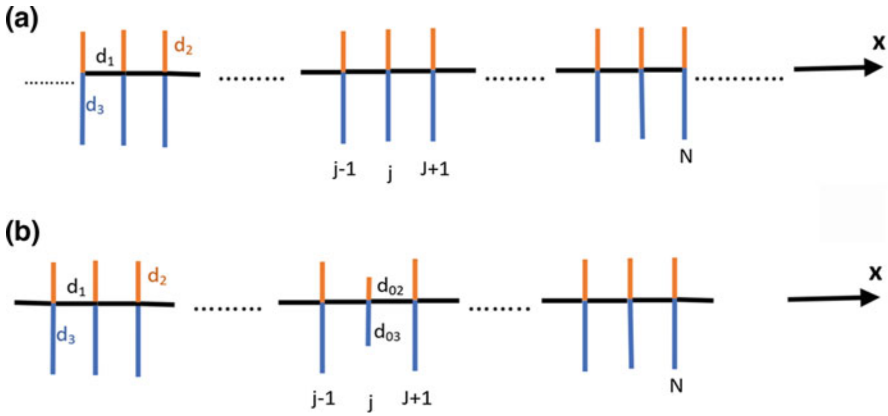


Fig. 1 (**a**) Schematic illustration of the one-dimensional waveguides structure with a periodic array of two grafted resonators of lengths d_2 and d_3, distant from each other by a length d_1. (**b**) Finite comb structure of the one-dimensional waveguides with a periodic array of two grafted resonators of lengths d_2 and d_3, distant from each other by a segment of length d_1 containing two defective resonators of lengths d_{02} and d_{03}

In this paper, we study a new structure composed by infinite or finite waveguide (segment) of length d_1 and grafted in each site by two lateral branches of different lengths d_2 and d_3, with $d_3 = \alpha d_2$ (Fig. 1a). We create inside this structure, two defectives resonators located in the same site of lengths d_{02} and d_{03}, with $d_{03} = \alpha' d_{02}$ (Fig. 1b). This geometrical defects can create two defects modes inside the band gaps; this modes can be used as electromagnetic filters with a high transmission coefficient and a high quality factor Q (the ratio of the central frequency and the full width at half-maximum of the transmittance modes).

2 Model and Formalism

In this framework, we developed the efficient model called the interface response theory to study the electromagnetic waves in photonic comb-like structure containing two geometrical defects. The object of this theory is to calculate the Green's Function of a composite system containing a large number of interfaces that separate different homogenous media. The knowledge of this Green's function enables us to obtain different physical properties of the system such as the reflection and transmission coefficients, the phase, and phase time of the waves. In this theory, the Green's function of a composite system can be written as [18]:

$$g(DD) = G(DD) + G(DM)\left\{[G(MM)]^{-1}g(MM) - [G(MM)]^{-1}\right\}G(MD) \quad (1)$$

where D and M are respectively the whole space and the space of the interfaces in the comb-like system. G is a block-diagonal matrix, where each sub-block G_{i} corresponds to the bulk Green's function of the subsystem i. One can notice that all the matrix elements $g(DD)$ of the composite material can be obtained from the knowledge of the matrix elements $g(MM)$ of the g in the interface space M. The latter are obtained by writing the matrix $g^{-1}(MM)$ as a sum of the sub-matrices $g_i^{-1}(MM)$ in each sub-block i considered separately. Within this theory, the reflected and transmitted waves u(D) resulting when a uniform plane wave incident U(D) upon a plane boundary between two different media are given by [18]:

$$\mathrm{u}\,(\mathrm{D}) = \mathrm{U}\,(\mathrm{D}) + G(DM)\left\{[G(MM)]^{-1}g(MM) * [G(MM)]^{-1}\right\}U(M) \qquad (2)$$

3 Results and Discussions

The proposed system is an infinite or finite one-dimensional photonic comb-like structure, constituted by a segment of length d_1 along which two finite lateral branches are grafted periodically with different lengths d_2 and d_3. This photonic system may contain two defects at the lateral branches level; these defects have lengths d_{02} and d_{03} and situated in the site j (see Fig. 1b). The materials constituting the waveguide and the two lateral branches of the structure are assumed homogeneous, nonmagnetic ($\mu_1 = \mu_2 = 1$), and are of identical relative permittivity, i.e., $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = 4$. The total number of sites (number of nodes) is taken $N = 7$, and $D = d_1$ is the period of the structure. In this work, we take the parameters $\alpha = d_3/d_2$ and $\alpha' = d_{03}/d_{02}$. The boundary condition is $H = 0$ (the magnetic field equal 0) and the reduced frequency is $\Omega = \omega\sqrt{\varepsilon_1 \mu_1}D/(c)$, where c is the velocity of light in vacuum for the electromagnetic waves and ω is the angular frequency (in s^{-1}).

3.1 Band Structure Versus the Length of the Lateral Branch d_3

First, we study in Fig. 2 the variation of the reduced frequency Ω versus d_3 for different values of d_2. The gray areas show the pass band regions while the white areas correspond to the gaps where the electromagnetic waves cannot propagate. The black dotted curves represent the variation of the maximum transmission for a perfect structure. In the case (a), we observe the existence of large gaps in same situations (for example when $d_3 = 0.55$) while there are small gaps for other situation (are good candidates for photonic band gaps (PBGs) materials). In the case (b), we observe that some large gaps and small gaps began to appear. In the case (c), we notice the creation of other gaps in comparison between the previous cases; these gaps become smaller (the gap is closed in comparison to the previous cases). For the case (d), we observe the creation of new band gaps; however, the width of these bands become smaller.

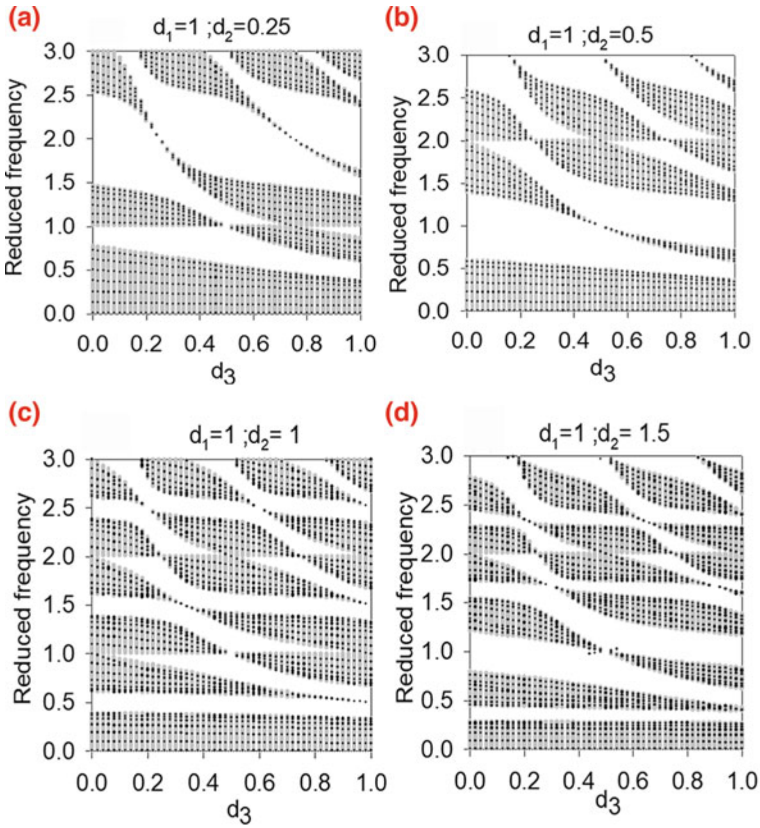


Fig. 2 Variation of the reduced frequency Ω as a function of the length d_3 for different values of d_2, namely, $d_2 = 0.25$D (**a**), 0.5D (**b**), 1D (**c**), and 1.5D (**d**)

So, we conclude in general that when we increase the value of d_2, new band gaps are created, and the width of these band gaps decreases. Let us notice that the system proposed possesses large gaps when we compared it with the multilayer photonic crystal band gaps [16].

3.2 Transmission Spectrum Through a Finite Structure with Different Values of $\alpha = d_3/d_2$

We study in this part in Fig. 3, the variation of the electromagnetic transmission versus the reduced frequency of a perfect finite structure for different values of the ratio between the lengths of the lateral branches α, with $d_3 = \alpha d_2$. For the case (a) where $\alpha = 0.1$, despite the finite number of lateral branches, one can notice that the transmission approaches zero in the areas of the gaps. Also, we observe very well the existence of four large gaps separated by pass bands and one semi-gap around $\Omega = 4$. For the case (b), where $\alpha = 0.5$, we observe the existence of six gaps. For the case (c), where $\alpha = 1$, we observe the existence of four large gaps while the existing twelve gaps where $\alpha = 2$ (case d). So, we deduce that the number and the width of the gaps are very sensitive to the parameter α describing the resonators asymmetric of the comb-like structure. Also, it appears clearly from these figures that more pass bands and more gaps occur in original gap with the increasing of α.

In this paragraph, we study the evolution of the gaps and the pass bands for the perfect structure. The gray areas represent the passbands, and the white areas represent the band gaps of perfect infinite are shown in Fig. 4. This figure represents the variation of the reduced frequency versus the ratio α. One can notice the dependence of the width and the number of the band gaps versus the variation of the ratio α. Also, we observe that the gaps move towards the low frequencies as long as we increase the parameter α. In addition, we find some gaps which decrease in the width with the increase α (for example, the first gap).

3.3 Transmission Coefficient as a Function of the Parameter $\alpha = d_3/d_2$ and $\alpha' = d_{03}/d_{02}$

We consider the comb-like structure composed of a waveguide along with two branched lateral branches, grafted onto N equidistant sites. We create inside this structure two geometrical defects of lengths d_{02} and d_{03} $d_{03} = \alpha' d_{02}$ and $d_3 = \alpha d_2$. In this section, we investigate in Fig. 5 the behavior of the induced defect modes shown in the transmission spectrum by choosing different values of the parameter α and α', the other parameters are taken to be $N = 7$ and $j = 4$. In the case (a), it is obvious that there are peaks corresponding to defects modes inside the gaps. These peaks have a high transmission and a good quality factor Q. For the first, second, fourth, and five gaps, we observe that it exists two defect modes while the other gaps contain a single defect mode (the superposition of the two defects modes). In the case (b), it is obvious that there is one defects mode inside the first and third gap, while two defects modes exist in the second, fourth, and fifth gap. In the case (c), it is obvious that there are twelve gaps with one defects modes in the first and eleventh gap, while two defects modes exist in eighth gap. In the case (d), we observe one defects mode in the eight gap, while two defects modes exist in the eleventh gap.

The lengths of the two defective lateral branches introduced in the perfect comb-like waveguides structure must be carefully designed to obtain isolated defect peaks within the band gaps. These results are improved compared to the results found by [18, 19, 20]

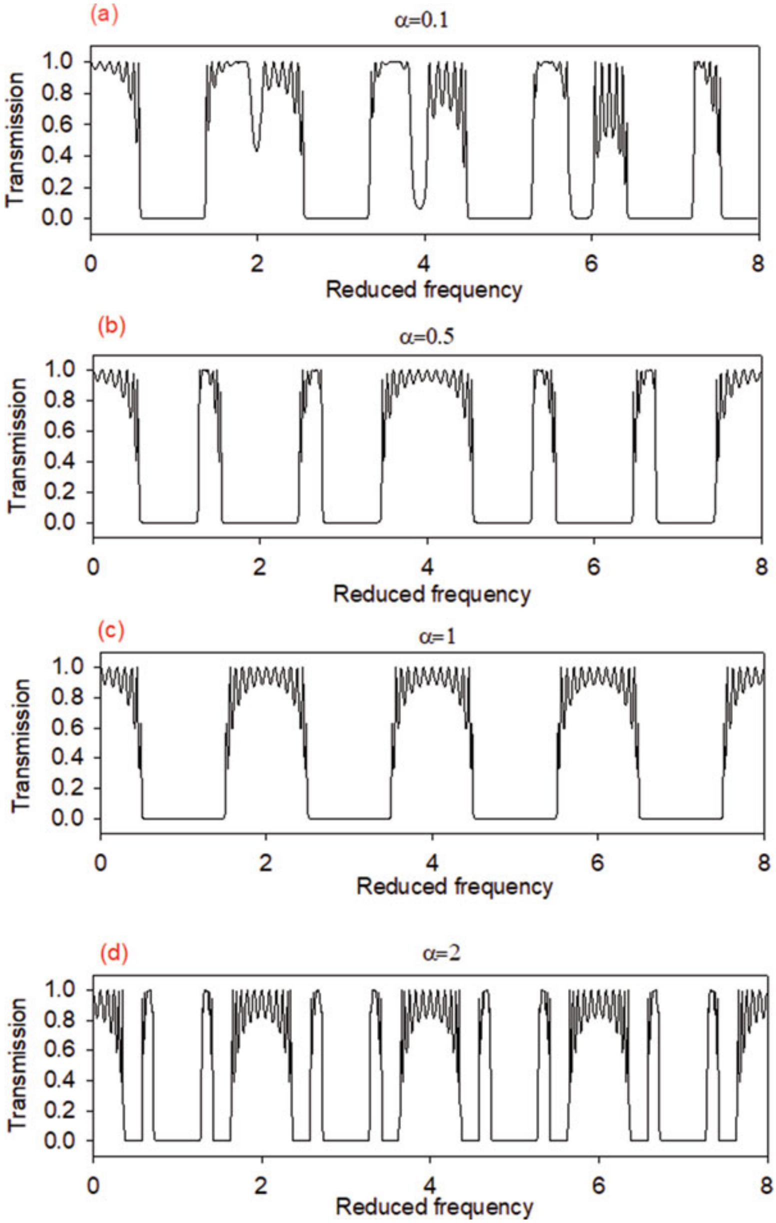


Fig. 3 Transmission coefficient frequency Ω through a perfect comb-like structure for different values of α, namely, $\alpha = 0.1$ (**a**), 0.5 (**b**), 1 (**c**), and 2 (**d**)

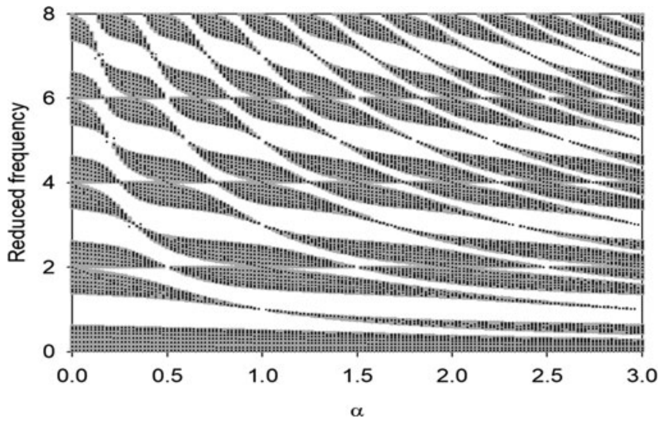


Fig. 4 Variation of reduced frequency Ω as a function of α with $d_3 = \alpha d_2$

3.4 *Transmission Spectrum of the Defective Asymmetric Comb-Like Structure as Function of the Parameter $\alpha' = d_{03}/d_{02}$*

To give a deeper insight into the existence and behavior of the defect's branches (defects modes) where the parameter α' ($d_{03} = \alpha' d_{02}$) is modified, we have performed calculations on the evolution of the transmission of the reduced frequencies as a function of the ratio of defects lengths α', in the photonic comb-like structure containing two asymmetric defects. In the curve (6a), the hatched areas correspond to the bulk bands (the modes located in the pass bands) of the perfect infinite comb-like structure and the white areas correspond to gap when exist the defect's branches. The frequencies of the localized modes are very sensitive to the defects lengths ratio α'. The localized modes emerge from the high bulk bands, decrease in frequency when increasing the lengths ratio α', and finally merge into a lower bulk band where they become resonant states. In each gap, we notice for a certain region that there is only one defect branch (it's the phenomenon of the superposition of the two defects branches), whereas there are two defects branches in otheregions as shown in the curve 6b. As a summary, one obtains one or two defects branches (defects modes) inside a gap by creating the two asymmetric defects in the photonic comb-like structure. We can also define the order of the gap where we are seeking two defects branches. To realize two photonic filters, one need to design a structure in which the transmission coefficient and quality factor Q are very high. Curve (c) shows that the two modes located at $\alpha' = 0.25$, in the curve (b) have a very high transmission and high quality factor (defects modes are very narrow). The curve (d) is the same as the curve (a), the curve (f) shows that the two defects modes (see the curve (e)) located at $\alpha' = 2.2$ have a maximum of transmission and a high quality factor Q. So, we conclude that we can obtain two electromagnetic filters with a high performance when $\alpha' = 0.25$ and the reduced frequency varied between 6.4 and 7.2, or when $\alpha' = 2.2$ and the reduced frequency vary between 4.7 and 5.3 (Fig. 6).

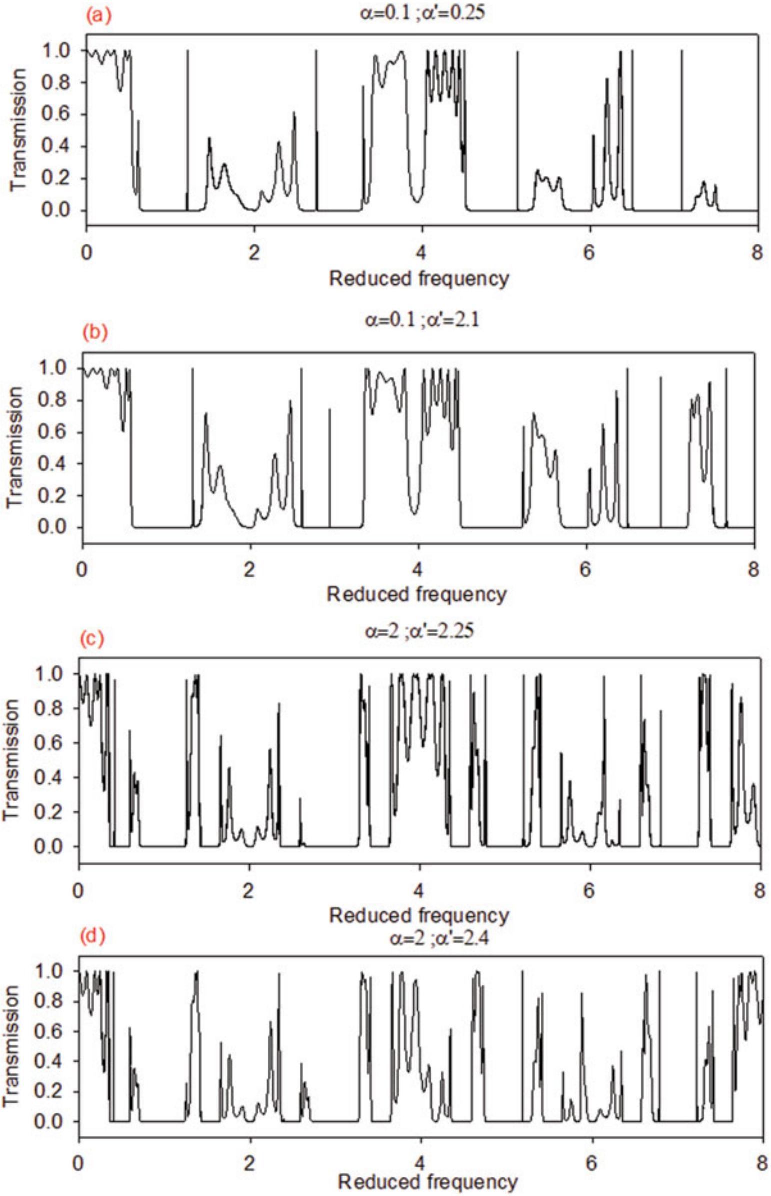


Fig. 5 Variation of the transmission spectrum as function of the reduced frequency Ω in the case where two defective resonators are introduced into the comb-like structure for different values of α and α' with $N = 7$ and $j = 4$

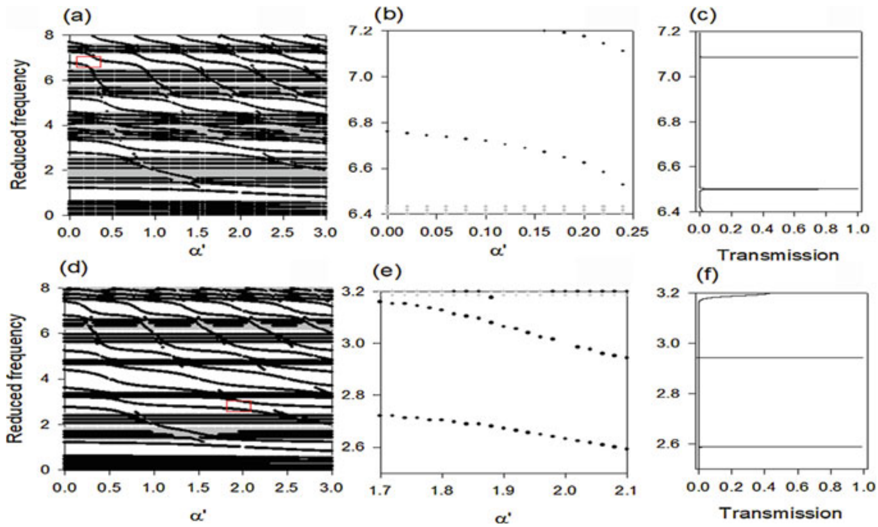


Fig. 6 Curve (**a**) represents the variation of the reduced frequency Ω as function of the α' for $\alpha = 0.1$. Curve (**b**) is same as curve (α') except $\alpha' = [0–0.25]$ and reduced frequency $\Omega = [6.4–7.2]$. Curve (**c**) represents the transmission when $\alpha' = 0.25$. Curve (**d**) is same as curve (**a**) except $\alpha' = 0.4$. Curve (**e**) is same as curve (**d**) except $\Omega = [4.7–5.3]$ and $\alpha' = [1.7–2.2]$. The curve (**f**) represents the transmission of the two modes in curve (**e**) for $\alpha' = 2.2$

4 Conclusions

In this paper, we have studied using the Interface Response Theory (IRT) a one-dimensional (1D) photonic comb-like structure formed by alternating medium of waveguide (segment) with length d_1 and grafted in each site by a finite number of two asymmetric lateral branches (resonators) with different lengths d_2 and d_3. These structures can create large band gaps. We have shown that defects modes appear inside these band gaps by introducing two geometrical defects. The band structure shows very well the appearance of double defects modes (two defects branches) in some particular band gaps. These modes have a very high transmission factor and a high quality factor Q. Therefore with these high characteristics, the structure proposed in this work could be used in the design of two electromagnetic filters for applications in telecommunication field.

References

1. Land, P. L., & Becker, R. J. (1993). *Comboptical interference filter*. Patents Number 5,225,930.
2. Land, P. L., & Becker, R. J. (1995). High-transmission comblike optical filters. *Journal of Optical Society A, 12*, 611–622.
3. Supradeepa, V. R., Long, C. M., Wu, R., Ferdous, F., Hamidi, E., Leaird, D. E., & Weiner, A. M. (2012). Comb-based radiofrequency photonic filters with rapid tunability and high selectivity. *Nature Photonics, 6*, 186–194.

4. Hamidi, E., et al. (2010). Tunable programmable microwave photonic filters based on an optical frequency comb. Submitted to the Joint Special Issue of *IEEE Transanction on Microwave Theory and Technolgy/Journal of Lightwave Technology on Microwave Photonics, 58*, 3269–3278.
5. Song, M., et al. (2011). Reconfigurable and tunable flat-top microwave photonic filters utilizing optical frequency combs. *IEEE Photonics Technology Letters, 23*, 1618–1620.
6. Chen, H. W., et al. (2010). Integrated microwave photonic filter on a hybrid silicon platform. *IEEE Transactions on Microwave Theory and Technolgy, 58*, 3213–3219.
7. Norberg, E. J., Guzzon, R. S., Parker, J. S., Jouhansson, L. A., Coldren, L. A., et al. (2011). Programmable photonic microwave filters monolithically integrated in InP-InGaAsP. *Journal of Lightwave Technology, 29*, 1611–1619.
8. Vasseur, J. O., Deymier, P. A., Dobrzynski, L., Djafari-Rouhani, B., & Akjouj, A. (1997). Absolute band gaps and electromagnetic transmission in quasi-one-dimensional comb structures. *Physical Review B, 55*, 10434–10442.
9. Ben-ali, Y., Tahri, Z., Falyouni, F., & Bria, D. (2018). Study about a filter using a resonator defect in a one-dimensional photonic comb containing a left-hand material. *Proceedings of the First International Conference on Electronic Engineering and Renewable Energy, 519*, 146–156.
10. Ben-ali, Y., Tahri, Z., Ouariach, A., & Bria, D.. (2019). Double frequency filtering by photonic comb-like. *IEEE Xplore* (pp. 1–6).
11. Ben-ali, Y., Tahri, Z., & Bria, D. (2019). Electromagnetic filters based on a single negative photonic comb-like. *Progress in Electromagnetic Research C, 92*, 41–56.
12. Dobrzynski, L., Akjouj, A., Djafari-Rouhani, B., & Vasseur, J. O. (1998). Giant gaps in photonic band structures. *Physical Review B, 57*, 9388–9391.
13. Jing, Y., Ma, Y., Li, Y., & Heine, T. (2017). GeP3: A small indirect band gap 2D crystal with high carrier mobility and strong interlayer quantum confinement. *Nano Letters, 17*, 1833–1838.
14. Wang, Z., Zhao, L., Mak, K. F., & Shan, J. (2017). Probing the spin-polarized electronic band structure in monolayer transition metal dichalcogenides by optical spectroscopy. *Nano Letters, 17*, 740–746.
15. Ben-ali, Y., Tahri, Z., Bouzidi, A., Jeffali, F., Bria, D., Azizi, M., Khettabi, A., & Nougaoui, A. (2017). Propagation of electromagnetic waves in a one-dimensional photonic crystal containing two defects. *Journal of Materials and Environmental Science, 8*, 870–876.
16. Singh, B. K., Tiwari, S., Chaudhari, M. K., & Pandey, P. C. (2016). Tunable photonic defect modes in one-dimensional photonic crystals containing exponentially and linearly graded index defect. *Optik, 127*, 6452–6462.
17. Bouzidi, A., Bria, D., Akjouj, A., Pennec, Y., & Djafari-Rouhani, B. (2015). A tiny gas-sensor system based on 1D photonic crystal. *Journal of Physics D: Applied Physics, 48*, 495102–495109.
18. Vasseur, J. O., Djafari-Rouhani, B., Dobrzynski, L., & Akjouj, A. (1999). Defect modes in one-dimensional comblike photonic waveguides. *Physical Review, 59*, 13446–13452.
19. Djafari-Rouhani, B., El Boudouti, E. H., Akjouj, A., Dobrzynski, L., Vasseur, J. O., Mir, A., Fettouhi, N., & Zemmouri, J. (2001). Surface states in one-dimensional photonic band gap structures. *Vacuum, 63*, 177–183.
20. Ben-Ali, Y., El Kadmir, I., Tahri, Z., Bria, D., (2020). Defects modes in one-dimensional photonic filter star waveguide structure. Materials Today, 27, 3042–3050.

Optimization of a Photovoltaic System by MPPT Strategy "Perturb and Observe"

Otman Oussalem, Mustapha Kourchi, Hicham Idadoub, Mohamed Ajaamoum, Azzdine Rachdy, Mhend Oubella, and Samia Jenkal

Abstract The work presented in this article deals with the realization of a control platform of a photovoltaic system through an Arduino Mega board via the Matlab-Simulink software. This experimental platform allows the photovoltaic system to extract the maximum power available using the "Perturb and Observe" algorithm. The results obtained from this experimental study have demonstrated the value of this platform as a cost-effective device for optimizing photovoltaic chain.

Keywords PV · DC-DC Converter · P&O · Arduino · Simulink

1 Introduction

As part of its energy strategy, Morocco has given priority to the development of renewable energy, specifically in solar thermal and photovoltaic energy.

Photovoltaic (PV) systems offer a very competitive solution as an alternative energy source, but they have low energy efficiency.

To overcome this problem, it is necessary to optimize the design of all parts of the solar PV chain.

Therefore, an adaptation stage will be inserted between the photovoltaic generator (GPV) and the load. This stage (DC-DC converter) which is driven by a microcontroller will allow the system to search for and reach the Maximum Power Point (MPP).

The search for this MPP is provided by strategies of Maximum Power Point tracking (MPPT) as "Perturb and Observe" (P&O). However, the choice of the DC-DC converter depends on the GPV and the load, and it is a key factor to increase the performances of power transfer between the GPV and the load [1].

O. Oussalem (✉) · M. Kourchi · H. Idadoub · M. Ajaamoum · A. Rachdy · M. Oubella · S. Jenkal
LASIME-National School of Applied Sciences, Ibn Zohr University, Agadir, Morocco
e-mail: m.kourchi@uiz.ac.ma; m.ajaamoum@uiz.ac.ma; a.rachdy@uiz.ac.ma; m.oubella@uiz.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_53

The objective of this study is the optimization of photovoltaic system with MPPT technology type P&O, through an Arduino board, handled by the Matlab-Simulink software. This optimization is based on the control of a DC-DC converter by a rectangular signal with Pulse width Modulation (PWM), generated by a special circuit, and controlled again by the same Arduino board. The latter also plays the role of an acquisition card.

The Matlab-Simulink interface allows generating the MPPT control and displaying system characteristics.

This article is divided into three parts: After the introduction, the first part gives a general description of a photovoltaic system. The second part deals with the interface based on the Arduino board under the Matlab-Simulink environment. The third one will be dedicated to the experimental results of the system studied. Finally, this study finishes with a conclusion.

2 Photovoltaic System

The photovoltaic system under study consists of four blocks as shown in Fig. 1. The first block represents the photovoltaic panel, the second block is the static converter DC-DC Buck-Boost, the third block represents the DC load, and the fourth one is the system controller.

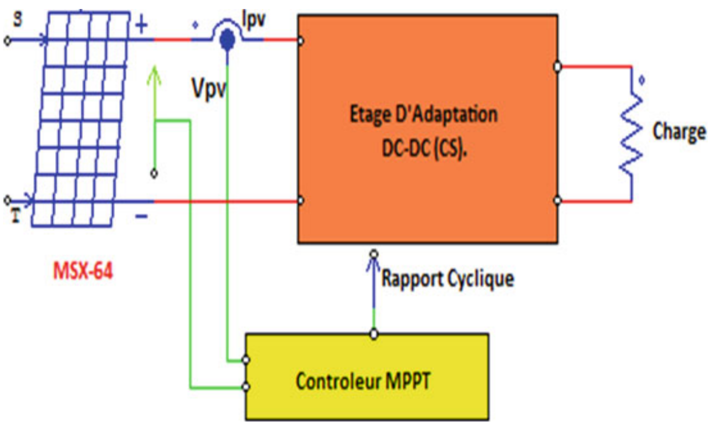


Fig. 1 The photovoltaic system

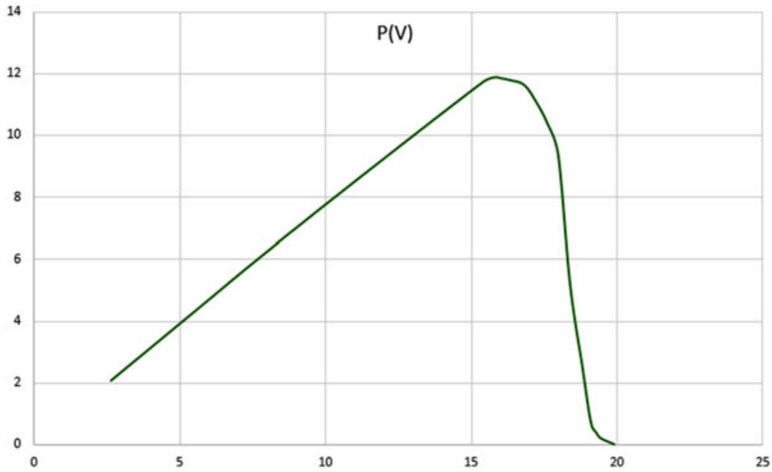


Fig. 2 The characteristic P(V) of the PV emulator for $G = 400$ W/m^2

2.1 PV Generator

The PV panels depend on climatic conditions, that's why we chose to employ a PV emulator.

This PV emulator (reference CO3208-1A) is an electronic power system able to reproduce the characteristics of the solar panel, and which has the following characteristics:

- It consists of three independent blocks emulating solar panels.
- Voltage of open circuit is 20 V.
- Short-circuit current up to 2 A.
- Displays voltage and current variables on integrated displays for each of the three blocks.
- Simulation adjustable intensities of solar irradiation, each of the three blocks.

For an irradiation of 400 W/m^2, the characteristics (I-V) and (P-V) of the photovoltaic emulator are illustrated in Figs. 2 and 3.

2.2 Static Converter

The main role of the Buck-Boost power converter is to ensure impedance matching, so that the output of the PV emulator delivers the maximum energy.

The electronic circuit corresponding to the Buck-Boost realized (Fig. 4), is essentially based around the power MOSFET transistor, type IRF730, driven by

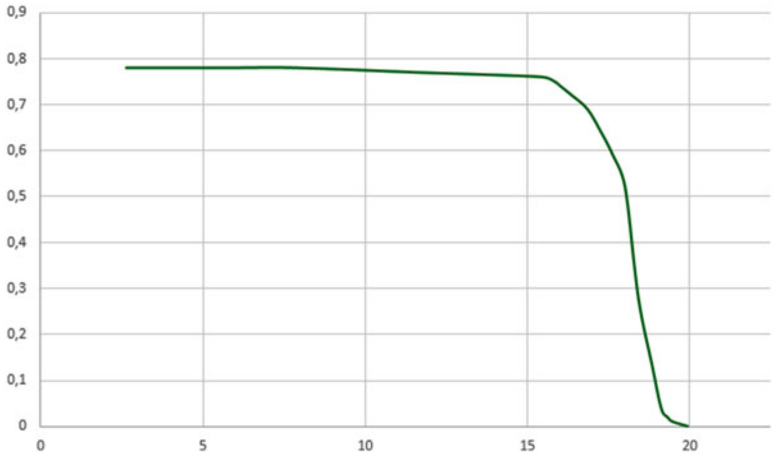


Fig. 3 The characteristic I(V) of the PV emulator for $G = 400$ w/m^2

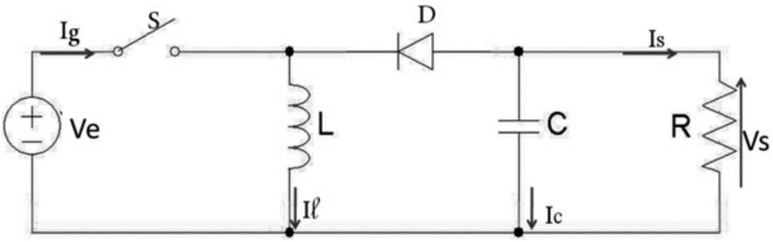


Fig. 4 Circuit of Buck-Boost converter

the Arduino control board via a driver, as well as a diode, and passive components (L, C) ensuring the smoothing and filtering of the current electric [2, 3].

The main variables characterizing the Buck-Boost converter [3] are:

The output voltage is:

$$V_S = \frac{\alpha}{1-\alpha} V_e \tag{1}$$

The output current is:

$$I_S = \frac{1-\alpha}{\alpha} I_e \tag{2}$$

The current ripple: ΔI_L

$$\Delta I_L = \frac{\alpha V_e}{Lf} \tag{3}$$

The voltage ripple: ΔV_C

$$\Delta V_C = \frac{\alpha I_S}{cf} = \frac{\alpha^2 V_e}{(1-\alpha)\, RCf} \tag{4}$$

With:

α: the duty cycle of the PWM signal or PWM.
f: the frequency of the PWM signal.
V_C: voltage across the capacitor.
I_L: current through the coil.
R: resistive load.
L and C: inductor and capacitor constituting the filter. The filter values are: $L \geq 0.9$ mH and $C \geq 27$ μF.

2.3 PWM Generator

There are a lot of generators of PWM signal, the used one is based on the TL494 component. This is an integrated circuit of pulse width modulation control for fixed frequency signals. The circuit accompanying the TL494 component is described in the electrical schema of Fig. 5 [4].

This generator can operate at frequencies up to 400 kHz. According to technical documents of TL494 builder, the approximate oscillation frequency is determined by [1]:

$$f_c = \frac{1.1}{R_t C_t} \tag{5}$$

The duty cycle of the PWM signal is controlled by a voltage that varies from 0 to 2.5 V.

2.4 Driver

The MOSFET used must be controlled by a PWM signal greater than 7 V, for that reason, a driver was realized for amplifying the signal of the PWM generator by the circuit shown schematically in Fig. 6.

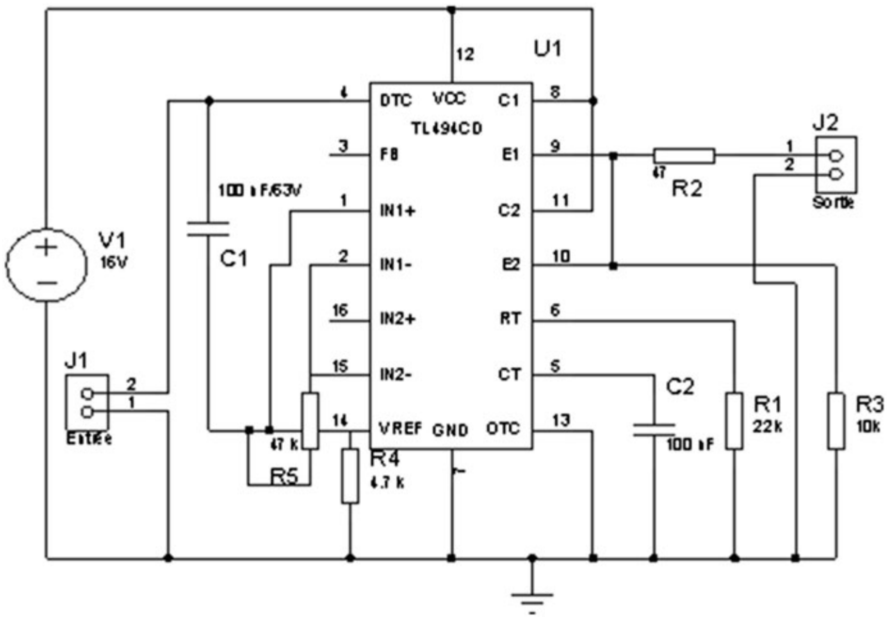


Fig. 5 PWM generator based on the TL494 circuit

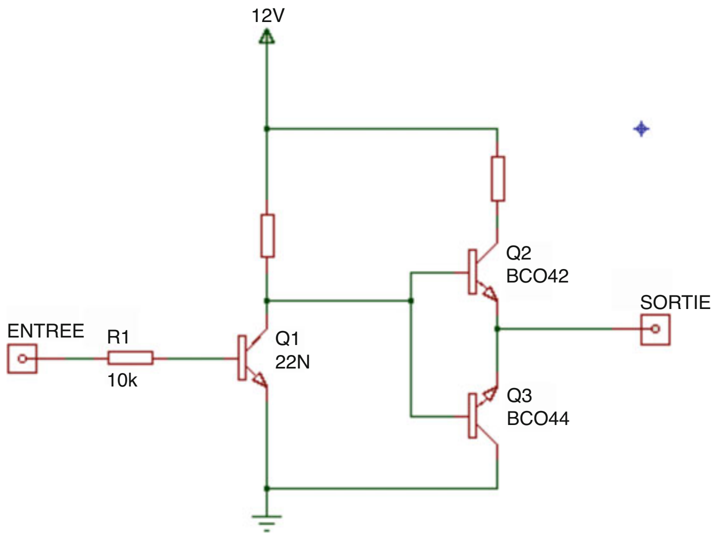


Fig. 6 The driver scheme used

2.5 Voltage Sensor and Current

To follow the MPP, the platform needs two sensors: current and voltage sensor.

The voltage measurement is performed from a voltage divider to have a voltage between 0 and 5 V. The output of this divider drives a follower amplifier realized by the circuit "LM324" to the impedance matching. However, the current measurement is done by a shunt resistor of 1Ω.

3 MPPT Strategy and Arduino Board

The strategy of maximizing the power from a photovoltaic source is to seek the optimum operating point. This technique is called the MPPT strategy.

There are several types of MPPT strategies, among which is the Perturb and Observe (P&O) command, which involves causing a low voltage perturbation, which results in a change in power.

Figure 7 shows that if a duty cycle decrease causes an increase in power, the operating point is on the left of the MMP, if on the contrary the power decreases, it is on the right. In the same way, we can make the same reasoning for a duty cycle increase [1–5].

The conventional flow chart of the MPPT control type P&O is represented in Fig. 7, where the evolution of the power is analyzed after each voltage disturbance. For this type of control, two sensors (current and voltage of the GPV) are needed to determine the power of the PV at every moment.

3.1 The Arduino Board

The Arduino is an electronic board. Open source, designed to control a system interactively from the program, which is defined and installed in his memory.

These boards allow simple and inexpensive access to prototyping embedded systems. The Arduino language differs from the language used in the computer industry embedded in its simplicity.

3.2 Matlab-Simulink

Matlab (Matrix laboratory) is a fourth-generation programming language, developed by MathWorks. Matlab has provided a tool book for Arduino called the support package, which includes a Simulink block library to configure and access the sensors, actuators, and Arduino communication interfaces [6, 7].

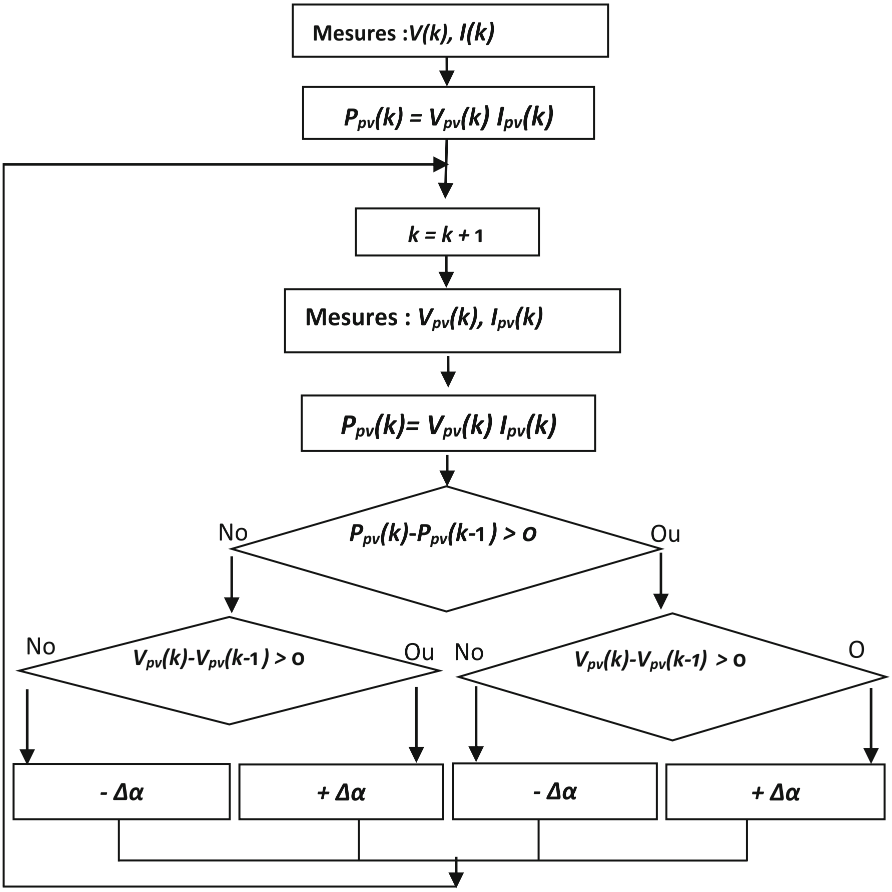


Fig. 7 Conventional flow chart of the MPPT control type P&O

This developed configuration (Fig. 8) is for the treaty system, based on the Simulink environment:

4 Experimental Result

Figure 9 shows the experimental realization of the photovoltaic system under study. The latter is composed of a PV emulator, a Buck-Boost chopper, sensors of Ipv (PV emulator current) and Vpv (PV emulator voltage), a resistive load, and an Arduino Mega board.

The experimental test of the studied system should start by attacking the Arduino board, by the program of the P&O strategy treated previously.

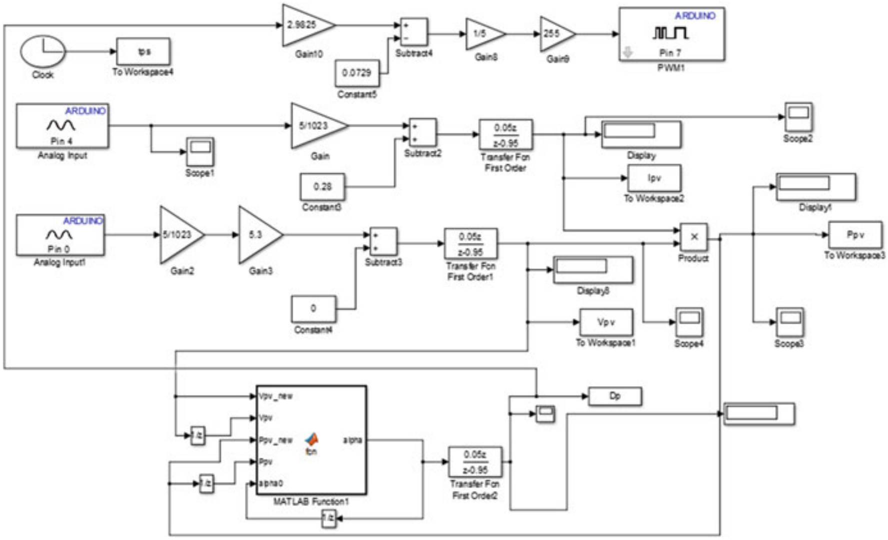


Fig. 8 Control Simulink interface

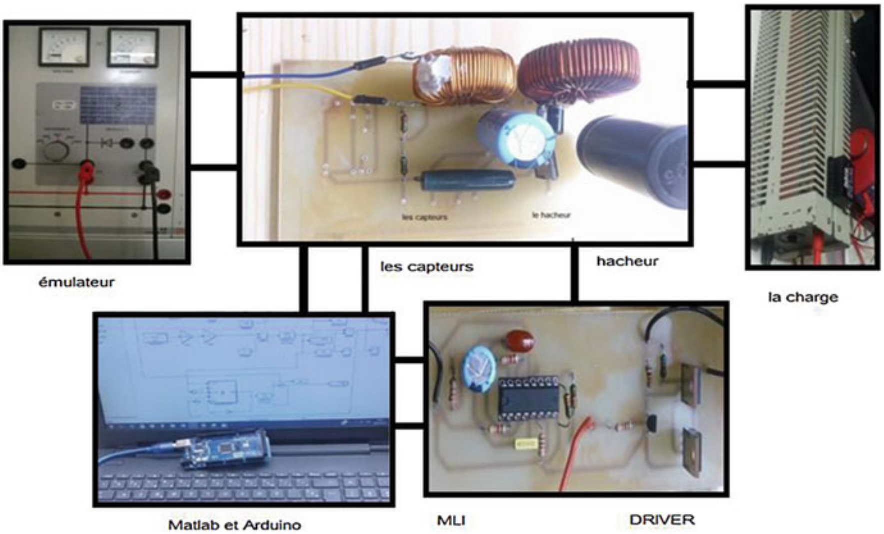


Fig. 9 The parts of the designed PV chain

It is noted that during the experimental process, some adaptations and calibrations were made between the various variables treated in the photovoltaic system. Thus the system can search for and converge towards the point of maximum power. The PV emulator offers the possibility to choose the G value of the irradiation by

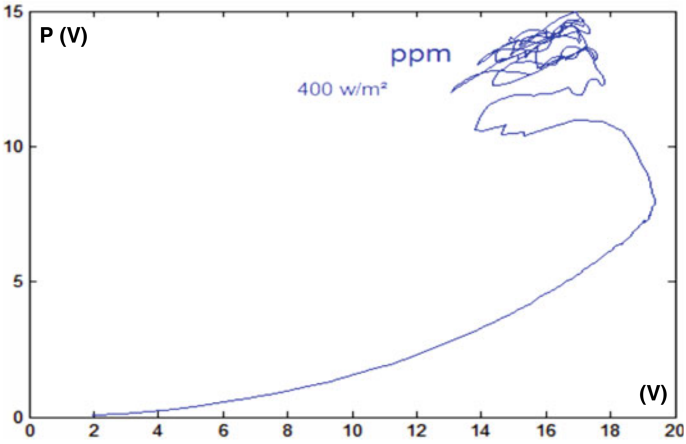


Fig. 10 The MMP of the system for $G = 400$ W/m^2

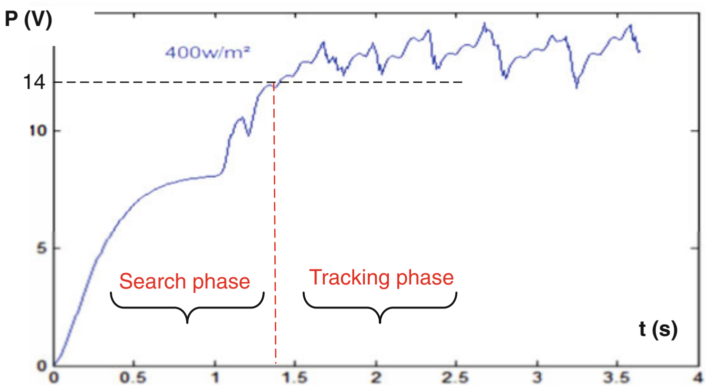


Fig. 11 The MMP of the system in function of time for $G = 400$ W/m^2

selecting one of the two values 400 and 600 W/m^2, in order to see the influence of the P&O strategy on the PV emulator's output power (Ppv).

Figure 10 describes the evolution of the PV emulator's output power for the irradiation $G = 400$ W/m^2.

This figure illustrates the recording values of the PV emulator's output power Ppv as a function of the voltage Vpv. It shows that the photovoltaic system realized was able to seek and follow the power supplied and to reach the maximum power at around 14 W for an irradiation of G = 400 W/m^2.

In Fig. 11, the recording values of the PV emulator's power supplied as a function of time are processed in order to show the search phase and the monitoring and stabilization of the output power around the maximum power.

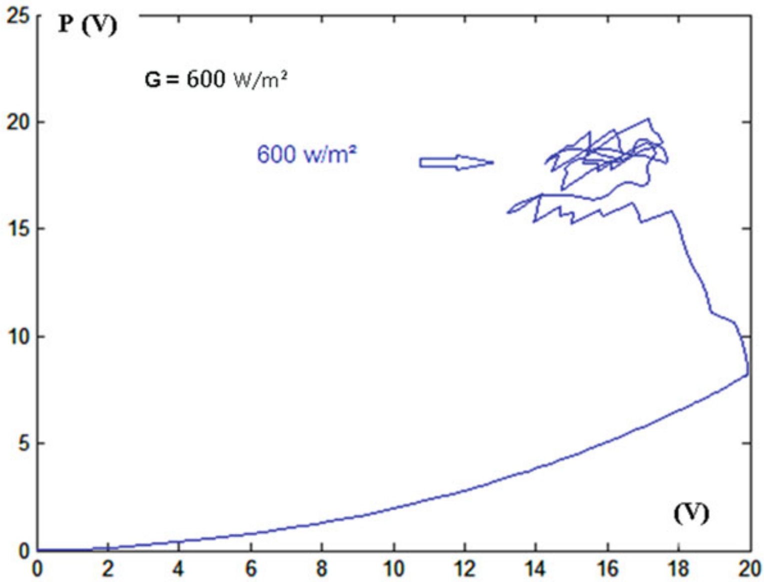


Fig. 12 The MPP of the system for $G = 600$ W/m^2

It is also remarkable in Fig. 12 that the system behaves in the same way in a second test for irradiation $G = 600$ W/m^2.

Figure 13 illustrates the recording values of the PV emulator's output power for the irradiation $G = 600$ W/m^2 during the search phase, tracking and stabilizing the round of the maximum power.

The goal in Fig. 14 is changing the irradiation G suddenly from 400 to 600 W/m^2 to investigate about the efficiency of the optimization platform carried out.

Following the given irradiations, the system then looks for and follows the new values of the maximum power. So the system responds with the same way.

Figure 15 also shows the recording values of PV emulator's output power as a function of time, to observe the system behavior at the immediate change of the irradiation from 400 to 600 W/m^2.

Figures 14 and 15 represent the variations of PV emulator's output power for irradiation values of 400 and 600 W/m^2. The power then reaches around the respective values 14 and 18 W. So those values are quite well correlated with the maximum powers in the static case presented previously.

The interest of these tests of sudden change of the irradiations values is to make sure of the performance of the control platform realized for the studied system.

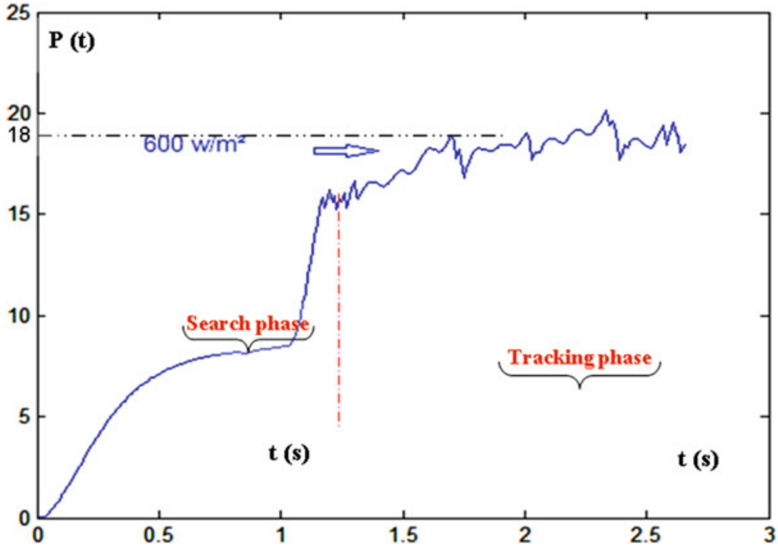


Fig. 13 The MMP of the system in function of time for $G = 600$ W/m^2

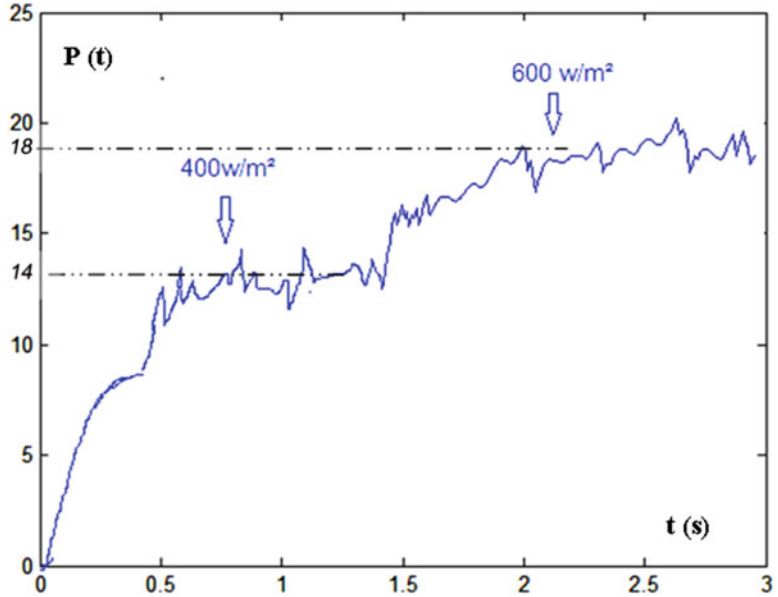


Fig. 14 MMP a function of time of the system when changing G from 400 to 600 W/m^2

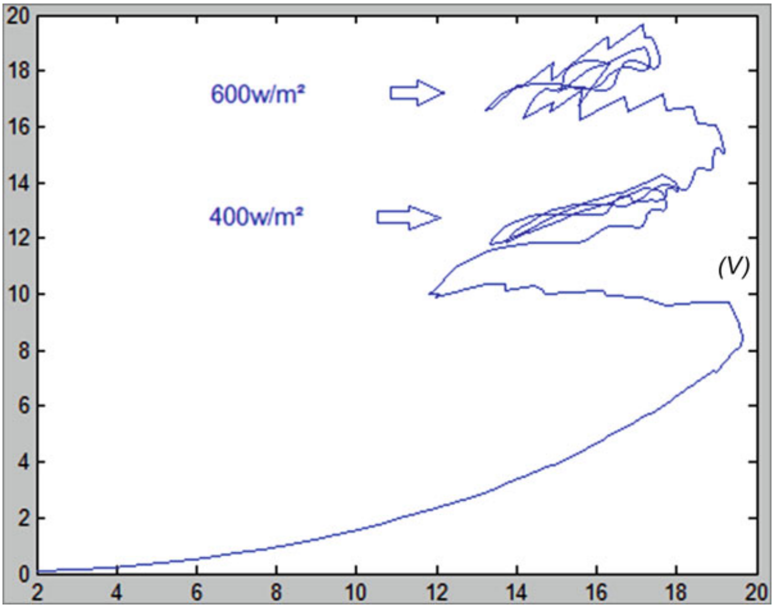


Fig. 15 MMP of the system when changing G from 400 to 600 W/m^2

5 Conclusion

The interest of this article is about the design and realization of a control platform of a PV system based on the Arduino board via Matlab simulink.

According to the results obtained from the practical tests of a control realized platform for the studied system.

The PV system transfers the maximum power supplied by the PV emulator to the load. Thus, the platform controlled by the “P&O” strategy, based on the Arduino Mega board via Matlab-Simulink, is corroborated as a cost-effective device for optimizing photovoltaic chain, following its performance in terms of MMP tracking of various irradiation values.

As a perspective to this work, it is recommended to develop this realization for different types of PV modules, for different MPPT strategies and via different electronic boards.

Bibliography

1. Ajaamoum, M., Kourchi, R. A., & Bouhouch, L. (2013). Fuzzy controller to extract the maximum power of a photovoltaic system. In *Renewable and Sustainable Energy Conference International (IRSEC)*, Ouarzazate, 7–9 March 2013 (pp. 141–146), IEEE. ISBN: 978-1-4673-6373-0.

2. Edric, C. (2008). *Energy optimization of the electronic adaptation floor dedicated to conversion*. PhD Thesis, University of Toulouse III Paul Sabatier, France.
3. Maçbahi, H. (1996). *Power factor correction in a resonant converter with active method*. Thesis Masters, Université du Québec à Trois-Rivières.
4. Debili, N. (2015). *Study and optimization in Matlab/Simulink environment of a photovoltaic pumping system*. Magister Thesis, University Mentouri, Constantine.
5. Abdelhalim, B. (2009). *Study and control of a circuit for extracting the maximum power from a solar panel*. Magister Thesis, Mentouri University of Constantine, Algeria.
6. Bouderhim Brahim and Salhi Younis' design and implementation of a chopper boost MPPT based on an Arduino board "professional master memory university kasdi merbah - ouargla 2016/2017.
7. Latrache soufian and Bouzid kheir Eddine. (2015/2016) Design and realization of a photovoltaic system based on an Arduino uno professional master memory university kasdi merbah ouargla.

Toward an Enhanced Learning Through Existing Massive Open Online Courses Platforms

Amal Battou

Abstract Massive Open Online Courses (MOOCs) attract many actors all over the world, especially those of higher education. So, there is a need to enhance and to tailor the instruction of MOOCs to meet the learner's needs. This chapter investigates the MOOCs system by reviewing the available literatures and suggesting a framework based on Analytic Hierarchy Process (AHP) algorithm for an adaptive learning through existing MOOCs. In this suggested framework, a variety of aspects including choice of learning path, learner satisfaction, and achievement are considered to present an adequate list of existing MOOCs to a learner. The main aim of this framework is to improve learning effectiveness and ensure learning quality.

Keywords Learning quality · Learner-centered design · Long-term learning · MOOC · Enhanced learning · AHP approach

1 Introduction

Undoubtedly, MOOCs platforms offer many opportunities to support e-learning and learning, from managing the learning and training process to monitoring the evaluation process. To make relevant MOOCs, diverse backgrounds such as content developers, domain experts, instructional designers, pedagogues, graphic designers and programmers, etc. are involved. MOOCs are expensive to produce. It involves considerable amount of time investment of several actors. If some large universities can afford them, it is not the case for smaller ones [4].

In addition, criticisms of the low rates of completion of MOOCs are still current. As with any training, the success of a MOOC is measured on the basis of what has been learned, understood, and internalized. Nevertheless, the task is particularly

A. Battou (✉)
IRF-SIC Laboratory, Faculty of science, Agadir, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_54

laborious because of the lack of individualized training and evaluation. This is mainly due to audience size [10].

Thus, we had the idea to reuse and capitalize existing MOOCs according to their reuse policy and our requirements. At the same time, we will develop courses that meet the skills of our team. Afterward, we will evaluate and choose courses that will be adapted to each learning profile.

Given the diversity of existing MOOCs, we thought about setting up an interactive and adaptive platform that represents the mediating layer between the learner and the MOOCs platforms. The purpose of this platform is to make the decision and to evaluate some amount of information to display the list of courses of a pedagogical course in adequacy with the information and the characteristics of each learner. In addition, this platform must be equipped with technological means enabling users to interact through the man–machine interfaces during the course. The questions that arises in this case is how the platform will evaluate multitude criteria (a list of existing MOOCs, learner needs, capacities, preferences, etc.) to choose the appropriate courses for a learner? How to implement the notion of adaptivity in the learning process in a pedagogical path?

Our readings allowed us to opt for *decision support systems* (DSS). These systems will allow us to take into account a multitude of criteria (information on existing MOOCs and open at time t, information about a student, information on the educational path, etc.) to choose the optimal solutions from a set of possible solutions. Also, the system must take into account information from user feedback and information from tools for tracking of actions performed by learners.

The remainder of this chapter is structured as follows. We will firstly present the AHP method. We discuss next the MOOC reusability concept. Afterward, we explore and present the framework and its features. Finally, we discuss the ability of this framework to improve learning effectiveness and ensure learning quality.

We will first begin by presenting an overview of the state-of-art literature on DSS. Then, we present the AHP method. We then discuss the concept of reusing existing MOOCs as a solution of effectiveness learning. Next, we explore and present the proposed framework architecture that implements the principles of AHP.

2 Multicriteria Decision Support Methods

Multicriteria decision support methods are relatively new and growing scientific approaches. These approaches are solicited where we are confronted with a complex situation, and the decision is based on several decision criteria and possible solutions.

The decision in the presence of multiple criteria is difficult because the criteria are often conflicting. For this, several multicriteria decision support methodologies have been developed, and we quote as an indication: Analytic Hierarchy Process (AHP), Analytic Network Process (ANP), Data Envelopment Analysis (DEA), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), etc.

[7, 11]. The diversity of these methods lies essentially on the synthesis of the information contained in each criterion [3].

The exhaustive study of these methods goes far beyond the scope of this work. Thus, we have relied on comparative studies of the most widespread methods and the choice of the best method to use in a given context. We quote as an example the works of Hammami [5] and Zouggari [16].

In general, the majority of these methods operate in 4 main steps [3]:

- **List the potential actions**: this step defines possible contributions to the overall decision that can be considered independently and serves as a point of application for decision support;
- **List the criteria to consider**: this step involves determining contribution and effects that impact the decision process;
- **Establish the table of performances**: this table is constituted, in rows, of the alternatives and, in columns, the criteria to be taken into account for the decision-making; and
- **Aggregating performance**: this step is about establishing a formalized representation of the appropriate global preferences to the problem of decision support.

The present work focuses on the AHP approach, which is the subject of the following paragraph, for the selection of the most appropriate courses taking into account a certain set of criteria (the learning profile, the existing MOOC data, etc.). Indeed, the final decision to present to the learner requires a comparison between all the input criteria in order to prioritize the courses and classify them according to their relevance.

2.1 The AHP Approach

2.1.1 Presentation of the Approach

Analytic Hierarchical Analysis (AHP) is a theory of complex decision analysis proposed by Thomas Saaty [13]. Recognized by its simplicity of application, this method allows the most credible decisions taking into account several factors. The AHP is considered one of the main mathematical models currently available to support the theory of decision [15].

The AHP structures the criteria in a hierarchical manner and then compares them in pairs to design, prioritize, justify, and choose the right solution for the most complex situations. This method has its advantage of its similarity to the decision-making mechanism of the human being, namely decomposition, judgment, and synthesis [2].

AHP is easy to implement, and it is widely used for solving multiple selection problems. This method makes it possible to split the most complex decision problems in the form of hierarchical levels. The scale of values chosen expresses the preferences of decision-makers, and it also makes it possible to rally qualitative and quantitative criteria [8], where each criterion contributes to the final decision [14].

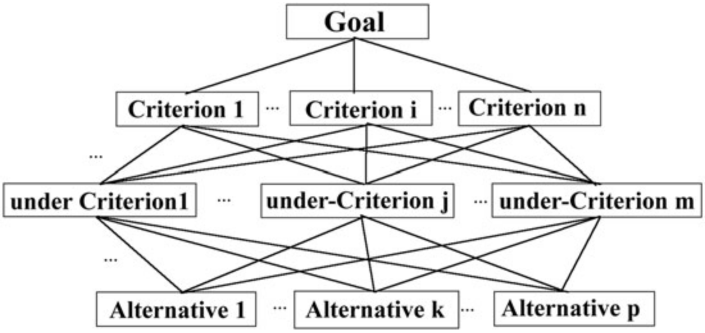


Fig. 1 Hierarchical architecture of the AHP approach

2.1.2 Principle of the Approach

The approach is to first designate the general objective on which we will make our decision and then the possible solutions or alternatives and criteria to consider [7]. In some cases (complex decisions), we can have several levels of criteria (criteria, sub-criteria, etc.) Figure 1 presents the hierarchical architecture of the AHP method.

After having determined the objective criteria, under criteria, and the various alternatives, the next step is to evaluate by peer all the criteria of each hierarchical level compared to the higher hierarchical level.

However, this method had some criticisms mainly on the fact that the association of a numerical scale with another semantics is restrictive; then, it introduces imprecise numerical values. The method has experienced several extensions to attempt to remedy some of the criticisms, the case of taking into account the uncertainty (stochastic AHP), and the blur (AHP blur) in the expression of the judgments [6].

In this work, we only present our system according to the AHP process. Other Phd-student work focuses on the use of fuzzy logic in combination with AHP. These works are subject of publication such as [1].

3 Toward Reusing Pre-existing MOOCs

3.1 Working Context and Demands

Several Moroccan universities, in particular IBN ZOHR University, have adopted several projects to implement and integrate information and communications technology (ICT) into the learning process in order to offer learners a variety of resources (software, multimedia, etc.). The main aim of this use is to increase learners' motivation to learn and contribute to improving the quality of learning and teaching. We also note the use of platforms and websites offering online courses (LMS, E-learning, etc.) and the proposal of diplomas and certificates via distance learning courses and the famous MOOCs that offer massive courses and open for everyone. These have revolutionized the world of education by combining technological and pedagogical aspects in their products (courses, assessments, etc.) and by imposing themselves as a solution to the problems of the low rate of supervision and the growing number of students.

However, all technologies have limitations and weaknesses [9]. As an example, we mention the MOOCs that have solidly established themselves in the educational panorama, especially in foreign universities (Fun, Coursera, etc.). Many works highlight the low completion rate which reflects the dropout rate recorded at the end of each course, while the number of enrolled in the latter at the start is very important.

Given the constraints highlighted in the introduction, we thought to capitalize the existing MOOCs in parallel and we will develop course materials in line with our specialties. Several questions arise and we quote them as an indication: how can we arrange a pile of courses from the MOOCs with the courses developed by our team? which are the criteria to take into account for the generation of the list of courses to select only course in adequacy with a given profile in a specific time? etc.

Below, we will try to bring some elements of answer to these questions.

3.2 Conception of the Proposed Solution

The main goal of our work is to design an intelligent platform to:

- Search courses that meet several criteria set at entry
- Generate a list of course choices relevant to a learning profile at a time t
- Ensure the interactivity of learners between themselves and with the teaching staff
- Track learner progress through tracking interactions with the system.

The learner is led, during his first visit to the system, to create his profile by filling information in a registration form. This profile contains:

- **Domain-independent data (DID)**: these data are rather permanent and include information about the learner's initial knowledge, purpose and plans, cognitive abilities, learning styles, preferences, academic profile (technological studies, knowledge of literature, artistic abilities, etc.), etc.
- **Domain-dependent data (DDD)**: these data are rather dynamic and change as the learner progresses in learning. They essentially contain information on the knowledge/skills acquired for a given field.

After initializing his profile, in a connected learner environment, learner will choose the concept he wants to study. If this concept is part of the courses developed internally, the content will be displayed to the learner taking into account his profile. Otherwise, the platform will launch a request concerning the concept requested

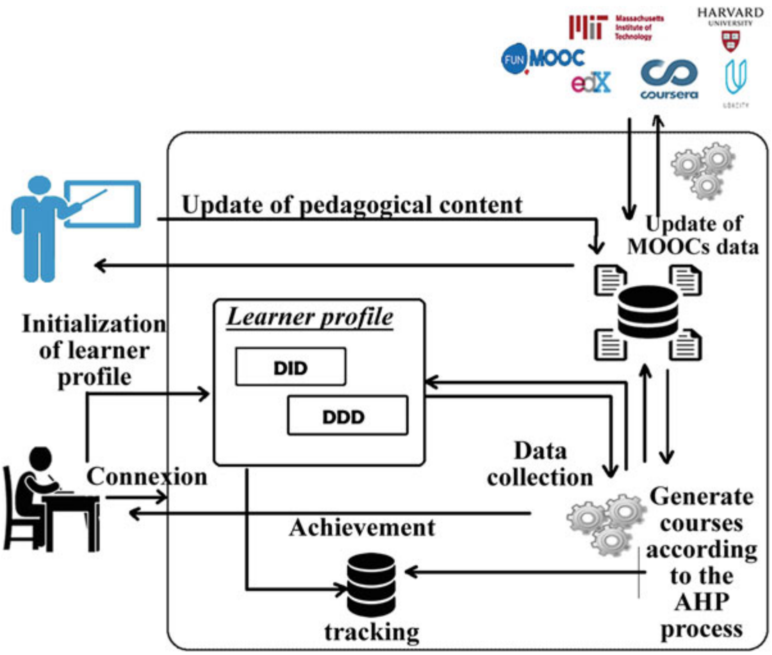


Fig. 2 Architecture of the proposed framework

to the mediation layer with the platforms of the existing MOOCs and chosen beforehand. All the results obtained by this request are transmitted to the generator that will apply the AHP process according to the criteria we have previously determined (learning profile, data on the MOOCs, etc.). The result of the optimal courses will be presented to the learner.

The proposed system architecture is given in Fig. 2.

The list of courses to be generated by the AHP process is based on the following criteria:

- Language of learning (L)
- Course Availability (D)
- Level of knowledge / skill (C)
- Institution responsible for the course (E)
- Prerequisites of a course (P)
- Free/paid course (GP)

Based on this information, we will extract all the data that meet the criteria and prioritize them using the AHP method. After completing his apprenticeship, the learner is asked to answer some questions (satisfied or not, suggestions to improve the platform, etc.) in the form of a survey to ensure the improvement of our platform.

3.3 A Worked-Through Example Showing the Use of AHP Approach

We will take as an example the "IT security" course provided for SMI6 students. The objective of this course is to acquire the fundamental concepts of the IT security.

To ensure such an objective, the criteria mentioned above must be taken into consideration, namely language of learning (L), availability of courses (D), institution responsible for the course (E), level of competence/knowledge (C) , prerequisites of a course (P), and free/paid course (GP). Suppose that the possible courses are cours_1, cours_2, and cours_3. The diagram corresponding to this objective given in Fig. 3.

The next step is to peer-evaluate all the criteria and determine the decision criteria matrix while determining the importance of each criterion over another, according to the following scale of value (see Table 1).

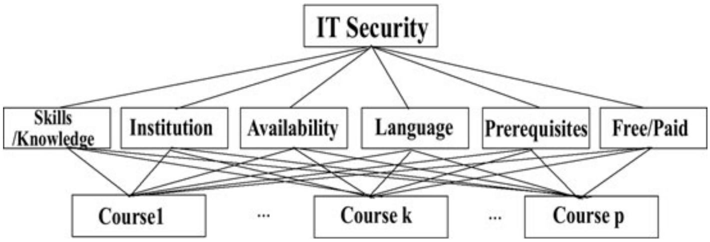


Fig. 3 A worked-through example of the course IT security

Table 1 Value scale proposed by Saaty [12]

Degrees of importance	Definition	Description
1	Equal importance of the two criteria	Two criteria contribute equally to the objective
3	One criterion is a little more important than the other	The experience and the personal judgment slightly favor one criterion compared to another
5	One criterion is much more important than the other	The experience and the personal judgment favor one criterion compared to another
7	One criterion is very strongly more important than the other	One criterion is strongly favored over another and its dominance is demonstrated in practice
9	One criterion is extremely more important than the other	The evidence favors one criterion over another
2,4,6,8	Intermediate values between two judgments	When we need to make a compromise between two criteria to refine the judgment

Table 2 Standardized matrix

	E	C	D	L	P	GP	Medium
E	0.043	0.044	0.015	0.015	0.531	0.206	0.142
C	0.257	0.265	0.131	0.149	0.177	0.265	0.207
D	0.386	0.265	0.131	0.522	0.035	0.088	0.238
L	0.009	0.132	0.026	0.075	0.059	0.147	0.075
P	0.300	0.265	0.654	0.224	0.177	0.265	0.314
GP	0.006	0.029	0.044	0.015	0.020	0.029	0.024

We notice that when a criterion "i" is given when it is compared to a criterion "j," it will have the opposite value when it will be compared to criterion "i." This allows us to build the following comparisons matrix:

$$\begin{array}{c|cccccc} & E & C & D & L & P & GP \\ \hline E & 1 & 1/6 & 1/9 & 1/5 & 3 & 7 \\ C & 6 & 1 & 1 & 2 & 1 & 9 \\ D & 9 & 1 & 1 & 7 & 1/5 & 3 \\ L & 1/5 & 1/2 & 1/5 & 1 & 1/3 & 5 \\ P & 7 & 1 & 5 & 3 & 1 & 9 \\ GP & 1/7 & 1/9 & 1/3 & 1/5 & 1/9 & 1 \end{array}$$

Subsequently, to obtain the weight of each of the criteria (the greater the value of the weight, the greater the importance of the criterion), it is necessary first of all to construct the standardized matrix; for this, it is necessary to calculate the sum of each column:

$$\sum_{i=1}^{n} C_{ij}$$

Then we divide each of the values of the column by this sum: $\frac{C_{ij}}{\sum_{i=1}^{n} C_{ij}}$. The standardized matrix is thus obtained as shown in the table below (see Table 2: standardized matrix)

Next, we calculate the weight of each criterion by calculating the average of the corresponding line:

$$\frac{\sum_{j=1}^{n} N_{ij}}{n}$$

with $n = 6$ in our case. We will obtain the following averages:

Criterion	*Average*
E	0.142
C	0.207
D	0.238
L	0.075
P	0.314
GP	0.024

The next step is to establish matrices for each decision criterion for each learner to compare courses in the area of network security.

Finally, we will obtain a matrix of solution that we will multiply by the matrix of averages. This last operation will allow us to classify the courses offered to each learner in priority order.

For the first version of the system, we emphasize that the framework allows to generate a list of MOOCs according to the needs of learners and the pedagogical criteria. In terms of the applicability of the approach, the preliminary results indicate that the method is useful and gave satisfactory result. In addition, we try in our framework to focus and to evolve learners. However, it was a bit difficult to ask them to use the framework and to be active in the process of learning in parallel with the presidential studies. We were forced to test our framework with a small number of learners.

We are aware that these preliminary results are not decisive. It remains for us to finalize the development of all the components of the system, add other courses, create courses, and involve many learners. Also, we will implement the fuzzy AHP and give a comparison of the two approaches.

4 Conclusions

In this work, we have proposed an overview of our framework based on Analytic Hierarchy Process (AHP) algorithm for an adaptive learning through the existing MOOCs. This suggested framework uses a variety of aspects including choice of learning path, learner satisfaction, and achievement to present an adaptive list of existing MOOCs to a learner.

We presented the preliminary results demonstrating the success of this framework in listing the most suitable MOOCs to a specific learner. For further validation, first, we plan to involve other members to our engineering team and implicate more learners in the evaluation of all components and improve our proposal based on the results of the assessments and feedback from these learners. Second, we plan to improve the pedagogical model, including more materials to make learning more efficient and attractive.

References

1. Afa, Y., Battou, A., & BAZ, O. (2017). Toward the MOOC capitalization using fuzzy analytical hierarchy process. *The International Journal of E-Learning and Educational Technologies in the Digital Media (IJEETDM)*, *3*(2).
2. Aissanou, F. (2012). *Décisions multicritères dans les réseaux de télécommunications autonomes*. HAL.
3. Ben Mena, S. (2000). Introduction aux méthodes multicritères d'aide à la décision. *Biotechnology, Agronomy and Society and Environment*, *4*(2), 83–93.
4. Bernardin, E. (2014). Business model for MOOCs: a perspective of reusability,. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*. New Orleans.
5. Hammami (2003). *Modelisation Technico-Economique D'une Chaine Logistique Dans Une Entreprise Reseau*. PhD thesis.
6. Hua, Z., Gong, B. X., & X. (2008). A DS–AHP approach for multi-attribute decision-making problem with incomplete information. *Expert Systems with Applications*, *34*, 2221–2227.
7. Keeney, R., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: J. Wiley.
8. Navneet, B., & Kanwal, R. (2004). *Strategic decision making: Applying the analytic hierarchy process* (1st edn.). Springer.
9. Park, Y. (2011). A pedagogical framework for mobile learning: Categorizing educational applications of mobile technologies into four types. *The International Review of Research in Open and Distributed Learning*, *12*(2), 78–102.
10. Reich, J. (2014). MOOCs and the science of learning. *Education Week*.
11. Roy, B. (1985). *Méthodologie multicritère d'aide à la décision*. Ed. Economica.
12. Saaty, R. W. (1987). The analytic hierarchy process – what it is and how it is used. *Mathematical Modelling*, *9*, 161–176.
13. Saaty, T. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, *48*, 9–26.
14. Saaty, T. (2007). Time dependent decision-making; dynamic priorities in the AHP/ANP: Generalizing from points to functions and from real to complex variables. *Mathematical and Computer Modelling*, *46*, 860–891.
15. Vargas, R. V. (2010). Using the analytic hierarchy process (AHP) to select and prioritize projects in a portfolio.
16. Zouggari, A. (2011). *Une approche couplant logique floue et capitalisation des connaissances pour la résolution du problème de choix des fournisseurs soutenues*.

Computational Fluid Dynamics (CFD) Analysis of the Effect of Anode Inlet Gas Humidity Variation on PEM Cell Performance Under $RH_C = 0\%$ and $RH_C = 100\%$

Yassine Amadane, Hamid Mounir, Abdellatif E. L. Marjani, and Samya Belarhzal

Abstract In this work, we consider a three-dimensional PEM fuel cell with a single channel. The cell is modeled by using computational fluid dynamics (CFD) simulation. The main goal of the present paper is to investigate the influence of the variation of cathode humidity under dry anode inlet gas on PEM fuel cell performance. The numerical results were in reasonable agreement with the experimental data. Our findings indicate that under $RH_C = 0\%$, the increased cell performance can be obtained by increasing the anode relative humidity from 10 to 90%. On the other hand, under $RH_C = 100\%$ the cell performance decreases when we increase the anode relative humidity from 10 to 90%.

Keywords PEM fuel cell · Anode relative humidity · Cathode relative humidity

1 Introduction

A proton exchange membrane fuel cell is a very important system that constituted different components: gas diffusion layers, catalyst layers, and membrane. Membrane plays a key role in PEM fuel cells. It has to be able to avoid the permeation of gases from the anode to the cathode but it also has to be able enough to limit ohmic losses resulting from protons transport [1]. A PEMFC utilizes hydrogen as fuel and oxygen as oxidant. Fuel cells have a large number of properties, such as high efficiency, no emissions, no noise, and potentially low cost, which may make them attractive in many applications [2]. In PEM fuel cells, a hydrogen humidification system may be required to avoid the dryness inside the PEM fuel

Y. Amadane (✉) · H. Mounir · A. E. L. Marjani · S. Belarhzal
EMISys Research Team, E3S Research Center, Mohammadia School of engineers, Mohammed V University in Rabat, Rabat, Morocco
e-mail: yassineamadane@research.emi.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_55

cell. Water management is a challenge in the PEM fuel cell because there is ohmic heating under high current flow, which will dry out the polymer membrane and slow ionic transport [3].

The first step in ANSYS Fluent Module is to define a three-dimensional geometry. The second step is to set up the mesh and then the simulation results. Many studies have investigated the effects of relative humidity on PEM fuel cell performance [4, 5].

In the present paper, this study seeks to investigate the influence of cathode relative humidity variation on PEM fuel cell performance under the dry anode.

Year	Author	Topic	Computational tools	Refs.
2011	Wong et al.	Inlet relative humidity control in PEM fuel cell	*ANSYSFLUENT*	[6]
2011	Jeon et al.	The effect of relative humidity of the cathode	*ANSYSFLUENT*	[7]
2014	Jian et al.	Influences of gas relative humidity	*ANSYFLUENT*	[8]
2016	Kim et al.	Numerical study on the effects of gas humidity	*ANSYSFLUENT*	[9]
2017	Cho et al.	Proton exchange membrane fuel cells (IT-PEMFCs) under reduced humidity conditions	*Experimental*	[10]

The overall paper is organized as follows: Section 2 presents computational fluid dynamics modeling. The simulation results and discussions are presented in Sect. 3. In this section, the simulation findings obtained are compared with the experimental data.

2 Computational Fluid Dynamics Modeling

2.1 Details of Geometry

The PEM fuel cell with a single channel was modeled using ANSYS 15 package software and the length of the cell considered is 50 mm [11–13]. The cell dimensions are presented in Table 1.

Figure 1 presents the three-dimensional geometry of the model created in ANSYS 15.

Table 1 Table captions should be placed above the tables

Components	Width (mm)	Height (mm)
Gas channels	2	0.9
Catalyst layers	4	0.015
Gas diffusion layers	4	0.4
Membrane	4	0.18

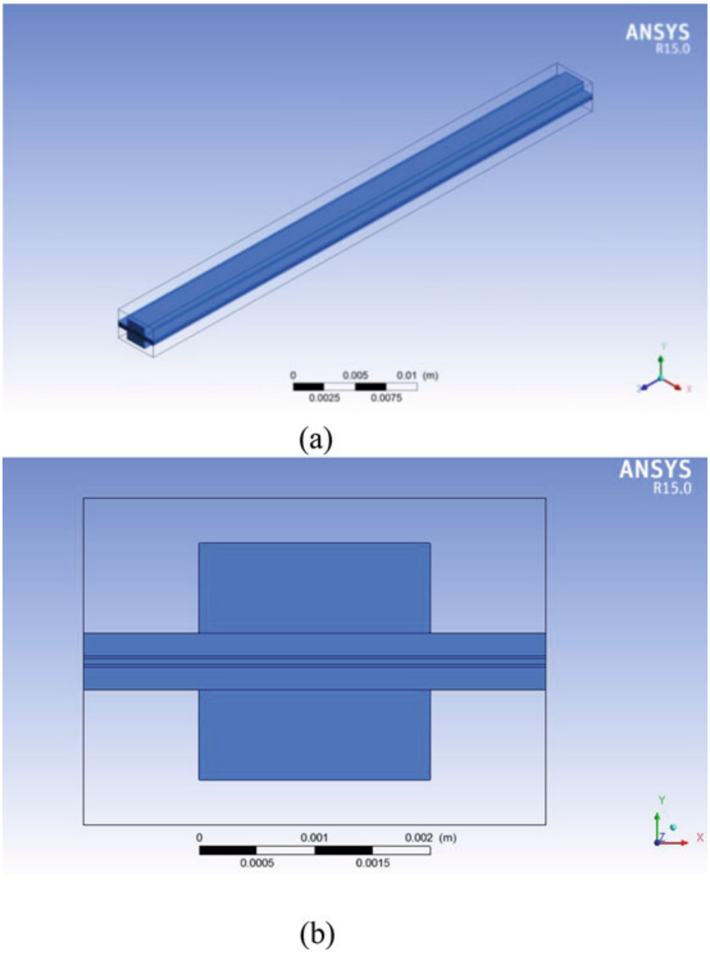


Fig. 1 Three dimensional design of a PEMFC, (**a**) 3-D geometry, (**b**) xy-plane

2.2 Mesh Report

The total number of elements in the simulation was 116,000. Table 2 presents the details of meshing.

Figure 2 shows the mesh created in ANSYS 15.

Table 2 The details of meshing

Domains	Nodes	Elements
Anode channel	9999	8000
Cathode channel	9999	8000
Anode catalyst layer	10,605	8000
Cathode catalyst layer	10,605	8000
Anode GDL	23,331	20,000
Cathode GDL	23,331	20,000
Membrane	14,847	12,000
All domains	143,319	116,000

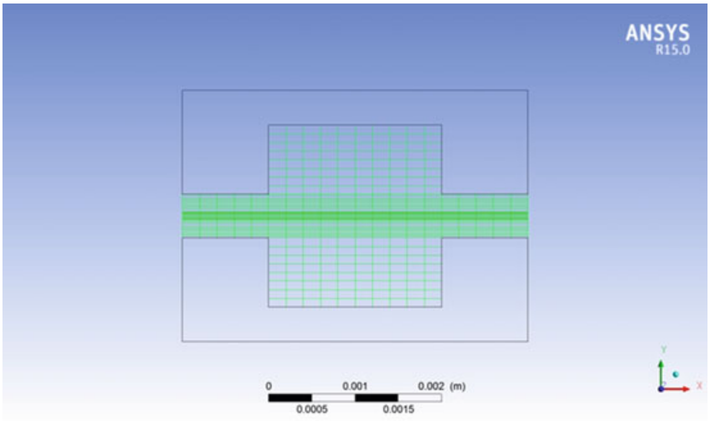


Fig. 2 The mesh of the cell simulated

Table 3 The governing equations

Description	Equations	Number
The mass conservation	$\nabla\left(\rho\vec{u}\right) = S_m$	1
The momentum conservation	$\nabla\left(\rho\vec{u}\,\vec{u}\right) = -\nabla p + \nabla\left(\mu\nabla\vec{u}\right) + S_u$	2
The species conservation	$\nabla\left(\vec{u}\,C_i\right) = \nabla\left(D_i^{\text{eff}}\nabla C_i\right) + S_i$	3
The energy conservation	$\nabla\left(\rho C_p\vec{u}\,T\right) = \nabla\left(k^{\text{eff}}\nabla T\right) + S_T$	4

2.3 Governing Equations

The following equations such as mass conservation, momentum conservation, species conservation, and energy conservation, respectively, are solved in ANSYS 15 fluent [16] (Table 3).

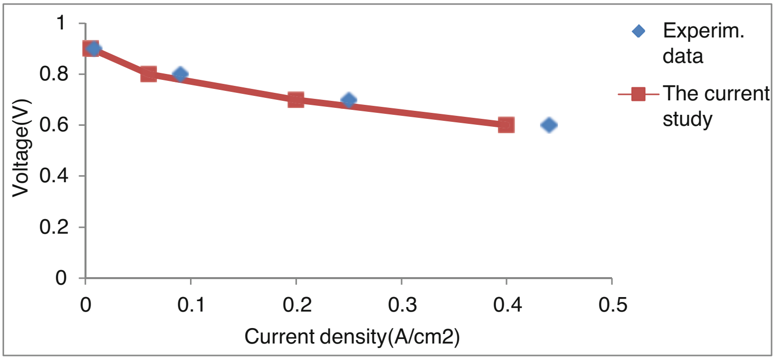


Fig. 3 The validation of the model with the experimental data

3 Results and Discussion

3.1 The Validation of the Numerical Model (Fig. 3)

The comparison of the current density and the experimental data shows that the simulation represents the performance of the cell. The maximum current density of simulation was 0.05 A/cm^2, while the maximum current density of experimental results was 0.06 A/cm^2 [14, 15].

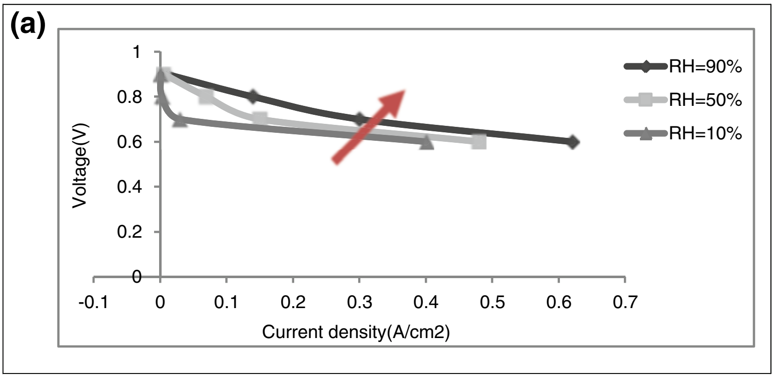
3.2 The Effect of Anode Relative Humidity

The present simulation findings show that the performance of PEMFC is influenced by the anode relative humidity when considering the following conditions $RH_C = 0\%$ and $RH_C = 100\%$, respectively [17–19].

From Fig. 4a it can be seen that increasing anode inlet humidity from 10 to 90% raises the polarization curve and then increases the cell performance while Fig. 4b shows that the cell performance decreases when we increase the anode inlet humidity from 10 to 90%.

4 Conclusion

Three-dimensional isothermal model was investigated by CFD simulation. This paper summarizes the main findings based on numerical simulation. The following conclusions are drawn:



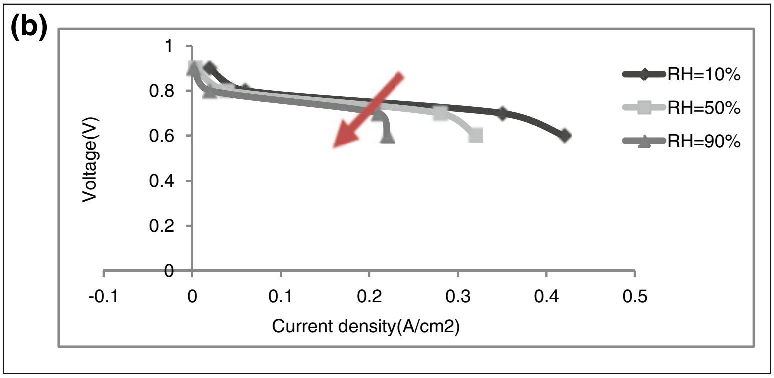


Fig. 4 The effect of the anode relative humidity under (**a**) $RH_C = 0\%$ (**b**) $RH_C = 100\%$

1. The simulation results were following in accordance with the experimental data.
2. The increased cell performance can be obtained by increasing the anode humidity from 10 to 90% under $RH_C = 0\%$.
3. The decreased cell performance can be obtained by increasing the anode relative humidity from 10 to 90% under $RH_C = 100\%$.

References

1. Barbir, F. (2005). *PEM fuel cells : Theory and practice*. Elsevier Academic press.
2. Perry, R. H., & Green, D. W. (1997). *Perry's chemical engineers' handbook* (7th ed.). McGraw Hill.
3. Wang, B., Wu, K., & Yang, Z. (2018). A quasi-2D transient model of proton exchange membrane fuel cell with anode recirculation. *Energy Conversion and Management, 171*, 1463–1475.

4. Li, Y., Pei, P., Wu, Z., Ren, P., Jia, X., Chen, D., & Huang, S. (2018). Approaches to avoid flooding in association with pressure drop in proton exchange membrane fuel cells. *Applied Energy, 224*, 42–51.
5. ANSYS, Inc. (2009). *ANSYS Fluent 12.0 fuel cells module manual*. ANSYS, Inc.
6. Wong, K., Loo, K., Lai, Y., Tan, S.-C., & Chi, K. T. (2011). A theoretical study of inlet relative humidity control in PEM fuel cell. *International Journal of Hydrogen Energy, 36*, 11871–11885.
7. Jeon, D. H., Kim, K. N., Baek, S. M., & Nam, J. H. (2011). The effect of relative humidity of the cathode on the performance and the uniformity of PEM fuel cells. *International Journal of Hydrogen Energy, 36*, 12499–12511.
8. Jian, Q. F., Ma, G. Q., & Qiu, X. L. (2014). Influences of gas relative humidity on the temperature of membrane in PEMFC with interdigitated flow field. *Renewable Energy, 62*, 129–136.
9. Kim, H.-Y., & Kim, K. (2016). Numerical study on the effects of gas humidity on proton-exchange membrane fuel cell performance. *International Journal of Hydrogen Energy*. https://doi.org/10.1016/j.ijhydene.2016.01.075
10. Cho, M. K., Park, H.-Y., Lee, S. Y., Lee, B.-S., Kim, H.-J., Henkensmeier, D., & Jang, J. H. (2017). Effect of catalyst layer ionomer content on performance of intermediate temperature proton exchange membrane fuel cells (IT-PEMFCs) under reduced. *Electrochimica Acta, 224*, 228–234.
11. Amadane, Y., Mounir, H., El Marjani, A., Ettouhami, M. K., & Awan, A. (2019). Numerical investigation of hydrogen consumption in proton exchange membrane fuel cell by using computation fuid dynamics(CFD) simulation. *Mediterranean Journal of Chemistry, 7*(6), 396–415.
12. Amadane, Y., Mounir, H., El Marjani, A., El Alaoui, R., Ettouhami, M. K., & Daoudi, K. (2019). Computational analysis of water distribution in a PEM fuel cell under different conditions. In *IEEE Conference (ICOA)*. https://doi.org/10.1109/icoa.2019.8727698
13. Amadane, Y., Mounir, H., El Marjani, A., & Ettouhami, M. K. (2018). A computational fluid dynamics (CFD) investigation of two PEM fuel cells with straight single-channel (SC): SC-40 mm and SC-50 mm. In *IEEE Conference (IRSEC)*. https://doi.org/10.1109/irsec.2018.8702826
14. Amadane, Y., Mounir, H., El Marjani, A., & Ettouhami, M. K. (2018). Numerical investigation of temperature and current density distribution on (PEM) fuel cell performance. In *IEEE Conference (IRSEC)*. https://doi.org/10.1109/irsec.2018.8702920
15. Amadane, Y., Mounir, H., El Marjani, A., Bouhrim, H., & Rafi, M. A. (2020). Taguchi approach in combination with CFD simulation as a technique for the optimization of the operating conditions of PEM fuel cells. *Arabian Journal for Science and Engineering, 45*, 7587–7597. https://doi.org/10.1007/s13369-020-04706-0
16. Belarhzal, S., & Boudi, E. M. (2020). Effect of profile shift factor on contact stress and natural frequencies of spur gear. *International Review on Modelling and Simulations (IREMOS), 13*, 6.
17. Amadane, Y., Mounir, H., El Marjani, A., & Ettouhami, M. K. (2019). Optimization of operating parameters affecting the performance of proton exchange membrane fuel cell (PEMFC) by using Taguchi method. In *Proceedings of 2019 International Conference of Computer Science and Renewable Energies, ICCSRE2019* (p. 8807782).
18. Amadane, Y., Mounir, H., El Marjani, A., & Ettouhami, M. K. (2020). *The effect of gas diffusion layer(GDL) porosity variation on oxygen distribution along the PEM fuel cell*. Lecture Notes in Electrical Engineering,624 LNEE (pp. 286–293).
19. Amadane, Y., Mounir, H., El Marjani, A., Ettouhami, M. K., & Atifi, A. (2020). *Using CFD simulation and Taguchi approach to optimize the parameters influencing the performance of PEM fuel cell with serpentine flow field*. Lecture Notes in Electrical Engineering,624 LNEE (pp. 306–312).

Analysis of Texture Feature Extraction Technique in Image Processing

Partha Chakraborty, Mohammad Abu Yousuf, Saidul Islam, Mahmuda Khatun, Aditi Sarker, and Saifur Rahman

Abstract Texture feature is one of the most common image segmentation, classification, extraction, and surface analysis techniques. In many image processing tasks, texture plays a vital role. The texture is defined by a neighborhood's spatial distribution of the gray level. For a point, however, texture cannot be described. The resolution at which an image is presented defines the perceived size of the texture. Image texture provides us with details on the color or intensity spatial arrangement in a picture or area chosen for a picture. Textures of images can be created artificially or found in natural scenes captured in an image. Nowadays, Gabor filtering has been widely used for texture feature extraction. In our paper, we presented a well-ordered extraction method of two-dimensional texture features and a comparison study between Gabor filer and Log Gabor filter. First, we converted the image to the gray level. Then on each converted part of the gray-level image, a two-dimensional Log Gabor filter with different frequencies decomposed with the SVD algorithm is applied to extract suitable distinctive texture information. We used SVD's unique values as a function vector to test the output of the proposed model. We used the Naïve Bayes classifier to train and test my experimental dataset for classifiers. We did the same things with Gabor filter and found a lower accuracy rate than Log Gabor filter.

Keywords Feature extraction. Image processing · Gabor filter · Singular value decomposition

P. Chakraborty (✉) · S. Islam · M. Khatun · A. Sarker · S. Rahman
Department of Computer Science and Engineering, Comilla University, Cumilla, Bangladesh
e-mail: partha.chak@cou.ac.bd; mahmuda@cou.ac.bd; aditisarker407@gmail.com; aditi1_cse@cou.ac.bd

M. A. Yousuf
Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh
e-mail: yousuf@juniv.edu

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_56

1 Introduction

Texture is diffusive in raw images and is a fast cue for a diversity of image analysis and computer vision applications like image segmentation, shape recovery from texture, and image retrieval. Texture primitives consist of micro-texture and macro-texture. Micro-texture is the smallest primitive while macro-texture is referred to larger primitive, i.e., macro-texture is composed of homogeneous aggregation of micro-texture. There is no clear criterion to differentiate micro-texture from macro-texture primitives. Texture plays an important role in many machine vision tasks such as surface inspection, scene classification, surface orientation, and shape determination. Feature extraction reduces the number of resources required to describe this thing with a large set of data. Generally, it does construct combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Texture is important for recognition, interpretation of data, computer vision, and image processing and pattern recognition [1, 2]. By using the extraction it is possible to solve some real-life problems. This problem is actually solved by a combination of feature detection, feature extraction, and matching [3, 4].

In recent years, big efforts have been devoted to the attempt to improve the performance of image retrieval systems and research has explored many different directions trying to use with profit results achieved in other areas. In recent years, texture feature is used in many works. In [5], they use texture feature extraction in region-based image retrieval system. Here their work is based on the Projection-Onto-Convex-Sets (POCS) theory. Texture feature extraction is widely used in the medical field nowadays. From [6], we find that fractal dimension is applied in texture feature extraction in X-ray chest image retrieval. They show that the retrieval system with the fractal algorithm has relatively high precision and recall up to about 85%. Based on a new Tamura feature extraction algorithm, the forensic image experiment happened in [7]. Similar to conventional 1D methods, nowadays, 2D texture feature extraction methods are used [8–14].

This paper specially targeted on the issue of establishing a new tactics which can help us to establish efficient operation of feature extraction and classification techniques. At first convert the sample to gray level image to achieve the texture information from the converted images. In this paper, we used 2D log Gabor and SVD for texture feature extraction. To extract detailed texture information, we applied a 2D log Gabor filter with various types of frequencies and different orientation angles are applied to each image to extract 2D log Gabor filtered images. Then decompose the SVD algorithm on each log Gabor filtered image to convert singular values of SVD as feature vectors and for classifier we used Naïve Bayes. In this paper we implemented a new image texture feature extraction process model. We used 2D Log Gabor filter (improved version of Gabor filter) and SVD to establish a new texture feature extraction method. We used six different category picture which was in total 1200 image to test, in proposed model. The categories we used are marble, upholstery, carpet, wallpaper, corduroy, and glass. First, we calculate how much accurate result actually the model show. We find very high accuracy percentage for each and every dataset. Like for marble we found

97.5% accuracy rate, upholstery also showed 97.5%, carpet showed 93.5% accurate result, wallpaper showed 91.5% accuracy rate, glass showed 95% correct result, and corduroy images showed 85.5% accurate result. We compared our Log Gabor and SVD based model with Gabor and SVD based texture feature extraction model and also some other existing model. After comparison with Gabor and SVD based model, the proposed model shows great result for each and every dataset. We then find some proved texture feature extraction model for same type of dataset. After comparing with that we found an extremely better result.

2 Related Work

Texture feature extraction is difficult task. Since texture is not a local phenomenon, one must consider a neighborhood of each pixel in order to classify that pixel exactly. In image processing and pattern recognition, feature extraction is an important step, which is a special form of dimensionality reduction. When the input data is too large to be processed and suspected to be redundant, then the data is transformed into a reduced set of feature representations. The process of transforming the input data into a set of features is called feature extraction. Features often contain information relative to color, shape, texture, or context. Nowadays we see there are many models for texture feature extraction. We found several papers [15, 16] about this topic. In every paper, we found different types of technique to implementing texture feature extraction. Like [17], they did design compact Gabor filter banks for an efficient texture feature extraction. In [18] they use Gabor filter and SVD based extraction method for state monitoring of induction motor. In [19] they did texture feature extraction using 2D Gabor Filter to extract the texture features of Inverse Fast Fourier Transform (IFFT), texture energy, and transformed IFFT. In [20] Gabor filter is used for efficient texture feature extraction. Firstly, to reduce computational complexity of texture feature extraction they use Gabor filter bank and after that Gabor filter bank thus designed produces low-dimensional feature representation with improved extract classification. In the paper [10], the author proposed a scale-invariant feature transform (SIFT) model. In [21, 22] spectral textural method is used for textural analysis and multi-spectral image classification. In [23] Wang used Gabor wavelet for feature extractor and for high level decision making they use Fuzzy Adaptive Resonance Theory (Fuzzy ART). In [24] they done comparison study between Gabor and log Gabor wavelet for texture analysis. In [25] M-band wavelet theory is introduced and for texture extraction M band wavelet histogram technology is applied. Every process has some defects like: SIFT has uncertainty in the number of key points and high computational cost [5], problem in wavelet texture analysis is that the number of features tends to become large, especially for wavelet packet decompositions [26]. An efficient image representation approach was proposed using a Log Gabor filter, where the Log Gabor filter possesses a number of interesting mathematical properties, e.g., it has a smooth and indefinitely differentiable shape, does not have side lobes, and generates high-dimensional representations. So, we used the Log Gabor and SVD based model to find the better result in CPU.

3 Methodology

Here is the basic workflow diagram of our model shown in Fig. 1. Firstly, we have collected a dataset of images with six categories. Then we have done feature extraction on those images. After that, we have kept those features extracted images on feature database. On the other side, we also collected several query images. Then done feature extraction of those images. After that we match the query images feature extracted sample with the feature database sample. And lastly, we tried to find how accurately this system defines each of the query images categories.

In Fig. 2 CPU implemented version. Here we first take images. After that we have checked images are in gray level or not. If they are not in gray level, we convert this image to gray level. Then we used 2D Log Gabor filter and SVD based model to extract feature information. Finally, the classification of the image shown on the CPU.

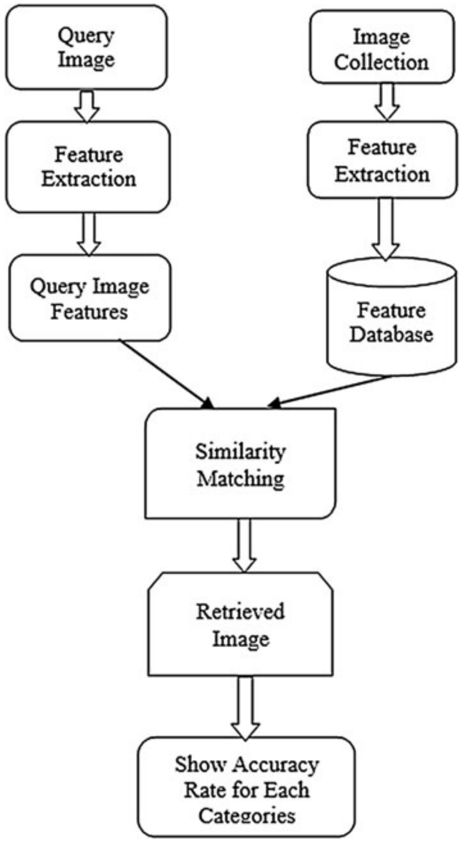


Fig. 1 Basic block diagram of model

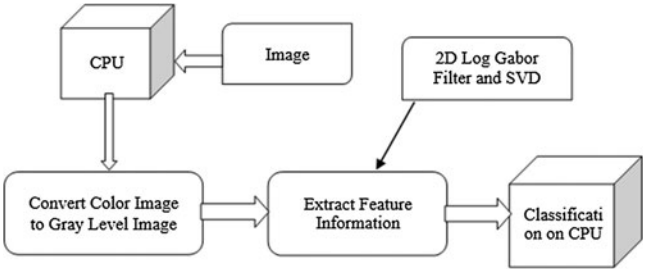


Fig. 2 CPU implemented workflow

3.1 2D Gabor Filter

Linear Gabor filter is used for identifying texture information in image processing. Frequency and orientation representations of Gabor filters are similar to those of the human visual system [33, 34], and they have been found to be particularly appropriate for texture representation. In this study, in order to extract meaningful texture information in a 2D gray level image, a 2D Gabor filter with various frequencies and orientation angles is used. In general, the 2D Gabor filter consists of a complex exponential centered at a given frequency and modulated by a Gaussian envelop. Because of the complex exponential, the filter has both real and imaginary parts [16, 25–27]. A Gabor filter can be formulated as:

$$g(x,y) = \frac{1}{\sqrt{\pi\sigma_x\sigma_y}} e^{-\frac{1}{2}\left[\frac{(x-x_0)_r^2}{\sigma_x{}^2}+\frac{(y-y_0)_r^2}{\sigma_y{}^2}\right]} e^{j(\omega_{0x}x+\omega_{0y}y+\theta)}, \tag{1}$$

where $\begin{bmatrix}(x-x_0)_r\\(y-y_0)_r\end{bmatrix} = \begin{bmatrix}\cos\alpha & \sin\alpha\\-\sin\alpha & \cos\alpha\end{bmatrix}\begin{bmatrix}(x-x_0)\\(y-y_0)\end{bmatrix}$, σ_x, $\sigma_y$are the variances along x and y axis, respectively, ω_0 is the angular frequency of the sinusoidal function, and θ represents the orientation of Gabor filter.

The filtering operation with a Gabor filter can be expressed as:

$$G(x,y) = I(x.y) * g(x,y), \tag{2}$$

where $G(x, y)$ denotes the complex convolution result. The magnitude of each pixel in a Gabor filtered image is calculated using the following equation [14].

$$m(x,y) = \sqrt{k_c^2(x,y) + k_S^2(x,y)}, \tag{3}$$

where $k_c = \mathrm{Re}\,(g(x,y) * i(x,y))$, $k_s = \mathrm{Im}\,(g(x,y)) * i(x,y)$ and * denotes a convolution operator. The frequency and orientation angle of the Gabor filter highly effect on the classification accuracy. Therefore, in order to extract more distinct texture features, a bank of Gabor filters with different scales and orientations are applied to the high frequency image to obtain Gabor filtered images.

3.2 2D Log Gabor Filter

2D Log Gabor filter is basically used for finding the texture information. 2D Log Gabor filter with various frequencies and orientation angles is used. It is the latest version of Gabor filter. It has complex exponential frequency.

The filtering operation with a log Gabor filter can be expressed as:

$$B = 2\sqrt{\frac{2}{\log(2)}}\left(\left\|\log\left(\frac{\sigma_f}{f_0}\right)\right\|\right). \tag{4}$$

The angular bandwidth is given by:

$$B_\theta = 2\sigma_\theta\sqrt{2\log 2}. \tag{5}$$

How 2D Log Gabor filter is implemented in our datasets are given below (Fig. 3)

3.3 Singular Value Decomposition

The singular value decomposition (SVD) is a factorization of a matrix (real or complex). SVD decreases a huge dimensional variable of data points to a lower dimensional space [8]. The singular values are used as a feature vector of the

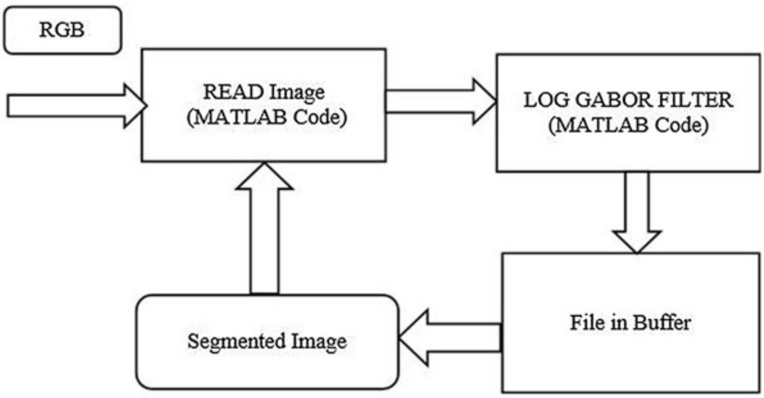


Fig. 3 Flowchart of 2D LOG Gabor

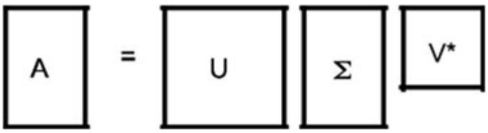
corresponding Log Gabor filtered image. If we have a rectangular matrix of size $m \times n$ that can be expressed by its SVD is:

$$A_{mn} = U_{mm} S_{mm} V_{nn}^{T}, \tag{6}$$

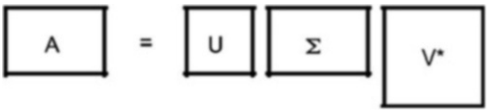
where $U^T U = I$, $V^T V = I$. Here we find columns of U are orthonormal eigenvectors of and the columns of V are orthogonal eigenvectors of $A^T A$ and is a diagonal matrix which actually contains the square roots of eigenvalues, $\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_m$ from U or V in descending order. The matrix S is represented as:

$$S = \begin{bmatrix} \sigma_1\, 0 \;\; 0\, 0\, 0 \\ 0 \;\; \sigma_2\, 0\, 0\, 0 \\ 0 \;\; 0 \;\; .\;\, 0\, 0 \\ 0 \;\; 0 \;\; 0\, .\;\, 0 \\ 0 \;\; 0 \;\; 0\, 0\, \sigma_m \end{bmatrix}.$$

The dimensions from σ_1 to σ_8 are stand for the computational complexity of a classifier, which increases with the increment of feature size. The first form of the singular value decomposition where $m < n$.



The second form of the singular value decomposition where $m \geq n$.

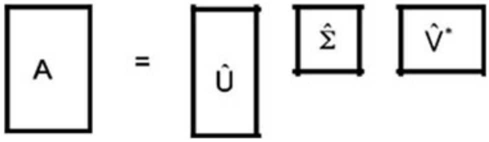


The third form of the singular value decomposition where $r \leq m < n$.

4 Experimental Setup and Result Analysis

4.1 Experimental Setup and CPU Implementation

During this part, we processed about 600 images on CPU to implement Log Gabor Filter and figure out the singular matrix of the image after implanting SVD



algorithm on the result of Log Gabor Filtered Image. Then, we used the Naïve Bayes Classifier to find out efficiency of our model. We used around 40% image of each category to test. After that, we used the entire image of each category to test to see how many of them the system can correctly guess which category they belong to.

4.2 *Result Analysis*

We showed and analyzed results for six categories of images here (Fig. 4). In Table 1 we showed result we got, while implementing on CPU.

Here, we have showed the result of processed image for marble category after implementing Log Gabor Filter on them (Fig. 5, Table 2).

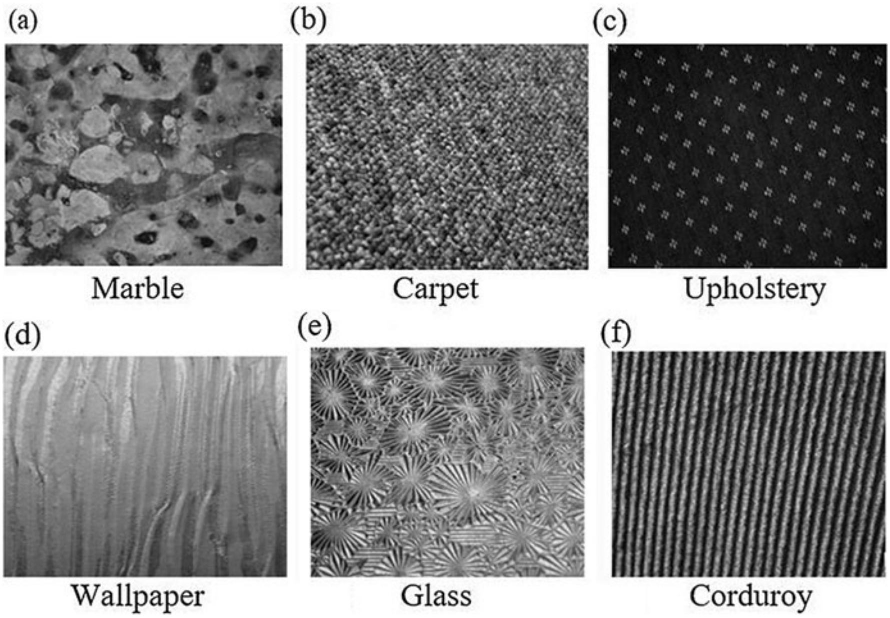


Fig. 4 Six types of images that were used in experimentation. (**a**) Marble, (**b**) carpet, (**c**) upholstery, (**d**) wallpaper, (**e**) glass, (**f**) corduroy

Table 1 Accuracy rate of the proposed model on testing different category images

Category	Total image	Percentage of image used in training (%)	Accuracy rate (%)
Carpet	100	40	93.5
Upholstery	100	40	97.5
Marble	100	40	97.5
Wallpaper	100	40	91.5
Glass	100	40	95
Corduroy	100	40	85.5

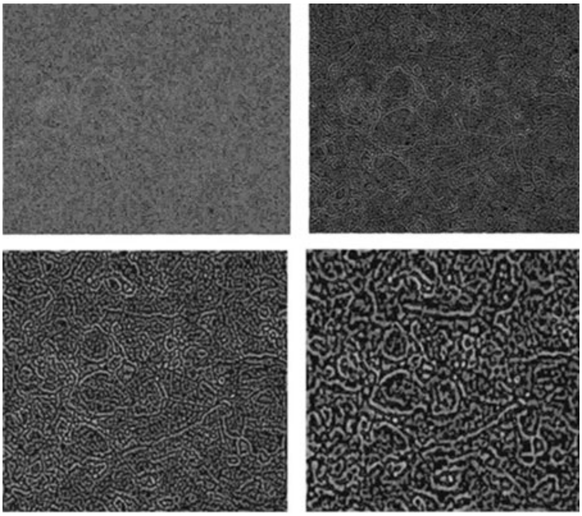


Fig. 5 Output of a processed image for marble after implementing Log Gabor Filter

Table 2 CPU image processing time on marble category

Image resolution	CPU (ms)
640 × 480	315.663417
480 × 360	294.648129
320 × 240	168.470668
240 × 180	157.740241

Table 3 CPU image processing time on carpet category

Image resolution	CPU (ms)
640 × 480	325.924430
480 × 360	298.564570
320 × 240	170.568921
240 × 180	158.452568

Then, we have processed the images for carpet using Log Gabor Filter (Fig. 6, Table 3).

After that, we have worked with the upholstery category. Results are shown below (Fig. 7, Table 4).

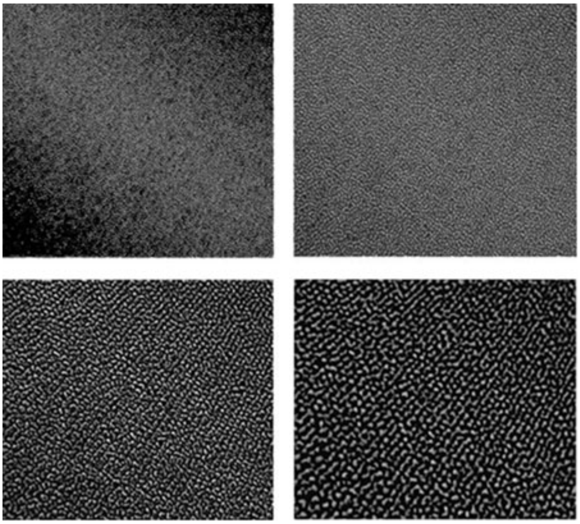


Fig. 6 Output of a processed image for carpet after implementing Log Gabor Filter

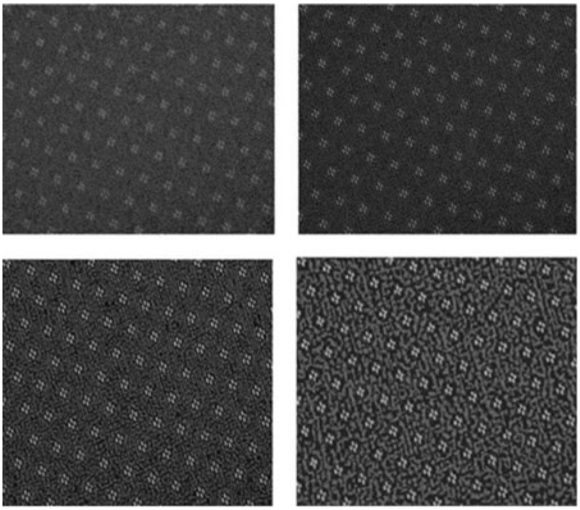


Fig. 7 Output of a processed image for upholstery after implementing Log Gabor Filter

Table 4 CPU image processing time on upholstery category

Image resolution	CPU (ms)
640 × 480	383.602666
480 × 360	315.258469
320 × 240	181.478963
240 × 180	165.213954

Fourthly, for wallpaper we did the same things as we used Log Gabor Filter on the images (Fig. 8, Table 5).

Finally, we filtered glass and corduroy images. We got almost same results for them.

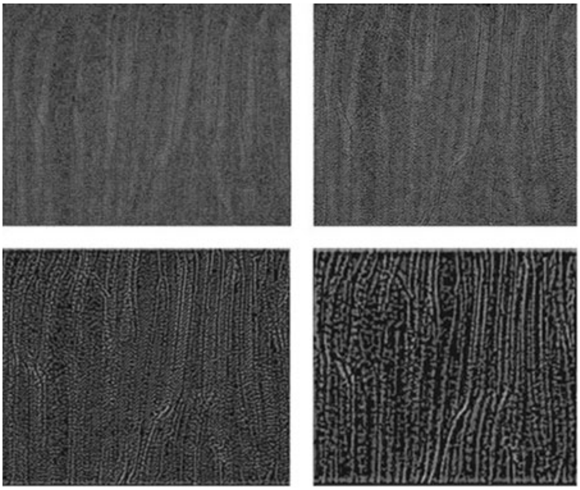


Fig. 8 Output of a processed image for wallpaper after implementing Log Gabor Filter

Table 5 CPU image processing time on wallpaper category

Image resolution	CPU (ms)
640 × 480	392.248952
480 × 360	318.54694
320 × 240	183.45644
240 × 180	167.413236

Table 6 Accuracy rate of Gabor filter and SVD based model on testing different category images

Category	Total image	Percentage of image used in training (%)	Accuracy rate (%)
Carpet	100	40	85.5
Upholstery	100	40	83.0
Marble	100	40	90.0
Wallpaper	100	40	82.0
Glass	100	40	89.25
Corduroy	100	40	81.65

4.3 *Result Comparison*

Now we will compare our proposed model to another similar model. We tested and trained with the same images that we used in our experiment with Gabor filter and SVD algorithm. We followed the same procedure that we used in CPU implementation of our model except a slight difference. Which is, instead of Log Gabor Filter we used here Gabor Filter. The rest remained the same. The difference is shown below in the Table 6.

Now, we will show some difference between some other models which used some of the very category of image that we used in our experimentation here. Firstly, we will talk about for marble type of images. The support vector machines (SVMs) method was originally invented by Vapnik [28]. In [27, 29] they used SVM classifier

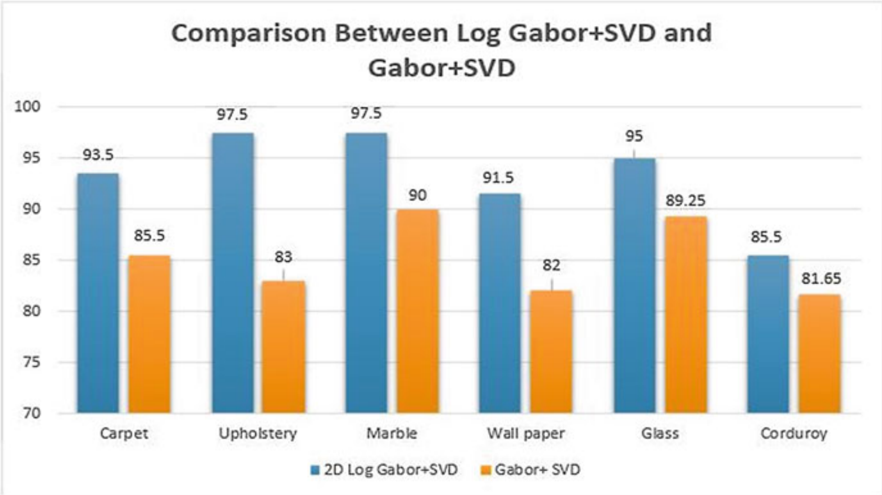


Fig. 9 Comparison between Log Gabor, SVD and Gabor, SVD model

for texture analysis. We found from [30] with this classifier they correctly guessed 96.01% using marble type of images. Also in [30] we found ANN classifier produce 94.44% correct result and MMD classifier will produce 83.61% correct result. From [31], we found the *K*-means algorithm produces correctly solved 74% of the total and the PAM (Partitioning around Medoids) algorithm shows a 75% average correct result. From [31], we found the *K*-means algorithm produced correctly solved 74% of the total and PAM The POM (Partitioning around Medoids) algorithm shows a 75% average correct answer.

But in our model, we showed for marbles our system accurately predicted 97.5% from all images. Secondly, for carpet wear we compared our system with existing system. In [32] we see that they use on support vector machine pattern recognition approach for carpet wear classification.

They show that SVM classifier is better than polynomial kernel approach because SVM show 92% correct labeling but in same place polynomial kernel approach only show 89.56% correct answer (Fig. 9).

Thirdly, we also found very high percentage of correct result for upholstery. Here we found 97.5% accuracy. But we cannot find any proposed model to compare with it. After that, same things goes for wallpaper. It shows 91.5% correct result but cannot find any existing model for comparison.

5 Conclusion

Due to continuous improvements of technology, usage of largest image dataset and resolution has been increasing rapidly. From the medical sector to geography or space, everywhere has its uses. Finding image from large dataset using old technology such as text-image retrieval becomes almost impossible. So in order to solve this problem, nowadays content-based image retrieval has been used instead of it. One of the many ways this technology is used is that it uses the texture feature for image classification and segmentation. There are many working models that use texture feature of image in image classification. But, in order to improve efficiency and save precious time on this model, we used a different combination of algorithms, which are the Log Gabor Filter and the SVD algorithm, to improve the efficiency rate and reduce time and computational complexity. As we demonstrated in the experimental results and analysis section.

With the techniques of texture representation and texture feature extraction in future, we would like to test the model with much larger dataset to see its efficiency. We would also like to test our dataset with GLCM approach and show an analysis between our model and GLCM.

References

1. Partha, C., Yousuf, M., Rahman, M. Z., & Faruqui, N. (2020). How can a robot calculate the level of visual focus of human's attention. In *Proceedings of International Joint Conference on Computational Intelligence. Algorithms for Intelligent Systems*. Springer. https://doi.org/10.1007/978-981-15-3607-6_27
2. Chakraborty, P., Yousuf, M. A., & Rahman, S. Predicting level of visual focus of human's attention using machine learning approaches. In *Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Advances in Intelligent Systems and Computing* (Vol. 1309). Springer. https://doi.org/10.1007/978-981-33-4673-4_5634
3. Sultana, M., Ahmed, T., Chakraborty, P., Khatun, M., Hasan, M. R., & Uddin, M. (2020). Object detection using template and HOG feature matching. *International Journal of Advanced Computer Science and Applications, 11*(7), 233–238. https://doi.org/10.14569/IJACSA.2020.0110730
4. Muzammel, C. S., Chakraborty, P., Akram, M. N., Ahammad, K., & Mohibullah, M. (2020). Zero-shot learning to detect object instances from unknown image sources. *International Journal of Innovative Technology and Exploring Engineering, 9*(4), 988–991.
5. Liu, Y., Zhang, D., Lu, G., & Ma, W.-Y. (2006). Study on texture feature extraction in region-based image retrieval system. In *12th International Multi-Media Modelling Conference Proceedings*.
6. Hong, S., & Huidong, D. (2012). Fractal dimension applied in texture feature extraction in x-ray chest image retrieval. In *2012 International Conference on Information and Automation (ICIA)*, Shenyang, June 6–8.
7. Liu, Y., & Li, Z. (2014). Study on texture feature extraction from forensic images with watermark. In *2014 IEEE 9th Conference on Industrial Electronics and Applications (ICIEA)*, Hangzhou, June 9–11.

8. Kang, M., & Kim, J. (2013). Singular value decomposition based feature extraction approaches for classifying faults of induction motors. *International Journal of Mechanical Systems and Signal Processing, 41*(1–2), 348–356.
9. Nguyen, D., Kang, M., Kim, C. H., & Kim, J. (2013). Highly reliable state monitoring system for induction motors using dominant features in a 2-dimension vibration signal. *New Review of Hypermedia and Multimedia, 13*(3–4), 245–258.
10. Do, V. T., & Chong, U. (2011). Signal model-based fault detection and diagnosis for induction motors using features of vibration signal in two dimension domain. *Journal of Mechanical Engineering, 57*(9), 655–666.
11. Shahriar, R., Ahsan, T., & Chong, U. (2013). Fault diagnosis of induction motors utilizing local binary pattern-based texture analysis. *EURASIP Journal on Image and Video Processing, 29*, 1–11.
12. Jang, W. C., Kang, M., & Kim, J. M. (2014). Fault classification of an induction motor using texture features of vibration signals. *LNEE, 314*, 177–183.
13. Nava, R. (2014). Overcomplete image representations for texture analysis. *Electronic Letters on Computer Vision and Image Analysis, 13*(2), 40–41.
14. Kang, M., & Kim, J. M. (2014). Reliable fault diagnosis of multiple induction motor defects using a two-dimensional representation of shannon wavelets. *IEEE Transactions on Magnetics* (accepted).
15. Chakraborty, P., Roy, D., Rahman, M. Z., & Rahman, S. (2019). Eye gaze controlled virtual keyboard. *International Journal of Recent Technology and Engineering, 8*(4), 3264–3269. https://doi.org/10.35940/ijrte.d8049.118419
16. Chakraborty, P., Muzammel, C. S., Khatun, M., Islam, S. F., & Rahman, S. (2020). Automatic student attendance system using face recognition. *International Journal of Engineering and Advanced Technology, 9*(3), 93–99. https://doi.org/10.35940/ijeat.b4207.029320
17. Li, W., Mao, K., Zhang, H., & Chai, T. (2010). Designing compact gabor filter banks for efficient texture feature extraction. In *2010 11th International Conference on Control Automation Robotics & Vision (ICARCV)*, Singapore, December 7–10.
18. Islam, M. R., & Kim, J.-M. (2014). An efficient Gabor filter and SVD based texture feature extraction method for state monitoring of induction motors. *International Journal of Multimedia and Ubiquitous Engineering, 9*, 10.
19. Roslan, R., & Jamil, N. (2012). Texture feature extraction using 2-D Gabor filters. In *2012 IEEE Symposium on Computer Applications and Industrial Electronics (ISCAIE)*, Kota Kinabalu, December 3–4.
20. Li, W., Mao, K., Zhang, H., & Chai, T. (2010). Selection of Gabor filters for improved texture feature extraction. In *2010 17th IEEE International Conference on Image Processing (ICIP)*, Hong Kong, September 26–29.
21. Liu, J., Zhu, W., Mou, M., & Wang, L. (2010). Cropland parcels extraction based on texture analysis and multi-spectral image classification. In *2010 18th International Conference on Geoinformatics*, Beijing, December 18–20.
22. Zortea, M., Tuia, D., Pacifici, F., & Plaza, A. (2010). Spectral-textural endmember extraction. In *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Reykjavik, June 14–16.
23. Wang, J., Naghdy, G., & Ogunbona, P. (1996). A new wavelet based ART network for texture classification. In *1996, Australian and New Zealand Conference on Intelligent Information Systems*, Adelaide, SA, November 18–20.
24. Comparison study of Gabor and log-Gabor wavelets for texture segmentation. In *2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA),* Dubrovnik, September 4–6, 2011.
25. Cheng, Q., Yang, C., Chen, F., & Shao, Z. (2004). Application of M-band wavelet theory to texture analysis in content-based aerial image retrieval. In *Proceedings. 2004 IEEE International Geoscience and Remote Sensing Symposium, 2004. IGARSS '04* (Vol. 3), Anchorage, AK, September 20–24.

26. Livens, S., Scheunders, P., Van de Wouwer, G., & Van Dyck, D. (2010). *Wavelets for texture analysis*. First International Conference on Networking and Computing.
27. Li, S., Kwok, J. T., Zhu, H., & Wang, Y. (2003). Texture classification using the support vector machines. *Pattern Recognition, 36*, 2883–2893.
28. Vapnik, V. (1995). *The nature of statistical learning theory*. Springer.
29. Alper Selver, M., & Akay, O. (2009). Evaluating clustering methods for classification of marble slabs in an automated industrial marble inspection system. In *International Conference on Electrical and Electronics Engineering, 2009. ELECO 2009*, Bursa, November 5–8.
30. Martinez-Alajarin, J. (2004). Supervised classification of marble textures using support vector machines. *Electronics Letters, 40*(11), 664–666.
31. Lazebnik, S., Schmid, C., & Ponce, J. (2005). A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(8), 1265–1278.
32. Bin, G. F., Gao, J. J., Li, X. J., & Dhillon, B. S. (2012). Early fault diagnosis of rotating machinery based on wavelet packets-Empirical mode decomposition feature extraction and neural network. *International Journal of Mechanical Systems and Signal Processing, 27*, 696–697.
33. Chakraborty, P., Ahmed, S., Yousuf, M. A., Azad, A., Alyami, S. A., & Moni, M. A. (2021). A Human-Robot interaction system calculating visual focus of Human's attention Level. IEEE Access, 9, 93409–93421.
34. Chakraborty, P., Yousuf, M. A., & Rahman, S. (2021). Predicting level of visual focus of human's attention using machine learning approaches. In Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Springer, Singapore 683–694.

Digital Filtering for Circumferential Wave Separation

Said Agounad, Younes Khandouch, and Abdelkader Elhanaoui

Abstract The excitation of a cylindrical structure by an acoustic wave gives rise to several circumferential waves. The received scattering signal is the result of different wave contributions. The analysis of this signal in time domain is difficult due to the superposition and overlapping of different wave contributions. In this chapter, we propose to use digital filtering in order to decompose the temporal signal into its components. This method is based on the isolation of the frequency content of each wave using the finite impulse response filter. The inverse Fourier transform of the isolated spectrum provides the contribution of the corresponding wave to the total temporal signal. Thereafter, the time–frequency representation is used as a tool for the verification of the isolated contribution. In addition, this representation can be used to identify the properties of the circumferential wave.

Keywords Numerical filtering · Circumferential waves · Time–frequency representation · Spectral analysis

1 Introduction

Due to the extensive use of cylindrical structures in engineering, several methods were developed to characterize these structures. The ultrasound method is one of the most employed methods [1–3]. The backscattering of an acoustic wave from cylindrical structures has been investigated theoretically [1, 4, 5] and exper-

S. Agounad (✉) · Y. Khandouch
Laboratory of Metrology and Information Processing, Department of Physic, Faculty of Sciences, Ibn Zohr University, Agadir, Morocco
e-mail: s.agounad@uiz.ac.ma; said.agounad@edu.uiz.ac.ma

A. Elhanaoui
Laboratory of Metrology and Information Processing, Department of Physic, Faculty of Sciences, Ibn Zohr University, Agadir, Morocco

ESIM, Faculty of Sciences and Techniques, Moulay Ismail University, Boutalamine, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_57

imentally [6–9]. Veksler [1] and Uberall [10] have studied the acoustic scattering of a plane wave by a cylindrical shell. They considered different angles of the emitter transducer with the normal to the shell. Mitri [11] analyzed the acoustic backscattering enhancements for incidence angles larger than 40° for different elastic and viscoelastic infinite cylinders. Rajabi et al. [12] revealed that the radiation force is a composition of three components: the background component, the resonant component and their interaction. Agounad et al. [3] used the time–frequency representation to investigate into the monostatic and bistatic evolutions of the acoustic scattering of a plane wave from cylindrical structures. Agounad et al. [13] employed Smoothed Pseudo Wigner–Ville to characterize an elastic tube from the predicted form function. They also used time–frequency properties, such as the group delay, to estimate the properties of the surface waves of the bilayered tube [14]. Maze [15] presented a spectroscopic method, which is based on the phase tangent of the Fourier transform. Agounad et al. [16] investigated the existence of zero group velocity waves (ZGV waves) in the submerged bilayered cylindrical shell. They concluded that the existence of these waves is a function of the thicknesses of materials constituting the shell. Elhanaoui et al. [17] analyzed the existence of the pseudo wave in the case of two-layered cylindrical shell. They explained the presence of this wave by the coupling and interaction phenomena between different circumferential waves.

This chapter proposes to isolate the contribution of each circumferential wave based on frequency filtering. The spectrum of the temporal signal can be clearly shown the resonances relative to each circumferential wave. Frequency filtering consists to choose a rectangular window with the cut-off frequencies fixed according to the properties of the wave which we want to isolate. A multiplication of the resonance spectrum by this window provides the isolated spectrum. Thereafter, the contribution of the circumferential wave is obtained by the inverse Fourier transform of the isolated spectrum. Among different time–frequency representations, we choose to use Smoothed Pseudo Wigner–Ville (SPWV) due to its interesting properties [5, 14]. This representation is used as a criterion for validation of the isolated contribution on the one hand and is employed to identify the properties of the circumferential waves on the other hand.

2 Materials and Methods

2.1 Access to the Temporal Signal

The temporal signal of the acoustic scattering from an elastic tube excited by an acoustic wave can be obtained either theoretically or experimentally (Fig. 1).

In the theoretical approach, the form function (f_∞) is calculated firstly from the following formula [1, 2]:

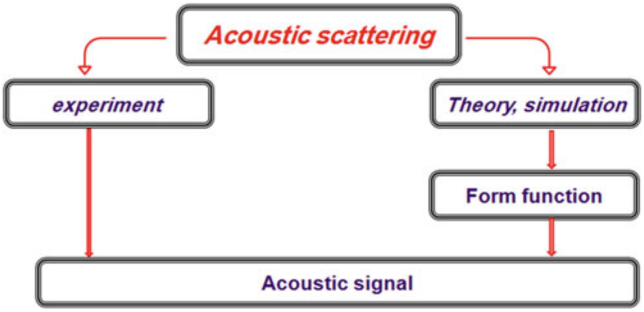


Fig. 1 Access to the temporal signal of acoustic scattering from cylindrical shell

$$f_\infty(\theta, x_1) \approx \lim_{r\to\infty} \sqrt{\frac{2r}{a_1}} \left|\frac{\phi_{scat}}{\phi_{inc}}\right|, \tag{1}$$

where ϕ_{inc} and ϕ_{scat} are the potentials of the incident and scattered waves, respectively. $x_1 = k_1 a_1$ is the reduced frequency, θ is the azimuthal angle, a_1 is the outer radius of the cylindrical shell, ϵ_n is the Neumann coefficient ($\epsilon_n = 1$ if $n = 0$ and $\epsilon_n = 2$ if $n > 0$) and r is the distance between the shell axis and the position of the emitter transducer. Replacing ϕ_{inc} and ϕ_{scat} by their expressions [1, 18, 19] and after some manipulations, the form function may be expressed as

$$f_\infty(\theta, x_1) = \frac{2}{\sqrt{\pi x_1}} \left|\sum_{n=0}^{\infty} \epsilon_n b_n \cos(n\theta)\right|, \tag{2}$$

where $b_n = D_n^{[1]}/D_n$ is an unknown coefficient, which characterizes the acoustic scattering by the composite cylindrical shell. $D_n^{[1]}$ and D_n are the two matrices computed from the solution of the wave equation and boundary conditions [1, 6, 18]. Secondly, the total acoustic signal is obtained by the inverse Fourier transform of the form function as

$$p_{scat}(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x_1) f_\infty(\theta, x_1) e^{(j\omega t)} d\omega, \tag{3}$$

where $f(x_1)$ is the bandpass of the transducer and generally is approximated by hamming function.

The experimental approach gives direct access to the acoustic signal. Figure 2 shows the experimental setup used to collect the temporal signal of the acoustic scattering from a cylindrical shell.

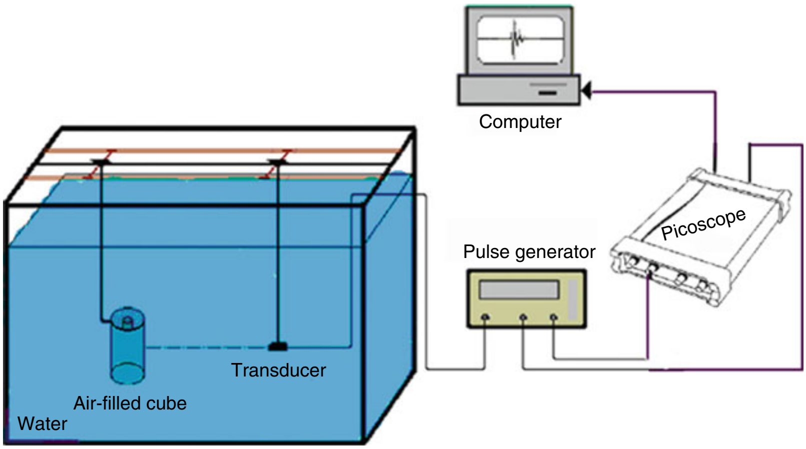


Fig. 2 Experimental setup

2.2 Numerical Filtering

The temporal signal of the acoustic scattering from cylindrical structure is a multi-component signal. Therefore, its analysis may be difficult due to the overlapping of different components (circumferential waves). Our goal in this chapter is to isolate the contribution of each circumferential wave to the total acoustic signal. For the considered shell, the resonances of different circumferential waves, which constitute the acoustic signal, are more or less distinguished in the frequency domain. By applying a frequency filtering, it is then possible to isolate the resonance frequencies relative to a specific circumferential wave. The partial temporal signal is then obtained from the inverse Fourier transform of the filtered resonance spectrum. This approach allows to decompose the temporal signal into its basic components which are the different circumferential waves. In this work, we propose to use the finite impulse response (FIR) filter because it is stable, characterized by a linear phase and simple to implement. The digital FIR filter is a discrete linear system and invariant in time, defined as [20]

$$z(n) = \sum_{i=-\infty}^{+\infty} h(i) p_{scat}(n - i), \tag{4}$$

where $p_{scat}(n)$ and $z(n)$ represent the input and output samples, respectively, and $h(n)$ is the impulse response of the filter. The design of the FIR filter is based on the identification of the impulse response h, which is real, causal and nonzero over finite duration. Several methods can be used to determine the impulse response h, such as the frequency domain method, the window design method, the Remez algorithm and the weighted least square design. In this work, we use the frequency domain method because it is the most efficient and accurate method. The transfer function of the FIR filter, $H(f) = FFT(h(n))$, is defined inside of a template which is determined as a function of circumferential wave characteristics to be eliminated.

2.3 Time–Frequency Representation

Many time–frequency methods have been proposed [21–24]. In this chapter, the Smoothed Pseudo Wigner–Ville (SPWV) and Gabor transform are used, since they are useful in acoustic application [5, 14].

The SPWV distribution has been proposed to reduce the interference terms that appear in the time–frequency plane of the Wigner–Ville distribution. The SPWV introduces two smoothing windows, one for frequency smoothing (h) and the second for temporal smoothing (g), in order to improve the readability of WV image and is given as

$$SPWV_{x_a}^{g,h}(t,\omega) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(\tau)g(t-\eta)x_a(\eta+\tau/2) \times x_a^*(\eta-\tau/2)e^{-j\omega\tau}d\eta d\tau, \tag{5}$$

where $x_a^*(t)$ is the complex conjugate of the analytical signal $x_a(t)$. The smoothing windows can be chosen as hamming, Blackman, etc.

The Gabor coefficients (G) for the signal $x(t)$ are calculated by the Gabor transform as

$$G(t,\omega) = \int_{-\infty}^{+\infty} x(\tau)e^{(-1/2((\tau-t)/\sigma)^2)}e^{-i\tau\omega}d\tau. \tag{6}$$

In practice, the time–frequency representation of the Gabor transform (GT) is given by the square modulus of these coefficients as $|G(t,\omega)|^2$.

3 Results and Discussion

Figure 3 shows the temporal signal of a stainless steel/polymer bilayered tube. The outer layer of this tube is characterized by the radius ratio $r_{21} = 0.92$ and the physical properties of the stainless steel (SST) (longitudinal velocity $C_{L,SST} = 5700$ m s^{-1}, transverse velocity $C_{T,SST} = 3100$ m s^{-1} and density $\rho_{SST} = 7900$ kg m^{-3}). The inner layer is identified by the radius ratio $r_{32} = 0.99$ and the physical properties of the polymer material ($C_{L,Pol} = 2500$ m s^{-1}, $C_{T,Pol} = 1200$ m s^{-1} and $\rho_{Pol} = 1500$ kg m^{-3}). The global radius ratio of the bilayered cylindrical shell is $r_{31} = 0.9108$.

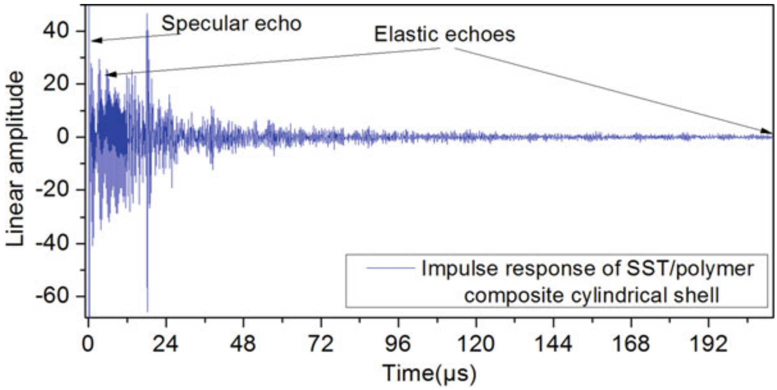


Fig. 3 Temporal signal of the stainless steel/polymer cylindrical shell

The first echo in Fig. 3 is the specular echo which is related to the reflection from the outer interface of the shell. This echo is saturated to observe the elastic echoes that are arrived later. These echoes are in relation with the radiation of the circumferential waves. Due to the overlapping of different echoes, it is difficult, perhaps impossible, to identify their arrival time which is very important for characterization and nondestructive testing of structures. This phenomenon constitutes the major drawback of the temporal analysis. To surmount this problem, we are going to isolate the contribution of each circumferential wave using digital filtering. The principle of this technique can be summarized in three steps.

- In the first step, the Fourier transform is used to calculate the resonance spectrum of the considered stainless steel/polymer cylindrical shell (Fig. 4b). This spectrum shows that the resonances of different circumferential waves are more or less isolated.
- The second step of this method aims to identify the low cut-off frequency (f_{lc}) and the high cut-off frequency (f_{hc}) of the filter template, $H(f)$ (rectangular window) from the resonance spectrum as a function of the properties of the wave which we want to conserve (e.g., for $M_{4,SST/Pol}$ wave, from Fig. 4b, we find $f_{lc,M_{4,SST/Pol}} = 26.02$ and $f_{hc,M_{4,SST/Pol}} = 79.03$, and for $M_{5,SST/Pol}$ wave, we find $f_{lc,M_{5,SST/Pol}} = 79.9$ and $f_{hc,M_{5,SST/Pol}} = 125.7$). The product of the filter template (Figs. 4a and 5a) by the resonance spectrum leads to isolate the resonances of each wave. Figures 4c and 5c show the filtered resonance spectra of $M_{4,SST/Pol}$ and $M_{5,SST/Pol}$, respectively.
- In the third step, the contribution of each circumferential wave to the total temporal signal of the considered structure is obtained by an inverse Fourier transform of the isolated spectrum (Figs. 6a and 7a depict the contributions of $M_{4,SST/Pol}$ and $M_{5,SST/Pol}$, respectively).

The time–frequency is thereafter used to compute the time–frequency image of each circumferential wave. Figures 6b and 7b show the time–frequency images of

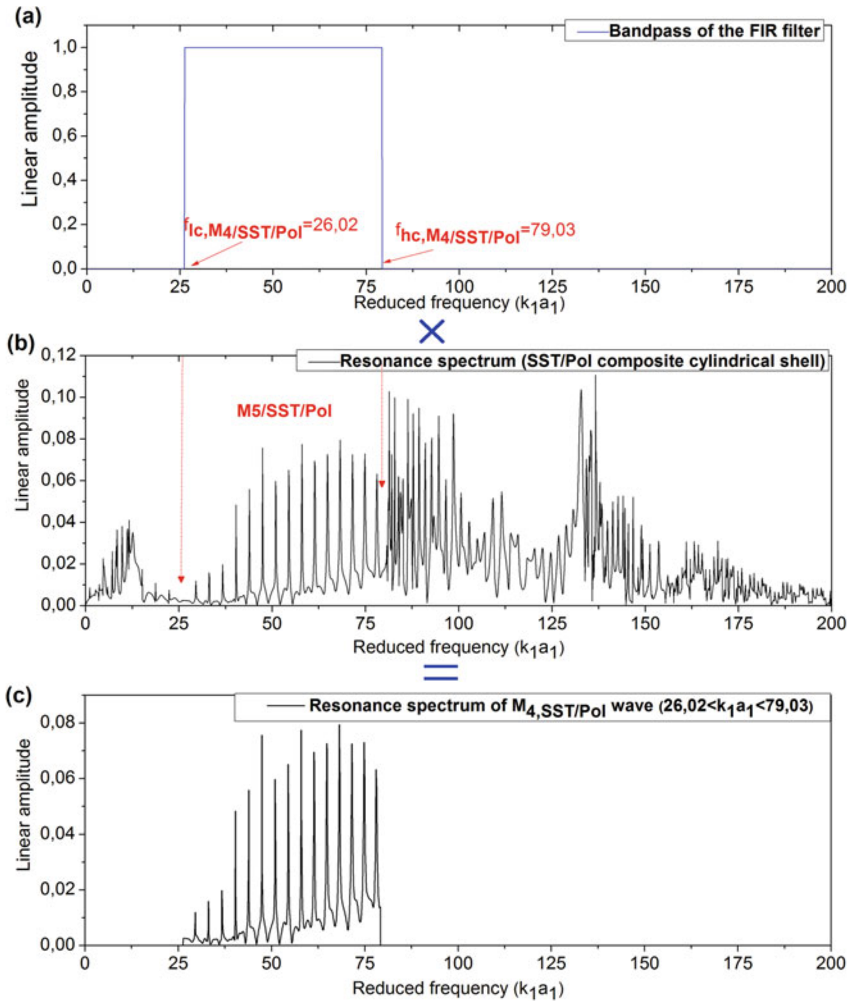


Fig. 4 Template of FIR filter (**a**), resonance spectrum (**b**) and isolated resonance spectrum of $M_{4,SST/Pol}$ wave (**c**) of the SST/pol bilayered tube

$M_{4,SST/Pol}$ and $M_{5,SST/Pol}$, respectively. The same process can be followed to isolate the contribution for each one of the other waves.

4 Conclusion

This chapter was concerned with the problem of echo overlapping in temporal plane. The frequency filtering was used to overcome this issue. The isolated

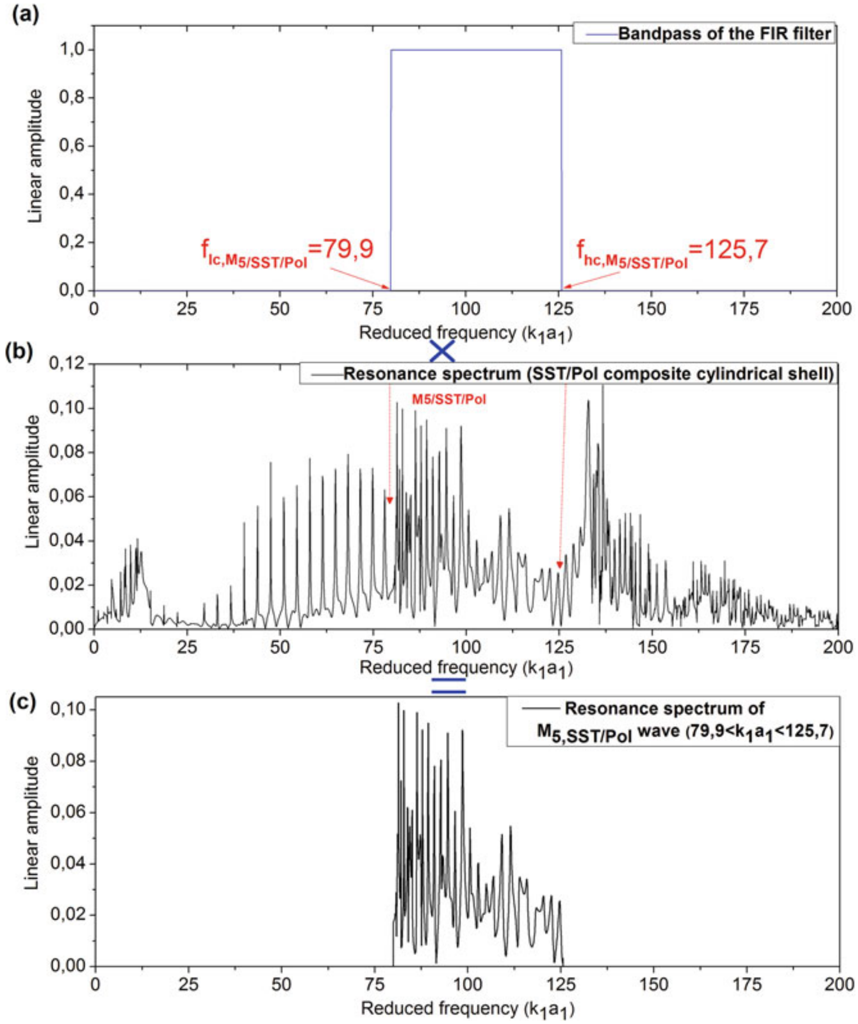


Fig. 5 Template of FIR filter (**a**), resonance spectrum (**b**) and isolated resonance spectrum of $M_{5,SST/Pol}$ wave (**c**) of the SST/pol bilayered tube

temporal signal of each wave showed a succession of echoes which more or less distinguished. This signal provides the possibility to identify the arrival time of echoes relative to a given wave. Moreover, this signal can be considered as an acoustic signature of the considered structure. The use of the time–frequency representation allows to obtain the time–frequency image of each circumferential wave.

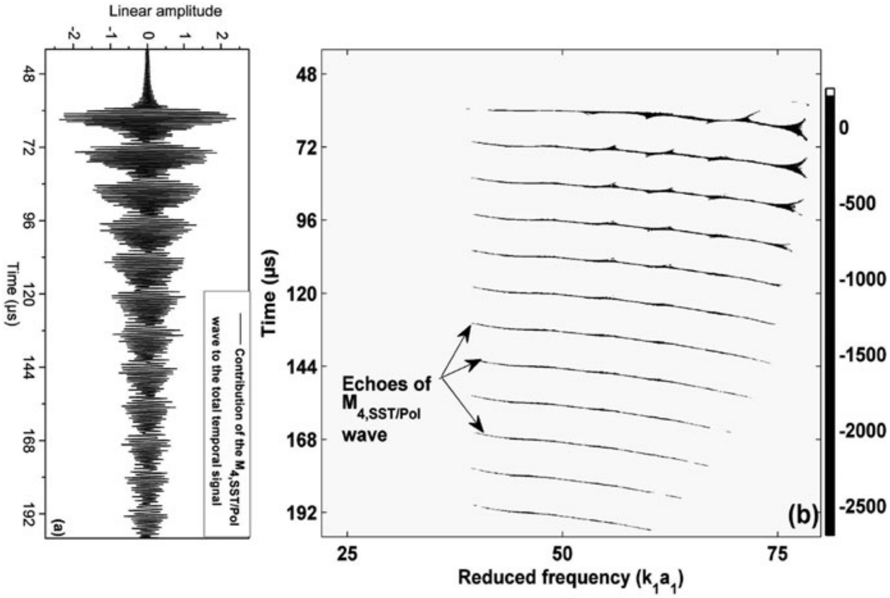


Fig. 6 (**a**) Contribution of $M_{4,SST/Pol}$ to the total temporal signal. (**b**) Time–frequency image of $M_{4,SST/Pol}$ obtained by the Gabor transform

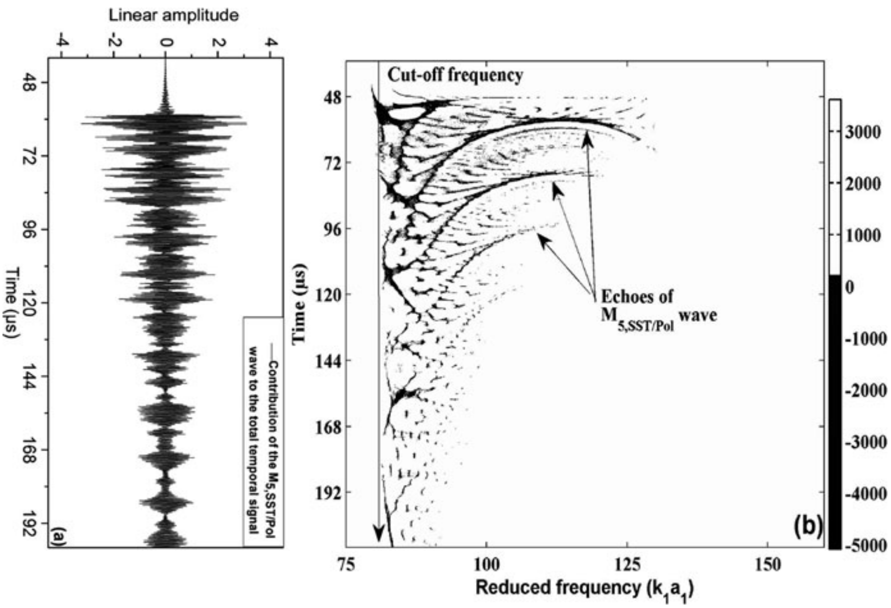


Fig. 7 Contribution of $M_{5,SST/Pol}$ to the total temporal signal (**a**). (**b**) Time–frequency image of $M_{5,SST/Pol}$ obtained by SPWV

References

1. Veksler, N. D. (2012). *Resonance acoustic spectroscopy* (vol. 11). Springer Science & Business Media.
2. Rajabi, M., Ahmadian, M. T., & Jamali, J. (2015). *Composite Structures, 128*, 395.
3. Agounad, S., Aassif, E. H., Khandouch, Y., Maze, G., & Décultot, D. (2018). *Journal of Sound and Vibration, 412*, 148.
4. Asadi, E., Wang, W., & Qatu, M. S. (2012). *Composite Structures, 94*(2), 494.
5. Dariouchy, A., Aassif, E. H., Decultot, D., & Maze, G. (2007). *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 54*(5), 1055–1064.
6. Chati, F., Léon, F., & Maze, G. (2005). *The Journal of the Acoustical Society of America, 118*(5), 2820.
7. Haumesser, L., Décultot, D., Léon, F., & Maze, G. (2002). *The Journal of the Acoustical Society of America, 111*(5), 2034.
8. Baillard, A., Chiumia, J., Décultot, D., Maze, G., Klauson, A., & Metsaveer, J. (2008). *The Journal of the Acoustical Society of America, 124*(4), 2061.
9. Van de Loock, J., Décultot, D., Léon, F., Chati, F., Maze, G., Rajaona, D. R., & Klauson, A. (2013). *The Journal of the Acoustical Society of America, 133*(1), EL26.
10. Überall, H. (1992). *Acoustic resonance scattering.* CRC Press.
11. Mitri, F. (2010). *Ultrasonics, 50*(7), 675.
12. Rajabi, M., & Behzad, M. (2014). *Ultrasonics, 54*(4), 971.
13. Agounad, S., Aassif, E. H., Khandouch, Y., Maze, G., & Décultot, D. (2018). *Ultrasonics, 83*, 222.
14. Agounad, S., El Houcein, A., Khandouch, Y., Décultot, D., & Maze, G. (2017). *2017 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)* (pp. 1–5). IEEE.
15. Maze, G. (1991). *The Journal of the Acoustical Society of America, 89*(6), 2559.
16. Agounad, S., Aassif, E. H., Khandouch, Y., Décultot, D., Maze, G., & Elhanaoui, A. (2017). *Composite Structures, 182*, 12.
17. Elhanaoui, A., Aassif, E., Maze, G., & Décultot, D. (2016). *Ultrasonics, 65*, 131.
18. Agounad, S., Aassif, E.H., Khandouch, Y., Décultot, D., & Maze, G. (2018). *Composite Structures, 187*, 439.
19. Ahmad, F., & Rahman, A. (2000). *International Journal of Engineering Science, 38*(3), 325.
20. Gan, W. S., & Kuo, S. M. (2007). *Embedded signal processing with the micro signal architecture.* John Wiley & Sons.
21. Flandrin, P., Auger, F., Chassande-Mottin, E., et al. (2003). *Applications in Time-Frequency Signal Processing, 5*(179–203), 102.
22. Prudhom, A., Antonino-Daviu, J., Razik, H., & Climente-Alarcon, V. (2017). *Mechanical Systems and Signal Processing*, *84*, 747.
23. Liu, N., Gao, J., Jiang, X., Zhang, Z., & Wang, Q. (2017). *IEEE Geoscience and Remote Sensing Letters, 14*, 127.
24. Pei, S. C., & Ding, J. J. (2007). *IEEE Transactions on Signal Processing, 55*(10), 4839.

A Simulation Study of PV Station Based on Polycrystalline Technology

Caouthar Bahanni, Mustapha Adar, Zakaria Khaouch, Mustapha Zekraoui, and Mustapha Mabrouki

Abstract In this work, a performance study of a grid-connected PV station under Fkih Ben Salah weather conditions was conducted using PVsyst software. The studied PV system is consisted of polycrystalline silicon (Pc-Si) solar cell technology. The results show that the total annual energy injected into grid was found to be around 779.8 MWh. The performance analysis depicted an annual average value of PR ratio with this technology of nearly 85.2%.

Keywords PV system · Pc-Si · Performance analysis · PVSyst

1 Introduction

Solar energy is a free and inexhaustible source of power that can provide alternative energy without polluting the environment; and therefore, its use reduces the decrease in energy reserves. Solar energy has a huge potential that exceeds fossil fuels and can meet the world's energy needs many times over.

Recently, research in the field of photovoltaic solar energy has been increasingly active. Most of the conducted researches are focused on two main areas: improving the conversion of solar radiation into electrical energy and DC to AC conversion at inverter level. As a result, the Photovoltaics become a fast-growing market with the compound annual growth rate of PV installations of 24% between years 2010 and 2017. Photovoltaic module production is in continuous growth, in 2017, China and Taiwan led with a share of 70%, followed by the rest of Asia-Pacific and Central Asia with 14.8%. Europe contributed 3.1%; the United States and Canada accounted for 3.7%. And the record lab cell efficiency is 26.7% for mono-crystalline and 22.3% for polycrystalline silicon wafer-based technology. The highest lab efficiency in thin film technology is 21.7% for CIGS and 21.0% for CdTe solar cells. PV system

C. Bahanni (✉) · M. Adar · Z. Khaouch · M. Zekraoui · M. Mabrouki
Industrial Engineering Laboratory, Faculty of Science and Technologies, Sultan Moulay Slimane University, Beni Mellal, Morocco
e-mail: mustapha.zekraoui@usms.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_58

performance has strongly improved. Before 2000 the typical performance ratio was about 70%, while today it is in the range of 80% to 90% [1].

Photovoltaic installations' performance analysis is very important because it ensures the monitoring of the installations by detecting anomalies. It reveals the impact of weather conditions, especially, temperature and dust, as well as losses at the level of inverters and cables.

The grid-connected photovoltaic system' electrical performance analysis is based on the international IEC 61724 [2] standard published by the International Electrotechnical Commission (IEC). It describes the performance parameters of photovoltaic installations including the Final Yield (Yf) and the Performance Ratio (PR). Final yield is used to compare the performance of PV systems installed at the same place using the same or different mounting structure. Performance ratio is widely used to analyze the performance and to compare PV systems located in various regions. Several studies are carried out to analyze the performance of photovoltaic installations based upon the performance ratio.

The aim of this paper is to simulate, using PVsyst software, the production of the photovoltaic station based on polycrystalline (pc-si). The performance analysis is made up based on performance parameters which have been specified by International Energy Agency (IEA) [2].

The remaining of this work is organized as follows: Sect. 2 provides a review of the aforementioned related works; Sect. 3 presents a detailed description of the PV plant; Sect. 4 introduces the expression and the definition of each performance parameter; Sect. 5 presents and discusses the results, and Sect. 6 concludes this work.

2 Review of Related Works

Monokroussos et al. [3], Huld et al. [4], Strobel et al. [5], and Zinßer et al. [6] show that the global radiance, ambient temperature, and the solar radiation spectrum are the most influential parameters on power generation. All of these parameters affect the operating conditions of PV modules, but the cells temperature is most crucial one. However, the effect of the temperature on operating solar cells is related to the temperature coefficient of each PV technology [7]. Furthermore, the accurate knowledge of the solar cells parameters dependence on irradiance and temperature is of vital importance for the performance assessment of photovoltaic modules [8]. The work by E. Radziemska [9] shows that for the crystalline silicon modules (c-Si), the performance decreases when temperature increases.

Performance Ratio (PR) is a globally accepted indicator to judge the performance of grid-connected PV plants [10].

It has been shown in paper [11] that the average value of the performance ratio of 993 residential PV systems in Belgium was found to be 78%. In Island, the performance ratio (PR) of a photovoltaic park, with a peak power of 171.36 kWp, ranged from 58 to 73%, giving an annual PR of 67.36% [12].

Table 1 Electrical characteristics

Modules	pc-Si	pc-Si
Module nominal power (W)	255	250
Module nominal open circuit voltage (V)	38	37.6
Module nominal voltage at maximum power (V)	30.0	30.5
Module nominal short circuit current (A)	8.88	8.81
Module nominal current at maximum power (A)	8.32	8.197
Temperature coefficient of power (per K)	−0.410%	−0.410%
Temperature coefficient open circuit voltage (per K)	−0.310%	−0.310%
Temperature coefficient short circuit current	0.051%	0.051%

Final yield is a good indicator to compare the performance of PV systems located at the same place and which use the same or differing mounting structure. PV plants are at the same location in Freiburg, Germany; with varying dates of installation, using different modules and inverters have shown a deviation of up to 10% in their final yield values [13].

The daily array yield is affected by types of PV module, solar radiation, and back surface module temperature. It is varied for different sites so it becomes important to predict daily array yield for grid and energy management [14].

In Emmanuel Kymakis [15], a correction to the module efficiency reduces the absolute percentage error between measured and predicted annual energy yield and performance ratio values to 4.89%, 4.94%, 1.16% and 4.34%, 4.93%, 1.88% for p-Si, HIT, and a-Si arrays, respectively. The performance comparison shows that HIT and a-Si arrays have performed better than p-Si array at this location. The energy yield of a-Si modules found to be 14% more in summer months and 6% less in winter months in comparison to p-Si modules. The HIT modules were found to consistently produce 4–12% more energy than p-Si modules. Table 1 gives a summary of all heading levels.

3 PV Plants Description

The polycrystalline photovoltaic system is installed in an olive farm located in Fkih Ben Salah, Beni Mellal, Morocco. This plant was designed for power generation in order to drive water pumps, for irrigation purposes.

Total installed panels are 1584 photovoltaic panels, including 12 rows of 3 lines, each of which consists of 44 solar panels. The panels of 6 rows (792 photovoltaic panels) have an output of 255 W, the other panels produce 250 W.

To drive the motors, 12 inverters are installed on site to convert the current from DC into AC.

The installed panels are from the Sunmodule plus SW from Solarworld brand. The modules, each of which containing 60 solar cells, are series-connected and present a 15% yield under standard test conditions. Every row is connected to the 3-phase Huawei SUN2000-33KTL inverter. Each row is formed by two strings connecting 66 modules in series. The unshaded modules were fixed with a tilt angle of 30°, facing south at an azimuth angle of 0°. More details are reported in Table 1.

4 Performance Parameters

In this section, we will introduce the performance parameter's expression and definition.

4.1 The Reference Yield (Yr)

The reference yield is the ratio of the total solar radiation Ht (kWh/m^2) arriving at the PV solar panels' surface and the reference radiation quantity G0 (kW/m^2). This parameter represents the number of hours during which the illumination is equal to that of the reference. It is recalled that Yr defines the solar resource for the PV system

$$\mathrm{Yr} = \mathrm{Ht}/\mathrm{G0} \tag{1}$$

4.2 The Array Yield (Ya)

The PV field efficiency is defined as the ratio between the total energy E_{DC} (kWh) generated by the PV system for a defined period (day, month, or year) and the rated power P_0 (kWp) of the system relative to the standard conditions (STC: irradiation: 1000 W/m^2, 25 °C, Ambient temperature and the reference spectrum AM 1.5-G). Algebraically, Ya is given by:

$$\mathrm{Ya} = E_{\mathrm{DC}}/P_0 \tag{2}$$

4.3 The Final Yield (Yf)

The final yield is the total energy produced by the PV system, E_{AC} (kWh) with respect to the nominal installed power P_0 (kWp). This quantity, which represents the number of hours during which the PV field operates at its nominal power, reads as:

$$\mathrm{Yf} = E_{\mathrm{AC}}/P_0 \tag{3}$$

4.4 Losses

The collection losses (LC): The collection LC losses are defined as the difference between the reference efficiency and the PV field efficiency. They represent losses due to panel temperatures, partial shading, spectral loss, staining, errors in research maximum power point, conversions (DC/AC), etc.

$$L_c = \text{Yr–Ya} \tag{4}$$

System losses (LS): The LS are due to the converting losses of the inverters (DC-AC). They are defined as the difference between the PV field yield (Ya) and the final yield (Yf).

$$L_s = \text{Ya} - \text{Yf} \tag{5}$$

4.5 The Performance Ratio (PR)

The PR indicates the overall effect of losses on the energy production of the PV system. The PR values indicate how a PV system approaches ideal performance under actual operating conditions. PR, which is a dimensionless quantity, is defined by the ratio between the final yield and the reference yield.

$$\text{PR} = \text{Yf}/\text{Yr} \tag{6}$$

5 Simulation Results and Discussion

PV_{syst} software has been used as a simulation tool to analyze the PV system station. It is the most used software for studying, designing, and analyzing data from various photovoltaic systems, including stand-alone, grid-connected, pumping and DC photovoltaic systems.

5.1 Meteorological Conditions

Figure 1 shows the monthly ambient temperature and the monthly horizontal global radiation. The average annual ambient temperature reported to be 18.32 °C. Temperature's maximum value is recorded to 23.47 °C in August and the lowest value was 12.84 °C in January. The monthly global horizontal radiance ranged from 78.2 kWh/m^2 in December to 219.8 kWh/m^2 in July.

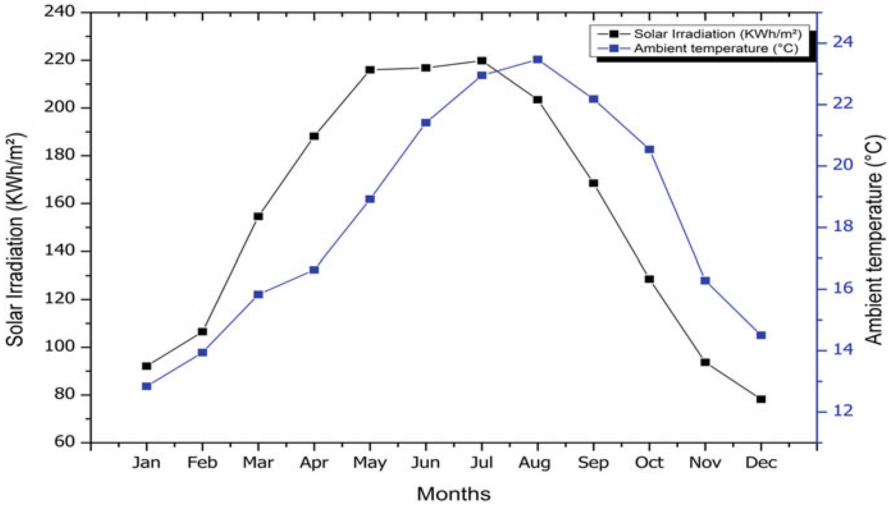


Fig. 1 Monthly ambient temperature and horizontal global radiation

5.2 PV Plants Production

Figure 2 depicts the monthly energy injected into the grid. It can be seen that the PV plant produces more energy in May and August with the values of 69.52 MWh and 68.29 MWh, respectively, while the lowest production was observed in December with the value 42.43 MWh. The annual energy injected into the grid by the PV station is 713.04 MWh.

5.3 Performance Analysis

Figure 3 shows that the annual average value of performance ratio is nearly 85.2%. The highest values of PR are observed in December and January with values of 88.3% and 88.2%, respectively, while its lowest values were observed in July, August, and September (summer) with the values of 83.1%, 82.9%, and 83.3, respectively. This can be explained by the decrease in temperature that minimizes the system losses.

Figure 4 presents system losses and collection losses. It appeared that the average daily energy losses are higher during the warmer months (summer), which explains the decreases in PR over this period. For losses in the system, the annual average value is about 0.08 KWh/KWp/day. The average annual value of collection losses is 0.77 KWh/KWp/day. It can be seen that the most important losses of the PV system lie at the capture level caused by the radiance and the array temperature.

The PV system's average annual final yield (YF) was found to be around 4.88 KWh/KWp/day.

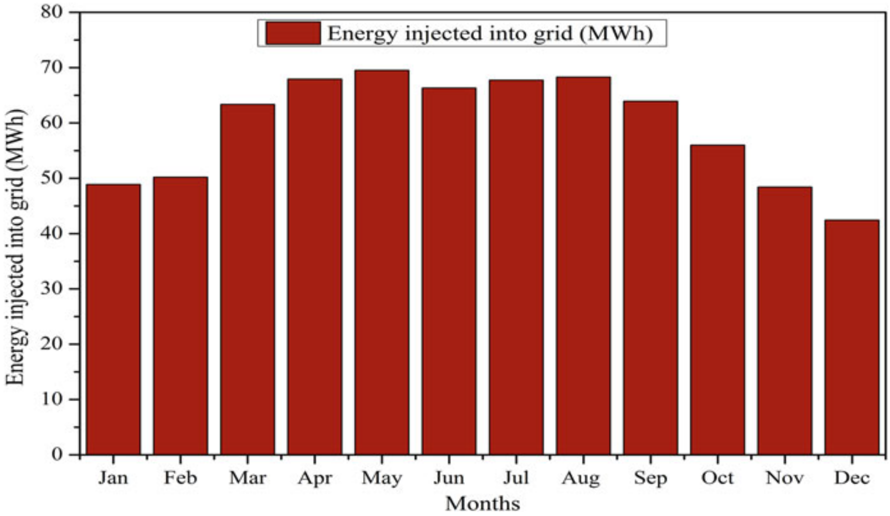


Fig. 2 Energy injected into the grid

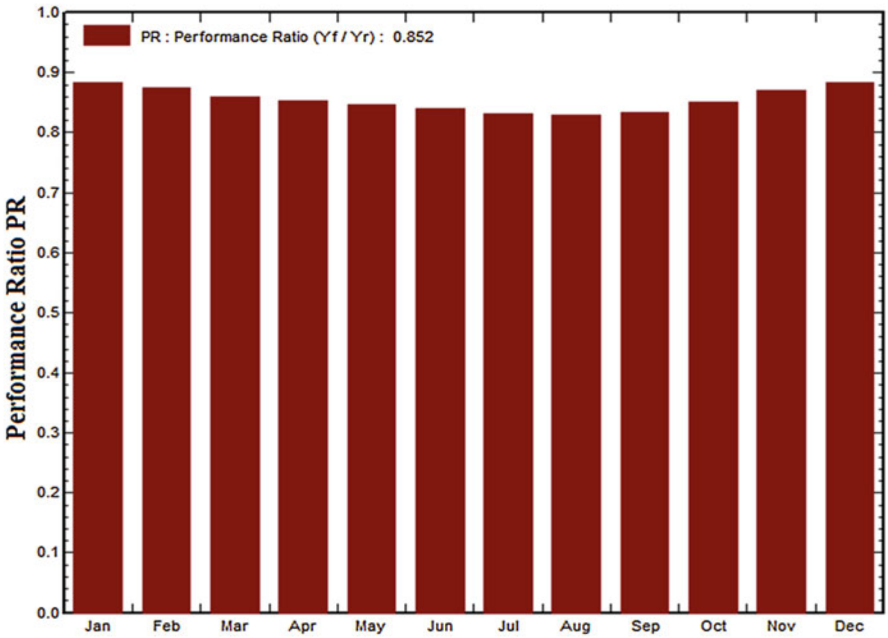


Fig. 3 Annual average value of performance

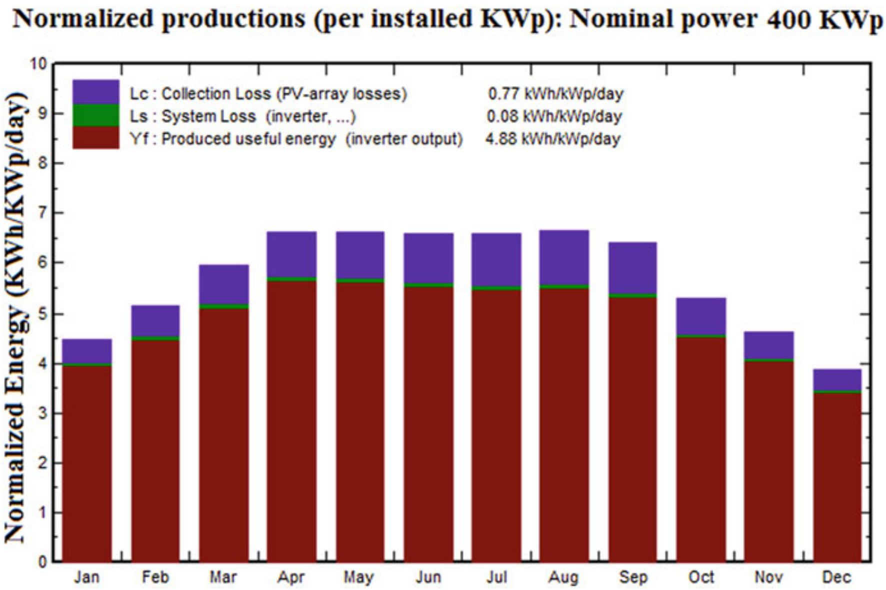


Fig. 4 Final yield and total losses of the system

6 Conclusions

In this paper, we have simulated the photovoltaic plant's electrical production behavior, based upon polycrystalline silicon technology panels. This study forecasts an annual production of around 779.8 MWh. The performance analysis was carried out, and it has showed that the annual average value of PR ratio for the PV station is nearly to 85.2%. The system losses' annual average value is about 0.08 KWh/KWp/day. The annual average value of the collection losses is found to be 0.77 KWh/KWp/day.

References

1. Fraunhofer ISE. (2018). *Photovoltaics report*, updated.
2. Anon. (1998). *Photovoltaic system performance monitoring-guidelines for measurements, data exchange and analysis*. IEC 61724.
3. Monokroussos, C., Bliss, M., Qiu, Y., Hibberd, C., Betts, T., Tiwari, A., & Gottschalg, R. (2011). Effects of spectrum on the power rating of amorphous silicon photovoltaic devices. *Progress in Photovoltaics: Research and Applications, 19*, 640–648I.
4. Huld, T., Gottschalg, R., Beyer, H. G., & Topic, M. (2010). Mapping the performance of PV modules, effects of module type and data averaging. *Solar Energy, 84*, 324–338.
5. Strobel, M. B., Betts, T. R., Friesen, G., Beyer, H. G., & Gottschalg, R. (2009). Uncertainty in photovoltaic performance parameters—Dependence on location and material. *Solar Energy Materials & Solar Cells, 93*, 1124–1128.

6. Zinßer, B., Makrides, G., Schubert, M., Georghiou, G., & Werner, J. H. (2008). Temperature and intensity dependence of twelve photovoltaic technologies. In *Proceedings of the 23rd European photovoltaic solar energy conference* (pp. 3249–3252).
7. Makrides, G., Zinsser, B., Georghiou, G. E., Schubert, M., & Werner, J. H. (2009). Temperature behaviour of different photovoltaic systems installed in Cyprus and Germany. *Solar Energy Materials and Solar Cells, 93*, 1095–1099.
8. Febba, D. M., Rubinger, R. M., Oliveira, A. F., & Bortoni, E. C. (2018). Impacts of temperature and irradiance on polycrystalline silicon solar cells parameters. *Solar Energy, 174*, 628–639.
9. Radziemska, E. (2003). The effect of temperature on the power drop in crystalline silicon solar cells. *Renewable Energy, 28*, 1–12.
10. Khalid, A. M., Indradip, M., Warmuth, W., & Schacht, V. (2016). Performance ratio—Crucial parameter for grid connected PV plants. *Renewable and Sustainable Energy Reviews, 65*, 1139–1158.
11. Makrides, G., Zinsser, B., Phinikarides, A., Schubert, M., & Georghiou, G. E. (2012). Temperature and thermal annealing effects on different photovoltaic technologies. *Renewable Energy, 43*, 407–417.
12. Leloux, J., Narvarte, L., & Trebosc, D. (2012). Review of the performance of residential PV systems in Belgium. *Renewable and Sustainable Energy Reviews, 16*, 178–184.
13. Photovoltaics Report. (2015). *Fraunhofer Institute for Solar Energy Systems*. ISE, Freiburg. https://www.ise.fraunhofer.de/de/downloads/pdf-files/aktuelles/photovoltaics-report-in-englischer-sprache.pdf.
14. Wan, C., Zhao, J., Song, Y., Xu, Z., Lin, J., & Hu, Z. (2015). Photovoltaic and solar power forecasting for smart grid energy management. *CSEE Journal of Power and Energy Systems, 1*(4), 38–46.
15. Kymakis, E., Kalykakis, S., & Papazoglou, T. M. (2009). Performance analysis of a grid connected photovoltaic park on the island of Crete. *Energy Conversion and Management, 50*, 433–438.

Investigation of Amazigh Speech Recognition Performance Based on G711 and GSM Codecs

Mohamed Hamidi, Hassan Satori, Ouissam Zealouk, and Khalid Satori

Abstract In this chapter, we evaluate the Amazigh alphabets speech performance through an interactive voice response (IVR) system based on the losses in the G711 and GSM coding schemes. To investigate the effect of voice coding on the speech recognition rate, our designed system was trained to recognize the 33 Amazigh alphabets by using 3 and 5 HMM states with different Gaussian mixture models (GMMs) and the mel-frequency spectral coefficients (MFCCs) used to extract the feature. The speech corpus contains a total of 15,840 alphabets recorded by 30 (15 male and 15 female) Amazigh native speakers. The proposed Amazigh speech recognition system is based on the Carnegie Mellon University (CMU) Sphinx tools. Our results indicate that the best system performance is found for the G711 codec, 3 HMM, and 16 GMMs.

Keywords Speech recognition · IVR · Telephony server · Asterisk server · Amazigh language

1 Introduction

In the last decade, modern human life was totally dependent on technologies such as interactive voice response (IVR) systems which become a central data source for natural language processing applications. Also, with the evolution of automatic speech recognition (ASR) and natural language processing, several spoken dialogue systems (SDS) appear in our lives as information assistants [1].

IVR is now a commonplace in different fields like medicine where Lieberman et al. [2] investigate the ability to track the illness progression by using interactive voice response technology. They developed a system which allows symptom monitoring and as an adjunct to the treatment of chronic pain. On the other hand, the authors investigate the possibility of conducting a high-quality survey about

M. Hamidi (✉) · H. Satori · O. Zealouk · K. Satori
LISAC, Department of Mathematics and Computer Science, FSDM, USMBA, Fez, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_59

problematic alcohol and drug use in the general population based on Internet and IVR technologies [3]. In addition, the IVR solution is used as a proactive outreach for engaging the low-income relapsed smokers in a new treatment cycle [4].

Rose et al. [5] present the integration of the ASR system with VoIP-based Asterisk PBX server. In their work, HTK is used with Asterisk server. IAX and SIP protocols are utilized to make calls and communicate naturally. Aust et al. [6] have created an automatic system that permits users to ask for train traffic information using the telephone. This system connects 1200 German cities. The caller can retrieve information talking fluently with the system which behaves like a human operator. The important components of their system are speech recognition, speech understanding, dialogue control, and speech output which is executed sequentially. Bhat et al. [1] created the Speech Enabled Railway Enquiry System (SERES) which is a system that permits users to get the railway information considering the Indian scenario, as a case study to define issues that need to be fixed in order to enable a usable speech-based IVR solution.

In another ASR study [7], authors have studied the classification of speech communication channels using MFCC features and GMM-based ML classifier. They utilize various databases from several sources to build and test models. Their obtained accuracy is about 95%. The researchers have examined the robust speech recognition in the GSM mobile environment. They are focused on voice degradation due to the losses in the GSM coding platform [8, 9]. Basu et al. [10] have described the real-time challenges of designing the telephonic automatic speech recognition system. In their study, authors have used the asterisk server to design a system that asks some queries, and the spoken responses of users are stored and transcribed manually for ASR system training. In this work, the speech data are collected from West Bengal.

Satori et al. [11] have created a system based on HMM (Hidden Markov Models) using the CMU Sphinx tools. The aim of this work is the creation of automatic Amazigh speech recognition system that includes digits and alphabets of Amazigh language. The system performance achieved was 92.89%. In [12, 13], we present our first experiment to integrate the ten first digits of Amazigh language in an interactive voice response (IVR) server where the users use speech (ten first Amazigh digits) to interact with the system. In general, the Amazigh speech recognition for different fields was targeted by researchers [14–17].

In this work, we compare the VoIP Amazigh ASR system performance by varying the values of their respective parameters as codecs, HMMs, and GMMs in order to determine the influence of codecs on the system recognition rates. Also, we aim to increase the recognition accuracy by varying the different automatic speech recognition parameters for both speaker-independent and speaker-dependent.

The rest of this chapter is organized as follows: Sect. 2 presents an overview of the VoIP system and protocols. Section 3 gives an overview of automatic speech recognition system. In Sect. 4, Amazigh language is explained. In Sect. 5, Telephony Amazigh speech recognition will be discussed. Finally, Sect. 6 is experimental results. We finish with some conclusion.

2 VoIP System and Protocols

VoIP (voice over Internet protocol) is a technology during the last decade. It provides audio and video streaming facility on successful implementation in the network.

2.1 Asterisk Server

Telephony Server Asterisk is an open source and a development environment for various telecommunication applications programmed in C language. It provides establishment procedures enabling to manipulate communication sessions in progress. Asterisk supports the standard protocols: SIP, H.323, and MGCP and transformations between these protocols. It can use the IAX2 protocol to communicate with other Asterisk servers [18, 19].

2.2 Session Initiation Protocol

The session initiation protocol (SIP) is a signaling protocol which is responsible for creating media sessions between two or more participants. SIP was defined by Internet Engineering Task Force (IETF) and is simpler than H.323 and adapted more specifically for session establishment and termination in VOIP [20]. In our word, SIP was used to create the user account and to assure internetwork communication.

2.3 Real-Time Transport Protocol

The real-time transport protocol (RTP) is an Internet protocol that allows transmitting real-time data such as audio and video. RTP is a protocol which facilitates the transport of data over a network in real-time applications. It is intended to be used for applications such as audio and video conferencing, real-time systems control, and unicast or multicast services [21].

2.4 Codec

The codecs are basically different tools of mathematical used for encoding or compressing the analogue voice signal into digital bit streams and back. The various

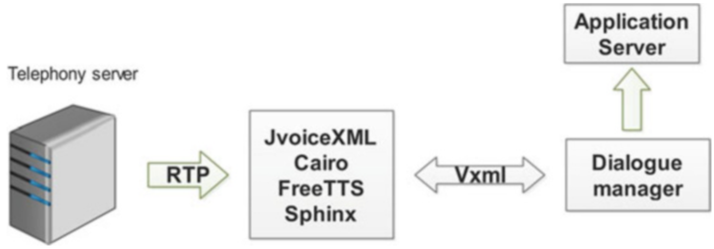


Fig. 1 Spoken dialogue system architecture

codecs are based on an algorithm of compression, data rate, and the sampling rate [22]. The VoIP codecs used in this work are G.711-u and GSM.

2.5 Spoken Dialogue System

A spoken dialogue system (SDS) is an interactive computer system, where a dialogue between the user and the computer is achieved [23]. The speech is received by the system from the user, and the response is given as an action or information. The spoken dialogue system architecture includes several components integrated together such as telephony, speech technologies, and web technologies. The scheme of spoken dialogue system is shown in Fig. 1.

3 Automatic Speech Recognition System

3.1 Speech Recognition

Speech recognition is the process of decoding the speech signal captured by the microphone and converting it into words [24]. The recognized words can be used as commands, data entry, or application control. Recently, this technology has reached a higher level of performance. The applications of speech recognition are found in several domains like healthcare, military, commercial/industrial applications, telephony, personal computers, and many other devices. Figure 2 shows the speech recognition system structure.

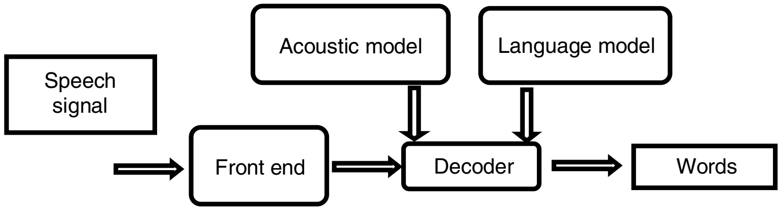


Fig. 2 Speech recognition system

3.2 MFCC Features

The mel-frequency cepstral coefficients are used in stressful speech and speaker recognition fields. These coefficients provide a high-level approximation of a human auditory perception, and they play a mainly role in the voice identification. The MFCC process details are explained in [25].

3.3 Hidden Markov Model

The Hidden Markov Model (HMM) [26] is a popular method in machine learning and statistics for modeling sequences like speech. This model is a finite ensemble of states, where each set is associated with a probability distribution. Transitions among the states are governed by a set of probabilities called transition probabilities. Markov models are excellent ways of abstracting simple concepts into a relatively easily computable form. It is used in data compression to sound recognition. The Markov Model makes the speech recognition systems more intelligent.

4 Amazigh Language

Before the implementation of a speech recognition voice response server system for any language, it is necessary to have a preliminary study of this language. In our case, we choose Amazigh which is a less-resourced Moroccan language. In the best of our knowledge, this is the first IVR using this language. The Amazigh language is widely spoken in a vast geographical area of North Africa. It is spoken by 28% of the Moroccan population. Our work is based on the 33 alphabets of Amazigh language which are consecrated by the foundation of the Royal Institute of Amazigh Culture (IRCAM).

5 Alphabets Speech Recognition

In this section, we describe our experience to create and develop a telephony Amazigh voice recognition system-based alphabets using Asterisk server and CMU Sphinx tools. Our experiments, both training and recognizing, were based on CMU Sphinx system, which is HMM-based, speaker-independent and speaker-independent, and continuous recognition system.

The system is created using Oracle virtual box tool on the host machine with 2 GB of RAM and an Intel Core i3 CPU of 1.2 GHz speed. The operating system used in our experiment was Ubuntu 14.04 LTS.

5.1 Speech Preparation

In this section, we describe our speech recognition database. The corpus consists of 33 spoken Amazigh letters collected from 30 Amazigh Moroccan native speakers aged between 16 and 50 years old. The audio data are recorded in wave format by using the recording tool WaveSurfer [27]. Each alphabet is pronounced 10 times. In our work, the database is partitioned to training 70% and testing 30% in order to ensure the speaker-independent aspect. For the speaker-dependent aspect, we use the same speakers in both phases training and testing.

5.2 Training Phase

In order to determine their optimal values for maximum performance, different acoustic models are prepared by varying HMMs (3–5) and GMMs (8–16–32–64). The wave recorded audio data is used in the training phase where the database is partitioned to 70% training and 30% testing in order to ensure the speaker independent aspect. In our work, we used a grammar file that includes the 33 isolated Amazigh alphabets.

5.3 Telephony Recognizing Phase

Our idea is to acquire and process audio signals from the transferred audio where the system was tested by coding audio data-based G711 VoIP audio codec. The prepared system includes two major threads: a coding–decoding audio stream and the alphabets speech recognition process.

In the first step, the audio is transferred via IVR service. In the next step, the audio signal is split into frames, and the MFCCs are calculated for each of them.

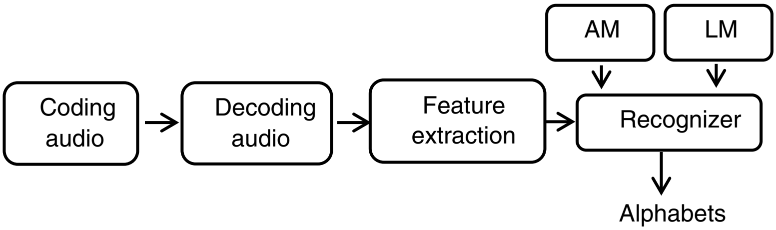


Fig. 3 The scheme of the telephony alphabets Amazigh speech recognition

Table 1 The training and testing data

Parameters	Values	Parameters	Values
Sampling rate	8–16 kHz	Token number	15,840
Number of bits	16 bits	HMMs	3–5
Audio format	WAV	GMMs	8–16–32–64
Number of speakers—Training	21	Number of g711 coded samples	2970
Number of speakers—Test	9	Number of GSM coded samples	2970

These MFCCs are expressed with GMMs parameters and compared with the stored database. The recognition rate for each alphabet was observed and recorded for each experiment. Figure 3 presents the main telephony ASR system components.

In this work, The CMU-Cambridge Statistical Language Modeling toolkit is used to generate language model of our system. Sphinx was utilized for the recognizer implementation because it is free and has been applied by many researchers all over the world.

6 Experimental Results

In order to evaluate the performances of the telephone ASR system, we train our system by the uncoded voice and we perform three recognition experiments (Exper1, Exper2, and Exper3). These experiments focused on testing the system using different subsets of the Amazigh alphabets corpus. In the first experiment, the system was tested by using the uncoded voice; the second experiment was working on the Amazigh alphabets voice decoded by the G711 audio codec; and in the third, the speech was decoded by using the GSM codec. More technical details about our system are shown in Table 1. The recognition rate for each alphabet was observed and recorded for each experiment. Our systems were trained using different HMMs and GMMs. The numbers of HMMs were 3 and 5, and Gaussian mixtures per model were 8, 16, 32, and 64. Table 2 presents the overall system recognition rate for different parameters for speaker-independent and speaker-dependent systems.

In general, the best rates are obtained when the system is tested by speaker-dependent speech using uncoded speech.

Table 2 System overall recognition rate

3 HMMs								
	8 GMMs		16 GMMs		32 GMMs		64 GMMs	
Audio codec	Diff speakers	Same speaker	Diff speakers	Same speaker	Diff speakers	Same speaker	Diff speakers	Same speaker
Exper1	88.99	97.00	87.91	96.68	86.84	94.44	86.63	93.88
Exper2	82.96	95.77	85.76	96.55	80.54	93.11	79.76	92.22
Exper3	79.39	94.44	82.19	96.22	76.97	92.22	76.20	90.78
5 HMMs								
	8 GMMs		16 GMMs		32 GMMs		64 GMMs	
Audio codec	Diff speakers	Same speaker	Diff speakers	Same speaker	Diff speakers	Same speaker	Diff speakers	Same speaker
Exper1	87.37	96.66	87.00	94.77	86.80	91.11	86.06	90.78
Exper2	80.67	94.44	83.47	93.11	78.25	90.44	77.47	90.33
Exper3	78.28	92.22	81.08	91.66	75.86	90.00	75.08	90.00

In the case of Exper1, the speaker-dependent system reaches the highest rate of 97.00% with 3 HMMs and 8 GMMs in comparison to the speaker-independent system rate which is lower by 8.01% in the same parametrization.

Concerning the second experience, the recognition rates have witnessed a decrease where the best rate for speaker-dependent amounting is 96.55% while the speaker-independent dropped at 3.23%. The lower recognition rates are obtained in the second experiment compared to the first experiment. That may be due to speech coding which degrades the voice quality.

For the Exper3, the speaker-independent system recognition rates were 79.39, 82.19, 76.97, and 76.20% for 8, 16, 32, and 64 GMMs, respectively. For 5 HMMs, the system performances were 78.28, 81.08, 75.86, and 75.08% for 8, 16, 32, and 64 GMMs, respectively. For speaker-dependent system recognition rates, the best rate is 96.22 with a difference of 0.78% with the first experiment.

Also, it is noted that in the first experiment, the best and higher accuracy was found with 3 HMMs and 8 GMMs. In the two last experiments, the system performance was better for 3 HMMs and 16 GMMs but lower for 5 HMMs and 64 GMMs.

By comparing the obtained results, we found that the voice coding has an effect on the recognition rates where a vast difference is observed between the rates achieved by using uncoded speech and decoded speech. Probably the degradation of recognition performance is due to the impact of data compression, transmission errors, or bandwidth. Based on the voice codec results, we can say that the G.711 audio codec performs better than GSM codec.

7 Conclusions

In this chapter, the Amazigh ASR system via the interactive voice response service was investigated using the sounds database corresponding to the Moroccan Amazigh language. The designed system was implemented based on the IVR system and ASR system. The results in this chapter compare the recognition performance of two audio codecs G711 and GSM based on speaker-independent and speaker-dependent approaches. The main aim of this chapter is studying the impact of data coding on the Amazigh alphabets speech recognition performance. Our findings show that the best performances for both speaker-independent and speaker-dependent are 85.76% and 96.55%, which were found by using the G711 codec with 3 HMMs and 16GMMs.

References

1. Bhat, C., Mithun, B. S., Saxena, V., Kulkarni, V., & Kopparapu, S. (2013, August). Deploying usable speech enabled ivr systems for mass use. In *Human computer interactions (ICHCI), 2013 international conference* (pp. 1–5). IEEE.
2. Lieberman, G., & Naylor, M. R. (2012). Interactive voice response technology for symptom monitoring and as an adjunct to the treatment of chronic pain. *Translational Behavioral Medicine, 2*(1), 93–101.
3. Sinadinovic, K., Wennberg, P., & Berman, A. H. (2011). Population screening of risky alcohol and drug use via internet and interactive voice response (IVR): A feasibility and psychometric study in a random sample. *Drug and Alcohol Dependence, 114*(1), 55–60.
4. Carlini, B. H., McDaniel, A. M., Weaver, M. T., Kauffman, R. M., Cerutti, B., Stratton, R. M., & Zbikowski, S. M. (2012). Reaching out, inviting back: Using interactive voice response (IVR) technology to recycle relapsed smokers back to Quitline treatment—A randomized controlled trial. *BMC Public Health, 12*(1), 507.
5. Rose, G. L., MacLean, C. D., Skelly, J., Badger, G. J., Ferraro, T. A., & Helzer, J. E. (2010). Interactive voice response technology can deliver alcohol screening and brief intervention in primary care. *Journal of General Internal Medicine, 25*(4), 340–344.
6. Aust, H., Oerder, M., Seide, F., & Steinbiss, V. (1995). The Philips automatic train timetable information system. *Speech Communication, 17*(3), 249–262.
7. Gao, D., Xiao, X., Zhu, G., et al. (2008). Classification of speech transmission channels: Landline, GSM and VoIP networks. In *Signal processing, 2008. ICSP 2008. 9th International conference on* (pp. 671–675). IEEE.
8. Salonidis, T., & Digalakis, V. (1998). Robust speech recognition for multiple topological scenarios of the GSM mobile phone system. In *Acoustics, speech and signal processing, 1998. Proceedings of the 1998 IEEE International conference* (Vol. 1, pp. 101–104). IEEE.
9. Kim, H. K., & Cox, R. V. (2001). A bitstream-based front-end for wireless speech recognition on IS-136 communications system. *IEEE Transactions on Speech and Audio Processing, 9*(5), 558–568.
10. Basu, J., Bepari, M. S., Roy, R., & Khan, S. (2013). Real time challenges to handle the telephonic speech recognition system. In *Proceedings of the fourth international conference on signal and image processing 2012 (ICSIP 2012)* (pp. 395–408). Springer.
11. Satori, H., & ElHaoussi, F. (2014). Investigation Amazigh speech recognition using CMU tools. *International Journal of Speech Technology, 17*(3), 235–243.
12. Hamidi, M., Satori, H., & Satori, K. (2016). Amazigh digits speech recognition on IVR server. *Advances in Information Technology: Theory and Application, 1*(1).
13. Hamidi, M., Satori, H., & Satori, K. (2016). Implementing a voice interface in VOIP network with IVR server using Amazigh digits. *The International Journal of Multi-Disciplinary Sciences, 2*(2), 38–43.
14. Satori, H., Zealouk, O., Satori, K., & ElHaoussi, F. (2017). Voice comparison between smokers and non-smokers using HMM speech recognition system. *International Journal of Speech Technology, 20*(4), 771–777.
15. Zealouk, O., Satori, H., Hamidi, M., Laaidi, N., & Satori, K. (2018). Vocal parameters analysis of smoker using Amazigh language. *International Journal of Speech Technology, 21*(1), 85–91.
16. Hamidi, M., Satori, H., Zealouk, O., Satori, K., & Laaidi, N. (2018). Interactive voice response server voice network administration using Hidden Markov Model Speech Recognition System. In *2018 Second world conference on smart trends in systems, security and sustainability (WorldS4)* (pp. 16–21). IEEE.
17. Hamidi, M., Satori, H., Zealouk, O., & Satori, K. (2019). Amazigh digits through interactive speech recognition system in noisy environment. *International Journal of Speech Technology, 23*, 101–109.

18. Spencer, M., Allison, M., & Rhodes, C. (2003). *The asterisk handbook*. Asterisk Documentation Team.
19. Madsen, L., Van Meggelen, J., & Bryant, R. (2011). *Asterisk: The definitive guide*. O'Reilly Media.
20. Handley, M., Schulzrinne, H., & Schooler, E., et al. (1999). *SIP: Session initiation protocol*.
21. Schulzrinne, H. (1996). *RTP: A transport protocol for real-time applications*. Internet Engineering Task Force, Audio-Video Transport Working Group. RFC 1889.
22. Karapantazis, S., & Pavlidou, F. N. (2009). VoIP: A comprehensive survey on a promising technology. *Computer Networks, 53*(12), 2050–2090.
23. Sharma, S. R. (2009). U.S. patent no. 7,502,737. U.S. Patent and Trademark Office.
24. Huang, X., Acero, A., Hon, H. W., & Reddy, R. (2001). *Spoken language processing: A guide to theory, algorithm, and system development* (Vol. 95). Prentice Hall PTR.
25. Muda, L., Begam, M., & Elamvazuthi, I. (2010). Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. In *arXiv preprint arXiv:1003.4083*.
26. Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257–286.
27. Wavesurfer. (2016). https://sourceforge.net/projects/wavesurfer/

Automatic Speech Recognition for Moroccan Dialects: A Review

Ouissam Zealouk, Hassan Satori, Hamidi Mohamed, and Khalid Satori

Abstract Speech is the naturally and effectively a form of communication among human beings; it is one of the most important fields of signal processing. Speech can be determined and converted to a machine-readable form via a technology named speech recognition. This chapter gives a technical overview of the speech recognition systems and their recent progress pertaining to feature extraction techniques, speech classifiers, and performance evaluation. The objective of this survey is to describe some used methods of the speech recognition system based on Moroccan dialects.

Keywords Automatic speech recognition system · Feature extraction techniques · Moroccan dialects · HMM · LPC

1 Introduction

The speech is the most available way of communication among humans being and also the most natural and efficient form of exchanging information between humans. This advantage has motivated several researchers to focus their interest in the speech systems design; these systems are known as automatic speech recognition systems (ASR) which allow the system to identify words and sentences from spoken speech and convert them into a text. This technology has a vast field of applications, e.g., command recognition, dictation, interactive voice response, learning foreign languages, and helping disabled people to interact with society. It is a technology which makes life easier and very promising [1–3]. ASR solution is rich in many languages like English, Arabic, Indian, and German [4, 5]. However, ASR and speech analysis for Moroccan dialects is a less researched area. Some efforts to develop ASR for Moroccan dialects in a noise-free environment have been reported

O. Zealouk (✉) · H. Satori · H. Mohamed · K. Satori
Computer Science, Signals, Automation and Cognitivism Laboratory, Sidi Mohammed Ben Abbdallah University, Fés, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_60

in the literature [6–9]. For many years, human speech interaction with computers was studied by researchers, they focused on the best ways to improve the technical speech system and almost all of the testing has centered on accuracy gains. Major developments are being observed regularly on both the technology and ASR system exploration [10].

In this chapter, a comprehensive survey on dialect Moroccan speech recognition is provided. The fundamentals and classification of the speech recognition systems have also been discussed. The review of different approaches of speech recognition system has been presented. The rest of this chapter is organized as follows. Sections 2 and 3 give related work and an overview of voice background. In Sect. 4, a brief description of the feature extraction is presented. Section 5 emphasizes the description of speech recognition classification. Section 6 shows the framework. The performance measuring parameters are presented in Sect. 7. Section 8 describes previous works. Conclusions are drawn in Sect. 9.

2 Related Works

This section offers some related works available in the literature which are similar to the presented chapter. Some of the works provide ASR system for others languages.

In [11], the authors have been created a Hindi isolated word speaker-dependent. Linear predictive cepstral coefficients (LPCCs) were used for feature extraction and recognition was based on the Hidden Markov model (HMM), and the database includes two speakers focused on Hindi digits. B. Al-Qatab et al. [12] have developed an Arabic automatic speech recognition tools using HTK. The system recognized both continuous speech and isolated words. The designed system used an Arabic dictionary built manually by the speech sounds of 13 speakers and consist 33 words. Silva et al. [13] have investigated the speech recognition system for Portuguese digit using line spectral frequencies (LSF). The results show that the LFS provide best results in compared to those obtained by using Mel-frequency cepstrum coefficient (MFCC).

3 Voice Background

3.1 Speech Types

Speech recognition system can be divided into several classes by defining what kind of utterances they can recognize [14, 15].

- *Isolated Word:*

  Isolated word recognition usually requires each utterance having calm on both sides of sample windows. It accepts a single word at a time. Often, these forms of speech have “Listen/Not-Listen cases.” Isolated speech describes best this dichotomy.

- *Connected Words:*
  Connected word system almost resembles isolated words but requires a minimum pause between utterances to form a clear flow of speech.
- *Continuous Speech:*
  Continuous speech recognizers permit the user to talk relatively in a natural fashion, while the computer defines the content. This kind of recognizers with continuous speech makes machine understanding much more difficult because they use special methods and a unique sound to determine utterance boundaries.
- *Spontaneous Speech:*
  At a basic stage, it can be thought of as speech that is natural sounding and not experimented out before. An ASR system with spontaneous speech ability must be capable to handle different words and diversity of natural speech feature such as words being run at one time.

3.2 Speaker's Modeling

Each speaker has specific voice features, due to his personality, gender, age, etc. In general, the speech recognition system is divided into two major axes based on speaker models, namely speaker-dependent and speaker-independent [16].

- *Speaker-Independent:*
  Speaker-independent systems are destined to a variety of speakers. These designed systems allow recognizing the speech patterns of a wide people set. In fact, these systems are more flexible, but they are the most complex and costly to develop. Also, they offer less performance compared to speaker-dependent systems.
- *Speaker-Dependent:*
  Speaker-dependent systems are constructed for a specific speaker. These systems are usually simpler to improve, economical, and more performing, but they are less flexible than speaker-independent systems. In general, these systems are more performant than speaker-independent ones.

4 Feature Extraction Technique

Feature extraction is the essential part of speech recognition since it is a principal factor to distinguish one speech from another [17]. The utterance can be extracted by a wide range of feature extraction techniques suggested and successfully used for speech recognition process. But the extracting of features should meet some criteria while dealing with the speech signal such as being:

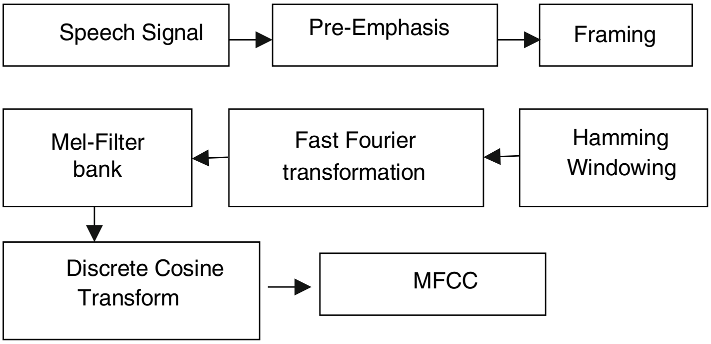


Fig. 1 Block diagram of MFCC feature extraction techniques

- Easy to calculate extracted speech features.
- Not be amenable to mimicry.
- Showing less change from one speaking environment to another.
- Steady over time.
- Occurring ordinarily and naturally in speech.

Different methods for feature extraction are MFCC, LPC, cepstral analysis, PCA, etc.

4.1 Mel-Frequency Cepstral Coefficient (MFCC)

MFCC is the most explicit and utilized speech recognition feature extraction technique. This method simulates the human system response more sincerely than any other system [18]. Also, in speaker recognition, the most used features are Cepstral coefficients due to the capability to contend with convolution channel distortion, the good vocal tract representation, and the robustness against noise. The main MFCC extraction components are preemphasis, framing, hamming windowing, fast Fourier transformation, Mel-filter bank, and discrete cosine transformation (*see* Fig. 1).

4.2 Linear Prediction Cepstral Coefficient (LPCC)

The feature extraction aim is to prove the speech signal by the finite measures of the signal numbers. Linear predictive coding is utilized to employ the LPCCs from the marked vocal words. The LPCC is then translated to cepstral coefficients which are regulated between 1 and −1 [19]. LPCC was performed based on the autocorrelation method. It is very sensitive to quantization; noise is the main drawback of it.

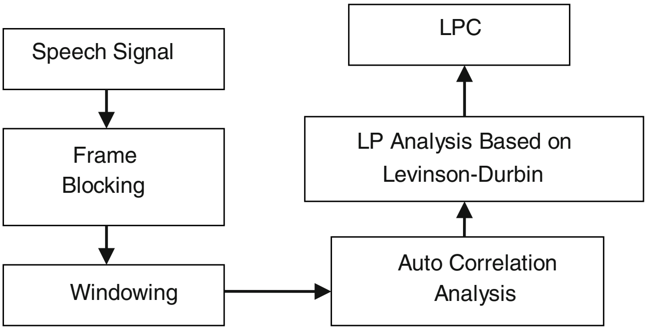


Fig. 2 LPC feature extraction process

4.3 Linear Prediction Coding (LPC)

LPC is one of the most robust signal analysis ways for linear prediction. It is a prevalent technique for determining the speech fundamental parameters and produces precise speech parameters assessment in addition to speech computational architecture. Speech sampling can be approximated as a linear combination of past speech pattern which is the main technique governing the LPC method. Following, Fig. 2 shows the steps involved in LPC feature extraction.

5 Speech Recognition Technique Classification

5.1 Acoustic-Phonetic Approach

Acoustic-phonetic approach for speech recognition is focused on finding the voice and providing adequate labels for these voices. The principle of the acoustic-phonetic method is based on the fact that there are limited and exclusive phonemes in spoken language and these are largely defined by a set of acoustic characteristics that are showed in the vocal signal over time [20]. With speaker and coarticulation impact, the phonetic units' acoustic properties change frequently; it is supposed, in this process, that the criteria controlling the instability are straightforward and can be easily learned by the machine. Steps included in acoustic phonetic approach are as follows; The first phase is the spectral analysis of speech which presents the wide acoustic properties of several phonetic units. The second phase is segmenting and labeling the speech that results in a phoneme lattice speech characterization. The third phase is determining the string of words or a valid word from phonetic label sequences transported out by the segmentation to labeling.

5.2 Pattern Recognition Approach

Two fundamental steps involved in pattern recognition method are speech pattern training and pattern comparison [2]. Where the mathematical framework and the start proportionate speech model representation are exploited for an effective pattern comparison, for a combination of labeled training examples over formal training algorithm is a primary feature of this approach. Within it we can distinguish between template-based approach and stochastic approach. In the template approach, the speech is compared with an ensemble of prerecorded words (templates) in order to find the best correlation. This has the advantage of using appropriate word models. Recognition is realized by matching an unknown spoken word with each of these reference templates and selecting the class of the better matching model. Usually, templates for entire words are built. Stochastic model is a more appropriate approach to speech recognition as it bases probabilistic models to deal with imperfect data. There are many methods in this approach like HMM, SVM, DTW, and VQ. The most popular stochastic approach today is Hidden Markov Model [21].

5.3 Artificial Intelligence Approach

The artificial intelligence approach is a combination of the pattern recognition process and acoustic-phonetic approach. This approach tries to mechanize the recognition procedure that is based on the following method: a person applies his intelligence in envisioning, analyzing, and finally making a decision on the measured acoustic features [22]. Expert system is used widely in this approach [23]. Acoustic-phonetic knowledge is utilized for speech classification rules developing where template-based technique provides less accuracy about human speech processing, but these techniques have been very productive in the different design of speech recognition system. This approach is not much successful as complexness in quantifying skillful knowledge. Artificial neural network (ANN) method is the most effective way used by the artificial intelligence approach. ANN is a method that consists of a big number of simple treatment elements called units or neurons that affect each other's performance via a network of excitatory or repressive weighting [24].

6 Framework

Given the importance of ASR, several free software have been developed, among the most famous are: CMU Sphinx and HTK.

6.1 CMU Sphinx4

The Sphinx4 is a system of speech recognition designed by Carnegie Mellon University, Sun Microsystems Laboratories, Mitsubishi Electric Research Labs, and Hewlett-Packard's Cambridge Research Lab. It is programmed totally in the Java TM language. Sphinx4 uses newer search strategies and is used internationally in its acceptance of several types of grammars and language models, kinds of acoustic models, and feature streams [25].

6.2 HTK

Hidden Markov Model toolkit (HTK) is designed by the Cambridge University Engineering Department (CUED). HTK is a set of library modules and tools available in language C. It is used in the speech recognition studies and many other applications as speech synthesis [26]. This framework provides system tools for both training and testing phase. Initially, HTK training tools are utilized to train HMMs using training speech from a stored database. Characteristics are extracted from these training sounds, and then, these features are used for modeling the system. Finally, HTK recognition tools are applied for transcribing the unknown sound. They use the system model generated during the training phase to test the system.

7 Performance Measuring Parameters

The performance of speech recognition system is mostly determined in terms of accuracy and speed. Accuracy may be calculated in terms of performance accuracy which is often rated with word error rate (WER) [27].

7.1 Word Error Rate (WER)

Word error rate is a prevailing measure of a speech recognition performance or translation device. The determining performance difficulty lies in the fact that the recognized word sequence can have several lengths from the reference word sequence. The WER is derived from the Levenshtein distance, this last is the number of deletions, insertions, or substitutions required to transform the source string into the target string. Word error rate can then be computed as:

$$\text{WER} = (S + D + I)/N$$

where

- S is the number of substitutions.
- D is the number of the deletions.
- I is the number of the insertions.
- N is the number of words in the reference.

When reporting the performance of a speech recognition system, sometimes word recognition rate (WRR) is used instead:

$$\mathrm{WRR} = 1 - \mathrm{WER} = (N - S - D - I)/N = (H - I)/N$$

where

H is $N - (S + D)$, the number of correctly recognized words.

Table 1 List of the summary of previous studies

Database	Language	Frame work	Feature extraction	Classification techniques	Recognition rate
2500 recorded pronunciation 1 h 40 min duration of training set	Arabic Moroccan dialect [28]	Sphinx4	MFCC	HMM and DTW	HMM 91% DTW 60%
20 speakers, 2000 utterances 1 h 45 min of pronunciation	Amazigh Language [29]	Sphinx4	MFCC	HMM and DTW	HMM 90% DTW 52%
60 speakers, 25,800 tokens 3 h 20 min recording time	Amazigh language [30]	Sphinx4	MFCC and GMM	HMM	92.89%
44 h 19 min duration of training set	Arabic Moroccan dialect [31]	PRLM	MFCC and GMM-UBM	HMM	83.92
8 speakers 400 waveform	Arabic Moroccan dialect (Darija) [32]	MATLAB	PLP and PCA	SVM and FFBPNN	SVM 97.5% FFBPNN 98.3%
36 speakers 3600 recorded pronunciation	Amazigh language [33]	IVR-ASR Astersik-Sphinx4	MFCC and GMM	HMM	90.67%
5935 wave file 8 h 8 min	Arabic language [34]	HTK	MFCC and GMM	HSMM and HMM	96%
34 speakers 11,200 audio files of training	Amazigh language [8]	Sphinx4	MFCC	HMM	90%

8 The Previous Work

In this chapter, there are various studies in Moroccan speech recognition, which summarize, based on database, feature extraction techniques, speech classification technique, and the speech recognition rate. The summary is presented in Table 1 as follows:

9 Conclusion

This review tried to cover the majority of the approaches used in ASR in both phases of the feature extraction and the classification. We focused on a selection of the previous studies for Moroccan dialects and the used techniques in order to determine the effectiveness of the different methods and comparing their performances based on the results that were obtained. The most widely used techniques by researchers for Moroccan dialects are the MFCC and HMM.

References

1. Bahl, L. R., DeGennaro, S. V., Mercer, R. L., et al. (1988). *Speech recognition system*. Patent no 4,718,094,5 Jan. U.S. Patent and Trademark Office.
2. Rabiner, L. R., & Juang, B. H. (1993). *Fundamentals of speech recognition* (Vol. 14). PTR Prentice Hall.
3. Chee, L. S., Ai, O. C., & Yaacob, S. (2009). Overview of automatic stuttering recognition system. In *Proceedings of the International Conference on man-machine systems no. October, Batu Ferringhi, Penang, Malaysia* (pp. 1–6).
4. Alotaibi, Y. A. (2005). Investigating spoken Arabic digits in speech recognition setting. *Information Sciences, 173*(1–3), 115–139.
5. Besacier, L., Barnard, E., Karpov, A., & Schultz, T. (2014). Automatic speech recognition for under-resourced languages: A survey. *Speech Communication, 56*, 85–100.
6. Satori, H., Harti, M., & Chenfour, N. (2007). Introduction to Arabic speech recognition using CMUSphinx system. In *arXiv preprint arXiv:0704.2083*.
7. Satori, H., Zealouk, O., Satori, K., & ElHaoussi, F. (2017). Voice comparison between smokers and non-smokers using HMM speech recognition system. *International Journal of Speech Technology, 20*(4), 771–777.
8. Telmem, M., & Ghanou, Y. (2018). Estimation of the optimal HMM parameters for Amazigh speech recognition system using CMU-Sphinx. *Procedia Computer Science, 127*, 92–101.
9. Zealouk, O., Satori, H., Hamidi, M., Laaidi, N., & Satori, K. (2018). Vocal parameters analysis of smoker using Amazigh language. *International Journal of Speech Technology, 21*(1), 85–91.
10. Anusuya, M. A., & Katti, S. K. (2010). Speech recognition by machine, a review. In *arXiv preprint arXiv:1001.2267*.
11. Pruthi, T., Saksena, S., & Das, P. K. (2000, August). Swaranjali: Isolated word recognition for Hindi language using VQ and HMM. In *International conference on multimedia processing and systems (ICMPS)* (pp. 13–15).

12. Al-Qatab, B. A., & Ainon, R. N. (2010, June). Arabic speech recognition using hidden Markov model toolkit (HTK). In *Information technology (ITSim). International Symposium* (Vol. 2, pp. 557–562). IEEE.
13. Silva, D. F., de Souza, V. M., Batista, G. E., & Giusti, R. (2012, November). Spoken digit recognition in portuguese using line spectral frequencies. In *Ibero-American conference on artificial intelligence* (pp. 241–250). Springer.
14. Das, S. (2012). Speech recognition technique: A review. *International Journal of Engineering Research and Applications, 2*(3), 2071–2087.
15. Prabhakar, O. P., & Sahu, N. K. (2013). A survey on: Voice command recognition technique. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*(5).
16. Saksamudre, S. K., Shrishrimal, P. P., & Deshmukh, R. R. (2015). A review on different approaches for speech recognition system. *International Journal of Computer Applications, 115*(22).
17. Desai, N., Dhameliya, K., & Desai, V. (2013). Feature extraction and classification techniques for speech recognition: A review. *International Journal of Emerging Technology and Advanced Engineering, 3*(12), 367–371.
18. Zulfiqar, A., Muhammad, A., & Martinez-Enriquez, A. M. (2009, November). A speaker identification system using MFCC features with VQ technique. In *2009 Third International Symposium on Intelligent Information Technology Application* (pp. 115–118). IEEE.
19. Dave, N. (2013). Feature extraction methods LPC, PLP and MFCC in speech recognition. *International Journal for Advance Research in Engineering and Technology, 1*(6), 1–4.
20. Narang, S., & Gupta, M. D. (2015). Speech feature extraction techniques: A review. *International Journal of Computer Science and Mobile Computing, 4*(3), 107–114.
21. Zerari, N., Yousfi, B., & Abdelhamid, S. (2016). Automatic speech recognition: A review. *International Academy of Research Journal of Business Technology, 2*(2), 63–68.
22. Gaikwad, S. K., Gawali, B. W., & Yannawar, P. (2010). A review on speech recognition technique. *International Journal of Computer Applications, 10*(3), 16–24.
23. Butzberger, J., Murveit, H., Shriberg, E., & Price, P. (1992, February). Spontaneous speech effects in large vocabulary speech recognition applications. In *Proceedings of the workshop on speech and natural language* (pp. 339–343). Association for Computational Linguistics.
24. Srivastava, N. (2007). Dr. Harsh Dev "Speech Recognition using MFCC and Neural Networks". *International Journal of Modern Engineering Research (IJMER)*.
25. Walker, W., Lamere, P., Kwok, P., Raj, B., Singh, R., Gouvea, E., . & Woelfel, J. Sphinx-4: A flexible open source framework for speech recognition Sun Microsystems (2004).
26. Woodland, P. C., Leggetter, C. J., Odell, J. J., Valtchev, V., & Young, S. J. (1995, May). The 1994 HTK large vocabulary speech recognition system. In *Acoustics, Speech, and Signal Processing, International Conference on* (Vol. 1, pp. 73–76). IEEE.
27. Kalamani, M., Valarmathy, S., Poonkuzhali, C., & Catherine, J. N. (2014, January). Feature selection algorithms for automatic speech recognition. In *Computer Communication and Informatics (ICCCI), International Conference on* (pp. 1–7). IEEE.
28. El Ghazi, A., Daoui, C., Idrissi, N., Fakir, M., & Bouikhalene, B. (2011). Speech recognition system based on Hidden Markov Model concerning the Moroccan dialect DARIJA. *Global Journal of Computer Science and Technology, 11*, 975–981.
29. El Ghazi, A., Daoui, C., Idrissi, N., Fakir, M., & Bouikhalene, B. (2011). Système de reconnaissance automatique de la parole Amazigh à base de la transcription en alphabet Tifinagh. *Revue Méditerranéenne des Télécommunications, 1*(2), 23–29.
30. Satori, H., & ElHaoussi, F. (2014). Investigation Amazigh speech recognition using CMU tools. *International Journal of Speech Technology, 17*(3), 235–243.
31. Lachachi, N. E., & Abdelkader, A. D. L. A. (2016). Identification Automatique des Dialectes du Maghreb. *Revue Maghrébine des Langues, 10*(1), 85–101.
32. Hassine, M., Boussaid, L., & Messaoud, H. (2016). Maghrebian dialect recognition based on support vector machines and neural network classifiers. *International Journal of Speech Technology, 19*(4), 687–695.

33. Hamidi, M., Satori, H., & Satori, K. (2016). Implementing a voice interface in VOIP network with IVR server using Amazigh digits. *The International Journal of Multidisciplinary Sciences, 2*(2), 38–43.
34. Khelifa, M. O., Elhadj, Y. M., Abdellah, Y., & Belkasmi, M. (2017). Constructing accurate and robust HMM/GMM models for an Arabic speech recognition system. *International Journal of Speech Technology, 20*(4), 937–949.

Modeling the Mean Stress Effect on the Fatigue Life of Welded Profiles Under Damped Loads for Automotive Applications

Imane Amarir, Hamid Mounir, and Abdellatif E. L. Marjani

Abstract Different parameters can contribute nowadays to the fatigue strength of welded structure. But generally, the cyclic loading history has the most important role in causing failures. For example, the case of stress cycle corresponding to zero-based and fully reversed loading. In the present chapter, a three dimensional of finite element simulation using the fatigue analysis and stress-life approach was investigated on a model based on welded rectangular profiles under damped loads for automotive utilization in order to show the effect of mean stress on fatigue life with different theories which are available in stress-life module. The Gerber, Goodman, and Soderberg relationships have been discussed. The results indicated that as the stress ratio is less or more than 1, the fatigue life shows differences along with correction theories. Whereas for stress ratio equal to −1, the fatigue life of our structure keeps same value whatever the correction theory because no mean stress exists in this case. Then, it is observed that Soderberg is the appropriate theory for our model that can give a significant result in term of durability.

Keywords Fatigue · Welded structure · Finite element simulation · Correction theory · Mean stress

1 Introduction

The welding assemblies are widely used in different industries, especially in transportation fields [1]. However, they are able to fail in any time if specified conditions and requirements have not been respected [2]. The most important factors are repeated cyclic loadings [3], weld process quality, concentration stress [4], size component, surface treatments, etc. Indeed, at the end of the nineteenth century, the analysis of fatigue was started by many authors and among them is Wohler [5],

I. Amarir (✉) · H. Mounir · A. E. L. Marjani
Research Team EMISys, Research Centre Engineering 3S, Mohammadia School of Engineers, Mohammed V University, Rabat, Morocco
e-mail: imaneamarir@research.emi.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_61

who described a curve called SN curve characterizing the alternating stress against numbers of cycles to failure. Besides, different theories applicable for tensile mean stress values have been proposed by designers [6] like Morrow [7], Gerber [8], Goodman [9], and Soderberg [10] relations.

Recently, some new approaches have been proposed for predicting fatigue life behavior such as the nominal stress approach [11], the hot spot stress approach, [12] and the notch stress method [13].

In this present work, a three-dimensional simulation using fatigue analysis and stress-life approach was developed on welded rectangular profiles under damped loads for automotive utilization [14] in order to estimate the fatigue life by post-processing finite element method results. Indeed, we have studied the same structure before in [15, 16] with other different simulation results. Moreover, in this chapter, we have demonstrated the vital role of mean stress in fatigue life estimation which we have evaluated with different values of stress ratio as a function of available number of cycles along with correction theories [17].

2 Numerical Modeling

A welded structure used frequently for automotive fields was developed with FEM software [18] in order to estimate the fatigue life using the stress-life method and mean stress correction.

2.1 Geometry

Fatigue analysis was built on a geometry based on four rectangular profiles welded together and connected to a longitudinal spring (body-ground), as shown in Fig. 1 [19].

2.2 Material Properties

Table 1 displayed the properties of structural steel as a material applied on our model:

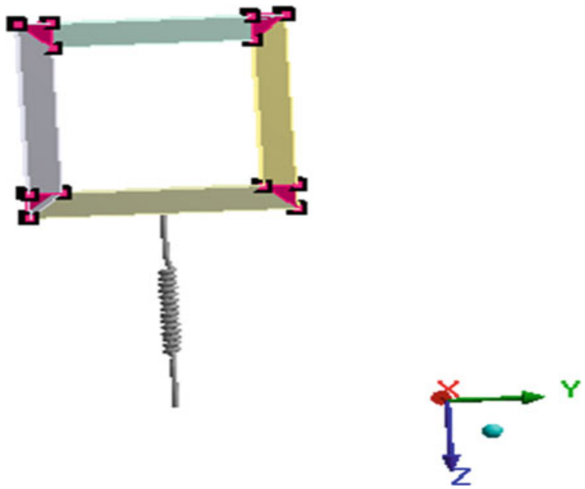


Fig. 1 Geometry of welded rectangular profiles

Table 1 Steel-applied properties

Type of material	Tensile yield strength	Tensile ultimate strength	Poisson's ratio	Mass density
Structural steel	250 MPa	460 MPa	0.3	7850 kg/m^3

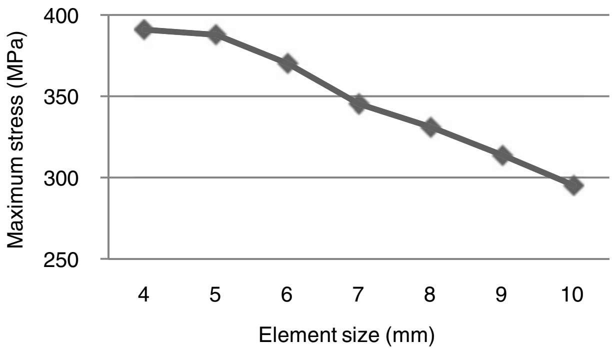


Fig. 2 Choice of the best mesh density

2.3 Loading and Boundary Conditions

After we have simplified the calculations using the meshing in 3D, the finest element size chosen is 5 mm because of the results convergence in terms of stresses as defined in Fig. 2.

We have then applied a load of 1500 N on the top of our structure and a displacement blocked along the *X* and *Y* axis (it is well shown in Fig. 3).

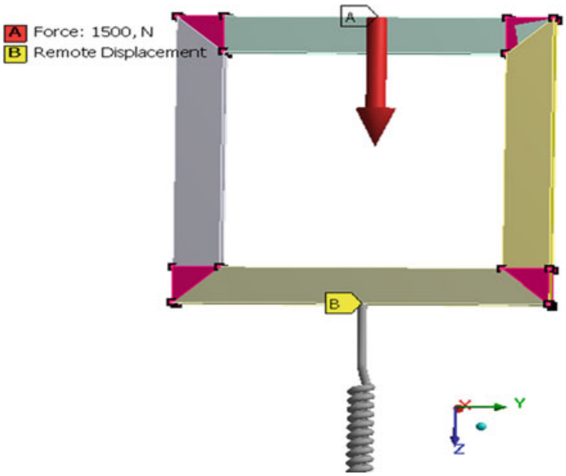


Fig. 3 Loading and boundary conditions applied on the model

3 Fatigue Analysis Properties

The repeated cyclic loading is one of the most factors that can affect the fatigue strength. The following Fig. 4 is used to determine a stress cycle as function of alternating and mean stress.

The mean stress σ_m is defined as the average of maximum and minimum stress during a load cycle as in Eq. (1).

$$\sigma_{m} = \frac{\sigma_{max} + \sigma_{min}}{2} \quad (1)$$

The alternating stress σ_a is expressed as one-half of stress range $\Delta\sigma$, as in Eq. (2).

$$\sigma_{a} = \frac{\sigma_{max} + \sigma_{min}}{2} \quad (2)$$

The stress ratio R led to measure the mean stress as in Eq. (3).

For example, $R = -1$ and $\sigma_m = 0$ is referred to as loading is fully reversed, and for $R = 1$ is corresponding to static loading, then $R = 0$ occurs where $\sigma_m = \sigma_a$ and the mean stress σ_m is tensile.

$$R = \frac{\sigma_{min}}{\sigma_{max}} \quad (3)$$

The fatigue analysis was built using the stress-life approach which is referred as an S-N approach. This method is based above all on high cycle fatigue application. Moreover, generally four different relations lines are available in the stress-life module. The Goodman [9], Gerber [8], Soderberg [10], and Morrow [7] lines

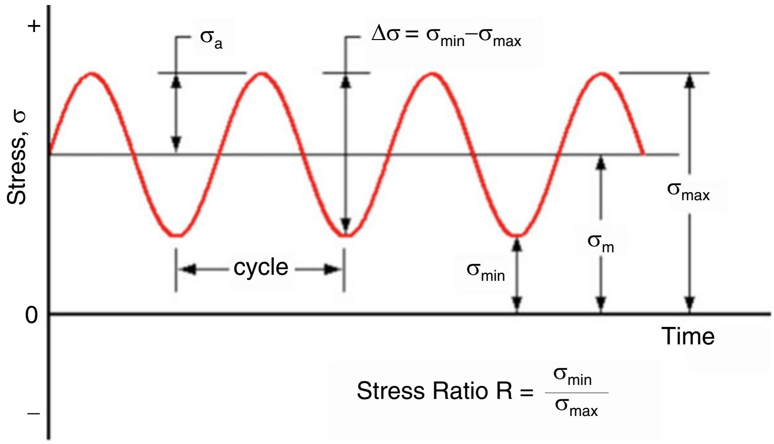


Fig. 4 Stress cycle [20]

expressed the relations between mean stress (σ_m), endurance limit (Se), tensile yield stress (Sy), alternating stress (σ_a), and ultimate stress (Su), as shown respectively in Eqs. (4), (5), (6), and (7).

$$\text{Gerber line}: \quad \frac{\sigma_a}{S_e} + \left(\frac{\sigma_m}{\mathrm{Su}}\right)^2 = 1 \tag{4}$$

$$\text{Goodman line}: \quad \frac{\sigma_a}{S_e} + \left(\frac{\sigma_m}{\mathrm{Su}}\right) = 1 \tag{5}$$

$$\text{Soderberg line}: \quad \frac{\sigma_a}{S_e} + \left(\frac{\sigma_m}{\mathrm{Sy}}\right) = 1 \tag{6}$$

$$\text{Morrow line}: \quad \frac{\sigma_a}{S_e} + \left(\frac{\sigma_m}{\sigma_f}\right) = 1 \tag{7}$$

where σ_f is called true fracture stress.

A schematic diagram of comparison of these equations is presented in the following Fig. 5.

Consequently, Goodman and Gerber are generally considered as the most widely used and accepted correction theories. Goodman relationship deals with simplicity in mathematics and can be used for brittle materials. Soderberg equation is often used for ductile material due to strongly conservative values [22]. And also, Gerber theory is generally used for ductile materials (σ_f > Su). The Morrow and Goodman

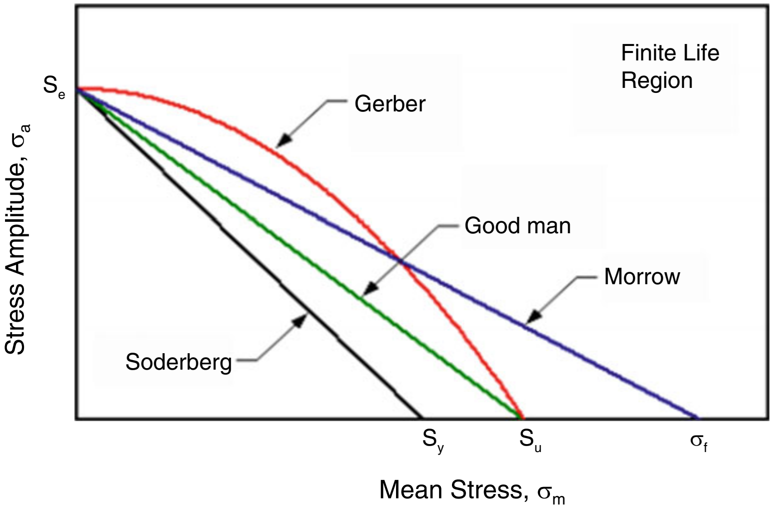


Fig. 5 A graphical comparison of Goodman, Gerber, Soderberg, and Morrow lines [21]

theories are similar for brittle steels. In fact, there is also the multiple SN curves or called mean stress curves that can be occurred if the test results are available [23].

4 Results and Discussion

Knowing that the model is subjected to constant amplitude and proportional loading, different values of stress ratio as function of correction theories and fatigue life have been analyzed. The fully reversed loading ($R = -1$ and $\sigma_m = 0$), then the zero-based loading ($R = 0$ and $\sigma_m = \sigma_{max}/2$), and the case of $R = 0.5$, $R = 0.5$, and $R = -0.5$ are as shown, respectively, in Figs. 6, 7, 8, 9, and 10.

In this part, we have studied the fatigue life with four cases: Goodman, Gerber, Soderberg, and the case of no correction theories (they are well shown in Fig. 11), using a FEM software to show the effect of means stress on the estimated fatigue life.

By a series of finite element analysis method, it is able to provide a summary of the results of number of cycles, Nc, corresponding to fatigue life in the form of graph and table, listing results by stress ratio and mean stress correction theories (*see* Table 2 and Fig. 12).

As you can see from Table 2, the fatigue life (number of cycles) [24] corresponding to none correction theories gives a high and unsatisfactory results which proves the necessity of one of the mean stress correction theories.

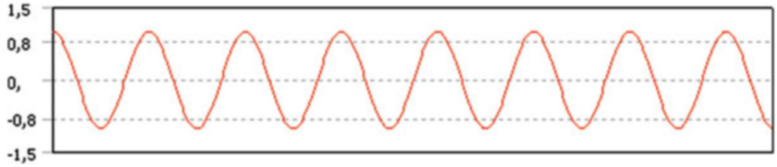


Fig. 6 Fully reversed loading

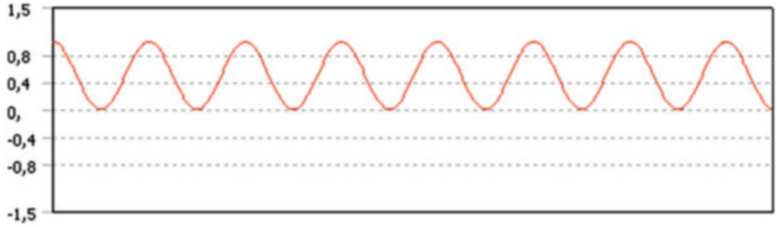


Fig. 7 Zero mean stress

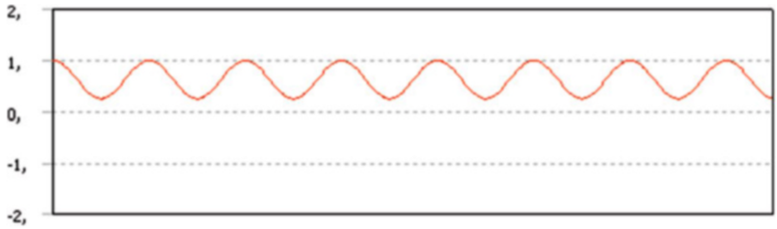


Fig. 8 $R = 0.25$

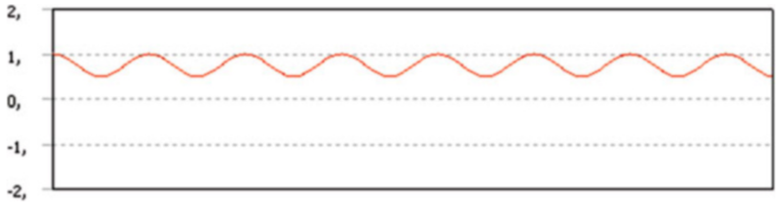


Fig. 9 $R = 0.5$

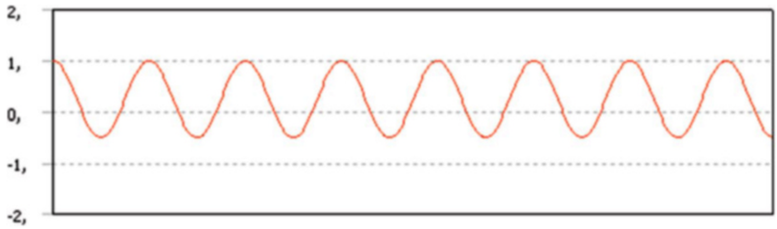


Fig. 10 $R = -0.5$

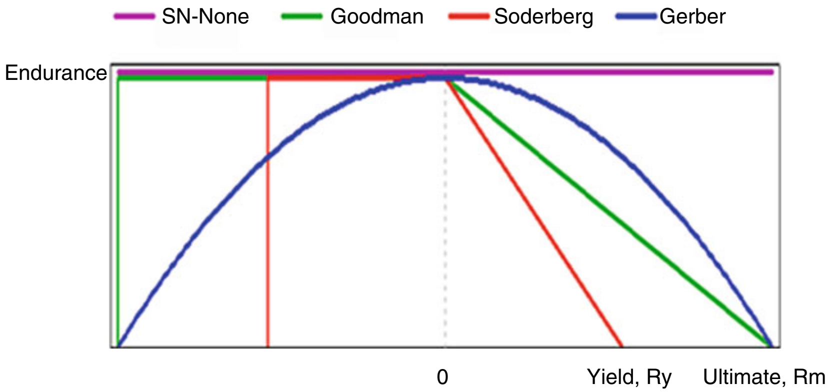


Fig. 11 Mean stress correction theories

Table 2 Fatigue life for different stress ratio

Theories	R	Nc (cycles)
None	−1	2975.8
	−0.5	7240.8
	0	28,720
	0.25	82,513
	0.5	5.0833e+005
Goodman	−1	2975.8
	−0.5	3483.9
	0	4670.1
	0.25	6103.2
	0.5	9808.2
Gerber	−1	2975.8
	−0.5	6292.1
	0	14,347
	0.25	2.50E+04
	0.5	56,138
Soderberg	−1	2975.8
	−0.5	1648
	0	347.87
	0.25	0
	0.5	0

Therefore, we have demonstrated that there is one exceptional case corresponding to means stress equal to zero ($R = -1$), where the fatigue life keeps the same value and no mean stress correction theory is required.

According to Fig. 12, the results of other cases of stress ratio indicated that Soderberg can be the adequate theory for our model which gives a significant and specific values of Nc compared to other theories. Besides, the stress ratio increases with the reduction of number of cycles, in contrast to Goodman and Gerber cases.

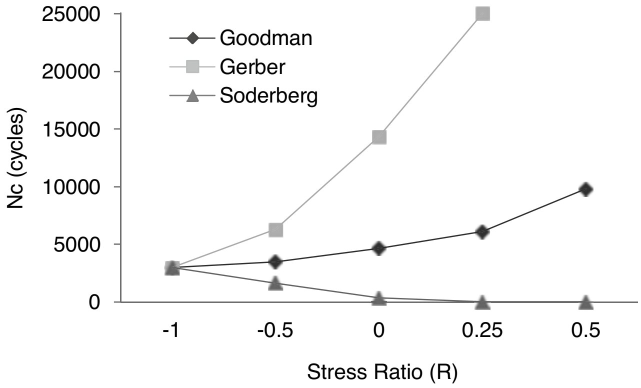


Fig. 12 Effect of mean stress on Nc

Then, it is observed that the fatigue life results showed a difference of nearly 44% between Goodman and Gerber. Moreover, there is a variation of 52% between Goodman and Soderberg. Consequently, the theories show slight differences, for cases where $R << 1$, whereas, there is large differences for cases where $R >> 1$. And obviously, there is no difference in the case of $R = -1$.

5 Conclusion

In this chapter, a welded rectangular profile under damped loads used for automotive domain was investigated with FEM software using fatigue analysis, stress-life method, and constant amplitude load in order to estimate the fatigue life of our structure as a function of mean stress correction theories and stress ratio.

Goodman, Gerber, and Soderberg are three correction theories which each one is appropriate with its condition of application.

The results indicated that for stress ratio equal to -1, there is no impact of correction theory on the fatigue life. Whereas, when stress ratio approaches 1, there is generally a large difference between the theories on the available life behavior of structural steel which is the effect of mean stress on the fatigue life of our structure. Then, Soderberg theory is considered as the best theory adequate with our model results. In future work, numerical results will be validated with experimental data.

References

1. Devarasiddappa, D. (2014). Automotive applications of welding technology. *International Journal of Modern Engineering Research (IJMER), 4*(9), 2240–2245.
2. Yildirim, H. C., & Marquis, G. (2012). Fatigue strength improvement factors for high strength steel welded joints treated by high frequency mechanical impact. *International Journal of Fatigue, 44*, 168–176.
3. Dowling, E. N. (2012). *Mechanical behaviour of materials* (4th ed.). Pearson Education.
4. Kawadkar, D. B., Bhope, D. V., & Khamankar, S. D. (2012). Evaluation of stress concentration in plate with cutout and its experimental verification. *International Journal of Engineering Research and Applications, IJERA, 5*, 566–571.
5. Strzelecki, P., & Sempruch, J. (2016). Verification of analytical models of the S-N curve within limited fatigue life. *Journal of Theoretical and Applied Mechanics, 54*(1), 63.
6. Dowling, E. N. (2004). Mean stress effects in stress life and strain life fatigue. In *2nd SAE Brasil International Conference on Fatigue*.
7. Morrow, J. (1964). *Fatigue properties of metals, Section 3.2 of Fatigue Design Handbook, Pub. No. AE-4*. Society of Automotive Engineers, Warrendale, Steel Technical Committee.
8. Beden, S., Abdullah, M., Kamal Ariffin, S., Al-Asady, N. A., & Mustafizur Rahman, M. (2009). Fatigue life assessment of different steel-based shell materials under variable amplitude loading. *European Journal of Scientific Research, 29*(2), 157–169.
9. Jelaska, D. (2011). On the Goodman's fatigue safety factor. *Advanced Engineering, 6*(1), 1–8.
10. Santhosh, N., Kempaiah, U. N., Sajjan, G., & Gowda, A. C. (2017). Fatigue behaviour of silicon carbide and fly ash dispersion strengthened high performance hybrid al 5083 metal matrix composites. *Journal of Minerals and Materials Characterization and Engineering, 5*(5), 274–287.
11. Al Zamzami, I., & Susmel, L. (2016). On the accuracy of nominal, structural, and local stress based approaches in designing aluminium welded joints against fatigue. *International Journal of Fatigue, 49*, 62–72.
12. Fricke, W. (2004). Fatigue design and structural hot-spot stress determination for welded joints. In *Proceeding of 9th Portuguese Conference on Fracture, Setubal*.
13. Lindqvist, A., & Nilsson, H. (2016). *Effective notch stress analysis of transverse attachments in steel bridges. A parametric fatigue life assessment*. Master's Thesis, Chalmers University of Technology.
14. Amarir, I., El Mounir, H., & Marjani, A. (2019). Modeling the effect of welding process on the durability of welded rectangular profiles under damped loads for automotive utilization. *International Journal of Mechanical and Production Engineering Research and Development (IJMPERD), 1*(9), 639–648.
15. Amarir, I., El Mounir, H., & Marjani, A. (2018). Effect of temperature and thickness of the weld bead on the durability of welded rectangular profiles under damped loads for electrical vehicles utilization. In *6th International Renewable and Sustainable Energy Conference (IRSEC)* (pp. 1–6). IEEE.
16. Amarir, I., El Mounir, H., & Marjani, A. (2018). Numerical modeling of fatigue analysis of welded rectangular profiles under damped loads for automotive utilization. In *4th International Conference on Optimization and Applications (ICOA)*. IEEE.
17. Bader, Q., & Kadum, E. (2014). Mean stress correction effects on the fatigue life behavior of steel alloys by using stress life approach theories. *International Journal of Engineering & Technology, 14*(04), 50–58.
18. Songra, K. (2017). Finite element analysis of composite material using ANSYS. *International Journal for Scientific Research and Development, 4*, 164–166.
19. Amarir, I., El Mounir, H., & Marjani, A. (2019). Modeling the damping effect of welded rectangular profiles under damped loads for automotive utilization on the durability. In *5th International Conference on Optimization and Applications (ICOA)*. IEEE.

20. Afkhami, S., Dabiri, M., Habib Alavi, S., Björk, T., & Salminen, A. (2019). Fatigue characteristics of steels manufactured by selective laser melting. *International Journal of Fatigue, 122*, 72–83.
21. Zhu, S. P., Liu, Q., Zhou, J., & Yu, Z. Y. (2018). Fatigue reliability assessment of turbine discs under multisource uncertainties. *Fatigue & Fracture of Engineering Materials & Structures, 41*(6), 1291–1305.
22. Grolier, P. (2000). A method to determine the safety coefficient in fatigue of compressor valves. In *International Compressor Engineering Conference*.
23. Hyun Jeong, S., Dong-Hoon, C., & Gil Ho, Y. (2015). Fatigue and static failure considerations using a topology optimization method. *Applied Mathematical Modelling, 39*(3–4), 1137–1162.
24. Amarir, I., Mounir, H., & El Marjani, A. (2020). Compilation of fatigue assessment procedures of welded structures for automotive utilization. *Mechanics Based Design of Structures and Machines, 1*(1), 1–13.

DTC Versus Vector Control Strategies for a Grid Connected DFIG-Based Wind Turbine

Khaled Kouider and Abdelkader Bekri

Abstract With the high level of wind power penetration, system executives have an increasing interest in investigating the wind power integration influence on power systems. The doubly fed induction generator (DFIG) is commonly employed in wind power generation systems. In this chapter, we focus on the direct torque and the classic vector control applied to the rotor side converter (RSC) of a grid-connected doubly fed induction generator (DFIG) using a detailed dynamic model under dq reference frame. The two strategies are compared considering many parameters as the rotor currents, the stator power, electromagnetic torque, and rotor flux to ensure the proper operation and to enhance the performance of the DFIG. Both control strategies of our machine are simulated using MATLAB/SIMULINK software package. Finally, the simulations results are displayed and well discussed.

Keywords Direct torque control · Hysteresis · Lock-up table · Vector control · Stator flux orientation

1 Introduction

Nowadays, the electrical power consumption increases proportionally with the human society. It is reported by Enerdata [1]. that the global net electricity consumption has risen from 21,463 TWh in 2016 to 22,015 TWh in 2017 (the power consumption accelerated again in 2017 (+2.3%)). Since the major part of electricity is generated from fossil fuels, the increase in net electricity consumption will result in significant greenhouse gas emissions, which could lead to global warming [2]. For this reason, the renewable energies occupied a crucial position in the last research. The wind power generation is regarded as the most widely used non-hydro renewable energy generation with a renewable and clean high reserve. Besides, it provides almost no greenhouse gas emissions [2]. In the same context,

K. Kouider (✉) · A. Bekri
Department of Electrical Engineering, University of Tahri Mohammed, Bechar, Algeria

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_62

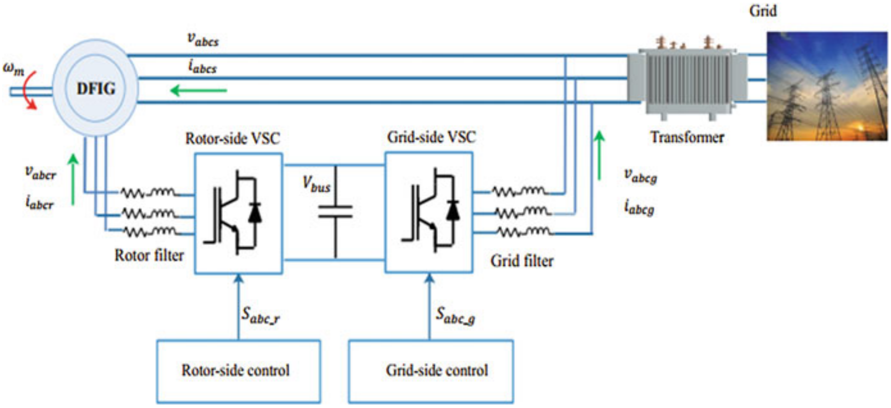


Fig. 1 A standard configuration of a doubly fed induction generator (DFIG) wind power system

and due to many technical and economic factors like the partial scale converting using an economical converter size to handle a fraction (20–30%) of the total power, the doubly fed induction generator (DFIG)-based wind turbine is considered one of the most practical solutions for the wind power conversion. A regular arrangement of a DFIG-WT system is illustrated in Fig. 1. A wound rotor induction generator with slip ring is essential to conduct current toward or out of the rotor winding at slip frequency.

The variable-speed operation is achieved with the help of injecting a controllable rotor voltage [3]. The DFIG could be operated under three modes: subsynchronous operation when $\omega_m < \omega_s$, hypersynchronous operation while $\omega_m > \omega_s$, and in synchronous mode if $\omega_m = \omega_s$, where ω_s and ω_m are the synchronous rotor electrical speeds allowing a bidirectional power flow from or into the back-to-back converter. The grid connection of the DFIG-WT receives a big intention in the last decade. Many works are fulfilled such as the classic vector control as in Dehong et al. [2], Fox et al. [3], Kerrouche et al. [4] and Fei et al. [5] to enhance the performance of the DFIG wind power systems. In the recent years, the direct control techniques have gained a huge importance as in Arnalte et al. [6], and a new improved direct torque control strategy is developed in Li et al. [7]. An ANFIS-based DTC for a DFIG is applied in Kumar et al. [8]. In this chapter, we concentrate on the comparison between the direct torque control and the classic vector control methods for a DFIG wind system. For this purpose, this chapter will be split into three subsections: firstly, the detailed DFIG-WT mathematical model is achieved, then the description of the two control methods laws and equations are established. Finally, the simulation and results are well explained. In this context, this chapter reveals that it is possible to compare the performance of both strategies using various parameters.

2 DFIG in Wind Energy Systems

In this section, the most essential electric and mechanic equations of the DFIG wind system are used to achieve an analytical model proper for the control of the machine. The mathematical modeling of the system is well shown in the two following subsections.

2.1 DFIG Mathematical Modeling

The dynamic model of the DFIG under dq reference frame can be reached using the stator and rotor windings equations as follows [9–11], where ψ_{sd}, ψ_{sq}, ψ_{rd}, ψ_{rq}, i_{sd}, i_E, i_E, i_{rq} are the stator and rotor flux and currents under dq components, respectively, and R_s, R_r, ω_s, ω_r, p are the stator and rotor resistances, speeds, and the number of poles pairs:

$$\begin{aligned} u_{sd} &= R_s i_{sd} + \frac{d\psi_{sd}}{dt} - \omega_s \psi_{sq} \\ u_{sq} &= R_s i_{sq} + \frac{d\psi_{sq}}{dt} + \omega_s \psi_{sd} \\ u_{rd} &= R_r i_{rd} + \frac{d\psi_{rd}}{dt} - \omega_r \psi_{rq} \\ u_{rq} &= R_r i_{rq} + \frac{d\psi_{rq}}{dt} + \omega_r \psi_{rd} \end{aligned} \tag{1}$$

After that the rotor and stator fluxes are given by Eq. (2), where L_s, L_r, L_m are the stator, rotor, and mutual inductances:

$$\begin{aligned} \psi_{sd} &= L_s i_{sd} + L_m i_{rd} \\ \psi_{sq} &= L_s i_{sq} + L_m i_{rq} \\ \psi_{rd} &= L_r i_{rd} + L_m i_{sd} \\ \psi_{rq} &= L_r i_{rq} + L_m i_{sq} \end{aligned} \tag{2}$$

And the electromagnetic torque:

$$T_{em} = \frac{3}{2} p \left(i_{rd} i_{sq} + i_{rq} i_{sd} \right) \tag{3}$$

2.2 Aerodynamic Turbine Modeling

The mechanical power P_m received from the wind can be denoted by the complex algebraic equation in function of blade pitch angle β, the rotor blades diameter R, the air density ρ, shaft and wind speed Ω_t, V_w, the tip speed ratio λ, and the power coefficient C_p as shown in Eq. (4):

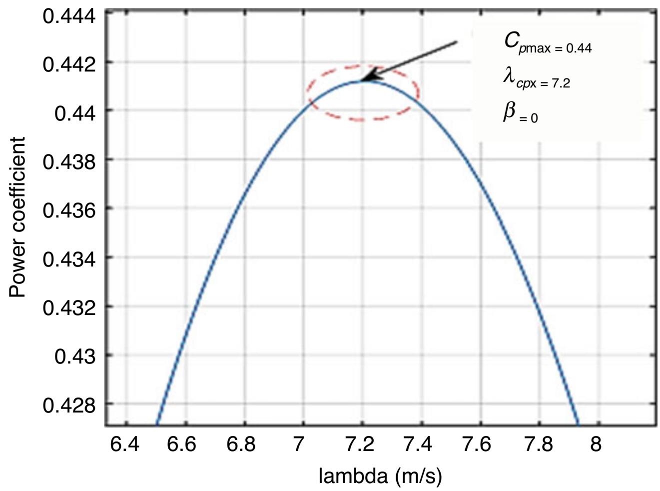


Fig. 2 The power coefficient C_p versus the tip speed ratio λ

$$P_m = \frac{1}{2}\rho\pi R^2 V_w^3 C_p(\lambda, \beta) \tag{4}$$

where λ is the tip speed ratio

$$\lambda = \frac{R\Omega_t}{V_w} \tag{5}$$

In this chapter, we use $C_{p\max} = 0.44$ and $\lambda_{opt} = 7.2$.

The gearbox is an essential part of the wind turbine systems (Fig. 2). It is usually used to adapt the low rotor shaft speed Ω_t to the high speed of the generator Ω_m. The following equations describe it in addition to the mechanical speed evolution:

$$\begin{cases} T_m = \frac{T_{aer}}{G} \\ \Omega_t = \frac{\Omega_m}{G} \end{cases} \tag{6}$$

$$J\frac{d\Omega_m}{dt} = T_m - T_{em} - f\Omega_m \tag{7}$$

3 Control Techniques of the DFIG

As mentioned before, the control strategies used in this chapter are the classic vector control and the direct torque control. Both approaches will be well described in the succeeding two subsections.

3.1 Classic Vector Control

Numerous control approaches are applied for the DFIG control; despite of that the vector control techniques were the most established ones. The use of the dq synchronous reference frame is ordinarily shared between the DFIG and the other machines [12]. From the previous dynamic model of the DFIG, and basing on the d-axis alignment of the stator flux ($\psi_{qs} = 0$), we obtain the two rotor voltage equations:

$$u_{rd} = R_r i_{rd} + \sigma L_r \frac{di_{rd}}{dt} - \omega_r L_r i_{rq} + \frac{L_m}{L_s} \frac{d}{dt} \overrightarrow{|\psi_s|} \tag{8}$$

$$u_{rq} = R_r i_{rq} + \sigma L_r \frac{di_{rq}}{dt} - \omega_r L_r i_{rd} + \omega_r \frac{L_m}{L_s} \frac{d}{dt} \overrightarrow{|\psi_s|} \tag{9}$$

where σ is the Blondel's coefficient, $\sigma = 1 - \frac{L_m^2}{L_r L_s}$.

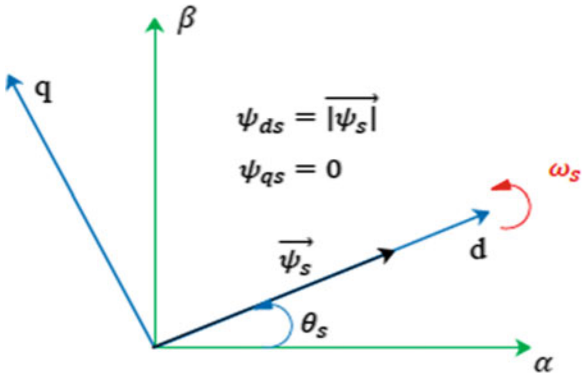
Basing on the d-axis alignment as shown in Fig. 3, we see that the reactive power Q_s is directly affected by the current I_{dr}, while the current I_{qr} modified the torque/active power [10] as shown in Eqs. (10) and (11):

$$T_{em} = \frac{3}{2} p \left(\psi_{qs} i_{dr} - \psi_{ds} i_{qr}\right) \Rightarrow T_{em} = -\frac{3}{2} p \frac{L_m}{L_s} \overrightarrow{|\psi_s|} i_{qr} \tag{10}$$

$$Q_s = -\frac{3}{2} \omega_s \frac{L_m}{L_s} \overrightarrow{|\psi_s|} \left(i_{dr} - \frac{\overrightarrow{|\psi_s|}}{L_m} \right) \tag{11}$$

The detailed explanations of the vector control technique and the current control loops used in this strategy are well established in [10]. Furthermore, the complete vector control schema used in this chapter is well drawn in Fig. 4.

Fig. 3 Stator flux-oriented in the dq reference frame



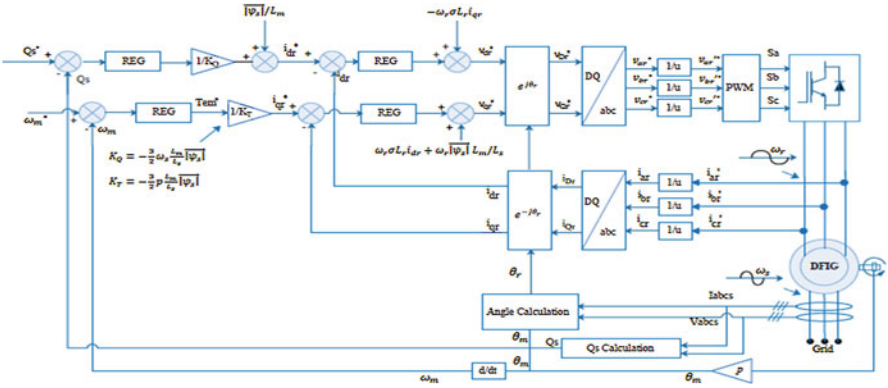


Fig. 4 The overall vector control diagram applied to the rotor side converter (RSC) of the DFIG

3.2 *Direct Torque Control Principles for the DFIG*

The direct control strategies are recognized like the modern solutions for the AC drives. Usually, from the side of the performance and global features, and compared to the vector control strategies, there are many differences. Several works have fulfilled in this area in Gonzalo et al. [9], Zhang et al. [13], Jihène et al. [14] and Xiong et al. [15], and as a consequence, various contributions to these new techniques are tested for a proper DFIG-based wind turbine applications. Referring to Haitham et al. [10], many facts about the DTC principle can be extracted: first, two variables of the DFIG have immediately commanded: the amplitudes of both rotor flux and the electromagnetic torque. Moreover, the angle δ denotes the distance between both stator and rotor flux vectors and by adjusting this angle we can control the torque. Furthermore, with the help of the injection of different voltage vectors to the rotor of the DFIG, the rotor flux trajectory and amplitude can be directly modified using a two-level voltage source converter (VSC). Eventually, the DTC is nearly related to the VSC that is manipulated. Therefore, the pulses are generated directly without any modulation schema.

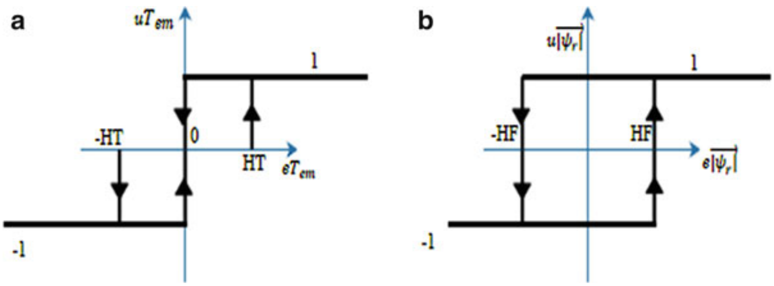


Fig. 5 (**a**) Three-level hysteresis comparator for the torque, (**b**) two-level hysteresis comparator for the rotor flux

Two hysteresis comparators are used in the classic DTC; one for the flux controller and it is based on two-level hysteresis comparator with the band HF and the other is based on three-level hysteresis comparator HT for the electromagnetic torque. Practically, the band values are usually restricted using the lowest switching sample time of the implementation tools [16]. Figure 5 shows the two hysteresis controllers for both flux and torque.

After recognizing the fundamental principle of the DTC, the next step is to estimate the main variables of the control technique (T_{em}, rotor flux and the rotor flux angle), which can be described directly by the following equations:

$$T_{\mathrm{em}} = \frac{3}{2} p \frac{L_{\mathrm{m}}}{\sigma L_{\mathrm{r}} L_{\mathrm{s}}} |\overrightarrow{\psi_{\mathrm{r}}}||\overrightarrow{\psi_{\mathrm{s}}}| \sin\ \delta \tag{12}$$

$$\psi_{\mathrm{rd}} = \int_0^t (V_{\mathrm{rd}} - R_{\mathrm{r}} I_{\mathrm{rd}})$$

$$\psi_{\mathrm{rq}} = \int_0^t \left(V_{\mathrm{rq}} - R_{\mathrm{r}} I_{\mathrm{rq}}\right) \tag{13}$$

$$|\overrightarrow{\psi_{\mathrm{r}}}| = \sqrt{\psi_{\mathrm{rd}}^2 + \psi_{\mathrm{rq}}^2} \tag{14}$$

Besides, after the computation of the rotor flux dq components, it is necessary to determine the rotor flux angle $\theta_{\psi\mathrm{r}}$, then the sectors can be simply calculated.

$$\theta_{\psi\mathrm{r}} = a \tan\left(\frac{\psi_{\mathrm{rq}}}{\psi_{\mathrm{rd}}}\right) \tag{15}$$

Table 1 Rotor voltage vector selection according to torque and flux errors

		uT_{em}		
		1	0	−1
$u\lvert\overrightarrow{\psi_r}\rvert$	1		$V_{(i-1)}V_0, V_7V_{(i+1)}$	
	0		$V_{(i-1)}V_0, V_7V_{(i+1)}$	

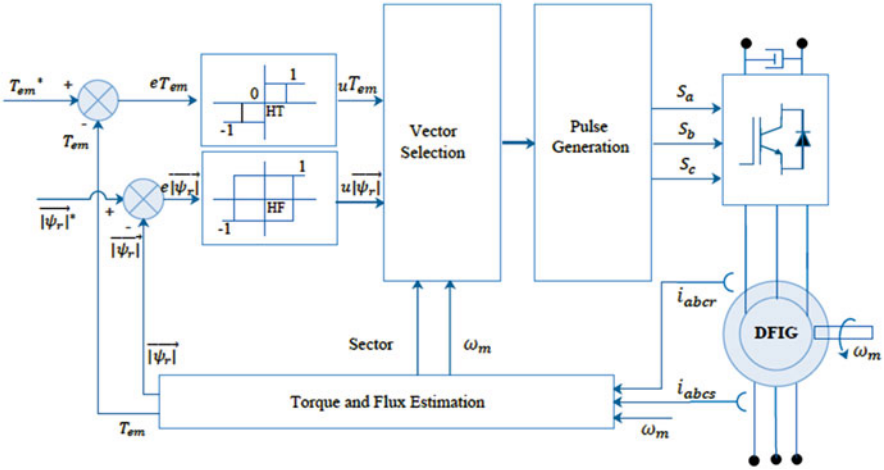


Fig. 6 DTC diagram applied to the RSC of the DFIG

The eight space voltage vectors are calculated using:

$$\begin{cases} V_i = \frac{2}{3} V_{dc} e^{-j\frac{\pi}{3}(i-1)} \\ V_0 = V_7 = 0 \end{cases} \tag{16}$$

where i is the sector number where $i = [1 \ldots 6]$.

Then, the selection of the proper rotor space vector is done based on Table 1.

Finally, the entire diagram of the DTC for DFIG wind system is well depicted in Fig. 6.

4 Simulation and Results

The simulation is made using the MATLAB/SIMULINK software package to examine the DTC and vector control strategies performance for a 2 MW DFIG-WT described in the Appendix. In this chapter, two control strategies (DTC and the classic vector control) will be compared to discuss the performance of a 2 MW DFIG-WT. Four parameters will be investigated (the electromagnetic torque T_{em}, the rotor flux ψ_r, the rotor current I_r, the stator power P_s). For the simulation of both control strategies, we use the magnetization through the stator ($I_{dr} = 0$), and all the parameters used in this simulation are well cited in the Appendix.

4.1 Vector Control Simulation

Figures 7, 8, 9 and 10 show the essential rotor side converter (RSC) variables (rotor currents, rotor flux, the electromagnetic torque, and the stator power). From Fig. 7a, b, we see that the direct and the quadratic components of the rotor currents follow the references after nearly 3 s after a transient.

The rotor flux amplitude and the electromagnetic torque are shown by the Fig. 8a, b; after the transient state, the two variables remain stable state after nearly 5 s and the electromagnetic torque coincide with the reference at that time.

Furthermore, the power delivered by the stator is well illustrated in the Fig. 9, we see that the power after the starting state increased until achieving 1.77 MW at permanent.

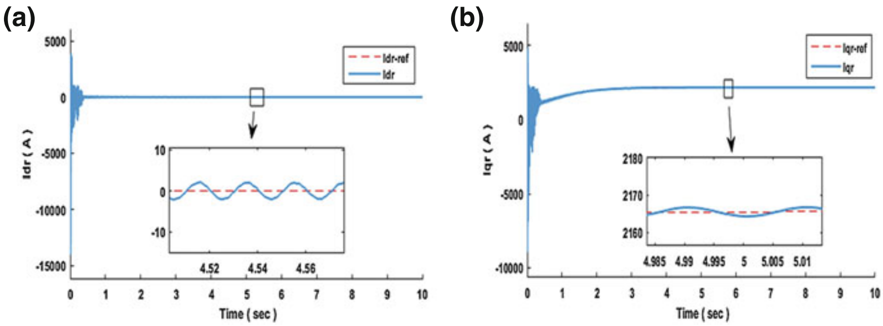


Fig. 7 Dq rotor current components versus references, (**a**) direct component, (**b**) quadratic component

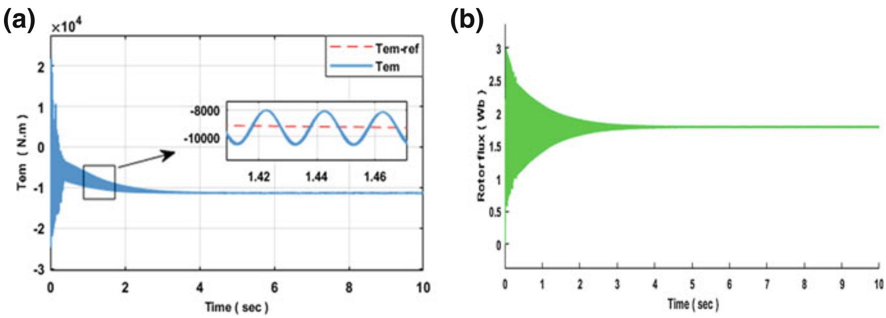


Fig. 8 (**a**) Electromagnetic torque versus the reference, (**b**) rotor flux magnitude

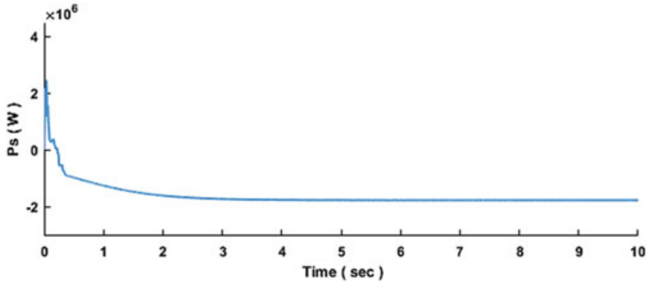


Fig. 9 The stator power

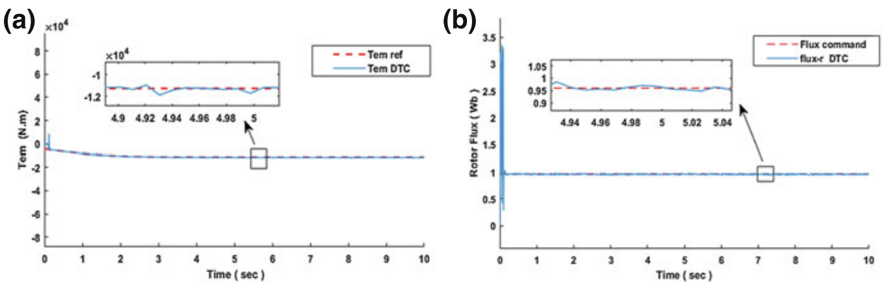


Fig. 10 (**a**) Electromagnetic torque versus the reference, (**b**) rotor flux magnitude versus reference

4.2 Direct Torque Control Simulation

In this part, the same parameters of the vector control except the rotor currents will be treated. Besides, the following Fig. 10a, b show the rotor flux amplitude and the electromagnetic torque, starting with rotor flux amplitude which follows the reference ($\psi_{\mathrm{r-ref}} = 0.96\,wb$) faster as shown in Fig. 10a. Moreover, the Fig. 10b demonstrates how the torque matches the reference rapidly with a neglected error, and both zooms confirm that. Figure 11 displays the stator power delivered to the grid, and we see that after a transient starting, the stator power achieves 1.65 MW at steady state.

4.3 Comparative Studies

In this section, only the electromagnetic torque and the stator power will be compared between both control techniques. The Fig. 12a illustrates the developed electromagnetic torque for the two strategies, and it is clear that the DTC technique coincides with the reference rapidly than the vector control. Besides, the power delivered from the stator for both strategies is shown by Fig. 12b. It is well

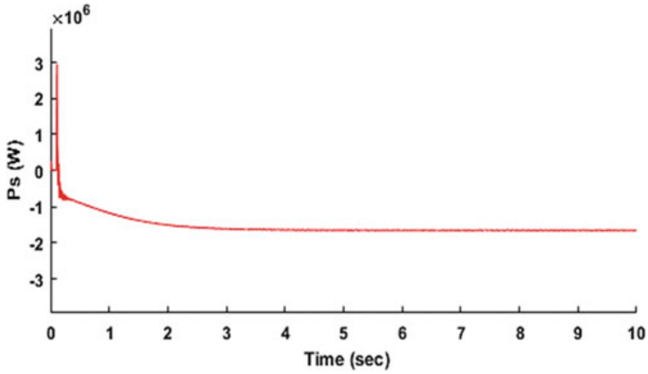


Fig. 11 The stator power

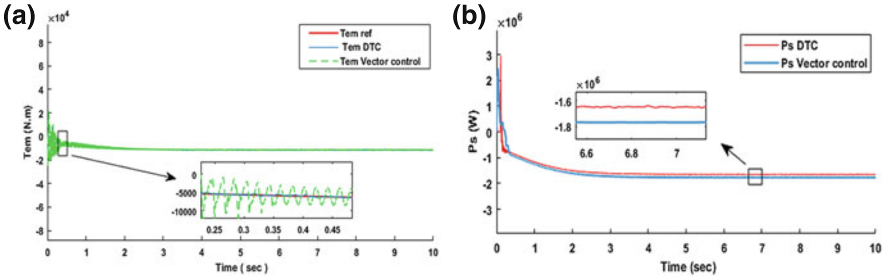


Fig. 12 Comparison of (**a**) the electromagnetic torque and (**b**) the stator power

demonstrated that the vector control has more power (1.77 MW) than the DTC (1.65 MW), knowing that in this chapter we use the generator convention, so the negative sign means generating mode.

5 Conclusion

In this chapter, a comparative study between the vector control and the direct torque control (DTC) strategies applied for the rotor side converter (RSC) has been demonstrated. The classic vector control needs many control loops with PI controllers to generate references, but in contrast, it delivered power more than the DTC control. With regard to the DTC, it has a simple configuration and a fast response for both flux and torque, but producing more power losses due to the variable switching frequency. Finally, this study can afford a good support for proper control and operation of the machine in term of performance and complexity of the chosen control techniques.

A.1 Appendix

A.1.1 DFIG-WT Parameters

V_s (line-line) =690 V, $f = 50$ Hz, $P_{nom} = 2$ MW, V_r (line-line) = 2070 V, $P = 2$, $u = 1/3$, $I_s = 1070$ A, (max slip) $s_{max} = 1/3$, (rated) T_{em}=12,732 N.m, $Fs = 1.9733$ Wb, $R_s = 0.0026\ \Omega$, $R_r = 0.0029\ \Omega$, $L_s = \mathrm{L_r} = 0.0026$ H, $L_m = 0.0025$ H, $\beta = 0$, $J = 127$ kg m^2, $f = 0.001$, $\sigma = 0.0661$, $V_{dc} = 1150$ V, $R = 42$, $\rho = 1.1225$, $G = 100$. $K_{opt} = 270{,}000$, ns = synchronous speed = 1500 rev/min, $s = 0.000002$ s, $V_w = 12$ m/s, (initial slip) = 0.2.

A.1.2 Parameters of the DTC

Ts_DTC = 0.00002 s
HT = HF = 1%.

A.1.3 Parameters of the PI Controllers

Kp_id = Kp_iq = 0.5771, Ki_id = Ki_iq = 491.6
Kp_n = 10,160, Ki_n = 406,400.

References

1. Enerdata. (2018). *The global energy statistical yearbook*. https://yearbook.enerdata.net.
2. Dehong, X., Frede, B., Wenjie, C., & Nan, Z. (2018). *Advanced control of doubly fed induction generator for wind power systems*. Wiley.
3. Fox, B., Leslie, B., Damian, F., Nick, J., David, M., Mark, O., Richard, W., & Olimpo, A. (2014). *Wind power integration: Connection and system operational aspects* (2nd ed.). Institution of Engineering and Technology.
4. Kerrouche, K., Mezouar, A., & Belgacem, K. (2013). Decoupled control of doubly fed induction generator by vector control for wind energy conversion system. In *Energy procedia* (pp. 239–248).
5. Fei, G., Tao, Z., & Zengping, W. (2012). Comparative study of direct power control with vector control for rotor side converter of DFIG. In *9th IET International Conference on advances in power system control, operation and management* (pp. 1–6).
6. Arnalte, S., Burgos, J. C., & Rodríguez-Amenedo, J. L. (2013). Direct torque control of a doubly-fed induction generator for variable speed wind turbines. *Electric Power Components & Systems, 30*(2), 199–216.
7. Li, Y., Hang, L., Li, G., Guo, Y., Zou, Y., Chen, J., Li, J., Jian, Z., & Li, S. (2016). An improved DTC controller for DFIG-based wind generation system. In *IEEE 8th International Power Electronics and Motion Control Conference (IPEMC-ECCE Asia)* (pp. 1423–1426).

8. Kumar, A., & GiriBabu, D. (2016). Performance improvement of DFIG fed Wind Energy Conversion system using ANFIS controller. In *2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)* (pp. 202–206).
9. Gonzalo, A., Jesus, L., Miguel, R., Luis, M., & Grzegorz, I. (2011). *Doubly fed induction machine modeling and control for wind energy generation*. Wiley.
10. Haitham, A., Mariusz, M., & Kamal, A. (2014). *Power electronics for renewable energy systems, transportation and industrial applications* (1st ed.). Wiley.
11. Zhaoyang, S., Ping, W., & Pengxian, S. (2014). Research on control strategy of DFIG rotor side converter. In *IEEE Conference and Expo Transportation Electrification Asia-Pacific (ITEC Asia-Pacific)* (pp. 1–5).
12. Bin, W., Yongqiang, L., Navid, Z., & Samir, K. (2011). *Power conversion and control of wind energy systems*. Wiley.
13. Zhang, Y., Li, Z., Wang, T., Xu, W., & Zhu, J. (2011). Evaluation of a class of improved DTC method applied in DFIG for wind energy applications. In *International Conference on electrical machines and systems* (pp. 1–6).
14. Jihène, B., Adel, K., & Mohamed, F. M. (2011). DTC, DPC and Nonlinear Vector Control Strategies Applied to the DFIG Operated at Variable Speed. *Journal of Electrical Engineering, 11*(3), 1–13.
15. Xiong, P., & Sun, D. (2016). Backstepping-based DPC strategy of a wind turbine driven DFIG under Normal and harmonic grid voltage. *IEEE Transactions on Power Electronics, 31*(6), 4216–4225.
16. Adel, K., Mohamed, F., & M. (2010). Sensorless-adaptive DTC of double star induction motor. *Energy Conversion and Management, 51*(12), 2878–2892.

Big Data Analytics, Reshaping the New Trends of Healthcare: Literature Review

Rabhi Loubna, Falih Noureddine, Afraites Lekbir, and Bouikhalene Belaid

Abstract Big Data in healthcare refers to the huge volume of heterogeneous medical data related to patient healthcare and well-being (social media post, lab reports, case history, sensor data, list of doctors and nurses in a particular hospital, etc.). These data are difficult to manage with traditional data management tools and methods. Analyzing these big data by specific tools and advanced analytics techniques offers immense potential to improve the quality of care and reduce waste, error, and the cost of care. Big data analytics in healthcare aims to provide relevant information from data to make decisions in real or near real time using various analytical concepts such as data mining, artificial intelligence, and predictive analytics. In this chapter, we present a literature review of big data analytics and its challenges for decision support in the healthcare field.

Keywords Big data · Big data analytics · Predictive analytics · Healthcare

1 Introduction

Big data in healthcare refers to the huge amount of electronic health data coming from heterogeneous sources at a very high speed [1] that it is not possible for traditional tools and techniques to analyze and extract value from it. Analytics of big data refers to the process of extracting insight from raw data by examining and analyzing its behavior and patterns using qualitative and quantitative techniques.

Recently, this large volume of big data cannot be analyzed by traditional tools, rather advanced techniques called big data analytics which has gained great

R. Loubna (✉) · A. Lekbir
Mathematics and Applications Laboratory, Faculty of Sciences and Technologies, Sultan Moulay Slimane University, Beni Mellal, Morocco

F. Noureddine · B. Belaid
Laboratory of Innovation in Mathematics, Applications and Information Technologies, Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco
e-mail: b.bouikhalene@usms.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_63

importance in the last years. Sophisticated analytics can substantially improve decision-making, minimize risks, and uncover valuable insights from the data. So, decisions need to be augmented by analyzing huge datasets using big data techniques and technologies.

This chapter tries to cover major aspects of big data analytics in smart health. It begins with a brief introduction of the topic. Important characteristics of healthcare big data is discussed in Sect. 2. Big data analytics in healthcare is discussed in Sect. 3. In Sect. 4, we discuss added values of big data analytics in healthcare. The challenges of these analytics are treated in Sect. 5, and the conclusion highlights future trends and perspectives of big data analytics approach in healthcare.

2 Related Works

The summary of success findings of big data in healthcare are summarized in Table 1. It explains the background of recent studies that treat the same subject of big data analytics applied in healthcare domain.

3 Big Data Definition

Big data is commonly defined through the 5Vs: volume, velocity, variety, veracity, and value. By applying this definition in health care, big data means [5]:

- **Volume**: Denotes the large amount of data which is being generated in every second from different sources. In healthcare, these data include personal information of patients, doctors and nurses, radiology images [10].
- **Variety**: Means that the incoming data can have different types as [11]:
  - *Structured*: Which is can be stored in table with rows and columns such as clinical data.
  - *Semi-structured*: Can be converted into structured data (e.g., doctor notes, paper prescriptions).
  - *Unstructured*: It has no standard structure, and it is very difficult to be stored in database (e.g., radiology images...).
  - *Multi-structured*: A mix of structured, semi structured and unstructured data.
- **Velocity**: Measures the speed at which data can be collected, analyzed, and exploited (i.e., real-time patient data).
- **Veracity**: Denotes the correctness and accuracy of the data. Veracity is of acute concern in healthcare because life or death decisions depend on having the accurate information [12, 13].
- **Value**: The useful data to the decision maker through the ability to take meaningful action based on insights derived from big data that we work with.

Table 1 Summary of findings of big data in healthcare

Code	Authors	Problem focused on	Work done	Outcome
1	Archenaa and Mary Anita [2]	Big data analytics in healthcare	Discussed how we can uncover additional value from the data generated by healthcare and government	Providing an effective data driven service to citizens by predicting their needs based on the analysis using Hadoop
2	Adriana Alexandru et al. [3]	– Advantages and challenges of big data in healthcare field – Big data architecture for healthcare	Discussed the required relation between big data and the smart health	Providing big data architecture for healthcare
3	Belle et al. [4]	Big data analytics in healthcare	Described big data analytics, characteristics, and potential and implementation of big data analytics in healthcare	Implementing big data analytics in healthcare
4	Ganjir et al. [5]	Big data analytics in healthcare	Big data analytics and its characteristics, advantages, and challenges in healthcare	Big data analytics applied for smart health
5	Wang et al. [6]	Big data analytics: Capabilities and potential benefits for healthcare organizations	Discussed implementation of big data analytics in healthcare organizations	Roles of big data analytics for healthcare organizations
6	Mehta and Pandit [1]	Concurrence of big data analytics and healthcare: a systematic review	Determined the scope of big data analytics in healthcare including its applications and challenges in its adoption in healthcare. It also intends to identify the strategies to overcome the challenges	Big data analytics finds its application for clinical decision support; optimization of clinical operations and reduction of cost of care
7	Raghupathi and Raghupathi [7]	Big data analytics in healthcare: Promise and potential	Described the nascent field of big data analytics in healthcare, the benefits, outlines an architectural framework and methodology Discussed the challenges	Providing a broad overview of big data analytics for healthcare researchers and practitioners

(continued)

Table 1 (continued)

Code	Authors	Problem focused on	Work done	Outcome
8	Kyoungyoung Jee and Gang-Hoon Kim [8]	Potentiality of big data in the medical sector: Focus on how to reshape the healthcare system	Providing an overview of the current state of big data applications in the healthcare environment, this study has explored the current challenges that governments and healthcare stakeholders are facing as well as the opportunities presented by big data	Insightful consideration of the current state of big data applications could help follower countries or healthcare stakeholders in their plans for deploying big data to resolve healthcare issues. The advantage for such follower countries and healthcare stakeholders is that they can possibly leapfrog the leaders' big data applications by conducting a careful analysis of the leaders' successes and failures and exploiting the expected future opportunities in mobile services
9	YichuanWang and Nick Hajli [9]	Exploring the path to big data analytics success in healthcare	Illustrated how healthcare organizations' big data analytics capabilities can be developed by big data analytics architecture	Providing new insights to healthcare practitioners on how to constitute big data analytics capabilities for business transformation and offer an empirical basis that can stimulate a more detailed investigation of big data analytics implementation

4 Big Data Analytics in Healthcare

Nowadays, the data that need to be analyzed are big, contain heterogeneous data types, and even including streaming data which may change the statistical and data analysis approaches. Therefore, traditional tools are not able to analyze this category of data. So, new approach of big data called big data analytics was born.

Big data analytics refers to advanced technologies designed to work with large volumes of heterogeneous data to build evidence for improved care delivery. They

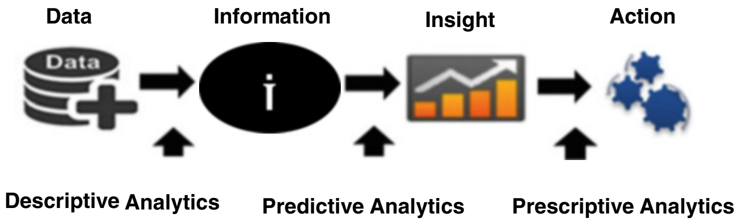


Fig. 1 Types of analytics techniques

can be classified into three types: descriptive analytics, predictive analytics, and prescriptive analytics (Fig. 1) [14]:

Descriptive Analytics Gives information about what happened. In this technique, based on historical data, new insights are developed using statistical descriptions (such as statistic summary, correlations and sampling . . .) and clustering (such as *K*-means . . .).

Predictive Analytics Predicts the future outcomes using new statistical methods and predictive algorithms such as "Decision Tree" [15]. It provides information on what will likely happen in the future and what actions can be taken [16].

Prescriptive Analytics It is a type of predictive analytics. It helps to derive the best possible outcome by analyzing the possible outcomes by answering the question so what? now what?

5 Benefits of Big Data Analytics on the Healthcare

Among the advantages of big data analytics in healthcare, we can cite:

5.1 Prediction of Disease Outbreaks

Big data analytics is used for monitoring diseases on networks like Facebook, Twitter, YouTube, etc. for healthcare. For example, using big data analytics can help to control the timing, location, and symptoms of search engine queries to predict disease outbreaks [2]. To date, as health demands from social networking sites continue to grow, it can potentially support key prevention programs such as outbreak management and disease surveillance [4].

5.2 Management of Diseases

Big data can also be used to check the effectiveness of a treatment. For example, in the field of vaccines, clinicians today measure hundreds of parameters during clinical trials: cell counts, cellular functionality, expression of genes of interest, etc., whereas a few years ago, was limited to the concentration of the antibodies of interest. Ultimately, this evolution, the massive data that it generates and the capacity to analyze them, could make it possible to verify that a vaccination worked well after only 1 h, starting from a micro drop of blood [17].

5.3 Support Clinical Decision

Big data analytics can help patients make the relevant and the right decision in real or near time. For example, analytics can contribute to proactive care in early stages of some diseases based on vast amounts of historical data [7].

6 Challenge of Big Data Analytics in Healthcare

6.1 Reliability of Data

Medical big data is increasingly powered by data that comes from multiple health apps [5]. Today, anyone can create a health app. And if the latter is not referenced as a medical device, we cannot appreciate the merits and sourcing of these masses of data, since we cannot consider them reliable.

6.2 Privacy and Security of Care Data

Big data challenges the paradigms of medicine as we know it. In fact, the use of personal medical data and big data imposes an ethical challenge to confidentiality, the right and freedom of access, commercialization, security, accountability, and medical confidentiality [17]. In particular, the development of e-Health, telemedicine, and big data medical involves legal changes, economic breaks. In addition, the doctor–patient relationship prevailed by a well-established order, values, principles, and rules is disrupted [18].

6.3 Risk of Patient Care

In view of the growing importance of big data in relation to health, it is questionable whether clinical and therapeutic practice can be replaced by pure big data analytics. Until the 2000s, the doctor–patient relationship was governed by a charitable principle, and the information was given and controlled by the doctor. But today, with social networks, health forums, patient associations, big data, and new technologies of information and communication (NTIC), the patient has become an active, free, and autonomous cyber-patient, and the doctor is no longer there to discover the medical information, but to explain it to the patient. Moreover, the proliferation of expert systems and health applications, which allow the patient to ask questions, benefit from a pre-medical consultation and to allow a follow-up of certain biometric and physiological data, risks transforming the doctor into "Data Manager" in charge of recovering, sorting, organizing, and explaining health data, and the patient became himself an actor of his health.

7 Big Data Analytics Architecture for Healthcare

Inspired by Adriana Alexandru et al. [3], the architecture for healthcare is presented in Fig. 2. Its components consist of four important layers:

The data layer: Includes hospital data sources which can have structured data (e.g., patient records), semi-structured data (e.g., the logs of health monitoring devices), or unstructured data (e.g., clinical images) [4].

The data aggregation layer: Performs three important operations on the data from various sources:

- *Data acquisition*: The process of collecting data from various sources.
- *Data transformation*: Detect, move, clean, split, and filter the unnecessary, inconsistent, and incomplete data using ETL (Extract, Transform, Load) tools. For example, structured data extracted from an Electronic Health Record (EHR) system could be changed, it could be sorted and then validated to respect quality requirements.
- *Data storage*: The persistently storing of the data into appropriate databases such as Hadoop distributed file systems (HDFS) or in a Hadoop cloud in order to be ready for further analysis. Traditional relational database systems based on SQL (e.g., MySQL, Oracle) are no longer appropriate due to their low scalability, extensibility, and velocity. They must be replaced by distributed, non-relational NoSQL databases which were built for large-scale data storage. Four types of database are available: document-stored (e.g., MongoDB, CauchDB), wide column-stored (e.g., HBase, Cassandra, DynamoDB), key value-stored (e.g., Redis, Riak), and graph-oriented (e.g., Neo4J) [6].

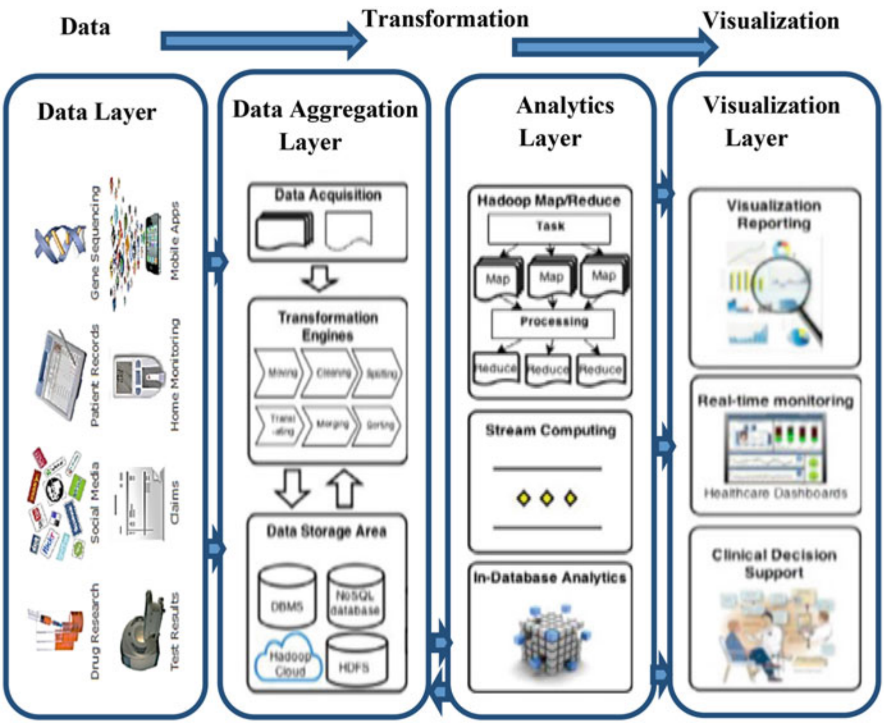


Fig. 2 Big data analytics architecture in healthcare. (Adapted from [3])

The analytics layer: The process of analyzing data coming from data aggregation layer uses the following types of data analysis:

- *Hadoop Map/Reduce*: MapReduce is the most commonly used programing model in big data analytics. It is designed to process huge datasets in batch form and analyze both structured and unstructured data [1, 19].
- *Stream computing*: Can make high-performance stream data processing in real or near real time. For example, stream computing can help prevent healthcare fraud by predicting the likelihood of illegal transactions or deliberate misuse of customer accounts. Thus, we can track data in motion, respond to unexpected events as they happen and quickly determine next-best actions.
- *In-database analytics*: Means data mining approach that is built on an analytic platform that allows data to be processed within the data warehouse. It allows for high-speed parallel processing, but its results are not in real time and thus only a static prediction can be possible. In-database analytics is useful for preventive healthcare and pharmaceutical management.

The information exploration layer can generate:

- *Visualization reports*: Representation of data in an interactive way in order to help healthcare decision-making [20].
- *Real-time monitoring of information* (e.g., of physiological readings, alerts): Can be used to assess a patient's health and help avoid exacerbations [19].
- *Clinical decision support*: Analysis of data about patients can lead to new discoveries which support evidence-based medicine.

8 Conclusion and Future Trends

The authors of this chapter have examined the innovative topic of "Big Data." They aim to reflect on big data analytics and its application in healthcare.

First, this chapter defined what is meant by big data in health industry to consolidate the divergent speech in this topic. Second, it puts the accent on "Big data analytics" where sophisticated analytics techniques are applied on healthcare big data in order to store, analyze, and treat this extensive heave of enormous data. Third, advantages and challenges of big data in smart health has been discussed. Fourth and finally, architecture of big data analytics in healthcare is presented.

Although, major innovations in analytical techniques for big data have not yet taken place. Our future work is planned in the sense of the contribution to the good governance of "Big Data Analytics Systems." For instance, real-time analytics will likely become a profitable field of research following the remarkable growth in location-aware social media and mobile apps.

References

1. Mehta, N., & Pandit, A. (2018). Concurrence of big data analytics and healthcare: A systematic review. *International Journal of Medical Informatics, 114*(January), 57–65.
2. Archenaa, J., & Anita, E. A. M. (2015). A survey of big data analytics in healthcare and government. *Procedia Computer Science, 50*, 408–413.
3. Alexandru, C. A. A. A., & Coardos, E. T. D. (2016). Healthcare, big data and cloud computing. *Proceedings of the VLDB Endowment, 4*(1–2), 2415–1521.
4. Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., Navidi, F., Beard, D. A., & Najarian, K. (2007). Big data analytics in healthcare. *Nature Protocols, 2*(12), 3270–3277.
5. Ganjir, V., Sarkar, B. K., & Kumar, R. R. (2016). *Big data analytics for healthcare* (Vol. 6, pp. 1–6).
6. Wang, Y., Kung, L. A., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change, 126*, 3–13.
7. Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems, 2*(1), 3.
8. Jee, K., & Kim, G.-H. (2013). Potentiality of big data in the medical sector: Focus on how to reshape the healthcare system. *Healthcare Informatics Research, 19*(2), 79.

9. Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research, 70*, 287–299.
10. Ta, V. D., Liu, C. M., & Nkabinde, G. W. (2016). Big data stream computing in healthcare real-time analytics. In *Proceedings of 2016 IEEE International conference on cloud computing and big data analysis* (pp. 37–42).
11. Zhang, Y., Qiu, M., Tsai, C. W., Hassan, M. M., & Alamri, A. (2017). Health-CPS: Healthcare cyber-physical system assisted by cloud and big data. *IEEE Systems Journal, 11*(1), 88–95.
12. Augustine, D. P. (2014). Leveraging big data analytics and Hadoop in developing India's healthcare services. *International Journal of Computers and Applications, 89*(16), 975–8887.
13. Lee, C., et al. (2017). *Big healthcare data analytics: Challenges and applications*. Springer.
14. Rai, B. K., Meshram, A. A., & Gunasekaran, A. (2018). Big data in healthcare management: A review of literature. *American Journal of Theoretical and Applied Business, 4*(2), 57–69.
15. Tan, S. S.-L., Gao, G., & Koch, S. (2015). *Big data and analytics in workplace learning* (pp. 546–547).
16. Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *IEEE Access, 5*(c), 8869–8879.
17. Shen, B. (2017). *Healthcare and big data management* (Vol. 1028). Springer.
18. Viceconti, M., Hunter, P., & Hose, R. (2015). Big data, big knowledge: Big data for personalized healthcare. In *Finns in the United States: A history of settlement, dissent, and integration* (Vol. 19, pp. 1209–1215).
19. Antony Basco, J., & Senthilkumar, N. C. (2017). Real-time analysis of healthcare using big data analytics. In *IOP conference series: Materials science and engineering* (Vol. 263).
20. Trifirò, G., Sultana, J., & Bate, A. (2018). From big data to smart data for pharmacovigilance: The role of healthcare databases and other emerging sources. *Drug Safety, 41*(2), 143–149.

Robust Admissibility of Uncertain Discrete-Time Switched Singular Systems with Time-Varying Delay

Mohammed Charqi, El Houssaine Tissir, Noureddine Chaibi, and Mohamed Ouahi

Abstract In this chapter, we focus on robust admissibility for uncertain discrete-time switched singular systems with time-varying delay for arbitrary switching law. The parametric uncertainties are assumed to be linear fractional perturbations. New matrix inequalities are proposed to establish new sufficient conditions that guarantee causality, regularity, and asymptotic stability of the nominal considered system in terms of linear matrix inequalities (LMIs). The approach is based on constructing a novel Lyapunov-Krasovskii functional, and the conditions proposed are dependent on the lower and upper delay bounds. Furthermore, slack variables are introduced for more relaxation. Finally, some numerical examples are provided to demonstrate the effectiveness of the proposed approach and to compare the obtained results with the existing ones in the literature.

Keywords Discrete-time switched singular systems · Switching signal · Lyapunov-Krasovskii functional · Time-varying delay · Robust admissibility · Linear matrix inequality (LMI)

1 Introduction

Switched systems are an important class of hybrid dynamical systems which consist of a family of continuous or discrete-time subsystems and a logical rule that orchestrates switching between these subsystems at each instant of time [1–3]. They

M. Charqi (✉) · El H. Tissir · N. Chaibi
Laboratory of Electronics, Signals Systems and Data Processing (LESSI), Department of Physics, Faculty of Sciences Dhar El Mehraz, Sidi Mohammed Ben Abdellah University, B.P.1796, Fez-Atlas Fez, Morocco
e-mail: mohammed.charq@usmba.ac.ma; elhoussaine.tissir@usmba.ac.ma

M. Ouahi
Lisa, National School of Applied Science (ENSA), Sidi Mohammed Ben Abdellah University, BP 72, Industrial District Ain Chkef, Bensouda, Fez, Morocco
e-mail: mohamed.ouahi@usmba.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_64

have received a growing attention during the past two decades [4–10], and the references therein. Furthermore, there are many practical systems and processes which can be presented by switched system models, such as manufacturing, automate highway systems, and they have several applications in the control of mechanical systems, automotive engine control, power systems, communication networks, robot manufacture, constrained robotics, stepper motors, power electronics, chemical process, and water quality control, e.g., see [2, 4–8, 11–19, 45], and the references therein.

On the other hand, singular systems, also called descriptor systems, implicit or generalized state-space systems, often appear in various areas and have extensive applications in several practical systems such as chemical process, electrical networks, aircraft dynamics, robotic systems, lossless transmission lines, space navigation, electronic systems, optimization problems, and many other areas [20–32]. It is well known that when time delays are introduced in singular systems, their study becomes more complex and induce instability or poor performances. Recently, a great number of results in the literature have been reported on the analysis of stability, stabilization, and H_∞ control of singular systems with time-delay [20–23, 27, 28, 30–32].

In the past few years, many efforts have been done to the study of switched singular systems (SSs) or switched singular time-delay systems (SSTDs), which have strong applications in various areas, like economic systems [33], electrical networks [34, 35], DC motor, network control systems. Note that the behavior and the structure of SSTD systems are more complicated than those of switched singular systems without or with time-delay. This complexity is due to the existence of coupling between the modes switching and the time-delay and as algebraic constraints in singular models [21]. Nowadays, there exist only few results in the literature that studies SSTDs, more especially in the discrete-time, and several existing results have been reported especially in the study of admissibility (regularity, causality, and stability), stabilization (control), and filtering of this type of systems, either in continuous case [4, 33, 36] or in the discrete case [2, 35, 37–43, 46].

In [38], the problem of state feedback stabilization of a class of discrete-time switched singular systems with time-varying state delay under asynchronous switching is studied by using the multiple Lyapunov function approach and the average dwell time technique. The delay-independent exponential stability for discrete-time switched descriptor systems with time varying delay by constructing a Lyapunov functional is studied in [36]. The approach of average dwell time is manipulated in [40] for synthesizing the exponential admissibility of discrete-time switched singular time-delay systems, by defining a properly constructed decay-rate–dependent Lyapunov function. The robust exponential admissibility of uncertain switched singular time-delay systems is synthesized in [39]. The fault detection for discrete-time switched singular time-delay systems has been studied in [43], by constructing an appropriate decay-rate piecewise Lyapunov function and by using the average dwell time scheme.

In this chapter, we are concerned with the robust admissibility analysis problem for a class of discrete-time switched singular systems with time-varying delay and linear fractional parametric uncertainties for arbitrary switching law. The time-delay considered here is assumed to be time-varying delay and has minimum and maximum bounds, which covers the constant delay as special case, and which is more general and practical. The uncertainties are assumed to be of structured linear fractional form, which includes the norm-bounded uncertainty as a special case. By constructing a novel Lyapunov-Krasovskii functional combined with LMIs techniques and by introducing some slacks variables for more relaxation, new sufficient conditions are derived in terms of linear matrix inequality (LMIs), which are dependent on the lower and the upper delay bounds, for the nominal system to be regular, causal, and asymptotically stable. The novel criteria do not involve any model augmentation transformation, which often leads to large computational burden. The developed method can be easily tested by using standard numerical software and provides simple calculations and fewer decision variables than other results in the literature [38, 39]. Moreover, this result has other several advantages over some existing results in the literature [7, 8, 12, 19, 38, 39, 43]. Thus, less conservative stability conditions are obtained. The proposed approach is then extended to the robust admissibility of uncertain switched singular delay system. Finally, numerical examples are presented to illustrate the effectiveness of the developed results. The proposed method provides some less conservative results.

2 Problem Formulation and Preliminaries

Consider the discrete-time switched singular system with time-varying delay described by:

$$\begin{cases} E_{\sigma(k)}x\,(k+1) = \overline{A}_{\sigma(k)}x(k) + \overline{Ad}_{\sigma(k)}x\,(k-d(k)) \\ \quad x(k) = \phi(k), \quad k = -d_M, d_M+1, \ldots, 0 \end{cases} \tag{1}$$

where $x(k) \in \mathbb{R}^n$ is the state vector, $\phi(k)$ is an initial state function, $d(k)$ represents the time-varying delay and satisfies $d_m \le d(k) \le d_M$, where d_m and d_M are constant positive scalars representing the lower and upper bounds of delay, respectively, $E_{\sigma(k)} \in \mathbb{R}^{n\times n}$ may be singular with rank $(E_{\sigma(k)}) = r \le n$.

$\sigma(k)$ is a piecewise constant function of time, called a switching signal, which takes its values in the finite set $I_N = \{1, 2, \ldots, N\}$ and assumed to be available in real time, N is the number of subsystems or modes, we can put: $\sigma(k) : [0, +\infty] \to I_N = \{1, 2, \ldots, N\}$.

Moreover, $\sigma(k) = i$ implies that the ith subsystem is activated at time k with $i \in I_N$. It follows that when the ith subsystem is activated $\sigma(k) = i$, and the system (1) can be rewritten as:

$$E_i x\,(k+1) = \overline{A_i}x(k) + \overline{Ad_i}x\,(k-d(k)) \tag{2}$$

$\overline{A_i}$ and $\overline{Ad_i}$ are uncertain matrices which are assumed to be of the form:

$$\begin{aligned} \overline{A_i} &= A_i + M_i \Delta_i(k) N_{1i}, \\ \overline{Ad_i} &= Ad_i + M_i \Delta_i(k) N_{2i} \end{aligned} \tag{3}$$

where $A_i \in \mathbb{R}^{n \times n}$ and $Ad_i \in \mathbb{R}^{n \times n}$ are known real constant matrices representing the nominal system for each $i \in I_N$. $\Delta_i(k)$ are the parameter uncertain matrices and satisfying:

$$\Delta_i(k) = [I - \Pi_i(k) J_i]^{-1} \Pi_i(k) \tag{4}$$

$$J_i J_i^T < I \tag{5}$$

where M_i, N_{1i}, N_{2i}, J_i, and $i \in I_N$ are some given real constant matrices with appropriate dimensions. $\Pi_i(k)$, $i \in I_N$, are some real unknown matrices representing the perturbations which satisfy:

$$\Pi_i(k) \Pi_i(k)^T < I \tag{6}$$

Remark 1 The fractional parametric uncertainties located in (3)–(6) and that have been investigated in this note are usually called linear fractional uncertainties perturbations. So, it is easy to conclude that if $J_i = 0$ for all $i \in I_N$, the form of linear fractional uncertainties reduces to norm-bounded uncertainties one, which presents a special case and which is widely studied in the literature [39]. Notice also, that conditions (5) and (6) guarantee that $I - \Pi_i(k) J_i$ is invertible.

Remark 2 The delay used in systems (1) and (2) is assumed to be time-varying delay *d(k)* and have the lower and the upper bounds d_m and d_M, respectively. Note that if $d_m = d_M = d$, the time-varying delay *d(k)* reduces to a constant delay d that has been studied in a vast literature. Also, if *d(k)* changes only when system mode is switched, then the time delay becomes mode-dependent constant. Thus, the time-varying delay considered here covers the previous two cases.

Assumption 1 The switching rule is not known a priori, but we assume that its instantaneous value is available in real time.

Now, considering the nominal system of (2) described by:

$$E_i x(k+1) = A_i x(k) + Ad_i x(k - d(k)) \tag{7}$$

Definition 2.1 [26, 39, 40] For the switching signal $\sigma(k)$, system (7) (or the pair (E_i, A_i)) is said to be:

1. Regular if $\forall\ \sigma(k) = i$, $i \in I_N$, there exists a constant scalar $z_i \in \mathbb{C}$ such that $det(z_i E_i - A_i)$ is not identically zero.

2. Causal if $\forall\ \sigma(k) = i$, $i \in I_N$, $deg(det(z_i E_i - A_i)) =$ rank (E_i), for all $z_i \in \mathbb{C}$.
3. The discrete-time switched singular system (1) is said to be admissible, if it is regular, causal, and stable.

Definition 2.2 [26] For given integers $d_m > 0$ and $d_M > 0$, the discrete-time switched singular time-delay system (2) is said to be regular and causal for any time delay $d(k)$ satisfying $d_m \le d(k) \le d_M$, if the pair (E_i, A_i) is regular and causal.

Lemma 1 [29] For any constant matrices $\begin{bmatrix} X_i & Y_i \\ * & X_i \end{bmatrix} \ge 0, \forall i \in I_N$, positive integers d_m, d_M, and $d(k)$ satisfying $d_m \le d(k) \le d_M$, and vector function $x(k + .) : \mathbb{N}[-d_M, -d_m] \to \mathbb{R}^n$, such that the sums concerned are well defined, then:

$$-d_{12} \sum_{s=k-d_M}^{k-d_m-1} \eta^T(s) X_i \eta(s) \le \varpi^T(k) \Omega_i \varpi(k)$$

where $d_{12} = d_M - d_m$; $\eta(s) = x(s + 1) - x(s)$:

$$\varpi(k) = \begin{bmatrix} \left[x\,(k - d_m)\right. \\ \left[x\,(k - d(k))\right. \\ \left[x^T\,(k - d_M)\right. \end{bmatrix}; \quad \Omega_i = \begin{bmatrix} -X_i & X_i - Y_i & Y_i \\ * & -2X_i + Y_i + Y_i{}^T & -Y_i + X_i \\ * & * & -X_i \end{bmatrix}$$

Lemma 2 [20 (Finsler Lemma)] Consider a vector $\chi \in \mathbb{R}^n$, a symmetric matrix $Q \in \mathbb{R}^{n \times n}$ and a matrix $B \in \mathbb{R}^{m \times n}$, such that rank($B$)<$n$. The following statements are equivalent:

1. $\chi^T Q \chi < 0, \forall \chi$ such that $B.\ \chi = 0$, $\chi \ne 0$.
2. $B^{\neg T} Q B^{\neg} < 0$.
3. $\exists \mu \in \mathbb{R}$; $Q - \mu B^T B < 0$.

3 Main Results

3.1 Admissibility of Nominal System

In this section, we investigate the admissibility of system of discrete-time switched singular system with time-varying delay described in (7).

Theorem 1 The switched singular system (7) with time-varying delay $d(k)$ satisfying $d_m \le d(k) \le d_M$, with d_m and d_M are two positive integers, is said to be admissible for arbitrary switching law $\sigma(k) = i$, $\forall i \in I_N$, if there exist positive definite matrices P_i, Q_{ji}, R_{ji} $(j = 1, 2, 3)$, real matrices T_i and S_i with appropriate dimensions and matrices F_{0i}, F_{1i}, and F_{2i}, such that the following LMIs hold for each $i \in I_N$:

$$\Upsilon_i = \begin{bmatrix} \Upsilon_{11i} & \Upsilon_{12i} & \Upsilon_{13i} & \Upsilon_{14i} & \Upsilon_{15i} \\ * & \Upsilon_{22i} & 0 & \Upsilon_{24i} & 0 \\ * & * & \Upsilon_{33i} & \Upsilon_{34i} & \Upsilon_{35i} \\ * & * & * & \Upsilon_{44i} & \Upsilon_{45i} \\ * & * & * & * & \Upsilon_{55i} \end{bmatrix} < 0 \tag{8}$$

$$\begin{bmatrix} R_{3i} & Y_i \\ * & R_{3i} \end{bmatrix} \geq 0 \tag{9}$$

where (*) denotes the symmetric part in a symmetric matrix.

$$\Upsilon_{11i} = -E_i{}^T P_i E_i + \tilde{d} Q_{1i} + Q_{2i} + Q_{3i} + \left(d_M^2 - 1\right) E_i{}^T R_{1i} E_i + \left(d_m^2 - 1\right) E_i{}^T R_{2i} E_i + d_{12}^2 E_i^T R_{3i} E_{(i)} + S_i L_i^T A_i + A_i^T L_i S_i^T + F_{0i} A_i + A_i^T F_{0i}^T \tag{10a}$$

$$\Upsilon_{12i} = -d_M^2 E_i^T R_{1i} - d_m^2 E_i^T R_{2i} - d_{12}^2 E_i^T R_{3i} + A_i^T T_i - F_{0i} + A_i^T F_{1i}^T \tag{10b}$$

$$\Upsilon_{13i} = E_i^T R_{2i} E_i \tag{10c}$$

$$\Upsilon_{14i} = S_i L_i^T Ad_i + F_{0i} Ad_i + A_i^T F_{2i}^T \tag{10d}$$

$$\Upsilon_{15i} = E_i^T R_{1i} E_i \tag{10e}$$

$$\Upsilon_{22i} = P_i + d_M^2 R_{1i} + d_m^2 R_{2i} + d_{12}^2 R_{3i} - T_i - T_i^T - F_{1i} - F_{1i}^T \tag{10f}$$

$$\Upsilon_{24i} = T_i^T Ad_i + F_{1i} Ad_i - F_{2i}^T \tag{10g}$$

$$\Upsilon_{33i} = -Q_{3i} - E_i^T R_{2i} E_i - E_i^T R_{3i} E_i \tag{10h}$$

$$\Upsilon_{34i} = E_i^T R_{3i} E_i - E_i^T Y_i E_i \tag{10i}$$

$$\Upsilon_{35i} = E_i^T Y_i E_i \tag{10j}$$

$$\Upsilon_{44i} = -Q_{1i} - 2E_i^T R_{3i} E_i + E_i^T Y_i E_i + E_i Y_i^T E_i^T + F_{2i} Ad_i + Ad_i^T F_{2i}^T \tag{10k}$$

$$\Upsilon_{45i} = -E_i^T Y_i E_i + E_i^T R_{3i} E_i \tag{10l}$$

$$\Upsilon_{55i} = -Q_{2i} - E_i^T R_{1i} E_i - E_i^T R_{3i} E_i \tag{10m}$$

And matrices $L_i \in \mathbb{R}^{(n-r)\times n}$ are any full row rank and satisfying $E_i^T L_i = 0$, for each $i \in I_N$, $d_{12} = d_M - d_m$, and $\tilde{d} = d_{12} + 1$.

Proof The proof is divided into two parts. The first one treats the regularity and causality, and the second one analyzes the asymptotic stability.

First**,** let us show that system (7) is regular and causal. Since (8) holds, it follows that:

$$\Omega_i = \begin{bmatrix} \Omega_{1i} & \Omega_{2i} \\ * & \Omega_{3i} \end{bmatrix} < 0 \tag{11}$$

where

$$\Omega_{1i} = -E_i^T P_i E_i + S_i L_i^T A_i + A_i^T L_i S_i^T + F_{0i} A_i + A_i^T F_{0i}^T$$

$$\Omega_{2i} = -d_M^2 E_i^T R_{1i} - d_m^2 E_i^T R_{2i} - d_{12}^2 E_i^T R_{3i} + A_i^T T_i - F_{0i} + A_i^T F_{1i}^T$$

$$\Omega_{3i} = -T_i - T_i^T - F_{1i} - F_{1i}^T$$

Let, $\Theta_i = \begin{bmatrix} I & A_i^T \end{bmatrix}$. Pre- and post-multiplying (11) by Θ_i and Θ_i^T, respectively, yield:

$$\begin{aligned} \boldsymbol{\Lambda_i} = & -E_i^T P_i E_i + S_i L_i^T A_i + A_i^T L_i S_i^T - d_M^2 A_i^T R_{1i} E_i - d_m^2 A_i^T R_{2i} E_i \\ & - d_{12}^2 A_i^T R_{3i} E_i - d_M^2 E_i^T R_{1i} A_i - d_m^2 E_i^T R_{2i} A_i - d_{12}^2 E_i^T R_{3i} A_i \end{aligned} \tag{12}$$

Since rank $(E_i) = r < n$, then there must exist two nonsingular matrices G_i and $H_i \in \mathbb{R}^{n \times m}$ such that: $G_i E_i H_i = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$. Then, L_i can be parametrized as$L_i = {G_i}^T \begin{bmatrix} 0 \\ \phi_i \end{bmatrix}$, $\phi_i \in \mathbb{R}^{(n-r) \times (n-r)}$ is any nonsingular matrix. Similarly we define:

$$G_i A_i H_i = \begin{bmatrix} A_{11i} & A_{12i} \\ A_{21i} & A_{22i} \end{bmatrix}, \quad {G_i}^{-T} P_i {G_i}^{-1} = \begin{bmatrix} P_{11i} & P_{12i} \\ {P_{12i}}^T & P_{22i} \end{bmatrix},$$

$${G_i}^{-T} R_{1i} {G_i}^{-1} = \begin{bmatrix} R_{111i} & R_{112i} \\ {R_{112i}}^T & R_{122i} \end{bmatrix},$$

$${G_i}^{-T} R_{2i} {G_i}^{-1} = \begin{bmatrix} R_{211i} & R_{212i} \\ {R_{211i}}^T & R_{222i} \end{bmatrix}, \quad {G_i}^{-T} R_{3i} {G_i}^{-1} = \begin{bmatrix} R_{311i} & R_{312i} \\ {R_{311i}}^T & R_{322i} \end{bmatrix},$$

$${H_i}^T S_i = \begin{bmatrix} S_{11i} \\ S_{21i} \end{bmatrix}, \quad \text{for each} \quad i \in I_N.$$

Pre- and post-multiplying (12) by ${H_i}^T$ and H_i, we can obtain the following result:

$${H_i}^T \Lambda_{1i} H_i = \begin{bmatrix} * & * \\ * & \Lambda_{122i} \end{bmatrix} < 0 \tag{13}$$

where (*) represents matrices that are not relevant in the following discussions, and Λ_{122i} is given by the following expression:

$$\Lambda_{122i} = A_{22i}^T \phi_i S_{21i}^T + S_{21i} \phi_i^T A_{22i}.$$

From (13), it is easy to see that:

$$A_{22i}^T \phi_i S_{21i}^T + S_{21i} \phi_i^T A_{22i} < 0 \tag{14}$$

And thus A_{22i} is nonsingular, which implies that the pair (E_i, A_i) is regular and causal. Hence, according to Definition 2.1 and Definition 2.2, system (7) is regular and causal.

Next, we prove the asymptotic stability of system (7). Choose the Lyapunov-Krasovkii functional candidate for system (7) as:

$$V_i\,(x(k)) = \sum_{j=1}^{7} V_{j\sigma(k)}\,(x(k)) \tag{15}$$

where

$$\boldsymbol{V}_{1\boldsymbol{i}}\left(\boldsymbol{x}\left(\boldsymbol{k}\right)\right)=x^{T}(k)E_{i}^{T}\mathrm{P}_{i}E_{i}x(k)$$

$$\boldsymbol{V}_{2\boldsymbol{i}}\left(\boldsymbol{x}\left(\boldsymbol{k}\right)\right)=\sum_{s=k-d(k)}^{k-1}x^{T}(s)\mathrm{Q}_{1i}x(s)+\sum_{\theta=-d_M+1}^{-d_m}\sum_{s=k+\theta}^{k-1}x^{T}(s)Q_{1i}x(s)$$

$$\boldsymbol{V}_{3\boldsymbol{i}}\left(\boldsymbol{x}\left(\boldsymbol{k}\right)\right)=\sum_{s=k-d_M}^{k-1}x^{T}(s)\mathrm{Q}_{2i}x(s)$$

$$\boldsymbol{V}_{4\boldsymbol{i}}\left(\boldsymbol{x}\left(\boldsymbol{k}\right)\right)=\sum_{s=k-d_m}^{k-1}x^{T}(s)\mathrm{Q}_{3i}x(s)$$

$$\boldsymbol{V}_{5\boldsymbol{i}}\left(\boldsymbol{x}\left(\boldsymbol{k}\right)\right)=d_M\sum_{\theta=-d_M}^{-1}\sum_{s=k+\theta}^{k-1}\eta^{T}(s)E_{i}^{T}\mathrm{R}_{1i}E_{i}\eta(s)$$

$$\boldsymbol{V}_{6\boldsymbol{i}}\left(\boldsymbol{x}\left(\boldsymbol{k}\right)\right)=d_m\sum_{\theta=-d_m}^{-1}\sum_{s=k+\theta}^{k-1}\eta^{T}(s)E_{i}^{T}\mathrm{R}_{2i}E_{i}\eta(s)$$

$$\boldsymbol{V}_{7\boldsymbol{i}}\left(\boldsymbol{x}\left(\boldsymbol{k}\right)\right)=d_{12}\sum_{\theta=-d_M}^{-d_m+1}\sum_{s=k+\theta}^{k-1}\eta^{T}(s)E_{i}^{T}\mathrm{R}_{3i}E_{i}\eta(s)$$

where $\eta(j)=x(j+1)-x(j)$ and $d_{12}=d_M-d_m$.

$$\xi(k)=\left[x^{T}(k)(E_{i}x\left(k+1\right))^{T}x^{T}\left(k-\right)x^{T}\left(k-d(k)\right)x^{T}\left(k-d_M\right)\right]^{T}.$$

Then the difference of $V_i(x(k))$ along the solution of (7) is given by:

$$\begin{aligned}\varDelta V_i\left(x(k)=\varDelta V_{1i}\left(x(k)\right)+\varDelta V_{2i}(\,x(k)+\varDelta V_{3i}\left(x(k)\right)+\varDelta V_{4i}\left(x(k)\right)\\+\varDelta V_{5i}\left(x(k)\right)+\varDelta V_{6i}\left(x(k)\right)+\varDelta V_{7i}\left(x(k)\right).\end{aligned}$$

where

$$\varDelta V_{1i}\left(x(k)\right)=x^{T}\left(k+1\right)E_{i}^{T}P_{i}E_{i}x\left(k+1\right)-x^{T}(k)E_{i}^{T}P_{i}E_{i}x(k)\tag{16}$$

also,

$$\Delta V_{2i}\left(x(k)\right)=\sum_{s=k+1-d(k+1)}^{k}x^{T}(s)Q_{1i}x(s)-\sum_{s=k-d(k)}^{k-1}x^{T}(s)Q_{1i}x(s)$$
$$+\sum_{\theta=d_M+1}^{d_m}\left[\sum_{s=k+1-\theta}^{k}x^{T}(s)Q_{1i}x(s)-\sum_{s=k+\theta}^{k-1}x^{T}(s)Q_{1i}x(s)\right]$$

$$\Delta V_{2i}\left(x(k)\right)\leq\tilde{d}x^{T}(k)Q_{1i}x(k)-x^{T}\left(k-d(k)\right)Q_{1i}x\left(k-d(k)\right)\tag{17}$$

where $\tilde{d}=d_M-d_m+1$.

$$\Delta V_{3i}\left(x(k)\right)=\sum_{s=k+1-d_M}^{k}x^{T}(s)Q_{2i}x(s)-\sum_{s=k-d_M}^{k-1}x^{T}(s)Q_{2i}x(s)$$

$$\Delta V_{3i}\left(x(k)\right)=x^{\mathrm{T}}(k)Q_{2i}x(k)-x^{T}\left(k-d_{\mathrm{M}}\right)Q_{2i}x\left(k-d_{\mathrm{M}}\right)\tag{18}$$

$$\Delta V_{4i}\left(x(k)\right)=\sum_{s=k+1-d_m}^{k}x^{T}(s)Q_{3i}x(s)-\sum_{s=k-d_m}^{k-1}x^{T}(s)Q_{3i}x(s)$$

$$\Delta V_{4i}\left(x(k)\right)=x^{T}(k)Q_{3i}x(k)-\left(k-d_m\right)Q_{3i}x\left(k-d_m\right)\tag{19}$$

$$\Delta V_{5i}\left(x(k)\right)=d_M^2\eta^{T}(k)E_i^{T}R_{1i}E_i\eta(k)-d_M\sum_{s=k-d_M}^{k-1}\eta^{T}(s)E_i^{T}R_{1i}E_i\eta(s)$$

since,

$$d_M^2\eta^{T}(k)E_i^{T}R_{1i}E_i\eta(k)=\xi^{T}(k)\Theta_{1i}\,\xi(k)\tag{20}$$

$$\Theta_{1i}=\begin{bmatrix} d_M^2E_i^{T}R_{1i}E_i & -d_M^2E_i^{T}R_{1i} & 0\;0\;0\\ * & d_M^2R_{1i} & 0\;0\;0\\ * & * & 0\;0\;0\\ * & * & *\;0\;0\\ * & * & *\;*\;0 \end{bmatrix}$$

By Jensen lemma, we have:

$$-d_M \sum_{s=k-d_M}^{k-1} \eta^T(s) E_i^T R_{1i} E_i \eta(s) \leq \begin{bmatrix} x(k) \\ x(k-d_M) \end{bmatrix}^T \begin{bmatrix} -E_i^T R_{1i} E_i & E_i^T R_{1i} E_i \\ * & -E_i^T R_{1i} E_i \end{bmatrix} \begin{bmatrix} x(k) \\ x(k-d_M) \end{bmatrix}$$

$$-d_M \sum_{s=k-d_M}^{k-1} x^T(s) E_i^T R_{1i} E_i x(s) \leq \xi^T(k) \Theta_{2i}\, \xi(k) \tag{21}$$

where

$$\Theta_{2i} = \begin{bmatrix} -E_i^T R_{1i} E_i & 000 & E_i^T R_{1i} E_i \\ * & 000 & 0 \\ * & *00 & 0 \\ * & **0 & 0 \\ * & *** & -E_i^T R_{1i} E_i \end{bmatrix}$$

From (20) and (21), we get:

$$\Delta V_{5i}(x(k)) \leq \xi^T(k) \Theta_{3i}\, \xi(k) \tag{22}$$

where

$$\Theta_{3i} = \begin{bmatrix} \left(d_M^2 - 1\right) E_i^T R_{1i} E_i & -d_M^2 E_i^T R_{1i} & 0 & 0 & E_i^T R_{1i} E_i \\ * & d_M^2 R_{1i} & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & -E_i^T R_{1i} E_i \end{bmatrix}$$

The same reasoning is followed for $\Delta V_{6i}(x(k))$, and then we have:

$$\Delta V_{6i}(x(k)) \leq \xi^T(k) \Theta_{4i}\, \xi(k) \tag{23}$$

$$\Theta_{4i} = \begin{bmatrix} \left(d_m^2 - 1\right) E_i^T R_{2i} E_i & -d_m^2 E_i^T R_{2i} & E_i^T R_{2i} E_i & 0\,0 \\ * & d_m^2 R_{2i} & 0 & 0\,0 \\ * & * & -E_i^T R_{2i} E_i & 0\,0 \\ * & * & * & 0\,0 \\ * & * & * & *\,0 \end{bmatrix}$$

$$\varDelta V_{7i}\left(x(k)\right) = d_{12}^2 \eta^T(k) E_i^T R_{3i} E_i \eta(k) - d_{12} \sum_{s=k-d_M}^{k-d_m-1} x^T(s) E_i^T R_{3i} E_i x(s)$$

$$d_{12}^2 \eta^T(k) E_i^T R_{3i} E_i \eta(k) = \xi^T(k) \begin{bmatrix} d_{12}^2 E_i^T R_{3i} E_i & -d_{12}^2 E_i^T R_{3i} & 0 & 0 & 0 \\ * & d_{12}^2 R_{3i} & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix} \xi(k) \tag{24}$$

(9) holds implies $\begin{bmatrix} R_{3i} & Y_i \\ * & R_{3i} \end{bmatrix} \geq 0$, then applying Lemma 1, we get:

$$-d_{12} \sum_{s=k-d_M}^{k-d_m-1} x^T(s) E_i^T R_{3i} E_i x(s) \leq \begin{bmatrix} x\left(k-d_m\right) \\ x\left(k-d(k)\right) \\ x\left(k-d_M\right) \end{bmatrix}^T \varGamma_i \begin{bmatrix} x\left(k-d_m\right) \\ x\left(k-d(k)\right) \\ x\left(k-d_M\right) \end{bmatrix} \tag{25}$$

where Γ_i is given by:

$$\varGamma_i = \begin{bmatrix} -E_i^T R_{3i} E_i & E_i^T R_{3i} E_i - E_i^T Y_i E_i & E_i^T Y_i E_i \\ * & -2E_i^T R_{3i} E_i + E_i^T Y_i E_i + E_i {Y_i}^T E_i^T & -E_i^T Y_i E_i + E_i^T R_{3i} E_i \\ * & * & -E_i^T R_{3i} E_i \end{bmatrix}$$

From (24) and (25) we have:

$$\varDelta V_{6i}\left(x(k)\right) \leq \xi^T(k) \begin{bmatrix} d_{12}^2 E_i^T R_{3i} E_i & -d_{12}^2 E_i^T R_{3i} & 0 \\ * & d_{12}^2 R_{3i} & 0_{(1\times 3)} \\ * & * & \varGamma_i \end{bmatrix} \xi(k) \tag{26}$$

where Γ_i is given by the above result and $d_{12} = d_M - d_m$.

Using the free weighting matrix method, notice that:

$$0 = 2x^T\left(k+1\right) E_i^T {T_i}^T \left[-E_i x\left(k+1\right) + A_i x(k) + Ad_i x\left(k-d(k)\right)\right],$$

is equivalent at:

$$0=\xi^T(k)\begin{bmatrix} 0 & A_i^T T_i & 0 & 0 & 0 \\ * & -T_i-T_i{}^T & 0 & T_i{}^T Ad_i & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix}\xi(k) \tag{27}$$

On the other hand, it is clear from $E_i^T L_i = 0$ that:

$$\begin{aligned} 0 &= 2x^T(k)S_i L_i^T E_i x(k+1) \\ &= 2x^T(k)S_i L_i^T \left[A_i x(k) + Ad_i x(k-d(k))\right]' \end{aligned}$$

$$0=\xi^T(k)\begin{bmatrix} S_i L_i^T A_i + A_i{}^T S_i L_i^T & 0 & 0 & S_i L_i^T Ad_i & 0 \\ * & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix}\xi(k) \tag{28}$$

where matrices $L_i \in \mathbb{R}^{(n-r)\times n}$ are any full row rank and satisfying $E_i^T L_i = 0$, for each $i \in I_N$.

From Eqs. (16)–(19), (22), (23), (26)–(28), we conclude that for any arbitrary switching law $\sigma(k) = i$, we have:

$$\Delta V_i\left(x(k)\right) \le \xi^T(k)\Psi_i\,\xi(k) \tag{29}$$

$$\boldsymbol{\Psi}_i = \left[\Psi_{jki}\right] \quad \text{and} \quad j,k = 1,2,3,4,5, \quad \text{and} \quad i \in I_N$$

$$\begin{aligned}\boldsymbol{\Psi}_{11\boldsymbol{i}} = {} & -E_i{}^T P_i E_i + \tilde{d}Q_{1i} + Q_{2i} + Q_{3i} + \left(d_M^2 - 1\right)E_i{}^T R_{1i}E_i + \left(d_m^2 - 1\right) \\ & E_i{}^T R_{2i}E_i + d_{12}^2 E_i{}^T R_{3i}E_i + S_i L_i{}^T A_i + A_i{}^T L_i S_i{}^T\end{aligned}$$

$$\boldsymbol{\Psi}_{12\boldsymbol{i}} = -d_M^2 E_i^T R_{1i} - d_m^2 E_i^T R_{2i} - d_{12}^2 E_i^T R_{3i} + A_i^T T_i, \quad \boldsymbol{\Psi}_{13\boldsymbol{i}} = E_i^T R_{2i}E_i,$$

$$\boldsymbol{\Psi}_{14\boldsymbol{i}} = S_i L_i^T Ad_i \quad \boldsymbol{\Psi}_{15\boldsymbol{i}} = E_i^T R_{1i}E_i,$$

$$\boldsymbol{\Psi}_{22\boldsymbol{i}} = P_i + d_M^2 R_{1i} + d_m^2 R_{2i} + d_{12}^2 R_{3i} - T_i - {T_i}^T, \quad \boldsymbol{\Psi}_{24\boldsymbol{i}} = {T_i}^T Ad_i,$$

$$\boldsymbol{\Psi}_{33i} = -Q_{3i} - {E_i}^T R_{2i} E_i - {E_i}^T R_{3i} E_i, \quad \boldsymbol{\Psi}_{34\boldsymbol{i}} = {E_i}^T R_{3i} E_i - {E_i}^T Y_i E_i,$$
$$\boldsymbol{\Psi}_{35\boldsymbol{i}} = {E_i}^T Y_i E_i,$$

$$\boldsymbol{\Psi}_{44\boldsymbol{i}} = -Q_{1i} - 2{E_i}^T R_{3i} E_i + {E_i}^T Y_i E_i + E_i {Y_i}^T {E_i}^T,$$
$$\boldsymbol{\Psi}_{45\boldsymbol{i}} = -{E_i}^T Y_i E_i + {E_i}^T R_{3i} E_i,$$

$$\boldsymbol{\Psi}_{55\boldsymbol{i}} = -Q_{2i} - {E_i}^T R_{1i} E_i - {E_i}^T R_{3i} E_i$$

Let $B_i = [A_i - I_n \; 0 \; Ad_i \; 0]$, and ${F_i}^T = \left[F_{0i}^T \, F_{1i}^T \; 0 \; F_{2i}^T \; 0\right]$. Then we can verify that $B_i.\, \xi(k) = 0$. The matrices Υ_i in (8) can be written as:

$$\Upsilon_i = \Psi_i + F_i B_i + {B_i}^T {F_i}^T < 0 \tag{30}$$

Applying Lemma 2, we have: $\xi^T(k)\Psi_i\, \xi(k) < 0$, which implies that $\Delta V_i(x(k)) < 0$. Thus, system (7) is asymptotically stable. This completes the proof.

Remark 3 In this remark, we give a comparison between our results and some existing results in the literature.

1. *Singular linear system with time-delay*:

   If the switched system has one subsystem ($N = 1$), the results developed in Theorem 1 reduce to the delay-dependent stability condition for the following singular time-delay system, $Ex(k+1) = Ax(k) + Adx(k - d(k))$, which is studied in several papers. If the slacks variables are not introduced from Theorem 1 of this chapter, this theorem reduces to that in [26]. Thus we can improve the results of the approach developed in [26] by introducing some slacks variables which give more relaxation for system. So, our result represents a general case.

2. *Switched time-delay system:*

   When $E_i = I$, the switched system (7) becomes a standard switched time-delay system that investigated in [7, 8] and other many results in the literature, which means that the system of our study is more general.

Remark 4 The regularity and causality of system (7) ensure that the solution to this system exist and is unique and causal, $\forall \sigma(k) = i \in I_N$.

3.2 Robust Admissibility of Uncertain Switched Singular System with Time-Varying Delay

In this section, we extend Theorem 1 in the previous section to obtain the corresponding results for uncertain switched singular systems (2) for any $\sigma(k) = i$, $\forall i \in I_N$.

Theorem 2 System (2) is robustly admissible for arbitrary switching signal, if there exist positive definite matrices P_i, Q_{ji}, R_{ji} ($j = 1, 2, 3$), real matrices T_i and S_i with appropriate dimensions, matrices F_{0i}, F_{1i}, and F_{2i}, and constant scalars $\epsilon_i > 0$ such that the condition (9) and the following LMIs hold for each $i \in I_N$:

$$\overline{\Upsilon_i} = \begin{bmatrix} \Upsilon_i & U_i & \varepsilon_i.{W_i}^T \\ * & -\varepsilon_i.I_2 & \varepsilon_i.{J_i}^T \\ * & * & -\varepsilon_i.I_2 \end{bmatrix} \tag{31}$$

where (*) denotes the symmetric part in a symmetric matrix and, Υ_i is defined in (8) for each $i \in I_N$.

Proof Replacing A_i and Ad_i by $A_i + M_i\Delta_i(k)N_{1i}$ and $Ad_i + M_i\Delta_i(k)N_{2i}$, respectively in (8), and putting:

$${U_i}^T = {M_i}^T \begin{bmatrix} L_i S_i^T & F_{1i}^T & 0 & F_{2i}^T & 0 \\ F_{01}^T & T_i & 0 & 0 & 0 \end{bmatrix}, \quad W_i = \begin{bmatrix} N_{1i} & 0 & 0 & N_{2i} & 0 \\ N_{1i} & 0 & 0 & N_{2i} & 0 \end{bmatrix},$$

We get,

$$\overline{\Upsilon_i} = \Upsilon_i + U_i \Delta_i(k) W_i + {W_i}^T \Delta_i^T(k) {U_i}^T < 0 \tag{32}$$

Then by Lemma 3, there exist scalars $\varepsilon_i > 0$ for each $i \in I_N$, such that:

$$\begin{bmatrix} \Upsilon_i & \mathrm{U}_i & \varepsilon_i.{\mathrm{W}_i}^{\mathrm{T}} \\ * & -\varepsilon_i.I_2 & \varepsilon_i.{J_i}^T \\ * & * & -\varepsilon_i.I_2 \end{bmatrix} < 0 \tag{33}$$

Thus, the result follows from Theorem 1. This completes this proof.

Remark 6 Theorem 2 presents sufficient conditions for robust admissibility of uncertain switched singular system with time-varying delay (2).

Table 1 The allowable upper bound d_M for different values of d_m

d_m	2	4	6	8	10
d_M in Th. 1 of [19]	5	6	7	9	10
d_M in Th. 1 of [12]	5	6	7	9	10
d_M in Th. 1 of [44]	11	11	12	12	13
d_M in Th. 1 of [8]	13	14	14	15	16
d_M in Th. 1 of this chapter	13	14	15	16	16

"Th." is the abbreviation of theorem

4 Numerical Examples

Example 1[8] (Asymptotic stability of switched system with TVD).
Consider system (7) with $N = 2$ and the following parameters:

$$E_1 = E_2 = I$$

$$A_1 = \begin{bmatrix} 0.7 & 0 \\ 0.08 & 0.95 \end{bmatrix}, \quad Ad_1 = \begin{bmatrix} 0.15 & 0 \\ -0.1 & -0.1 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 0.7 & 0 \\ 0.08 & 0.9 \end{bmatrix}, \quad Ad_2 = \begin{bmatrix} 0.14 & 0 \\ -0.1 & -0.05 \end{bmatrix}$$

Applying Theorem 1 with $L_1 = L_2 = [0; 0]$, the allowable delay upper bound d_M that guarantees the asymptotic stability, for different values of the minimum delay d_m are provided in Table 1. This shows that our result is less conservative than that in the selected literature.

Then, for an arbitrary switching law shown in Fig. 1, the time-varying delay $d(k)$ is given by: $1 \leq d(k) \leq 3$. The simulation results of state vector of the **switched system with time-varying delay** with the initial condition $x(0) = [0.8 - 0.7]^T$ are realized by Matlab Toolbox R2008a and are shown in Fig. 2. From this figure, we can see that the switched system is asymptotically stable.

Example 2[40] *(Asymptotic admissibility of SSTD system).*
Consider system (7) with $N = 2$ and the following parameters:

$$E_1 = E_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

$$A_1 = \begin{bmatrix} 0.9 & 0 \\ 0 & 0.7 \end{bmatrix}, \quad Ad_1 = \begin{bmatrix} -0.1 & 0.1 \\ -0.1 & -0.1 \end{bmatrix},$$

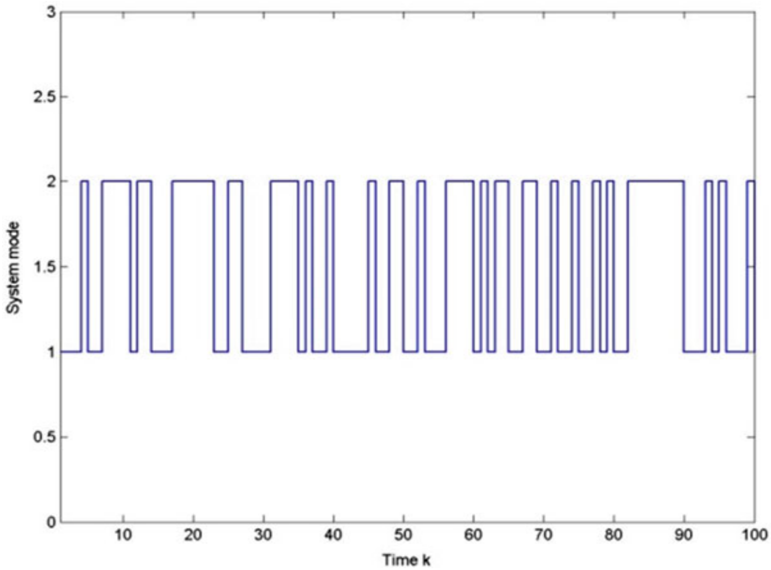


Fig. 1 The switching signal

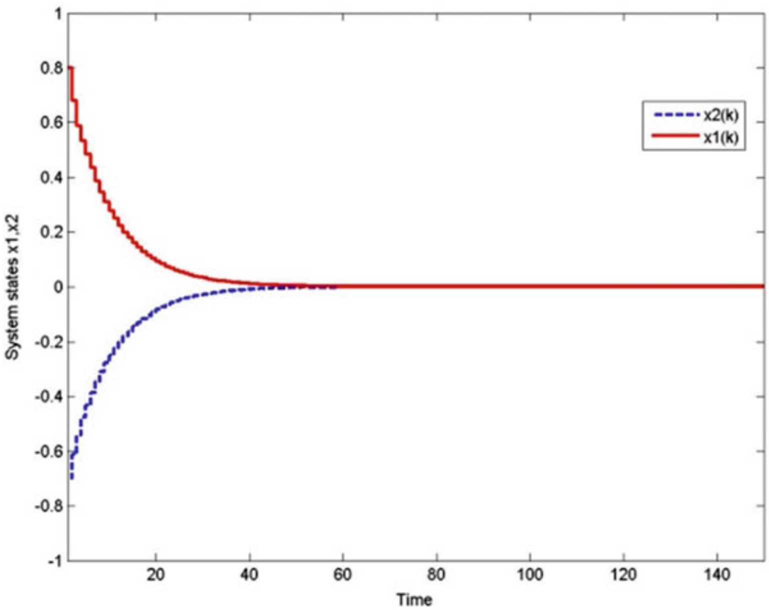


Fig. 2 State trajectories of switched system

$$A_2 = \begin{bmatrix} 0.8 & 0.1 \\ 0 & 0.8 \end{bmatrix}, \quad Ad_2 = \begin{bmatrix} -0.2 & -0.1 \\ -0.2 & -0.1 \end{bmatrix}.$$

Table 2 Maximum bound d_M for different values of d_m

d_m	1	3	5	7	9	10
d_M in Th. 1 of [39]	12	13	Inf	Inf	Inf	Inf
d_M in Lem. 1 of [38]	10	11	11	11	12	13
d_M in Th. 3.1 ($\lambda = 1.1$) of [43]	6	6	6	6	7	8
d_M in Th. 3.1 ($\lambda = 1.2$) of [43]	5	6	7	8	9	9
d_M in Th. 1 of this chapter	10	11	12	12	13	14

"**Lem**" design Lemma and "**inf**" design infeasible

Applying Theorem 1 with $L_1 = L_2 = [0\ 1]^T$, the allowable delay upper bound d_M that guarantees the asymptotic admissibility of system (7), for different values of the minimum delay d_m is provided in Table 2. For comparison, we reported in the same table the results obtained by the approaches of [38, 39, 43]. From (2), it is obvious that of Theorem 1 gives larger delay bounds $d_M = \{10, 11, 12, 12, 13, 14\}$. This means that the proposed method in this chapter is less conservative than that in the selected literature.

Example 3[39] *(Robust admissibility).*

Consider the robust admissibility problem of system (2) with $J_1 = J_2 = 0.1$ and two modes, i.e., $N = \{1, 2\}$, and the following parameters:

$$E_1 = E_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

$$A_1 = \begin{bmatrix} 0.4 & 0.9 \\ -1 & -2 \end{bmatrix}, \quad Ad_1 = \begin{bmatrix} 0.05 & -0.1 \\ -0.05 & -0.1 \end{bmatrix}, \quad M_1 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 1.6 & 1 \\ -2 & -2.5 \end{bmatrix}, \quad Ad_2 = \begin{bmatrix} 0.1 & -0.05 \\ -0.05 & -0.1 \end{bmatrix}, \quad M_2 = \begin{bmatrix} -0.01 \\ -0.02 \end{bmatrix},$$

$$N_{11} = N_{12} = N_{21} = N_{22} = \begin{bmatrix} 0.02\ 0.01 \end{bmatrix},$$

In this example, we assume $d_m = 1$, choose also $L_1 = L_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $J_1 = 0.05$, $J_2 = 0.1$. By Theorem 2, the system (2) is admissible for $d_M = 10$, but in Theorem 2 of [39] the same system is admissible for $d_M = 2$. Moreover, if we set $J_1 = J_2 = 0$, then, by Theorem 2, the system (2) is admissible for $d_M = 9$, and the admissibility of this systems by Theorem 2 of [39] is assured for $d_M = 4$. It can be seen that our method is less conservative.

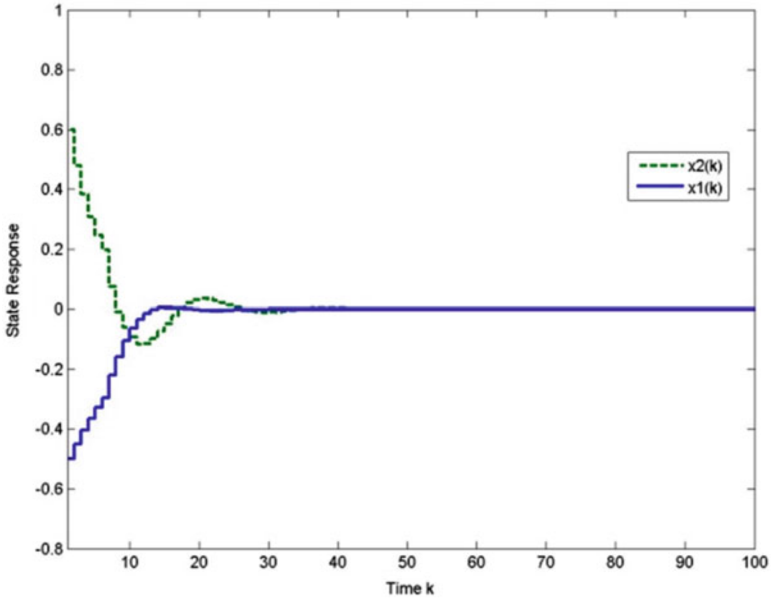


Fig. 3 State trajectories of switched singular system with time-varying delay

For a switching signal shown in Fig. 1, the delay bounds by $d_m = 1,\ d_M = 4$, and the initial condition is $(0) = [-0.6 \quad 0.8]^T$. The simulation results of state vectors of the **switched singular system with time-varying delay** for Example 3 can be realized by MATLAB Simulink and are shown in Fig. 2. It is observed that both of the trajectories converge to the origin and then we can conclude that the system (7) is regular, causal, and asymptotically stable (Fig. 3).

5 Conclusion

In this chapter, we have studied the problems of robust admissibility analysis for uncertain switched singular systems with time-varying delay for any arbitrary switching signal. Sufficient conditions for nominal switched singular delay systems to be admissible are given in terms of LMIs. The established conditions in this chapter are extended to the robust admissibility of uncertain discrete-time switched singular system with time-varying delay and linear fractional perturbations. Finally numerical examples were given to demonstrate the effectiveness of the proposed approach.

References

1. Liberzon, D. (2003). *Switched linear systems and control*. Birkhausser.
2. Lin, J., & Fan, C. (2010). Exponential admissibility and dynamic output feedback control of switched singular systems with interval time-varying delay. *Mathematical Problems in Engineering, 2010*, 1–21.
3. Sun, Z., & Ge, S. S. (2011). *Stability theory of switched dynamical systems*. Springer-Verlag.
4. Ahmida, F., & Tissir, E. H. Exponential stability of uncertain T-S fuzzy switched system with time delay. *International Journal of Automation and Computing, 10*(1), 32–38. https://doi.org/10.1007/s11633-013_0693-1
5. Ahmida, F., & Tissir, E. H. (2015). Stabilization of switched T-S fuzzy systems with additive time varying delays. In *Part of book, Lecture notes in Electrical Engineering* (pp. 401–408). Springer International. https://doi.org/10.1007/978-3-319-30301-7_42
6. Charqi, M., Tissir, E. H., Chaibi, N., & Ouahi, M. Robust stability of uncertain discrete-time switched systems with time-varying delay. In *International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Mo, April 02–04*. https://doi.org/10.1109/ISACV.2018.8354061(2018
7. Chen, J.-D., Wu, I.-T., Lien, C.-H., Lee, C.-T., Chen, R.-S., & Yu, K.-W. (2014). Robust exponential stability for uncertain discrete-time switched systems with interval time-varying delay through a switching signal. *Journal of Applied Research and Technology, 42*, 1187–1197.
8. Lien, C.-H., Yu, K.-W., Yang, J.-H., & Chou, I.-C. (2015). Novel delay-partitioning approach on stability of uncertain discrete switched time-delay systems via switching signal design. *IMA Journal of Mathematical Control and Information*, 1–15.
9. Magdi, S. M. (2010). *Switched time-delay systems: Stability and control*. Springer-Verlag.
10. Tissir, E. H. (2011). Exponential stability of switched linear systems with time-varying delays. *Applied Mathematical Sciences, 5*, 395–405.
11. Ahmida, F., & Tissir, E. H. (2015). A new approach to H_∞ filtering for switched T-S fuzzy systems with interval time varying delays and parameter uncertainties. *British Journal of Mathematics & Computer Science, 6*, 215–232.
12. Du, D., & Zhang, J. (2010). Novel robust stabilization and H_∞ control results for discrete-time switched systems with time-varying delays. In *International Conference on Computer and Communication Technologies in Agriculture Engineering* (Vol. 1, pp. 93–96).
13. Ibrir, S. (2008). Stability and robust stabilization of discrete-time switched systems with time-delays: LMI approach. *Applied Mathematics and Computation, 206*, 570–578.
14. Lien, C.-H., Yu, K.-W., Chung, Y.-J., & Chang, H.-C. (2011). Exponential stability and robust control for uncertain discrete switched systems with interval time-varying delay. *IMA Journal of Mathematical Control and Information, 7*, 433–444.
15. Lien, C.-H., Yu, K.-W., Chung, Y.-J., Chang, H.-C., & Chen, J.-D. (2011). Switching signal design for global exponential stability of uncertain switched nonlinear systems with time-varying delay. *Nonlinear Analysis: Hybrid Systems, 5*, 10–19.
16. Phat, V. N., & Ratchagit, K. (2011). Stability and stabilization of switched linear discrete-time systems with interval time-varying delay. *Nonlinear Analysis: Hybrid Systems, 5*, 605–612.
17. Rajchakit, G., Rojsiraphisal, T., & Rajchakit, M. (2012). Robust stability and stabilization of uncertain switched discrete-time systems. *Advances in Difference Equations, 1*(134), 1–15.
18. Zhang, W.-A., & Yu, L. (2009). Stability analysis for discrete-time switched time-delay systems. *Automatica, 45*, 2265–2271.
19. Zhang, L., Shi, P., & Basin, M. (2008). Robust stability and stabilization of uncertain switched linear discrete time-delay systems. *IET Control Theory and Applications, 2*(7), 606–614.
20. Chaibi, N., & Tissir, E. H. (2013). Delay dependent robust stability of singular systems with additive time varying delays. *International Journal of Automation and Computing, 10*(1), 85–90.
21. Dai, L. (1989). *Singular control systems*. Springer-Verlag.
22. Du, Z., Zhang, Q., & Liu, L. (2011). New delay-dependent robust stability of discrete singular systems with time-varying delay. *Asian Journal of Control, 13*(1), 136–147.

23. Ech-charqy, A., Ouahi, M., & Tissir, E. H. (2018). Delay-dependent robust stability criteria for singular time-delay systems by delay-partitioning approach. *International Journal of Systems Science, 32*(14), 2957–2967.
24. Ech-charqy, A., Ouahi, M., & Tissir, E. H. (2018). Robust observer-based control design for singular systems with delays in states. *Internatonal Journal of Automation and Smart Technology, 8*(3), 1–11.
25. Fang, M. (2010). Delay-dependent stability analysis for discrete singular systems with time-varying delays. *Acta Automatica Sinica, 36*(5), 751–755.
26. Jarina, B. L., & Balasubramaniam, P. (2015). Admissibility analysis for discrete-time singular systems with time-varying delays by adopting the State-space Takagi-Sugeno Fuzzy model. *Iranian Journal of Fuzzy Systems, 12*(3), 1–16.
27. Jiao, J. (2012). Robust stability and stabilization of discrete singular systems with interval time-varying delay and linear fractional uncertainty. *International Journal of Automation and Computing, 9*(1), 8–15.
28. Krirm, S., Hmamed, A., & Tadeo, F. (2015). Analysis and design of H_∞ controllers for 2D singular systems with delays. *Circuits, Systems, and Signal Processing, 35*(5), 1579–1592.
29. Wu, Z.-G., Park, J. H., Su, H., & Chu, J. (2012). Admissibility and dissipativity analysis for discrete-time singular systems with mixed time-varying delays. *Applied Mathematics and Computation, 218*, 7128–7138.
30. Xin, S., Qing-Ling, Z., Chun-Yu, Y., Youg-Yun, S., & Zhan, S. (2010). Delay dependent stability analysis and stabilization of discrete-time singular delay systems. *Acta Automatica Sinica, 36*(10), 1477–1483.
31. Xu, S., & Lam, J. (2006). *Robust control and filtering of singular systems*. Springer.
32. Zhang, X., & Zhu, H. (2013). Robust stability and stabilization criteria for discrete singular time-delay LPV systems. *Asian Journal of Control, 15*(3), 1–11.
33. Boukas, E. K. (2008). *Control of singular systems with random abrupt changes*. Springer.
34. Ma, S., Zhang, C., & Wu, Z. (2008). Delay-dependent stability and H_∞ control for uncertain discrete switched singular systems with time-delay. *Applied Mathematics and Computation, 206*(1), 413–424.
35. Zamani, I., & Shafiee, M. (2013). Stability analysis of uncertain switched singular time-delay systems with discrete and distributed delays. *Optimal Control Applications and Methods, 2013*, 1–28.
36. Ding, X., Liu, X., & Zhong, S. (2014). Delay-independent criteria for exponential admissibility of switched descriptor delayed systems. *Applied Mathematics and Computation, 228*, 432–445.
37. Jin-Xing, L., & Shu-Min, F. (2010). Robust exponential admissibility of uncertain switched singular time-delay systems. *Acta Automatica Sinica, 36*(12), 1773–1779.
38. Lin, J., Fei, S., & Shen, J. (2010). Robust stability and stabilization for uncertain discrete-time switched singular systems with time-varying delays. *Journal of Systems Engineering and Electronics, 21*(4), 650–657.
39. Lin, J., Fei, S., & Gao, Z. (2012). Stabilization of discrete-time switched singular time-delay systems under asynchronous switching. *Journal of the Franklin Institute, 349*, 1808–1827.
40. Lin, J., Gao, Z., Zhou, Y., & Gai, S. (2012). Exponential admissibility of discrete-time switched singular time-delay systems: an average dwell time approach. In *Proceedings of 24th Chinese control and decision conference (CCDC)* (pp. 1805–1810).
41. Lin, J., Fei, S., Gao, Z., & Ding, J. (2012). Fault detection for discrete-time switched singular time-delay systems: an average dwell time approach. *International Journal of Adaptive Control and Signal Processing, 2012*, 1–28.
42. Lijun, Z., Jiemei, Z., Xue, Q., & Fanbiao, L. (2011). Exponential stability for discrete-time singular switched time-delay systems with average dwell time. In *Proceedings of the 30th Chinese control conference, Yantai, China, July 22–24* (pp. 1789–1794).
43. Zhang, L., & Zhao, J. (2013). Delay-dependent exponential stability for discrete-time singular switched systems with time-varying delay. *Asian Journal of Control, 15*(2), 630–635.

44. Li, Z., Gao, H., & Karimi, H. R. (2014). Stability analysis and H_∞ controller synthesis of discrete-time switched systems with time delay. *Systems and Control Letters, 66*, 85–93.
45. Sun, Z., & Shuzhi, S. G. (2005). *Switched linear systems: Control and design*. Springer-Verlag.
46. Zhai, G., & Xu, X. (2010). A unified approach to stability analysis of switched linear descriptor systems under arbitrary switching. *International Journal of Applied Mathematics and Computer Science, 20*(2), 249–259.

Fuzzy MAS for Power Prediction Based on the Markov Chains: Modeling and Simulation in a HEV

Rachid El Amrani, Ali Yahyaouy, and Hamid Tairi

Abstract Electric and hybrid vehicles have obtained an important place in the field of sustainable transport. Thanks to their advantages of structure and operation, they are the most present and studied by the research works. Energy management, in this sort of vehicle, is a vital tool for guaranteeing a good operation during the driven cycles by controlling available energy sources and likewise forecasting future energy demand. In this chapter, we propose a method based on the Markov chains to optimize and predict the energy, which necessitates the construction of a proper system to control the modeled electric traction chain. The vehicle in question has a hybrid-electric source and an electric motor, so it is advantageous to design a fast and efficient management system capable of processing the available components online. So, we use an MAS to manage the energy exchange in the primary and secondary source (fuel cell and supercapacitor). In the simulation step, we are looking to apply an online forecast and see the reaction of the system as far as improved power. Finally, we compare the results found by our proposed method with other previous tests of the study case and discuss what makes it possible to determine the best energy management method (s) appropriate for an HEV.

Keywords Hybrid electric vehicle · Energy management · Optimization · Prediction · Markov chain · Fuzzy control · Multiagent system

1 Introduction

The technological innovation in the new generation of vehicles has achieved an interesting development, in the use of unconventional and non-GHG (*greenhouse gas*) engines, electric vehicles, and vehicles with hybrid electric sources (*HES*) that have become one of the relevant and sustainable solutions.

R. El Amrani (✉) · A. Yahyaouy · H. Tairi
LISAC Laboratory, Department of Computer Sciences, Dhar Mehraz College of Sciences, University of Sidi Mohamed Ben Abdellah (U.S.M.B.A), Fez, Morocco
e-mail: rachid.elamrani@usmba.ac.ma; ali.yahyaouy@usmba.ac.ma; hamid.tairi@usmba.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_65

The International Organization for Standardization is laying the groundwork for these technologies by providing ISO standards (electrically propelled vehicles: ISO/TC 22/SC 37 and hydrogen technologies: ISO/TC 197) and useful tools for supporting its development, namely ISO 8714 for measuring the baseline energy consumption of electric road vehicles and ISO 23274 for measuring fuel consumption of hybrid vehicles and exhaust emissions, see [1]. Also, the ISO 6469 series of standards establishes the specifications of safety for rechargeable energy storage systems for electric vehicles [1].

Particularly concerning hybrid vehicles and the reason why it is made, the subject of this chapter, is the most present and studied by scientific researches, as a result of their advantages in terms of structure and operation.

At this type of vehicle, energy management is a key tool to ensure smooth operation during rolling cycles while optimizing the available energy sources (fuel cell and supercapacity) and even predicting future demand for energy.

The optimization and energy prediction involves developing a suitable system to control the electric drivetrain. As we know the vehicle concerned has a hybrid source and the treatment is online, it is interesting to define a flexible, fast management system and capable of handling available subcomponents.

This approach has been validated in a system implemented in Java. We, therefore, used a computer model of processing based on the distributed approach (*MAS*). First, we studied the main management strategies applied to HEVs, and then we developed the energy management model. The work consists of developing algorithms for optimizing and predicting vehicle energy consumption. These algorithms are based on fuzzy logic and the Markov chain.

To perform real-time simulations on the proposed multiagent architecture, the last phase is the implementation of management laws on the built model, to assess the performance of the adopted strategy.

Speed profiles are simulated. They correspond to the current *NEDC* cycle which is updated in 2017 by the new *WLTP* cycle (or Worldwide harmonized Light vehicles Test Procedures) [2] for all new models of vehicle.

The found results have shown that the combination of methods is efficient, and the integration of the Markov process model prediction algorithm showed a significant improvement of system performance and a fairly accurate prediction of the power Pfc to be provided by the fuel cell. This approach is characterized by an architecture easy to implement and a reduced calculation time, which facilitates its use in real time.

2 Strategies and AI

2.1 Existing Management Strategies

Various strategies of energy management have been proposed to reduce the consumption of the main source and control the sharing power between these sources of energy and the load for known routes.

In different research communities, we find several online and offline methods. These strategies were classified in [3, 4]. A new strategy was recently proposed in previous works [5, 6]. So, we have in [5] three main strategies: rule-based strategies, optimization-based strategies, and hybrid strategy. These methods are developed for the sources of hybrid storage [7], and also applicable for hybrid electric vehicle (*HEV)* and all-electric vehicles (*AEV)*.

For the strategy based on rules, it belongs to the field of artificial intelligence, A detailed description has been made in [5]. The rule-based strategy was applied in [6] for intelligent energy management of an AEV. For the strategy based on optimization, it belongs generally to the field of automatic control. It was already well defined in [5]. The optimization-based strategy was also tested in combination with the first method.

2.2 Online Optimization Methods: Advantages and Disadvantages

In our case, we are interested in the online methods; given the online optimization requires a multiagent system based on a method of optimization and online prediction.

The strategies used in real time are classified in [8] in two families, optimization and prediction. The first is based in particular on dynamic programming (*DP*), optimal control, and the combination of it with algorithms such as genetic algorithms, the Pontryagin's minimum principle, and particles swarm.

Other optimization methods based on prediction have been suggested in the literature [8]; namely stochastic dynamic programming (*SDP*), auto-optimization, Pareto optimum, predictive command, and Markov chain. There is also a combination of these methods with deterministic or fuzzy rules.

The main weak point of the recently referenced methods is, during the treatment, the management process is divided into several sequential tasks, due to the combination between two or more different methods (inline and/or offline) applied to manage, optimize, or predict energy in EV studied. This leads to a long-running time on the one hand and the question of system consistency on the other hand.

To avoid these inconveniences, a new strategy has been proposed by El Amrani et al. [5, 6], it is a hybrid strategy based on multiagent systems, combining adaptive fuzzy rule-based methods and evolutionary optimization methods. The evolutionary optimization method based on a genetic algorithm is executed at the beginning only once to obtain optimized membership functions; as a result, the fuzzy controller is launched directly to manage the power between the source and the demand through the multiagent system which links each component of the chain of traction. This method has given good results and the adopted architecture works in all simplicity.

In [8], Himi et al. summarize the advantages and disadvantages of existing offline and real-time strategies. The strategy used in this work should consider the advantages and disadvantages of each of the strategies according to the results of existing research. Therefore, the major challenge remains to develop an online system that optimizes and predicts the power demanded by the EV throughout a traction cycle. In the next part of this work, we introduce our method as a solution that overcomes the inconvenience thus mentioned. The fuzzy rules-based management strategy in combination with the Markov chains has been chosen in this chapter and will be tested on two different cases.

2.3 Proposed Approach: Fuzzy MAS Based on Markov Chains

In the hybrid strategy that we have adopted, optimization is multiobjective because it combines several objectives to make the system more efficient. Therefore, two objectives have been set. The first is to minimize the hydrogen consumption of the main energy source and to improve the life of the secondary source. The second is to maintain the state of charge of the bounded secondary source. In this case, the road course, the current and future driving conditions are assumed to be unknown.

As previously mentioned, the proposed method combines three strategies to optimize the energy consumption of the fuel cell vehicle; this approach includes:

– A real-time method: It is based on a multiagent system, and it distributes the management process and makes local control.
– A local optimization strategy applied in line: It is based on the fuzzy logic of order 1, and it ensures the distribution of the power between the components of the system and minimizes, at every moment, the consumption of hydrogen.

A prediction algorithm: This one uses the Markov chains and makes it possible to predict future driving conditions and hydrogen consumption.

3 Modeling

Generally, the process of management is the same in all strategies, and these methods are flexible and can be applied to an HES in an electric vehicle, as in a hybrid vehicle. The management procedure can be divided into two main steps, as explained in detail in the sizing process of energy management by [9].

In this section, it is important to mention that a set of factors listed in [10] and quoted in [6] must be taken into account in the management procedure in the HV, namely the technology of energy storage, the autonomy in electrical energy, the hybridization rate, the autonomy fuel fossil, the topology of the vehicle drivetrain, and the mission profile. The modeling of vehicle dynamics has already been done in previous work [6], of which we have realized a simulator to manage the energy in an electric vehicle with a battery (Bat) and a supercapacitor (*SC*). The difference is that this vehicle is made of a fuel cell (*FC*) instead of a battery.

3.1 HES Model

Fuel Cell Model A fuel cell produces electrical energy from oxygen (usually taken from ambient air) and hydrogen (*H2*) stored in a fuel tank. The only products of the electrochemical reactions are, besides the electric energy produced, water and heat [11]. The embedded FC is therefore considered to be zero-emission and is used as the main source [12]. The cell is *PEM* (proton exchange membrane) type [12], and its inputs–outputs are defined in Fig. 1.

According to the static model describing the polarization curve of the PEM cell [13], the total voltage supplied by the fuel cell, neglecting the resistance of the connectors between the cells, is expressed by:

$$V_{\mathrm{pac}} = N_{\mathrm{cell}} * V_{\mathrm{cell}} \tag{1}$$

where V_{cell} is the voltage of a cell, and N_{cell} is the number of available cells.

Supercapacitor Model A supercapacitor is a very strong source of power and is the most widely used component for storing electrostatic energy, by accumulating charges in these terminals that polarize an electrolytic solution. The adopted model was well detailed in [5], and it is double electric supercapacitor, named model of Zubieta (or *RC Model*).

From [6], the stat of charge (*SOC in %)* of the supercapacitor is defined by the percentage ratio between the energy that remained in the supercapacitor at a time t and the maximum energy:

$$\mathrm{SOC} = \frac{E_{\mathrm{sc}}(t)}{E_{\mathrm{max}}} \tag{2}$$

with $E_{\mathrm{sc}}(t) = E_{\mathrm{sc\,max}} - E_{\mathrm{sc}}$ and the energy E_{sc} is proportional to the capacity and the maximum energy which corresponds to its maximum vacuum voltage V_{max}, given by:

$$E_{\mathrm{sc}} = \frac{1}{2} C.U_{\mathrm{c}} \quad \text{implies} \quad E_{\mathrm{sc\ max}} = \frac{1}{2} C.U_{\mathrm{c}}^2 \tag{3}$$

By the law of Kirchhoff, the voltage of supercapacitors is defined as:

$$U_{\mathrm{sc}} = U_{\mathrm{c}} - R_{\mathrm{sc}}.I_{\mathrm{sc}} \tag{4}$$

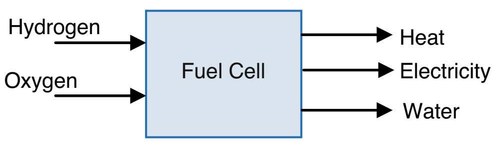


Fig. 1 The inputs–outputs of the PEMFC [12]

Consequently, the power is therefore given by:

$$P_{sc} = U_{sc}.I_{sc} = \frac{Q_{sc}}{C_{sc}}.I_{sc} - R_{sc}I_{sc}^2 \tag{5}$$

where Q_{sc} *is* quantity of charge stored in supercapacitors and is given in Columb by:

$$Q_{sc} = \int I_{sc}\, dt \tag{6}$$

3.2 Model of the Traction Chain

In this case of an embedded application integrated with EV and linked to an HES, among the factors mentioned by [10] in (2nd ¶, § 3) that must be applied is the knowledge of the vehicle dynamics. Because it is necessary to do the sizing of HES. Thus, this vehicle modeling step was realized in [5] and considered completed.

The forces applied to the vehicle are given by:

$$\overrightarrow{F}_{acc} = \overrightarrow{F}_{air} + \overrightarrow{P_g} + \overrightarrow{F}_{rol} + \overrightarrow{R}_{w} + \overrightarrow{F}_{t} \tag{7}$$

where $\overrightarrow{F}_{acc}$ is the acceleration force and $\overrightarrow{F}_t$ is the tensile force of the front and rear tires.

And the external forces of resistant are:

- $\overrightarrow{F}_{air}$: Resistance force of the air on the vehicle.
- $\overrightarrow{P}_{g}$: Resistance force of the slope (gravity).
- $\overrightarrow{F}_{rol}$: Rolling resistance (of the wheels on the ground).
- $\overrightarrow{R}_{w}$: Ground reaction on the wheels [N].

From the dynamic model of EV [6], the vehicle is, therefore, a moving mass, and its action is determined by all the forces that apply to it in one direction.

On the Y-axis, the vector of the traction force F_t is zero. In this case, we take the equations along the X-axis.

Therefore, force expressions are given as follows:

$$\begin{cases} \overrightarrow{F}_{acc} = M_v \frac{dv}{dt} \underset{x}{\rightarrow} \\ \overrightarrow{F}_{air} = -\frac{1}{2}\rho_{air} v^2 S C_x \underset{x}{\rightarrow} \\ \overrightarrow{P}_{gx} = -M_v g \ \sin\alpha \underset{x}{\rightarrow} \\ \overrightarrow{F}_{rol} = -M_v g C_r \cos\alpha \underset{x}{\rightarrow} \end{cases} \tag{8}$$

The expression of the forces (7) becomes:

$$M_v \frac{dv}{dt} = -\frac{1}{2}\rho_{\text{air}} V^2 S C_x - M_v g \ \sin\alpha - M_v g C_r \cos\alpha + F_t \tag{9}$$

By definition, the mechanical power P_{m} required to drive the vehicle is equal to the traction force F_t multiplied by the speed V, with an efficiency of the transmission system η_{r}:

$$P_{\text{m}} = \frac{1}{\eta_{\text{r}}} \cdot F_t \cdot V \tag{10}$$

The required power is therefore expressed as follows:

$$P_{\text{dem}} = \frac{1}{\eta_{\text{r}}} \left(M_v \frac{dv}{dt} + \frac{1}{2}\rho_{\text{air}} V^2 S C_x + M_v g \ \sin\ \alpha + M_v \ g \ C_r \ \cos\ \alpha \right) V \tag{11}$$

where all parameters of P_{m} formula are explained in [12, 14].

Thereby, the charging current is formulated by:

$$I_{\text{ch}} = \frac{P_{\text{dem}} - \text{Losses}}{U_{\text{ch}}} \tag{12}$$

with the cancellation of Losses and U_{ch} is the DC bus voltage, $U_{\text{ch}} = 400v$. The sizing step is a key element in the management process, and it allowed us to determine a limit condition on the product $N_{\text{s}} * N_{\text{p}}$ (number of elements of serial storage multiplied by the number of elements in parallel).

The dimensioning step is also completed for the vehicle (Toyota Prius type II) which uses a PEMFC pack for the fuel cell and an EDLC pack for the supercapacitor. Thus, based on the manufacturer's data (*see* Table I in [5]), the results of the design are summarized in Table 1 with changes in the PEM fuel cell.

Table 1 Results of sizing

Electric motor	P_{max}: 50 kW, Cycle: NEDC with 20 min = 1200 s G_{EV}: 1320 kg, V_{nom}: 400 V (DC bus), D_{max}: 150 km Speed$_{\text{max}}$: 150 km/h, from 0 to100 in 15 s
Fuel cell	E: 110 Ah, 30 cells of 12 v, P_{mon}: 45 kW, P_{max}: 75 kW in 30 s
EDL capacitor	E: 2600 F, 20 cells in series of 2.7, P_{max}: 80 kw
Method	Fuzzy MAS for power prediction based on MC

4 Optimization and Energy Management in HV

4.1 Energy Management Based on MAS

The vehicle configuration is constituted by the supercapacitor and the fuel cell.

The strategy used in our case is part of the strategies based on real-time prediction. This method is called "Fuzzy MAS for power prediction based on the MC." Knowing that the losses in the DC/DC and DC/AC converters and the power of the auxiliaries are negligible. This method optimizes the powers distributed by HES by Eq. (13):

$$P_{\text{fc}} + P_{\text{sc}} = P_{\text{ch}} \tag{13}$$

This method aims to improve the performance of the hybrid power supply [15]. The purpose of each storage component is:

- The fuel cell assures the autonomy. During steady-state conditions, P_{fc} is limited at the continuous bus to maintain it constantly, which compensates the low voltages of the fuel cell.
- The supercapacitor provides supplementary power during the changing phases. It ensures the buffer storage of energy. During transient conditions (traction/brake cycles), the supercapacitor exchanges quantities of energy. To meet these power demands, its SOC requires real-time monitoring. The SOCsc value should be back to the reference value at the end of any circle.

The elaborated management system is based on intelligent agents. In this *MAS*, we have two types of agents: reactive agents related to *HES* (*SC* agent, *FC* agent) and electric motor (motor agent) and cognitive agents (such as the control agent and command agents) of the embedded system. The control step is distributed and the command step is therefore local. Thus, the sequence of interactions between agents is modeled in the sequence diagram designed with UML2 [16] like in [6]. There is:

Motor Agent To control the engine in traction or braking, this agent is linked to the DC/AC inverter which receives the parameters of an accelerator and brake pedal sensors. The motor agent extracts at every moment the acceleration A_{cc} and speed V of EV from a database, then calculates the P_{dem} for the electric motor (Eq. 11). It then sends these values to command and control agents.

Fuel Cell Agent During operation, it acquires the discharge voltage U_{fc} and the current I_{fc} of the fuel cell. It instantly checks these parameters and calculates its state of charge SOC_{fc} from Eq. (1), and then displays its chronological variations in an HMI. The *FC* agent exchanges also a data with the control agent, which allows it to make a local command of *DC/DC* converter to meet the energy demands in permanent regimes.

Supercapacitor Agent It does the same work as the fuel cell agent.

Its role is to extend the existing power peaks in the NEDC cycle by Eq. (13), to load by the FC or to recover the braking energy through Eq. (2).

Control Agent It is a calculator that predicts the power of $P_{\text{fc-MC}}$ and $P_{\text{sc-Mc}}$ by Markov chains, and its role is the predictive sharing of power. It is the core of the system and exchanges several data with the other agents to cooperate and coordinate the actions to the fact. It displays $P_{\text{fc-MC}}$ and $P_{\text{sc-Mc}}$ on its interface as a result.

Local Command Agent It locally controls DC/DC converters associated with the HES, by checking the current of charge $I_{\text{ch-ref}}$, U_{ch}, and P_{dem} (Eq. 12). So, the command agent calculates the reference currents $I_{\text{fc-MC}}$ and $I_{\text{sc-MC}}$, applying the converter control rule below:

$$I_{\text{sc\_MC}} = I_{\text{ch\_Calculted}} - I_{\text{fc\_MC}} \tag{14}$$

where $I_{\text{fc\_max}} = 140\,A$

These calculated values are visualized on a graphical interface and communicate to the fuel cell and supercapacitor agents, which makes it possible to locally command their converters. This last task done by the command agent is the end of the management process launched by *MAS*.

4.2 Energy Optimization Based on Fuzzy Controller

The fuzzy logic was founded by [17]. He realized the concept of a fuzzy set. This is a concept that enables given a degree of membership of an element to a class; in this way, this element can belong more or less to a defined class.

To optimize the distributed power in an adaptive way, we used a fuzzy inference system [18]. It is composed of a fuzzifier based on the fuzzy C-means algorithm, a fuzzy inference system (FIS) where its rules base is determined according to the limitation constraints of power for each component of HES, and a defuzzifier based on the center of gravity (*CoG*).

The goal is to make a fuzzy limitation of power, that is to say, a fuzzy optimization of energy. From these constraints and Eq. (13), we have:

$$P_{\text{dem}}(t) = P_{\text{fc}}(t) + P_{\text{sc}}(t), \forall t \tag{15}$$

This equation means that the sum of the powers supplied by the HES must reply equally to the load power at time t, where each source run at any time in a minimum and maximum power interval:

$$\begin{cases} P_i^{\min} \le P_i(t) \le P_i^{\max} \\ P_i^{\min} \le 0 \le P_i^{\max} \end{cases}, \quad i \in \{\text{fc}, \text{sc}\}, \quad \forall t \tag{16}$$

Table 2 Inference matrix of *FIS*

P_{fc}		P_{dem} N	VW	W	A	H
SOCsc	VW	VW	O	H	H	H
	W	VW	W	A	A	A
	A	VW	W	O	O	O
	H	VW	VW	W	W	O

The controlled parameters are the following: P_{dem} is the power demanded by the EV, SOC_{sc} is the stat of charge of the supercapacitor, and P_{fc} is the power delivered by the fuel cell.

The symbolic values of each parameter are five in number and of trapezoidal shape:

- P_{dem}: N or "negative" if $P_{dem} < 0$, VW or "very weak" if $0 < P_{dem} < 15$, W or "weak" if $10 < P_{dem} < 25$, *A* or "average" if $20 < P_{dem} < 35$, and *H* or "high" if $30 < P_{dem} < 50$.
- SOC_{sc}: VW or "very weak" if $SOC_{sc} < 25\%$, W or "weak" if $20\% < SOC_{sc} < 55\%$, *A* or "average" if $50\% < SOC_{sc} < 75\%$, and *H* or "high" if $SOC_{sc} > 75\%$.
- P_{fc}: VW or "very weak" if $0 < P_{fc} < 15$, W for "weak" if $10 < P_{fc} < 25$, O or "optimum" if $20 < P_{fc} < 35$, *A* or "average" if $30 < P_{fc} < 45$, and *H* for "high" if $40 < P_{fc} < 75$.

The rules base finally built are cast in the inference matrix shown in Table 2.

4.3 Power Prediction Based on Markov Chain Model

A Markov chain is a Markov process named after their inventor, Andrei Markov in 1906. It is a sequence of random variables (X_n, $n \in N$) based on evolutionary logic, and its objective is to model the dynamic behavior of a random system.

Where X_n represents the state of the system at instant n. In a Markov chain, "the future evolution of Xn variables depends only on the past by its present value," it is its fundamental property known as Markov property. In another way, the variables X_n, of (X_0, ..., X_n) and ($X_n + k$, $k \in N$), are independent.

Markov chains were therefore applied in our case as a stochastic algorithm of energy optimization by simulating the hybrid vehicle behavior, in terms of speed or acceleration and the power demanded.

To apply the real-time prediction to our system during the temp interval [t_k, t_{k+1}], we have adopted the steps of the Markov process defined by Himi et al. [8, 23] and adapted to the case studied as follows:

- The vehicle speed and acceleration are the files managed in real time. *These data are then analyzed* to have a distribution grated to the following conditional function:

Table 3 NEDC course sequence [23]

A_v vs. P_{dem}	$[P_{dem1}P_{dem2}]$	$[P_{dem2}P_{dem3}]$	$[P_{dem3}P_{dem4}]$	$[P_{dem4}\ P_{dem5}]$
$[A_{v1}A_{v2}]$	S1	S2	S3	S4
$[A_{v2}A_{v3}]$	S1	S3	S4	S4

```
    If(A=0 and Pdem=0)then State='O';
Else If(A>0 and Pdem>0)then State='A';
Else If(A=0 and Pdem>0)then State='C';
Else If(A<0 and Pdem<0)then State='R';
Else State ='X';
```

– This analysis of the results leads to divide this parse into sequences with *O* stationary state, *A* state of acceleration (SC provide demanded power), *C* state of constant energy (supply of energy by FC), *R* state of braking (energy recovery, by SC), and *X* change of state (intermediate state inflection point).
– The principal sequences are presented in Table 3 as follows: *O* is the first sequence S1, *A* the second S2, *C* is S3, and *R* is S4.

– The relationship between the future state and the current state of the system is given by the model of the Markov chain and is presented by the Eq. (17):

$$P\left(X_{(k+1)} = j|X_{(k)} = i\right) = P_{ij} \tag{17}$$

– The general process of Markov and the interaction between states is illustrated in [19]. The state of the system is chosen as a vector *V* composed of the acceleration (Acc) and the requested power (P_{dem}) [8]. The transition probability giving information on the next (future) state is represented by the Eq. (18):

$$P\left(P_{dem\ (k+1)}|P_{dem\ (k)}, A_{(k)}\right) = P_{k,k+1} \tag{18}$$

– The probability in the next step is given by:

$$P\left(P_{dem\ (k+n)}|P_{dem\ (k)}, A_{v(k)}\right) = P^n_{k,k+1} \tag{19}$$

This probability is calculated by aggregating these two parameters A and Pdem. The transition probability matrix is derived from sequence tables; this matrix has the role of estimating future energy request.
– The transition probabilities are determined for each vehicle speed sequence, by calculating the number of occurrences of each transition, as shown in the following Table 4 [8]:

Table 4 Number of occurrences for each sequence [23]

(S1, S1)	(S2, S1)	(S3, S1)	(S4, S1)	(S1, S2)	(S2, S2)	(S3, S2)	(S4, S2)
$m_{s1,\, s1}$	$m_{s2,\, s1}$	$m_{s3,\, s1}$	$m_{s4,\, s1}$	$m_{s1,\, s2}$	$m_{s2,\, s2}$	$m_{s3,\, s2}$	$m_{s4,\, s2}$
(S1, S3)	(S2, S3)	(S3, S3)	(S4, S3)	(S1, S4)	(S2, S4)	(S3, S4)	(S4, S4)
$m_{s1,\, s3}$	$m_{s2,\, s3}$	$m_{s3,\, s3}$	$m_{s4,\, s3}$	$m_{s1,\, s4}$	$m_{s2,\, s4}$	$m_{s3,\, s4}$	$m_{s4,\, s4}$

– Eq. (20) represents the calculus matrix of these occurrences:

$$N = \begin{pmatrix} m_{s1,s1} & m_{s1,s2} & m_{s1,s3} & m_{s1,s4} \\ m_{s2,s1} & m_{s2,s2} & m_{s2,s3} & m_{s2,s4} \\ m_{s3,s1} & m_{s3,s2} & m_{s3,s3} & m_{s3,s4} \\ m_{s4,s1} & m_{s4,s2} & m_{s4,s3} & m_{s4,s4} \end{pmatrix} \tag{20}$$

Besides, Eq. (21) represents the calculation method used, with $m_{si,\, sj}$ is the number of transition occurrences for $P_{\text{dem}\,(k)}$ until $P_{\text{dem}\,(k\,+\,1)}$ at $A_{v(k)}$:

$$P_{i,j} = P_{k,k+1} = \frac{m_{si,sj}}{m_{si}} \tag{21}$$

with m_{si} is the total transition number of $P_{\text{dem}\,(k)}$and we have:

$$m_{si} = \sum_{sj=1}^{n} m_{si,sj} \tag{22}$$

– Thus, we obtain the following transition matrix [23]:

$$P = \begin{pmatrix} \frac{m_{s1,s1}}{m_{s1,s1}+\cdots+m_{s1,s4}} & \cdots\cdots & \frac{m_{s1,s4}}{m_{s1,s1}+\cdots+m_{s1,s4}} \\ \vdots & \ddots \cdots & \vdots \\ . & \cdots \ddots & . \\ \frac{m_{s4,s1}}{m_{s4,s1}+\cdots+m_{s4,s4}} & \cdots\cdots & \frac{m_{s4,s4}}{m_{s4,s1}+\cdots+m_{s4,s4}} \end{pmatrix} \tag{23}$$

5 Simulation and Results

5.1 Simulation

The vehicle speed and acceleration are collected for the following *NEDC* course [2]. The controlled parameters are the powers of the fuel cell and supercapacitor. They are calculated by the controller agent using a fuzzy controller. Besides, this agent predicts also the power requested at the future stage; that is to say after three-time steps.

Figure 2 hereafter shows the fuzzy multiagent system for real-time managing and predicting implemented in the JADE environment [20]. A function of the jMarkov [21] library realizes the Markov chain with the predefined transition matrix, in Table 5. This matrix is intended to predict the requested power. A fuzzy controller using the jFuzzylogic library [18] calculates the optimum Pfc power while maintaining the SOCsc bounded.

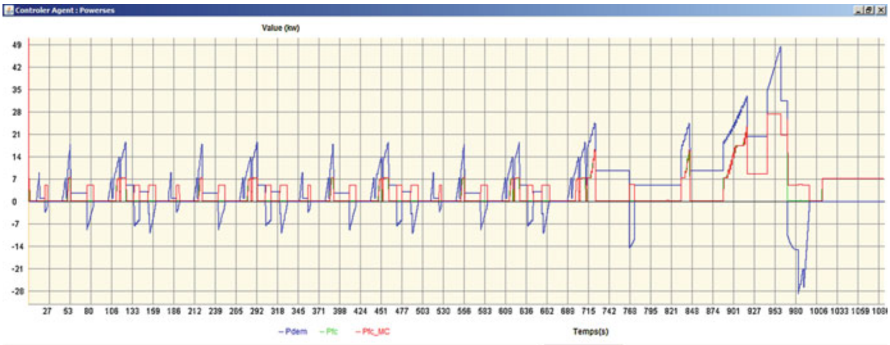


Fig. 2 Power demand and fuel cell without (P_{fc}) and with Markov chain (P_{fc}_*MC*)

Table 5 Predefined transition matrix

p	*x*	*o*	*a*	*c*	*r*
x	0.04	0.30	0.27	0.00	0.41
o	0.30	0.23	0.30	0.00	0.00
a	0.00	0.00	0.22	0.70	0.00
c	0.41	0.00	0.05	0.36	0.05
r	0.30	0.30	0.00	0.11	0.15

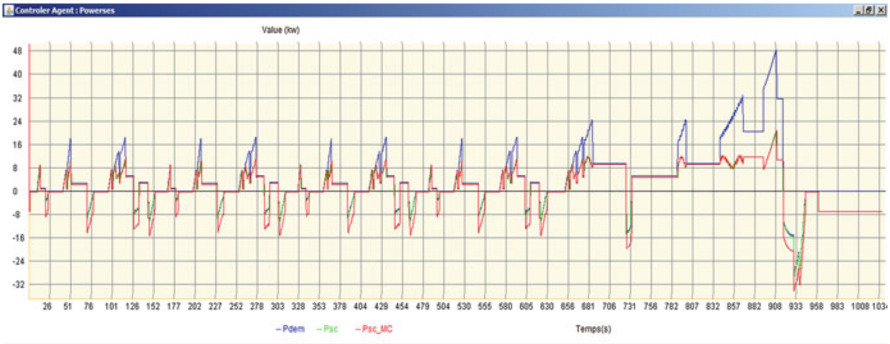


Fig. 3 Power demand and supercapacitor without (P_{sc}) and with Markov chain (P_{sc}_MC)

5.2 Results

Figures 2 and 3 show the results of the simulation for the proposed strategy applied to a hybrid electric vehicle. We compared the results of two models, the first model is without the Markov chain and the second is with the Markov chain. In both cases, the speed of the vehicle follows the predicted speed. In this simulation test, the problem of adaptive control is resolved with a type 1 fuzzy logic system.

The controller calculates the power of the fuel cell P_{fc} during vehicle operation (Fig. 2).

The P_{sc} power (Fig. 3) is also calculated from the difference between P_{dem} power and P_{fc} power [22]. The requested power of the electric motor is shared between the fuel cell pack and the supercapacitor pack. The latter respects the constraints imposed by the fuzzy controller algorithm [23].

6 Discussion

Simulation Without Markov chains During the simulations and in the case of "Without Markov Chain" phenomena were observed. It occurs when the power demand exceeds the maximum power of the fuel cell; that is to say during a sudden acceleration and a fast demand for power. The dynamics of the *HES* components were not identical (response of the fuel cell is slow relative to the supercapacitor) [23], so the system reacts quickly to stabilize this power spike. In this case, the fuel cell cannot provide more than its maximum power, and the supercapacitor continues to support to compensate for this lack of power despite its state of charge.

Simulation With Markov chains To remedy this problem, predicting the power demand is sufficient. So, the integration of an algorithm of Markov chain fulfills this role. The effectiveness of this algorithm lies in predicting the sequence where

P_{dem} can be elevated and likely to create these problems. The simulation results for this case show that with the addition of the *MC* algorithm, the previously mentioned problems disappear completely.

Comparison The idea of predicting the power demand is that the *PaC* cannot provide high power in a very short time. Thus, with the *MC* algorithm, the *FC* is forced before three sequences to activate and provide the P_{fc} by multiplying its real power (calculated by the fuzzy controller) by the probability of the transition matrix. The ease of integration with other optimization strategies is a major advantage of this algorithm. In Fig. 3 illustrates the consumption of *FC*, the Markov chain consumptions are slightly larger than those of "no Markov chain," because of the supercapacitor charge made by the FC during P_{dem} close to zero. But the variations in the supercapacitor state of charge disappeared through brake energy recovery, or during deceleration.

7 Conclusion

In this work, we have presented, tested, and compared two real-time energy management strategies by using a fuzzy *MAS*.

The first strategy is the Markov chains, and the second is with fuzzy *MAS* only.

The *MC* algorithm is added into the system model to predict the future P_{dem} and avoid some mismatches due to its changes.

The two models are implemented in an *HEV* with *FC/SC*.

The conditions of the road course are considered hidden, and the nature of the P_{dem} is therefore unknown. The *NEDC* road course was chosen to carry out the simulations because the speed changes it contains are numerous and more real than other cycles.

These mean more power losses by the *HES* of the vehicle and a decrease in its energy efficiency.

Thus, the system operates at a higher speed and extreme power, which makes it possible to properly study the behavior of the *FC* and *SC* in the face of these disturbances.

The results of the simulation confirmed that the stat of charge of *SC* remains limited by the fuzzy controller, and the *FC* consumption is almost equivalent to the model with fuzzy *MAS* only.

We deduce, also, the handy implementation of the model with the Markov chain by using the *MAS*, which allows separating the fuzzy controller and the *MC*.

In the end, the proposed multiagent system is the subject of an intelligent and integrated application. In perspective, this developed model can be incorporating with other optimization methods in real-time, on recent and at more true driven cycles.

References

1. ISO International Organization for Standardization. (2018). *Energy and ISO, ISO/TC 22/SC 37: Electrically propelled vehicles*. http://www.iso.org.
2. UNECE: United Nations Economic Commission for Europe, Information Unit. (2012). *WLTC cycles for validation 2, based on the following preliminary vehicle classification, WLTP-DHC-12-07*. www.unece.org/trans/main/wp29/wp29wgs/wp29grpe/wltp_dhc12.html.
3. Hankache, W. (2008). Optimized power management of a hybrid fuel cell generator. In *Electric energy*. National Polytechnic Institut of Toulouse.
4. Salmasi, F. R. (2007). Control strategies for hybrid electric vehicles: Evolution, classification, comparison, and future trends. *IEEE Transactions on Vehicular Technology, 56*(5), 2393–2404. https://doi.org/10.1109/TVT.2007.899933
5. El Amrani, R., Yahyaouy, A., & Tairi, H. (2017). Hybrid strategy based on MAS for intelligent energy management: Application to an electric vehicle. In *2017 Intelligent systems and computer vision conference (ISCV), Fez, Morocco* (pp. 1–8). https://doi.org/10.1109/ISACV.2017.8054917
6. El Amrani, R., Faquir, S., Yhayaouy, A., & Tairi, H. (2018). Modelling and implementation of an energy management simulator based on agents using optimized fuzzy rules: Application to an electric vehicle. *International Journal of Innovative Computing and Applications, 9*(4), 203–215.
7. Tie, S. F., & Tan, C. W. (2013). A review of energy sources and energy management system in electric vehicles. *Renewable and Sustainable Energy Reviews, 20*, 82–102. https://doi.org/10.1016/j.rser.2012.11.077
8. Hemi, H., Ghouili, J., & Cheriti, A. (2015). Combination of Markov chain and optimal control solved by Pontryagin's minimum principle for a fuel cell supercapacitor vehicle. *Energy Conversion and Management, 91*, 387–393. https://doi.org/10.1016/j.enconman.2014.12.035
9. Goua, Y. (2014). *Mathematical models and nonlinear and combinatorial optimization techniques for the energy management of a multi-source system: towards real-time implantation for different electrical structures of hybrid vehicles* (p. 39). Ph.D. thesis, Toulouse University.
10. Butterbach, S. (2012). *Electrical power storage by association of lead acid batteries and supercapacitors for heavy vehicles*. Ph.D. thesis, University of Technology, National No: 2012COMP2017.
11. Kermani, S. (2009). *Hybrid vehicle energy management: From simulation to real-time control*. Ph.D. thesis, Valenciennes et du Hainaut Cambrésis University.
12. Hattab, M. (2008). *Optimization and energy management for a hybrid system: Combination supercapacitor fuel cell*. Ph.D. thesis, Belfort Montbéliard University.
13. Yahyaouy, A., Sabor, J., Gualous, H., & Lamrini, M. (2012). Modelling and implementation of a multi-agent architecture for intelligent energy management. *Journal of Theoretical and Applied Information Technology (JATIT), 37*(2), 177–187.
14. Sadoun, R. (2013). *The interest of a hybrid electrical power source for urban electric vehicles—dimensioning and cycling tests (Intérêt d'une Source d'Energie Electrique Hybride pour véhicule électrique urbain - dimensionnement et tests de cyclage)*.
15. Allegre, A. L. (2010). *Modeling and energy management methodologies of mixed storage systems for electric and hybrid vehicles*. Lille University—Sciences and Technologies.
16. Hanandeh, F. A., Alsmadi, I., Al-Shannag, M. Y., & Al-Daoud, E. (2015). Mobile agents modelling using UML. *International Journal of Business Information Systems, 19*(4), 419–432. https://doi.org/10.1504/IJBIS.2015.070202
17. Zadeh, L. (1965). Fuzzy sets. *Information and Control, 8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X
18. Cingolani, P., & Alcala, J. (2013). jFuzzyLogic: A Java library to design fuzzy logic controllers according to the standard for fuzzy control programming. *International Journal of Computational Intelligence Systems, 6*(Supplement 1), 61–75. https://doi.org/10.1080/18756891.2013.818190

19. Gubner, J. A. (2006). *Probability and random processes for electrical and computer engineers*. Cambridge University Press.
20. Bellifemine, F. L., Caire, G., & Greenwood, D. (2007). *Developing Multi-Agent Systems with JADE* (p. 300). Wiley. http://jade.tilab.com/
21. Pérez, J. F., Silva, D. F., Góez, J. C., Sarmiento, A., Sarmiento-Romero, A., Akhavan-Tabatabaei, R., & Riaño, G. (2017). Algorithm 972: jMarkov: An integrated framework for Markov Chain Modeling. *ACM Transactions on Mathematical Software, 43*(3), 29. https://doi.org/10.1145/3009968
22. Hankache, W., Caux, S., Fadel, M., & Hissel, D. (2011). Real-time fuzzy energy management of fuel cell and ultracapacitor powertrains. In *Fundamentals and Developments of Fuel Cell Conference–FDFC 2008, Nancy, France*.
23. Himi, H. (2015). *Real-time energy optimization of a fuel cell vehicle and supercapacitors* (261 p). Ph.D. thesis, Québec à Trois-Rivières University.

A Modified Cultural-Based Genetic Algorithm for the Graph Coloring Problem

Mohamed Amine Basmassi, Lamia Benameur, and Jihane Alami Chentoufi

Abstract In this chapter, a modified cultural-based genetic algorithm (MCBGA) is presented to solve the graph coloring problem. The algorithm proposed here is an implementation of a modified cultural algorithm, which uses the genetic algorithm to represent the population space, besides the clique number and the max degree graph formulas are deployed to initialize the belief space. Moreover, the communication protocols between the two spaces (belief and population) are ensured by the best individual's depending on their fitness and number of used colors. Experiments on a set of seven well-known DIMACS benchmark instances show that the proposed algorithm achieves competitive results in comparison with two state-of-the-art algorithms in terms of both success rate and solution quality.

Keywords Cultural algorithm · Genetic algorithm · Graph coloring problem · Evolutionary intelligence · Metaheuristics

1 Introduction

1.1 Graph Coloring Problem

The graph coloring problem (GCP) is a well-known NP-complete problem. It includes both vertex coloring and edge coloring. Though, the graph coloring term in general goes to vertex coloring more than edge coloring [1].

For a specified set of vertices that form a connected graph, the first aim is to color vertices, on condition that no edge connects two identically colored vertices. It also aims to find the minimum number of colors required to color the graph without

M. A. Basmassi (✉) · J. A. Chentoufi
Laboratory of Informatics, Systems and Optimization, Department of Computer Science, Faculty of Sciences, Ibn-Tofail University, Kenitra, Morocco

L. Benameur
Laboratory LIROSA, Faculty of Sciences, Abdelmalik Essaadi University, Tetouan, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_66

violating the adjacency constraint. That number is called “the chromatic number χ(G)” [2].

For a given graph G, $P(G,c)$ is the number of possible solutions for coloring the graph G with c colors and $c = 1, 2, 3...$; then, the chromatic number is given by the following formula:

$$\chi(G) = \min\ (c : P\ (G, c) > 0) \tag{1}$$

The graph coloring problems belong to the NP-complete complexity classes that entertain many practical applications in addition to theoretical challenges and can be modeled as constraint satisfaction problems [3].

1.2 Cultural Algorithm

CA is a class of evolutionary algorithms which operates at two evolutionary levels: the micro and macro-evolutionary levels [4]. At the micro-evolutionary level, behavioral traits represent the principal description of individuals (that could be socially acceptable or unacceptable). Behavioral traits are transferred from generation to another using several socially motivated operators. At the macro-evolutionary level, individuals generalize descriptions from their experiences and can generate a mappa. Also, a group mappa can be formed by combining and modifying individual mappa employing a set of generic- or problem-specific operators.

The two levels share a communication bond which constitutes an interface between them, and the best individuals of the population can readjust the belief space and lead to the elimination of useless “or wrong” traits. Also, the contained knowledge of the belief space can affect the population component to guide and influence the evolution of the population at the next step. So, cultural algorithms can be described in terms of three basic components: the belief structure, the population structure, and the communication protocol [5]. The basic principles of CA are depicted in Fig. 1.

The population space consists of a set of possible solutions to the problem and can be modeled using any population-based technique. In this work, the genetic algorithm is selected as the framework to describe the population. One reason for this choice is that the rate at which genetic algorithms can explore a space of possibilities is very satisfying and adequate to the studied problematic.

1.3 Genetic Algorithm

Genetic algorithm is one of the simplest metaheuristic that appertains to random-based evolutionary algorithms, inspired from natural selection process, for solving both constrained and unconstrained optimization problems. The genetic algorithm

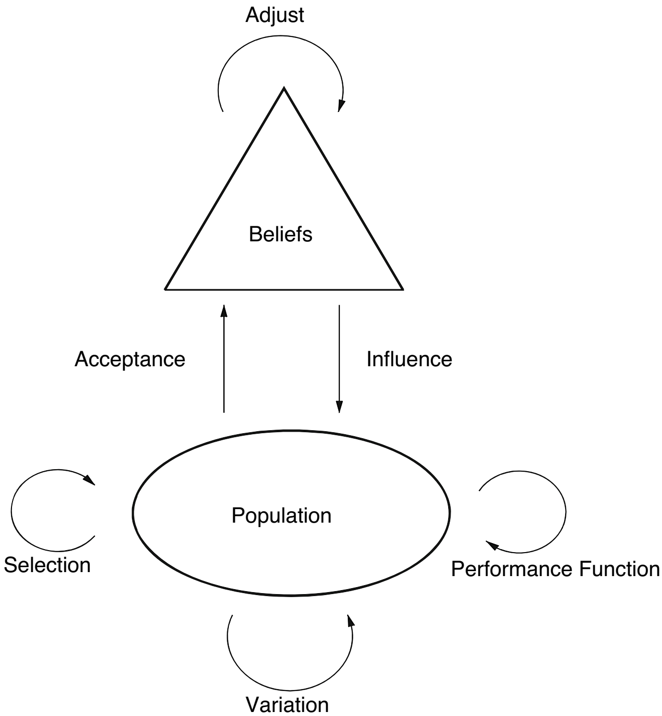


Fig. 1 Cultural algorithm

[6] is commonly used to generate high-quality solutions. The process generally starts with a population of random generated individuals, and then repeatedly modifies a population of individual solutions in an iterative process. The fitness is the worth of objective function that tests how much the individual is near from the optimal solution. In each generation, the selection of individuals is done at random from the current population to be parents. These parents are used to produce new individuals, after that the mutation modified randomly a minority of them depending on the coefficient of mutation. Then, the new generation of candidate solutions is employed in the following iteration of the algorithm. Commonly, the algorithm ends once a maximum number of generations has been reached, or a satisfying fitness stage has been produced for the problematic.

In the next section, a short view on prior works used to solve the GCP is presented. The proposed algorithm is presented in details in Sect. 3. Finally, in Sect. 4, experimental results are presented and compared to other methods reported in the literature [7, 8].

2 Prior Works

The GCP is one of the combinatorial optimization problems. While, a lot of metaheuristics and heuristics algorithms were innovated to resolve GCP, they have several limitations in one way or another.

In [9], a combined genetic algorithm with a special crossover operator and tabu search algorithm was presented. The Branch-and-Cut algorithm based on an integer programming formulation is proposed for solving graph coloring instances that are out of the reach of Dsatur method [10]. Also, a memetic algorithm [11] integrating a tabu search procedure with genetic algorithm to solve the GCP called MACOL algorithm is presented in [12]. Furthermore, many algorithms based on local search metaheuristic methods including iterated local search [13], greedy randomized adaptive search procedure [14], and reactive partial tabu search [15] were implemented to resolve the GCP. Interested readers are directed to [16] for an inclusive review of the local search approaches for graph coloring. Recently, in [17], swarm intelligence-based technique called as particle swarm optimization is employed. In [18], a recursive largest first-like algorithm gives good results, allowing to obtain a reduction of more than half of the gap between the number of colors used and the best-known upper bound of the chromatic number.

This chapter presents the resolution of the GCP by merging a modified cultural algorithm, which uses the clique number and the max degree formula to initialize the belief space, with the genetic algorithm. The suggested procedure is performed on multiple instances of graphs imported from the DIMACS library [19], and the attained results are compared with others reported in the literature [7, 8].

3 The Proposed Approach MCBGA

The proposed approach combines cultural algorithm (CA) and genetic algorithm (GA) to solve the GCP. In MCBGA, the purpose of the belief space in cultural algorithm is to guide the research procedure in order to find the chromatic number. The clique number and the max degree are used to initialize the belief space, while the population space is initialized depending on the knowledge from the belief space, then the genetic algorithm is applied. The two spaces intercommunicate using a communication protocol (via acceptance and influence functions) and when the maximum number of generations is reached, the optimal solution is obtained.

The pseudo-code of the proposed method is as follows:

```
Begin
     Initialize the Belief space BLS(0)using Eq.3;
     Generate the initial population POP(0)using BLS(0);
     Evaluate POP(0);
     Repeat
        Select parents to be used by genetic operators;
        Generate new chromosomes using crossover;
```

```
        Applied mutation on the new chromosomes;
        Evaluate POP(t);
        Acceptance (POP(t),BLS(t));
        Update(BLS(t));
        Influence (BLS(t),POP(t));
        t <- t+1;
        Until (Terminating condition is reached);
End
```

3.1 The Belief Space

In this work, the belief space is only represented by the normative knowledge, which stores intervals of form $[L_c U_c]$ (L_c and U_c are respectively the lower and the upper bound) for the decision variables of the problem that correspond to the regions where good results were found [5]. In the studied problem, the decision variable represents the chromatic number.

For a graph G, the chromatic number, $\chi(G)$, is the smallest number of colors required to color G [2]. The calculation of χG is NP-complete.

The clique number of a graph G, $\omega(G)$, is the number of vertices of the maximum clique that G has. The chromatic number of a graph is greater than or equal to its clique number [2]. For any graph G with maximum degree Δ, the chromatic number of G is always less than plus one [20]. Therefore:

$$\Delta + 1 \geq \chi(G) \geq \omega(G) \tag{2}$$

$$\begin{cases} U_c = \Delta + 1 \\ L_c = \omega(G) \end{cases} \tag{3}$$

3.2 The Population Space

Initial Population The proposed algorithm begins with a population P of n feasible coloring graphs. This population can be acquired via any graph coloring algorithm which is capable to produce different appropriate coloring for a graph. In this case, the belief space is used to generate different chromosomes with different number of colors randomly.

Parents' Selection The tournament selection is used as the selection method, where two individuals are randomly selected from the population, and then the best out of these is selected to become a parent.

Crossover In this operator, two parents are selected and two children are produced using the genetic material of the parents. In this approach, a three points crossover is used.

Mutation The uniform mutation is used to maintain and introduce diversity in the genetic population and it is applied with a small probability.

Child Generation This population includes$n - 1$ new child chromosomes gets by the modified genetic process (where n is the population size) and the best individual from the parent generation (Elitism strategy), which guarantees that the best solution obtained by the GA so far will be preserved among the generations.

The Communication Protocol *The Acceptation function.* The acceptation function: in this method, all child's generations chromosomes are evaluated and the best individual is selected to update the belief space.

The influence function. The knowledge stored in the belief space is used to adapt all individuals with unacceptable traits to the new beliefs.

4 Experimental Results

4.1 Data

The data employed to test the proposed approach are derived from the DIMACS benchmarking graph collection. DIMACS is the Center for Discrete Mathematics and Theoretical Computer Science [19].

This collection is habitually used in experiments involving constraint satisfaction problems. For the genetic algorithm used at this work, the crossover probability used is $P_{\mathrm{c}} = 90\%$, and $P_{\mathrm{m}} = 1\%$ for the mutation probability. The algorithm is run for 100 generations, and the population size used is 200.

4.2 Results and Discussion

The obtained results using the proposed approach are depicted in Table 1. Each line in the table corresponds to a particular graph. The first columns indicate, respectively, the name, the number of vertices, and the number of edges of the considered graph. The next column displays the chromatic number. KMCGBA represents the obtained results using our suggested algorithm, and K [5] and K [15] display the best results presented, respectively in [7, 8].

A brief explanation was done to give an idea about algorithms that we chose for comparing results.

In [8], a column generation-based algorithm is proposed to solve the robust graph coloring problem which is a generalization of the original graph coloring problem,

Table 1 Results

Graph	V	E	$\chi(G)$	KMCGBA		K [5]		K [15]	
				K	Suc/15	K	Suc/15	K	Suc/15
Queen7_7	49	476	7	7	4/15	11	–	7	15/15
Huck	74	301	11	11	15/15	17	–	11	15/15
Jean	80	254	10	10	15/15	15	–	10	15/15
David	87	406	11	11	15/15	17	–	11	14/15
Miles250	128	387	8	8	3/15	12	–	8	11/15
2_Insertion_4	149	541	4	6	1/15	6	–	–	–
Anna	138	493	11	11	15/15	17	–	11	12/15

and the algorithm is composed of two phases. The column generation phase requires constantly generating columns with negative reduced cost, that corresponds to solve a pricing sub-problem. The pricing sub-problem is resolved at each iteration, and it also generates new columns to expand restricted master problem or indicate that an optimal solution for the linear programming relaxation of master problem is obtained. In second phase, a branch-and-bound method is employed along with the column generation procedure for each node of the branch-and-bound tree. This method is called branch-and-price. Information from the pricing sub-problem and the optimal solution to master problem are used to find an integer feasible solution.

In [7], a hybrid genetic algorithm is based on a local search heuristic called DBG. The initialization of the number k of colors is carried out by the DBG algorithm to give an approximate value of χG. Then the genetic algorithm is used to find and improve an optimal solution for GCP in experimental parameter; the authors functioned with a population equal to 120, a crossover probability $P_c = 0.7$, a mutation probability $P_m = 0.15$, and iteration number varies between 250 and 2000.

The challenge tacked in this work is choosing very different graphs for test from different selection of graph, for that two recent state-of-the-arts are selected to show how our algorithm is flexible to give optimum result.

Comparisons between these algorithms show that the MCBGA gives better results than the column generation-based algorithm in most of the cases. Furthermore, results obtained by the hybrid genetic algorithms [7] are the same results as the proposed algorithm unless for the Queen7_7 and Miles250 where the it does not succeed to surpass [7], but for the David and Anna Graphs the proposed algorithms have been successful to surpass it in the rate of trials run.

5 Conclusion

In this chapter, a modified cultural-based genetic algorithm is proposed to solve the graph coloring problem in small, medium problem instances. The hybridization of the algorithms used in this work shows a harmonic combination. Genetic algorithm explores the space of solutions and cultural algorithm supervises it and intelligently adjusts its individuals to avoid blindness searching and to reduce computation cost, all that contribute to a fast convergence and localization of the optimal solution. The computational experiments, carried out on a set of seven benchmark graphs, show that our algorithm is able to provide strong performance on the majority of instances. The results obtained, using the proposed approach, are competitive compared to the current best-known results reported in state-of-the-art.

References

1. Jensen, T. R., & Toft, B. (2011). *Graph coloring problems*. Wiley.
2. Erciyes, K. (2013). *Distributed graph algorithms for computer networks. Computer communications and networks* (1st ed.). Springer.
3. Garey, M. R., & Johnson, D. S. (1979). *Computersand intractability: A guide to the theory of NP-completeness*. W. H. Freeman.
4. Reynolds, R. G. (1994). An introduction to cultural algorithms. In A. V. Sebald & L. J. Fogel (Eds.), *Evolutionary programming—Proceedings of the Third Annual Conference* (pp. 131–139). World Scientific Press.
5. Basmassi, M. A., Benameur, L., & Chentoufi, J. A. (2019). A hybrid cultural-hierarchical clustering algorithm for the graph coloring problem. *Journal of Advanced Research in Dynamical and Control Systems, 11*(6-SI), 1834–1842.
6. Mitchell, M. (1996). *An introduction to genetic algorithms*. MIT Press.
7. Douiri, S. M., & Elbernoussi, S. (2015). Solving the graph coloring problem via hybrid genetic algorithms. *Journal of King Saud University Engineering Sciences, 27*(1), 114–118.
8. Yceoglu, B., Ahin, G., & van Hoesel, S. P. (2017). A column generation based algorithm for the robust graph coloring problem. *Discrete Applied Mathematics, 217*(P2), 340–352.
9. Galinier, P., & Hao, J. K. (1999). Hybrid evolutionary algorithms for graph coloring. *Journal of Combinatorial Optimization, 3*(4), 379–397.
10. Mndez-Daz, I., & Zabala, P. (2006). A branch-and-cut algorithm for graph coloring. *Discrete Applied Mathematics, 154*(5), 826–847.
11. Moscato, P. (1999). *New ideas in optimization chapter. Memetic algorithms: A short introduction* (pp. 219–234). McGraw-Hill.
12. Lü, Z., & Hao, J. K. (2010). A memetic algorithm for graph coloring. *European Journal of Operational Research, 203*(1), 241–250.
13. Chiarandini, M., & Sttzle, T. (2002). An application of iterated local search to the graph coloring problem. In *Proceedings of the computational symposium on graph coloring and its generalizations* (pp. 112–125).
14. Laguna, M., & Mart, R. (2001). A GRASP for coloring sparse graphs. *Computational Optimization and Applications, 19*(2), 165–178.
15. Blochliger, I., & Zufferey, N. (2008). A graph coloring heuristic using partial solutions and a reactive tabu scheme. *Computers and Operations Research, 35*(3), 960–975.
16. Galinier, P., & Hertz, A. (2006). A survey of local search methods for graph coloring. *Computers and Operations Research, 33*(9), 2547–2562.
17. Agrawal, J., & Agrawal, S. (2015). Acceleration based particle swarm optimization for graph coloring problem. *Procedia Computer Science, 60*, 714–721.
18. Adegbindin, M., Hertz, A., & Bellaiche, M. (2016). A new efficient RLF-like algorithm for the vertex coloring problem. *Yugoslav Journal of Operations Research, 26*(4), 441–456.
19. DataSet: Computational series: Graph coloring and its generalizations (2002). http://mat.gsia.cmu.edu/color02/
20. Feige, U., Halldrsson, M. M., Kortsarz, G., & Srinivasan, A. (2003). Approximating the domatic number. *SIAM Journal on Computing, 32*(1), 172–195.

Artificial Immune System and Artificial Neural Network in Intrusion Detection System

Hanane Chliah, Amal Battou, and Omar Baz

Abstract With the increasing worldwide network attacks, intrusion detection (ID) has become a popular research topic in the last decade. Several artificial intelligence techniques such as neural networks and artificial immune system have been applied in ID. The results are varied. The intrusion detection accuracy is the main focus for intrusion detection system (IDS). Most research activities in the area are aiming to improve the ID accuracy. This chapter begins with a review of the different categories of intrusion detection systems, especially when they run in cloud computing environments. Then, we present a description of the most used approach in IDS: *artificial immune system* and *artificial neural network* and their application.

Keywords Intrusion detection system · Cloud computing environment · Artificial immune system · Artificial neural network

1 Introduction

Computer security domain covers all the processes and instruments by which computer-based information, services, and equipment are guarded from involuntary/voluntary unauthorized access, change, or destruction. With the expansion of the networks, computer security is facing great challenges. Several research projects have been carried out, and numerous solutions have been proposed in the literature which have tried to provide some answers to the question of preventing and detecting at the appropriate time any unauthorized use misuse, overindulgence in a computing environment [1–3], etc.

H. Chliah (✉) · A. Battou
IRF-SIC Laboratory, Faculty of Science, Ibn Zohr University, Agadir, Morocco

O. Baz
Higher School of Technology of Agadir, Ibn Zohr University, Agadir, Morocco
e-mail: o.baz@uiz.ac.ma

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_67

Cloud computing has become a gorgeous target for attackers as the mainstream technologies extend computing, storage, and software as a service or as a utility to consumers, and thereby posing the so-called problem of internal facing security [4]. Security and confidentiality of cloud services are the major problems to meet, nowadays. The major concern besides data security is the detection and prevention of intrusions in the cloud based on virtualization methods. Similar to other systems, cloud computing also suffers from various conventional attacks such as IP spoofing, address resolution protocol spoofing, routing information protocol attack, DNS cache poison, denial of service (DoS) attacks, and distributed denial of service (DDoS) attacks [5].

Intrusion detection systems (IDSs) is one of the solutions that have become an essential component to the infrastructure security. IDSs have been proposed to detect and identify abnormal activities through different approaches.

Anderson [2] was among the pioneers to propose solution to intrusion detection (ID) problems in 1972. This author suggested the concept of IDS in 1980 which was one of the earliest works on ID. Between 1984 and 1987, a first IDS model was proposed by Denning [1]. This prototype was named as the intrusion detection expert system (IDES). 1990 is a watershed in IDS development story. White et al. [6] developed the network security monitor (NSM) [6]. Then IDS was formally developed as two camps: network-based IDS (NIDS) and host-based IDS (HIDS). Different approaches and methods have been adopted for the design of IDS.

This chapter debates various IDS methods using both artificial immune system (AIS) and neural networks. The rest of this chapter is organized as follows. In Sect. 2, we present the related work on IDS. In Sect. 3, we provide a detailed discussion about the conjunction of IDS and AIS. Section 4 provides a background to intrusion detection system and artificial neural networks. Finally, Section 5 draws conclusions and future research directions.

2 Related Work

2.1 Host-Based Intrusion Detection System (HIDS)

A host-based IDS is located on the monitored system and tracks changes to important files and directories. It takes a snapshot of existing system files and matches it to the previous snapshot. If the critical system files have been modified or deleted, the alert is sent to the administrator to inquire. Zirkle described host-based IDS as “loading software onto the system to be monitored.” This software, usually defined as a host band/personal firewall or agent-based software, performs the following operations:

- Uses log files and/or system audit agents as data sources and incoming and outgoing traffic from a single computer.

- Verifies the integrity of system files and tickers for suspicious processes, including changes to system files and user privileges.

2.2 Network-Based Intrusion Detection System (NIDS)

A network-based intrusion detection system (IDS) monitors and examines the traffic on its network segment to spot intrusion assails. IDS can be made of many sensors, each sensor being in charge of monitoring the traffic passing through its own segment.

The detectors cannot supervise anything outside their own segment or switch. Northcutt described network-based intrusion detection system (NIDS) as an ID system that monitors the traffic on its network segment as a data source. Implementation involves:

- The network interface card is disposed in promiscuous mode to capture all network traffic that crosses its network segment and packets passing on that network segment.
- A sensor, which controls the objective, must determine whether the packet flow corresponds to a known signature.
- There are three signatures that are especially major: first, string signatures that look for a text string that denotes a possible attack. Second, port signatures simply monitor connection essay to known and frequently attacked ports. Third, header signatures that monitor heavy or illogical combinations in packet headers.

2.3 Detection Types

2.3.1 Anomaly Based Detection

Anomaly detection systems are also famous as behavior-based systems. They rely on the fact that intrusions can be sniffed by noticing deviations from the expected behaviors of the system supervise. This "normal" attitude can either conform to some observations made in the past or to some predicts, produced by several techniques. Everything that does not accord to this "normal" pattern will be signalized as unnatural. The kernel process of anomaly detection is not to comprehend what is anomalous but to comprehend what is normal or predicted.

Scientific research has focused on the anomaly detection area due to its ability to detect unknown intrusions. Patcha and Park [7], present a study of current anomaly detection techniques, focusing on intrusion detections in computer networks. NIDES proposed by Lunt [8] is an anomaly detection system using a standard user profile measured statistically. Zainal et al. [9], introduced the feature selection algorithm into the anomaly detection system (Table 1).

Table 1 Advantages and disadvantages of anomaly based detection

Advantages	Disadvantages
They can detect unknown intrusions because a priori knowledge about specific intrusions is not required	The difficulty to build a normal behavior profile that contains every possible normal behavior
By adding the statistical-based algorithm, the anomaly detection system will be adaptive to the changing circumstance because it is relatively easier to update the statistical measures	The limit between normal and abnormal behavior is not precise in some time
	Malicious opponents will often adapt to appear normal, making it more difficult to distinguish
	Normal behavior may continue to progress, so that current knowledge of normal behavior may not be sufficient to represent all normal future behaviors

Table 2 Advantages and disadvantages of misuse-based detection

Advantages	Disadvantages
Permit to define an attack precisely and to give it a name	Only known attacks detection
Helps system administrators without great security background to understand what happened and if needed inform the security team	
When the rules are correctly defined, the relatively low rate of false alarms	
The signatures which are used in rules must be as specific as possible to prevent false alarms	

2.3.2 Misuse-Based Detection

This type of intrusion detection system contains a database of known vulnerabilities. It supervises traffic and searches a reason or a signature match. It manages in much the same way as a virus scanner, by searching for a known conformity or signature for every special intrusion event. It can be located on a host or on a network to monitor network vulnerabilities [10]. Signature-based IDS examines in progress traffic, transaction, activity, or performance for corresponds with known patterns of even specific to known attacks, and it increases alarm only when the so-called match is detected. Which the number of false-positive alarms is comparatively less in signature-based IDS (Table 2).

Intrusion detection system is an important part of the network security that researchers have developed many systems to detect intrusion using the algorithms like immune system and neural network. The following sections present the main approaches of the artificial immune system and artificial neural network.

3 Artificial Immune System

This section begins by explaining how the biological immune system works. Thereafter, it discusses the building blocks of AISs with more detail.

3.1 Concepts of the Biological Immune System

The human immune system has been effective at protecting a human body against a huge variety of foreign pathogens or organisms [11]. This outstanding property is attractive to computer security researchers and artificial intelligence researchers. Based on the studies by immunologists, an increasing number of computer scientists have suggested several different computer immune models [12]. The artificial immune system is designed for the computational system and inspired by the HIS; it is applied to solving various problems in the field of information security, particularly intrusion detection systems [11, 13]. Moreover, it incorporates many attributes of the HIS, including diversity, error tolerance, dynamic learning, adaption, and self-monitoring [12, 14]. The AIS has the capability to differentiate between the "self" (cells that are owned by the system) and "nonself" (foreign entities to the system) as intrusions. Likewise, detectors similar to lymphocytes are deployed in computer system nodes to intercept and report any malicious activities.

AIS research began in the mid-1980s with Farmer, Packard, and Perelson's study [15]. Their study implied that computer science might borrow from the immune system. The wide formative AIS researches for computer security were those that suggested the immune system as an analogy for IDSs. One of the classical theories is negative selection (NS) [16] which is abstract model of biological NS. In this theory, the detector model generated in censoring phase is planned to monitor the self-state and detect whether or not self has been changed. Then they estimated the method feasibility as a change-detection method on the problem of computer virus detection. Based on the above analysis, Kephart successfully applied immune mechanisms to antivirus problems [17]. With the development of HIS principle, negative selection algorithm (NSA) [18], clonal selection algorithm (CSA) [19], immune network algorithm (INA) [12], and danger theory algorithm (Aickelin et al., 2003) become the most representative algorithms in the AIS theory. Aickelin et al. [20] provided a detailed overview of immune system approaches to ID. They gave a review of methodologies, algorithms, and research groups in the application of AISs to ID. Kim et al. summarized six immune features that are desirable in an effective IDS [21]. They provided an overview in the view of the research development history.

3.2 Properties of the Immune System

AISs are inspired in the modus operandi of the biological immune system. Computer scientists imitate the behavior and mechanisms observed or theorized by immunologists, computationally modeling only a few aspects that are sufficient to address a specific problem. In the context of the metaphor that enables the translation of those mechanisms to the computer science field, a useful breakdown of the characteristics of the immune system is as follows [22]:

- Uniqueness: The immune system of each individual is unique.
- Distributed sensing and self-regulation: No centralized coordination and control is present in a scalable, parallel operation.
- Diversification: Clonal selection and hypermutation continually test detector configurations;
- Anomaly detection: Unknown patterns are distinguished from known models.
- Learning and memorization: Information is saved to optimize future responses to models already encountered in the past.

The common model known by the framework of AIS defines the rules to comply by AIS and the process for developing new approaches. The necessary conditions are [23]:

- The representation of system components.
- Adapting procedures to monitor the evolution of the system.

Then, the form of an antibody as a set of l parameters. These parameters may be represented by a point in a space of l dimensions. A prime mark is that in this plan, those antibodies are close to each other. Population or repertoire of N individuals is modeled as a space forms a finite volume V containing N points. An antigen is represented by the point Ag = <Ag1, Ag2, ..., Agl>, an antibody is also represented by a point Ab = <Ab1, Ab2, ..., Abl>. To measure the degree of completeness between the antigen and the antibodies, various techniques can be used. More frequently the distances are used [24]:

Euclidean distance:

$$D = \sqrt{\sum} li = 1 \text{ (abi} - \text{agi)}^2$$

Manhattan distance:

$$D = \sum\nolimits^{l} i = 1 \text{ |abi} - \text{agi|}$$

Hamming distance:

$$D = \sum\nolimits^{l} i = 1 \; \delta_i \text{ with } \delta_i = \begin{cases} 1 & \text{if abi} \neq \text{agi} \\ \delta_i = 0 & \text{if not} \end{cases}$$

$$\text{if } D \downarrow == \text{»} \text{ Affinity } \uparrow$$

Thus, the antigen-antibody affinity is relative to the distance in space between them. Once the antigens and antibodies are represented, the quantitative function of the degree of compliance defined between them remains only to implement the immune theories.

3.3 *Immune Algorithms*

3.3.1 Clonal Selection Algorithm

This theory is based on the principle that only the cells having the antigen acknowledge the antigen proliferate and become memory cells. The clonal selection algorithm is based on the following:

- Maintaining a set of memory cells.
- Selection and cloning of the most stimulated antibodies.
- Reselection clones proportionally to the affinity with the antigen.
- Removal of unstimulated antibodies.

The maturation of their affinity [25].

```
Begin
   P = set of shapes to recognize
   M = Population random individuals
   while (A minimal form is not
          recognized)
     for i de 1 à taille(P)
       aff = affinite(Pi, Mi)
     end for
      Select n1n1 Elements with the best affinity with the
      elements of M.
      Generate duplicates of these elements in proportion to their
      affinity for the antigen.
      Mutate all copies in proportion to their affinity with the
      P assembly forms.
      Add mutated individuals to the M population.
      Choose n2n2 of these mutated elements (optimized) as memory.
    end while
End
```

3.3.2 Negative Selection Algorithm

This notion is very interesting, particularly for systems monitoring applications and detecting and preventing abnormal or unusual uses [23]. The problem of protection of computer systems is the learning problem of distinguishing between self and nonself. Instead, they compare the charge detection problem inside systems to the process of negative selection takes place in the thymus [26]. Summary of the negative selection algorithm.

```
Begin
 S = set of elements of the self
 D = A detector array
  SeuilAff = affinity sill
       while (i < nbDetecteurs)
  Generating a di detector so that
  it has no affinity with a member S
    If(affinity(di,Si)>SeuilAff) Then
     classified Sii as non-self
    else if
      classified Si as self
    end if
  end while
return A set of detectors D
end
```

4 Artificial Neural Network

An ANN is an information-processing system inspired by the manner of biological nervous systems, such as the brain, process information. It is composed of a high number of interconnected processing elements (neurons) that work together to solve specific problems. Each processing element (neuron) is essentially a summation element held by an activation function. The output of each neuron (after applying the weight parameter associated with the connection) is transmitted as input to all the neurons of the next layer. The learning process is basically an optimization process in which the parameters of the best set of connection coefficients (weight) to solve a problem are found and includes the following basic steps [27]:

- Modify the parameters of the neural network (weight) to better approach the outputs.
- Verify how the real output generated for a specific input matches the desired output.
- Present the neural network with a number of entries (vectors each representing a pattern).

ANN is one of the best practices and has proven useful for intrusion detection [28–30]. According to different types of ANN, these techniques can be categorized into the following three categories: ANN-based intrusion detection supervised, unsupervised ANN-based intrusion detection, and ANN hybrid-based intrusion detection.

The supervised ANNs applied to IDS principally includes multilayer feed-forward (MLFF) neural networks and recurrent neural networks [10, 29, 31] employed MLFF neural networks to detect anomalies based on user behavior. But in practice, the number of training sessions is very wide and the distribution of training is imbalanced. MLFF neural networks are easy to reach at least local and stability is less. Especially for unusual attacks, the detection accuracy is very low. Some researchers have compared the effectiveness of supervised ANN with other methods similar to MVC and multivariate adaptive regression splines [10, 32]. Supervised ANNs have been shown to have lower detection performance than SVM and MARS.

The second category uses the unsupervised ANN to classify input data and separate normal behaviors from abnormal or intrusive behaviors. Use of unattended ANN in intrusion detection has several advantages. The main benefit is that unsupervised ANN can improve their analysis of new data without recycling. Fox was the first to apply a self-organized map (SOM) to learn the characteristics of normal system activity and to identify statistical variations from normal trends. Just like using supervised learning ANN, the performance of unsupervised ANN is also lower. Notably for low-frequency attacks, unsupervised ANN also gets a lower detection accuracy [33].

The third category is hybrid ANN which combines supervised ANN and unsupervised ANN, or combine ANN with other data mining techniques to detect intrusion [34, 35]. The reason to involve the hybrid ANN is to master the limitations of individual ANN. Jirapummin suggests employing a hybrid ANN for both visualizing intrusions using Kohenen's SOM and classifying intrusions using a resilient propagation neural networks [28] used a combination of SOM and radial basis function (RBF) networks. Aneetha and Bose [36] proposed the combined approach for anomaly detection using clustering and neural network techniques. The modified SOM or self-organized map is used to create the network using the distance, connection power, and location functions, and the *k*-means clustering algorithms group nodes from the network to the network using similarity measures. It is exposed when the learning rate increases the number of output nodes decreases. Shraddha Surana [37] provided intrusion detection using the fuzzy clustering network and the artificial neural network, where it attest to be the best precision for using both the techniques. But different ways to construct hybrid ANN will highly influence the performance of intrusion detection.

The system offers generally better results than IDS based on RBF networks alone. Sang-Jun Han and Sung-Bae Cho [34], proposed an intrusion detection technique based on evolutionary neural networks in order to establish the structure and weights of the appeal sequences.

Chen et al. [38], proposed hybrid flexible neural tree-based IDS based on flexible neural tree, evolutionary algorithm, and particle swarm optimization (PSO).

Cannady [39] employed a three-layer neural network for offline classification of connection archives in normal and misuse classes. The system designed in this study was planned to work as an autonomous system (and not as a basic classifier whose result could be used in a rule-based system). The feature vector used [39] consisted of nine features, all describing the current connection and the commands used. A dataset of 10,000 connection records, including 1000 simulated attacks, was used. The training set included 30% of the data. The end result is a two-class classifier that has been able to classify normal and attack records in 89 to 91% of cases.

Ryan et al. [31] reported an offline anomaly detection system (NNID) using a back-propagated MLP neural network. The MLP was formed to identify the user profile, and at the end of each log session, the MLP evaluated user commands for possible (offline) intrusions. The authors described their research in a small computer network of 10 users. Each feature vector describes the connections of a single user for an entire day. 100 of the most important commands are used to describe the behavior of a user. They used a three-layer MLP (two hidden layers). The MLP correctly identified the user at 90%.

5 Conclusion

Prevention of security breaches completely using the existing security technologies is unrealistic. As a result, intrusion detection is an important component in network security. IDS offers the potential advantages of reducing the manpower needed in monitoring, increasing detection efficiency, providing data that would otherwise not be available, aiding the information security community learn about new vulnerabilities, and providing legal evidence. This chapter describes a detailed overview of IDS, the background of the artificial immune system-based IDS face many difficulties like self-set constantly changes, and detection is in real time, and the artificial neural networks which improve the performances of IDS system which are based on the misuse detection model and the anomaly detection model. In order to resolve all these issues and make progress for this research, our future IDSs should focus on the questions of quick response and less false alarm and false negative. In the future, depending on the biological immune mechanism, we will be able to propose efficient identification models and algorithms, for the implementation of IDS in computational computing using the AIS algorithm.

References

1. Yu, Z., & Tsai, J. J.-P. (2011). *Intrusion detection: A Machine learning approach*. World Scientific.
2. Pathan, A.-S. (2014). *The state of the art in intrusion prevention and detection*. Auerbach Publications.
3. Elhoseny, M., Elminir, H., Riad, A., & Yuan, X. (2016). A secure data routing schema for WSN using elliptic curve cryptography and homomorphic encryption. *Journal of King Saud University—Computer and Information Sciences, 28*(3), 262–275. https://doi.org/10.1016/j.jksuci.2015.11.001
4. Rezvani, M. (2018). *Journal of Artificial Intelligence and Data Mining, 6*(2).
5. Chliah, H., Battou, A., & Baz, O. (2018). Performance security in distributed system: Comparative study. *International Journal of Computer Applications, 179*(15), 29–33.
6. White, G. B., Fisch, E. A., & Pooch, U. W. (2017). *Computer system and network security*. CRC.

7. Patcha, A., & Park, J.-M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks, 51*(12), 3448–3470.
8. Lunt, T. (1993). Detecting intruders in computer systems. In *Proceedings of the 1993 conference on auditing and computer technology*.
9. Zainal, A., Maarof, M. A., & Shamsuddin, S. M. (2006). *Feature selection using rough set in intrusion detection, 1–4*. IEEE.
10. Mukkamala, S., Janoski, G., & Sung, A. (2002). *Intrusion detection using neural networks and support vector machines* (pp. 1702–1707). IEEE.
11. Harmer, P. K., Williams, P. D., Gunsch, G. H., & Lamont, G. B. (2002). An artificial immune system architecture for computer security applications. *IEEE Transactions on Evolutionary Computation, 6*(3), 252–280.
12. Castro, L. N., de Castro, L. N., & Timmis, J. (2002). *Artificial immune systems: A new computational intelligence approach*. Springer Science & Business Media.
13. Hosseinpour, F., Bakar, K. A., Hardoroudi, A. H., & Kazazi, N. (2010). *Survey on artificial immune system as a bio-inspired technique for anomaly based intrusion detection systems* (pp. 323–324). IEEE.
14. Hofmeyr, S.A. & Forrest, S., 2000, Architecture for an artificial immune system.
15. Farmer, J. D., Packard, N. H., & Perelson, A. S. (1986). *The immune system, adaptation, and machine learning*. North-Holland.
16. Forrest, S., Perelson, A. S., Allen, L., & Cherukuri, R. (1994). *Self-nonself discrimination in a computer* (pp. 202–212). IEEE Computer Society Press.
17. Kephart, J. O. (1994). *A biologically inspired immune system for computers* (p. 10).
18. Pöllä, M. (2009). A generative model for self/non-self discrimination in strings. In M. Kolehmainen, P. Toivanen, & B. Beliczynski (Eds.), *Adaptive and natural computing algorithms* (Vol. 5495, pp. 293–302). Springer.
19. Castro, L.N. de & Von Zuben, F. J. (1999) Artificial immune systems: part I—Basic theory and applications 99.
20. Aickelin, U., Greensmith, J., & Twycross, J. (2008). Immune system approaches to intrusion detection—A review. In *International conference on artificial immune systems* (Vol. 6, p. 14).
21. Kim, J., Bentley, P. J., Aickelin, U., Greensmith, J., Tedesco, G., & Twycross, J. (2007). Immune system approaches to intrusion detection—A review. *Natural Computing, 6*(4), 413–466.
22. Dasgupta, D., & Nino, F. (2011). *Immunological computation* (p. 139).
23. de Castro, L. N., & Timmis, J. I. (2003). Artificial immune systems as a novel soft computing paradigm. *Soft Computing—A Fusion of Foundations, Methodologies and Applications, 7*(8), 526–544.
24. Gharbi, M. (2006). *Systèmes Immunitaires Artificiels et Optimisation* (p. 31).
25. Haktanirlar Ulutas, B., & Kulturel-Konak, S. (2011). A review of clonal selection algorithm and its applications. *Artificial Intelligence Review, 36*(2), 117–138.
26. de Castro, L. N., & Von Zuben, F. J. (2000). The clonal selection algorithm with engineering applications. In *Proceedings of the genetic and evolutionary computation conference (GECCO '00). Workshop on artificial immune systems and their applications* (p. 37).
27. Theodoridis, S., & Koutroumbas, K. (2009). *Pattern recognition* (4th ed.). Elsevier Academic Press.
28. Horeis, T. (2003). *Intrusion detection with neural networks—combination of self-organizing maps and radial basis function networks for human expert integration* (p. 24).
29. Tan, K. (1995). The application of neural networks to UNIX computer security. In *Proceedings of ICNN'95 - international conference on neural networks* (Vol. 1, pp. 476–481). IEEE.
30. Joo, D., Hong, T., & Han, I. (2003). The neural network models for IDS based on the asymmetric costs of false negative errors and false positive errors. *Expert Systems with Applications, 25*(1), 69–75.
31. Ryan, J., Lin, M.-J., & Miikkulainen, R. (1997). Intrusion detection with neural networks. In *Proceedings of AAAI-97 workshop on AI approaches to fraud detection and task management* (pp. 92–97).

32. SciTePress—Science and and Technology Publications. (2004). *Intrusion detection systems using adaptive regression splines*. In *Proceedings of the 4th International Workshop on Pattern Recognition in Information Systems* (pp. 26–33).
33. Beghdad, R. (2008). Critical study of neural networks in detecting intrusions. *Computers & Security, 27*(5–6), 168–175.
34. Han, S.-J., & Cho, S.-B. (2005). Evolutionary neural networks for anomaly detection based on the behavior of a program. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics), 36*(3), 559–570.
35. Jirapummin, C. (2002). Hybrid neural networks for intrusion detection system. In *Proceedings of the IEEK Conference* (p. 5).
36. Aneetha, A. S. (2012). The combined approach for anomaly detection using neural networks and clustering techniques. *Computer Science & Engineering: An International Journal, 2*(4), 37–46.
37. Surana, S. (2013). Intrusion detection using fuzzy clustering and artificial neural network. *Advances in Neural Networks, 9*.
38. Chen, Y., Abraham, A., & Yang, B. (2007). Hybrid flexible neural-tree-based intrusion detection systems. *International Journal of Intelligent Systems, 22*(4), 337–352.
39. Cannady, J. (1998). Artificial neural networks for misuse detection. In *Proceedings of the National Information Systems Security Conference* (p. 14).

An Integrated Control to Enhance Dynamic Performance of the Off-Grid Electrical Installation With Renewable Power Sources: Photovoltaic Source Case

Kamal Hirech, Mustapha Melhaoui, and Khalil Kassmi

Abstract The optimal management of production and storing of photovoltaic energy increases the life duration of equipment and accordingly reduces the cost of installation and exploitation. In this chapter, we present the operation of a stand-alone PV installation equipped with solar lead batteries. Our first aim is to ensure a longer batteries life and to protect them against the causes of premature aging, also to optimize the installation operation. All of that in order to minimize the cost and, also, to reduce the negative effect of manufacturing and the impact of using toxic materials such as lead, this allows to contribute to the environmental protection. Therefore, the experiment shows that the system concerned: guarantees mainly the state of charge estimation by a reliable algorithm in real time. The estimation is performed by a dynamic method to reduce the estimation errors caused by classical and static systems. These errors can reach 6% in this study; they are required into consideration and compensated during the PV installation operation. The charge/discharge process control allows the protection of the batteries, and the regulation system realized optimizes the number of charge/discharge cycles and, consequently, optimizes the batteries lifetime.

Keywords PV panels · MPPT · Control · State of charge · Batteries operation · Off-grid · Electrical grid

1 Introduction

The energy that comes from fossil sources presented disadvantages in terms of environment protecting, while that of renewable type and especially the photovoltaics ones appear as alternative sources saw their cleanliness and availability. However, despite the enormous development known by photovoltaics, the cost of

K. Hirech (✉) · M. Melhaoui · K. Kassmi
Electromagnetism, Signal Processing and Renewable Energy “LESPRE”, Department of Physics, Faculty of Sciences, University Mohamed I Oujda, Oujda, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_68

energy produced by operating PV systems is relatively high compared to traditional ones [1]. For this and to optimize the exploitation of this energy, PV systems must be reliable, good sized, and have a long service life. However, one of the main problems associated with the use of renewable energy is not available in continuous mode [2, 3]. The stand-alone PV installations are equipped with storage devices [4] that appear as a solution to problems of energy availability during non-sunny periods. However, these batteries or storage devices present the weakest element in installations and also the most polluting and toxic considering the materials used and the manufacturing process. Therefore, any optimization of batteries lifetime is through the provision of a system that is able to reveal the actual state in terms of state of charge [1] for these devices in order to control their processes of charge/discharge optimally and protect them against the causes of aging. The state of charge estimation systems available on the market are often integrated in circuits or devices named battery management systems (BMS) [5] or solar controller [6] that suffer mainly from problems related to the estimation methods and implementation techniques [7, 8].

In this chapter, we present the operation of stand-alone PV system equipped with solar batteries and a DC/DC boost converter. The overall operation is ensured in optimum conditions by:

- Maximizing the power supplied by a proposed MPPT improved control.
- Dynamic estimation of the state of charge by combining two conventional methods (OCV and Coulomb counting) to improve the precession, and correct the estimation errors in real time.
- Optimal transfer of energy between the installation blocs by using a control block that optimizes control of the system switches in opening/closing.

Management and supervision of data acquired by serial link with a human–machine interface (HMI).

2 The PV System Operation

2.1 The Architectural of the Installation

We represent in Fig. 1 the block diagram of the stand-alone PV installation created during this work which is mainly the charging of two solar batteries in series (24 V) from two panels in series through a boost converter. The installation is equipped with a monitoring system based on the use of a 16F877 microcontroller. Its objective is based on the execution of various algorithms to achieve all predefined tasks in this study that are:

- Optimizingoperation of a proposed MPPT control.
- Estimating in real time the batteries state of charge by execution the algorithm proposed.
- Controlling the energy flow by execution of the algorithms of Fig. 1.

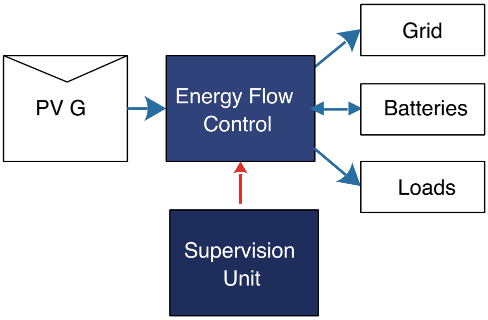


Fig. 1 Synoptic diagram of the designed system

2.2 *Mixed Dynamic Method for the State of Charge Estimation*

For best accuracy of the state of charge estimation SoC, taking into consideration the Faradic efficiency η [9] during the charge/discharge process, we adopted a strategy, which is mainly based on the combination of two methods:

- The first method, based on the use of the open circuit voltage (OCV) [7], determines the initial state of charge SoCi when the batteries are at rest from the Eq. (1). This equation presents the variation of the initial state of charge SoCi in function of the OCV.

$$\mathrm{SoCi}\ \left(25^{\circ}\mathrm{C}\right) = 50 * \mathrm{OCV}{-}1170 + \varepsilon 1 \tag{1}$$

- The second method consists to determine the quantity of charge absorbed or supplied by the batteries during operation by the Coulomb counting method [8, 10]. The quantity of charge absorbed or supplied is measured by an algorithm that integrates the current of charge (SoCc (%)) (Eq. 2) and the current of discharge (SoCd (%)) (Eq. 3), relative to time, and taking into consideration the faradic efficiency η [9]. According to this method, the measurement and conversion errors will be accumulated during the integration of the charging current I_{BAT} and discharging current I_{DECH}. These errors highly depend on the current acquisition circuit and their tolerance, as well as the estimation algorithm and the speed of the system; consequently, more the error is greater, more the estimation is incorrect, hence the presence of estimation error $\varepsilon 2$ for this method. The SoC(t)a (added to SoCi) estimated by this method is given by the Eq. (4):

$$\mathrm{SoCc}(t) = 100\% * C(t) * \int \eta * I_{\mathrm{BAT}}(t) * dt \tag{2}$$

$$\mathrm{SoCd}(t) = 100\% * C(t) \times \int I_{\mathrm{DECH}}(t) * dt \tag{3}$$

Table 1 Electrical characteristics of the charge/discharge parameters of batteries at 25 °C

Characteristics and conditions of charge	Batteries of 12 cells
Nominal voltage	24 V
Nominal capacity	110 Ah
Voltage of regulation (VR)	28.4 V
Voltage of floating (VFLT)	27.3 V
Current of over charge terminate (IOCT)	1 A
Limit voltage of discharge (LVD)	22.8 V
State of charge minimum to disconnect the batteries (SoC_{min})	40%
State of charge maximum to charge batteries (SoC_{max})	100%

$$\mathrm{SoC}\ (t)a = \left(\frac{100\%}{C(t)}\right) * \left[\int (\eta * I_{\mathrm{BAT}}(t) - I_{\mathrm{DECH}}(t)) * dt\right] + \varepsilon 2 \tag{4}$$

In the charge/discharge process, the Eq. (5) gives the expression of the state of charge SoC according to the proposed method, at the t instant

$$\mathrm{SoC}\ (t) = \mathrm{SoCi} + \left(\frac{100\%}{Qo}\right) * \left[\int (\eta * I_{BAT}(t) - I_{\mathrm{DECH}}(t)) * dt\right] + \varepsilon 1 + \varepsilon 2 \tag{5}$$

The batteries used during the experimentation are characterized by the data in Table 1.

3 Experimental Results

3.1 Experimental Procedure

The proposed system is implemented using a digital bench as shown in Fig. 2. The system consists of:

- Two PV panels mounted in parallel [11].
- Two lead acid solar batteries mounted in series [12], each one of 12 V voltage, 110 Ah capacity, and with a deep of discharge (DoD) equal to 60%.
- A DC/DC boost converter [13] and a human–machine interface developed to acquire, record, and trace the various electrical quantities, and also communicate with the microcontroller and manage all tests realized.
- A weather station equipped with a pyranometer, temperature sensor, Keithley multimeter, and an interface in PC to acquire and trace the variation of irradiance and temperature versus time.
- The supervision and monitoring unit which allows managing the acquired data, operation, and control of the PV system blocks.

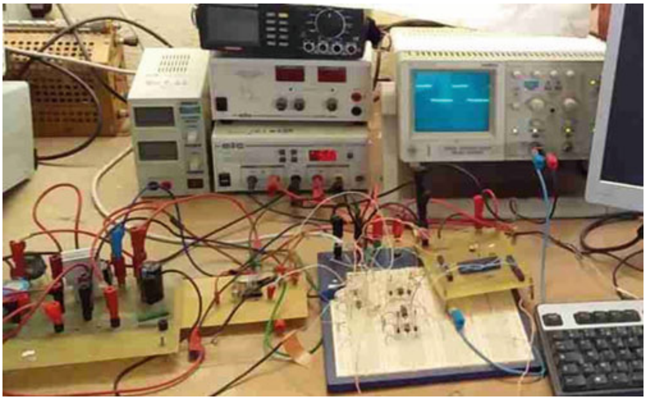


Fig. 2 The photovoltaic system designed

3.2 *The State of Charge Estimation*

We have charged the batteries used in this study completely for 5 days of experimentation, and we have estimated the state of charge by two methods: the first one is the OCV method (conventional) that consists to estimate the SoC when the batteries are in a state of rest [14]. The second method is presented and detailed in this study. The results produced by these two methods are compared to SoC estimation provided by the manufacturer of the solar batteries used (ideal SoC), depending on their type, temperature, and voltage; this is for periods of rest that are greater than 24 h.

We represent on Fig. 3a the estimation of the state of charge by the proposed method and the OCV method after a rest of battery of 12 h and the ideal SoC provided by the manufacturer. In Fig. 3b, we present the estimation error of each method, and the Fig. 3c presents the difference in % between the proposed method and the conventional one (OCV), these results show:

- The proposed method can estimate the state of charge of the batteries in real time. This estimation is nonlinear relative to time; this is explained by the fact that the irradiance changed when the quantity of charge absorbed by the batteries during the experimental periods also changed. Contrary, the OCV method can be used only if the battery voltage has stabilized.
- Either method, the proposed or the OCV, presents errors in estimation of the SoC; however, the errors introduced by the proposed method are of the order of 3%, this value decreases remarkably in the end of charging. However, the OCV method presents an error considerably higher in estimation (average value of the order of 5.5%). The errors depend on voltage value stabilization.
- The difference in the estimation between the OCV and the proposed methods compared to the ideal SoC (Fig. 3c) can reach a value around 2% at the beginning

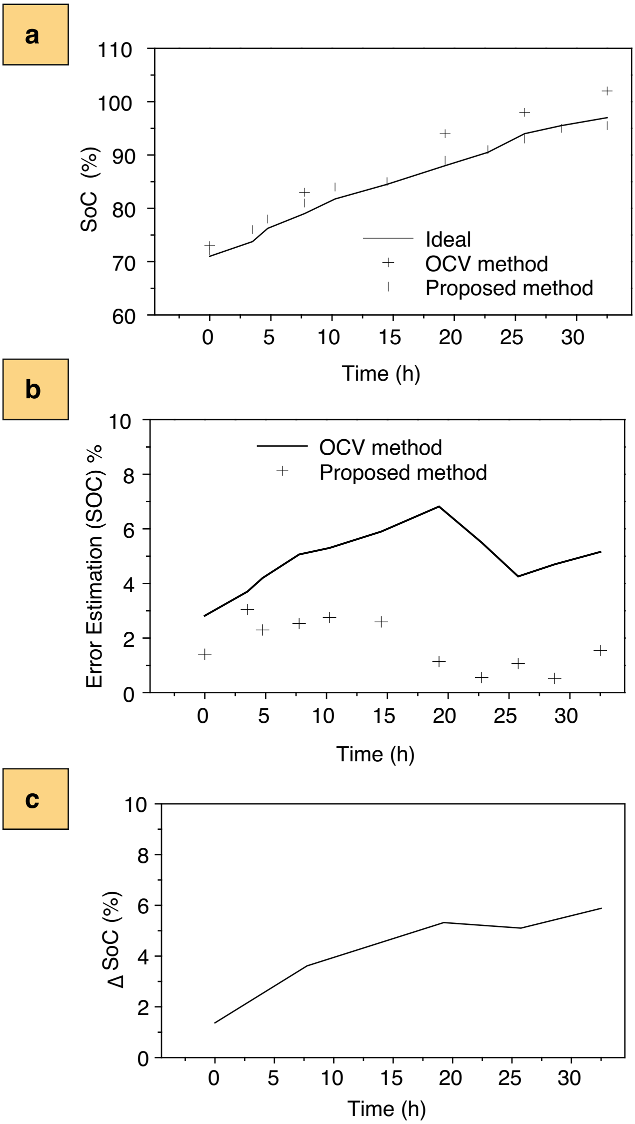


Fig. 3 The variations of: (**a**) SoC, (**b**) errors estimation, (**c**) difference in the estimation of SoC between the OCV and the proposed method compared to the ideal SoC

of charging, this value increases with time and reaches a value in the range of 6% due to estimation errors that accumulate and committed by the OCV method. However, the SoC estimation by the proposed method is always close to the ideal one. Hence, an average improvement contributed in estimating.

These results show that the proposed method allows estimating the SoC dynamically and with precision. Consequently, it optimizes the charge/discharge process and improves the of lifetime of the batteries.

3.3 Management Strategy of the Energy During Battery Operation

3.3.1 During the Batteries Charge Mode

We have tested and analyzed the system (Fig. 4) where the irradiance reaches 800 W/m^2. The results obtained show that the batteries are charged during three phases in order to fully charge and protect them against overcharging. In the first phase (0 h to 17 h), they are charged with the max of current until a limit value of voltage V_R. However, in the second phase (17 h to 26 h), the batteries voltage remains fixed and the charging current degrades until the full charge of these batteries. The third charging phase (26 h to 32 h) consists to compose the state of charge and maintaining the SoC at 100%.

3.3.2 During the Operation Modes of the Batteries

The batteries are tested during different modes of charge/discharge. The results obtained (Fig. 5) are concerning the evolution of the voltage of the batteries V_{BAT} and the state of charge SoC. During the charging, the voltage and the state increase. However, during the discharge mode, these parameters decrease. The control system stops the discharging when one and/or both protection conditions are validated (for example, at $t = 7.5$ h V_{BAT} = LVD, the system has stopped the discharging)

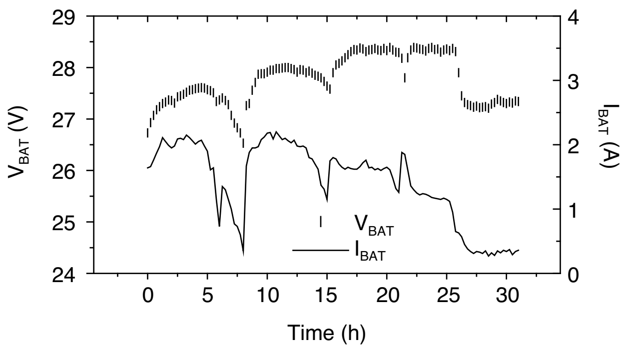


Fig. 4 Variations of the charging voltage V_{BAT} and current I_{BAT} as a function of time

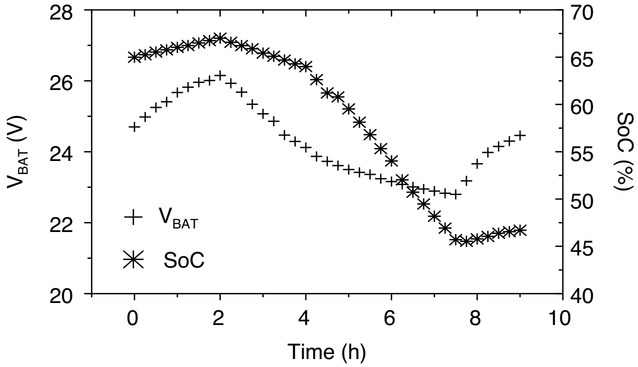


Fig. 5 The evolution of electrical characteristics of the batteries (the voltage V_{BAT} and the state of charge SoC)

to protect the batteries against deep discharges and following their characteristics presented in Table 1.

3.3.3 The Optimal Control and Operation of the PV Installation

To show the proper functioning of the electrical system, we present in Fig. 6 the evolution of powers and yields as a function of time. These results show (Fig. 6a) that the energy supplied by the photovoltaic generator is absorbed by the batteries, and the surplus is used to feed another load or inject into the grid via an inverter. So, we noticed a very good conversion efficiency that exceeds 90% (Fig. 6b). These results increase the proper functioning of the electrical system.

4 Conclusions

In this chapter, an off-grid PV installation equipped with a dynamic SoC estimator, MPPT controller, and a human machine interface is realized and studied to manage all data and electrical quantities acquired. The experimentation shows that:

- The proposed method of the SoC estimation overcomes the limits and imprecision in conventional method, and it can be used in installations that are equipped with battery for an estimation in real time.
- The MPPT controller forces the operation of the stand-alone installation to work under optimal conditions and to charge the batteries completely.
- The system of supervision communicates with the human–machine interface to treat different electrical quantities acquired and manage the energy supplied by the PV panels. All that is with a good efficiency.

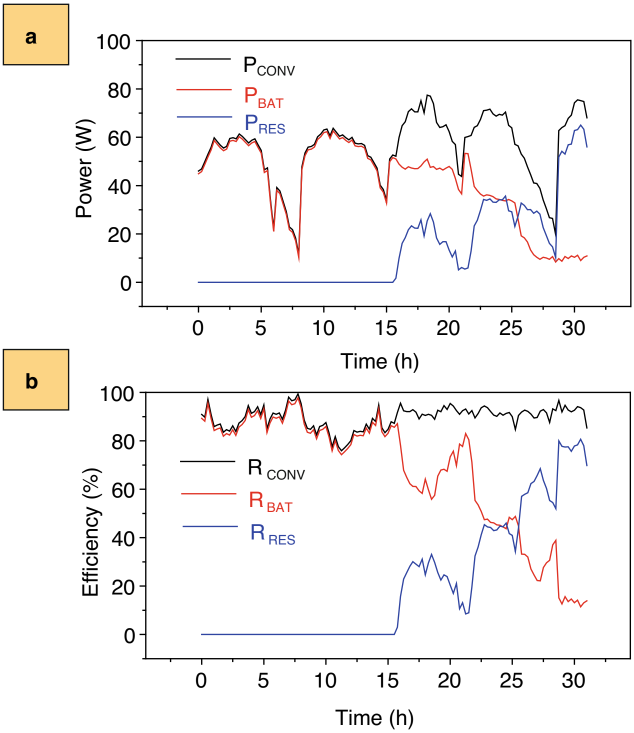


Fig. 6 The evolution of (**a**) P_{CVN}, P_{BAT}, and P_{RES} and (**b**) R_{CVN}, R_{BAT}, and R_{RES} during the operation

References

1. Lai, C. S., & McCulloch, M. D. (2017). Levelized cost of electricity for solar photovoltaic and electrical energy storage. *Applied Energy, 190*, 191–203.
2. Kumar, A., Sah, B., Singh, A. R., Deng, Y., He, X., Kumar, P., & Bansal, R. C. (2017). A review of multi criteria decision making (MCDM) towards sustainable renewable energy development. *Renewable and Sustainable Energy Reviews, 69*, 596–609.
3. Aguado, J. A., Racero, A. J. S., & de la Torre, S. (2018). Optimal operation of electric railways with renewable energy and electric storage systems. *IEEE Transactions on Smart Grid, 9*(2), 993–1001.
4. Lu, X., Sun, K., Guerrero, J. M., Vasquez, J. C., & Huang, L. (2015). Double-quadrant state-of-charge-based droop control method for distributed energy storage systems in autonomous DC microgrids. *IEEE Transactions on Smart Grid, 6*(1), 147–157.
5. Armengaud, E., Kurtulus, C., & Macher, G. (2015). A smart computing platform for dependable battery management systems. In *Advanced microsystems for automotive applications* (pp. 239–250). Springer.
6. Eldahab, Y., Saad, N., & Zekry, A. (2016). Enhancing the design of battery charging controllers for photovoltaic systems. *Renewable and Sustainable Energy Reviews, 58*, 646–655.
7. Stetzel, K., Aldrich, L., & Trimbpli, M. (2015). Electrochemical state and internal variables estimation using a reduced-order physics-based model of a lithium-ion cell and an extended Kalman filter. *Journal of Power Sources, 278*, 490–505.

8. Sepasi, S., Ghorbani, R., & Liaw, B. (2014). A novel on-board state-of-charge estimation method for aged li-ion batteries based on model adaptive extended Kalman filter. *Journal of Power Sources, 245*, 337–344.
9. Manhor, A., Kim, K., & Plichta, E. (2016). A high efficiency iron-chloride redox flow battery for large-scale energy storage. *Journal of the Electrochemical Society, 163*(1), A5118–A5125.
10. Jeong, Y., Yong-Ki, A., & Jung, H. (2014). Enhanced Coulomb counting method with adaptive SOC reset time for estimating OCV. In *Energy Conversion Congress and Exposition (ECCE), 2014 IEEE* (pp. 1313–1318). IEEE.
11. Hirech, K., Melhaoui, M., Yaden, F., Baghaz, E., & Kassmi, K. (2013). Design and realization of an autonomous system equipped with a charge/discharge regulator and digital MPPT command. *Energy Procedia, 42*, 503–512.
12. Melhaoui, M., Baghaz, E., Hirech, K., & Kassmi, K. (2014). Contribution to the improvement of the MPPT control functioning of photovoltaic systems. *International Review of Electrical Engineering (IREE), 9*(2), 393–400.
13. Hirech, K., Melhaoui, M., Atmane, I., Kassmi, K., & Bachiri, N. (2017). Photovoltaic energy: Research, training and development of rural areas in eastern of Morocco. *Journal of Communications Technology, Electronics and Computer Science, 14*, 6–10.
14. Snihir, I., Rey, W., Verbitskiy, E., Belfadhel-Ayeb, A., & Notten, P. H. L. (2006). Battery open-circuit voltage estimation by a method of statistical analysis. *Journal of Power Sources, 159*, 1484–1487.

Systemic Disruptive Events and Sustainability in Supply Chains: Assessments by Mediating Effect of Vulnerability and Resilience—A Study on Health Sector in Morocco

Mohamed El Abdellaoui, Youssef Moflih, Mostapha Amri, Mohamed Hansali, Abdelwahed Gourch, Meriem El Mountasser, Tarek Zari, and Abdelmounaim Aggour

Abstract The viable resilience of supply chains, in which resilience, durability, and performance remain unchanged at acceptable levels in the presence of potential disruptions by supporting performance while minimizing the spread and severity of risks through practical capacities for regulated but balanced resilience and sustainability. Through the 17 hypotheses, built based on the relevant theoretical frameworks (normal accident theory, systemic risk theory, resource dependence theory, and high reliability of organization theory), we seek to justify the meaning of complementarity between the concepts of vulnerability, resilience, and sustainability of supply chain. With a questionnaire survey administered during the period from January 2018 to April 2018 to public and private health establishments in the kingdom, thus making it possible to establish a database of 109 establishments. For evaluation, we used structural equation modeling and partial square analysis to

M. El Abdellaoui
Research Laboratory on New Economy and Development, Faculty of Law, Economic and Social Sciences, Hassan II-University, Ain-Sebaâ, Casablanca, Morocco

Research Laboratory on Transport, Logistics and Organization Management, National School of Commerce and Management, Ain-Sebaâ, Casablanca, Morocco

Y. Moflih
Research Laboratory on New Economy and Development, Faculty of Law, Economic and Social Sciences, Hassan II-University, Ain-Sebaâ, Casablanca, Morocco

M. Amri
Research Laboratory on Transport, Logistics and Organization Management, National School of Commerce and Management, Ain-Sebaâ, Casablanca, Morocco

M. Hansali · A. Gourch · M. El Mountasser · T. Zari
Research Laboratory on Modeling Applied to Economics and Management, Faculty of Law, Economic and Social Sciences, Hassan II-University, Ain-Sebaâ, Casablanca, Morocco

A. Aggour
Management Sciences, Hassan II-University, Casablanca, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_69

explore all possible links between the five variables that make up our conceptual research model. However, our results empirically justify this advantageous link, which remain even less explored.

Keywords Supply chain risk management · Risk in supply chain management · Sustainable supply chain · Supply chain performance · Supply chain resilience · Supply chain vulnerability · Moroccan hospital sector

1 Introduction

Competitive contexts become more turbulent and complex, making them more vulnerable to disruptive events characterized by dynamic propagation, uncertainty, and diffusion effects [1]. Alternatively, supply chains that are factors of competitiveness are required to evolve their managerial practices by integrating new approaches focused on risk and sustainability into the management of their logistics operations. This change or reconfiguration of global value chains will have a definite impact on public policies, especially those of social voices (for example: health policy), with more pressure also due to the restriction of public spending which does not seem to increase resources. All these conditions combined erode efforts to make hospital systems more responsive in terms of the level of services and care provided, which can lead to a deterioration in the organization and flow control methods and impact the continuity of operations in hospitals, particularly in developing countries. As a result, these organizations are faced with a challenge that does not correspond to their cultures that of developing models of reflection that allow them to maintain themselves in difficult contexts and to improve their governance systems. In this respect, these premises make it possible to guide our research thinking by admitting that supply chains constitute benchmarks to reveal the capacity of resilient and sustainable practices to support performance levels, to control both the degree of exposure to risks and the factors of supply chains vulnerability. The objective of this study is to propose a confirmatory analysis of the possible causal links between the five concepts that make up our conceptual model on risk management and supply chain sustainability (risk, vulnerability, resilience, sustainability, and performance).

2 Literature Review

2.1 Systemic Disruption in Supply Chain

Supply chains are becoming more complex, interdependent, and vulnerable, but at the same time ensuring competitiveness and value creation [2]. It is true that this structural evolution has increased the efficiency of supply chain operations through dependent and narrow interconnections allowing secure continuity in the management of flows and operations resulting in a balanced level of performance and customer satisfaction [3, 4]. In this context, supply risks are characterized by cumulative, retroactive, and reciprocal disruptions of a speed, extent, and magnitude of loss that vary according to the time of disruption and the complex design of supply chain structures [5, 6]. In these circumstances, supply risk disruptive events are likely to influence the flexibility and responsiveness of operations, thus being considered as a source of vulnerability that generates potential losses throughout the supply chain [7, 8]. In this sense, we recognize that supply chain risk factors and events are interrelated and that an incident of a first-tier supplier or transport provider potentially affects the end customer, even more serious are their negative impacts increasing in severity even if they result from an initial small disruption of significant probability [5]. This interconnection of structures and risks affects the inherent capabilities of resilient and sustainable supply chains necessary to safeguard sustainable competitive advantages [9, 10]. This contextual evolution of supply chains in an unpredictable and turbulent environment is due to the adaptive capacity to challenges resulting from downstream supply chain operations that increase the vulnerability of supply chains through increased instability and uncertainty [11, 12]. The literature on systemic risk confirms that the characteristics and locations of disruptive events change unexpectedly and that a small event can turn into a more potential failure with serious consequences on supply chain performance [4, 13]. This depends largely on the variability of the disruption time and appropriate mitigation measures put in place to control recovery time through more responsiveness required for sustainable management of flows and logistics operations [12, 14]. Whose transport risks cannot be treated in isolation given their impacts on the sustainability of upstream and downstream logistics operations, this is supported by the example of the two authors or a macro natural disaster risk, regulatory change or micro modification of transport standards and processes can change the time it takes to complete the shipment of orders upstream and downstream, thus creating an unforeseen bottleneck, with changes in demand having a direct effect on transport modes, inventory levels, performance levels. In this example, several interdependent and interconnected supply chain risks are mutually exacerbated even if it is an isolated disruption of an interface or part of the supply chain. To summarize, our first hypothetical formulation suggests that micro systemic risks have negative effects on supply chain performance (H1). It is subdivided according to the nature of the risks studied into four sub hypotheses (H1.1, H1.2, H1.3, and H1.4).

2.2 Supply Chain Performance

As much as a dominant performance dimension resulting from customer value resilience relies on the speed in the execution of their requirements through a certain visibility, flexibility, and controllability of logistics processes through indicators to quantify and measure the performance of logistics operations. Representing a significant part of the overall performance with a significant impact on economic, marketing, and operational performance [15]. Supply chain performance is measured by a wide range of criteria evaluation models and measurement scales that vary according to the contexts and issues studied. For example, used the criteria of customer satisfaction, speed, flexibility, and reliability of delivery. While, used seven items: reactivity, cost reduction, delivery reliability, lead time, specification compliance, process improvement, and time to market. More recently, some authors have combined several scales of measures for supply chain performance assessment (such as [15]. For this study, we will use a combined measurement scale of 16 items with a particular focus on the social aspect. So, the second, third, and fourth hypotheses will be: supply chain performance is more affected by the supply chain vulnerability factors (H2). Supply chain performance improves when resilient capacities are present (H3). Supply chain performance improves when sustainable capacities are available (H4).

2.3 *Supply Chain Resilience*

We believe that supply chains must further introduce mechanisms of resilience and agility with balanced or regulated levels in order to improve their adaptive capacities in complex and vulnerable contexts. As a result, the structural complexity and unpredictability of supply chain forecasts upstream and downstream have made them less responsive to change and therefore less resilient and less sustainable [16, 17]; this is in line with the observation of the two authors, where supply chains optimized to be less costly and faster have the opposite effect by becoming more vulnerable to disruptive incidents of evolving probability and severity [18, 19]. Faced with this propagation of a nature, progressive attribute, and a cumulative and dynamic effect, the importance of protecting the continuity and security of operations through safeguard measures was necessary to mitigate before, during, and after the occurrence of a potential disruption remains [20]. And whose authors argue that often the supply chains that are likely to survive are not necessarily the most robust but those that are resilient and able to adapt, grow, and recover quickly in the face of turbulence [21]. Indeed, resilience refers to both an inherent capacity or property of the interfaces of the supply chain and an emerging part of their network; this double level or skills at which they can manifest themselves implicitly reflects their intrinsic attributes and mechanisms, which remains exclusive for this shifting concept. In this sense, the extent, location of propagation, and correlation of structures, systemic, and reciprocal disruptive events make the choice of properties of resilient strategies necessary to mitigate them ineffective [6, 10]. So, fifth and sixth hypotheses will be: more supply chains adapt resilient practices, the more their contextual factors are controllable (H5). More supply chains are less resilient, the more sustainable supply chains capacities are questioned (H6).

2.4 Supply Chain Vulnerability

Several recent examples show the extent and effects of dynamic profit disruptions on the capacity of the resilience tools and practices needed for viable and sustainable supply chains. In the same sequence of ideas, this viability translates into sustained levels of resilience, sustainability, and performance even in the presence of potential disruptions due to increased sensitivity to sources of risk. This sensitivity or degree of risk exposure characterizes vulnerable supply chains whose structures, characteristics of their supply chains, internal and external environments, as well as their processes, infrastructures, and processes, are considered by researchers to be the very background or factors of their vulnerability [22, 23]. And of which we assume that the understanding of vulnerability requires an understanding of the sources of risk; this reveals that there is a close relationship between the two concepts of risk and vulnerability. In parallel, we agree with the conclusions of the authors [5, 24], that supply chain resilience increases as vulnerability decreases with a positive effect on supply chain performance, contributing through balanced resilience to the sustainability of logistics operations at all levels. This can be complemented by the ability to properly assess the vulnerability profile in order to compensate it with the appropriate resilient capacity necessary to improve sustainability, at a cost-benefit ratio, balanced without eroding logistical performance [16]. So, the seventh and eighth hypotheses will be: more supply chains context is characterized by increased vulnerability, the more resilient capacities implemented to make supply chains viable are challenged (H7). More supply chains context is characterized by increased vulnerability, the more sustainable capacities implemented to make supply chains viable are challenged (H8).

2.5 Sustainable Supply Chain

The sustainability of the supply chain as a field of reflection has developed remarkably over the past two decades with a particular focus on the sustainable design of operations and supply chains. Unlike traditional ones that focus on economic, financial, and logistics performance, sustainable supply chains extend beyond the economic dimension to integrate social and environmental ones or what is briefly called the Triple Bottom line (TBL) in planning design, and whose holistic approaches that reflect these three dimensions remain relatively limited in the literature. Currently, the issue of adaptability and complementarity in the design of sustainable and resilient supply chains remains an interesting new avenue for research in the sense of deploying more viable chains in which performance, sustainability, and resilience remain stable in the face of disruptions, also balanced by resilient capacities to keep vulnerability at manageable levels [16]. One of the problems justified in our study is the interconnectedness of performance, risk management, and sustainability in our field of research [16] and whose potential causal links between risk, vulnerability, resilience, and sustainability we wish to assess by arguing that sustainability ensures business continuity and reduces long-term risks.

2.6 Theory and Conceptual Model

We adapt three theoretical domains that are still inspiring researchers based on the theory of normal accidents (NAT), high reliability theory (THR), and resource dependence theory (DVR) that are useful to explain why the active association of the two interfaces of resilience and sustainability can absorb the effects of potential disruptions caused by vulnerability factors, thus stabilizing levels of sustainability, resilience, and performance on systems characterized by interactive complexity and close coupling. Due to the nature and structure of systems, failures represent unexpected but common defects in out-of-control operation resulting from the interaction of interface dependency and the interactive complexity of supply chains. In this sense, efforts to design sustainable supply chains taking into account certain sustainability measures to increase the reliability of system security are of great importance because of the inherent ability of supply chain structures and complexity to prevent failure detection and control. This idea is of crucial importance given that sustainable and resilient practices focused on efficiency must provide a balance between these resilient capabilities that ensure greater reliability and vulnerability factors due to the complexities and dynamic nature of the disruptions that influence the validity of risk management tools. This viability balance [5, 16], synonymous with the security reliability of systems, will improve the detectability of failures, and the controllability on potential new sources of risks due to interactive complexity or dependence on other interfaces of the system, thus reducing uncertainty in the decision-making process and ensuring the sustainability of the competitive advantage that guides strategic action throughout the logistics system.

3 Research Objective and Methodology

3.1 Research Objective

Several studies in the academic and professional fields emphasize the dynamic extent of sustainable and resilient practices in complex business contexts, enabling supply chain managers to defend themselves and recover quickly in the face of the complexity and dynamism that characterize both the nature of risks and the structure of supply chains. So, our ultimate objective is to propose some premises on the fertile, possible links between risk management and supply chain sustainability. Otherwise, the viable resilience of supply chains, at which resilience, sustainability, and performance remain unchanged at acceptable levels in the presence of potential disruptions by supporting performance (logistics, operational, and social) while minimizing the spread, severity of risks through regulated but surely balanced resilient and sustainable practices. Based on these findings, we have developed a conceptual research model that attempts to examine the links between the concepts of supply chain risk management (risk, vulnerability, and resilience), supply chain sustainability, and supply chain performance. Through the 17 hypotheses, built on the basis of the relevant theoretical frameworks (normal accident theory; resource dependence theory, and high reliability of organizations theory), we seek to validate the meanings of influences considered important.

3.2 Research Methodology

Through a hypothetical-deductive approach, we seek to validate the 17 hypotheses that constitute the present conceptual research model. For this purpose, a five-part questionnaire survey on concepts related to risk and sustainability management in the supply chain was conducted: risk, vulnerability, resilience, sustainability, and performance. Participants were asked for their opinions based on their expertise in situations concerning their logistics operations over the past 4 years, using two types of the five- and seven-point Likert scales: the first from [15] "Strongly disagree" to [25] "Agree" for the concepts performance and sustainability of the supply chain, and the second from [15] "Strongly disagree" to [26] "Strongly agree" for the concepts risk, vulnerability, and resilience in the supply chain, to be discussed in more detail in the next section (Table 1).

Table 1 Construct origins and natures

Variable natures	Variable origins	Types
Supply risk	[19, 23, 25]	Exogenous
Demand risk	[16, 19]	Exogenous
Infrastructural risk	[23]	Exogenous
Transport risk		Exogenous
Supply chain resilience	[6, 16, 21]	Exogenous
Supply chain vulnerability	[22]	Exogenous
Supply chain performance		Exogenous
Sustainable supply chain	[9, 10]	Endogenous

3.3 Justification for PLS-SEM Choice

Because of the exploratory and confirmatory nature of our study, which focuses on the cause-and-effect relationships that can express and predict the relationships between risk management concepts (risk, vulnerability, and resilience) and sustainability in the supply chain, with a particular focus on supply chain performance that are not adequately explored empirically. As a result, the mobilized conceptual model consists of nine latent variables of a reflective and formative nature, making it more complex, with moderation and mediation analyses for the indirect relationships necessary for the development of the theory that combines risk management and sustainability in the supply chain, tested via the hospital and pharmaceutical sector on a sample of 109 institutions and companies at national level, which represents a fairly homogeneous sample despite its small size. These results are consistent with related studies in supply chain management other than other management disciplines. So, we chose the SEM-PLS approach for this study because it is adapted both to the characteristics of the complex conceptual model and to the size of our sample.

4 Analysis and Results

The research model mentioned in Fig. 1, which latent variables micro-macro risks related to the supply chain (micro supply risk, micro demand risk, micro transport risk, micro infrastructure risk, and macro ecological risk), resilience, vulnerability, and sustainability in the supply chain are all modeled as reflective measurement variables. Except that of performance in the supply chain is modeled by variables and formative measures. So, to operationalize the scales of the conceptual research model, SmartPLS version 3 software is used on a database of 109 institutions and organizations whose characteristics are thus collected (Table 2).

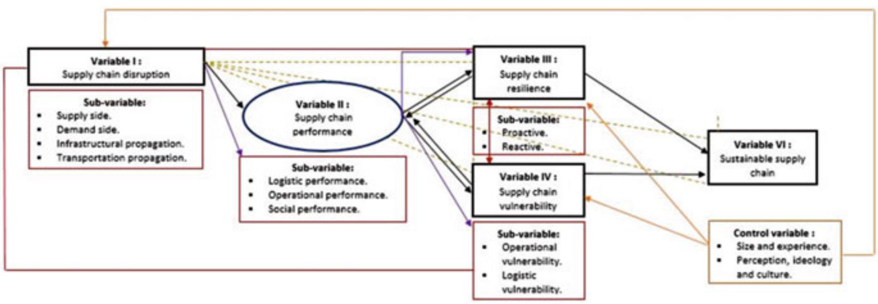


Fig. 1 Research model

Table 2 Demographic characteristics of hospital sample

Offerors	Number of respondents	Percentage	Cumulation
Director of establishment or organization	12	11%	11%
Administrative manager	43	39%	50%
Purchasing manager	23	21%	72%
Procurement and inventory managers	15	14%	85%
Others	16	15%	100%
Total	109		100%

4.1 Measurement Model Validity

4.1.1 Convergent Validity of Reflective Measurements

We begin our analysis by assessing the internal consistency and convergent validity of the measurement model through the loads of the composite measurements, the CR composite reliability, and the mean variance extracted from AVE and Goldstein Rho-A index of reflective items whose estimates are grouped in Table 3 or the composite reliability values and the Rhô-A index for each variable are greater than 0.7. To verify the correlation between the items measuring a construct on the one hand, and the constructed items to be measured on the other hand, the average variance extracted from AVE is used that remains above the 0.5 threshold whose results are acceptable, which suggests a high convergent validity of the reflective measures.

4.1.2 Discriminatory Validity

Following the acceptable level of convergent validity, we can continue the evaluation of the measurement model by verifying the absence of correlation between the items of the constructs, i.e., we try to verify that the items are well represented on their constructs to which they belong. As a result, indicators such as the Fornell-Larcker criterion, the heterotrait-monotrait (HTMT) ratio, and of course cross loading are used. The authors indicate that each latent variable must display a higher value on its line than the remains of the constructed ones but also on its column for items of the same constructed one. Thus (Tables 4, 5, and 6) successively display higher values for each construct between 0.86 and 0.71 for the Fornell-Larcker criterion measuring the correlation between the constructs and 0.85 and 0.36 for the HTMT ratio, as well as higher cross loading values for each measure of the same construct with satisfactory results. Thus, all these conditions together prove the discriminating validity of our external reflective measurement model.

Table 3 Measurement model results—converging validity

Constructs	Measures	Factor loading	Composite reliability	rho-A	Variance average extracted
Supply chain disruption	Disruption 1	0.646	0.878	0.874	0.543
	Disruption 2	0.701			
	Disruption 3	0.546			
	Disruption 4	0.581			
	Disruption 5	0.715			
	Disruption 6	0.67			
	Disruption 7	0.74			
	Disruption 8	0.512			
	Disruption 9	0.681			
Supply chain performance	Performance 1	0.904	0.935	0.929	0.646
	Performance 2	0.783			
	Performance 3	0.828			
	Performance 4	0.788			
	Performance 5	0.724			
	Performance 6	0.874			
	Performance 7	0.792			
	Performance 8	0.717			
Supply chain resilience	Resilience 1	0.713	0.93	0.938	0.695
	Resilience 2	0.622			
	Resilience 3	0.793			
	Resilience 4	0.838			
	Resilience 5	0.773			
	Resilience 6	0.762			
	Resilience 7	0.752			
	Resilience 8	0.713			
	Resilience 9	0.792			
	Resilience 10	0.514			
	Resilience 11	0.719			
	Resilience 12	0.548			
	Resilience 13	0.764			
Supply chain vulnerability	Vulnerability 1	0.685	0.916	0.933	0.625
	Vulnerability 2	0.811			
	Vulnerability 3	0.729			
	Vulnerability 4	0.801			
	Vulnerability 5	0.714			
	Vulnerability 6	0.694			
	Vulnerability 8	0.725			
	Vulnerability 9	0.816			
	Vulnerability 10	0.723			

(continued)

Table 3 (continued)

Constructs	Measures	Factor loading	Composite reliability	rho-A	Variance average extracted
Sustainable supply chain	Sustainability 1	0.882	0.931	0.912	0.731
	Sustainability 2	0.816			
	Sustainability 3	0.891			
	Sustainability 4	0.846			
	Sustainability 5	0.838			

4.2 Internal Measurement Model Evaluation

Once the external measurement model is satisfied, the next step is to evaluate the internal structural model. Then, to test the hypothetical relationships formulated on the research model, start by evaluating the importance of path coefficients using bootstrap techniques, thus specifying for each hypothesis the standard beta, standard deviation, and the *T* and *P* value. We evaluate both direct and indirect relationships on endogenous variables, through the total effect and the specific indirect effect, so predictive power is measured by the determination coefficient R2, the size effect f2, and the cross-redundancy coefficient Q2 to validate the research model. For the rest of our evaluation, we use the bootstrap method recommended by the authors Preacher and Hayes, 2008, to test mediation; this nonparametric statistical procedure samples several times all the data to calculate indirect effects, in the same sense a moderation test between latent variables is used using the procedures of Aiken et al. [27], Bearman and Dawson [28], and Dawson and Richter [29] to interpret and plot the effects of bidirectional interactions. Towards the end of our structure model, an unobserved heterogeneity analysis of the data structure remains essential to reinforce the cause-prediction meanings of our results and to avoid the risk that they will not be significant on aggregate data. Table 7, indicate the values relating to the path coefficient, total effect, indirect special effect of the assumptions of our model the determination coefficient R2, the cross redundancy Q2, effect size f2, and of course the GoF adjustment quality criterion. For the coefficients, the literature suggests that they must be significant with *t* and *p* value lower than 0.05–0.1, to evaluate the model's ability to predict latent variables that the determination coefficient, effect size f2, and cross redundancy Q2 are calculated by successively displaying values between 0.19 and 0.67 for R2, 0.02 and 0.35 for f2, and above zero for Q2. The analysis shows that 14 of the 17 hypotheses are validated (H1 to H7, H9, H11 to H16) and that the five hypotheses are rejected (H8, H10, and H17). Illustrate the values for the path coefficient, the total effect, the indirect special effect, the determination coefficient R2, the effect size f2, the cross redundancy Q2, and of course the GoF adjustment quality criterion. They range from 0.19 to 0.67 for R2, from 0.02 to 0.35 for f2, and above 0 for Q2, which corresponds to quite satisfactory values. For GoF, the conceptual model displays a value that is

Table 4 Discriminant validity—cross loading

Construct	Prfm SCM	Res SCM	Disrp SCM	Sus SCM	Vul SCM
Performance 1	**0.904**	0.323	0.188	0.679	−0.296
Performance 2	**0.782**	0.161	0.099	0.508	−0.105
Performance 3	**0.83**	0.306	0.149	0.548	−0.213
Performance 4	**0.784**	0.03	0.106	0.549	−0.012
Performance 5	**0.718**	0.325	0.111	0.502	−0.219
Performance 6	**0.875**	0.312	0.268	0.692	−0.436
Performance 7	**0.797**	0.121	0.02	0.68	−0.251
Performance 8	**0.718**	0.253	0.079	0.564	−0.171
Resilience 1	0.34	**0.706**	−0.133	0.237	−0.396
Resilience 2	0.163	**0.619**	−0.009	0.215	−0.305
Resilience 3	0.423	**0.788**	0.034	0.493	−0.436
Resilience 4	0.227	**0.834**	−0.075	0.317	−0.406
Resilience 5	0.293	**0.772**	0.022	0.316	−0.509
Resilience 6	0.35	**0.757**	0.11	0.369	−0.436
Resilience 7	0.165	**0.751**	−0.04	0.16	−0.421
Resilience 8	0.079	**0.722**	−0.058	0.224	−0.528
Resilience 9	0.319	**0.791**	0.009	0.344	−0.377
Resilience 10	−0.07	**0.551**	0.206	0.201	−0.248
Resilience 11	0.125	**0.721**	0.012	0.199	−0.391
Resilience 12	−0.077	**0.56**	0.026	0.158	−0.315
Resilience 13	0.152	**0.768**	−0.047	0.263	−0.493
Disruption 1	0.242	0.112	**0.693**	0.172	−0.053
Disruption 2	−0.016	0.064	**0.711**	0.049	−0.208
Disruption 3	0.078	0.124	**0.754**	0.058	0.002
Disruption 4	−0.001	−0.135	**0.711**	0.066	0.093
Disruption 5	−0.0.001	−0.135	**0.711**	0.066	0.093
Disruption 6	0.257	−0.042	**0.831**	0.243	−0.03
Disruption 7	−0.106	−0.246	**0.699**	−0.179	0.053
Disruption 8	−0.016	0.089	**0.701**	0.049	−0.214
Disruption 9	0.075	0.135	0.478	0.059	−0.005
Sustainability 1	0.709	0.401	0.058	**0.882**	−0.308
Sustainability 2	0.589	0.306	−0.014	**0.816**	−0.32
Sustainability 3	0.611	0.349	0.03	**0.891**	−0.374
Sustainability 4	0.686	0.313	0.134	**0.846**	−0.267
Sustainability 5	0.564	0.271	−0.022	**0.838**	−0.393
Vulnerability 1	−0.152	−0.409	0.18	**−0.116**	0.722
Vulnerability 2	−0.197	−0.365	0.049	**−0.215**	0.809
Vulnerability 3	0.042	−0.347	0.145	**−0.245**	0.719
Vulnerability 4	−0.299	−0.637	0.216	**−0.339**	0.8
Vulnerability 5	−0.267	−0.395	0.029	**−0.281**	0.721
Vulnerability 6	−0.176	−0.279	0.051	**−0.25**	0.695
Vulnerability 8	0.107	−0.218	0.059	**−0.059**	0.478
Vulnerability 9	−0.294	−0.32	0.095	**−0.393**	0.729
Vulnerability 10	−0.349	−0.571	−0.096	**−0.432**	0.821

Table 5 Latent variables correlation—Fornell-Larcker criterion

Construct	Perfm SCM	Disp SCM	Res SCM	Sus SCM	Vul SCM
Performance	**0.804**	0	0	0	0
Disruption	0.173	**0.791**	0	0	0
Resilience	0.334	−0.029	**0.787**	0	0
Sustainability	0.742	0.145	0.395	**0.855**	0
Vulnerability	−0.301	−0.024	−0.582	−0.394	**0.744**

Table 6 Discriminant validity—heterotrait-monotrait ratio (HTMT)

Construct	Perfm SCM	Disp SCM	Sus SCM	Vul SCM
Performance	1	0	0	0
Disruption	**0.715**	1	0	0
Resilience	0.335	**0.75**	0	0
Sustainability	0.801	0.588	1	0
Vulnerability	0.313	0.314	**0.579**	1

also consistent with the literature, since traditionally, if the GoF value is less than 0.1, the model is strictly rejected between 0.1 and 0.25, the quality of the model is considered poor; between 0.25 and 0.36, the quality of the model is considered average; and above 0.36, the quality of the model is considered strong which is the case here, and in which the literature suggests that R2 values (0.19–0.33–0.67) are significantly (low–medium–high). According to the same author, the values of f2 (0.02–0.15–0.35) are considered as (low–medium–strong), which must be greater than zero for the accepted model.

5 Conclusion

This research aims to assess the causal relationship between risk management, performance and sustainability in supply chain with a particular focus on the complementarity of resilience and sustainability practices in supply chain. In this research, structural equation modeling and partial least square regression analysis were designed to assess the association links between risk, vulnerability, resilience, sustainability and performance concepts. Seventeen hypotheses have been tested as part of our conceptual model and the results of which justify that it has indirect associations between risks (of supply, demand and infrastructure); vulnerability; resilience; performance and sustainability of the supply chain. We agree with author Ivanov on the potential of this research topic, which requires a particular intention by continuing to assess the dimensions and interfaces of resilience and sustainability in other business contexts. And to adopt a new approach to risk management practices based on a compromise between sustainability and resilience such as Shopfloor and/or Time driven activity based costing using real data from hospitals in Morocco.

Table 7 Path coefficient of research hypotheses

Hyp	Relationship	Standard beta	Standard error	*T* value	*P* value	Decision
H01	Supply risk–performance	−3.586	1.418	2.529	0.012	Accepted
H02	Demand risk–performance	−2.944	1.245	2.365	0.018	Accepted
H03	Infrastructural risk–performance	−0.853	0.45	1.893	0.059	Accepted
H04	Transport risk–performance	−2.801	2.51	2.311	0.021	Accepted
H00	Disruption–performance	−10.029	4.248	2.361	0.019	Accepted
H05	Performance–vulnerability	−0.33	0.187	1.761	0.079	Accepted
H06	Performance–resilience	0.35	0.151	2.317	0.021	Accepted
H07	Performance–sustainability	0.765	0.087	8.765	0	Accepted
H08	Disruption–vulnerability	−8.367	4.795	1.745	0.082	Rejected
H09	Disruption–resilience	9.211	4.965	1.855	0.064	Accepted
H10	Disruption–sustainability	−1.832	2.06	0.889	0.374	Rejected
H11	Resilience–sustainability	0.145	0.152	0.957	0.339	Accepted
H12	Vulnerability–resilience	−0.567	0.136	4.18	0	Accepted
H13	Vulnerability–sustainability	−0.225	0.114	1.981	0.048	Accepted

References

1. Ivanov, D., Mason, S. J., & Hartl, R. (2016). Supply chain dynamics, control and disruption management. *International Journal of Production Research, 54*(1), 1–7.
2. Christopher, M., & Holweg, M. (2011). "Supply chain 2.0": Managing supply chains in the era of turbulence. *International Journal of Physical Distribution and Logistics Management, 41*(1), 63–82.
3. Tuncel, G., & Alpan, G. (2010). Risk assessment and management for supply chain networks: A case study. *Computers in Industry, 61*(3), 250–259.
4. Zhao, L., et al. (2013). The impact of supply chain risk on supply chain integration and company performance: A global investigation. *Supply Chain Management: An International Journal, 18*(2), 115–131. https://doi.org/10.1108/13598541311318773
5. Pettit, T. J., Croxton, K. L., & Fiksel, J. (2013). Ensuring supply chain resilience: Development and implementation of an assessment tool. *Journal of Business Logistics, 34*(1), 46–76. https://doi.org/10.1111/jbl.12009
6. Chowdhury, M. M. H., & Quaddus, M. (2017). Supply chain resilience: Conceptualization and scale development using dynamic capability theory. *International Journal of Production Economics, 188*, 185–204. https://doi.org/10.1016/j.ijpe.2017.03.020

7. Zsidisin, G. A., & Wagner, S. M. (2010). Do perceptions become reality? The moderating role of supply chain resiliency on disruption occurrence. *Journal of Business Logistics, 31*(2), 1–20. https://doi.org/10.1002/j.2158-1592.2010.tb00140.x
8. Manuj, I., Esper, T. L., & Stank, T. P. (2014). Supply chain risk management approaches under different conditions of risk. *Journal of Business Logistics, 35*(3), 241–258. https://doi.org/10.1111/jbl.12051
9. Das, D. (2017). Development and validation of a scale for measuring sustainable supply chain management practices and performance. *Journal of Cleaner Production, 164*, 1344–1362. https://doi.org/10.1016/j.jclepro.2017.07.006
10. Hong, J., Zhang, Y., & Ding, M. (2018). Sustainable supply chain management practices, supply chain dynamic capabilities, and enterprise performance. *Journal of Cleaner Production, 6526*(17), 31263–31265.
11. Pereira, C. R., Christopher, M., & Da Silva, A. L. (2014). Achieving supply chain resilience: the role of procurement. *Supply Chain Management: An International Journal, 19*(5/6), 626–642. https://doi.org/10.1108/SCM-09-2013-0346
12. Gligor, D. M., Esmark, C. L., & Holcomb, M. C. (2015). Performance outcomes of supply chain agility: When should you be agile? *Journal of Operations Management, 33–34*, 71–82. https://doi.org/10.1016/j.jom.2014.10.008
13. Sodhi, M. S., Son, B. G., & Tang, C. S. (2012). Researchers' perspectives on supply chain risk management. *Production and Operations Management, 21*(1), 1–13. https://doi.org/10.1111/j.1937-5956.2011.01251.x
14. Gligor, D. M., & Holcomb, M. C. (2012). Antecedents and consequences of supply chain agility: Establishing the link to firm performance. *Journal of Business Logistics, 33*(4), 295–308.
15. Ivanov, D., Sokolov, B., & Dolgui, A. (2014). The ripple effect in supply chains: Trade-off 'efficiency-flexibility-resilience' in disruption management. *International Journal of Production Research, 52*(7), 2154–2172.
16. Ivanov, D. (2017). Revealing interfaces of supply chain resilience and sustainability: A simulation study. *International Journal of Production Research*, 1–17. https://doi.org/10.1080/00207543.2017.1343507
17. Tang, C., & Tomlin, B. (2008). The power of flexibility for mitigating supply chain risks. *International Journal of Production Economics, 116*(1), 12–27. https://doi.org/10.1016/j.ijpe.2008.07.008
18. Craighead, C. W., Blackhurst, J., Rungtusanatham, M. J., & Handfield, R. B. (2007). The severity of supply chain disruptions: Design characteristics and mitigation capabilities. *Decision Sciences, 38*(1), 131–156. https://doi.org/10.1111/j.1540-5915.2007.00151.x
19. Thun, J. H., & Hoenig, D. (2011). An empirical analysis of supply chain risk management in the German automotive industry. *International Journal of Production Economics, 131*(1), 242–249. https://doi.org/10.1016/j.ijpe.2009.10.010
20. Tang, C. S. (2006). Perspectives in supply chain risk management. *International Journal of Production Economics, 103*(2), 451–488. https://doi.org/10.1016/j.ijpe.2005.12.006
21. Ponomarov, S. Y., & Holcomb, M. C. (2009). Understanding the concept of supply chain resilience. *The International Journal of Logistics Management, 20*(1), 124–143. https://doi.org/10.1108/09574090910954873
22. Wagner, S. M., & Neshat, N. (2010). Assessing the vulnerability of supply chains using graph theory. *International Journal of Production Economics, 126*(1), 121–129. https://doi.org/10.1016/j.ijpe.2009.10.007
23. Wagner, S. M., & Bode, C. (2008). An empirical examination of supply chain performance along several dimensions of risk. *Journal of Business Logistics, 29*(1), 307–325. https://doi.org/10.1002/j.2158-1592.2008.tb00081.x
24. Pettit, T. J., Fiksel, J., & Croxton, K. L. (2010). Ensuring supply chain resilience: Development of a conceptual framework. *Journal of Business Logistics, 31*(1), 1–21. https://doi.org/10.1002/j.2158-1592.2010.tb00125.x

25. Manuj, I., & Mentzer, J. T. (2008). Global supply chain risk management strategies. *International Journal of Physical Distribution and Logistics Management, 38*(3), 192–223.
26. Tang, S. Y., & Gupta, H. G. D. (2014). Managing disruptions in decentralized supply chains with endogenous supply process reliability. *Production and Operations Management, 23*(7), 1198–1211.
27. Aiken, L. S., West, S. G., & Reno, R. R. (1991). Multiple regression: Testing and interpreting interactions. *Sage*.
28. Bearman, M., & Dawson, P. (2013). Qualitative synthesis and systematic review in health professions education. *Medical Education, 47*, 252–260.
29. Dawson, J. F., & Richter, A. W. (2006). Probing three-way interactions in moderated multiple regression: Development and application of a slope difference test. *Journal of Applied Psychology, 91*(4), 917–926. https://doi.org/10.1037/0021-9010.91.4.917

Big Data Analytics Opportunities and Challenges for the Smart Enterprise

Brahim Jabir, Falih Noureddine, and Khalid Rahmani

Abstract Today, the real challenge is examining large and varied data sets to get the value in order to uncover information including hidden patterns, unknown correlations, market trends, and customer preferences that can help organizations make informed business decisions. Big data analytics is a form of advanced analytics, which involves complex applications with elements such as predictive models, statistical algorithms, and what-if analysis powered by high-performance analytics systems, but the problem is the field still suffering from gaps and problems, and there is no a complete framework that can achieve the business objectives. In this chapter, we present a literature review, explain how big data analytics helps improving business objectives, and describe some analytics systems used in this goal.

Keywords Business analytics · Big data analytics · Smart enterprise

1 Introduction

Organizations handle and collect large volume of data and try to exploit it to achieve their business goals. So that makes data pass through several organizational, strategical, and procedural stages. Over the years, various applications and approaches [1] appeared and helped to collect and analyze the right data to make the right decisions.

Lately big data analytics comes as a new strategy of analyzing large amount of data, or big data [2]. This big data is gathered from a wide variety of sources, including social networks, videos, digital images, sensors, and sales transaction records, that might provide valuable insights. Through this insight, businesses may be able to gain an edge over their rivals and make superior business decisions. This

B. Jabir · F. Noureddine
LIMATI Laboratory, Polydisciplinary Faculty Mghila, BP 592, BeniMellal, Morocco

K. Rahmani (✉)
ERPTM Laboratory, Polydisciplinary Faculty Mghila, BP 592, BeniMellal, Morocco

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7_70

chapter highlights the challenges and pressures that the healthcare systems face, identifies its opportunities, and discusses the important role of analytics and its framework.

2 Business Analytics

Business analytics is an advanced concept emerged from business intelligence. It is an approach combining different disciplines to extract value from data in order to make a valuable decision [3] via regular analysis with different plans and strategies. So, BA is used by organizations for the aim to develop existing processes, identify new opportunities, discover more product features, change and evolve new services and systems, better understand the behavior of customers, and expect problems before they happen. These disciplines are computer science, statistics, data management, decision science, and scientific research methods.

2.1 Existing Type of Analytics

To better explain this concept of analytics, it is necessary to understand it through its three types [4], which shorten its role and purposes:

2.1.1 Descriptive Analytics

Descriptive analytics is a type of analytics known as business reporting, provides an interpretation and extrapolation of historical data, and helps to understand the significant change in the enterprise [5]. Its main result is making the raw data understandable for the various components of the organization (employees, managers, investors, supplier, etc.), this type enables the company to answer the questions of “what happened” and/or “what happening” [6] like:

- How many stocks have been delivered last year?
- What is the medium sales volume for the last year?
- What is the kind of the products returned for last month?
- What are the best-selling products?
- How many customers purchased last month?
- How much paid for the general costs last year?

This analytics type uses many techniques and tools such as data mining and data aggregation to provide information, create a report of available data, and prepare it for further processing in order to provide insights and predictions, so that to help understand why and how some event happened and explain why some results occur, all while trying to improve employee engagement and productivity.

2.1.2 Predictive Analytics

Predictive analytics is a branch of analytics comes as a kind of analytic modeling and requires several statistical tools that can analyze current and historical events to provide insights and make predictions about unknown events about future [7].

Experts use this type to deploy future business planning to predict the problems before they happen, discover new services and more opportunities to reduce time, increase productivity, and minimize risks. Its principal outcome is to answer the question of "what will happen?" or/and "why will it happen?" Examples are:

- Who is the most likely employee to leave our organization?
- What is the risk of losing money on new project investment?
- What will be the revenue if sales service decreases by X percent?
- What will be the revenue in case of a boycott for an X time?
- What will happen if supplier prices grow by an X percent?
- What do we expect to pay for X services over the next year?

By answering these questions, the enterprise examines the results to detect new patterns and links to enhance their performance through its different business areas, operations, finance, and marketing.

2.1.3 Prescriptive Analytics

This analytic is the final step of the analytics process. It defines the actions to be taken to avoid future risk or to take full advantage of a promising trend. It uses also historical data and external information due to the nature of statistical algorithms to identify opportunities and identify the reasons behind failure or success. Prescriptive analytics uses sophisticated tools and technologies, like machine learning, business rules, and algorithms. It answers the question of "what I should do?" and/or "why should I do it?" [8]. Examples are:

- What is the alternative plan to maintain maximum profit if X employee leaves?
- How many products do we need to sell to maximize revenue?
- What is the best way to minimize costs and fees?

The answers obtained by this processing help the organization to set new criteria for success and failure in order to reconstruct the business with reliable predictions to develop efficiency and reduce costs.

Figure 1 shows the analysis levels of the business analytics with its three types, and it clarifies that descriptive analytics provide insights into the past. Predictive analytics helps to understand the future and prescriptive analytics to advise on possible outcomes.

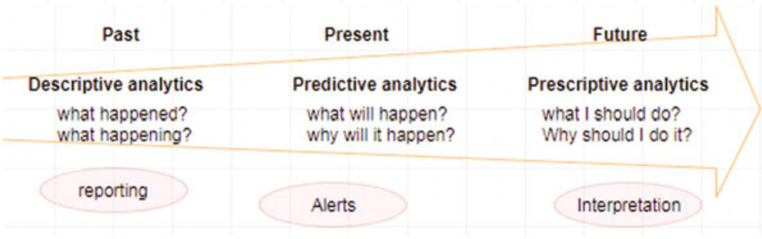


Fig. 1 Business analytics types view

3 Big Data Overview

3.1 What Is Big Data?

From all existed definitions for "big data," we can choose that big data is data that is "too big," "too hard," and "too fast" [9]. "Too big" indicates the large available quantities of data, known as "volume." "Too hard" relates to the different type and several format of data and also a variety of sources, which need some kind of analysis and some kind of tools to process, that is known as "variety." "Too fast" means that data need a high speed to process and that refers to "velocity." These three Vs are the big data dimensions and attributes (volume, velocity, variety) [10] (Fig. 2).

3.2 Defining Big Data Analytics

According to the big data research article [2], the big data analytics is the point where advanced analytics technique has been emerged and runs in the big data to process data for extracting value [11]. This analytics adoption drive potential benefits, and it is the key for the enterprises to exploit data-dependent capabilities and provide insight and direction to enhance and help making decision. Big data analytics unites different tools and disciplines using statistical algorithms, predictive models, data mining, and various other tools, in order to provide valuable information from a large amount of data (Fig. 3).

4 Smart Enterprise

Mellote affirms that smart enterprise is an organization use and is a kind of exploiting technologies to answer key challenges, and also to manage and extract meaning from the diversity and volume of data that is available to them. A second

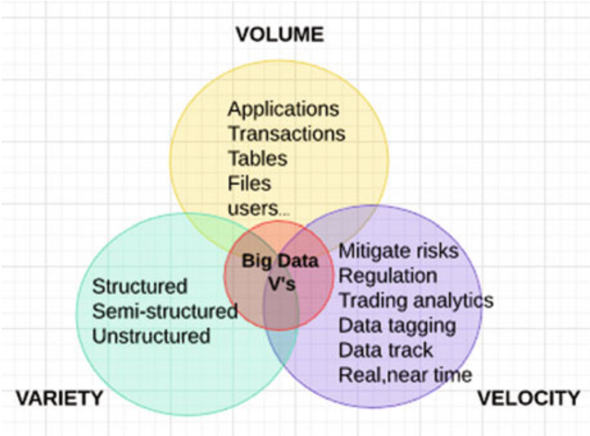


Fig. 2 The three Vs of big data

Fig. 3 Big data analytics for enterprise

group of definitions [12] emphasis that "smart enterprise" is an organization that offers an integration of all recent analytics technological advances to acquire, transfer, interpret, and analyze the information. A third definition focuses on the insight aspect and accord that smart enterprise is an organization that embeds analytics to transform information into insights and predictions and then into action. From all those definitions, we can conclude that smart enterprise is a new enterprise performance optimization strategy, enable some methods and approaches defined under a solution which is big data analytics and come to turn data into value in order to make a decision and affects all enterprise areas to enhance their competitiveness (Fig. 4).

4.1 Enterprise Big Data Sources

The available information varies significantly in its volume, its format, and the speed depending on the type of the company [13]. It is necessary that information combine data typically managed by the HR department, customer satisfaction, and operational data. This data expanded, linked, and analyzed with several tools to find what is really happening in the organization and discover what will happen and what should the organization do. Some examples of the types of data are described in Table 1.

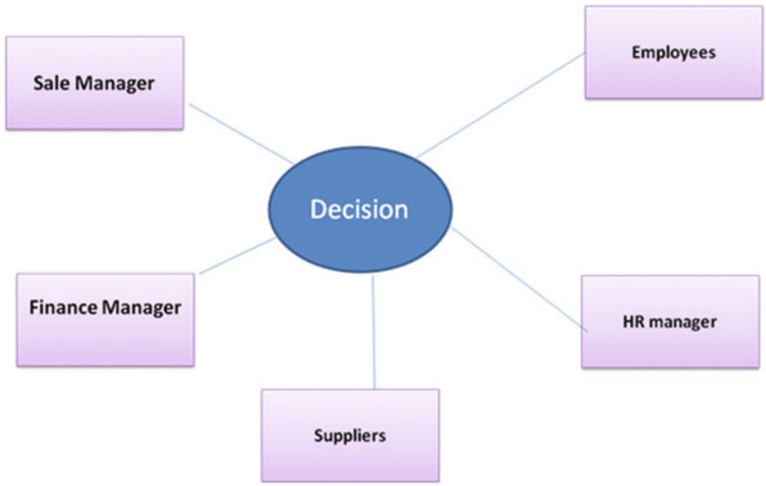


Fig. 4 Making decision in enterprise areas

5 Challenges and Opportunities of Big Data Analytics

Big data analytics has become essential as it helps in improving business, decision makings and providing the significant advantage over the competitors. This applies for organizations as well as professionals in the analytics field. For smart enterprises, who are skilled in big data analytics, there is a wide range of opportunities out there.

Decision-making is a major challenge [14] given the many changes that have occurred in all areas of the enterprise; to this end, big data analytics has been emerged as a set of strategies, tools, and methods to facilitate decision-making. It makes enterprise ensure the following benefits:

- Better targeted social influencer marketing.
- More numerous and accurate business insights.
- Segmentation of customer base.
- Recognition of sales and market opportunities.
- Automated decisions for real-time processes.
- Definitions of churn and other customer behaviors.
- Detection of fraud.
- Greater leverage and ROI for big data.
- Quantification of risks.
- Trending for market sentiments.
- Understanding of business change.
- Better planning and forecasting.
- Identification of root causes of cost.
- Understanding consumer behavior from click-streams.

Table 1 Several information sources used by enterprise

Information source	Description	Examples
HR database	Data collections contain information about employees, customers, products, etc., such as employee personal details, performance, diversity data, and promotion details	Database: Oracle, SAP, etc. Information: Age, gender, salary, department, performance rating, sickness absence, location, team, price, etc.
Employee attitude survey data	A range of information usually stored in survey programs and exported to files contains the attitude of employees and their engagement data, usually managed by providers organization	Job strain level, employee engagement, employee performance, satisfaction, perception of justice, stress level, etc.
Customer satisfaction survey data	Also stored in survey programs, provides information about customers' preferences, customers' experiences, customers' satisfaction, etc.	Customer rating Customer loyalty Preferences Satisfaction Purchases Likelihood of further business
Sales performance data	Information usually owned by the sales function, recording details of sales performance and revenues. It is a useful information that help to determine how the organization reached the business goal	Sales of month New purchases Revenue attained Best-selling Products' characteristics
Operational performance data	Information refers to the efficiency of the organization. It is about measuring the successful running of the business	Number of complaints resolved Number of calls dropped out Number of queries resolved Time consumed in some operations

- Manufacturing yield improvements.
- Discover new facts about their customers, markets, partners, costs, and operations.

Data generated through big data analytics sources can help companies better understand their performance than previous technologies [15], but there is no doubt that big data presents technical challenges due to its volume, variety, and velocity. Data volume alone is a showstopper for some organizations, and most of them face the very real risk of information overload generated by the different systems, the complexity of data, the lack of experts in this field, and the costs of systems, and there is not yet a complete framework that can offer a solution for this major problems and help enterprise for a sample transformation to be a smart enterprise that adopt analytics to get value from data in order to take the right decision (Fig. 5). So, finding a way to exploit the data at their disposal and leverage

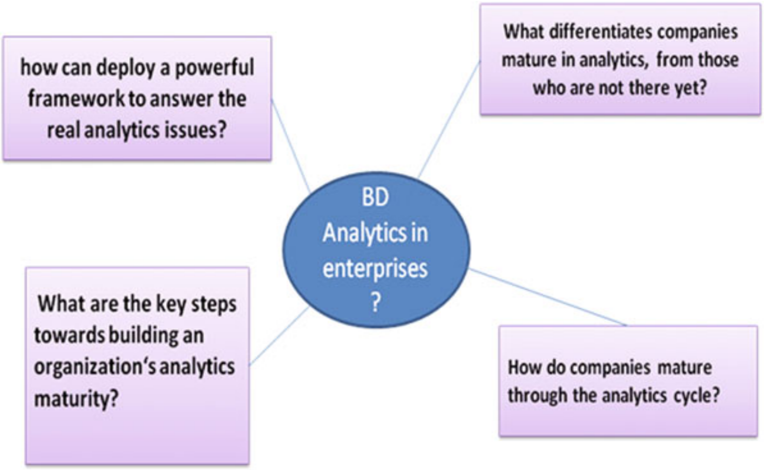


Fig. 5 Big data analytics issues in enterprise

them to improve business and organizational performance becomes a necessity [16], this can only be achieved when analytical tools and techniques are combined and integrated in a structured and rigorous framework, and this combination can identify new opportunities for improvement or suggest innovative ways to address old challenges...

5.1 Big Data Analytics Technique Solutions

In this section, we describe the current analytics solutions that can be a major key to create and deploy an efficient predictive management model. This analytic management model will produce predictions and insights and help HR managers make better decisions for the company. Of course, most of them are expensive but there is a lot of similarity between the packages, and the analysis generally produces the same results. So, all you need is analytic methods and skills, and it is possible to apply them in most other systems. We consolidate in Table 2 the most popular analytics system with some details:

5.2 Components of Big Data Analytics

The levels and layers of abstraction presented in the Fig. 6 show components of big data analytics, forms of stacks, and their integration with each other; in most of the cases, HDFS/Hadoop forms the core of most of the big data-centric applications, but that is not a generalized rule:

Table 2 Popular big data analytics systems with details

Analysis software system	Details
Apache zookeeper	Apache zookeeper [17] is a framework that can federate the communications between distributed systems. It functions by providing a memory space shared by all the instances of servers. These several machines connected to each other solve problems together and process large quantity of data together that accelerates the processing, offers real-time access, and handles system breakdown problems
Apache storm	Apache storm is an open-source project started with the idea of developing a stream processing system came in December 2010, by Nathan Marz and adopted by apache on 17 September 2014 [18], this system offer a real time distributed processing systems that can process the unbounded streams of data very fast than ever [19], it is easy to use and provide low latency with guaranteed data processing, it support communication over a JSON-based protocol. It offers services of filtering, aggregation, join, read/write to and from a several sources
KAFKA	The Kafka project used to build real time data pipelines and streaming applications [20], it is a distributed, partitioned and replicated service, it support parallel data loading provide by a various producers (Frontend web applications, services, adapters...), consumed by real-time consumers (filter and sift information in databases and trigger alert), near real time consumers (save data in any NoSQL system) and offline consumers (storing information in traditional data warehouse for offline analysis). This following diagram shows a typical big data analysis and aggregation scenario supported by the Apache Kafka system
SPARK	Apache spark is a powerful open source framework for smart data processing, developed by AmPlab, in 2009, adopted by apache in 2010 [21], makes sophisticated analysis and designed for speed and ease, it has several advantages over other technologies, it is an efficient solution for smart data analytics offers text and graphics visualization, and supports streaming requests
HADOOP	Apache Hadoop is an open source designed for the distributed processing of large amount data sets across clusters of computers using simple programming models [22], created to scale up from single servers to lots of machines, each offering local storage and computation. Rather than it can deliver high-availability, detect and handle failures at the application layer, it is composed from several components, that work together to process data: HDFS (the distributed file system layer that coordinates storage and replication across the cluster nodes) YARN (the cluster coordinating component of the Hadoop stack), MapReduce (the native batch processing engine for Hadoop)
CASSANDRA	Apache Cassandra is a powerful distributed database of the NoSql family designed to collect a large amount of data from multiple sources [23], data stored is automatically replicated to multiple physical instance which is nodes offering none downtime and offers also a high availability and scalability, architecture of this database constitutes from nodes uses peer-to-peer connection, also from clusters, data centers and a partitions

Original data	Storage systems	Task trackers	Higher level language	Security and management	Modeling	Loading analytic databases	Analytics applications	presentation
Files	Hadoop				ETL modeling tools	e.g., greenplum netezza	Merced,	Reports
external data bases	file system,e.g.,HDFS	MapReduce engine	Hive	Cascading,			ClickFox	dashbords
Surveys	NoSqlDB,e.g.,Hbase Cassandra		Pig	Kerberos				Alerts
business apps								
media								

Fig. 6 Big data analytics components

6 Proposed Big Data Analytics Framework

In socioeconomic world, the decision-making challenge turn around type and quality of data, people, business planning, business objectives, etc., and not only around technological side, that is why it is necessary to think about a comprehensible multidimensional framework that project on all the dimensions that can extract value from large amount of data and help to turn this value into action in order to take the right decision. We proposed a simple framework based on the existed ones that are relatively limited in scope, this framework presents a complete model that combines several dimensions to interact with each other (Fig. 7):

- Actors: Include software developers, personnel, customers, managers.
- Interaction models: Interface between user and system aspects.
- Computing infrastructure: Devices and the software required.
- Communication: Business workflow as a collaboration that requires a significant two-way communication.
- Enterprise content: Information on attitude of employees, personnel details and their performance, promotions details, customers information.
- Data sources: The source systems such data warehouses, supply chain systems, and other operational systems, surveys.
- Internal organizational policies, procedures, and culture: Purchase of hardware and software, data backups, etc.
- External forces: External rules, regulations, and pressures that place constraints or help on the deployment.
- Regular basis measuring and monitoring of the effects of information technology.

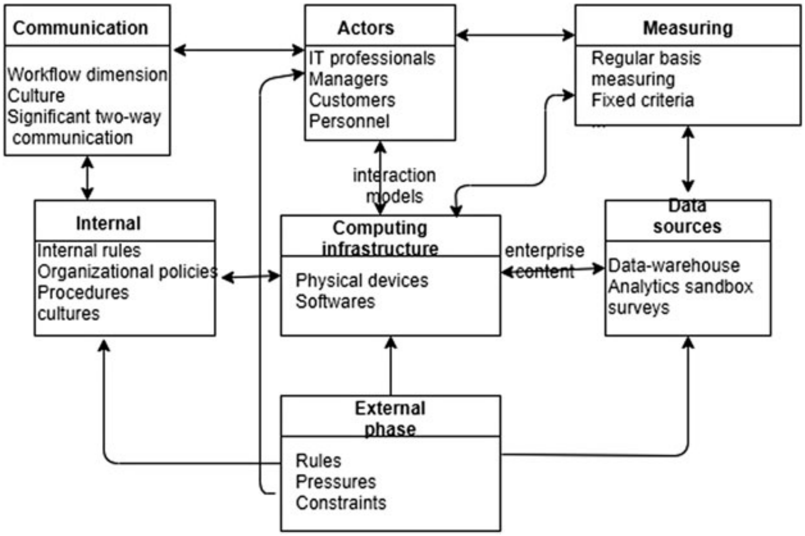


Fig. 7 Proposed analytics framework

6.1 Discussion

6.1.1 Benefits

The model offers the possibility of its hierarchical decomposition; in other words, the possibility of breaking down a complex process, system, or device into its components, this offers the possibility to study them and then integrate the results trying to understand the functioning of complete system. This model can offer the capability to feed management reports and dashboards with deep insight into past, current, and even future performance.

6.1.2 Barriers

There is always a gap between analytics development and the use of analytics within enterprises so the challenge is not to find a complete framework [24], but the analytics deployment is confronted by the following barriers:

- Lack of experienced people that can understand the analytical systems.
- Distrust of the information and gaps to extract correct data.
- Models are expensive and complex to deploy.
- Turning information and insights into decision requires an immense experience.

7 Conclusion

In this chapter, we described big data analytics approach for the company as a set of analysis of a large amount of data that comes to drive business planning and deploy the future business planning as well as a proposed framework that support big data analytics aspect.

Big data analytics generates potential benefits for the company. It is the major key behind the reaching of business goals. For this, our future contribution will be about concretizing this notion of big data analytics by a specific and original approach about scenario modeling, even predicting employee performance.

References

1. Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences, 1*(1), 83–98.
2. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144.
3. Kohavi, R., et al. (2002). Emerging trends in business analytics. *Communications of the ACM, 45*(8), 45–48.
4. Van Barneveld, A., et al. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE Learning Initiative, 1*(1), 1–11.
5. Banerjee, A., et al. (2013). Data analytics: Hyped up aspirations or true potential? *Vikalpa, 38*(4), 1–12.
6. Delen, D., & Demirkan, H. (2013). Data, information and analytics as services. *Decision Support Systems, 55*(1), 359–363.
7. Eckerson, W. W. (2007). Predictive analytics. Extending the value of your data warehousing investment. *TDWI Best Practices Report, 1*, 1–36.
8. Ransbotham, S., et al. (2015). Minding the analytics gap. *MIT Sloan Management Review, 56*(3), 63.
9. Madden, S. (2012). From databases to big data. *IEEE Internet Computing, 3*, 4–6.
10. Jothimani, D., et al. (2014). *A big data analytical framework for portfolio optimization*. Cornell University.
11. Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.
12. Springett, J. (2007). U.S. patent application no. 11/613,847.
13. Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on* (p. 4).
14. Harrison, E. F. (1995). *The managerial decision-making process* (Vol. 4, pp. 1–39). Houghton Mifflin.
15. Marcu, O. C., et al. (2016). Spark versus flink: Understanding performance in big data analytics frameworks. In *CLUSTER computing (CLUSTER), 2016 IEEE international conference on* (pp. 433–442). IEEE.
16. Grigoroudis, E., et al. (2012). Strategic performance measurement in a healthcare organisation: A multiple criteria approach based on balanced scorecard. *Omega, 40*(1), 104–119.
17. Haloi, S. (2015). *Apache ZooKeeper essentials*. Packt Publishing.
18. Iqbal, M. H., & Soomro, T. R. (2015). Big data analysis: Apache storm perspective. *International Journal of Computer Trends and Technology, 19*(1), 9–14.

19. van der Veen, J. S., et al. (2015). Dynamically scaling apache storm for the analysis of streaming data. In *Big data computing service and applications (BigDataService), 2015 IEEE first international conference on* (pp. 154–161). IEEE.
20. Ranjan, R. (2014). Streaming big data processing in datacenter clouds. *IEEE Cloud Computing, 1*(1), 78–83.
21. Xin, R., et al. (2014). *Graysort on apache spark by databricks*. GraySortCompetition.
22. Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.
23. Chebotko, A., et al. (2015). A big data modeling methodology for apache Cassandra. In *Big data (BigData congress), 2015 IEEE International Congress on* (pp. 238–245). IEEE.
24. Williams, I., & Bryan, S. (2007). Understanding the limited impact of economic evaluation in health care resource allocation: A conceptual framework. *Health Policy, 80*(1), 135–143.

Index

M. Elhoseny et al. (eds.), *Distributed Sensing and Intelligent Systems*, Studies in Distributed Intelligence, https://doi.org/10.1007/978-3-030-64258-7