Advances in Intelligent Systems and Computing 1184

Xin-She Yang
Simon Sherratt
Nilanjan Dey
Amit Joshi *Editors*

Proceedings of Fifth International Congress on Information and Communication Technology

ICICT 2020, London, Volume 2

Springer

Advances in Intelligent Systems and Computing

Volume 1184

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within "Advances in Intelligent Systems and Computing" are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink ****

More information about this series at http://www.springer.com/series/11156

Xin-She Yang · Simon Sherratt ·
Nilanjan Dey · Amit Joshi
Editors

Proceedings of Fifth International Congress on Information and Communication Technology

ICICT 2020, London, Volume 2

Springer

Editors
Xin-She Yang
Middlesex University
London, UK

Simon Sherratt
The University of Reading
Reading, UK

Nilanjan Dey
Techno India Institute of Technology
Kolkata, West Bengal, India

Amit Joshi
Global Knowledge Research Foundation
Ahmedabad, Gujarat, India

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-981-15-5858-0 ISBN 978-981-15-5859-7 (eBook)
https://doi.org/10.1007/978-981-15-5859-7

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This AISC volume contains the papers presented at ICICT 2020: Fifth International Congress on Information and Communication Technology in concurrent with ICT Excellence Awards. The conference was held during February 20–21, 2020, London, UK, and collaborated by Global Knowledge Research Foundation, City of Oxford College. The associated partners were Springer, InterYIT IFIP and Activate Learning. The conference was held at Brunel University, London. This conference was focused on e-business fields such as e-agriculture, e-education and e-mining. The objective of this conference was to provide a common platform for researchers, academicians, industry persons and students to create a conversational environment wherein topics related to future innovation, obstacles to be resolved for new upcoming projects, exchange of views and ideas. The conference attracted immense experts from more than 45 countries, the conference was involved in deep discussion and issues were intended to solve at international level. New technologies were proposed, experiences were shared and future solutions for design infrastructure for ICT were also discussed. Research submissions in various advanced technology areas were received and then were reviewed by the committee members; 120 papers were accepted. The conference was overwhelmed by the presence of various members. Amit joshi, organizing secretary, ICICT 2020, gave the welcome speech on behalf of the conference committee and editors. Our special invitee guest —Sean Holmes, Vice Dean International, College of Business, Arts and Social Sciences, Brunel University London, UK—also addressed the conference by a speech. The conference was also addressed by our inaugural guest and speakers—Mike Hinchey, President, International Federation for Information Processing (IFIP), Xin-She Yang, Professor, Middlesex University, UK, Jyoti Choudrie, Professor, University of Hertfordshire, UK, Milan Tuba, Vice-Rector for International Relations, Singidunum University, Serbia. There were 14 technical sessions in total and talks on the academic and industrial sector were focused on both the days. We are obliged to Global Knowledge Research Foundation for their immense support to make this conference a successful one. A total of 105 papers were presented in technical sessions and 120 were accepted with strategizing on ICT and intelligent systems. At the closing ceremony, 10 Best Paper awards by

springer were announced among the best selected and presented papers. On behalf of editors, we thank all sponsors, press, print and electronic media for their excellent coverage of this conference.

London, UK — Xin-She Yang
Reading, UK — Simon Sherratt
Kolkata, India — Nilanjan Dey
Ahmedabad, India — Amit Joshi

Contents

About the Editors

Xin-She Yang obtained his D.Phil. in Applied Mathematics from the University of Oxford, and subsequently worked at the Cambridge University and the National Physical Laboratory (UK) as a Senior Research Scientist. He is currently a Reader in Modelling and Optimization at Middlesex University London and an Adjunct Professor at Reykjavik University (Iceland). He is also an elected Bye-Fellow at the Cambridge University and the IEEE CIS Chair for the Task Force on Business Intelligence and Knowledge Management. He was included in the "2016 Thomson Reuters Highly Cited Researchers" list.

Simon Sherratt was born near Liverpool, England, in 1969. He is currently a Professor of Biosensors at the Department of Biomedical Engineering, University of Reading, UK. His main research area is signal processing and personal communications in consumer devices, focusing on wearable devices and health care. Professor Sherratt received the 1st place IEEE Chester Sall Memorial Award in 2006, the 2nd place in 2016 and the 3rd place in 2017.

Nilanjan Dey is an Assistant Professor at the Department of Information Technology, Techno India College of Technology, India. He has authored/edited more than 75 books with Springer, Elsevier, Wiley, CRC Press and published more than 300 peer-reviewed research papers. Dr. Dey is the Editor-in-Chief of the International Journal of Ambient Computing and Intelligence; a Series Co-Editor of Springer Tracts in Nature-Inspired Computing (STNIC); and a Series Co-Editor of Advances in Ubiquitous Sensing Applications for Healthcare, Elsevier.

Amit Joshi is Director of the Global Knowledge Research Foundation and the International Chair of InterYIT at the International Federation of Information Processing (IFIP, Austria). He has edited more than 40 books for Springer, ACM and other reputed publishers. He has also organized more than 50 national and international conferences and workshops in association with the ACM, Springer and IEEE in e.g. India, Thailand and the UK.

Novel Methods Based on CNN for Improved Bacteria Classification

Chahes Chopra[1](✉) and Rohit Verma[2]

[1] BNY Mellon, 240 Greenwich Street, Manhattan, NY, USA
chahes96@gmail.com
[2] ABV-IIITM, Gwalior, India

Abstract. Recent times have witnessed extensive use of deep learning in both supervised and unsupervised learning problems. One of these models is convolution neural networks (CNN) which has outperformed all others for object recognition and object classification task. Although these convolution neural networks have achieved exceptional accuracies, still a huge amount of iterations create chances of getting stuck in local optima makes it computationally expensive to train. To handle this issue, we have developed some hybrid methods using, particle swarm optimization (PSO), genetic algorithm, and autoencoders for training CNN. We have taken the images of bacteria and classified them into four different genera (*E. coli*, *Listeria*, *Salmonella*, and *Staphylococcus*) to measure the performances of these models. Through this study, we concluded that evolutionary techniques can be used to train CNN more efficiently.

Keywords: Convolution neural networks · Particle swarm · Optimization · Genetic algorithm · Autoencoders

1 Introduction

Many of the recent spectacular successes in machine learning involve learning one complex task very well, through extensive training over thousands or millions of training examples [1–3]. After learning is complete, the agent's knowledge is fixed and unchanging; if the agent is to be applied to a different task, it must be re-trained (fully or partially), again requiring a huge number of new training examples.

Deep convolutional networks (CNN) have been crucial to the success of deep learning. Architectures based on CNN have achieved unprecedented accuracy in domains ranging across computer vision [1], speech recognition [4], natural language processing [5–7], and recently even the board game Go [3].

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_1

The performance of deep convolutional networks has improved as these networks have been made ever deeper. For example, some of the best-performing models on ImageNet [8] have employed hundreds or even a thousand layers [9]. However, these extremely deep architectures have been trainable only in conjunction with techniques like residual connections [9] and batch normalization [10]. It is an open question whether these techniques qualitatively improve model performance or whether they are necessary crutches that solely make the networks easier to train.

In this work, we developed some hybrid CNN models using autoencoders, PSO, and genetic algorithm individually to optimize the model and improve the classification accuracy. It has already been stated that the CNN networks might be trained end to end in a supervised method while learning highly discriminative features; this eliminates the requirement of handcrafting descriptors. Even though these networks admit a large variety of specific architectures and hyperparameters, it is challenging and expensive to identify these optimal hyperparameters manually. Therefore, we propose a deep learning approach of the convolutional neural network in conjunction with the particle swarm optimization algorithm for hyperparameter optimization. We named our first model Auto-PSO CNN. We utilized convolutional neural networks by taking use of PSO algorithm to evolve the parameters of processing layers internally. We then compared the Auto-PSO CNN accuracy with another models.

The second hybrid model has been built with a motive to use genetic algorithm to learn the best hyperparameter for CNN. Genetic algorithm is used to optimize the weights of the fully connected layer and then L-BFGS algorithm is used to optimize the best candidate produced by GA. We observed that second order-based optimization methods result in faster convergence to the best known optimal value when applied on the weights that are obtained as an output of GA (instead of randomly initialized weights). To keep chromosome length small and computation cost minimal, we used genetic algorithm only to find the best weights for the fully connected layer of the convolution neural network.

Before diving into the actual implementation of our experiment, let us first see what an autoencoder is. The first encoding layer might learn to encode easy features like edges, corners, the second encodes a bit complex features and so on until when the last layer would learn to encode the entire input image. Then the decoder will try to gradually decode it back to the original image. So autoencoders are neural networks models whose point is to recreate their information/input: this is insignificant if the system has no imperatives, however in the event that the network has constraints the learning process turns out to be intriguing. Learning the identity function alone is useless on the grounds that the network will never learn to extract useful features rather it will simply pass forward the input data to the output layer. In order to learn useful features, constraints must be added to the network. Here, the identity function is related to reproducing the image back. Here, we have used the sparsity constraint. For our experiment, we used a type of autoencoder known as sparse convolutional autoencoder. Sparse coding is defined as learning an over-complete set of basis vectors to represent input vectors. Here, we are interested in using an autoencoder for feature extraction from our images and for pretraining of convolutional neural network in later stages. The pretraining of

CNN means that the CNN would be using the same architecture as well as the trained weights of the encoder part of the CAE.

This paper is organized in the form of sections: Section 2 includes the related works. Section 3 gives the proposed approach for bacteria classification into four different classes. Section 4 discusses and describes the results. Finally, Section 5 acknowledges the concluding remarks of this study.

2 Literature Review

In the field of image recognition, one of the noteworthy explorations was done by Krizhevsky et al. [1]. The study made a neural network of 650,000 neurons having more than 60 M parameters. It is worth noting that the model had five convolution layers having max-pooling capacity among a significant number of the layers. Due to a large number of parameters, the training time of the model was very long, creating opportunities to improve the model further.

Particle swarm optimizer was introduced with the motive of reducing the training time without affecting the accuracy [11]. PSO depends on fish swarming and bird flocking hypotheses. Being a straightforward calculation, it is computationally economical regarding both space and speed. In this calculation, every particle is given a local best and global best value. The particle then chooses new solutions from these values.

Another intriguing application was proposed in 2016 by Syulistyo in a research paper. To improve the performance of CNN in the recognition task, they connected PSO to optimize output vectors from CNN [12]. The algorithm so planned would refresh the CNN weights based on a value acquired from standard backpropagation training joined with a delta factor, got from PSO. This algorithm displayed a significant accuracy of 95.08% with just four epochs. The proposed strategy showed outperformed CNN, DBN, however, took somewhat longer to create than CNN.

Fitness evaluation and encoding of the problem are two crucial factors on which the performance of the genetic algorithm depends. Since the neural networks are mostly used in large datasets, GA becomes less efficient, when used for initializing weights and optimizing the architecture of neural networks. This is because the chromosome length becomes very large, making it difficult for GA to evaluate the fitness function.

Slow convergence and local minima are some of the major problems faced in ANN; however. genetic algorithm is utilized for artificial neural network's weight initialization with a motive of solving the above mentioned issues [13, 14]. It has also been used to get the architectural design for the ANN [15]. Similarly for getting the initial weights in each layer of the stacked autoencoder, GA has shown exceptional results [16]. Utilization of GA has additionally been done in CNN for the weight's initialization [15, 17, 18] and to optimize architecture too [19, 20]. Numerous variations of GA have likewise been proposed in recent past [21, 22]. The impact of elite count on the behavior of GA has been talked about in [23].

3 Methodology

Our full experimentation consisted of five different models:

- Simple convolutional neural network (CNN)
- Simple convolutional neural net optimized using particle swarm optimization
- Autoencoders pretraining the convolutional neural network (CNN)
- Autoencoders pretraining the convolutional neural net optimized by PSO
- Convolutional neural network model optimized using genetic algorithms (GA).

A total of 1833 images were present in the dataset out of which 606 were used for training while 1227 images belonged to test set. The images corresponds to bacteria from four different genuses (*E. coli*, *Listeria*, *Salmonella*, and *Staphylococus*). The ultimate task was to classify the bacteria at the genus level effectively (Fig. 1).

Fig. 1 *Listeria*, *Salmonella*, *E. coli*, and *Staphylococcus*

This experiment utilized the fivefold cross-validation algorithm. The training set consisted of 4-folds (484 images), and the validation set consisted of onefold (122 images) in each iteration of cross-validation analysis. Since the dataset was balanced and contained roughly equal number of images of each class, we ensured that our training set and validation set for CNN as well as for autoencoder also have no major class imbalance. This made the learning of CNN and autoencoder fair enough and treated all the classes equally.

The first classification model is a simple CNN model. We tried various architectures with a different number of filters in the convolution layer and multiple sizes. Our experiment also consisted of varying optimizers, dense layers, dropout values, and the best result this architecture could fetch us was 87.1% accuracy and 0.841 as Avg. F1 score. The model that we come up with in tuning this simple CNN comprised of four convolutional layers. The number of filters in each convolutional layer is 16, 64, 64, and 96 in order. Also, 2D filter shaped 3×3 are used throughout the network. The activation functions used in between are ReLU while the final activation function for the classification task is softmax. The above system was compiled with accuracy as metric and Rmsprop as optimizer with a low initial learning rate of 0.0005. The best possible batch size was configured to 96.

Better approaches are discussed below.

Our second approach was derived from the first one, and it consisted of PSO optimizing the above simple CNN rather than the Rmsprop.

3.1 Particle Swarm Optimization

Purposed in 1995, particle swarm optimization (PSO) algorithm is one of evolutionary algorithm. It is an optimization strategy created by Kennedy and Eberhart [11]. This

technique is propelled by the social conduct of creatures that do not have a leader in their group. PSO comprises of a swarm of particles, where the particles represent a potential solution. The algorithm includes many consecutive steps. As a matter of primary importance, initialization which randomly select the particles as searching agents (x) as well as the velocities (v). The algorithm of PSO used is shown in Fig. 2.

Algorithm 1: Particle Swarm Optimization

```
% Initialization
Number_Interation_PSO;
Number_PSO_Swarm;
Determine gbest from PSO_Swarm;
Determine pbest from PSO_Swarm;
calculate fitness_old;

for 1 to Number_Interation_PSO do
 for 1 to Number_of_Particle do

    %Update Velocity
    v(n)t = v(n)t+1 + c1*r1.*(pbest-
    x(n) +c2*r2.*(gbest-x(n))

    %Update position
    x(n+1) = x(n) + v(n)t

    %Evaluate the objective function
    fitness_new = f(x(n+1))
    if (fitness_new< fitness _old)
     fitness _old = fitness _new;
     x = x(n+1);
     else
     fitness_new = fitness _old;
     x(n+1) = x;
    end if
  end for;
  Index = min (fitness_new);
  pbest = x(index);
end for;
```

Fig. 2 PSO algorithm

Furthermore, the particles at that point inserted into cost function to discover local bests (p_{best}) and global best (g_{best}). Local best is characterized as the area on which the expense is the least for each particle, while global best is the area where the cost is least among the local bests. Thirdly, the particles are updated according to Eqs. (1) and (2) mentioned below.

$$v_{n+1} = v_n + c_1 r_1 (p_{\text{best}} - x_n) + c_2 r_2 (g_{\text{best}} - x_n) \tag{1}$$

$$x_{n+1} = x_n + v_{n+1} \tag{2}$$

PSO is used to ameliorate the performance of CNN by optimizing the output vector. The utilization of PSO is due to its exemplary performance on the optimization problems. Sparse convolutional autoencoder is used to pretrain the CNN weights, CNN in this case

is the encoder part of the autoencoder and PSO is used to optimize the model for better accuracy and lesser loss function value.

Figure 3 shows the proposed method used in this Hybrid model. Here r_1 and r_2 are random numbers, c_1 and c_2 are the constants, and n is iteration.

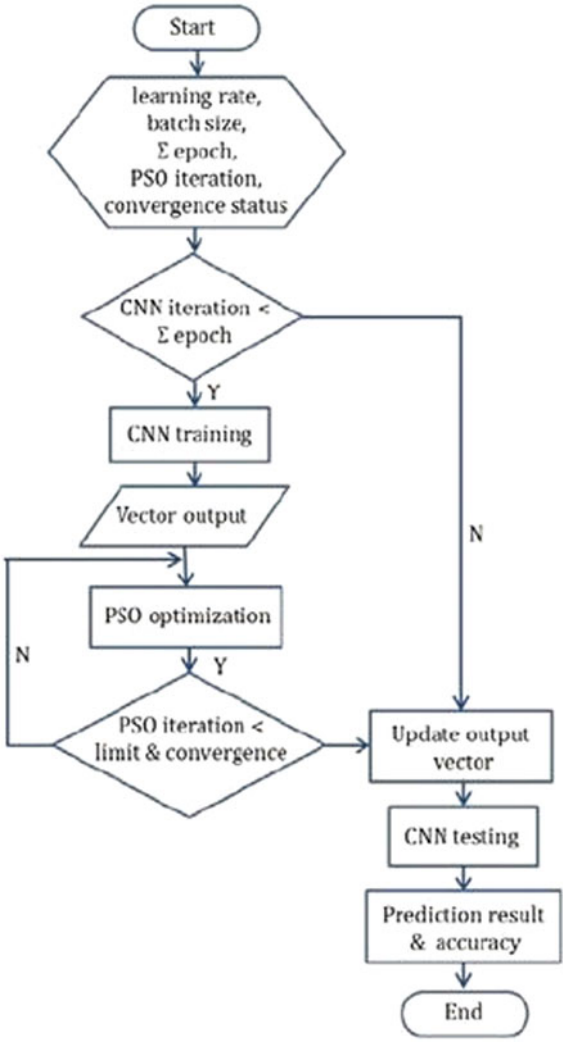


Fig. 3 Flowchart of the purposed method

3.2 Architecture and Training of Convolutional Autoencoders (CAE)

In the third approach, we trained the CAE model. We opted for 12 convolutional layers. The best architecture after multiple runs is stated as follows. Each convolutional block

consisted of two convolutional layers of the same number of filter shapes and sizes; next is the max-pooling layer followed by batch normalization layer. We used consecutive three convolutional blocks in the encoder part. The 2D filter size used is of shape (3, 3) throughout while the number of filters used in the first block is 32, then 64 and 128. The decoder part had the 2D filters in the corresponding decreasing order with the upsampling layer and the batch normalization layer in each convolution block. The activation function used is the same as were in the previous two models, i.e., ReLU, and for the last activation function, softmax. The above network was compiled using the loss as mean square error and optimized using an Adam optimizer with a low initial learning rate. The batch size used for CAE is 128 while the number of epochs is 120. We targeted a smaller learning rate and a more significant number of epochs throughout our experiment for better learning of the network and better convergence.

Since the learning rate is a very crucial hyperparameter, we took care of it using an explicit callback that was called after each epoch to change the learning rate according to the decaying exponential function that we chose for our experiment. The decaying function took the current epoch number as the input to calculate the new decayed learning rate. The initial learning rate that we used was 0.0001, and this decreased exponentially according to $lr * e^{-(k * \text{epoch number})}$, where '$k$' is a constant and was initialized to 0.01. The latter value was suited as we did several runs of the CAE.

Apart from above, we used another callback including one model checkpoint, to save the best model with least validation loss, which in this case was the mean square error or reconstruction cost. In this case, we saved the weights of the CAE prior to the state it starts overfitting. We tried several different runs with varying initial learning rate, different batch sizes, and optimizer's combination. The above stated are the best we achieved. We also used data augmentation for better training of the CAE. The augment properties that we considered were featurewise standard normalization, width shift, height shift, horizontal flip, zca whitening, shear range, and brightness range.

The above properties gave us an edge on training and increased the performance of CAE slightly for good. The final least validation loss that we incurred was of 0.00125 considering the time constraints as with each epoch proceeding the learning was getting slower too. As we stated earlier, we initialized the weights of CNN with the weights of the encoder part of the sparse convolutional autoencoder trained in the previous step. This also infers that the feature extracting part of the architecture of CNN used in this approach is the same as that of the encoder part of CAE.

The methods above were used for CNN training as well as for CAE training. Next, we one-hot encoded the labels resulting in a vector of size four since we had four classes. Once the extracted features by the encoder part of CAE are flattened, the remaining architecture of CNN comprised of two dense layers, one of 128 neurons and the other of four neurons for the final classification task with a dropout of 0.5 in between, then compiled the model with optimizer as Adam and loss as categorical cross-entropy.

Besides the above methods common to both, we also tried to customize our CNN training a bit more by increasing the batch size in between the training process. We trained our model in such a way that the batch size was doubled after every 20 epochs. We opted for increasing batch size in between since larger batches mean fewer parameter updates per epoch, so training is potentially much faster. Also with small batch size, the

gradients cannot be approximated (closely) to the true gradients. So it would take a lot longer to find the right solution. But larger batches also consume higher memory, so considering the memory constraints and the advantages of large batch size, we did not allow to get the batch size to exceed the limits beyond our storage resources but still furnishing us with the benefits. This training can also be taken as an ensemble of batch sizes that could find a trade-off between the memory availability and more accurate approximations of the true gradients. We also monitored how increasing the batch size at some point in training favored validation accuracy.

The fourth approach in our experimentation was derived from the third one but this time the model optimization was done using particle swarm optimization (PSO) algorithm rather than an Adam optimizer as in our previous model.

3.3 Optimization of CNN Architecture Using Genetic Algorithm

The final path that we worked upon was the use of the genetic algorithm in the optimization of CNN given some constraints. This resulted in an entirely new architecture to work best on the given dataset. The genetic algorithm is a metaheuristic motivated through natural selection process. Genetic algorithm has performed well when it comes to finding solution to search and optimization problems [2, 4, 11, 24]. The main reason for such efficiency is their reliance on bio-inspired operators namely mutation, crossover, and selection.

Next, we carried out a genetic search after fixing the number of generations to search for, the initial population size, fixed epochs, and the metric. It returned the top performing model upon completion. Since the classes were roughly equal, we set the metric for optimization as classification accuracy for each generation. We let 95% of the population from the crossover; best models were to survive automatically and allowed random mutations as well for continued diversity. For the search of the best architecture of CNN for this work, we provided maximum convolution layers possible, maximum dense layers, and maximum allowable filters considering the time and memory constraints. Besides that, we permitted the following available options while optimizing the network:

- Optimizers: Adam, rmsprop, adagrad, adadelta.
- Activation functions: relu, sigmoid, softmax, tanh.
- Layers: conv2D layer with various valued filters, batch normalization, dropout and max-pooling.

The best architecture that the genetic algorithm could come upon is as follows:

The model consisted of three convolutional blocks. Each block consisted of two convolutional layers with different number of filters, followed by max-pooling layer, batch normalization layer, and a dropout layer. The first block consisted of 32 and 64 filters; while second possessed 64 and 96 and the third block had 96 and 128 (filters). The dropout value was set to 0.4 and throughout the architecture ReLU was used as an activation function. Next (after flattening), three dense layers, two of sizes 128 while one of size 4 for the classification purpose are used respectively. A dropout of 0.5 is used between the second and the third dense layer. The best optimizer resulted is Adam optimizer, and a softmax activation function was placed after the last dense layer.

4 Results

In this chapter, we showed the experimental results achieved on different hybrid deep learning models used on our dataset. The overall accuracies and corresponding precision and recall of the models can be seen in Table 1. Following that, we also presented the loss curves for all the five models. Though the accuracies do not have much difference, it is essential to note that PSO, when used with autoencoders (aka Auto-PSO CNN), gave the best results when trained for the same number of epochs.

Table 1 Accuracies and F1 scores of various hybrid models

Models	Accuracies (%)	F1 Score (average)	Precision (average)	Recall (average)
Simple CNN	87.1	0.841	0.870	0.869
Simple CNN with PSO	90.7	0.890	0.890	0.887
Autoencoders with CNN	89.3	0.912	0.909	0.912
Auto-PSO CNN	94.9	0.956	0.939	0.939
CNN with genetic algorithm	93.8	0.941	0.937	0.950

4.1 Training and Testing Accuracies for Purposed Hybrid Models

See Figs. 4, 5, 6, 7 and 8.

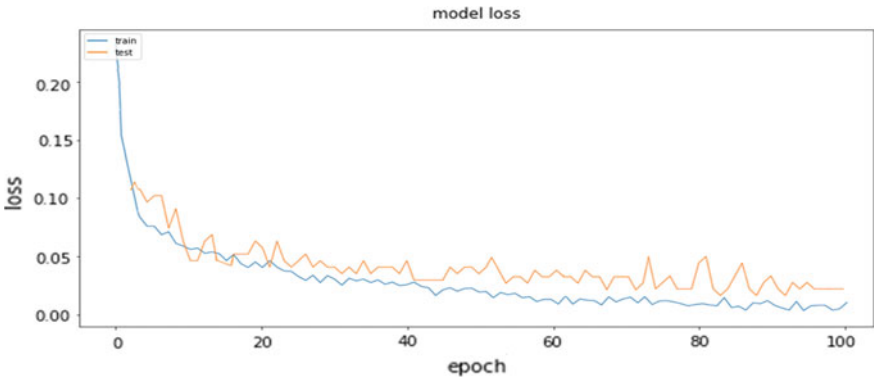


Fig. 4 Accuracy graph for simple CNN

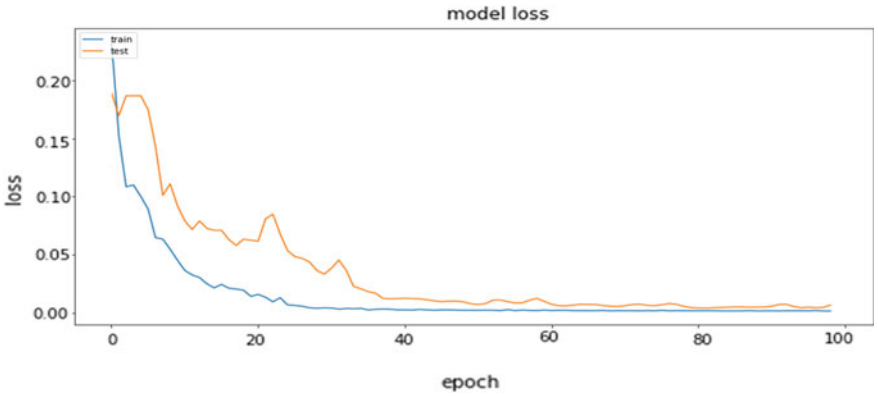


Fig. 5 Accuracy graph for simple CNN with PSO

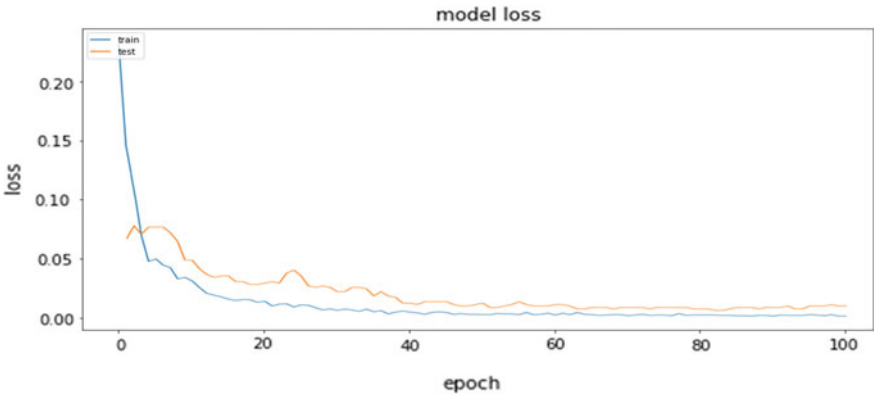


Fig. 6 Accuracy graph for CNN with autoencoders (CAE)

4.2 Confusion Matrices of Various Hybrid Models

See Figs. 9, 10, 11, 12 and 13.

4.3 Accuracies and Standard Deviation of Genetic CNN Model After Every Five Iterations

Table 2 shows that the best accuracy achieved with our last hybrid model, i.e., genetic convolution neural networks, is 93.81% after 45 generations. The respective standard deviation is 0.23. Though the difference between Auto-PSO CNN and genetic CNN accuracy is not huge in this case, it is worth noting that the training time for the latter is almost half as compared to that of Auto-PSO CNN.

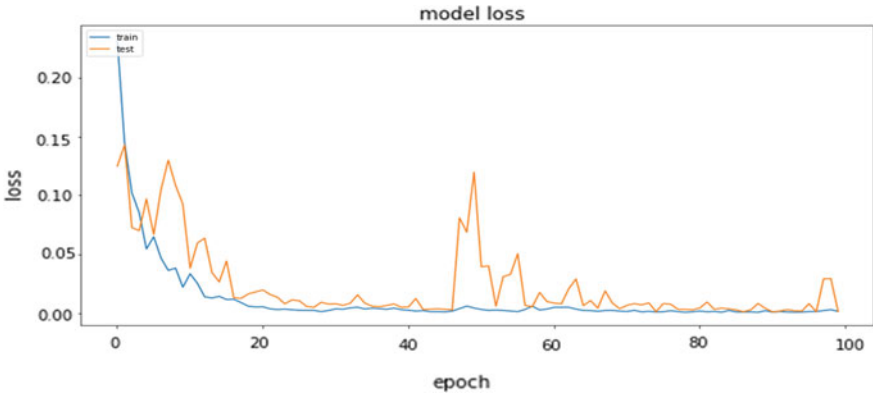


Fig. 7 Accuracy graph for auto-PSO CNN

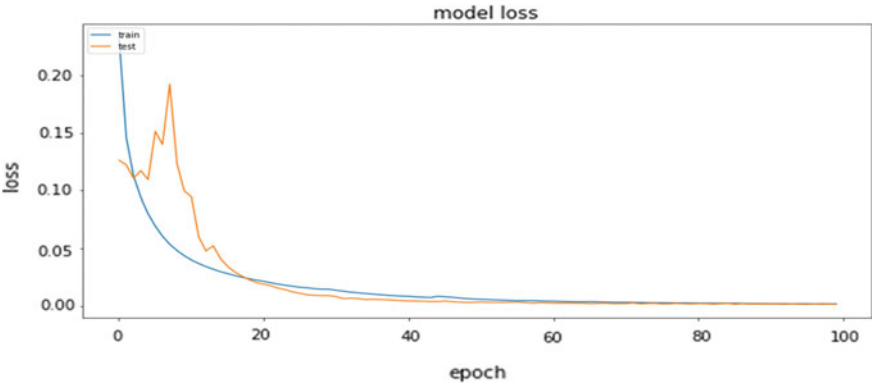


Fig. 8 Accuracy graph for CNN with genetic algorithm

5 Conclusion

The paper compared performances of five different convolution neural network models for bacteria detection out of which four were hybrid CNN models, and one was simple convolution neural networks. We approached the problem with five different and unique manners and classified bacteria images into four different classes. Based on the experiment result collected, it can be deduced that Auto-PSO CNN outperforms every other model on this dataset with 94.9% as accuracy and 0.956 average F1 score which is quite respectable considering the little data that we had while training. The important and primary reason behind it is pretraining the CNN using autoencoders instead of randomly initializing the weights. We tried achieving the minimal representation loss with autoencoders considering the time and training constraints. The secondary logical reason behind the success of Auto-PSO CNN is the customized training with callbacks, which lead to impressive results. This included special attention to the decay in learning rate using an exponential function, checkpointing model to save the best model with

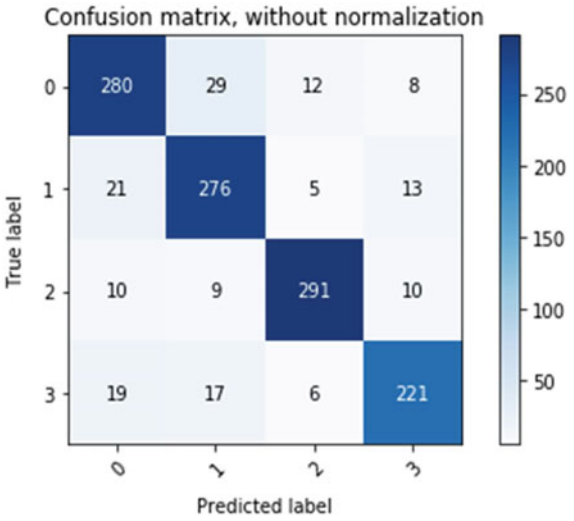


Fig. 9 Simple CNN

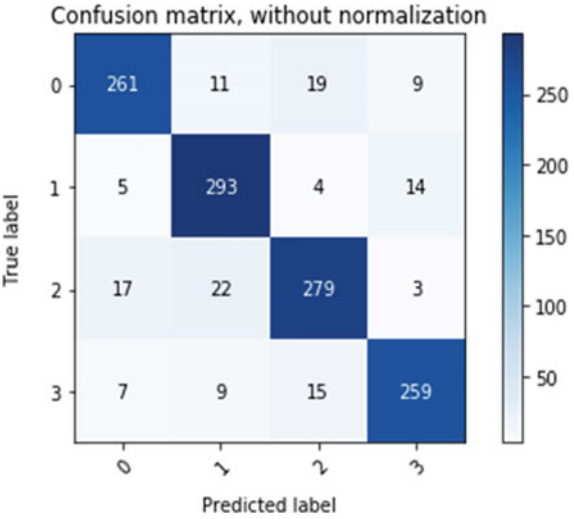


Fig. 10 CNN with PSO

preferred metric value and ensemble of variable size batch training as explained above. Lastly, the use of robust PSO combined with all above gave even more enhanced results.

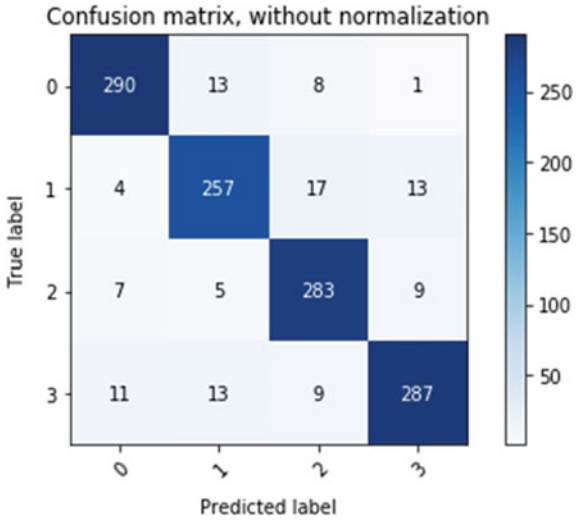


Fig. 11 CNN with autoencoders (CAE)

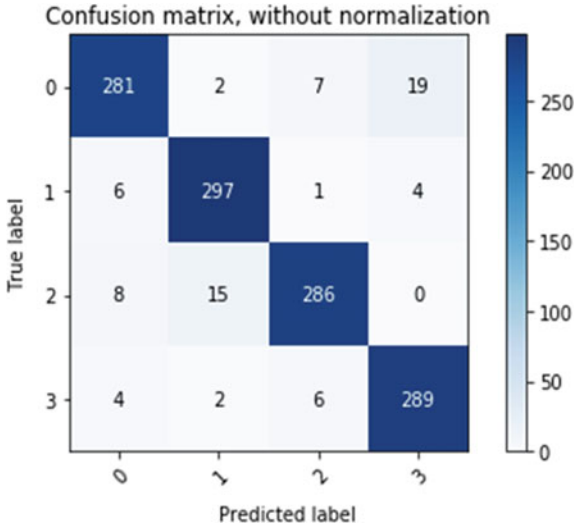


Fig. 12 Auto-PSO CNN

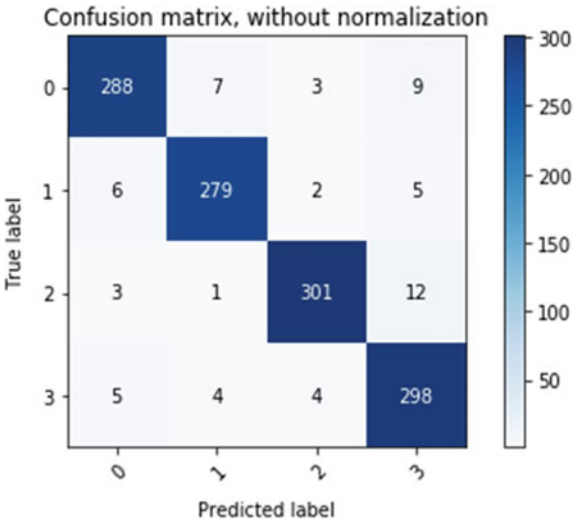


Fig. 13 CNN with genetic algorithm

Table 2 Genetic CNN performance

Generations	Best accuracy	Average accuracy	Standard deviation
5	0.6639	0.4041	0.1670
10	0.6971	0.4246	0.3015
15	0.7451	0.3922	0.2533
20	0.8567	0.5410	0.2745
25	0.8120	0.7107	0.2702
30	0.8325	0.8157	0.2469
35	0.8726	0.9270	0.0872
40	0.8921	0.8033	0.2930
45	0.9381	0.8484	0.2289
50	0.9012	0.8475	0.2216

References

1. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
2. Mnih, V., Badia, A.P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., Kavukcuoglu, K.: Asynchronous methods for deep reinforcement learning. In: International Conference on Machine Learning, pp. 1928–1937 (2016)

3. Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al.: Mastering the game of go with deep neural networks and tree search. Nature **529**(7587), 484–489 (2016)
4. Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.-R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., et al.: Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process. Mag. **29**(6), 82–97 (2012)
5. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. J. Mach. Learn. Res. **12**(Aug), 2493–2537 (2011)
6. Kalchbrenner, N., Grefenstette, E., Blunsom, P.: A convolutional neural network for modelling sentences. arXiv preprint arXiv: 1404.2188 (2014)
7. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv: 1408.5882 (2014)
8. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: CVPR09 (2009)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
10. Loffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning, pp. 448–456 (2015)
11. Eberhart, R., Kennedy, J.: A new optimizer using particle swarm theory. In: Proceedings of the Sixth International Symposium on Micro Machine and Human Science 1995. MHS'95, pp. 39–43 (1995)
12. Syulistyo, A., Rachmad, D.M., Purnomo, J., Rachmadi, M.F., Wibowo, A.: Particle Swarm Optimization (PSO) for training optimization on Convolution Neural Network (CNN). Jurnal Ilmu Komputer dan Informasi **9**(1), 52–58 (2016)
13. Kalderstam, J., et al.: Training artificial neural networks directly on the concordance index for censored data using genetic algorithms. Artif. Intell. Med. **58**(2), 125–132 (2013)
14. Yuan, Z., Wang, L.-N., Ji, X.: Prediction of concrete compressive strength: Research on hybrid models genetic-based algorithms and ANFIS. Adv. Eng. Softw. **67**(January 2014), 156–163 (2014). http://dx.doi.org/10.1016/j.advengsoft.2013.09.004
15. Oullette, R., Browne, M., Hirasawa, K.: Genetic algorithm optimization of a convolutional neural network for autonomous crack detection. In: Evolutionary Computation, 2004. CEC2004 (2004). https://doi.org/10.1109/cec.2004.1330900
16. David, O.E., Greental, I.: Genetic algorithms for evolving deep neural networks. In: Proceedings of the Companion Publication of the 2014 Annual Conference on Genetic and Evolutionary Computation (GECCO Comp '14). ACM, New York, NY, USA, 1451–1452 (2014). http://dx.doi.org/10.1145/2598394.2602287
17. Ijjina, E.P., Krishna Mohan, C.: Human action recognition using genetic algorithms and convolutional neural networks. Pattern Recogn. (2016). https://doi.org/10.1016/j.patcog.2016.01.012
18. Zhining, P., Yunming, P.: The genetic convolutional neural network model based on random sample. Int. J. u- and e- Serv. Sci. Technol. **8**(11), 317–326. (2015). http://dx.doi.org/10.14257/ijunesst.2015.8.11.31
19. Albeahdili, H.M., Han, T., Islam, N.E.: Hybrid algorithm for the optimization of training convolutional neural network. Int. J. Adv. Comput. Sci. Appl. **6**(10) (2015)
20. Xie, L., Yuille, A.: Genetic CNN (2017). arXiv:1703.01513
21. Mishra, A., Shukla, A.: Mathematical analysis of the cumulative effect of novel ternary crossover operator and mutation on probability of survival of a schema. Theoret. Comput. Sci., 1–11 (2016). http://doi.org/10.1016/j.tcs.2016.07.035

22. Mishra, A., Shukla, A.: Mathematical analysis of schema survival for genetic algorithms having dual mutation. Soft Comput., 1–9 (2017). https://doi.org/10.1007/s00500017-2650-3.24
23. Mishra, A., Shukla, A.: Analysis of the effect of elite count on the behavior of genetic algorithms: a perspective. In: 7th IEEE International Advance Computing Conference (IACC-2017), VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India, January 2017. https://doi.org/10.1109/iacc.2017.0172
24. Beasley, J., Chu, P.: A genetic algorithm for the set covering problem. Eur. J. Oper. Res. **94**(2), 392–404 (1996)

Plasma Temperature Classification for Cancer Treatment Based on Hadron Therapy

Vala Mehryar Alviri[1], Sheida Asad Soleimani[2], and Morteza Modarresi Asem[1](✉)

[1] Department of Clinical Biomedical Engineering, Tehran Medical Sciences Branch, Islamic Azad University, Tehran, Iran
mehryarvala@gmail.com, Modaresi.bme@gmail.com

[2] Department of Biomedical Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran
Sheidasoleimani95@gmail.com

Abstract. A plasma is an ionized gas that consists of free electrons, ions, and molecules, or atoms. Plasmas are characterized by many characteristics, including temperature, ionization intensity, and density, where plasmas can also be graded in a number of ways by their severity and estimations of the type they represent. Here, we classified four types of plasma based on the temperature; Hot Plasma (Thermal Plasma), Warm Plasma, Cold Plasma (Non-thermal Plasma), and ultra-cold plasma in the presence of three ions; Carbon, Neon, and Oxygen, respectively. Then, we consider the limiting factors of laser-plasma acceleration based on those four plasma types. By using laser-plasma acceleration, however, Hadron therapy reflects the significant contribution to patient care and many exciting developments. Hadron treatment, in a wide range of cancers, can generate enduring loco-regional disease control. The laser-plasma acceleration is more advantageous due to compact size compared to conventional accelerators of charged particles used in the Hadron therapy centers of today. In this paper, we demonstrate that oxygen ion in attendance of ultra-cold plasma leads to better cancer treatment by increasing the delivered energy to the tumor.

Keywords: Plasma-type classification · Cancer treatment · Hadron therapy · Laser-plasma accelerator · Ideal plasma ionization

1 Introduction

Interestingly, speaking, Irving Langmuir invented the word “plasma” to underline the contrast of plasma in the physical sciences with features of ionic fluids omnibus in biology and medicine. Despite this historical association, until lately there have been few utilizations of plasma in medicine. This condition is evolving quickly and the purpose of this analysis is to update the latest studies on the applications of plasma in medical

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_2

purposes and possible mechanisms for plasma-living interactions [1]. In this paper, we study the plasma in four groups based on its temperature; "Hot Plasma", "Warm Plasma", "Cold Plasma", "ultra-cold plasma". Hadron therapy increases survival, targets tumors and cancer cells, decreases general toxicity, provides a perfect radiation dose to the tumor, and increases the quality of life during and following treatment compared to other radiotherapy procedures [2]. The use of strong Laser-Plasma accelerators to irradiate the objective provides an effective solution for a compact accelerator for particle transmission to the patient. Plasma electrons rise to induce powerful electrical fields from the back of the source which further attracts particles out to be transmitted to the destination [3]. Cancer ion therapy is preferred because of some physical attributes compared to other techniques of Hadron therapy. Ions, at the end of their spectrum, deposit maximum energy density of Bragg peak [4].

Here, we illustrate the Bragg curve for these three ions; Carbon, Neon, and Oxygen in the existence of four mentioned plasma types. The Bragg peak is an important crest of the Bragg curve that tracks the reduction of energy from ionizing radiation while passing across the material. The maximum happens instantly before the particles settle for protons, α-rays, and other ion beams [5]. Then, we investigate the effect of plasma types on controlling the Laser-Plasma accelerator limiting factors which are named as; "energy spread", "rapid dephasing", and "pump depletion". Also, the effect of plasma density on electron energy gain in an accelerator is investigated. Finally, we proved that oxygen ion in the presence of an ultra-cold plasma in a Laser-Plasma accelerator can enhance the energy of Hadron, prolong the particle path length, and transfer sufficient energy to the cancerous target and made better treatment.

1.1 Motivation

It is evident that a fresh machine idea with personality, determined by the omnibus nature of the laser itself, is embraced in the embedded laser-driven particle system. To motivate, enable, and lead to the continuous development of laser-driven accelerator systems towards greater technical competence, creating scientific significant or influential applications. In addition, laser-driven applications, especially in the case of Hadron therapy for treating cancer, can use the singularity. Hence, using Laser-Plasma accelerators and improving them, by applying suitable low-density plasma is the main aim of this paper. Also, choosing the appropriate ion which can traverse more path through the patient's body is important.

The following paper is structured as follows. Section 2 gives includes an overview of ideal plasma contribution. Section 3 presents Plasma Classification Based on Plasma Temperature. In Sect. 4 Hadron therapy Based on Laser-Plasma Accelerators for Cancer Treatment is demonstrated. In Sect. 5 Simulations and Experimental Results are expanded. Finally, Sect. 6 is allocated to the Conclusion and Prospective Works.

2 Ideal Plasma Contribution

The ratio of $n_{z+1,1}$ and $n_{z,1}$ particles of the ions in the ground states of the ionizing stage $z+1$ and z, respectively, are the ratio ionization in thermodynamic balance, with the

density of the electron n_e multiplying, that's conveyed through Saha's equation [6]:

$$\frac{n_{z+1,1}}{n_{z,1}} n_e = \frac{g_{z+1,1}}{g_{z,1}} \frac{(4\pi m_e k_B T_e)^{3/2}}{h^3} \exp\left(\frac{-\chi_z}{k_B T_e}\right) \tag{1}$$

where $\frac{g_{z+1,1}}{g_{z,1}}$ and $g_{z,1}$ are statistical weights of the ionization stages ground states $z+1$ and z respectively, m_e is the electron mass, k_B is the Boltzmann constant, T_e is the electron temperature, h is the Planck's constant, and χ_z is the energy of ionization in stage z. The remainder of plasmas in nature is in the ideal condition. A plasma becomes ideal if the particles' total thermal efficiency, $E_{\text{thermal}} = 3/2\, k_B T$, crosses the mean electrostatic energy of contact [6].

Figure 1 illustrates the proposed ideal plasma ionization in terms of the variation of ionization energy of the ion and electron temperature. Based on this figure, Coulomb collisions are insignificant for an ideal plasma. This figure shows that the proportion of particle densities to ionization energy changes in low temperatures immediately, but changes in the environment at elevated temperatures rise. Also, Fig. 1, demonstrates that at low-temperature partially ionized plasma can be considered a combination of the ideal electron, atom, and ion gasses, at low-density, that defines the "Ideal Plasma".

As is shown in this figure, a plasma can be characterized as an ideal gas if the electron and ion shared potential energy is small in comparison with median kinetic energy. The ionization of the optimal plasma is the combination of electron, proton, neutral species, and photon (thermal radiation) ideal gas free energies, respectively. The physical image or the chemical image of the plasma can be based on a theoretical depiction of optimal ionized plasmas. Bound species (atoms, etc.) are regarded with nuclei and free electrons as primary entities in the chemical image. Nuclei and free electrons are the only fundamental components in the actual sense of the thermodynamic system (free and bound) [7].

3 Plasma Classification Based on Temperature

3.1 Hot Plasma (Thermal Plasma)

There is no clear definition of the word "Hot Plasma". It is, however, often used for fully ionized plasmas with long medium free paths in those particles. Furthermore, non-neglectable kinetic stress is often a Hot plasma signature. More exactly, with regard to the typical energy, duration, or time scales of the item under accounts, we can express the above requirements. The Sun is an example of Hot plasma in nature [6, 8]. The situation of complete ionization generally implies that the particle's thermal energy (E_{therm}) exceeds the ionization energies (E_{ion}) of the plasma electrons is formed [6]:

$$E_{\text{therm}} = \frac{3}{2} k_B T > E_{\text{ion}} \tag{2}$$

where k_B is the Boltzmann constant and T is the temperature. In the event of a pure hydrogen plasma, $E_{\text{ion}} = 13.6$ eV, but we often have a limited quantity of light elements like C or O in hydrogen plasma in useful applications. In order to completely ionize,

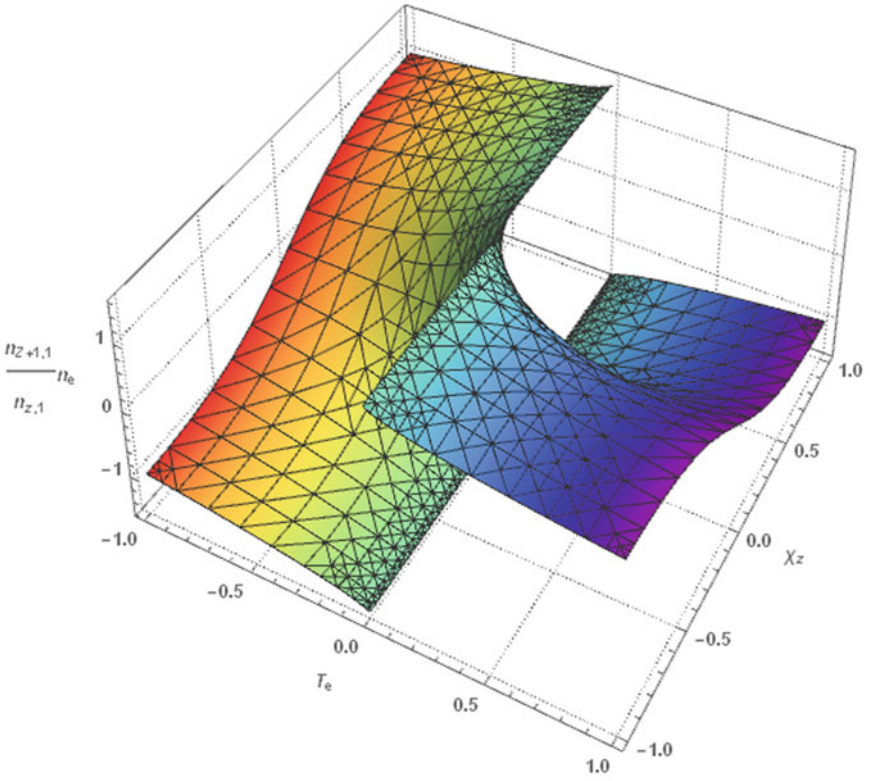


Fig. 1. Proposed Ideal plasma ionization state in terms of ionization energy of the ion and electron temperature

the temperature should be more than 100 eV. Excess energy is required to support such a plasma below these temperatures, as a significant loss channel is line radiation from partly ionized C and O atoms. In order of 1 keV$\left(\sim 1.1 \times 10^7 \text{ K}\right)$ or above temperatures, there is probably a low collisionality. That is to say, the mean free path of the particles λ_{mfp} exceeds the system level length of L [6]:

$$\lambda_{\text{mfp}} = \frac{T^2}{n_{\text{e}}} > L \tag{3}$$

where T is the temperature and n_{e} is the density of electron. For example, the mean free path is $\lambda_{\text{mfp}} \sim 20$ km and a participle goes around the normal $L = 10$ m rim before a crash, for instance, in a plasma used in the laboratory fusion studies, at $T = 10$ keV and $n_{\text{e}} = 10^{20}$ m^{-3}. This limited collision also means elevated conductivity because the collisions are accountable for plasma resistivity. As a result, the present distribution in a Hot plasma may only rapidly differ [6, 9].

3.2 Warm Plasma

The incorporation of heat impacts presents fresh forms of wave in the plasma and permits free energy needed for plasma imbalances. Thermal impacts can be taken into consideration in a warm plasma by, for example, an oscillation or adiabatic method. A kinetic strategy for the overall therapy of warm plasma instabilities is necessary. If the wave frequency is complicated and the imaginary part is the disturbance will increase exponentially. The Vlasov theory states, in addition, that all the distribution functions which decrease monotonously ($\partial f/\partial v < 0$) are stable. A positive slope, possibly makes the distribution unstable but does not ensure stability. A positive slope makes the distribution potentially unstable but does not guarantee instability. The comparative movement between the ion and electron beams can produce similar imbalances. Imbalances in plasma are a rich and diverse field. A detailed treatment of imbalances needs techniques further than the linear plasma theory, even affecting existing knowledge limits. In a warm plasma, the plasma temperature is about 10^5 K [10].

3.3 Cold Plasma (Non-thermal Plasma)

The electrical gas discharge non-thermal plasma (e.g., low-pressure glow discharge) is poorly ionized (degree of ionization 10^{-6}–10^{-4}) and significant. The temperature of the electron is significantly higher than ion and gaseous neutrals, $T_e \approx 10^4$ K $\gg T_{ion} \approx T_{gas} \approx 300$ K. The factors include the low kinetic energy conversion in collisions between electro and heavy particles and the duration of the confinement of the heavy particles. The "cold plasma" can be identified at room temperatures with heavy particles, due to the small intensity with electrons and heavy particles with small friction. Within an unbalanced plasma, electrons become charged particles with a temperature of 10^4 K or standard kinetic energy of a few eV [6, 11].

3.4 Ultra-Cold Plasma

The ultra-cool plasma first comes from the ionization of an ultra-cold gasses that electron and ion temperature are characteristic are around 1–1000 K and 1 K, respectively [12]. The complicated interaction of many bodies leads, as were many other procedures, to the sudden diffusion delay and rate delay occurrence of ultra-cold plasma. The ultra-cold system particles (plasma bundles in a loop) developed across two phases of relaxation during selective ionization of cold electrons. In the first phase, the metastable state is formed at an insignificant change in plasma density, and the balance between free and Rydberg atoms is characterized. There is a basic interest in studying the parameters and kinetic qualities of such plasma. In the second stage of development, the plasma bunch extends into an atmospheric space, is supercooled further, and certainly the transformation of the plasma phase to the crystal phase [13].

Ultra-cold plasma greatly extends the bounds of traditional neutral plasma physics and has been recently tested empirically. The most popular methods for the ultra-cold plasma research include the use of excellently-defined optical study methods including laser-inducing fluorescence imaging, optical absorption, and even recombination. Optical materials can be activated or ingested for Rydberg ions and atoms. In a system based

on ultra-cold plasma, the radiation electric field ($E_r(R, \theta)$), is given by [12]:

$$E_r(R, \theta) = \frac{e^2 n_e}{4\pi \varepsilon_0 c^2 R m_e} E_0 \sin(\theta) \tag{4}$$

where e is the electron charge, n_e is the electron density, E_0 is the initial radiation electric field, θ is the angle between the electric field polarization and the x-axis, ε_0 is the electric constant permittivity, c is the speed of light in vacuum, R is the distance, and m_e is the electron mass.

Figure 2 illustrates the radiation electric field in terms of electron density and the angle between the electric field polarization and the x-axis. Ultra-cold plasmas extend typical plasma physics' limits. This figure allows the electric radiation field to be ratcheted up to large levels, and all loaded particles can be rapidly dumped. The described electrical radiation field is then converted into real particle orbits, enabling all particle poses and speeds to eventually improve as necessary according to Eq. 4.

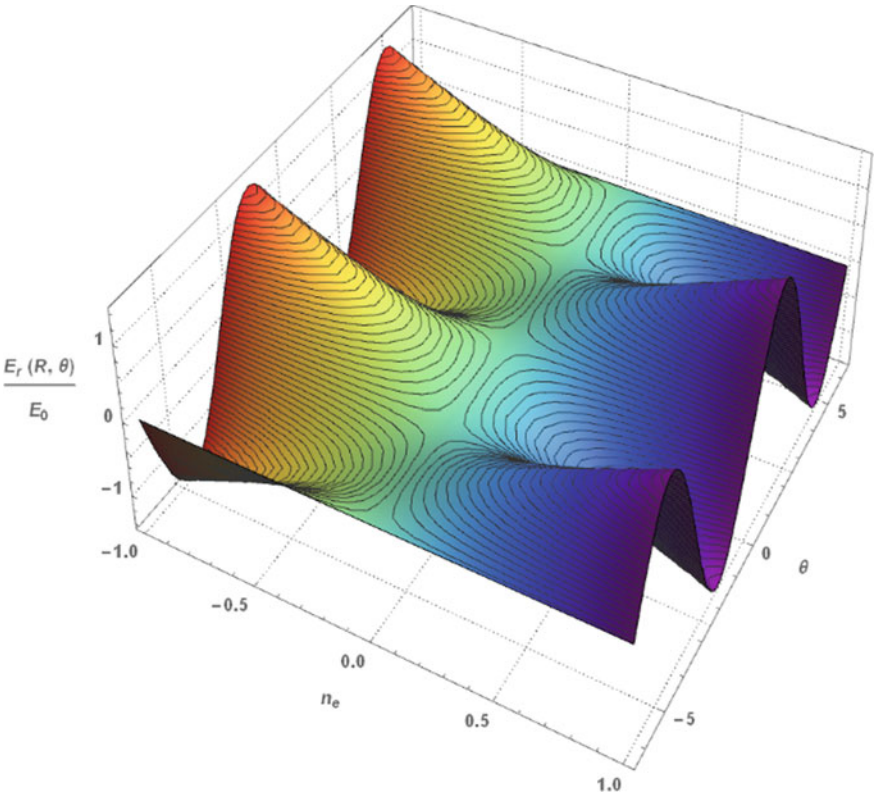


Fig. 2. Proposed electric field emission for the intensity of magnetic electron and the distance between the orientation of the electric field and the x-axis

Additionally, Fig. 3 indicates the possible difference in the strength of the ultra-cold plasma electric radiation region. Depending on this calculation, atoms are scattered in

each generation first and adjust to the final value of the electrical radiation field. Then, the radiation electric field is increased. Once the ionic potential has captured electrons, they are released to create the second curve maximum. Also, when radiation intensity is so small, no electron is captured and the second peak lacks that never deeper than the previous kinetic energy.

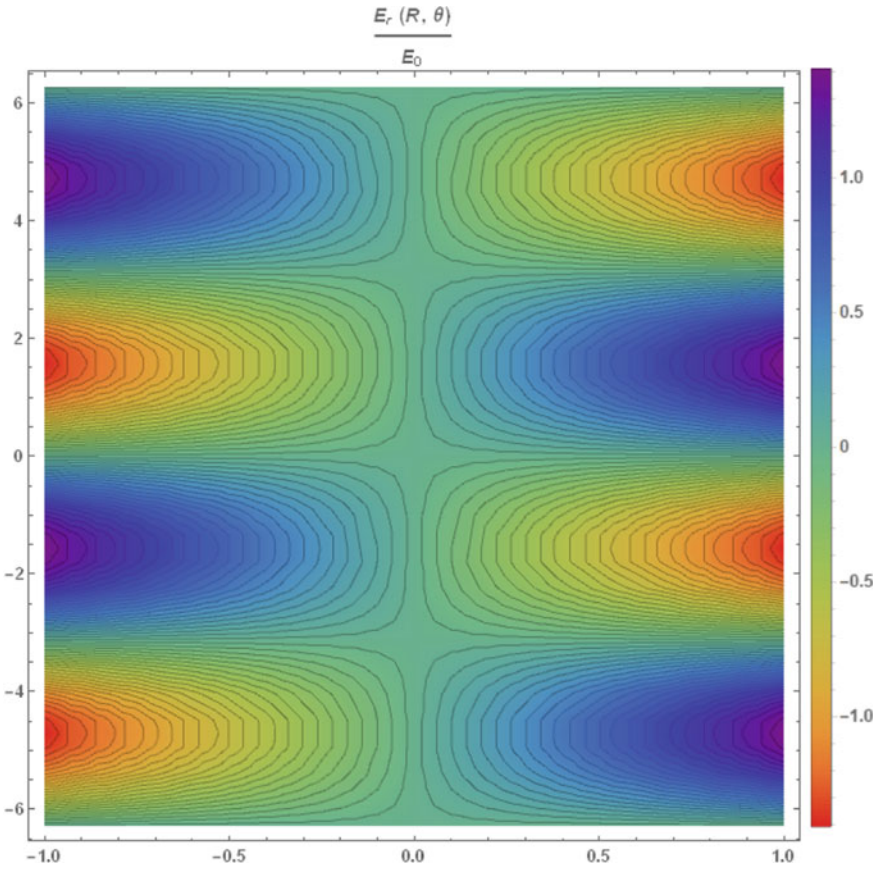


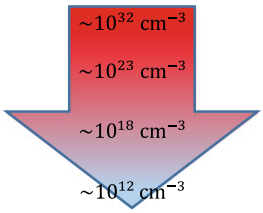
Fig. 3. Proposed intensity variation of the radiation electric field of ultra-cold plasma

With a growing electron temperature, the density of the electron has been shown to decline exponentially, defined as [14]:

$$\log(n_e) = -2\log(T_e) + 0.6 \tag{5}$$

where n_e is the density of electron and T_e is the electron temperature. Table 1, shows the electron density based on electron temperature based on Eq. 5. In this table, the electron density was shown to reduce correspondingly with decreasing electron temperature. Investigating the most efficient plasma-type building on the plasma density is the main aim of this paper with the goal of better results in Hadron therapy and complete cancer treatment with fewer side effects.

Table 1. Calculated electron temperature and ion temperature for mentioned plasma types

Plasma-type	Electron temperature (T_e)	Ion temperature (T_i)	Electron density (n_e)
Hot plasma (thermal plasma)	10^7 K or above	10^7K or above	
Warm plasma	10^5 K	10^5 K	
Cold plasma (non-thermal plasma)	10^4 K	300 K	
Ultra-cold plasma	1–1000 K (~500 K in previous calculations)	1 K	

4 Hadron Therapy Based on Laser-Plasma Accelerators for Cancer Treatment

Hadron therapy is a particular form of ecological radiation that uses quick Hadrons to achieve better particle distribution than the X-ray of standard radiotherapy (non-elemental particulars of quarks and antiquarks) [15]. It is a well-established cancer approach due to the beneficial ballistic properties of ion beams which facilitate the treatment of one of the key problems in radiation therapy: irradiation of a malignant tumor at a dose which is adequately high and uniform, ensuring that surrounding healthful tissues and bodies are irradiated to a minimum [16]. The method focuses on the effective particle speed found in high-power laser interaction simulations with the matter, influenced by plasma surroundings.

The laser Particle Acceleration is more beneficial than conventional accelerators for loading particulate matter in today's Hadron therapy facilities due to its compactness and comparatively low price of laser accelerators and because the optical whole design with gantries can be developed. Throughout experimental experiments for the integration of high-power laser radiation with different targets, successful particle acceleration enables laser-plasma to be used as a high-energy portable Hadron therapy particle source [16].

4.1 Laser-Plasma Accelerators Mechanism

The ion acceleration may be regulated if the electrical field is created and, at the same moment, the electrons are not heated too much, so that it is neutralized at the end of the acceleration process, the principal concept being the acceleration of laser radiation [17]. Collision ionization happens when an electric current passes across gasses at low pressures. If the present electrons have enough energy, they pull other electrons from the neutral gas molecules and produce ion clusters that compose separately of the new positive ion and separate negative electron. Negative ions are also produced as neutral gas molecules are attached to certain electrons. Intermolecular collisions at elevated

temperatures can also ionize gasses [18]. Finally, the possible acceleration can be created by ionization.

A simple and direct strategy of a laser-plasma ion accelerator in Hadron therapy is shown in Fig. 4. The high-intensity laser field induces relative collective plasma movement, extreme physical circumstances observed in space, and energy particle manufacturing. In Fig. 4, a terawatt laser pulse focuses on the front of the source where a cut-off plasma is generated and electrons are accelerated immediately. The electrons enter the origin and ions the hydrogen and other atoms on the reverse to create a sheath of Debye. The regular propagation of the hot electron cloud produces a cross-section rapid flow (TNSA). The distribution on the back wall of a tiny hydrogen-rich spot improves the proton efficiency of the central part of the accelerator, where it has become identical. Such protons are the community of about mono-energy group [19].

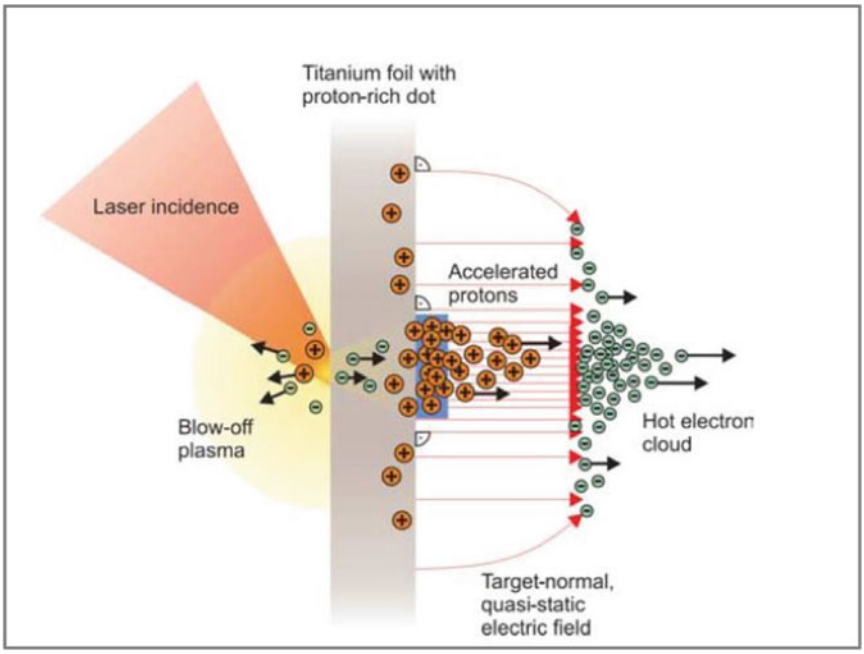


Fig. 4. The schematic of laser-plasma acceleration of particles [19]

4.2 Ion Viewpoint Cancer Treatment

- **Carbon Ion**: Carbon treatment is more enormous in atoms than in protons and neutrons. Carbogenic emission has been drawn research through technical supplies advances as well as clinical trials that show preventive advantages for different cancers, for example, breast, neck-and-mouth, lung and liver cancer, bone, and fragile sarcomas, locally occurring rectal cancer and pancreatic cancer. The possible effects

of carbon ions are contrasted to protons are presented: They include a wider distribution of physical doses because side dispersal is decreased; and higher comparative biological efficiency and a reduced rate of oxygen improvement; desirable radioresistant, hypoxic tumor eradication characteristics. The disparity between strictly ionizing nuclei and relatively ionizing X-rays and protons is more possible radiological advantages, include greater recovery efficiency and greater period dependency. [20].

- **Neon Ion**: The biological and physical benefit of neon ion radiation over mega-voltage X-rays is transplanted. The neon beam biologically decreases the percentage of oxygen rise and improves the comparative biological efficacy. Neon ion-radiated cells show less radiosensitivity and far less radiation exposure substitutions correlated with cell cycles. High-loaded particles spatial action helps high-dose radiation to be delivered reliably to tumors and lowers cell radiation [21].
- **Oxygen Ion**: They have much more benefits than radiotherapy light waves. Linear Energy Transfer (LET) and Relative Biological Effectiveness (RBE) ions have a relatively higher oxygen ion than lighter ones. Low-intensity ionizing rays provide an effective solution to radiative therapy with the introduction of a dosage of substance and Bragg peak production in a particular area. Furthermore, the oxygen radiation dose towards the close of the Bragg peak results in a big path which prohibits unwanted dosages from absorbed in delicate organ [1, 22].

In contrast to other techniques for Hadron therapy, ion cancer treatment is therefore preferred. Ions provide a better selection of the physical dose and a greater comparative bio-effectiveness as well as a reduced oxygen improvement rate. The side effect of ion therapy is significantly lower than that of other techniques. The variety of cancer treatments are also covered by ion therapy [23]. Other Hadron therapy techniques like quick-neutron treatment, proton therapy, and pion therapy are definitely more effective and effective than conventional X-ray therapies and light-beam treatment.

5 Simulations and Experimental Results

5.1 Comparison the Path Length of Tree Ions in Cancer Treatment by Hadron Therapy Through Patient's Body

In this part, by using the Bethe–Bloch equation and comparing the Bragg curve for three ions, the researchers evaluate the powerful collisions of heavy ions with matter, such tissue; Carbon ion, Neon ion, and Oxygen ion, in four plasma types; Hot plasma, Warm plasma, Cold plasma, Ultra-cold plasma.

The path length is defined as the route in which the transferred particle goes to the cancerous tissue across the patient's body. This approach investigates total dose absorbed by carrier ion in a target, in the most efficient plasma environment, through matter which can lead to maximize stopping power in Hadron therapy. When charged parts such as protons and heavy ions travel through the material mainly via ionization, the energy is lost. The authors used the Bethe–Bloch equation to measure the energy deficit for this reason [24]. The relativistic version of the formula reads Bethe–Bloch equation

in SI units as follows, for a component of speed v, charge z (multiple of the charging electron) and energy E, which moves from distance x through a range of electron number density n and mean excitation potential I [24]:

$$-\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = \frac{4\pi}{m_e c^2} \cdot \frac{N_A \cdot Z \cdot \rho}{A \cdot M_u} \cdot \frac{z^2}{\beta^2} \cdot \left(\frac{e^2}{4\pi\varepsilon_0} \right)^2 \cdot \left[\ln\left(\frac{2m_e c^2 \beta^2}{I \cdot \left(1-\beta^2\right)} \right) - \beta^2 \right] \quad (6)$$

where c is the speed of light, N_A the Avogadro number, Z is atomic number, ρ is the density of the material, A its relative atomic mass, M_u the Molar mass constant, ε_0 the vacuum permittivity, $\beta = \frac{V}{c}$, e and m_e the electron charge and rest mass, respectively. "n" is defined for each of the ions as an atomic number and atomic mass and a differentiating factor between ions. So, in contrast to neon and carbon ion, the importance of n is low to oxygen ion. Therefore, the loss of oxygen ion energy is limited to a minimal as a dosage converter. The oxygen molecule then travels further into a plasma with low intensity. The Bragg curve is a graph of the energy loss per direction in stop-format. Based on Eq. 6, energy depletion is principally characterized by the nuclear charge square, Z, and the particle velocity reverse section, β. The Bragg curve assumes its common shape which occurs in very low energy before the projectile ends. It is this Bragg peak, which benefits ion treatment over X-ray therapies. β is the ratio of the speed of particle (v) to the speed of light (c) ($\beta = v/c$) and alters between 0 and 1 ($0 < \beta < 1$). Equation 8, presents the velocity of the particle (u) as following [25]:

$$u = \frac{m\sigma\,\Delta V}{\rho esl} \quad (7)$$

where m is the molecular mass, σ is the electric conductivity, ΔV is the voltage applied across the conductor, ρ is the density, e is the elementary charge, s is the number of free electrons per atom, and l is the length of the conductor. Based on Eq. 7, velocity is inversely proportional to density. Consequently, particle speed growths at low-density ranges. The mentioned rapidity causes rises in β value. Therefore, in advance of Eq. 6, the increase of β lead to decrease of $\frac{-\mathrm{d}E}{\mathrm{d}x}$. Hence, plasma presence decreases the energy loss generated by the particle-matter interactions. Thus, by varying the β parameter, the result of the plasma types can evaluate.

When applying the stopping power over the whole course of the ion and going through the substance the accumulated energy may be obtained. The particle's initial energy lost from their contact with the material. The forms of plasma activate the electrical field which can increase carrier operators' energy for purposes of improved dose distribution to the cancer tissue. The Bragg peak summit is also held in a low-density plasma in the deeper path through the target, thus reducing energy dissipation through the material and achieving an adequate dose. Figure 5 illustrates the proposed comparison of Bragg curve manner for Carbon ion (Black), Neon (Orange), and Oxygen ions (Blue) in four plasma types; Hot Plasma (Fig. 5a), Warm plasma (Fig. 5b), Cold plasma (Fig. 5c), and Ultra-cold plasma (Fig. 5d). According to these figures, For cases of ion oxygen for four forms of plasma, a Bragg peak is observed deeper around the tumor aim. Also, based on Fig. 5a–d, the length of the route through the patient's body is longer with ultra-cold plasma having the lowest density. Therefore, the dissipation of energy through the matter decreases in ultra-cold plasma environment with oxygen ion as energy transferor to the

target. Accordingly, Table 2, demonstrates the numerical values of three ions in four different plasma types. As can be seen in this table, for oxygen ion, in the existence of ultra-cold plasma, the path length through a patient's body is 42.44 cm, which is the longest path. Based on Fig. 5 and Table 2, the stopping power of oxygen ion is the least For equivalent boundary conditions where ultra-cold plasma is present (0.82 MeV/cm). This means that The oxygen ion energy dissipation is the least across the content to the target in a molecule; This ion travels a greater distance and transfers more radiation to the focal tumor. In the presence of ultra-cold plasma, however, the oxygen ion causes acceptable outcomes in Hadron therapy.

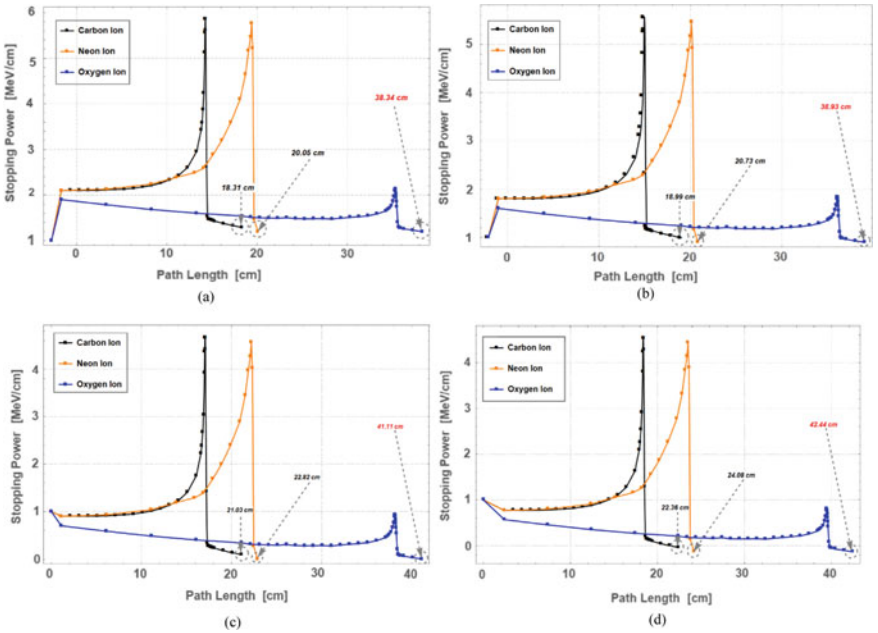


Fig. 5. Proposed Bragg curve of **a** Hot plasma, **b** warm plasma, **c** cold plasma, and **d** ultra-cold plasma for Carbon, Neon, and Oxygen ions

5.2 Electron Energy Gain in Laser-Plasma Accelerator

For various electron density values, we have tested the energy of speeding Hadrons in a Laser-Plasma accelerator. Essential research on accelerators based on laser-plasma over the last few years has taken a lot of steps towards producing high-energy Hadron therapy beams [26]. Today the energy particles over 1 TeV are produced in the industry experimentally. For these methods, the energy limits of Hadron therapy and surgical procedures are critical to high-energy particle beams with powerful laser pulses, because they can facilitate sufficient and man-able strength. Changes can be achieved in heavy ion therapy with lower cost acceleration technologies. The energy gain of an electron in

Table 2. Proposed path length and maximum stopping power (Bragg peak) of Carbon, Neon, and Oxygen ions in four plasma types

Plasma types	Hot plasma			Warm plasma			Cold plasma			Ultra-cold plasma		
Ions	Carbon	Neon	Oxygen	Carbon	Neon	Oxygen	Carbon	Neon	Oxygen	Carbon	Neon	Oxygen
Path length	18.31 cm	20.05 cm	38.34 cm	18.99 cm	20.73 cm	38.93 cm	21.03 cm	22.82 cm	41.11 cm	22.36 cm	24.08 cm	**42.44 cm**
Maximum stopping power	5.89 MeV/cm	5.78 MeV/cm	2.16 MeV/cm	5.57 MeV/cm	5.46 MeV/cm	1.82 MeV/cm	4.67 MeV/cm	4.58 MeV/cm	0.93 MeV/cm	4.55 MeV/cm	4.44 eV/cm	**0.82 MeV/cm**

the Laser-Plasma accelerator, E_e, can be approximately defined as follows [27]:

$$E_e \cong m_e c^2 \left(\frac{e^2 P}{m_e^2 c^5}\right)^{1/3} \left(\frac{n_c}{n_e}\right)^{2/3} \tag{8}$$

where m_e is the electron mass, c is the speed of light in vacuum, e is the electron charge, P is the laser power, n_c is the critical density, and n_e is the electron density. This concept demonstrates that the plasma electron density is an essential variable for changing the electron bunch strength; a great energy gain over the dephasing period is achievable for low electron density and high laser power. Figure 6 shows the electron energy gain as a function of electron density, critical density, and laser power, calculated from Eq. 8. For a laser power of 100 TW, Eq. 8 predicts that electrons can be accelerated up to a maximum value at a low-density plasma environment. The energy gain is more dependent than on laser power on the electron density. A lower electron plasma density is therefore beneficial to achieve higher electron energy. Consequently, using ultra-cold plasma, which has the lowest electron density can lead to more initial energy for transferring the Hadron to the target.

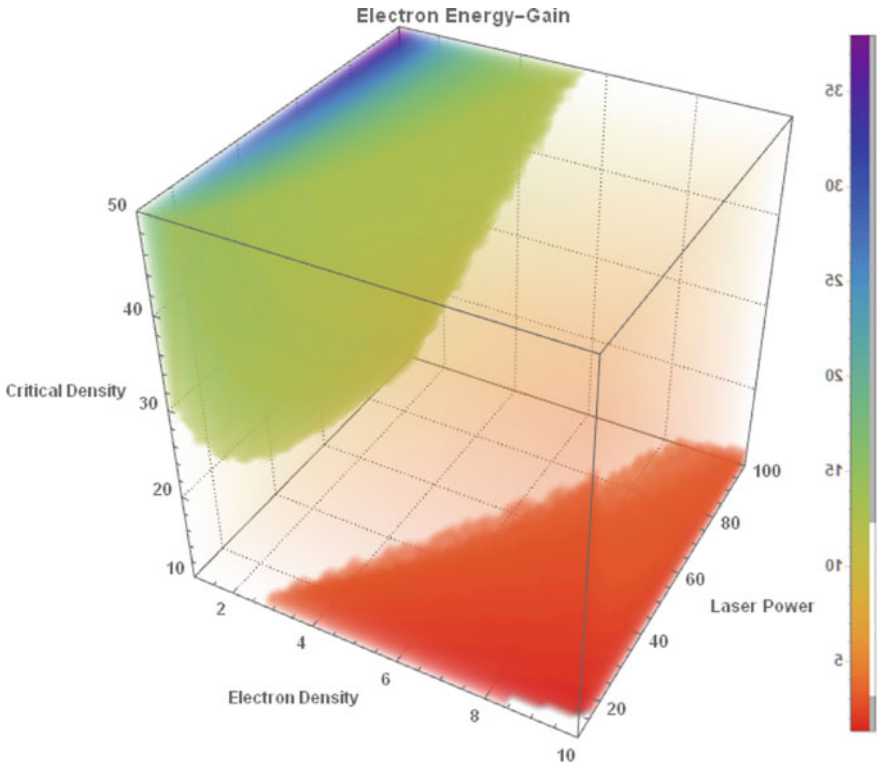


Fig. 6. Proposed electron energy gain in a laser-plasma accelerator in terms of electron density, critical density, and laser power

5.3 Investigation of Laser-Plasma Acceleration Limitations for Tree Ions Based on Four Plasma Types

5.3.1 Energy Spread

The energy spread comes first with regard to the phase reliance of the Wakefield across the finite bunch together with the dilution throughout acceleration. This may be offset by the beam loading effect which then results in an energy spread that is almost independent of the bunch length [28]. The relative energy spread $\left(\frac{\Delta\varepsilon}{\varepsilon}\right)$ is defined as [29]:

$$\frac{\Delta\varepsilon}{\varepsilon} = \frac{\eta\cos\left(\frac{kL}{2}\right)\left(\psi_{\text{ext}} - \psi_{\text{inj}}\right)}{\sin\left(\psi_{\text{inj}}\right) - \sin(\psi_{\text{ext}})} \tag{9}$$

Also, the beam loading (η) is defined as [29]:

$$\eta = 2kb\sin\left(\frac{kL}{2}\right)\frac{n_{\text{e}}}{n_0 A_0} \tag{10}$$

where L is the bunch length, k is a coefficient, ψ_{ext} is the extracted phase, and ψ_{inj} is the injected phase, n_{e} is the electron density, n_0 is the initial density, and A_0 is the laser Wakefield amplitude. Figure 7 shows that Minimization of proportional energy distribution by cold plasma locates the particle at the purpose (tumor) and does not affect tissue confinement on the application of the particle beam. Thus, in Hadron therapy, the process described can trigger improve dose delivery to cancerous tissue.

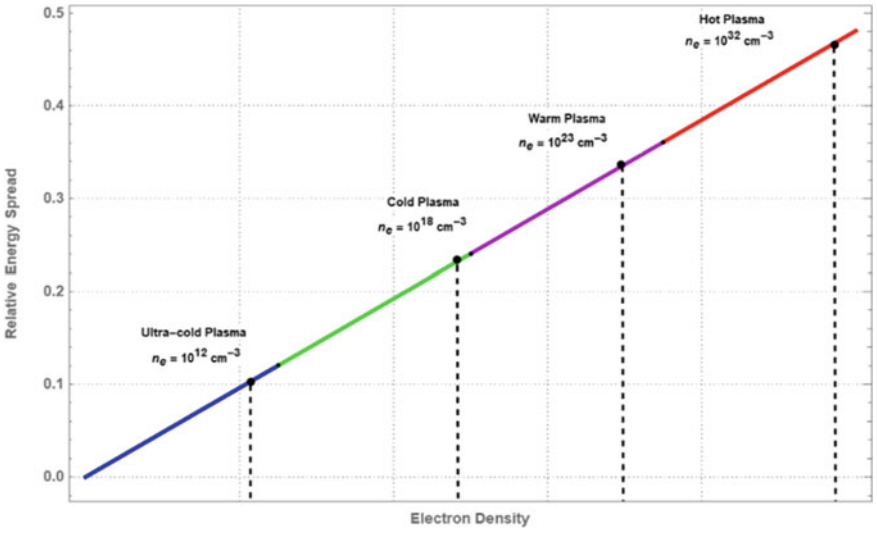


Fig. 7. Proposed comparison of the relative energy spread for four plasma types

Additionally, Fig. 8 illustrates the relative energy spread as a function of beam loading and bunch length at $\psi_{\text{inj}} = \pi$ (constant). According to Fig. 8, at first, there is

no variation of relative energy spread, then began to decrease and the minimum value of the energy spread appears. After mentioning minimization, began to decrease. Figure 8 shows that for some specific values of beam loading and bunch length the relative energy spread is minimized. Hence, the optimal values of the three defining parameters can be selected to minimize or even eliminate the supply of energy. Furthermore, this process can apply for constant values of beam loading and in terms of bunch length and injection phase. In this case, it can be observed that at least one concavity appears in "bunch length-injection phase-relative energy spread" graph. Based on previous statements and implementations, we demonstrate that the well-known question of energy spreading laser-plasma acceleration will solve by actively selecting optimum values for bunch frequency, beam loading, and the injection process, which can impact negative local cancer care and even energy gain. Also, we show that for optimal amounts of three mentioned parameters, the energy spread eliminates completely, in theoretical studies and simulations. In an ultra-cold plasma setting, the energy delivery minimizes and Hadron therapy produces optimum results by choosing carefully optimized bunch length, beam charge, and the injector step.

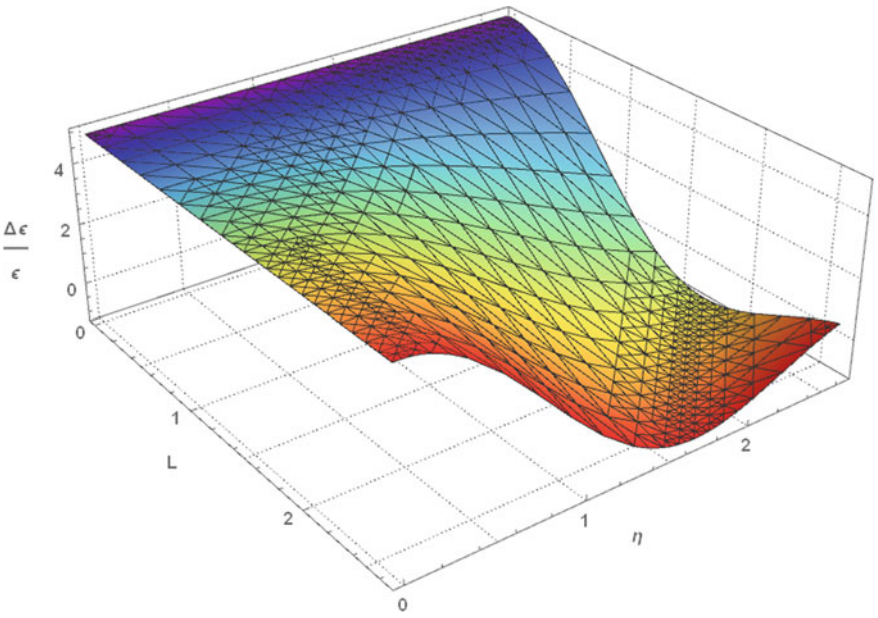


Fig. 8. Proposed relative energy spread as a function of beam loading (η) and bunch length (L) at $\psi_{\text{inj}} = \pi$

5.3.2 Rapid Dephasing

As the plasma wave phase develops at the group laser velocity in the plasma, an accelerated relativistic electron can examine various phases of the accelerating field. The

dephasing length, L_d, is defined as the length for an electron to be speeded up before the electric field decelerates. L_d is evaluated as follows [27]:

$$L_d = \frac{4}{3}\frac{\omega_0^2}{\omega_p^2}\frac{\sqrt{a_0}}{k_p} \tag{11}$$

Also, the plasma frequency (ω_p) is defined as [27]:

$$\omega_p = \sqrt{\frac{n_e e^2}{m_e \varepsilon_0}} \tag{12}$$

where ω_0 is the laser frequency, ω_p is the plasma frequency, a_0 is the laser pulse envelope, k_p is the plasma wave number, n_e is the electron density, e is the electron charge, m_e is the electron mass, and ε_0 is the permittivity of free space. Figure 9a and b shows the dephasing length in terms of laser pulse envelope and laser frequency for four plasma types. Based on Fig. 9, a low-density plasma is used to delay the electron dephasing. Therefore, ultra-cold plasma can counteract the negative and unfavorable effect of dephasing. As an important consequence, in lengthy electron dephasing, the maximum energy gain in tumor, increases. Hence, particles convey much more energy through the body to the tumor.

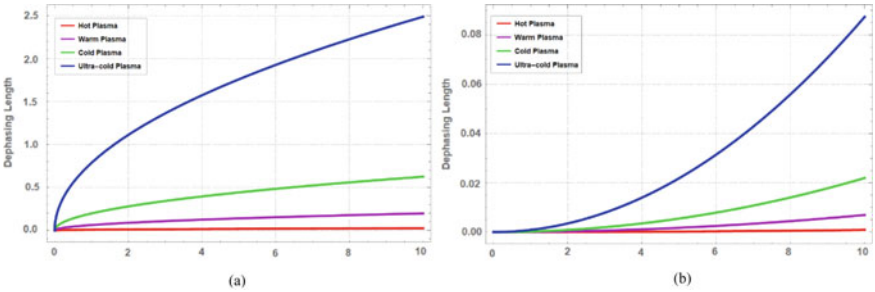


Fig. 9. Proposed dephasing length for four plasma types in terms of **a** laser pulse envelope and **b** laser frequency

As density decreases, the dephasing length and maximum energy increase so that accelerator lengths of the several cm and energies up to many GeV per level can then be achieved. Consequently, the maximum dose can deliver to the target, in Hadron therapy, for precise cancer treatment. Getting an adequate dose in the tumor brings successful and complete treatment with no remaining cancerous tissue.

5.3.3 Pump Depletion

Another limit for laser-plasma acceleration is the pump depletion (L_{pd}), use that depletes the driving laser. As a laser pulse passes through the plasma, energy is transferred to the plasma wave. The length above which energy in the plasma wave corresponds to the

driver laser is described as the length of pump depletion. The pump depletion length (L_{pd}) is described as [27]:

$$L_{pd} = \frac{\omega_0^2}{\omega_p^2} c\tau_0 \tag{13}$$

where c is the light velocity in the vacuum, ω_p is the plasma frequency, ω_0 is the laser frequency, and τ_0 is the laser period. Figure 10, demonstrates the pump depletion length, for four plasma types on our simulation. Depending on this figure the length of the pump depletion depends on the plasma intensity and for higher densities, it gets longer. For some situations, the pump degradation is interrupted by a low-density plasma. Therefore, the proposed approach to resolve laser loss is ultra-cold plasma. In Hadron therapy, the transferor particle delivers more energy to the target by introducing ultra-cold plasma.

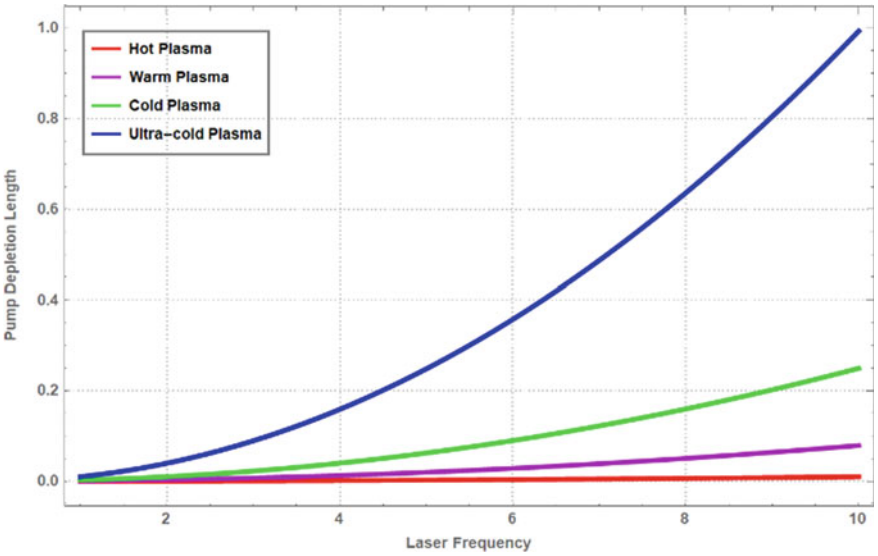


Fig. 10. Proposed pump depletion length for four plasma types

6 Conclusion and Prospective Works

A wide variety of neutral substances, which include many interacting free electrons and ionized atoms or molecules, are used to describe the term plasma, because of its individual conduct. Hadron therapy is a high accuracy cancer treatment method allowing preferable conformal therapy with conventional methods of radiation therapy. Ions consider practical uses in the fields of interdisciplinary therapy like Hadron therapy, fusion, space program, etc. The principal common element in all these applications is particular transport physics with the main constituents atomic and plasma physics. The

high-energy ion beams provide favorable situations in the treatment of local deep-seated tumors contrasted with traditional photons and electron beams. Their physical distribution of the depth of the tissue is distinguished towards the end by a high distal lower rim attributable to the minimal entry dose and a distinct limit (Bragg peak). The established range and the small side beam distribution allow for a millimeter-precision distribution of the dose. Ions, in fact, are enhanced by heavy ionization and decreased cellular repair level in the Bragg peak area, making these very desirable for the cure of radio-resistant localized cancers.

In this study, the authors discuss three ions: Carbon, Neon, and Oxygen in the existence of four forms of plasma; "Hot plasma", "Warm plasma", "Cold plasma", and "Ultra-cold plasma". Then, three limiting factors of Laser-Plasma accelerators are presented; "Energy spread", "Rapid dephasing", and "Pump depletion", which can control based on laser-plasma environment density. This study indicates that reducing or even eliminating the energy spread can be achieved by selecting the optimum values of both determinative variables in a low-density ultra-cold plasma environment.

Finally, we proved that Oxygen ion in the existence of ultra-cold plasma in a Laser-Plasma accelerator can enhance the energy of Hadron, prolong the particle path length, and transfer sufficient energy to the cancerous target. In Hadron therapy, authors have tried to present new approaches, for increasing energy and deliver a sufficient dose to the target due to boost cancer therapy with minimal damage to safe tissues and overcome the existential challenges.

References

1. Friedman, G., Gutsol, A., Shekhter, A.B., Vasilets, V.N., Fridman, A.: Applied plasma medicine. J. Plasma Process Polym. (2008). https://doi.org/10.1002/ppap.200700154
2. Mohan, R., Grosshans, D.: Proton therapy—present and future. J. Adv. Drug Deliv. Rev. (2018). https://doi.org/10.1016/j.addr.2016.11.006
3. Orecchia, R., Zurlo, A., Loasses, A., Krengli, M., Tosi, G., Zurrida, S., Zucali, P., Veronesi, U.: Particle beam therapy (hadron therapy): basis for interest and clinical experience. Eur. J. Cancer (1998). https://doi.org/10.1016/S0959-8049(97)10044-2
4. Enferadi, M., Sadeghi, M., Shirazi, A.: Radiation therapy in the early 21st century: technological advances. Curr. Cancer Ther. Rev. J. (2011). https://doi.org/10.2174/157339411797642614
5. Ludwiga, N., Kusumotoab, T., Galindoa, C., Peaupardina, P., Pin, S., Renault, J.-P., Mullerd, D., Yamauchib, T., Kodairae, S., Barillona, R., Raffya, Q.: Radiolysis of phenylalanine in solution with Bragg-peak energy protons. Radiat. Meas. (2018). https://doi.org/10.1016/j.radmeas.2018.07.007
6. Dinklage, A., Klinger, T., Marx, G., Schweikhard, L.: Plasma Physics Confinement, Transport and Collective Effects. Springer, Berlin (2005). ISBN: 978-3-540-25274-0
7. Potekhin, Y., Chabrier, G., Shibanov, Y.A.: Partially ionized hydrogen plasma in strong magnetic fields. J. Phys. Rev. E. (2000). https://doi.org/10.1103/PhysRevE.60.2193
8. Keidar, M., et al.: Plasmas for treating cancer: opportunities for adaptive and self-adaptive approaches. J. Trends Biotechnol. (2018). https://doi.org/10.1016/j.tibtech.2017.06.013
9. Ramazanov, T.S., Kodanova, S.K., Bastykova, N.Kh., Maiorov, S.A.: Investigation of hydrodynamic properties of hot dense plasma. Phys Wave Phenom. (2018). https://doi.org/10.3103/S1541308X18040131

10. Kilpua, E., Koskinen, H.: Introduction to Plasma Physics. (2017). ISBN: 978-951-745-228-1
11. Metelmann, H.-R., von Woedtke, T., Weltmann, K.-D.: Comprehensive Clinical Plasma Medicine—Cold Physical Plasma for Medical Application. Springer, Berlin (2018). https://doi.org/10.1007/978-3-319-67627-2
12. Lu, R., Guo, L., Guo, J., Han, S.: Microwave diagnostics of ultracold neutral plasma. J. Phys. Lett. A (2011). https://doi.org/10.1016/j.physleta.2011.04.027
13. Maiorov, S.A.: On the ultracold plasma evolution. Bull. Lebedev Phys. Inst. (2013). https://doi.org/10.3103/S1068335614030075
14. Edberg, N.J.T., Wahlund, J.-E., Agren, K., Morooka, M.W., Modolo, R., Bertucci, C., Dougherty, M.K.: Electron density and temperature measurements in the cold plasma environment of titan: implications for atmospheric escape. Geophys. Res. Lett. (2010). https://doi.org/10.1029/2010gl044544
15. Rezaee, L.: Design of spread-out Bragg peaks in hadron therapy with oxygen ions. Rep. Pract. Oncol. Radiother. (2018). https://doi.org/10.1016/j.rpor.2018.08.004
16. Bulanov, S.V.: Laser ion acceleration for hadron therapy. In: International Conference Laser Optics. IEEE (2014). https://doi.org/10.3367/UFNr.0184.201412a.1265
17. Tikhonchuk, V.T.: Physics of laser-assisted ion acceleration. Nucl. Instrum. Meth. Phys. Res. Sect. A (2010). https://doi.org/10.1016/j.nima.2010.01.051
18. Miyamoto, K.: Plasma Physics and Controlled Nuclear Fusion. Springer, Berlin (2005). https://doi.org/10.1007/3-540-28097-9
19. Schwoerer, H., Pfotenhauer, S., Jackel, O., Amthor, K.-U., Liesfeld, B., Ziegler, W., Sauerbrey, R., Ledingham, K.W.D., Esirkepov, T.: Laser–plasma acceleration of quasi-monoenergetic protons from microstructured targets. Nat. Int. J. Sci. (2006). https://doi.org/10.1038/nature04492
20. Kamada, T., Tsujii, H., Blakely, E.A., Debus, J.: Policy review carbon ion radiotherapy in Japan: an assessment of 20 years of clinical experience. J. Lancet Oncol. (2015). https://doi.org/10.1016/S1470-2045(14)70412-7
21. Mukherji, A.: Particle Beam Therapy: A Quick View. Springer, Berlin (2018). https://doi.org/10.1007/978-981-10-6659-7_18
22. Xu, Y., Lin, Y., Gao, S., Shen, J.: Study on mechanism of release oxygen by photo-excited hemoglobin in low-level laser therapy. J. Lasers Med. Sci. (2018). https://doi.org/10.1007/s10103-017-2363-y
23. Blyth, B.J., Cole, A.J., MacManus, M.P., Martin, O.A.: Radiation therapy-induced metastasis: radiobiology and clinical implications. J. Clin. Exp. Metastasis. (2018). https://doi.org/10.1007/s10585-017-9867-5
24. Belkić, D.: Advances in Quantum Chemistry. Springer, Berlin (2013). https://doi.org/10.1016/B978-0-12-396455-7.00001-7
25. Boylestad, R.L., Nashelsky, L.: Electronic Devices and Circuit Theory. Prentice Hall, Upper Saddle River (2012). ISBN: 978-0132622264
26. Nunes, M.D.: Hadron Therapy Physics and Simulations. Springer, Berlin (2013). ISBN: 978-1461488989
27. Cros, B.: Laser-driven plasma wakefield: propagation effects. CERN J. (2016). https://doi.org/10.5170/CERN-2016-001.207
28. Li, X., Nghiem, P.A.P., Mosnier, A.: Toward low energy spread in plasma accelerators in quasilinear regime. Phys. Rev. Accel. Beams J. (2018). https://doi.org/10.1103/PhysRevAccelBeams.21.111301
29. Reitsma, A., Trines, R., Goloviznin, V.: Energy spread in plasma-based acceleration. IEEE Trans. Plasma Sci. (2000). https://doi.org/10.1109/27.893302

On the Specific Role of Electronic Document and Record Management Systems in Enterprise Integration

Ingrid Pappel[1], Sidra Butt[1(✉)], Ingmar Pappel[2], and Dirk Draheim[1]

[1] Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia
{ingrid.pappel,sidra.butt,dirk.draheim}@taltech.ee
[2] Interinx Ltd., Kadaka tee 44-35, 12915 Tallinn, Estonia
ingmar.pappel@interinx.ee

Abstract. The paper aims at providing practical insight into the specific role of electronic document and record management systems (EDRMSs) in the process of enterprise integration (EI) to provide organizational data flow. Based on a case study with the ERDMS Amphora, we identify enabling factors for enterprise integration in order to provide interoperability, where we consider both technological and organizational factors. An EDRMS is not a standalone system, and it must interact with many systems used in organization. In this paper, we discuss some certain information flows based on the EDRMS Amphora and related system.

Keywords: Information systems · Enterprise application integration · Electronic document · Record management · MoReq

1 Introduction

In today's enterprise information system landscapes, there co-exist a large variety of different integration techniques and technologies: from the classical service-oriented architecture (SOA) [1,2] with its subsequent, heavy-weight enterprise service bus (ESB) [3] approach, over business process management (BPM) suites [4] to the lightweight integration with ad-hoc glue clode [5] that is still used in many scenarios today.

However, the role of enterprise integration that is based on document exchange is often underestimated. Electronic document and record management systems, such as defined by the EU Commission's standard Model Requirements for the Management of Electronic Records (MoReq) [6], play an important role in today's organizations; whenever it comes to compliance, their contribution is usually mission critical. Therefore, the need for an EDRMS-based data exchange emerges naturally in today's organizations. Often, it is realized in an ad-hoc man-

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_3

ner. But what are the enablers for a systematic, stable EDRMS-based enterprise integration? This is the main question that we target in this paper. In service of this, we report our experience with the EDRMS system Amphora. From 2003 till now, Amphora has been introduced as the EDRMS in most Estonian local governments, which constitute more than 70% of Estonia's local governments. We identify both technological and organizational factors for an EDRMS-based EI integration. We include inter-organizational EI into our considerations and discuss what type of information flows can be managed by that. In Estonian public sector, most of the official documents are exchanged digitally. In addition, we will discuss integration and interoperability based on real-life example based on e-invoicing in local governments as in Estonia, all invoices are exchanged digitally in public sector [7].

In Sect. 2, we provide with the description for the basic principles of an EDRMS, In Sect. 3, we provide a brief description of the case of the EDRMS Amphora in Estonia. In Sect. 4, we explain the identified enablers for EDRMS-based EI. Furthermore, interoperability is discussed in Sects. 5 and 6. Then, in Sect. 7, business process re-engineering implementation has been highlighted. We proceed with a discussion of possible future direction in EDRMS in Sect. 8. We finish the paper with a conclusion in Sect. 9.

2 Basic Principles for an EDRMS

A review in 2007 concluded that the old document management system and process was highly inefficient and that the users were dissatisfied with the features and reliability [8]. A large quantity of the paper-based files and documents make it harder for the organizations to manage the workflow efficiently. For example, an EDRMS system that is implemented and designed for private sector is able to manage files constituting government records in the required format. Most of the organizations rely communication and document sharing through shared drives, which negatively impacts the effectiveness of the management as well as the assessment of the sensitive business information. Improper handling of sensitive information by government organizations leads to discredited reputation. An EDRMS allows these organizations to make sure that information saved in an efficient manner, which in turn guarantees elevated performance. This also serves as a concrete basis for the process of digital archiving of the data as all the content is now available in the proper format through the implementation of an EDRMS system. Hence, it can be said that EDRMS is a founding pillar for efficient record and information management. Digital archiving of the records is important to ensure the protection, maintenance, as well as the accessibility of the information. Furthermore, the technology changes quite often to adapt to the dynamic requirements and needs of an organization; archiving solutions helps preserve the critical information outside of all the software applications that might go obsolete with the passage of time [9].

Some of the tools are necessary in order to make sure that the public information is handled properly and effectively. According to Morissey; "the corporate

information repository of organizations to increase the capacity of information services to integrate the vast array of corporate knowledge being generated by organizations and their employees" [10].

2.1 Importance of Implementing EDRMS

An EDRM is a system that constitutes of all the required attributes of a quality management system necessary for both the public and private sector. A centralized database system enables access control, tracking and trace-ability while increasing efficiency of processes and record management. It is through the integration of an EDRMS that confidentiality of the retained information is maintained while also being able to share non-sensitive information for transparency and enhanced efficiency. A successful integration of EDRMS within the business processes (BP) enables organizations to follow the principles of interoperability [11]. This is why an EDRMS is essential for the organizations that have data stored in shared drives, making it easier for employees to retrieve required information while reducing the misuse and loss of data.

An EDRM system also makes sure that resources are utilized efficiently by reducing the duplication of data. One of the main reasons of implementing an EDRMS is to make sure that duplication is eliminated, apart from its benefits of sharing data across different platforms. A decrease in data duplication in an organization improves accuracy as well as the work efficiency. The retention of a knowledge base is also maintained and enhanced through EDRMS [12]. For example, if an organization requires to replace its employee, the new appointment employee can easily find the documents and files.

2.2 EDRMSs for E-Government

There are countless benefits that an EDRMS system can provide for a successful e-government. Electronic documents reduce cost while ensuring standardized format for information storing while enhancing transparency. Documents can be shared instantly and more efficiently, which improves the workflow and enables a unified interface. The loss of any document is prevented through a log maintaining history, location, and the status of these document which simplifies the tracking process. Security is maintained at high level. The implementation of digital signatures and encryptions while making use of public key infrastructure (PKI) maintains security. These digital signatures allow users to maintain the authenticity and the integrity of a document, while eliminating the need for hard copies. Government can be held accountable with a successful integration and implementation of an EDRMS within its services and processes. It is through the digital signatures based on PKI that non-repudiation of documents is ensured. An EDRMS also enables integration with the E-archiving. The metadata for each document is available in the EDRMS, which means that any document can be retrieved from one location. Such example can also be seen through the e-government systems in Estonia where the government enforced the implementation of digital signatures through the support of law and legislations [13].

A digital government or an e-government concept is introduced and adapted in many first world countries in order to improve the efficiency of government processes while also improving the efficiency and increasing the transparency for decision-making. A strong emphasis remains on the fact that it is through the EDRMs that information remains reliable and secure, which is critical for the smooth functioning of an e-government. Estonia is a perfect case, where paperless government meetings were first implemented in 2001. It then further developed into the development of an e-LocGov model in order to introduce e-governance, with the cooperation of local governments. It is through the successful integration of EDRMSs for the public sector that Estonia developed ideal e-government model [14].

3 Case Description

Today, the EDRMS Amphora is used in the majority of local governments in Estonia [7]. Beyond handling electronic documents and record management and a series of other basic information management tasks, a series of electronic service delivery processes have been implemented on the basis of the EDRMS Amphora—including interaction with both internal and external entities. The system was initially developed primarily for application in local governments, which typically operate—in accordance with the Estonian "no legacy" principle [15]—leading-edge ICT infrastructure and solutions. Beyond that purpose, Amphora the system is applied more and more also in the private sector due to its flexibility, modularity, and enterprise content management (ECM) functionality. The main contributions of Amphora to digitalization of local governments in Estonia are [16]: creation of electronic documents exchange within and between governmental agencies; improvement of public servants' electronic working space (including the integration with different office software solutions); improvement of business processes through the integration of document workflow management; provision of basic functionality for interaction with citizens (e.g., by allowing public servants to publish documents for public access); an increased quality of public services. Over the years, Amphora has become an example of applying an EDRMS for integration with other information systems, both inside organizations and cross-organizational.

4 Enabling Factors of EDRMS-Based in EI

4.1 Technological Enabling Factors

One of the most important factors was service-oriented thinking [17] in the development of Amphora that integrates EDRMS. It was clear from the beginning that merely a tool for document management would not be sufficient. What was needed, was a tool that is tightly integrated in the e-services delivery process. In particular, its features for the integration with other information system made Amphora particularly useful for efficient service delivery.

Here, "service-oriented thinking" does not stand for service-oriented architecture (SOA) [2,3,18]. Rather, it is about a certain requirement elicitation attitude, i.e., a permanent mode of thinking about what is useful in e-service delivery processes. From technical perspective, the most important ingredient of Amphora's success was the continuous enterprise integration with other systems, which allows to extend functionality of the EDRMS in a well-defined way. Integration is basically the process of combining smaller components into a system that functions as one. It is important to emphasize that the main aspect is not about interoperability (in the sense of interaction between different systems), but an integration that allows to add new functionalities to the EDRMS or other information system: The actual integration is achieved not only through direct data exchange between information systems, but also through the unification of meta-data, i.e., the "normalization" of data. Such functions are promising for the perspective application of EDRMSs as IS ensuring exchange between all other ISs in an organization. Another technical aspect is what we call extension modularity of the solution, which allows to deliver functionality for different deployment instances of the product through well-defined extension points. This way Amphora avoids artificial complexity due to excessive functionality. With this measure, Amphora decouples the maintenance of a product baseline from the maintenance of its several product instances—a lightweight and flexible way to approach the product line engineering problem in software engineering, compared with [19,20].

4.2 Organizational Enabling Factors

One of the main organizational success factors of the EDRMS Amphora was the principle of strict efficiency orientation pursued by the local governments, i.e., local governments always target a maximum efficiency, in particular, cost-efficiency, in the procurement of software systems. In times of sparse funding and a legacy of several different types of existing information system, local governments were eager to go for any information system that supports the integration with other information systems. This feature was enabled through Amphora where any development was automatically scaled to all of the Amphora users. Another crucial organizational factor was that the EDRMS deployment was tightly coupled with a streamlining conceptual framework of paperless management in the local governments, called the e-LocGov model [21]. This rigorous framework was developed and used by Amphora developers for the improvement of the EDRMS implementation process. This meant that Amphora was delivered with business re-engineering features (such as the flexible adoption of workflows) plus instructions for out-of-the-box organizational transformations, supplemented by corresponding e-training modules [22].

4.3 Beyond Intra-organizational EI

The backbone of Estonian e-governance is the cross-organizational data exchange layer X-Road (X-tee in Estonian) [23,24]. X-Road facilitates the cooperation

between affiliated organizations and databases and serves as an explicit tool for cross-organizational collaboration. X-Road relies on a mature eID [25] infrastructure and builds on the protocols and patterns of the standard SOA stack, i.e., Simple Object Access Protocol (SOAP) and Web Service Description Language. X-Road is the enabler for all of the Estonian e-assets: digital signing [26], digital archiving [27], e-voting, e-banking, e-health, e-school, e-taxation, etc., via its document exchange center (DEC) [28] component, and X-Road is also the crucial enabler for digitally exchanged documents. During the initial stage of this study, the document exchange center (DEC), which is an interoperability solution for digital document exchange, was designed and realized. Thanks to the DEC, the rate of paperless official communications reached 93% in 2015 [26,28] and by now reaching more than 97%.

5 The EDRMS Amphora: Interoperability and Integration

An Enterprise Resource Planning (ERP) software has changed the dynamics of managing a business. The ERP systems are used by organizations in order to make sure that their businesses are managed through an integrated and centralized system [14]. Organizations that need to manage their customers, employees, and their inventories can benefit a lot from the features of an ERP. It is a software that basically helps different organizations in order to collect the required data, store it, manage it accordingly, and then interpret. This data is then utilized for several different functions, for example, Amphora is introduced to manage all the functions such as accounting, for the payments, sales, and most importantly the HR functions, in order to keep track of the official interactions and communications etc. The advent of digitalization through the introduction of EDRMS has changed the way business conducted in several ways. Organizations now find themselves overwhelmed with huge amount of paperwork which makes the storage as well as the maintenance of the document a huge challenge. Manual documenting management requires long hours, typing of redundant information, correcting the errors, misfiling, managing the physical storage locations of these documents, and also locating these documents all across different email inboxes of the employees. It is through the integration of both the EDRMS and the ERP system in an organization that can help maintain the paperwork as it aids in increasing efficiency while making the data more manageable and easily accessible.

5.1 Machine-Readable Documents as the Basis for Business Process Automation

The integration of EDRMS and ERP also gives rise to the need of creating more machine-readable documents. A document managed by an EDMRS system and then further readable by machine can help automate lengthy processes

while reducing the risk of discrepancies and loss of crucial data. The machine-readable documents are semantic documents. It is important to note that the machine-readable documents are not the same as digital documents. Introduction of systems such as Amphora can serve as the basis for the transformation to machine-readable data (Fig. 1).

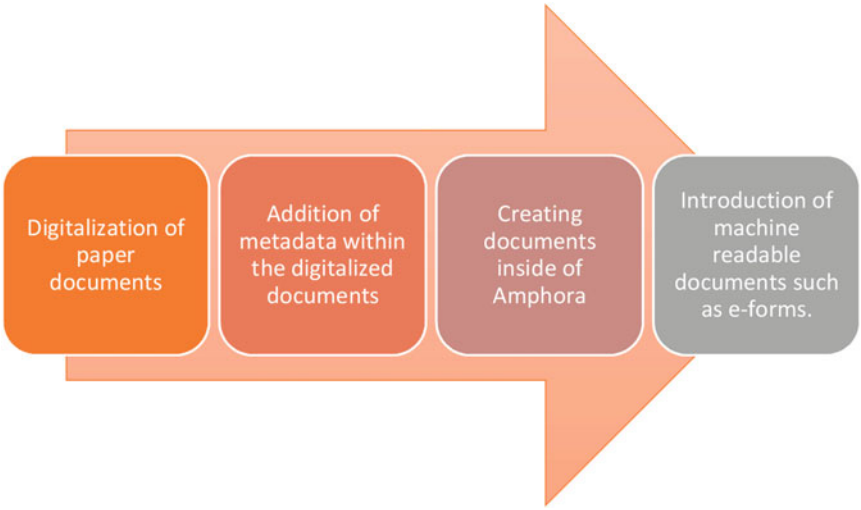


Fig. 1. Digitalization through Amphora

The difference between the semantic documents and the digital documents is even more so as compared to the difference between the digital documents and the paper-based documents. Digital document can be defined as the human-readable version of a semantic document. Therefore, it can be said that semantic documents are from the generation of documents that can be electronically processed and understood. The digital documents, in their current form, do not meet the requirements of the Semantic Web. The resources for Semantic Web need to be unique in order to make sure that they are identified uniquely with content that is accessible easily and annotated with the meta-level descriptions which can be processed in the machines. The contents of the digital documents mostly include schema elements which are often not easily accessible from outside of that document. The need to transform the digital documents into documents which can serve as a resource for the Semantic Web leads to the requirement of semantic documents, which can be termed as the new form of document. There are many different technologies that exist and can be used as a bases for the formation of semantic documents, such as the RDF and ontologies. The most common strategy to convert digital documents into semantic resources includes annotating these digital documents with the ontologies. Please check the clarity of the sentence ‘It is important to understand that in order to take advantages

of the semantic documents...'.It is important to understand that in order to take advantages of the semantic documents, is not to just add in annotations for the digital documents but also integrating the two representations of the knowledge in order to create documents that are both readable by humans and process able by machines. This will help in the creation of the tools that are independent of one unified view for the document data [29].

5.2 Importance of EDRMS and ERP Integration: Example of E-Invoicing

E-invoicing can be explained as an exchange of the invoice document between the buyer and the supplier through an electronic format which is integrated. It is not similar to sending and receiving digital invoices. An e-invoice is basically a document or a data file which is standardized and machine readable so that it can be processed by the ERPs system without the use of any additional software. This helps in eliminating the use of paper and the physical distribution of invoices which also helps for a greener environment [30]. It is important to understand and eliminate all the barriers that are present for the implementation of e-invoicing in particular. There are several benefits of integrating the EDRMS and ERP in an organization to automate its processes. For example, Estonia used EDRMS in order to implement a crucial process of e-invoicing. The approach utilized in this case was the integration of the workflow models with the ERP interfaces. With this integration, organizations do not require financial systems separately, and there is an increase in the efficiency of the system, whereas the expenses are visibly reduced. The integration of efficient document management system with the ERP system of an organization can help transform the way the employees work with documents. Required documents can be linked to make sure that the invoices as well as the orders are regrouped easily which makes the workflow faster and easier to manage. Once all of the important information is stored on an EDRMS, manual processes are eliminated; for example, the xml e-invoice systems as introduced in Estonia are based on the standardized data. Payments received while giving the capability of tracking the overall processes and costs have been made easy [31]. There are several added benefits that come with the integration of EDRMS and ERP; for example, less time is wasted in looking for the required document or data. Employees can enjoy the benefits of working remotely while also benefiting from the eco-friendly nature of electronic documents. ERP and EDRMS integration not only help organizations achieve a greener workplace but also helps in reducing costs. Audits, both external from the government and internal from the organization, are easier as company's compliance to the legal and regulatory structure is revealed immediately. The companies can also easily trace their profits, their payment, and employee contributions.

6 Interoperability and E-Government

E-government interoperability can be defined as a process that enables the heterogeneous information systems or the components comprising them to be managed by external partners or different jurisdictions according to the predefined terms and specifications [31]. It is through interoperability that effectiveness is enhanced through an interconnected solution. It not only reduces cost but also provides with prompt responses due to easy access to required information.

6.1 Principles of Interoperability

There are three different dimensions that help analyze the principles related to Interoperability.

6.1.1 The Organizational Interoperability It describes the capability that an organization or a government has to utilize in order to complete the information systems [32]. For example, the EDRMS Amphora as implemented by the most of the local governments in Estonia is used for interaction by both internal and external entities. EDRMS Amphora has contributed dynamically for the creation of electronic documents, digital signing, and their exchange between the government agencies which improves the business process and the document management. For example, it has created an interface for communicating with the national registries which has also enabled the government to manage e-invoice workflow. The work routines of employees are improved which results in high quality of services with the same number of officials. This also increases the speed in which the applications by citizens and entrepreneurs are processed. Hence, it improves the efficiency of government processes which in return is beneficial for the citizens of the state.

6.1.2 The Semantic Interoperability It indicates the capability that different organizations or a government have to understand the data that has been exchanged in the same manner. This presumes that there needs to be a proper mechanism followed for the presentation of the data including the data definitions [33]. For example, the document exchange center (DEC) component of X-Road allows the use of documents and data stored digitally. DEC is the interoperability solution for the exchange of digital information. It is through DEC that the paperless communication is enabled and enhanced.

6.1.3 The Technical Interoperability It represents the infrastructure and the software applied. Infrastructure interoperability is the capability of the hardware that is being used by varying organizations in order to work in an interconnected manner. This can be made sure through the PKI infrastructure, whereas the software interoperability can be defined as the ability of the software that is being used in varying organizations in order to successfully exchange data.

In order to achieve the software interoperability, it is of eminent importance to make sure that there are data protocols being followed and that a software is developed which is necessary for managing these defined data connections, while also creating user interfaces to facilitate communication among different organizations involved [34].

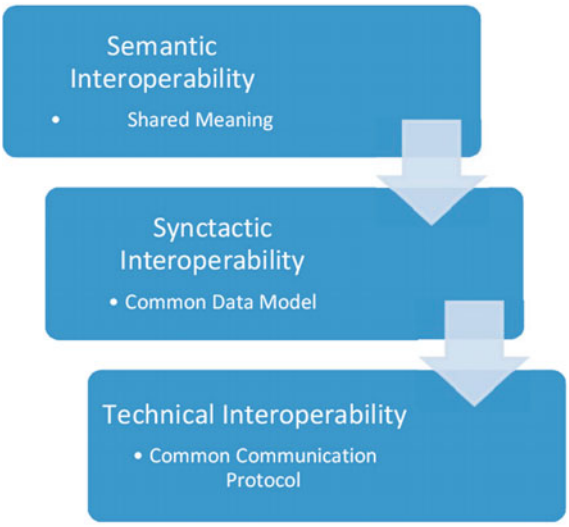


Fig. 2. Electronic government and the information systems perspective[35]

6.2 The Complexity of Choice

It is important to make sure that interoperability is maintained at a high level in organization in order to provide support to the record keeping goals as set by the management. This is because the EDRMS allows organizations to utilize software that is both flexible and configurable. It is also important to understand that there are several options and fields which are user-definable in EDRMS. The challenge is to figure to the congruence in the metadata elements across the varying systems, to manage random changes and configurations. The advantages of interoperability and integration cannot be gained if the configuration design is poor and not proper. Organizations need to take into consideration the way forward in order to adhere to the standards that are relevant and best for the standardization process in an organization (Fig. 2).

6.3 EDRMS as an Essential Factor for Interoperability

The interoperability of document management systems represents the capability that these systems have in order to exchange as well as manage the digital

documents. It is through EDRMSs that information is shared without any need for the postal services or paper forms. In order to make sure that these systems are interoperable within the local and the central government, there are specific considerations that are required to be followed.

Further development of the EDRMS is important in order to make sure that the data and documents can be transported. It is important to have the descriptions for documents which are XML-based and also the metadata for the document completion. It is required to make sure that the public sector document management systems have the ability to communicate with its environment of the citizens. The establishment of the e-services which are integrated and advanced solutions require high level of e-government interoperability. There are several political barriers such as the inadequate definition of the required guidelines and policies which need to be overcome. At the organizational level, employees often lack the required skills and experience, and hence, successful integration of an EDRMS helps. There is also a lack of resources often at times, which limits the provision of required information.

7 Business Process Re-engineering: Implementing EDRMSs

The business standards as well as the required processes of re-engineering stress on the important role of the document control and data flow. The documents that are used by the processes or generated by them need to be managed in a way that they reduce the non-conformance, while providing updated information and ensuring that only the approved documents are utilized by organizations [34]. An EDRMS constitutes of all the tools necessary for managing the businesses and the internal flow of information. There are several roles that EDRMS can play in business process re-engineering such as the storing, managing, and archiving of data. As the complexity within a document increases, it becomes more important for organizations to introduce automation of document processing. These engineering processes should include the documents, design, and data relations. In order to integrate and also conform to the standards, companies need to make sure that they have employed the required EDRMS solutions which effectively manage the documents (Fig. 3).

7.1 EDRMS to Redesign and Automate Business Processes

The automation of business processes means that documents are available where required in an easy to access and quick manner. This helps the employees to work smartly while getting things done and approved faster. It also allows the employees to view their steps and track them accordingly. However, the first essential step in automating is the redesigning of the processes. The most inefficient of the processes can be eliminated by flowcharting the process and implementing a proper workflow, through mapped information. It is through an EDRMS that the necessary information is then available at the desktop of the worker. The

retrieval as well as the research capabilities of an EDRMS lets the employee win back time by not wasting it to find paper-based files. It is through the utilization of electronic documents and automation systems that the business processes can be revolutionized. It is currently being used in several industries and government agencies. EDMS is used at both the strategic level and the operational level.

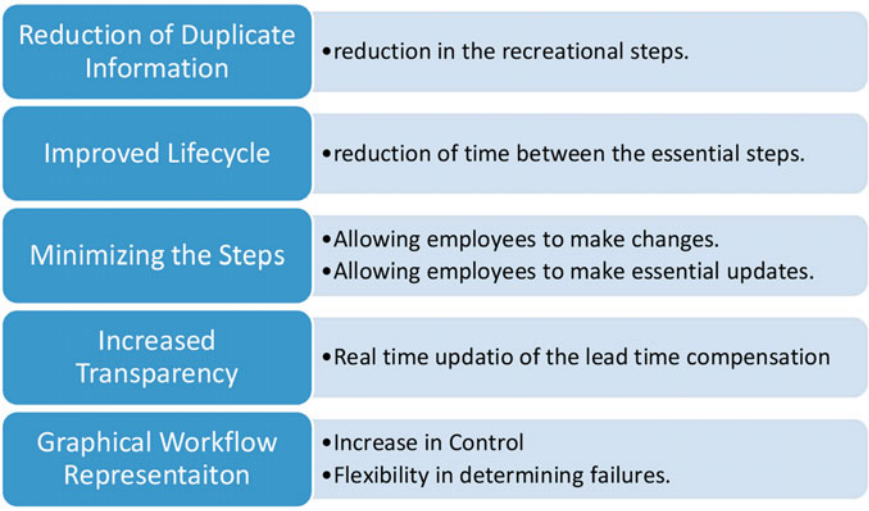


Fig. 3. A philosophy of the electronic document management [36]

8 Future Directions in the EDRMS

Over the last decades, the implementations of the EDRMS has seen different stages of maturity. In the early stage, the introduction of EDRMSs was about ad-hoc, one-to-one transformation of paper-based processes into e-document-based processes (often simply on the basis of scanned documents); the second phase was about introducing genuine electronically exchanged documents-based processes; the third phase, that we are currently still in, is about systematically moving toward structured data and meta-data. Beyond the current stage, the next phase will be in AI-based, automated decision support. AI is the future, and in order to adapt to the future concepts, it is important to stay updated through the use of machine-readable data which has the ability to be automated. Automatic decision support requires a new architecture of EDRMSs. It needs a stringent, fully orthogonal design [37], i.e., all essential building notions (versioning, access rights management, collaborative editing, meta-data, etc.) need to show in each respect, mutually and the finest granularity level. Furthermore, a fully typed, transclusion-based [38] EDRMS is needed (a calculation graph of information) to enable the full potential of AI tools for EDRMSs.

9 Conclusion

The integration efforts of the EDRMS Amphora made it crossing the borders of its original functional domain—electronic document and records management. The application case Amphora in the Estonian local governments is an example of an efficient EDRMS-based EI. We identified that the technological and organizational enablers for EDRMS-based EI include service-oriented thinking, continuous enterprise application integration, extension modularity, strict efficiency orientation, and finally the conceptual framework of paperless management. In the investigated case, intra-organizational EI and cross-organizational EI are mutually dependent. The enablers for cross-organizational EDRMS-based EI include cross-organizational data exchange layer and document exchange center. Given that machine learning techniques become ever more mature, we view the integration of automatic decision support into an EDRMS—based on a more stringent and orthogonal EDRMS architecture—as the most important future direction in the development of EDRMSs.

References

1. Schulte, R., Natis, Y.: Service Oriented Architectures, Part 2. Technical Report Gartner Research ID Number SPA-401-069. Gartner Group (1996)
2. Schulte, R., Natis, Y.: Service Oriented Architectures, Technical Report Gartner Research ID Number SPA-401-068. Gartner Group (1996)
3. Draheim, D.: The service-oriented metaphor deciphered. J. Comput. Sci. Eng. **4**(4), 253–275 (2010)
4. Draheim, D.: Business Process Technology—A Unified View on Business Processes, Workflows and Enterprise Applications. Springer, London (2010)
5. Ousterhout, J.K.: Scripting: higher level programming for the 21st century. Computer **31**(3), 23–30 (1998)
6. European Commission Model Requirements for the Management of Electronic Records. European Communities (2008)
7. Draheim, D., Pappel, I., Tsap, V., Pappel, I.: Exploring e-services development in local government authorities by means of electronic document management systems. In: Electronic Governance and Open Society: Challenges in Eurasia: 5th International Conference, pp. 14–16 (2019)
8. Morrissey, J.: Regimes of performance: practices of the normalised self in the neoliberal university. Br. J. Soc. Educ. **4**, 614–634 (2015)
9. Pappel, I., Pappel, I., Tampere, I., Draheim, D.: Implementation of e-invoicing principles in Estonian local governments. In: The Proceedings of 17th European Conference on Digital Government ECDG, p. 127 (2007)
10. Vakola, M., Yacine, R.: Critique of existing business process re-engineering methodologies: the development and implementation of a new methodology. Bus. Process Manage. J. **6**, 238–250 (2000)
11. Pappel, I., Tsap, V., Draheim, D.: The e-LocGov model for introducing e-governance into local governments: an Estonian case study. IEEE Trans. Emerging Topics Comput. 1 (2019)
12. Pappel, I., Pappel, I., Tepandi, J., Draheim, D.: Systematic digital signing in Estonian e-government processes. In Transactions on Large-Scale Data-and Knowledge-Centered Systems, pp. 31–51. Springer (2017)

13. Vernadat, F.: Technical, semantic and organizational issues of enterprise interoperability and networking. Annu. Rev. Control **34**, 139–144 (2010)
14. Burr, B., Akshay, P., Kevin, S., Miyake, N.: System and method for parameterizing documents for automatic workflow generation. U.S. Patent 9,880,987 (2018)
15. E. M. of Economic Affairs and Communications. Digital Agenda 2020 for Estonia. MKM (2013)
16. Pappel, I., Pappel, I., Saarmann, M.: Development of information society and e-government by improving electronic records management solutions at Estonian local authorities. In: IADIS International on Conference e-Society—IADIS, pp. 457–462 (2012)
17. Pappel, I., Pappel, I.: The service-oriented state and local authority: service orientation in public administration, pp. 111–116, July 2012
18. Atkinson, C., Bostan, P., Draheim, D.: A unified conceptual framework for service-oriented computing. In: Hameurlain, A., Kung, J., Wagner, R. (eds.) Transactions on Large-Scale Data and Knowledge-Centered Systems VII—Lecture Notes in Computer Science. vol. 7720, pp. 128–169. Springer, Berlin, Heidelberg (2012)
19. Atkinson, C., Bayer, J., Bunse, C., Laitenberger, E., Laqua, R., Muthig, D., Peach, B.: Component-Based Product Line Engineering with UML. Addison-Wesley (2001)
20. Atkinson, C., Draheim, D.: Cloud aided-software engineering—Evolving viable software systems through a web of views. In: Mahmood, Z., Saeed, S. (eds.) Software Engineering Frameworks for the Cloud Computing Paradigm, pp. 255–281. Springer, New York, Berlin, Heidelberg (2013)
21. Pappel, I., Pappel, I.: Methodology for measuring the digital capability of local governments, pp. 357–358, Sept 2011
22. Pappel, I., Pappel, I.: Using e-learning methods in (e-)implementation of e-governance software at local governments. Case study in Estonian way. Feb 2011
23. Kalja, A., Robal, I., Vallner, U.: New generations of Estonian e-government components. In: 2015 Portland International Conference on Management of Engineering and Technology (PICMET), pp. 625–631, Aug 2015
24. Paide, K., Pappel, I., Vainsalu, H., Draheim, D.: On the systematic exploitation of the Estonian data exchange layer x-road for strengthening public-private partnerships. In: Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance, ICEGOV '18, pp. 34–41. ACM, New York, NY, USA (2018)
25. Tsap, V., Pappel, I., Draheim, D.: Key success factors in introducing national e-identification systems. In: Dang, T., Wagner, R., Kung, J., Thoai, N., Takizawa, M., Neuhold, E. (eds.) Future Data and Security Engineering, pp. 455–471. Springer (2017)
26. Pappel, I., Oolu, K., Saarevet, K., Lauk, M., Draheim, D.: The digital archiving process in Estonia—Assessment and future perspectives, pp. 472–481, Nov 2017
27. Draheim, D., Koosapoeg, K., Lauk, M., Pappel, I., Pappel, I., Tepandi, J.: The design of the Estonian governmental document exchange classification framework. In: Ko, A., Francesconi, E. (eds.) EGOVIS'16—The 5th Conference on Electronic Government and the Information Systems Perspective. vol. 9831, pp. 33–47. Springer, New York, Berlin, Heidelberg (2016)
28. Tunjic, C., Atkinson, C., Draheim, D.: Supporting the model-driven organization vision through deep, orthographic modeling. EMISAJ (Enterp. Model. Inf. Syst. Archit. J.) **13**, 1–39 (2018)
29. Koch, B.: E-invoicing/e-billing. Digitisation Autom. (2016)

30. Chen, Y., Chen, H., Huang, W., Ching, R.K.: E-government strategies in developed and developing countries: an implementation framework and case study. J. Glob. Inf. Manage. (JGIM) **14**, 23–46 (2006)
31. Pearson, S.: Privacy, security and trust in cloud computing. In: Privacy and Security for Cloud Computing. Springer (2013)
32. Gottschalk, P.: Maturity levels for interoperability in digital government. government information quarterly. Govern. Inf. Q. **26**, 75–81 (2009)
33. Hodges, J., Kimberly, G., Ray, S.: Semantic development and integration of standards for adoption and interoperability. Computer **50**(11), 26–36 (2017)
34. Downing, L.: Implementing EDMS: putting people first. Inf. Manage. **4**, 44–46 (2006)
35. Draheim, D., Koosapoeg, K., Lauk, M., Pappel, I., Pappel, I., Tepandi, J.: The design of the Estonian governmental document exchange classification framework. In: Electronic Government and the Information Systems Perspective: 5th International Conference, pp. 33–47 (2016)
36. Volarevic, M., Strasberger, V., Pacelat, E.: A philosophy of the electronic document management. In: ITI 2000. Proceedings of the 22nd International Conference on Information Technology Interfaces (Cat. No. 00EX411), pp. 141–146, June 2000
37. Nelson, T.: The heart of connection: hypermedia unified by transclusion. 31–33 (1995)
38. Siu, B.: EDRMS—Answers and solutions for better records and information management. IQ: RIM Q. **33**, 24–26 (2017)

Robotic Tool to Multi Time-Lapse Videos Creation

João Roberto de Toledo Quadros(✉) and Roberto de Castro Souza Pinto

Department of Computer Science, Federal Center for Technological Education Celso Suckow da Fonseca, Maracanã Ave. 229, Rio de Janeiro, Brazil
joao.quadros@cefet-rj.br, robertcsp@gmail.com

Abstract. There is scientific research in several areas of knowledge that make use of the time-lapse technique. Tidal, rocks, and glaciers movements, which take up to months to occur, can be seen in a few seconds with this technique. Swamps, deserts, and even submerged areas are places of difficult access and need to be observed. There is always a high cost involved in the process of capturing time-lapse images. In this paper, a robotic tool is presented that meets the operational requirements to support these tasks. Among these new features is the ability to use idle time in the process of capturing images for this type of media. Its built-in camera is moved in different directions, being able to capture images for the generation of multiple time-lapse videos, using only one device. Because of this, this robotic tool is a facilitator to the construction of scientific, entertainment, and educational projects that make use of this type of unconventional data.

Keywords: Robotics · Entertainment · Time-lapse video

1 Introduction

Time-lapse videos have been widely used around the world as a tool to aid scientific research and entertainment projects, since this kind of image is an important means of obtaining data of various natural phenomena [1]. For example, flower growth and tidal movement are essentially very slow, take a long time to change from one state to another, and are best observed through time-lapse video [2, 3].

Time-lapse videos are multimedia data, which use similar technology applied to cinematography, in which motion is created by a sequence of frames captured over an extended period and visualized in a shorter interval [1]. Research areas that make use of this technology include geology, oceanography, meteorology, biology, and astronomy.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_4

There are some examples to time-lapse videos usage, such as rocks movement and desert whirlwinds (dust devil) observation research in California's Death Valley [1, 4], monitoring of glacier movements, eruptions, and outburst floods in Fuego-Patagonia in Central Chile [5], observation of the ammunition dumping effects on Hawaii's marine life after WWII [6], observation of marine species habits and tidal movement in the mangrove ecosystem in Japan [7], and others.

To capture the images and produce a time-lapse video, it is necessary to use cameras configured in continuous shooting mode and use special equipment with sensors and other components to control the camera to take the shots. As seen in Lorenz et al. [1], the cost of the entire process of producing these videos is a relevant factor in a monitoring project. Most of the time, the equipment is installed in remote and hostile locations for the camera and other hardware components. Deserts, swamps, mangrove areas, and even the seabed are environments that hinder the installation process, equipment maintenance, exchange of batteries, and acquisition of the data, due to the difficulty of moving people and vehicles in these environments. As seen in the papers that were analyzed, the maintenance and support costs of the tools used to obtain this multimedia data are the main difficulties in this type of research. A tool that enables research teams to reduce time and cost of maintenance becomes essential.

In Lorenz et al. [1], it was also observed that 45 s was the shortest time interval between frames of time-lapse videos analyzed in this research, which means that equipment used to capture images is idle 45 s after each shot. This time interval was the shortest found in the time-lapse video papers observed in the literature review.

From the aforementioned requirements, it was defined that one aim of this work was to develop a robotic tool that optimizes this idle time, using the remaining time for new photographs. This is due by re-positioning the tool and the camera, which is on the same platform, without manual movements. Thus, this tool can capture new images from different angles and perspectives and produce multiple time-lap videos (Fig. 1).

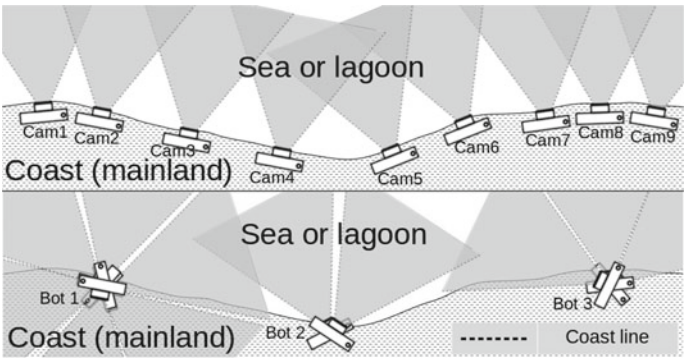


Fig. 1. Two different projects configuration to monitor a coastline area. One with the standard configuration and the other with the solution proposed in this work

This way, it is possible to accomplish time-lapse video projects with greater accuracy and cost optimization using one robotic tool but creating several time-lapse videos, rather

than just one. In this figure, two typical patterns of use of photographic devices to obtain time-lapse videos are shown. In the first one, nine photographic cameras with an external gadget to control them are necessary to take photos to produce time-lapse videos, with each camera being allocated at only one angle, in need of manual help to re-allocate the camera. In the other scenario, only three devices must be used, each one with an allocated camera. As the tool moves by itself, there is no need for local manual assistance to change the camera position. The tool has been developed to be able to move horizontally and vertically so that it is possible to take many time-lapse videos from the same coordinate.

In this paper, Time-Lapse Robot (TilBOT) is described as a robotic tool designed to accomplish XML photographic projects, control and move an attached photographic camera, and capture images to generate multiple time-lapsed videos, with only one equipment and from the same coordinates. The goal is to show that this tool can be a low-cost alternative and easy manipulation for works that require both scientific and entertainment time-lapse videos. TilBOT will allow to optimizing the use of time-lapse video equipment to reduce the cost of this kind of project.

In addition to this introduction, Sect. 2 is a brief review of related work. Multi time-lapse video description is presented in Sect. 3, while TilBOT hardware and software are described in Sect. 4. Section 5 presents applied test data and a comparison with some time-lapse videos. Finally, Sect. 6 gives a final consideration of this work.

2 Related Work

The study of geophysical and meteorological phenomena using time-lapse video is addressed in Lorenz et al. [1]. This work describes the difficulties of obtaining images, studying, and understanding short-lived phenomena such as desert dust devils, a whirlwind rendered visible by lofted dust in Gates Pass in Tucson, Arizona, the USA, that are responsible for damages and even aircraft accidents [4]. Transient flooding and sliding rocks of Racetrack Playa in Death Valley, California, United States, are also covered in this paper. This phenomenon occurs due to the action of strong winds and a thin layer of ice during the winter in a lakebed 1130 m above sea level that is dry almost all over the year.

According to the authors, using a video camera is not convenient to be in an observation station, and this type of camera has a higher frame rate than necessary for their kind of research. Using a PC webcam is not appropriate as well, due to the necessity to have a power connection and a shelter for all components. A homemade solution with a modified digital camera would be suitable, due to the large capacity of memory cards and the use of a compact battery without the necessity of an AC power. The digital camera must have their internal circuits modified to support its activation by some external device. Modifications must be made to the camera's shooting mechanism and battery connection, to trigger the camera according to the project's necessity and to turn the battery on and off during camera idle time to save energy, respectively.

An electronic device was built and equipped with Picaxe microcontroller and transistors to control the camera accordingly. This microcontroller is programmed in basic programming language and is designed to control camera switches to take pictures and to turn the camera on and off enabling great flexibility to alter software parameters easily without the impact of hardware modifications.

Initially, two options were considered, the Brinno GardenWatchCam camera, designed for horticultural observation projects, and a more expensive Harbotronics solution. They were discarded due to lack of a motion trigger in the first case and Harbotronic because of the high cost of investment and equipment large size. Racetrack Playa, for example, is a national park, and camera location must be approved by park staff to avoid visual impact. Experiments were conducted with the homemade hardware connected to Flycam-One in Racetrack Playa to observe dust devils. And then, new images were captured successfully in Arizona desert with VistaQuest VQ1005 camera.

Another paper investigates the marine life observation by two time-lapse camera systems in a site south of Pearl Harbor, O'ahu, Hawaii, [6]. The effort to understand the natural behavior of fauna in contact with ammunition and chemical agent disposed of this area toward the end of World War II is the main purpose of this work. A time-lapse system from Woods Hole Oceanographic Institution Instrumentation Multidisciplinary in Support of Oceanography (WHOI-MISO) equipped with Nikon™ Coolpix 995 camera and KidCam and a low-cost time-lapse equipment, developed by three high school students and equipped with a Canon™ SD1000 camera, has been deployed under the sea to capture images of many nearby species in contact with materials and debris thrown into the sea

In Llanos et al. [5], homemade equipment is described and used to monitor glacier behavior and the phenomena associated with it in Grey Glacier in Torres del Paine National Park, Chilean Patagonia. Instead of using a commercial time-lapse system, they opted to build a low-cost solution inside a weatherproof plastic box with a Nikon™ D5500 camera connected to a Raspberry Pi 3 microcontroller and a WiFi access point to transmit pictures to a hotel 15 km away and so, allow real-time observation.

3 Multi Time-Lapse Videos

Multi time-lapse video is an innovative technique performed by TilBOT that uses its capacities to move the camera horizontally and vertically for different points of view. Images are captured with a single camera from the same coordinate in motion cycles. This technique is very similar to time-lapse video and takes advantage of the equipment idle time, inherent to the time-lapse videos to move the camera and capture images to create more than one video.

In the diagram, in Fig. 2, a sequence of the times of a multi time-lapse project with total N cycles of movements of three shots each that generates three time-lapse videos is shown. In this figure, t_{p} is the time the photographic camera takes to take each picture. t_m and t_n are the time interval TilBOT waits to move the camera to the next position to take the next shot, t_i is the time interval for TilBOT to move the camera to the initial position and to start a new cycle of shots. T_1 is the time from the beginning of a shot to the beginning of the next one, while T_{C} is the total time of a cycle. $T_{\mathrm{T}} = N \times T_{\mathrm{c}}$ is the total time of a multi time-lapse project (time intervals to take a photo are disproportional on purpose to allow easy visualization).

Figure 3 shows an example of a multi time-lapse project in which TilBOT is configured to capture images to generate three time-lapse videos that cover an area 180° horizontally. It moves the camera and takes pictures at 0°, 60°, and finally 120°, completing one cycle. Then, it moves the camera back to 0° (i.e., initial position), occasionally

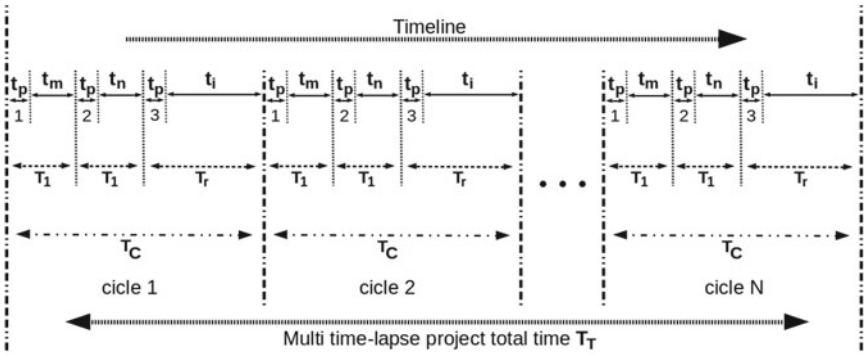


Fig. 2. Example of multi time-lapse video, made by TilBOT

waits some time, and starts the next cycle again. Table 1 shows how pictures from a multi time-lapse project are selected to generate three time-lapse videos, 0°, 60°, and 120° angles, respectfully.

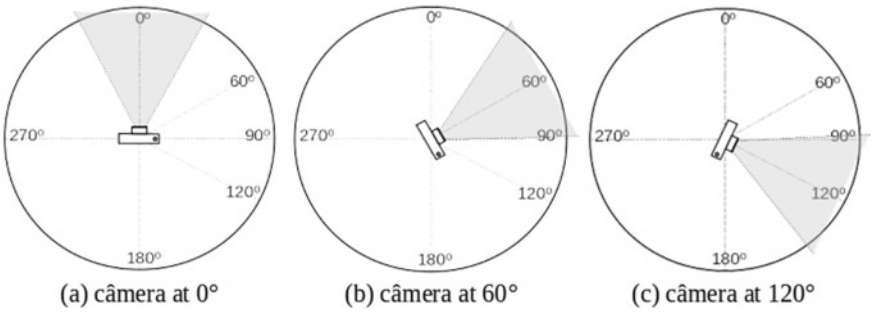


Fig. 3. a–c Shows the superior view of how only one TilBOT would position its camera to capture images to the generation of three time-lapse videos from the same coordinates. The cycle of shots begins at 0°, 60°, and finally 120° when TilBOT returns to the initial position of 0° and waits the programmed time to restart a new shot cycle

Table 1. Multi time-lapse video with N pictures dismembered into three time-lapse videos

Three time-lapse videos dismembered						
Dismembered	Selected pictures					
First time-lapse video	1	4	7	10	13	$N-2$
Second time-lapse video	2	5	8	11	14	$N-1$
Third time-lapse video	3	6	9	12	15	N

4 Hardware and Software Design

TilBOT is composed of two main parts, the support and movement module and the camera support module. The latter is equipped with support where the photographic camera is screwed on it. The software consists of two parts, a mobile application in which the user creates the photographic project and embedded software on the TilBOT circuit responsible for the execution of the project.

4.1 Hardware Shield

Figure 4 shows two modules, which are the two most important and largest modules. They are the basis for both the electronic components and the camera. The support and movement module is equipped with two DC motors with wheels on each side of the module responsible for TilBOT traction. There is also a wheel on the back of this module and besides making the third support with the ground; it is responsible for controlling direction. An electronic circuit is docked inside this module composed of microcontroller, connections for all motors and actuators, and a camera control cable. All TilBOT components are powered by three 3.7 V lithium-ion batteries, were also stored inside this module, and allow a range of ten continuous hours.

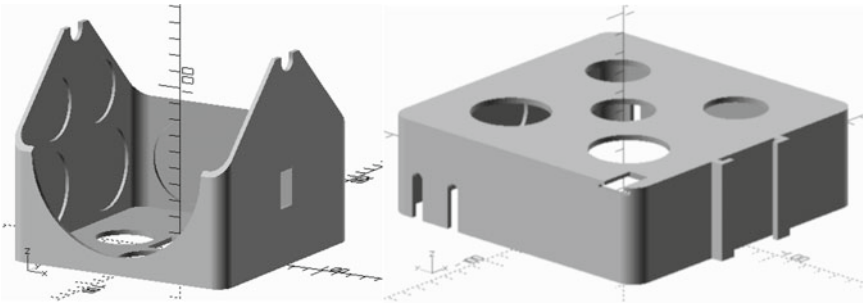


Fig. 4. TIlBOT support and movement module designed to fit DC motors and a driving wheel. The center hole on top of it is used to fit an axle to connect cables, while the other four holes on it are intended for weight reduction only

The camera support module is located above the support and movement module with a bearing between them, allowing this module to rotate 360° gently. The first prototype was built to support a Canon™ PowerShot G15 compact camera. This prototype used an ATMega328p microcontroller. Figure 5 shows the prototype of the TilBOT configuration with all hardware installed. It was equipped with a PowerShot G15 Canon camera in all experiments for this work.

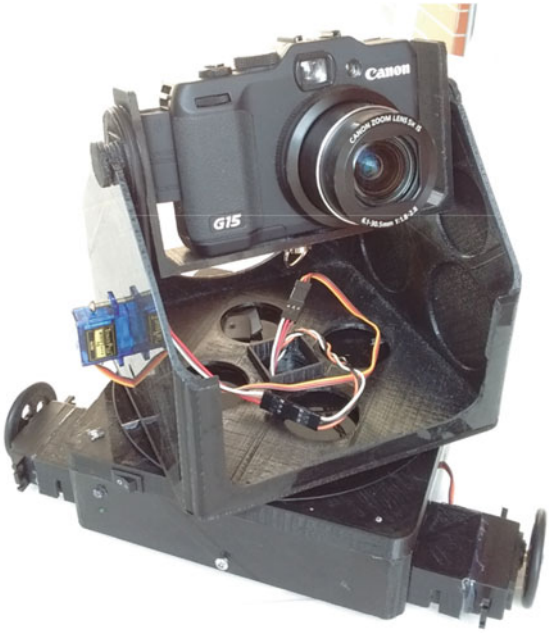


Fig. 5. TilBOT prototype assembled with all its actuators

4.2 TilBOT's Software

There are two main components in TilBOT's software, an application for a mobile device called AppTilBOT and a "C" software embedded in the microcontroller board on support and movement module. The MIT App Inventor platform [8] is a code Web application designed to allow the development of software for mobile applications with Android operational system, through the Google Chrome browser interface. It is a graphical environment, through which the developer, instead of coding the programs by command lines, does it moving graphic blocks representing the language functions. The App-TilBOT application was developed on the MIT App Inventor platform and runs as an Android app [9]. The user configures the photographic project and sends it to TilBOT software to be executed.

4.3 Photographic Project

Photographic project is a term defined in this work and consists of an XML message with a group of instructions to be accomplished by TilBOT, to use its camera to take the pictures that will make up a multi time-lapse video. These instructions are commands for the TilBOT to perform actions such as direction and speed of TilBOT itself, horizontal and vertical camera movement, and the order to take a picture.

Figure 6 shows the photographic project for the generation of 360° videos. This data will be used by the TilBOT embedded system to make movements and controls the camera to perform the videos. A photographic project is configured by the mobile

device application AppTilBOT and then transferred to TilBOT embedded software to be executed.

```
<?xml version = "1.0" encoding = "UTF-8"?>
<proj-foto-360>
   <general>
      <shutter-speed>1/125</shutter-speed>
      <set-focus>true</set-focus>
   </general>
   <upper-vertical-angle>20</upper-vertical-angle>
   <lower-vertical-angle>140</lower-vertical-angle>
   <number-vertical-pictures>2</number-vertical-pictures>
   <number-horizontal-pictures>6</number-horizontal-pictures>
</proj-foto-360>
```

Fig. 6. This is a photographic project to create 360° videos. Camera shutter speed was adjusted to 1/125 and with vertical camera movement between 20° and 140°

5 Experimental Evaluation

The evaluation process, for this paper, lasted five months and executed 23 multi time-lapse projects generating 67 videos. TilBOT camera took a total of 1889 photographs in the whole evaluation process. These tests showed that TilBOT was able to generate multiple time-lapse videos with a time interval between frames shorter than all time-lapse video intervals observed in the related works. Even when a multi time-lapse project configured with six time-lapse videos, TilBOT was able to generate time-lapse videos with intervals of 34 s as shown in project number 22 in Table 2. This table shows all data regarded to evaluation tests performed to create multiples time-lapse videos for this work.

Evaluation tests numbers 1–9, 11–13, 15, 16, 21, and 22 were performed only to take pictures to collect timestamp data and so calculate the interval between shots and elapse time. The other time-lapse projects were performed to generate time-lapse videos, post-process them, and improve TilBOT software and hardware to solve errors found. Figure 7 shows the scenario created to perform photographic project number 23 whose characteristics were shown in Table 2. In this scenario, two clocks were laid 60 cm far from TilBOT which was in the middle of them so that, after taking a shot of one clock, it had to turn 180° to position its camera toward the other clock. This is the worst-case scenario to perform a photographic project to generate two time-lapse videos mainly because the robot tool has to turn the camera the maximum angle, that is, 180°.

Table 3 shows two columns with five images of the two time-lapse videos generated from multi time-lapse project #23 shown in Fig. 7. In this project, TilBOT captured images in an angle 180° from each other without interval between cycles of captures, that is, after taking the two first shots, it turned to the initial horizontal position immediately. The interval between images in both videos was 20 s. This is due to the time the tool takes to move the camera to the initial point.

Table 2. Multi time-lapse project performed, with #23 in the spotlight

Multi time-lapse project performed						
# Project	# Time-lapse videos	Camera angle (°)	Interval between shots (s)	Total # of shots	Project elapse time	Camera focal length
1	2	90	11	22	00:02:29	140
2	3	45	19	12	00:01:08	140
3	6	60	34	30	00:02:44	140
4	2	90	11	10	00:00:49	140
5	2	180	15	10	00:01:09	140
6	2	180	17	10	00:01:16	140
7	3	45	21	15	00:01:38	140
8	6	60	33	30	00:02:32	140
9	6	60	38	30	00:02:58	140
10	6	60	37	600	01:02:25	140
11	3	45	24	15	00:01:48	140
12	6	60	22	24	00:01:27	140
13	6	60	26	24	00:01:41	140
14	2	180	15	160	00:20:18	140
15	2	180	15	8	00:00:54	140
16	2	90	9	8	00:00:32	140
17	2	180	48	194	01:17:00	140
18	2	180	23	200	00:36:47	140
19	2	180	27	164	00:36:14	28
20	2	180	27	160	00:36:14	140
21	3	120	24	9	00:01:03	28
22	6	60	34	18	00:01:36	28
23	*2*	*180*	*20*	*200*	*00:31:56*	*140*
Totals	**67**			**1889**		

6 Conclusion

In this work, a robotic tool to capture images used to generate multiples time-lapse videos were presented. According to the various videos created, especially the ones generated in the photographic project test #23, the quality of these two videos allowed observers to identify details in the images, in spite of the worst scenario mounted for the test, two clocks 180° far from each other and the camera lens set to maximum focal length (140 mm).

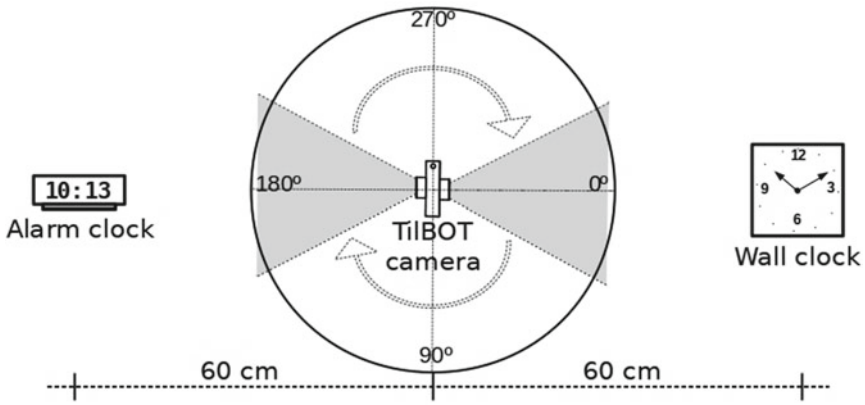


Fig. 7. About project #23. The first shot is taken at 0°, and then, it turns to 180° to take the other shot to complete the first cycle. Then, it turned again to 0° to start the next cycle

In addition to the multi time-lapse functionality, the use of photographic projects created flexibility to exchange projects between the mobile application and TilBOT itself, which will facilitate the change in the present projects and the creation of other types of photographic projects in the future.

The camera was connected through a remote terminal without a protocol between them, which prevented the robotic tool to know the camera precise state. So, delays were added before and after any shot to avoid camera shaking and camera movements before shots were completed.

In these evaluation tests and through the precise movement of the photographic camera, TilBOT were capable not only to capture the images used to generate time-lapse videos in a post-processing way but did this in a short enough time to meet the video generation needs of the works related and studied here, as well. In the future, with the use of a more modern camera and a better connection between the two devices, TilBOT can control it avoiding unnecessary delays, executing photographic projects with a shorter interval between frames.

Table 3. About project #23, which is composed of two time-lapse videos with one hundred images each, the two columns show the 1st, 25th, 50th, 75th, and last frame of the videos

Sample images from a multi time-lapse project #23		
# of frame	Frames from video 1	Frames from video 2
1st frame		
25th frame		
50th frame		
75th frame		
100th frame		

Acknowledgements. We appreciate the support provided to this research by CEFET/RJ, CAPES, and FAPERJ.

References

1. Lorenz, R.D., Jackson, B., Barnes, J.W.: Inexpensive time-lapse digital cameras for studying transient meteorological phenomena: dust devils and playa flooding. J. Atmos. Oceanic Technol. **27**(1), 246–256 (2010)
2. Bennett, E.P., McMillan, L.: Computational time-lapse video. ACM Trans. Graph. **26**(3), 102 (2007)
3. Liu, W., Li, H.: Time-lapse photography applied to educational videos. In: 2nd International Conference on Consumer Electronics, Communications and Networks, CECNet Proceedings, vol. 2(1), pp. 3669–3672 (2012)

4. Lorenz, R.D., Myers, M.J.: Dust devil hazard to aviation: a review of United States air accident reports. J. Meteorol. **30**(298), 179–184 (2005)
5. Llanos, P., Noceti, R., Casassano, C.: Real-time transmission of time-lapse imagery of glaciers in the southern Andes. In: IEEE International Symposium of Geoscience and Remote Sensing, vol. 1(1), p. 1517 (2017)
6. Edwards, M.H., Fornari, D.J., Rognstad, M.R., Kelley, C.D., Mah, C.L., Davis, L.K., Flores, K.R., Main, E.L., Bruso, N.L.: Time-lapse camera studies of sea-disposed chemical munitions in Hawaii. Deep Sea Res. Part II **128**, 2533 (2016)
7. Tezuka, N., Sea, I., Agency, E., Uchida, M., Sea, I., Agency, E., Shigeta, T., Sea, I., Agency, E., Fukatsu, T., Agriculture, N.: Tidal flat observation and monitoring using still video and network cameras. In: Techno-Ocean Conference, pp. 532–535 (2016). https://doi.org/10.1109/Techno-Ocean.2016.7890711
8. Kamriani, F.: App Inventor 2 Essentials a Step-by-Step Introductory Guide to Mobileapp Development with App Inventor 2. Packt Publishing, Birmingham (2016)
9. Johnston, C.: My Android Phone. Packt Publishing, Birmingham (2016)

Accomplishment of Multimedia Ontology for the Tunisian Archaeology Field

Degachi Haifa(✉), Yengui Ameni, and Neji Mahmoud

MIRACL, University of Sfax, Sfax, Tunisia
Haifa.degachi@gmail.com, Ameni_iseah@yahoo.fr,
Mahmoud.neji@fsegs.rnu.tn

Abstract. Interoperability is a key concept in what is known today as "data web" because it is by releasing the data from the application constraints that it is possible to easily navigate and react to additional information that complements the initial resource. However, while being released from the web applications, the information loses its contextualization and it will then be essential to document it by the addition of metadata. To cover the semantic richness of an object and the diversity of the information attached to it, there are computer ontologies designed to provide more meaning to the beacons and contextualize more information about a domain. Nevertheless, until now, we have not found any ontology for the Tunisian archaeological field. In fact, we aim to develop a multimedia ontology for the Tunisian archaeology domain, in order to structure the diversified knowledge involved in this field. The development of ontology will be realized in six steps which are described in this paper.

Keywords: Archaeology · MOWL · Multimedia ontology

1 Introduction

For an information retrieval project, we are in the process of developing an ontology of the Tunisian archaeology field, a termino-ontological resource that ensures (a) the model role of the domain that list all the relevant concepts and (b) the link between the concepts and the way they are named in the archaeologist's documents. This double function allows the annotation and the indexing of images of an archaeological site and the search for information in annotated images.

In addition, Tunisian cultural heritage is generally considered diversified, as well as the local monuments which are seen as rich and varied [1]. Thus, when speaking about the archaeology field, we are addressing a huge variety of documents. Thus, it becomes mandatory to organize and manipulate this huge quantity of information.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_5

Due to its advantages such as expressiveness and reasoning, the ontology is considered as the best solution to manipulate and structure data [2].

Gruber define ontology as "an explicit specification of a conceptualization" [3]. The ontologies are an artifact designed to express a common understanding of a set of concepts. Indeed, this conceptualization concerns the extraction of the vocabulary of a domain and gives a simplified view of the world which we want to represent. Through this conceptualization, each of the concepts and their mutual relationships make it possible to describe the real world [3].

Our approach shows that (1) the viability of such resource presupposes a precise articulation of the concepts and terms, and (2) such a prerequisite can be attained by the implementation of procedures based on the architecture of meta-modeling, which enables to model the set of the Knowledge Organization Systems and the structures of the necessary knowledge.

In the following section we detail the process that aims to allow these requirements.

2 General Process

Our objective is to create ontology of archaeology domain which enables to describe all the Tunisian archaeological sites. We benefit from the existence of several sources that describe this domain. Our proposal consists in constructing ontology and transforming non-ontological sources. This general process consists of three steps: (1) Analysis of Sources, (2) Transformation of Sources, and (3) Knowledge Base Merge.

2.1 Analysis of Sources

This preliminary phase of the process makes it possible to list the sources which will be considered in the continuation in future steps. In order to select the most adapted sources, it is important to specify why a source is suitable and which criteria to observe on a source to select it. We defined four principal criteria which we use to interact with the expert: (1) The reputation of the source, (2) The freshness of the source, (3) The adequacy of the source to the target, and (4) The clarity of the source mode.

Once the sources are selected, we analyze them and represent them according to the different levels of abstraction. In other words, this step aims at identifying what concerns the modeling of the data, the structuring of the implementation or of the syntax of implementation. Our objective in this process is to find common knowledge from several sources. The various sources do not necessarily have the same interest according to their quality. Therefore, the knowledge resulting from a good quality source will have more weight than the knowledge deriving from a lower quality source.

2.2 Transformation of Sources

Transformation Process This second step helps with the transformation of the non-ontological source (RNO) according to the elements determined in the first step. This transformation will be made according to the existing transformation patterns (if possible).

In order to obtain a source knowledge base (SKB) from any source, we define this process involves four steps:

- **Alignment**: the module and the source are aligned. In this way, it is possible to determine which elements in the source correspond to the classes of the module.
- **Extraction**: by using these correspondences, all the elements of the source being part of the studied domain are extracted.
- **Syntactic transformation**: in order to obtain only one format (RDF), a syntactic transformation is operated by following the selected transformation pattern.
- **Reengineering**: the extracted and transformed elements are associated with the ontological module in order to enrich it, by using a relation labeled "subClassOf" or "instanceOf". In this way, the extracted elements are represented by exploiting the modeling of the ontological module.

Transformation Pattern We try to use these transformation patterns by the implication of the ontological module in the process. We propose a disambiguation of the elements to be transformed by reusing rules of disambiguation adapted to the domain. Moreover, we propose to represent these rules in the form of an ontological module modeled by an expert and a transformation pattern of the methodology. Then, we adapt the creation of a transformation pattern by using the specific disambiguation in a domain [4, 5] by the use of a module. A transformation pattern is in the form of an algorithm of source transformation that can be assimilated to a set of rules to be applied in this algorithm in order to obtain the desired transformation. The expert defines a transformation pattern for each source by adapting the used ontological module. In this way, a source knowledge base (SKB) is generated for each considered source. These SKB represent the transformed sources according to the objective identified by the module.

2.3 Knowledge Base Merge

As presented in the general process, once the source knowledge bases are obtained, it is necessary to merge them to obtain the final knowledge base.

This phase of the SKBs merging is decomposed into several steps. The input of this process consists in the various SKBs that we consider to be merged. We want to obtain, as an output, the final knowledge base which is the result of the merger of the different SKBs. For this purpose, four steps are present in this process:

- **Alignment of the source knowledge bases**: This step makes it possible to create correspondences between all the pairs of the considered SKB.
- **Candidates' Generation**: From the correspondences present between the SKB, the candidates of ontological elements or the relations are generated. These candidates are potential elements that can be a member of the final knowledge base.
- **Calculation of the confidence**: A confidence score is associated with each candidate representing its consensual trust. This score is higher when the element is found in several sources.

- **Discovery of the optimal extension**: We consider a notion of incompatibility between candidates when they share common ontological elements. From these incompatibilities, we define an extension as a subset of non-incompatible candidates. This step allows the discovery of the optimal extension, which is the extension maximizing the confidence of the candidates.

The following section presents the six steps of creating the multimedia ontology oriented to the Tunisian archaeology field.

3 Steps of Creating Our Ontology

3.1 Step 1: Filtering on the Domain

The general idea of this automatic filtering would be the use of external resources to disambiguate the membership of a class to the studied domain. This domain would be limited by created the module. The extraction would be initiated by an alignment between the module sheets and the resource. This alignment can be made either by a terminological alignment technique or manually, for greater reliability. In addition, a lot of ontologies are already available in electronic versions with various forms directed towards several fields. Thus, we can benefit of those existing ontologies by using them as core or reference ontology. For our approach we use the CIDOC-CRM [6] ontology and the CARE ontology [7] as reference ontologies. According to this initial step, the extraction of the classes would be done by the same simple method. For each extracted class, a disambiguation by a generic resource would be made to determine the rising list of hyper names. If one of these hyper names is also one of the elements aligned with the module, the extracted class is a part of the domain. Otherwise, this class would not be exported to the final document because it is not a part of the domain. In this way, we would obtain a list of potential classes belonging to the domain of study. These classes describe the different existing periods in history of Tunisia, for example, the Roman period, Islamic period, etc.

3.2 Step 2: Validation of Classes

After the previous step, we cannot guarantee the existence of a class belonging to the list. It is possible that one of the classes present in the previous list has not grounds for being. For example, a class extracted from a database would not represent a class which is a part of the ontology but a relation between other classes. This kind of problem is due to the nature of the studied sources, which is different from the ones faced in ontology. To validate these classes, we plan to intersect the treatment of several sources and resources during this step. For example, if a class extracted from a first source is similar to a class extracted from a second source, then the confidence of the existence of this class increases. Moreover, during the previous step, if the class is aligned with an element of a generic resource, its confidence rises.

We define seven main classes for our ontology: (1) Anthroponyme (“Anthroponyms”), (2) Appellation, (3) Peuple (“People”), (4) Entité-spatiale (“spatial (locative) entity”), (5) E2-Entité-Temporelle (“Temporal Entity”), (6) Objet-Physique (“physical object”), (7) E28-Objet-Conceptuel (“conceptual object”).

3.3 Step 3: Extracting of the Hierarchy (subClassOf)

We propose the following hierarchy for the classes defined for our multimedia ontology:

Anthroponyme It includes all the names of persons (gods, heroes, etc.) related to such monument. We have associated to this class the subclasses: "*Héro*" ("hero"), "*Personnage-Réligieuse*" ("religious person"), and "*Divinité*" ("Divinity").

Appellation It is equivalent to the E41-Appellation class from CIDOC-CRM. Along the class "Appellation" we have associated the subclasses "*Identifiant*", "*Nom-Actuel*" ("current name"), and "*Nom-Ancien*" ("old name").

Peuple It designs the names of inhabitants of a city mentioned in a historical document. It includes the subclasses "*Appellation-Peuple*" ("name of the people"), "*Civilization*" ("civilization"), and *Empereur* ("Emperor").

E2-Entité-Temporelle It is a mapped class from CIDOC-CRM ontology. It includes the subclasses: "*E4-Période*" (Period): it includes sets of coherent phenomena or cultural events it also integrates the subclass "*E5-Evénement*" ("Event").The subclass "E5-Evénement" defines the subclass "*E7-Activité*". *Activity* it is defined as the action or the sequence of actions carried out by agents who pursue a determined purpose. In fact, this purpose leads generally to a change of the state of the cultural, social and material systems that interest us. It includes as subclasses: "*E11-Evénement-De-Modification*" ("Event of modification"), "*E9-Déplacement*", "*E63-Début-Existence*" ("Beginning of Existence"), "*E64-Fin-Existence*" ("End of Existence"), and "*E81-*Transformation" ("Transform"). E2-Entité-Temporelle class includes also the subclass "*Tranche-Chronologique*" ("Chronological Tranche") which includes the subclasses: "*Datation-Absolue*" ("absolute dating"), "*Periodisation*" ("periodization"), and "*Siècle*" ("century").

Entité-Spatial It refers to any entity related to the location of a monument or site. We have associated the following subclasses to the "Entité-Spatiale" class: "*E47-Coordonnées-Géographiques*" ("Geographic coordination"), "*Nom-Lieu*" ("Name-Location"), "*Supérficie*" ("Surface"), "*Délimitation*" ("Delimitation"), "*Municipalité*" ("Municipality"), "*Structure*", and "*Occupation-Actuelle*" ("Current-Occupation").

Objet-Physique It is equivalent to the class "E24-Quelques-Choses-De-Matériel-Et-De-Fabrique" from CIDOC-CRM. We have associated to this class the subclass "*Elément-Architectural*" ("architectural element") mapped from the CARE ontology, the subclass "*E57-Matériaux*" ("materials") mapped from the CIDOC-CRM ontology, and the subclass "*Statut-Juridique*" ("Legal-Status") that is equivalent to E72-Objet-Juridique subclasses from the CIDOC-CRM ontology.

E28-Objet-Conceptuel It is a mapped class from CIDOC-CRM. It includes, on the one hand; the subclass *E73-Objet-Informatif*. The documentary object could be a visual element described by the subclass *E36-Elément-Visuel* or, it could be a document described by the subclass *E31-Document*. On another hand, E28-Objet-Conceptuel includes the subclass "*Description-De-Site*" that includes the different subclasses that

aim to describe the site or the monument: "*Catégorie*" ("Category"), "*Données-Générales*" ("General-Data"), "*Patrimoine-Mondial*" ("World-Heritage"), "*Nombre-De-Monuments*" ("Number-Of-Monuments"), "*E3-Etat-De-Conservation*" ("State-Of-Preservation"), and "*Etat-De-Recherche*" ("State-Of-Research").

The following steps are still under study and are not detailed in this paper.

3.4 Step 4: Disambiguation of the Relations

According to our study, the discovery of the existence of a relation between two classes is simple. The real challenge consists in managing to disambiguate these relations. The most efficient methods of extracting relationships are based on a corpus study. The only research studies that dealt with the automatic extraction of relations from a thesaurus [8] also use a corpus. We should therefore consider an automatic processing of the corpus so that we can disambiguate these relationships.

Nevertheless, we intend to validate the disambiguation carried out via an alignment of the relations. The more relations are present in several sources and resources, the more confidence in this relation will be large.

3.5 Step 5: Define the Axioms

In the search of [9], it is specified that it is not advised to explicitly express the hierarchical relations (subClassOf) but rather to use a reasoner to deduce them. Therefore, it is more important to focus on the definition of the axioms present in the ontology than on the hierarchy (subClassOf) between the classes. For this reason, we propose to define some axioms in our module that will enable us to infer the hierarchical relations between the extracted classes.

3.6 Step 6: Define the Instances

Instances are used to represent elements. In addition, creating an instance of such entities, consists of creating a real object from abstract objects, therefore this gives a values to this entities and defines their relations [2].

4 Multimedia Ontology Representation Through MOWL

We have used the Multimedia Web Ontology Language [10] for representing the multimedia ontologies used in our experiments. An ontology encoded in a traditional ontology language, e.g., OWL, uses text to express the domain concepts and the properties. Thus, it is quite straightforward to apply such ontology for semantic text processing. Semantic processing of multimedia data, however, calls for ontology primitives that enable modeling of domain concepts with their observable media properties. This kind of modeling is called Perceptual Modeling, an example of which is shown in Fig. 1a. Such modeling needs to encode the inherent uncertainties associated with media properties of concepts media manifestation.

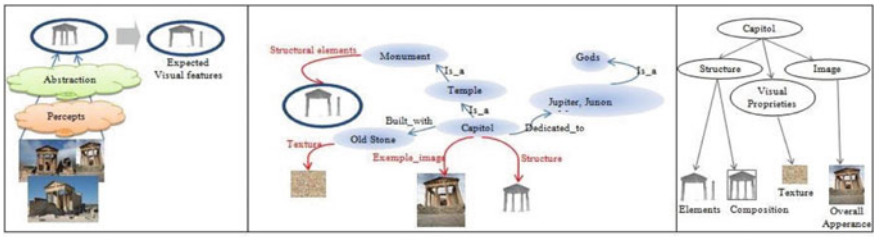


Fig. 1. **a** Perceptual modeling, **b** multimedia ontology of Tunisian monuments, **c** observation model of Capitol

MOWL Relations Relations between the concepts play an important role in concept-recognition. For example, an important clue to the recognition of a medieval monument can be the visual properties of the stone it is built with (as shown in Fig. 1b). In order to enable such reasoning, MOWL allows definition of a class of relations that imply "propagation" of media properties.

Specifying Spatio-temporal Relations Complex media events can be defined in MOWL with constituent media objects and their spatio-temporal relations with formal semantics which is consistent with and can be executed with an extended MPEG-7 Query Engine proposed in [11]. A multimedia ontology should be able to specify such concepts in terms of spatial/temporal relations between the components. MOWL defines the subclass <mowl:ComplexObject> which represents composition of media objects related through spatial or temporal relations. Every complex object is defined by a spatial or temporal relation or predicate and two media objects—one the subject of the predicate relation and the other the object of the predicate.

Uncertainty Specification The relations that associate concepts and media objects are causal relations and are generally uncertain in nature. Thus, these associations are probabilistic in nature. MOWL provides for specification of uncertainty of these associations in a multimedia domain by providing special constructs for defining Conditional Probability Table (CPTs) and associating them with concepts and media objects.

Reasoning with Bayesian Networks The knowledge available in MOWL ontology is used to construct an observation model (OM) for a concept, which is in turn used for concept-recognition. This requires two stages of reasoning:

- Reasoning for derivation of observation model for a concept: This requires exploring the neighborhood of a concept and collating media properties of neighboring concepts, wherever media property propagation is implied.
- Reasoning for concept-recognition: Once an observation model for a concept is created, it can be used for concept-recognition. We use an abdicative reasoning scheme that exploits the causal relations captured in the observation model (Fig. 1c).

5 Conclusion

To conclude, data related to the Tunisian archaeology field are so diversified. Thus, the design of multimedia ontology for the Tunisian cultural heritage domain seems to be the best solution in order to structure these data.

Ontology is not dependent only on the domain but also on the target task. Indeed, it is the context of the task that makes it possible to fix the relevant features of the significance of the semantic concepts so as to cancel the effects of the context.

Through the description of the construction process, we have shown: (1) the necessity of exactly articulating the concepts and the terms within such a resource; (2) the necessity of developing a meta-modeling architecture to model all the necessary sources and the knowledge structures.

The encountered main difficulties are: (1) the definition of the conceptual field of the RTO; (2) the lack of representativeness of the original corpus which obliged us to boost a partially manual conceptualization process.

Our work perspectives are: (1) a complete reorganization of the concepts in order to improve the efficiency of the ontology, and (2) work on the completeness of the terms with a test of the ontology in an application in progress.

References

1. Culture de la Tunisie, Wikipédia: l'encyclopédie libre [en ligne], Dernière modification de la page 22 janvier 2019 à 19:20, [Consulté le 22 janvier 2019]. Disponible à l'adresse: https://fr.wikipedia.org/w/index.php?title=Culture_de_la_Tunisie&oldid=156066915 (2019)
2. Ben Salah, M., Yengui, A., Neji, M.: Construction of OTunAr: ontology of Tunisian archeology. In: The International Arab Conference on Information Technology (ACIT 2017), 22–24 Dec 2017, Tunisia (2017)
3. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. Int. J. Human Comput. Stud. **43**(5–6), 907–928 (1995)
4. Villazón-Terrazas, B.C., Suárez-Figueroa, M., Gómez-Pérez, A.: Pattern-based method for re-engineering non-ontological resources into ontologies. Int. J. Semant. Web Inf. Syst. **6**, 27–63 (2010)
5. Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., Katz, S.: Reengineering thesauri for new applications: the AGROVOC example. J. Digit. Inf. **4**(4). cited by (since 1996) 51 bibtex: soergel_reengineering_2004 (2004)
6. Doerr, M.: The CIDOC CRM, an ontological approach to schema heterogeneity. In: Dagstuhl Seminar Proceedings, Schloss Dagstuhl-Leibniz-Zentrum für Informatik (2005)
7. Granjon, P.L., Leclercq, F., Millereux, A., Savonnet, M., Sapin, C.: Base de Données annotées et Wiki pour la constitution du corpus numérique CARE. Hortus Artium Medievalium **18**(1), 27–35 (2012)
8. Chrisment, C., Haemmerlé, O., Hernandez, N., Mothe, J.: Méthodologie de transformation d'un thesaurus en une ontologie de domaine. Revue d'Intelligence Artificielle **22**(1), 7–37. bibtex: chrisment_methodologie_2008 (2008)
9. Rector, A.L.: Modularisation of domain ontologies implemented in description logics and related formalisms including OWL. In: Proceedings of the 2nd International Conference on Knowledge Capture, pp. 121–128. ACM. bibtex: rector_modularisation_2003 (2003)

10. Mallik, A., Ghosh, H., Chaudhury, S., Harit, G.: MOWL: an ontology representation language for web-based multimedia applications. ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) **10**(1), 8 (2013)
11. Wattamwar, S.S., Ghosh, H.: Spatio-temporal query for multimedia databases. In: Proceeding of the 2nd ACM Workshop on Multimedia Semantics (MS'08), pp. 48–55. ACM, New York (2008). https://doi.org/10.1145/1460676.1460686

Ontology-Based Information Extraction Framework for Academic Knowledge Repository

Veena Jose[1], V. P. Jagathy Raj[2(✉)], and Shine K. George[1]

[1] Master of Computer Applications Programme, Union Christian College, Aluva, Kerala, India
veenajose@mcaucc.edu.in, shineucc@gmail.com

[2] School of Management Studies, Cochin University of Science and Technology, Cochin, Kerala, India
jagathyraj@gmail.com

Abstract. Extracting required information from the huge academic repository in an efficient way is a tedious job as the repository grows day by day. Getting user satisfaction in searching is the primary concern always. Academic search engines are mostly used by research scholars, scientists, faculty, and students for their academic purposes. The use of ontology in information extraction helps an academic search engine to perform its operation effectively. As the research activities are going on in an extensive pace in all domains, new specialized areas are evolved. This paper proposes a novel framework that incorporates a dynamically evolving ontology that focuses on expanding the existing Computer Science Ontology based on the new findings in existing areas. The Word2Vec model helps to identify new keywords for the expansion of the ontology. The keywords given by the Word2Vec should satisfy a confidence score above a threshold value and only these keywords will be used for updating the Computer Science Ontology. The dynamic semantic web helps a user to retrieve the specific document, and thus provides a high degree of personalization experience.

Keywords: Ontology · Information extraction · Semantic web · Word2vec · Confidence score

1 Introduction

Extracting correct information from the internet is found to be difficult nowadays, as the knowledge repository is growing instantaneously. Retrieval of information is efficient if storage is perfect. Consider the case of adding new journals to the existing academic repository.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_6

As far as an academic repository is concerned, the development is fast and specializations are more. The context can be considered for a particular branch—Computer Science. It is a major research area, and hence a large number of journals are added often. Searching for a particular journal from this repository is very difficult. Normally, in searching, a list of possible links is given back as the result of the user search query. This result may not be exact. This paper proposes a novel framework which retrieves the specific journal from the academic repository with the help of ontology by adding dynamism. The paper [1] proposes that the existing Computer Science Ontology identifies 26 k semantic topics and 226 relationships. New specializations are getting introduced day by day in the area—Computer Science, and hence the semantic relationship between the keywords is to be updated dynamically. These updations are to be reflected in the archive also, that only helps to retrieve more accurate results in searching.

In Fig. 1, SubDomain 1, when introduced under the domain, the semantic relationship between domain and SubDomain1 is to be identified between them which is essential in journal storage and retrieval.

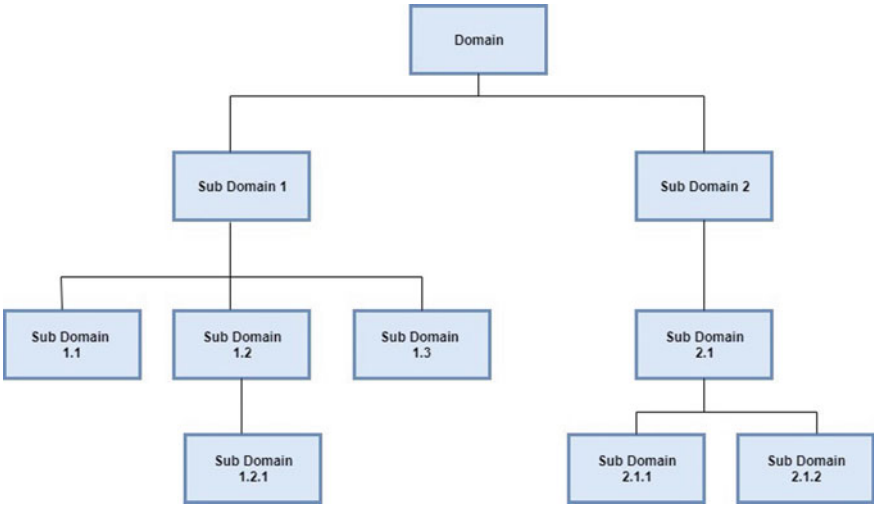


Fig. 1. Representation of progress in domains in the academic repository

To maintain the hierarchy of newly added content or metadata to the existing one, ontology plays a big role. All users demand personalization, for which the ontology is to be expanded dynamically.

The framework proposed by this paper focuses on enhancing the personalization in academic search by adding Word2Vec and Confidence Score based on Word2Vec model.

The outline of the paper is as follows: Sect. 2 identifies the related work in the area and the proposed dynamic framework is discussed in Sect. 3. The processes in the framework are listed in Sect. 3.1. Section 3.1.1 lists the steps to be carried out in the storage and Sect. 3.1.2 lists the steps for retrieval in the framework. Section 4 projects the experimental analysis and Sect. 5 concludes the concept.

2 Related Work

Ontology-based semantic web information retrieval using the Jena framework [2] is suggested as a superior method to keyword-based search in the paper. Here, the search engine interprets the meaning of the query submitted by users and also the relationship among the concepts specified in the document and the domain. The paper concludes that instead of listing the possible links, the exact information the user searches for is retrieved.

Searching documents based on Word Sequence Graph (WSG) is discussed in paper [3] where it proposes a fast and efficient lengthy content search by converting an unstructured document into a structured format. Sentences in the document are represented as a graph model. The documents are represented using WSG on the assumption that documents are parsed into a set of sentences and in turn into a set of words. The directed graph consists of a list of nodes that are connected by edges. The paper points out that the advantage is obtained by extending the graph model by the incremental approach.

Semantic Information extraction in the University domain [4] is purely confined to the University where semantic information extraction done using ontology is proposed by the author.

Searching and extracting specific knowledge directly from unstructured documents using ontology [5] is explained in the paper. Here, the author proposes a knowledge extraction tool linked with ontology to obtain continuous knowledge support which helps to a great extent in retrieving the information.

From the various resources, the semantics of information is retrieved and that can be used to define the ontology for the library system [6] is discussed in the paper. Existing ontologies are reused and these ontologies are adapted to certain domain-specific tasks.

Deep learning enhances text processing, analysis, and classification [7] is explained by the author of the paper. The success of this is mainly due to the quality of word representation. The quality of word representation is enhanced using the Word2Vec approach. Feature selection removes redundant and irrelevant features and reduces the dimension of features [8] is the theme of the paper. A decrease in feature dimension improves the performance of the classifier in text categorization. Word2Vec model identifies the semantic meaning through vector representation of words. Considering the semantic meaning of words instead of counting the word frequency, an approach followed by the Word2Vec model [9] is discussed here. The weight of the words is calculated; the semantic relationship is identified and the final weight of vectors is also formed in the model. Semantic similarity computation helps to get the word set that belongs to the target domain. Experiments show that domain clustering based on semantic similarity using Word2Vec achieves higher accuracy [10].

Achieving dynamic ontology through IoT [11] is the thread of the paper. To realize the capability of connected cars, the dynamic knowledge of the world is to be acquired by vehicles. The power of semantics is enhanced through the ontology for accessors and services provided by the vehicles are identified by the author.

In the scholarly domain, ontology supports the integration of research data, information extraction from articles, exploring the research possibilities, etc. [12] is discussed in the paper. The paper suggests that Computer Science Ontology (CSO) is a large-scale and automatically generated ontology based on the research areas considered up

to a particular period. The CSO performs an automatic ontology generation of research areas using the Klink-2 algorithm. A list of relevant concepts or keywords is identified by mapping n-grams from texts to concepts in the CSO. The CSO classifier removes English stop words and gathers together unigrams, bigrams, and trigrams. Then, for each n-gram, the similarity with the labels of the topics in CSO is computed. If the similarity of the research topics is greater than or equal to a minimum threshold value with an n-gram, they are added to a final set of topics is proposed in the paper.

Personalized information extraction is possible by identifying semantically matching information through ontology [1] is the finding in the paper. The effectiveness of information extraction solely depends upon the capability of the machine to understand the semantics of the search query is discussed by the author.

3 Proposed Dynamic Ontology Framework

The web search can be made efficient by dynamically extending the ontology. To achieve this dynamism, Word2Vec is suggested by the framework.

The framework initially identifies the candidate word set or keywords by removing stop words from the title and abstract of the new journal submitted to the archive. These identified keywords and keywords from the keyword section of the journal are checked with the Computer Science Ontology (CSO). The CSO cannot recognize all the keywords as it cannot include all specializations or new subdomains. The keywords not recognized by CSO are forwarded to Word2Vec. Word2Vec generates vector representation of words that carry semantic meaning. The numerical value of these vector representations is the confidence score of the words. Only the words with a confidence score greater than 75% will be taken in the framework which indicates the proximity of these keywords in semantic meaning with the keywords already identified by the Computer Science Ontology. The new keywords identified through Word2Vec are updated to the CSO ontology. The journal, title of the journal, and the set of keywords are stored in the repository.

The retrieval of information will be efficient as the storage follows the above method. When searching for a journal in the repository, the candidate words are extracted from the given query and it is checked with the Computer Science Ontology which helps to retrieve the exact information than the list of possible documents. Thus, this approach provides a fast and efficient document retrieval in academic search compared to the existing methods.

3.1 Framework

The processes in the framework can be categorized into two parts: storage and retrieval.

3.1.1 The Storage Part of the Framework

1. The input to the storage part of the framework is the journal.
2. Fetch the keywords, title of the journal, and abstract of the journal.
3. Identify and extract the keywords from the title and abstract of the journal.

4. The fetched keywords and extracted keywords are to be checked with the existing Computer Science Ontology.
5. Keywords not recognized by the Computer Science Ontology are to be given to the Word2Vec model which identifies the semantic meaning of the keywords through the vector representation.
6. The confidence score of these identified keywords is compared with a specific threshold value. The threshold value suggested in this paper is 75%.
7. The keywords with a confidence score greater than the threshold value are forwarded for updating the Computer Science Ontology.
8. The journal, keywords, and title of the journal are stored in the repository.

3.1.2 Steps to Be Followed—Retrieval Part of the Framework

9. The input to the retrieval part of the framework is the search query.
10. Extract keywords by removing stop words.
11. Check the CSO for semantic relationship.
12. Retrieve the matching journal from the repository.

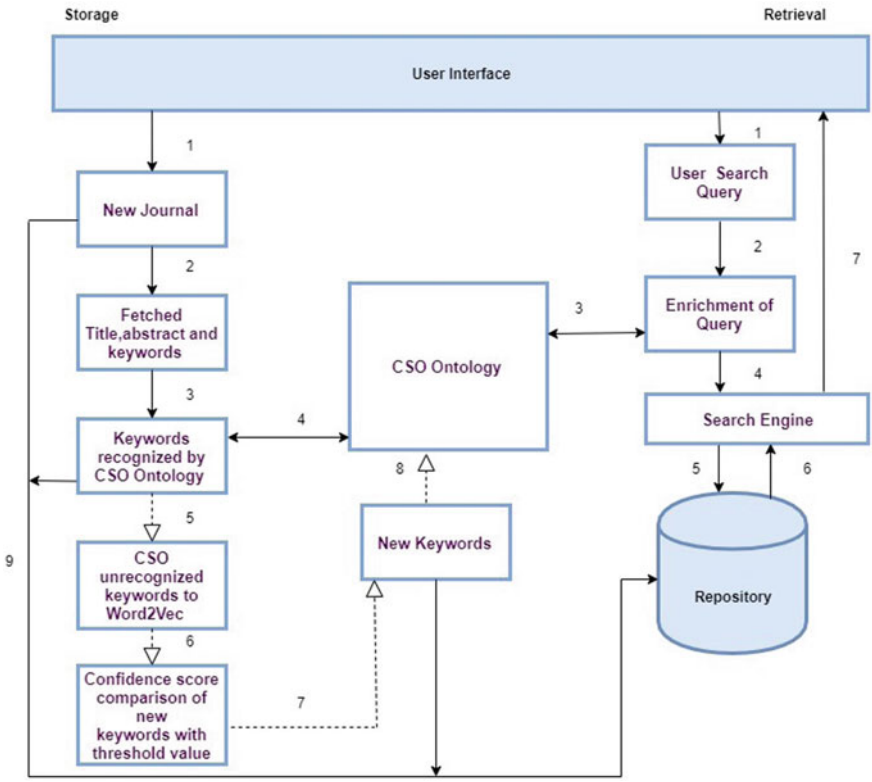


Fig. 2. Conceptual framework for dynamic ontology-based information storage and extraction

The dotted line contributes to the framework for achieving dynamic ontology.

3.2 Word2Vec and Confidence Score

The framework uses the flavor-Continuous Skip-gram of the Word2Vec approach suggested by [13]. As the new keywords are selected by ensuring the threshold value of the Confidence Score based on Word2Vec, the semantic relationship between the existing domains and newly added subdomains is promising.

4 Experimental Analysis

Extracting the exact journal from the academic repository for academic and research purposes is the concept behind the framework introduced here.

The Latent Semantic Analysis (LSA) algorithm identifies the semantic similarity between the search query and the extracted information by combining the Vector Space model with Singular Value Decomposition [12].

Table 1 represents the semantic similarity between the concepts given in the document and the search query. The first column gives the semantic similarity in Computer Science Ontology and the second one gives the semantic similarity in Dynamic Computer Science Ontology.

Table 1. Comparison of semantic similarity for sample queries

Sl. No.	Query	Average semantic similarity in percentage	
		Using CSO ontology	Using dynamic CSO ontology
1	Machine learning and python	75.2	97.1
2	Natural language processing and Word2Vec	72.8	97
3	Ontology-based information retrieval system for academic library	77	97.7
4	Ontology and word2vec	71	97.5
5	Dynamic updation of semantic relationship	78	98
6	Stop word removal and keyword extraction	77.9	97.9

The values given in tabular format in Table 1 are represented using a line graph in Fig. 3. In the graph, the blue line indicates the percentage of semantic similarity using Computer Science Ontology and the yellow line shows the percentage of semantic similarity between the search query and concepts given in the documents using Dynamic Computer Science Ontology.

In the above graph, the horizontal axis represents the search query given in tabular format (Table 1). The vertical axis represents the percentage of average semantic similarity between the search query and the concepts specified in the journals available in the repository.

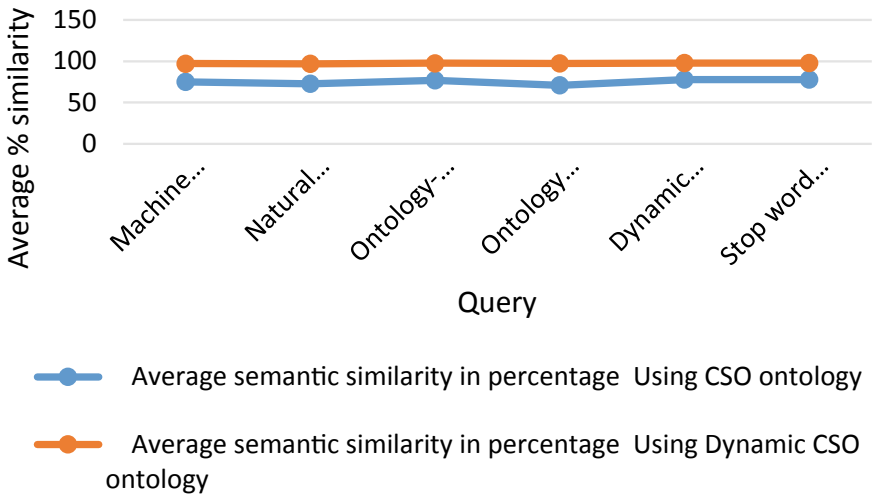


Fig. 3. Semantic similarity evaluation

5 Conclusion

Getting a personalization experience for retrieval of information in academic search is highly appreciable. The effective information retrieval is possible through effective storage mechanism. The new subdomains added to the existing domains always places a major challenge in maintaining the semantic relationship between them. Extending the ontology in a dynamic aspect is the topic of discussion in this paper where the extensibility is achieved through Word2Vec. This development is not to be restricted in academic search but can be used in all fields where searching expects personalization like medical fields, agriculture, etc.

Achieving a high degree of completeness is still found to be a challenge. Other methods like deep neural networks can also be tried for dynamically extending the ontology but it requires huge computational power. In the mere future, there may be alternatives for this limitation also.

References

1. George, S.K., Jagathy Raj V.P., Santhosh Kumar, G.: Personalized news media extraction and archival framework with news ordering and localization. In: Information and Communication Technology for Sustainable Development, pp. 463–471. https://doi.org/10.1007/978-981-13-7166-0_46
2. Jamgade, A.N., Karale, S.J.: Ontology based information retrieval system for Academic Library. In: International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). IEEE (2015). https://doi.org/10.1109/iciiecs.2015.7193106
3. George, S., Sudheep Elayidom, M., Santhanakrishnan, T.: A novel sequence graph representation for searching and retrieving sequences of long text in the domain of information retrieval. Int. J. Sci. Res. Comput. Sci., Eng. Inf. Technol. **2**(5) (2017), www.ijsrcseit.com. Journal No: 64718

4. Rajasurya, S., Muralidharan, T., Devi, S., Swamynathan, S.: Semantic information retrieval using ontology in university domain. In: https://www.researchgate.net/publication/229556823 (2012)
5. Alani, H., Kim, S., Millard, D.E., Weal, M.J., Hall, W., Lewis, P.H., Shadbolt, N.R.: Automatic ontology-based knowledge extraction from web documents. IEEE Intell. Syst. **18**(1), (2003). https://doi.org/10.1109/mis.2003.1179189
6. Ying Ding, Dieter Fensel: Ying Ding & Dieter Fensel: Ontology Library Systems. In: https://www.researchgate.net/profile/Ying_Ding/publication/221466741_Ontology_Library_Systems_The_key_to_successful_Ontology_Reuse/links/0deec52a8c9d27995b000000.pdf
7. Tezgider, M., Yıldız, B., Aydın, G.: Improving word representation by tuning Word2Vec parameters with deep learning model. In: International Conference on Artificial Intelligence and Data Processing (IDAP), (2018). https://doi.org/10.1109/idap.2018.8620919
8. Tian, W., Li, J., Li, H.: A method of feature selection based on Word2Vec in text categorization. In: 37th Chinese Control Conference (CCC), (2018). https://doi.org/10.23919/chicc.2018.8483345
9. Ogul, I.U., Ozcan, C., Hakdagli, O.: Keyword extraction based on word synonyms using WORD2VEC. In: 27th Signal Processing and Communications Applications Conference (SIU), (2019). https://doi.org/10.1109/siu.2019.8806496
10. Luo, J., Wang, Q., Li, Y.: Word clustering based on word2vec and semantic similarity. In: Proceedings of the 33rd Chinese Control Conference (2014). https://doi.org/10.1109/chicc.2014.6896677
11. Liu, M., Jin, C.-B., Yang, B., Cui, X., Kim, H.: Online multiple object tracking using confidence score-based appearance model learning and hierarchical data association. IET Comput. Vis. **13**(3), 312–318 (2019). https://doi.org/10.1049/iet-cvi.2018.5499
12. Salatino, A.A., Thanapalasingam, T., Mannocci, A., Osborne, F., Motta, E.: Classifying research papers with the computer science ontology. In: International Semantic Web Conference (ISWC), California, USA. (2018). https://doi.org/10.5281/zenodo.2660820
13. https://towardsdatascience.com/an-implementation-guide-to-word2vec-using-numpy-and-google-sheets-13445eebd281

PLS Approach to Measure the Impact of Digital Transformation on User Satisfaction: Case of Local Public Services in the Northern Region of Morocco

Fadoua Laghzaoui, Najoua El Abbas El Ghaleb(✉), and Mohcine Ben Sabih El Imrany

Laboratory of Territorial Governance and Sustainable Development (GT2D), University Abdelmalek Essâdi, Tangier, Morocco
laghzaouifadoua@yahoo.fr, elababsnajoua@gmail.com, mohcinebsei@gmail.com

Abstract. This article means to propose a theoretical framework model to quantify the effect of digital transformation on the satisfaction of users of public administrations. In this way, we proceed with a hypothetic-deductive reasoning. A survey was carried out among 110 users of the public services, and the evaluation of our model was carried out using the partial least squares (PLS) structural equations approach via the software SPSS 21 and XL-stat (version 2018) for exploratory factor analysis and verification of reliability and validity. The results of our study show that digital transformation of the public administration positively affects the satisfaction of its users.

Keywords: Digital transformation · Users · Satisfaction · PLS

1 Introduction

With the emergence of digitalization, public administrations are invited more than ever to strengthen their resources, in particular their own skills to serve their stakeholders citizens and other government agencies [1].

Moreover, the Moroccan cultural context is considered as a strong and complex cultural context [2], and the success of such initiatives depends not only on the support of the government, but also on the willingness of citizens to adopt and use these services [3].

The objective of this paper is to study how the digital transformation impacts the satisfaction of the users without taking into account only the functional determinants.

The first section of this work will present our conceptual model with particular domain of our constructs; the second section presents the results confirming the validity and reliability of our empirical study to finish with the interpretation of the results and the proposition of perspectives' research.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_7

2 The Influence of Digital Transformation on Users Satisfaction: A Conceptualization

In order to evaluate and measure the influence of digital transformation on user satisfaction in the Moroccan context, our study was based on some numbers of theoretical hypotheses that the field analysis will try to test.

2.1 The «Digital Transformation» Construct

The digital transformation has become a flagship concept, very fashionable, which knows a craze on the part of the practitioners and an interest more and more crossing on the part of the researchers. Based on various studies examined, among the components of the digital transformation in the public administration, we find the behavioral choices «willingness» [4] «adaptation» [5] and «ease of use» [6], Fig. 1.

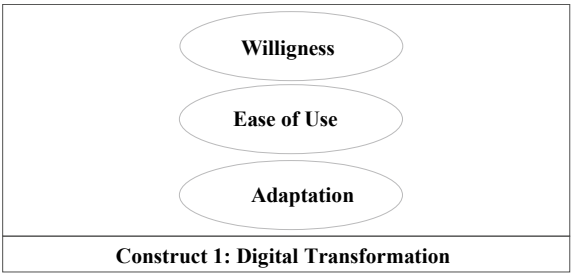


Fig. 1. Components of digital transformation

Willingness: In the context of the administrative services, the willingness is the intention of the users to accept the digital transformation being that element which influences their satisfaction.

Adaptation: It assumes that the process of judging a new object is based on judgments already made about ancient objects. Indeed, the adaptation theory models the various emotional traces resulting from past experiences likely to disturb the balance of the central nervous system.

The ease of use: Defined as «the degree to which an individual accepts that an authoritative and specialized foundation exists to help their utilization of a framework» [7]. In this study, enabling conditions were measured according to the view of having the option to get to the necessary assets.

2.2 The «User Satisfaction» Construct

User satisfaction has become an indicator of public administration performance. This is one of the explicit objectives of public sector reform programs [8].

Among the variables that influence user satisfaction in the context of Moroccan public administration, we find three typologies «**perceived confidence**» [9], «**perceived utility**» [10] and «**perceived pleasure**» [11], Fig. 2.

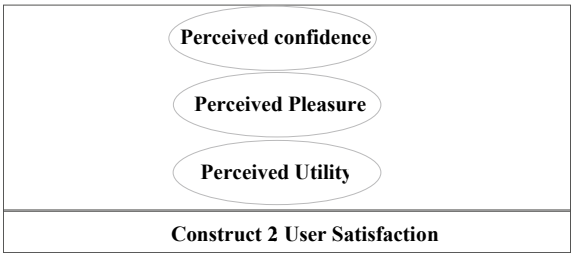


Fig. 2. Components of user satisfaction

Perceived confidence: Several studies confirm the importance of the trust factor because it helps to understand the behavior of citizens in relation to their uses of electronic services [12]. These criteria differ between cognition and affect [13].

Perceived utility: In this study, perceived utility is measured according to perceptions of the user's use of services in terms of benefits, such as time savings, communication with the administration.

Perceived pleasure: It measures the degree of perceived agreeableness when using digital services. Indeed, the satisfaction itself corresponds to an emotional state made of pleasure and relief [14].

3 The Research Model and the Hypotheses

Analysis of both concepts: Digital transformation and user satisfaction led us to the development of the conceptual model below, Fig. 3.

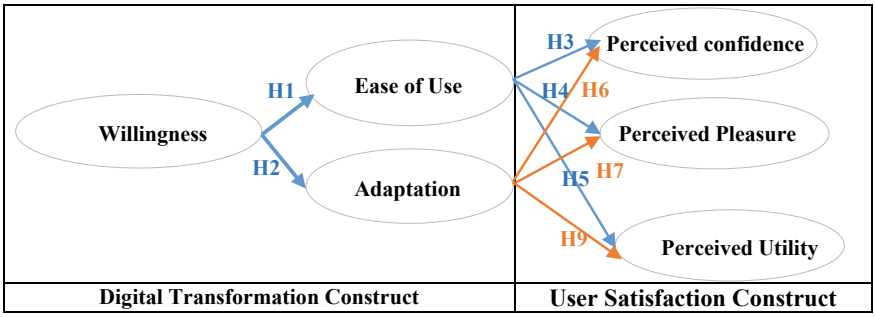


Fig. 3. Conceptual model of research the impact of digital transformation on user satisfaction

Following our literature review, we identified **nine research hypotheses** (**H1, H2, H3, H4, H5, H6, H7, H8, and H9**) generated from a **global hypothesis** that supposes digital transformation of public administration positively affects user aatisfaction.

4 Empirical Studies and Results

4.1 Sampling

According to Churchill's (1979) paradigm [15], our empirical study is carried out using a survey composed of items collected on a fifth-degree Likert scale ranging from «strongly disagree» to «totally agree administered to 110 end users in public administrations of the studied region.

4.2 Reliability of Measurements

The reliability of the measurement scale is verified by two measures: Cronbach Alpha and Rhô of D.G. According to the majority of researchers [16], when the indicator is greater than or equal to 0.6, there is a good level of reliability.

The reliability indices of the factors selected in our study are greater than 0.6 (Table 1), Cronbach's alpha is satisfactory [17], so the items are correlated and consistent with each other, and they can be summed to form a scale score, Table 1.

Table 1. Reliability of measures

Latent variables	Items	Cronbach alpha	Rho of D.G.
WI	2	0.6	0.831
EU	2	0.697	0.868
AD	4	0.772	0.862
PC	4	0.794	0.868
PU	5	0.851	0.894
PP	4	0.762	0.856

For each block, the largest eigenvalue must represent at least 50% of the sum of all the values belonging to this block [18]. This is the case for our results found in Table 2 which confirm the **unidimensionality of the blocks**.

Table 2. Eigenvalues of the latent variables of the causal model

WI	EU	AD	PC	PU	PP
1.423	1.535	2.531	2.505	3.149	2.497
0.577	0.465	0.931	0.735	0.616	0.948
		0.371	0.479	0.51	0.316
		0.166	0.282	0.446	0.239
				0.279	

4.3 Evaluation of the Measurement Model

To analyze our results, we used modeling by structural equations according to the partial least square (PLS) approach which is especially when the examination is exploratory [19]. In our case, our variables proved reflexive [20].

The relation between the latent variable and all the manifest variables associated with it is written:

$$\mathbf{x_{kj}} = \boldsymbol{\pi}_{\mathbf{kj}} * \boldsymbol{\xi}_{\mathbf{k}} + \mathbf{kj}$$

with:

$\mathbf{x_{kj}}$ Vector associated with the Jth manifest variable of the latent variable ξ_k
$\boldsymbol{\xi}$ Latent variable
$\mathbf{k}$ Index of latent variables
$\mathbf{kj}$ Index of the manifest variables of the block k
$\boldsymbol{\pi}$ Loading associated with x_{kj}; $\mathbf{kj}$: Error term

4.4 Convergent Validity

This validity represents the average of the variances between the construct and its items, and it must be greater than 0.5. The convergent validity of the measurement model was evaluated by examining its mean extracted variance (AVE) value [17].

Table 3 shows that all variables have AVE values that exceed the recommended threshold of 0.5 [21]. This result shows that the measurement model of the study has demonstrated adequate convergent validity; D.G's Rhô is also >0.7.

Table 3. Eigenvalues of the latent variables of the causal model

Latent variable	Average variance extracted (AVE)	Rho of D.G.
WI	0.69	0.831
EU	0.767	0.868
AD	0.631	0.862
PC	0.589	0.868
PU	0.629	0.894
PP	0.623	0.856

4.5 The Divergent Validity

Since each latent variable has more saturation with its measurement variables than with the measurement variables of the other constructs, the discriminating validity of our search model is ensure (Table 4).

Convergent and divergent validity confirm the validity of our **measurement model**.

Table 4. Divergent validity

	WI	EU	AD	PC	PU	PP	AVE
WI	*0.830						0.690
EU	0.506	*0.875					0.767
AD	0.373	0.467	*0.794				0.631
PC	0.126	0.196	0.461	*0.767			0.589
PU	0.369	0.422	0.542	0.356	*0.793		0.629
PP	0.280	0.309	0.549	0.434	0.410	*0.789	0.623

*Square root of the AVE

4.6 Goodness of Fit Index (GoF)

According to the results obtained (Table 5), our model expresses a good level of adjustment of the global model [22].

Table 5. Adjustment indices

	GoF
Absolute	0.563

4.7 Coefficient of Determination R^2

See Table 6.

Table 6. Results R^2 and R^2-adjusted

Latent variable	Type	R^2	R^2 adjusted
WI	Exogenous	–	–
EU	Endogenous	0.506	0.506
AD	Endogenous	0.373	0.373
PC	Endogenous	0.462	0.457
PU	Endogenous	0.582	0.578
PP	Endogenous	0.554	0.55

The coefficient R^2 is greater than 0.1 [23], we can conclude that our model has a strong predictive validity.

Our empirical results clearly show the validity of **the (external) measurement model of the (internal) structural model**, which allows us to proceed to validating the hypotheses derived from our model.

4.8 Structural Equations of the Conceptual Framework

The generation of the five structural equations representing the eight assumptions derived from our conceptual model was realized using SPSS 21 and Xl-stat software (version 2018) as follows:

$$\mathbf{EU} = \mathbf{0.711} * \mathbf{WI}; \quad \mathbf{AD} = \mathbf{0.610} * \mathbf{WI}$$
$$\mathbf{PC} = -\mathbf{0.0411} * \mathbf{EU} + \mathbf{0.707} * \mathbf{AD};$$
$$\mathbf{UP} = \mathbf{0.275} * \mathbf{EU} + \mathbf{0.547} * \mathbf{AD};$$
$$\mathbf{PP} = \mathbf{0.092} * \mathbf{EU} + \mathbf{0.678} * \mathbf{AD}$$

4.9 Assumptions Tests

Our global hypothesis assumes that **the digital transformation** of the public administration positively impacts **the satisfaction of its users**. This assumption generates **eight derived hypotheses** according to our proposed model that will be subject to confirmatory testing as follows (Table 7):

Table 7. Research hypothesis tests

Hypothetical relationship	Path coefficient (β)	Student T*	Effect size	Validation	Degree of association
H1:WI → EU	0.711	10.519	1.024	Yes	Strong
H2:WI → AD	0.61	8.012	0.594	Yes	Strong
H3:EU → PC	**−0.041**	**−0.423**	**0.002**	No	Invalid
H4.EU → PU	0.275	3.221	0.097	Yes	Lower
H5:EU → PP	**0.092**	**1.041**	**0.01**	No	Invalid
H6:AD → PC	0.707	7.281	0.495	Yes	Strong
H7:AD → PU	0.547	6.394	0.382	Yes	Strong
H8:AD → PP	0.678	7.664	0.549	Yes	Strong

*Student's t-test values are greater than |2.775|(|1.960|) indicate significant parameters u threshold of 1% (5%)

The effect size (0.02: low, 0.15 medium, 0.35 wide) from Ref. [24] ensures the validity and the magnitude of the structural coefficients.

4.10 Indirect Effects

The indirect links in Table 8 confirm the existence of the strong relations between the exogenous variable '**willingness**' and all the variables of the construct '**user satisfaction**' **including** 'confidence,' '**utility**' **and** '**pleasure.**'

Table 8. Indirect effects between latent variables

	WI	EU	AD	PC	PU	PP
PC	**0.403**	0.000	0.000			
PU	**0.530**	0.000	0.000	0.000		
PP	**0.480**	0.000	0.000	0.000	0.000	

4.11 Results Discussions

Based on the results obtained, we conclude the following:

The exogenous latent variable '**willingness**' of the construct '**digital Transformation**' has positive and statistically significant influences ($T > 1.96$) and has strong effects on the latent variables 'ease of use' and 'adaptation' two endogenous variable belonging to the same construct.

The latent variable '**adaptation**' has positive and statistically significant influences and has significant effects on latent variables '**perceived confidence,**' '**perceived utility**' and '**perceived pleasure**' three component of the **user satisfaction** construct.

The latent variable 'ease of use' has **insignificant** influences on '**perceived confidence**' and '**perceived pleasure,**' which proves the invalidity of these two links.

5 Conclusion and Perspectives

The empirical study conducted with 110 users of public services carried out in the northern region of Morocco generated six valid hypotheses derived out of eight confirming that the digital transformation impacts positively the satisfaction of public service users.

Nevertheless, the existence of some invalid links between «**user facility**» and «**perceived confidence**» are justified by the fact that perceived trust is supposed to be impacted by culture and not by functional conditions [17]. Similarly, the existence of an invalid link between «**ease of use**» and «**perceived pleasure**» can be justified to the fact that the national culture evolves and one attends a standardization of the behaviors toward the search for a perceived pleasure.

Our point of view is to loosen up the example size past 280, so as to re-test our model utilizing the two PLS and LISREL basic demonstrating techniques and think about the outcome got. It will also be possible to add other constructs to our model to be able to treat several aspects of the digital transformation of public administrations in Morocco.

References

1. Teo, T., Srivastava, S., Jiang, L.: Trust and electronic government success: an empirical study. J. Manag. Inf. Syst. **25**, 99–131 (2008)
2. Brown, K.L.: people of salé: traditions and change in a Moroccan city 1830–1930, p. 265. Harvard University Press, Cambridge, Mass (1976)
3. Carter, L., Belanger, F.: A citizen adoption of electronic government initiatives. Hawaii International Conference on System Sciences (2004)
4. Atarodi, S., Berardi, A.M., Toniolo, A.M.: Le modèle d'acceptation des technologies depuis 1986:30 ans de développement. Psychologie du travail et des organisations, Elsevier Masson, pp. 18 (2018)
5. Oliver, R.L.: Measurement and evaluation of satisfaction processes in retail setting. J. Retail. **3**, 25–48 (1981)
6. Colesca, S.E., Dobrica, L.: Adoption and use of e-government services: the case of Romania. J. Appl. Res. Technol. **6**(3), 204–216 (2008)
7. Nisson, C., Earl, A.: The theories of reasoned action and planned behavior: examining the reasoned action approach to prediction and change of health behaviors. University of Michigan (2015)
8. Mayers, R., Lacey, R.: Satisfaction du consommateur, performance et responsabilité au sein du secteur public. Revue Internationale des Sciences Administratives **3**, 395–419 (1996)
9. Derbaix, C., et P.: Pham. «Pour un développement des mesures de l'affectif en marketing: synthèse et prérequis». Recherche et applications en Marketing, pp. 71–87 1989
10. Cadotte, E.R., Woodruff, R.B., Jenkins, R.L.: Expectations and norms in models of consumer satisfaction. J. Market. Res. **24**, 305 (1987)
11. Teo, T.S., Srivastava, S.C., Li, J.: Trust and electronic government success: an empirical study. J. Manag. Inf. Syst. **25**(3), 99–131 (2008)
12. Atarodi, S., Berardi, A.M., Toniolo, A.M.: Le modèle d'acceptation des technologies depuis 1986:30 ans de développement» Psychologie du travail et des organisations, Elsevier Masson (2018)
13. Gefen, D., Straub, D.W., Boudreau, M.C.: Trust and TAM in online shopping: an integral model. MIS Q. **27**, 51–90 (2003)
14. Venkatesch, V., Morris, M., Davis, G., Davis, G., Davis, F.: User acceptance of information technology: toward a unified view. MIS Q. **27**, 425–478 (2003)
15. Churchill, G.A.: Marketing research: methodological foundations, 6th edn. South-Western/Thomson Learning, Mason, Ohio (1995)
16. Hulland, J.: Use of partial least squares (PLS) in strategic management research: a review of four recent studies. Strateg. Manag. J. **20**(2), 195–204 (1999)
17. Sahaf, K., Ali, B.M., Rifai, S., Bouksour, O., Adri, A.: 'Improvement of the hospital supply chain and its impact on reduction of patient waiting times. Case of the Oncology department of University Hospital IBN ROCHD. In: 5th International Conference on Control, Decision and Information Technologies (CoDIT), Thessaloniki, pp. 222–227 (2018)
18. Loewenthal, K.M.: An introduction to psychological tests and scales, 2nd edn. Psychology Press, Hove, UK (2004)
19. Gefenet, D., Straub, D.W., Boudreau, M.-C.: Structural equation modelling and regression: guidelines for research practice. Commun. Assoc. Inf. Syst. **4**(7), 1–79 (2000)
20. Ben Ali, M., Rifai, S., Bouksour, O., Barrijal, S.: Young manufacturing enterprises in the growth phase located in the Wilaya of Tetouan, Morocco: a model for measurement of the impact of quality on their growth. Asian J. Manag. Res. **6**(1), 208–222 (2015)
21. Nunnally, J.C., Bernstein, I.H.: Psychometric Theory, 3rd edn. McGraw-Hill, New York (1994)

22. Chin, W.W.: How to write up and report PLS analyses. Handbook of Partial Least Squares, pp. 655–690. Springer, Berlin, Heidelberg (2010)
23. Wetzels, M., Odekerken-Schroder, G., Vanopen, C.: Using PLS path modelling for assessing hierarchical construct models: guidelines and empirical illustration. MIS Q. **33**(1), 177–195 (2009)
24. Falk, R.F., Miller, N.B.: A Primer for Soft Modelling. The University of Akron Press, Akron, Ohio (1992)
25. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences (2e édition). Lawrence Erlbaum Associates, Hillsdale, New Jersey (1988)

SB-Spectral Clustering Algorithm for Clustering and Validating Sensor

Abdo M. Almajidi(✉) and V. P. Pawar

School of Computational Science, SRTM University, Nanded 431606, India
abdomtn@gmail.com, vrushvijay@yahoo.co.in

Abstract. Due to the use of WSN technologies these days, there have been many developments in WSN. And many types of researches have been proposed for the sensor nodes validation in the spots. The efficiency of the WSN is strongly relying on the lifetime of the network. Validation is the most common approach used in WSN to increase the lifetime, the efficiency and the performance of the wireless network. In this paper, a validation and clustering mechanisms are presented by using SB-SPC algorithm for specified the bad sensor nodes in the applications and excluding them from the application to prolong the functioning of WSN. SPC algorithm determines the most accurate not working sensors with the help of the eigenvalues and eigenvectors by using receive signal strength (RSS). The result of the SB-SPC algorithm shows that this approach specifies the almost not working sensors properly, deleting them from the domain, and achieve good performance in the wireless network.

Keywords: Spectral clustering (SPC) · Distributed wireless network (DWN) · Validation · Bad sensor nodes · Target sensor (TS)

1 Introduction

Wireless sensor networks are new technology that is very much used in our daily life with a large number of civilian and military applications. WSNs consist of a group of connected sensor nodes spatially scattered in the area of interest for monitoring physical or environmental conditions. These sensor nodes, sensing the information from surrounding area, process it using aggregating schemes and transmit it to an external a sink node through the network [1]. This sink node is similar to a gateway in a conventional network. WSNs are predicated to be used in some real-world range applications, such as monitor environment, tunnel, home care, bridges, weather monitoring.

The key negatives in WSNs are the energy use, bandwidth and CPU power, where inter-nodes conserve more power in the communication. To save energy and prolong the functional lifetime of the entire network, inter-nodes communication should be reduced between nodes or sleeping mode should be active while

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_8

the sensor nodes are not needed [2]. The procedure of organizing of the sensor nodes into a group of disjoint sets is the clustering in WSN; this technique can be employed to effectively handle network energy, do data aggregation method and sensor nodes validation in the network. Data aggregation and validation sensors are essentially employed in the clustering WSN, where data aggregation is used to decrease the number of transmitted to sink node, and validation is used to check the validity of each sensor nodes in the domain without using any technical teams.

Clustering techniques can impact in extending the lifetime of the whole network, but not all clustering schemes can equally impact [2,3]. Validation and grouping sensor nodes into clusters have been widely studied by the researcher to achieve a network stability objective. Every cluster may have one target sensor (TS) or more. The objective of TS is to check the validation of the sensor nodes in the wireless network [4–6]. The conventional spectral clustering algorithm is likely the most common technique which has an intense link with graph theory. Spectral clustering is one of the most popular methods of clustering algorithms, with wide applications in distributed computing systems; spectral clustering is used to estimate the wireless transmission cost in wireless networks. In this paper, we present two scenarios based on SPC method to validate the sensor nodes in the spots without attending any technical person; this method takes the advantage of using eigenvalues and eigenvectors for specifying the bad nodes in the wireless network; those nodes can be eliminated by replacing them in the application. Sensor nodes are grouped into smaller groups based on the location, where every group has one reference target (RT) for conducting the validity of the sensors.

The rest of the paper is arranged as follows: Sect. 2 presents the preliminaries. Section 3 explains the sensor validation problem. Section 4 presents the suggested SB-spectral clustering in DWN. Simulation result and validation are presented in Sect. 5. Summary is presented in Sect. 6.

2 Preliminaries

2.1 Model Problem

In this subsection, we demonstrate a graph model for sensor nodes clustering and validation; this model will be used to disclose faulty sensor nodes.

Let assume that the directed graph can be represented by $G = (V, T, W)$ where V is the vertex group (sensor nodes) and T indicates the group of all directional edges (target nodes), and W represents the (sensor nodes and target nodes) weights. G represents the graph model; this graph can be represented by adjacency sensor matrix X, and the entry of this matrix is the weight between sensor node a_i and target node b_j.

2.2 Scoring Function

The technique of *step function* is used for the function of measuring the **sensor nodes** and **reference targets** by location so that visualization and interpre-

tation are gained easily. Every sensor nodes and reference targets pair which show the measurement as a function of a *difference* between the location of the sensor nodes and reference targets are determined by measuring the distance. This measurement data can be implemented based on the scoring function. The sensor targets weight X is driven by applying the **scoring function** of Fig. 2.

3 Sensor Validation Problem

Given a sensor nodes sets with fixed number of sensor nodes, referred as $V=\{x_1, \ldots, x_n\}$ that is intensely connected, each sensor x_i can communicate to x_j a number of hops. They usually sent to their one-hop neighbors, where the received signal strength values are high. Therefore, the weighted adjacency matrix $W=[w_{ij}]$ catches the connection of sensor nodes in the sensor network, employing the RSS as the goodness of the connection between two sensor nodes. This is referred to as w_{ij}. We refer to $G = (V, T, W)$ as the directed graph set up by the sensor network. V is the group of all sensor nodes, T is the group of all target nodes to examine among sensor nodes (member), and W is the weighted adjacency matrix. The node in the environment and power usage will report a like measurement of RSS, supposing that a properly working sensor of the same property such as radio and antenna characteristics. Therefore, the working sensors are identified as a good sensor, whereas other sensors which do not report property measurements are called bad sensors. This terminology is shown by using a simple example in which the sensors are indexed by their antenna direction. Thus, adjacency sensors are known as those sensors which have common antenna direction [1]. The purpose of sensor nodes validation model is to identify those bad sensors. This aims a fulfilled by detecting abnormal connection patterns attained from bad sensors. To put it in another exact word, we should solve the problem of identifying if a sensor part of the group of the adjacent sensor or not to determine bad sensors. We suppose that cluster of the adjacent sensor will be large because there are a lot of sensors. A sensor is taken to be as bad sensors in two cases:

1. If the sensor is in unique smell cluster or
2. If the sensor is in a smell out of the side of the large cluster.

4 SB-Spectral Clustering Algorithm

In this section, we describe the SB algorithm for sensor nodes validation in DWN by using SB-spectral clustering (SPC) algorithm. SPC functions nonlinear transformation of low measurement data from the input space R^n into high-dimensional space named noted as the new feature. A nonlinear not dividing in the input space is resulted from computing a linear dividing in the feature space. This transformation increases the divisibility of the input sensor nodes measurement data in the high-dimensional space.

The SB-SPC algorithms comprise of six steps and can be concluded as follows:

Algorithm 1: The SB-SPC Algorithm

1 Given a sensor nodes set N={(s_i,t_j)|s_i,$t_j \in R^{n \times n}$, i, j=1,...,N}.

2 Construct the sensor-targets weight matrix W using the Euclidean distance.
3 Construct sensor-to-sensor matrix B.
4 Compute the eigenvalues v_1,..., v_n and eigenvectors u_1,..., u_n of B.
5 Let $X \in R^{n \times k}$,be the new matrix containing the vectors u_1,..., u_k as columns.
6 For $i = 1,...,n$, let $s_i \in R^k$ be the vector corresponding to the i-th row of X.
7 Cluster the sensor nodes (s_i) i= 1,. . . ,N in R^K with sign based clustering into clusters ($+_i, \ldots, +_n$).

8 K clusters C_1,..., C_K.

5 Simulation Study

In this paper, the network is intensely connected because the sensor in the topology of uniform deployed, the network with size $s = 46$ sensor nodes, and this sensor is put on one line with equal distance to produce the topology among them. This scenario has 46 sensor nodes and 7 sensor targets which are distributed in the line. The other parameters set for this algorithm are presented in Table 1. After gaining the topology RSS values between the sensors and targets, pairs are irregularly produced the adjacent sensor matrix by location index. Thus, a group of bad sensor less than 46 is irregularly selected to be bad sensor nodes, and their RSS values are altered in the entry matrix W to their neighbors importantly. Therefore, the sensors in the largest eigenvectors are clustered by measurement values.

We apply the first method in the suppose scheme graph-based SPC algorithm on the basis of value and the largest eigenvector of the inverse matrix $W^T W$.

Table 1. Network parameters of SB-SPC algorithm

Parameters	Values
Network type	stationary
Distributed type	Random
Network size	90 m
Number of target nodes	7
Number of bad node	1, 5
Number of sensor nodes	46
Initial power energy	0.5 J
Packet size	1024 bits

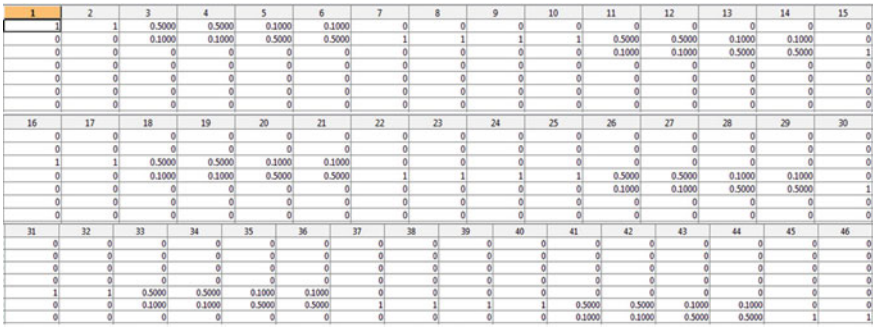


Fig. 1. Main entry weight matrix W composed of **46** sensor nodes and **7** target nodes.

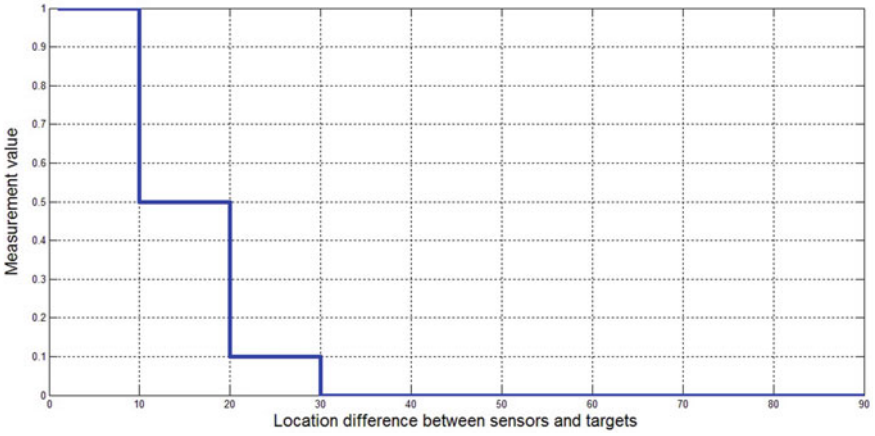


Fig. 2. Step function

5.1 Simulation Results and Discussion

Step function [0–4] = 1, [4–8] = 0.5, [8–12] = 0.1, otherwise = 0; Fig. 2 shows the step function. Reference sensors = (7) with reference polarizations as shown in Table 2.

Sensor nodes (1–46) = 46 sensors with different polarizations between 0 and 90. The first approach of clustering is called as valued-based grouping; first, second, third and fourth eigenvectors groups are demonstrated in Fig. 3 the four vectors. This method employs only four eigenvectors to show the grouping of sensor nodes in several clusters. Each one of this eigenvector may have positive or negative value; the first vector has only positive value, whereas the second, third and fourth have both values (Fig. 1).

The values-based clustering algorithm shows only 19 clusters: $\{[1,2],[3,4],[5,6],[7,8,9,10],[11,12],[13,14],[15,16,17],[18,19],[20,21],$

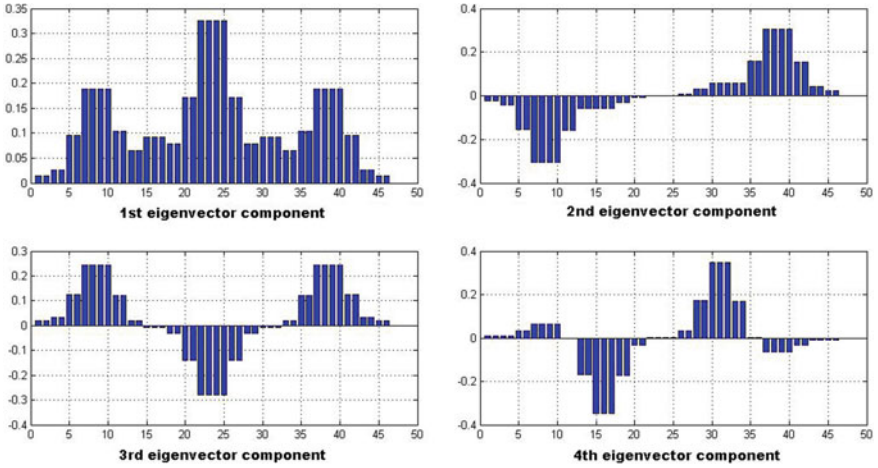


Fig. 3. Four eigenvectors of the value-based methods

Table 2. Polarization value

No	1	2	3	4	5	6	7
Polarization	0	15	30	45	60	75	90

[22, 23, 24, 25], [26, 27], [28, 29], [30, 31, 32], [33, 34], [35, 36], [37, 38, 39, 40], [41, 42], [43, 44], [45, 46]}.

The SB cluster algorithm shows four clusters only; Fig. 4 shows that, and Table 3 summary the clustering results in the second method as the following:

Clustering result from Table 3 in sign-based method with four clusters, and these group are $1 = (+-+)$, $0 = (+- -)$, $2 = (++-)$ and $3 = (+++)$.

5.2 Validation Studies

Assume five sensor nodes as bad, i.e., (8, 16, 26, 35 and 45).
Now in the second validation scenario of sensor nodes, five sensor nodes enter as bad to the simulation where $B = 5$ and will change the value of eigenvector $K = 3$, $K = 4$, $K = 6$; in each scenario of the validation study, we increase the number of eigenvectors successively to re-identify the incorrectly misclassified the sensor as good sensor by the algorithm which should be classified as died sensor in the scenario.

Let us assume that sensors 8, 16, 26, 35 and 45 show malfunctions and gains error measurements. Sensor 8 has error measurements on targets 1 which are same to those of sensors 1 and 2. Sensor 16 obtains error measurements on targets 2 which are like to those of sensors 7, 8, 9 and 10. Sensor 26 obtains

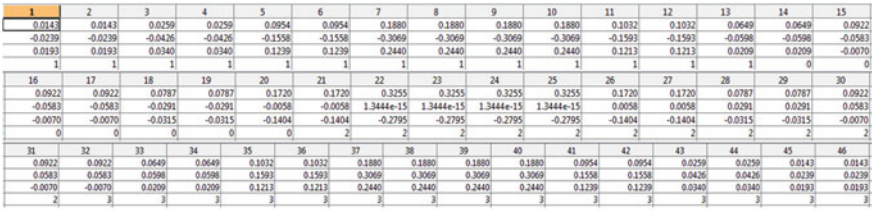


Fig. 4. Sign-based method with four groups (0, 1, 2, 3)

Table 3. Sign-based result

Sensor node	1–13	14–20	21–31	32–46
Signs-based	+ − +	+ − −	+ + −	+ + +
Cluster nodes	1	0	2	3

incorrect measurements on targets 3 that are related to those of sensors 15, 16 and 17 in the original data measurement matrix Fig. 1. Sensor 35 obtains error measurements on targets 4 which resemble sensor 27. Sensor 45 obtains error measurements on targets 5 which are like to those of sensors 33 and 34. Fig. 5 demonstrates the entrance W matrix of malfunctioning sensor nodes.

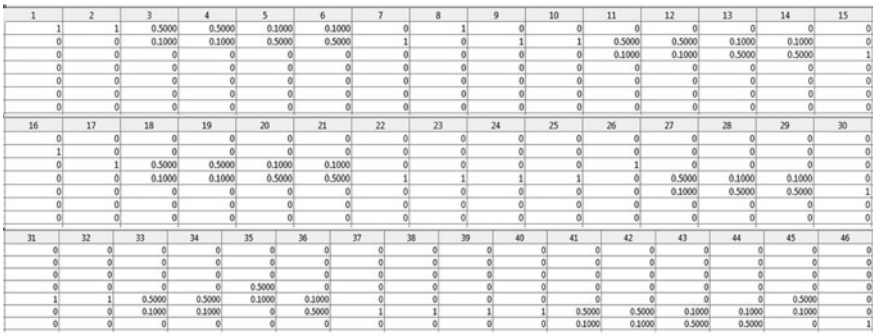


Fig. 5. W, Weight matrix with five sensor nodes as bad, i.e., (8, 16, 26, 35 and 45)

The first scenario $K = 3$.
Firstly, in the three leading eigenvectors of matrix $W^T W$, we see that now sensor 16 is grouped in a group classified as 2 while its right near sensor belongs to a big group classified as 3. Therefore, sensor 16 is in a small outside from its element of a big group. This designates that our spectral grouping technique has accurately specified sensor 16 as a malfunctioning sensor. Unlikely, sensor 17 which also grouped in a group classify as 3 while its right close to sensors belongs to a big group classify as 1. Therefore, sensor 17 is in a small out from its element of a

big group. In the wrong direction, this indicates that our SB-spectral grouping algorithm has wrongly misidentified sensor 17 as a not working sensor. But if we see to the left nearby after one position (small filter f), we can note that the sensor 17 is not in a small out of place its element of a big group which leads to recognizing this sensor as a working sensor Fig. 6 presents all these exchanges happen in the data measurement.

A similar scenario of sensors 16 and 17 has appeared for sensors 26 and 27 where our spectral grouping approach has perfectly recognized sensor node 26 as a malfunctioning sensor and recognized sensor 27 as a good sensor. However, we note that now sensor 35 is grouped in a group classified as 1 while its adjacent sensors belong to a big group assorted as 0. So, sensor 35 is in a small outside from its element of a big group. This confirms that our SB-spectral grouping technique has accurately determined sensor 35 as a malfunctioning sensor. Finally, sensors 8 and 45 have clustered as good sensors which add another fail of small **k** methods. Fig. 6 shows both bad and non-bad nodes and illustrates all alters of five sensors as deemed as bad. This erroneous in the result can be overcome with raising the value of k as we will see.

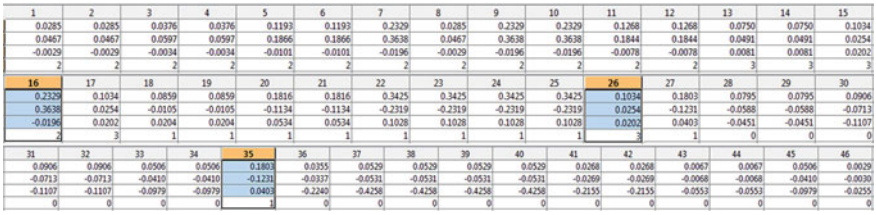


Fig. 6. Sign-based with five modify sensors as bad and $K = 3$

From sign-based at $k = 3$, we found clustering result such as 10 clusters under 4 groups, and these groups are $0 = (+\,-\,-)$, $1 = (+\,-\,+)$, $2 = (+\,+\,-)$, $3 = (+\,+\,+)$. Table 4 presents result of both cluster and major group.

Table 4. Clustering results in sign-based where $k = 3$

Sensor nodes	1–12	13–15	16	17	18–25	26	27	28–34	35	36–46
Signs-based	++−	+++	++−	+++	+−+	+++	+−+	+− −	+−+	+− −
Cluster-Label	2	3	2	3	1	3	1	0	1	0

A result obtained in sign-based of $K = 3$, 10 clusters under four groups, we insert B=5, and we found B=3.

The second scenario $K = 4$.
In this validation, we perform the validity study on four eigenvectors of matrix $W^T W$. From these matrices, we note that sensor 16 is grouped in a group classified as 2 while its right adjacent sensor nodes connect to a small group labeled

as 3. Consequently, sensor 16 is in a small outside of its region of a big group. This indicates that our spectral grouping approach has perfectly defined sensor 16 as a malfunctioning sensor. Unlikely, sensor 17 which also grouped in a group is classified as 3 while its right close sensor nodes belong to a big group classified as 1. Therefore, sensor 17 is in a small outside from its region of a big group. In the wrong direction, this signifies that our SB-spectral grouping technique has wrongly defined sensor 17 as a malfunctioning sensor. But if we see to the left nearby after one position (small filter f) we can note that the sensor 17 is not in a small out from its place of a big cluster which leads to identifying this sensor as a function sensor node.

In the previous scenario, we see the result of both sensor nodes (26 and 27) where our method has correctly determined sensor 26 as bad labeled as 3, and sensor 27 as the no-working sensor in a large group labeled as 1. In this scenario of $k = 4$, the sensors (26 and 27) has been studied and the result our SB-spectral grouping technique has accurately determined sensor 26 as a not working sensor labeled as 3, and identified sensor 27 as a good sensor in the new small group labeled as 4. However, we observe that now sensor 35 is grouped in a group labeled as 1 while its close sensors belong to a big group labeled as 5. Therefore, sensor 35 is in a small an away from its region of a big group. This emphasizes that our SB-spectral grouping algorithm has rightly specified sensor 35 as a malfunctioning sensor. Lastly, sensor 8 has been grouped as working sensor, and 45 has clustered as fault sensor labeled as 0, and sensor 45 has clustered as fault sensor in a large group classified as 0 while its close sensors belong to a big group labeled as 5 Fig. 7 shows all bad sensor nodes with blue color. Sensor 8 has added another fail of small k methods figure shows both bad and non-bad nodes. This wrong result can be overcome by increasing the value of K to 6.

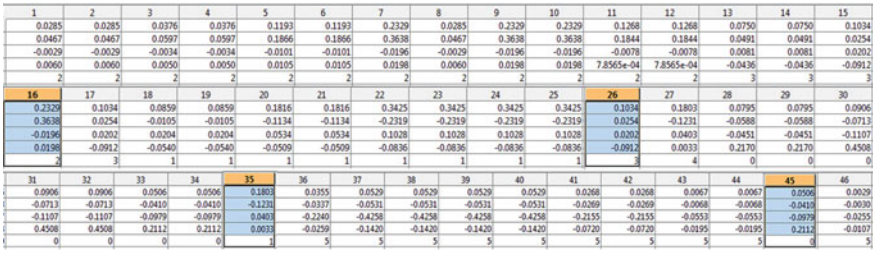


Fig. 7. Sign-based with five modify sensors as bad and $K = 4$

Result obtained from sign-based algorithm at $k = 4$; we gained 12 clusters under 6 groups, Table 5 illustrates these groups. All these groups are presented with signs. First group is $0 = (+ - - +)$, second group is $1 = (+ - + +)$, third group is $2 = (+ + - +)$, forth group is $3 = (+ + + -)$, fifth group is $4 = (+ - + +)$, and sixth group is $5 = (+ - - -)$.

The best result we got from $k = 4$ is that sensor 45 is grouped in a group classified as 0 while its close to a sensors link to a big group labeled as 5. So,

Table 5. Clustering results in sign-based where $k = 4$

Sensor nodes	1–12	13–15	16	17	18–25	26	27	28–34	35	36–44	45	46
Signs-based	++−+	+++−	++−+	+++−	+−+−	+++−	+−++	+−−+	+−++	+−−−	+−−+	+−−−
Cluster-Label	2	3	2	3	1	3	4	0	1	5	0	5

sensor 45 is in a small out from its element of a big group labeled as 0. This indicates that our SB-spectral grouping algorithm has correctly classified sensor 45 as a malfunctioning sensor. Even truly specified sensor 27 as a good sensor while its a good sensor. Sensor 8 remains without any clustering which needs to go to higher K eigenvector value.

The third scenario. $K = 6$

In this scenario, we study the validation of the five sensor nodes that entered to the simulation by altering the signal of these sensors in weight matrix W as shown in Fig. 5. Thus, to prove that the altered receive signal strength (RSS) of sensor nodes can put the wireless network down, this erroneous may happen due to bad sensor node present in the wireless network. Sensor 8 enters as malfunctioning and link to a big group classified as 3; this sensor determines with value 3, and its nearby sensors are grouped with classified as 2. Consequently, sensor 8 is in a small outside from its region component of a big group. This indicates that our spectral grouping approach has perfectly classified sensor 8 as a malfunctioning sensor with increasing the value of k eigenvectors to $k = 6$ while it was not classified as a not working sensor in $k = 3$, and $k = 4$. Sensor 16 marks as bad sensor by exchanging the value in weight matrix from (0 0 1 0 0 0 0) to new value such as (0 1 0 0 0 0 0); this sensor is clustered into group as 2 while its near sensors have different group values such as 0. Sensors 26 and 35 have been specified as a bad sensor by our method, where sensor 26 is grouped in a group with class value labeled as 0 and its left neighbor sensor 25 has different class value labeled as 6. And the right neighbor sensor 27 has new class value labeled as 7. Therefore, sensors 26 and 35 are specified as bad nodes due to its labeled values far away from its components. Finally, sensor 45 has been determined as a bad node in $k = 6$; this new identification has added new success of large k methods Fig. 8 shows all five bad sensors.

The good result we obtained so far in $k = 6$ is to prove that increasing the value of k can be useful in showing clearly the working sensor from non-working sensor in distributed wireless network, and also to support that our sign-based algorithm is strongly recommended for large wide area of wireless network to identify a good and bad sensor node in the wireless network. While the value-based method is not effective for large wireless network validation study due to the wrong grouping of sensors.

From Fig. 9, we see 19 clusters and 10 subgroups when $K = 6$, and this table shows clearly that sensor 8 and sensor 45 are bad nodes where each sensor

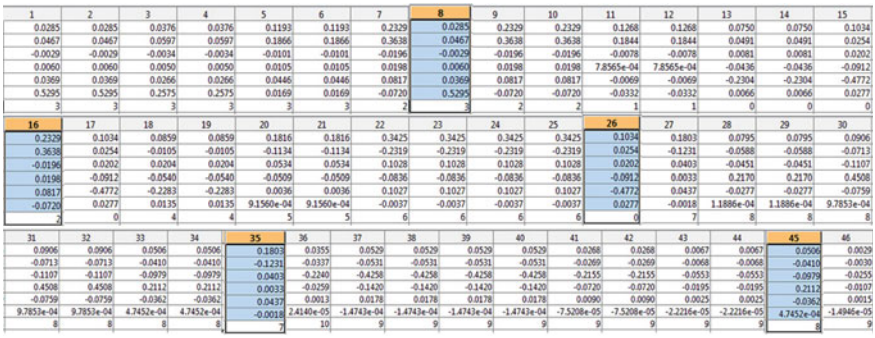


Fig. 8. Sign-based with five modify sensors as bad and $k = 6$

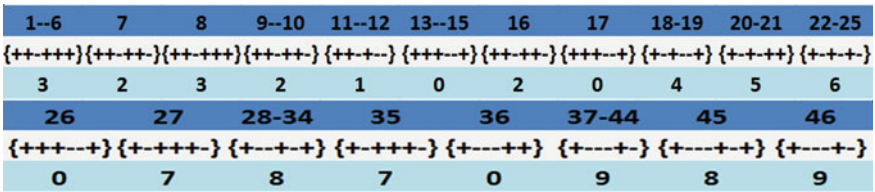


Fig. 9. Clustering results in sign-based with 19 clusters and 10 major groups

carries different labeled values. Result was obtained from sign-based method at $k = 6$; we gained 19 clusters under 10 groups, Figure 9 reveals these groups. This group is presented with various signs. First group is $0 = (+++- \; -+)$, second group is $1 = (++-+- \; -)$, third group is $2 = (++-++-)$, forth group is $3 = (++-+++)$, fifth group is $4 = (+-+- \; -+)$, sixth group is $5 = (+-+-++)$, seventh group is $6 = (+-+-+-)$, eighth group is $7 = (+-+++-)$, ninth group is $8 = (+- \; -+-+)$, and last group is $9 = (+- \; - \; -+-)$.
From previous validation studies and graphs, we see clearly this important result in Table 6.

Table 6. Malfunction sensor

Sensors	Number of bad nodes	Number of eigen-vector	Number of sensor	Total number of bad nodes
8, 16, 26, 35, 45	5	$k = 3$	3	3
8, 16, 26, 35, 45	5	$k = 4$	4	4
8, 16, 26, 35, 45	5	$k = 6$	5	5

6 Conclusion

In this paper, SPC is a good and simple method which is used in DWN. There are two methods (value-based and sign-based) used in SPC for grouping sensor nodes and validation of sensor node by entering different number of bad sensor nodes into the simulation. The performance of SB-SPC has been tested on different number of sensor nodes as died sensor entered into the simulation with increasing the number of eigenvectors, respectively, in each scenarios in order to identify those bad nodes. The performance of developed method is contrasted with state-of-the-art method. The empirical results show that the developed method perfectly identifies the exact number of bad sensor nodes as entered into the simulation as we consider more eigenvectors in sign-based method, and gain good performance in WSNs. The obtained result from different scenarios demonstrate that by including more eigenvectors can be useful in specifying the bad sensor node, especially in sign-based technique which is recommended for large WSNs.

The scope of future work, we plane to integrate the SPC technique with ABC metaheuristic technique and compare the performance of the network with result of SPC technique.

References

1. Kung, H.T., Vlah, D.: Validating sensors in the field via spectral clustering based on their measurement data. In: Proceedings of the IEEE Military Communications Conference (MILCOM), pp. 1–10 (2009)
2. Kung, H.T., Vlah, D.: Sign-based spectral clustering. In: Proceedings of the 25th Biennial Symposium on Communications (BSC), pp. 32–39 (2010)
3. Nasser, A.M.T., Pawar, V.P.: Machine learning approach for sensors validation and clustering. In: Proceedings of the 8th IEEE International Conference on Emerging Research in Electronics Computer Science and Technology (ICERECT), pp. 370–375 (2015)
4. Almajidi, A.M., Pawar, V.P.: A new system model for sensor node validation by using OPNET. Wirel. Pers. Commun **108**(4), 2389–2401 (2019). https://doi.org/10.1007/s11277-019-06527-3
5. Almajidi, A.M., Pawar, V.P.: ABC-based algorithm for clustering and validating WSNs. In: International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), pp. 1–9 (2019)
6. Kung, H.T., Vlah, D.: A spectral clustering approach to validating sensors via their peers in distributed sensor networks. In: Proceedings of the 18th IEEE International Conference on Computer Communications and Networks (ICCCN), vol. 8, pp. 1–7 (2009)
7. Almajidi, A.M., Pawar, V.P.: K-means-based method for clustering and validating wireless sensor network. In: International Conference on Innovative Computing and Communications (ICICC), vol. 55, pp. 251–258 (2018)

Analysis of Precoding-Based Pilot Decontamination Methods in Massive MIMO Systems

Angelina Misso(✉), Baraka Maiseli, and Mussa Kissaka

University of Dar es Salaam, Dar es Salaam P.O. Box 33335, Tanzania
{amisso,barakaezra,kissaka.mussa}@udsm.ac.tz

Abstract. During transmission of information, massive multi-input multiple-output (MIMO) antenna system encounters inter-cell interference, called pilot contamination that limits the capacity. Recent developments in massive MIMO technology have provided a plethora of methods to reduce pilot contamination using precoding in massive MIMO. However, a trade-off exists between the system performance based on achieved spectral efficiency in relation to the sumrates and bit error rates and complexity in terms of iterations count, the number of optimization variables and the relative time consumed to resolve the optimization problems. The current literature contains insufficient information to address this trade-off, a consequence that hinders the advancement of research in pilot contamination. This study covers the gap through a detailed review of various linear and nonlinear schemes centered on the two contending metrics, namely spectral efficiency and complexity. A systematic review approach is adopted to analyze different related studies and their associated methodologies, results and limitations. Moreover, we provide recommendations for future research.

Keywords: Massive MIMO · Pilot contamination · Linear precoding · Nonlinear precoding

1 Introduction

Massive multiple-input multiple-output (MIMO) system consists of several numbers of antenna arrays at the base station, to improve the spectral efficiency [1, 2]. The capacity enhancement in massive MIMO is a result of spatial multiplexing [1]. Nevertheless, in a multiple cell scenario, inter-cell interference called pilot contamination limits massive MIMO's capacity [3].

The current studies have explored various methods to reduce pilot contamination in massive MIMO using precoding schemes [4–10]. However, limited literature exists on the comparative analysis of various precoding techniques [2, 11–13], most of which focus on linear precoding schemes. Therefore, an extensive study is essential to compare and quantify the performance of linear [14] and nonlinear [3] precoding schemes.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_9

The current study, therefore, analyzes the linear and nonlinear massive MIMO precoding schemes to reduce pilot contamination using computational complexity and spectral efficiency metrics. This study applies a systematic approach to examine different works of the literature that utilize precoding to alleviate pilot contamination in massive MIMO. Specifically, the current study (1) analyzes related works and the respective methodologies, results and limitations on pilot decontamination; (2) compares related works based on achieved results using computational complexity and spectral efficiency and (3) provides recommendations on future research direction.

The rest of the paper is structured as follows: Sect. 2 contains the literature review and Sect. 3 covers the systematic review approach used in this study to select the works of literature. Section 4 analyzes the related works on pilot contamination precoding. Section 5 presents the comparative analysis of related works based on achieved results using complexity and system performance as metrics. Section 6 presents the recommendations for further research and Sect. 7 concludes the paper.

2 Literature Review

Pilot contamination is a result of pilot reuse during estimation of channel owing to the inadequate number of orthogonal pilots under constraints of coherence interval [3]. However, using precoding, pilot contamination can be mitigated [4, 5]. Consider a network with m antennas at the base station, l number of cells, k number of user terminals, channel matrix $\boldsymbol{H}$, $\boldsymbol{n}$ as the noise matrix, ρ is the signal to noise ratio, $\boldsymbol{w}$ being the precoding matrix, $\boldsymbol{x}$ and $\boldsymbol{y}$ being the transmitted and received signal matrix, respectively. In a multi-cell scenario with pilot contamination,

$$\boldsymbol{y} = \sqrt{\rho} \sum_{k=1}^{K} \boldsymbol{h}_{lk}^{lm} \boldsymbol{x}_{lk} \boldsymbol{w}_{\boldsymbol{l}} + \sqrt{\rho} \sum_{l'=1}^{L} \sum_{k=1}^{K} \boldsymbol{h}_{l'k}^{lm} \boldsymbol{x}_{l'k} \boldsymbol{w}_{\boldsymbol{l}} + \boldsymbol{n} \tag{1}$$

Pilot contamination precoding (PCP) utilizes precoding to mitigate pilot contamination by exploiting collaboration, though channel state information (CSI) sharing among base station in massive MIMO network [5]. Using the principle of channel hardening that occurs in massive MIMO, the channel characteristics can be reduced to large-scale parameters only; thus, the decoding matrix is computed. By applying the law of large numbers, with increased number of antennas, the transmitted signal can be recovered.

3 Methodology

This research adopts a systematic review approach for analysis of recently published research. The following databases were used to search for relevant literature: Scopus, IEEE Xplore, arXiv and Science Direct. The search included journal and conference articles published in the English language only within the last ten years (2009–2019), as massive MIMO popularity has experienced widespread research in this period. The following keywords were used to search for appropriate literature relevant to this study: massive MIMO, pilot decontamination, precoding and pilot contamination, and 326

papers were found. The screening was performed based on titles and abstracts; then, 202 papers were included and 124 papers were left out. The papers were further filtered and 78 papers that focused on linear and nonlinear precoding methods remained. A further detailed analysis was performed based on the results and limitations; hence, 22 papers remained for comprehensive analysis.

4 Related Works on Pilot Decontamination Precoding

To assess the performance of linear and nonlinear precoding schemes, twenty-two works of the literature were examined. This section presents an analysis of related works by reviewing the methodologies, results and limitations. Early research on pilot contamination precoding was conducted by Jose et al. [4], whereby a precoding method based on MMSE scheme was developed. The precoding method reduced intra-cell and inter-cell interference, and numerical results showed significant performance gains. Moreover, a PCP study was conducted by Ashikhmin and Marzetta [5] for user stations from different cells that were reusing similar pilot sequence. The results showed that pilot contamination could be suppressed using precoding schemes.

A study by Xu et al. [6] explored different approaches to reduce pilot contamination using MMSE estimator. Also, the authors in [7] examined the massive MIMO downlink performance in terms of spectral efficiency for the ZF precoding. Moreover, the authors [8] analyzed the performance of ZF precoding scheme in comparison with MRT precoding scheme, whereby ZF achieved higher rates than MRT.

Furthermore, the author [9] proposed the use of MRT as a low-complexity linear precoding scheme to achieve a better bit error rate (BER) at a lower computational complexity. Further research by Tan et al. [14] provided a comparative analysis between ZF, MRT and MMS, whereby MRT was inferior to the ZF and MMSE precoding schemes. However, MRT precoding scheme outperformed ZF precoding schemes at low SNR.

The authors in [15] used a spatial filter, and the achievable rate increased monotonically with achieved SINR. However, the proposed method was only useful for cell-edge users. Furthermore, research by Shahabi et al. [16] proposed a constant envelope precoding (CEP) using limited cooperation among cells to reduce overheads and at improved performance. However, as the number of base station increased, the performance gap between imperfect and perfect CSI scenarios increased.

Research by Hawej and Shayan [17] proposed a novel Iterative Weighted Nuclear Norm which showed better performance in terms of the normalized mean square error compared to the least square method even with high pilot contamination. However, in comparison, the scheme had high computational complexity. Another novel precoder based on singular value decomposition was proposed in [18]. The proposed scheme exhibited greater BER performance over the existing techniques. However, the application of the proposed scheme was limited to visible light.

The authors in [19] proposed a heuristic statistical beamforming method. The proposed interference-suppressed precoding methods achieved a higher system sumrate than conventional methods; nevertheless, the method ignored the CSI acquisition overhead. The proposed scheme in [20] based on MMSE has better performance compared to the conventional MMSE and ZF. However, the scheme was based on single-cell, assumed perfect CSI and limited feedback.

Research by Zhao et al. [21] proposed a Joint Pilot Assignment and Pilot Contamination Precoding based on ZF-PCP for improved spectral efficiency at low computational complexity. Additionally, a low-complexity MMSE detector algorithm was proposed in [22]. The proposed algorithm was based on damped Jacobi and achieved a performance closer to MMSE in terms of BER, however, showed increased complexity.

The authors in [23] proposed an adaptive joint reduced-length feedback filter/Tomlinson–Harashima precoding which reduced the computational complexity by 90%. However, the performance improvement is negligible for a smaller number of user terminals. Moreover, a low-complexity SSOR-based precoding scheme is proposed in [24] with reduced complexity compared to classical ZF. Nevertheless, classical ZF had better performance in comparison. Furthermore, a method called precoding assisted sum rate maximization of all users was proposed in [25] which outperformed the existing ZF-PCP in terms of sumrates but was more complex.

Improved multiple cell MMSE channel estimation was proposed in [26] which saved 61.2% computational cost compared to the MMSE scheme. Nevertheless, the system assumes large-scale fading which is known as a priori. Furthermore, Kronecker-type approximation is proposed in [27]. The results indicated that the limited feedback precoding provided improved achievable sumrate. However, the CSI feedback overhead is significantly high that may influence the capacity gain.

Furthermore, the authors in [28] proposed a first-order Neumann series for improved performance-complexity trade-off over the ideal ZF, but the performance deteriorates as the number of base station antennas increases. Furthermore, the authors in [37] proposed discrete-phase CEP approach, with achieved performance comparable to the conventional approaches at the cost of increased computational time.

The increase in the number of antennas in massive MIMO influences the complexity of precoders, even further difficult for nonlinear precoders. Researchers have strived to improve the performance of nonlinear precoders at reduced complexity, at least closer to that of linear precoders, while others have compromised the complexity for better spectral efficiency.

5 Analysis of Computational Complexity and Spectral Efficiency

To compare related works based on achieved results using complexity and spectral efficiency as metrics, this section presents a quantitative analysis of linear and nonlinear precoding schemes. MRT, ZF and MMSE schemes have a computational complexity of $O(\mathrm{MK}^2)$, $O(\mathrm{MK}^3)$ and $O(\mathrm{MK}^3)$, respectively [15, 22, 28]. Table 1 shows an analysis based on the complexity and massive MIMO performance for linear precoding schemes. From the literature, MMSE precoding provides an optimal performance compared to ZF and MRT, while ZF has better performance at higher computational complexity compared to MRT.

Table 2 presents a comparative analysis of studies based on spectral efficiency and computational complexity. The studies have been arranged based on the publication year, starting with the most recent work. From the table, K is the number of users, M is the number of base station antennas, T_c is the coherence interval, τ is the number of symbols per user, T is the total number of iterations and r is a real variable.

Table 1. Analysis based on the complexity and massive MIMO performance for linear precoding schemes

Linear precoding scheme	MRT	ZF	MMSE
Methodology	Complex the conjugate of the estimated matrix	Channel matrix pseudo-inverse	Channel matrix MSE
Complexity	Low	Medium	High
Spectral efficiency	Low	Medium	High

Table 2. A comparative analysis of various studies based on the complexity and massive MIMO performance

Study	Computational complexity analysis	Performance in terms of spectral efficiency
[15]	$O(\mathrm{MK}^2)$	Five percent of users will have an NMSE greater than 10 dB, compared with MF precoding
[16]	$O(\mathrm{MKL}^2)$	Achieves sumrates of 1.4 bps/Hz with 100 antennas at the base station
[29]	Switches between $O(MK^2)$ and $O(1/2\mathrm{MK}^2)$	Achieves sumrates of 110bps/Hz at SNR of 110 dB
[30]	$O(\mathrm{Tr}^2)$	Achieves NMSE of -8 dB at SNR of 20 dB and NMSE of -6 dB at SNR of 4 dB
[18]	$O(\mathrm{M}^2)$	Achieves a BER of 2×10^{-2} at 85 dB and BER of 10^{-5} at 105 dB
[28]	$O(4\mathrm{K}^2)$	Achieves near ZF performance with sumrates of 120 bps/Hz with 5000 antennas at the base station
[20]	$O(\tau M(2K-1))$	Achieves sumrates of 5.5 bps/Hz using 50 antennas and sumrates of 6 bps/Hz using 300 antennas
[22]	$O(K^2)$	Achieves near MMSE performance at reduced complexity with a BER of 10^{-1} at 5 dB
[24]	$T_c O(\mathrm{K}^2 + \mathrm{NK})$	BER of 10^{-5} at 26 dB and BER of 10^{-1} at 10 dB.
[25]	$O(\mathrm{TKL}^3)$	MSR-PCP has better performance compared with ZF—PCP with sumrates up to 74 pbs/Hz compared to 65 bps/Hz at 1000 antennas

The complexity in estimation of the channel is attributed to a large channel matrix in massive MIMO. As shown in Table 1, linear precoders are less complex but have limited spectral efficiency compared with nonlinear precoders [2, 11]. In Table 2, many nonlinear precoder designs strive to reduce the complexity close to linear precoders [22, 28], while others have improved performance at the cost of increased complexity [16]. Interestingly, other researchers have explored combined adaptive precoders [29]. Due to difficulty in determining the suitable precoding scheme at reasonable complexity, optimization is crucial to leverage between the performances metrics.

6 Recommendations for Future Research

The analysis shows a compromise between spectral efficiency performance parameters and computational complexity. Research has shown that nonlinear precoding schemes offer optimal capacity compared with linear schemes but are difficult to implement due to high computational complexity [11, 20]. Further analysis is required to ascertain efficient precoders that will balance between performance and complexity [11]. Researchers could explore a combination of multiple precoding schemes that adapt to the channel conditions through analysis of the CSI to determine which precoder to use at a particular instance and thus improve the overall system performance [29].

7 Conclusions

This paper presents an analysis of the literature based on linear and nonlinear precoding methods. A systematic approach is used to analyze twenty-two different published researches on linear and nonlinear precoding based on the respective methodologies, results and limitations associated with pilot contamination. Furthermore, ten studies were selected for further analysis based on the two trade-off parameters, computational complexity and spectral efficiency. The analysis depicted that nonlinear precoding schemes have better spectral performance compared to linear precoding schemes at the cost of increased complexity. As a recommendation, further analysis is required to determine efficient precoders that will balance between the two metrics.

References

1. Larsson, E.G., Edfors, O., Tufvesson, F., Marzetta, T.L.: Massive MIMO for next generation wireless systems. IEEE Commun. Mag. **52**, 186–195 (2014). https://doi.org/10.1109/MCOM.2014.6736761
2. Hassan, N.: Massive MIMO wireless networks : an overview. Electronics. 1–29 (2017). https://doi.org/10.3390/electronics6030063
3. Zhao, L., Zhao, H., Zheng, K., Xiang, W.: Massive MIMO in 5G networks: selected applications. (2018)
4. Jose, J., Ashikhmin, A., Marzetta, T.L., Vishwanath, S.: Pilot contamination and precoding in multi-cell TDD systems. IEEE Trans. Wirel. Commun. **10**, 2640–2651 (2011). https://doi.org/10.1109/TWC.2011.060711.101155

5. Ashikhmin, A., Marzetta, T.: Pilot contamination precoding in multi-cell large scale antenna systems. In: IEEE International Symposium on Information Theory Proceedings, pp. 1137–1141 (2012). https://doi.org/10.1109/ISIT.2012.6283031
6. Xu, P., Wang, J., Wang, J.: Effect of pilot contamination on channel estimation in massive MIMO systems. In: 2013 International Conference on Wireless Communications and Signal Processing, WCSP (2013). https://doi.org/10.1109/WCSP.2013.6677112
7. Zhao, L., Zheng, K., Long, H., Zhao, H.: Performance analysis for downlink massive MIMO system with ZF precoding. Trans. Emerg. Telecommun. Technol. **25**, 1219–1230 (2014). https://doi.org/10.1002/ett.2745
8. Parfait, T., Kuang, Y., Jerry, K.: Performance analysis and comparison of ZF and MRT based downlink massive MIMO systems. In: 2014 sixth international conference on ubiquitous and future networks, pp. 383–388 (2014). https://doi.org/10.1109/ICUFN.2014.6876818
9. Zhang, Y., Gao, J., Liu, Y.: MRT precoding in downlink multi-user MIMO systems. Eurasip J. Wirel. Commun. Netw. (2016). https://doi.org/10.1186/s13638-016-0738-6
10. Zhao, Y., Lin, Z., Liu, K., Zhang, C., Huang, L.: Pilot contamination reduction in massive MIMO system. In: 2018 IEEE International Conference on Consumer Electronics, pp. 1–2 (2018). https://doi.org/10.1109/ICCE-China.2018.8448529
11. Fatema, N., Hua, G., Xiang, Y.: Massive MIMO linear precoding: a survey. IEEE Syst. J. **12**, 3920–3931 (2018). https://doi.org/10.1109/JSYST.2017.2776401
12. Mazrouei-sebdani, M., Melzer, J.: Massive MIMO with nonlinear precoding: large-system analysis. IEEE Trans. Vehicular Technol. **65**, 2815–2820 (2016). https://doi.org/10.1109/TVT.2015.2425884
13. Mukubwa, E., Sokoya, O.A., Ilcev, D.S.: Comparison and analysis of massive MIMO linear precoding schemes in the downlink. In: 2017 IEEE AFRICON: Science, Technology and Innovation for Africa, AFRICON 2017. pp. 187–191 (2017)
14. Tan, W., Huang, W., Yang, X., Shi, Z., Liu, W., Fan, L.: Multiuser precoding scheme and achievable rate analysis for massive MIMO system. Eurasip J. Wirel. Commun. Netw. 1–12 (2018)
15. Gong, Z., Li, C., Jiang, F.: Pilot decontamination in noncooperative massive MIMO cellular networks based on spatial filtering. IEEE Trans. Wirel. Commun. **18**, 1419–1433 (2019). https://doi.org/10.1109/TWC.2019.2892775
16. Shahabi, S.M., Ardebilipour, M., Omid, Y., Kazemi, M.: Low-overhead constant envelope precoding in multi-cell massive MIMO systems with pilot contamination. IET Commun. **13**, 926–933 (2019). https://doi.org/10.1049/iet-com.2018.5196
17. Hawej, M., Shayan, Y.R.: Pilot decontamination in massive multiuser MIMO systems based on low-rank matrix approximation. IET Commun. **13**, 594–600 (2018). https://doi.org/10.1049/iet-com.2018.5654
18. Jain, S., Mitra, R., Bhatia, V.: Adaptive precoding-based detection algorithm for massive MIMO visible light communication. IEEE Commun. Lett. **22**, 1842–1845 (2018). https://doi.org/10.1109/LCOMM.2018.2855975
19. Qiu, S., Chen, D., Qu, D., Luo, K., Jiang, T.: Downlink precoding with mixed statistical and imperfect instantaneous CSI for massive MIMO systems. IEEE Trans. Veh. Technol. **67**, 3028–3041 (2018). https://doi.org/10.1109/TVT.2017.2774836
20. Ge, Z., Haiyan, W.: Linear precoding design for massive MIMO based on the minimum mean square error algorithm. Eurasip J. Embed. Syst. (2017). https://doi.org/10.1186/s13639-016-0064-4
21. Zhao, M., Zhang, H., Guo, S., Yuan, D.: Joint pilot assignment and pilot contamination precoding design for massive MIMO systems. In: 2017 26th Wireless Optical Communication Conference, pp. 1–6 (2017). https://doi.org/10.1109/WOCC.2017.7929001

22. Minango, J., De Almeida, C.: Low-complexity MMSE detector based on refinement Gauss-Seidel method for massive MIMO systems. In: 2017 IEEE 9th Latin-American Conf. Commun. LATINCOM 2017. 2017-Janua, 1–5 (2017). https://doi.org/10.1109/LATINCOM.2017.8240166
23. Xiong, D., Peng, W., Chen, D., Jiang, T.: Co Co. Wirel. Commun. Mob. Comput. pp. 3190–3200 (2012). https://doi.org/10.1002/wcm
24. Xie, T., Dai, L., Gao, X., Dai, X., Zhao, Y.: Low-complexity SSOR-Based precoding for massive MIMO systems. IEEE Commun. Lett. **20**, 744–747 (2016). https://doi.org/10.1109/LCOMM.2016.2525807
25. Liu, J., Li, Y., Zhang, H., Guo, S., Yuan, D.: Pilot contamination precoding assisted sum rate maximization for multi-cell massive MIMO systems, pp. 60–65 (2016)
26. Li, K., Song, X., Omair Ahmad, M., Swamy, M.N.S.: An improved multicell MMSE channel estimation in a massive MIMO system. Int. J. Antennas Propag. (2014). https://doi.org/10.1155/2014/387436
27. Kazemi, M., Aghaeinia, H., Duman, T.M.: Discrete-phase constant envelope precoding for massive MIMO systems (2017)
28. Zhang, C., Luxi, Y., Jing, Y., Huang, Y., Yang, L., Zhang, C., Huang, Y., Yang, L.: Performance analysis for massive MIMO downlink with low complexity approximate zero-forcing precoding. IEEE Trans. Commun. (2018). https://doi.org/10.1109/TCOMM.2018.2823715
29. Farsaei, A., Willems, F., Gustavsson, U.: A reduced-complexity linear precoding strategy for massive MIMO base stations. In: 2018 25th International Conference on Telecommunications (ICT). pp. 121–126. IEEE (2018)
30. Su, X., Zeng, J., Li, J., Rong, L., Liu, L., Xu, X., Wang, J.: Limited feedback precoding for massive MIMO. Int. J. Antennas Propag. **2013**, 1–9 (2013). https://doi.org/10.1155/2013/416352

A Quantitative Assessment of the Television White Space in Tanzania: A Case of Coastal and Morogoro Regions

Moses Ismail(✉) and Mussa Kissaka

University of Dar Es Salaam, P.O. Box 33335, Dar Es Salaam, Tanzania
moses82004@gmail.com, mkissaka@yahoo.com

Abstract. In this paper, we establish suitable propagation models and use them to estimate the potential of TVWS for Coastal and Morogoro regions in Tanzania. Four mostly used propagation models, Free Space, Egli, Ericsson and Hata were analyzed. We established an AM-Ericsson model from the Ericsson model, which was closest to the measured results in rural areas but still with a 25 dB difference. The results showed that Hata was the suitable model for semi-urban areas and the established AM-Ericsson model had more than 20 dB improvement from the Ericsson model for rural areas. The study found that there is up to more than 95% TV channels availability for Coastal region and 97% availability for Morogoro region. This research provides a baseline information to facilitate the estimation of the potential TV channels for secondary use in Tanzania. The study also guides a proper regulation and planning of spectrum in the country.

Keywords: TVWS · Radio propagation models · Broadband · Access network technologies

1 Introduction

1.1 Background

Tanzania, like most developing countries, faces a great digital divide, with an Internet penetration rate of only 43% [1]. The rural community is most affected by this poor penetration rate and the communities represent more than 86% of the unconnected population [2]. On the other hand, Tanzania has a fiber backbone that provides broadband access to the district levels with a capacity of 4.8 Tbps and a span of more than 7560 km [3]. However, access technology to extend this coverage into rural areas remains a challenge [3]. There is a need for affordable and efficient access technology to provide Internet access to rural areas.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_10

With a growing demand for service provision, broadband connectivity has become a key area for research. Various connectivity technologies are available to provide broadband connectivity in rural and remote areas. Of these techniques, Television White Space (TVWS) is the latest candidate. TVWS refers to coverage holes created by unused TV channels in a specific geographic area. Due to its range and penetration benefits, successful implementation of the TVWS ensures a cost-effective broadband connectivity to areas that would otherwise be too expensive to serve. TVWS also allows a variety of applications because of the potentially huge frequency spectrum that is available.

Globally, the TVWS deployment has drawn a great interest from researchers and regulators. Since the decisions to open the TV band for secondary use by the Federal Communication Commission (FCC) in the USA and the Office of Communications (Ofcom) in the UK [4], many other countries have explored the possibility [5]. Some of the opportunities and applications that can be supported are described in the previous works by [6, 7]. In this study, we quantified the potential availability of the TVWS in Coastal and Morogoro regions, in Tanzania.

1.2 Tanzania Digital Terrestrial Television Broadcasting

Tanzania commenced the switch off of its analogue terrestrial television signals in December 2012—the first country in mainland sub-Saharan Africa to do so. Its analogue terrestrial broadcasting had a population coverage of about only 24%, with many viewers using free-to-air satellite television instead. With this switch off being policy rather than market driven, Tanzania managed a complete switch off in most areas ahead of the agreed ITU deadline of June 2015. The DVB-T2 was chosen as the standard for digital terrestrial television broadcasting (DTTB) technology [8]. By the beginning of 2015, Tanzania had completed the digital switch over [9].

In Tanzania, the deployed TV network for digital terrestrial transmission broadcast (DTTB) is a multi-frequency network (MFN) as presented in [10]. The operators assigned channels in Coastal and Morogoro regions are Star Media (T) Limited, Agape Associates Limited and Basic Transmission Limited [10].

2 Literature Review

We reviewed different studies conducted previously on the establishment of suitable propagation models for TV signals broadcasting, especially in the developing world where regions and countries share similar geographic features. The study also looked into studies on the quantitative assessment of the TVWS conducted elsewhere.

The USA, UK, India, Nigeria, South Africa and others have assessed the availability of the TVWS to properly plan and protect the incumbent TV operators [11]. In most of studies, the Hata and Egli models have shown better performance over others. Hata-Davidson model which was an improved version of Hata model showed better performance over longer distances greater than 50 km; however, it was still outperformed by Hata and Egli over shorter distances. In all these cases, Hata and Egli models were the closest to the measured results in most of these studies [12, 13].

Another study by Maloku et al. which was conducted in a mountainous rural areas showed that Hata-Davidson model was close to measured results for distances greater than 30 km but it was outperformed by Hata model for distances below 30 km [14]. The study by Pathania et al. conducted the performance evaluation of different path loss models for broadcasting applications [15]. Although the study found similar findings on the Hata and Hata-Davidson model like previous studies, COST-231 was the most suitable for the plane areas like border district of Punjab where the study was conducted. In [16], ten path loss models were compared and the Ericsson model showed good paths loss performance especially for rural areas.

Studies on quantifying the TVWS in many developing countries show a large amount of TVWS available for secondary deployment [17, 18]. In India, more than 90% of the channels were available in Pune city, while similar findings were obtained for Vietnam, Malawi and Nigeria. However, for Tanzania, there has not been specific research conducted to quantify the potential availability of TVWS in the country. The motivation for this research comes from the ongoing research in Tanzania on how to provide broadband connectivity to the rural population.

3 Methodology

In this study, we identified suitable propagation models for the semi-urban and rural areas and used them to assess each TV channel's availability. We describe the process in this section.

3.1 Choice of the Study Area

Since we aimed at quantifying the TVWS for Tanzania, we considered areas with two important features that characterize semi-urban and rural areas, which are mostly affected by the digital divide due to lack of broadband connectivity. The areas in the outskirt of Dar es Salaam and Pwani regions are semi-urban and are served by the same TV transmitter stations. We are referring to these two regions as Coastal region. A mountainous Morogoro region, which consists of several remote areas, characterizes most rural areas in the country. In this study, it represents the rural scenario, which is a suitable candidate for TVWS deployment.

3.2 Data Collection

Data were obtained from a previous study conducted in 2017 by the College of Information and Communication Technologies (CoICT), together with TCRA. The data include location name, latitude and longitude data, signal strength measurements, DTT operators, frequency/channels of operation, antenna height and transmission power. The transmitters' data in the study areas from Kisarawe and Makongo stations located in Coastal region and Kola Hill transmitter in Morogoro region can be found in [1]. The signal strength measurements in dBμV/m were later converted to dBm further exploitation.

3.3 Establishment of Suitable Propagation Models

We established the appropriate propagation model by calculating the deviation of the predicted value (from the model) to the measured received signal strengths. We used the spread correlated root mean square error (SCRMSE) as a benchmark for comparison in the same way it was used in [19]. The tools used were MATLAB; and we used Excel for data and results analysis. The propagation equations for Free Space, Egli and Hata can be found in [20]. The established A-M Ericsson model takes the same shape as the Ericsson model presented in [20], with $a_1 = 15.60$, a new factor obtained after regression analysis using measurement data obtained in Morogoro region.

3.4 Calculation of Distance, Elevation and TV Channels' Availability

We used the distance calculator tool obtained from [21] to calculate the distance from each measurement location in different areas in the study areas to the DTTB transmitters located in Kola Hill, Makongo and Kisarawe areas. The input was latitude and longitude values from each measurement location and the DTTB transmitter sites. TV channel availability was calculated using a model presented in [12]. In this study, the total areas for Coastal and Morogoro regions are 33,940 km^2 and 70,624 km^2, respectively [22].

4 Results and Analysis

The results focused on three areas, which are, features of the study areas, suitable propagation models and TV channels availability.

4.1 Elevations Versus Distance

To relate elevation and distance in the study area, the MATLAB code was written based upon the Shuttle Radar Topography Mission elevation data, SRTM 1 Arc-Second, obtained from the US Geological Survey (USGS) website [23]. Figure 1 shows the variation of elevation for various locations in Morogoro region where some of the data were collected. The variations in altitude between different measurement locations corroborate the mountainous nature of the place depicted in the figure. The figure also depicts high altitude of the Morogoro region with most of areas ranging between 300 and 700 m and mountainous nature of the place resulting to altitude as high as over 2000 m in some places.

4.2 Suitable Propagation Models

The appropriate propagation models were established by examining the three commonly used propagation models: Free Space, Egli, and Hata by calculating the error of deviation from the measured signal strengths. Figure 2 shows the plots of the path loss performance for different propagation models in Morogoro (left) and Coastal (right) regions. The Free Space model represents the ideal scenario when there is only 1/r power loss. The results show that the Hata model is close to the measured values compared to Egli model in Coastal region, while the A-M Ericsson model is the closest for Morogoro region.

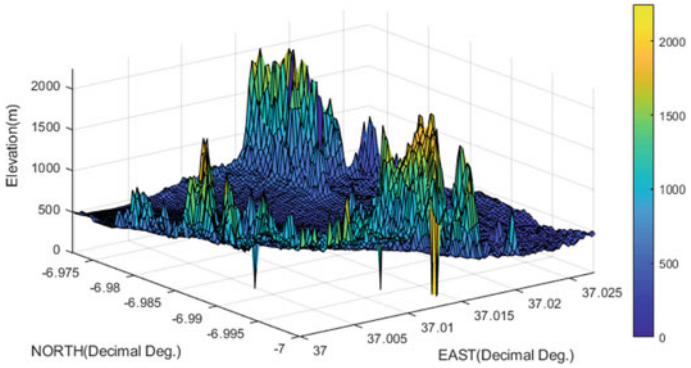


Fig. 1. Elevation from the DTTB for the part of the location of measurement sites

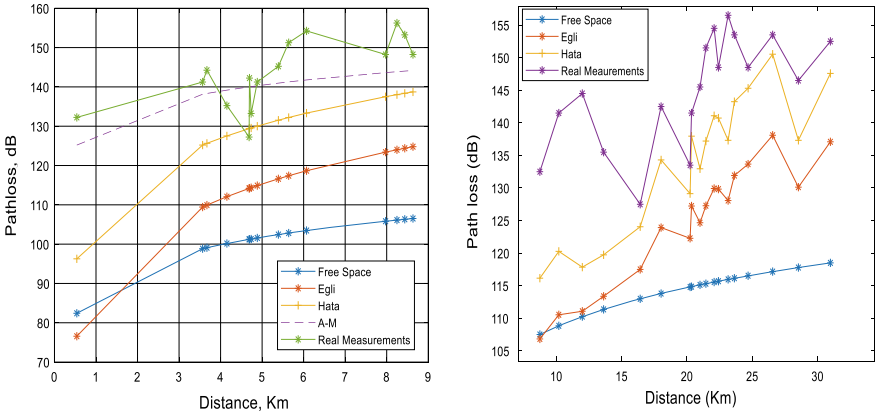


Fig. 2. Propagation models path loss performance for Morogoro (left) and Coastal (right) regions

Error Deviation. The A-M Ericsson and Hata model had the lowest values of SCRMSE at 7.62 dB and 7.86 dB for Morogoro and Coastal regions, respectively. In both cases, the free space model had—as expected—the largest values at 35.90 and 24.67 for Morogoro and Coastal regions, respectively.

4.3 Channels Availability

Figure 3 shows the percentage channel availability in Morogoro and Coastal regions. As expected channels in Morogoro region have higher availability compared to Coastal region. This can be attributed to the fact that Coastal region is a lot smaller and composed of urban and semi-urban places where channels usage is huge.

5 Discussions

The main objective of this study was to establish a suitable propagation model and to use it to estimate the TVWS in the Morogoro region in Tanzania. The results of the study show

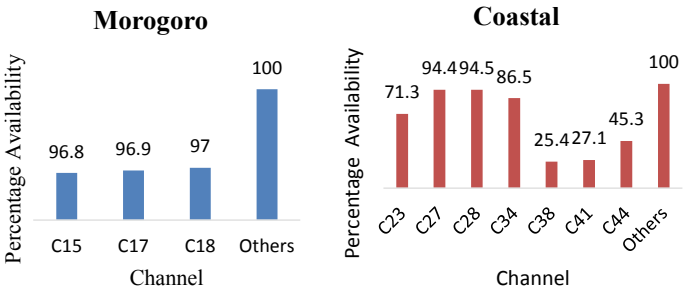


Fig. 3. Percentage channels availability for Morogoro region

that the proposed A-M Ericsson model exhibits lower values of RMSCE for Morogoro region while Hata model exhibits lower RMSCE values for Coastal region. The findings better those obtained in the studies done in [19, 24] where the Hata model outperformed others. However, while in those studies, Davidson-Hata model outperformed the Hata model for longer distances, in this study A-M Ericsson model outperforms Hata model for shorter distance within which the study was conducted. The study by Maloku et al. was conducted in a mountainous area similar to that of Morogoro region and showed that Davidson-Hata model was close to measured results for distances greater than 30 km but it was outperformed by Hata model for distances below 30 km. The A-M Ericsson model outperforms both in this study. These results corroborate the findings of previous work in [16].

The study showed that the channel availability for Morogoro and Dar es Salaam is huge—as high as 97%—despite most frequencies not being used in the areas. This indicates that, barring co-channel interference, the overall availability can be even higher. Similar studies conducted in Vietnam India, UK, USA, Singapore, Malawi, South Africa and Japan have suggested the same for developing countries [11].

6 Conclusion and Recommendations

This research identified a TVWS estimation model, which was then used to estimate the availability of the TV channels in Morogoro, Tanzania. Results show that there is approximately 97% channel availability of the three channels in use in Morogoro and a maximum of 95% availability among deployed TV channels in Coastal region. We recommend more studies to evaluate the availability of the TVWS in the whole country while taking into account more accurate propagation models.

References

1. Tanzania Communications Regulatory Authority: Quarterly Communications Statistics, https://www.tcra.go.tz/images/TelCom_Statistics_June_2019.xlsx.pdf
2. Metri, N.: Piloting “White Space” to connect the underserved of rural Tanzania, https://www.internetsociety.org/blog/2018/04/piloting-use-tv-white-space-community-networks-rural-tanzania/

3. The World Bank: Innovative Business Models for Expanding Fiber-Optic Networks and Closing the Access Gaps, http://documents.worldbank.org/curated/en/674601544534500678/pdf/132845-7-12-2018-17-20-11-InnovativeBusinessModels.pdf
4. Federal Commun. Commission: Second Memorandum Opinion and Order. Washington, D.C. (2010)
5. Zhao, Z., Vuran, M.C., Batur, D., Ekici, E.: Shades of white: impacts of population dynamics and TV viewership on available TV spectrum. IEEE Trans. Veh. Technol. **68**, 2427–2442 (2019). https://doi.org/10.1109/TVT.2019.2892867
6. Ismail, M., Kissaka, M.M., Mafole, P.: Television white space opportunities and challenges: what next for the developing countries? In: Social Implications of Technology, pp. 1–9. IEEE, Nairobi (2019)
7. Masonta, M.T., Ramoroka, T.M., Lysko, A.A.: Using TV white spaces and e-learning in South African rural schools. In: 2015 IST-Africa Conference, pp. 1–12. IIMC International Information Management Corporation Ltd. (2015)
8. ITU: Digital switchover in Africa, https://itunews.itu.int/En/5773-Digital-switchover-in-Africa.note.aspx
9. Tanzania Daily News: Tanzania Completes Switch to Digital Broadcasting, https://allafrica.com/stories/201504030144.html, (2015)
10. Tanzania Communications Regulatory Authority: Tanzania Network Facilities, https://tcra.go.tz/images/documents/licensinginformation/NETWORK_FACILITIES_updated_2.pdf
11. Ifeoluwa, O., Nassir, F., Olayiwola, W.B., Ayeni, A..: Recent trends on TV white space deployments in Africa. Niger. J. Technol. **34**, 556–563 (2015). https://doi.org/10.4314/njt.v34i3.19
12. Hessar, F., Roy, S.: Capacity considerations for secondary networks in TV white space. IEEE Trans. Mob. Comput. **1233**, 6–10 (2014). https://doi.org/10.1109/TMC.2014.2374159
13. Surajudeen-bakinde, N.T., Faruk, N., Popoola, S.I., Salman, M.A., Oloyede, A.A., Olawoyin, L.A., Calafate, C.T.: Path Loss Predictions for multi-transmitter radio propagation in VHF bands using adaptive neuro-fuzzy inference system. Eng. Sci. Technol. Int. J. **21**, 679–691 (2018). https://doi.org/10.1016/j.jestch.2018.05.013
14. Anggraeni, M.E., Dwiyanti, N.: Performance analysis of empirical path loss models for UHF TV broadcast on mountainous area. In: 2018 International Electronics Symposium on Engineering Technology and Applications (IES-ETA), pp. 247–252. IEEE (2018)
15. Pathania, P., Kumar, P., Rana, S.B.: Performance evaluation of different path loss models for broadcasting applications. Am. J. Eng. Res. **03**, 335–342 (2014)
16. Igbinosa, O.G., Okpeki, U.K.: Performance investigation of different pathloss models for a wireless communication system in Nigeria. Heliyon **5**, e01656 (2019). https://doi.org/10.1016/j.heliyon.2019.e01656
17. Naik, G., Singhal, S., Kumar, A., Karandikar, A.: Quantitative assessment of TV white space in India. In: 2014 Twentieth National Conference on Communications (NCC), pp. 1–6. IEEE (2014)
18. Hieu, D.C.: Assessment of TV White Space in Vietnam, pp. 637–640 (2014)
19. Fanan, A., Riley, N., Mehdawi, M., Ammar, M., Model, B.D.H.: Comparison of propagation models with real measurement around Hull, UK. 2016 24th Telecommun. Forum, pp. 1–4 (2016). https://doi.org/10.1109/TELFOR.2016.7818749
20. Mauwa, H., Bagula, B., Zennaro, M., Lusilao, G.: On the impact of propagation models on TV white spaces measurements in Africa. In: 2015 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), pp. 148–154. IEEE (2015)
21. Hedges, A.: Calculating Distances and Elevation, https://andrew.hedges.name/experiments/haversine/
22. City Population: Tanzania Regions Map, https://www.citypopulation.de/en/tanzania/cities/

23. USGS: Science for a Changing World, http://www.usgs.gov/default.asp
24. Maloku, H., Fazliu, Z.L., Sela, E., Ibrani, M.: Path loss model fitting for TV bands based on experimental measurements for urban environments in Kosovo. In: 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2019, pp. 480–485 (2019)

Potentials of Digital Business Models in Tourism—Qualitative Study with German Experts

Ralf Härting(✉), Raphael Kaim, Lisa Fleischer, and Alexander Sprengel

Aalen University of Applied Science, 73430 Aalen, Germany
{ralf.haerting,raphael.kaim}@hs-aalen.de,
{kmu,alexander.sprengel}@kmu-aalen.de

Abstract. Digitization is one of the most discussed topics in recent years. The traditional business models are also massively affected by digitization. Still, the impact of digitization and the potentials for business is only being analyzed in detail by few. Therefore, this research work focuses on the potential of digital business models in the tourism industry. The results are based on empirical data from German experts. The data were collected and analyzed using grounded theory. The effects of digitization, such as new technologies or changed customer needs, change the usual business models. In the future, many business models will concentrate on digital technologies. In particular, rating platforms or sharing economy platforms will play an even greater role in tourism. In addition, social networks such as Twitter, Facebook or Instagram will strongly influence the tourism industry. These new technologies will change and shape the tourism industry in the future. More and more offers are available for customers who enjoy traveling. The aim of this qualitative study is to identify the potential of new digital business models in the respective markets.

Keywords: Digital business model · Digitization · Tourism industry · Potentials · Qualitative study

1 Introduction

Today, over 20 billion devices and machines are already connected to each other via the Internet. A trend will continue to rise to half a trillion by 2030 [1]. Digitization is causing an enormous upheaval in the economy and society, in work, consumption, cooperation and communication [2]. The Internet, mobile technologies and general digitization seem to play a major role here. Travelers can plan, book and manage their holidays around the clock or even report on their holidays in the social media and rate them online. The current distribution of roles between customers and providers is undergoing a major upheaval. On the provider side, much is changing due to digitization.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_11

Although the term digitization is widely used, there is no uniform definition in the literature. Digitization has already appeared as a data conversion process from analogue to digital in the course of the third Industrial Revolution. Today, however, the understanding of the term digitization is associated with digital technologies such as big data, Internet of things, cloud and mobile computing [3]. These approaches change value creation processes, which in turn enable the emergence of new business models.

The new digital technologies and tightened competitive conditions in the tourism industry are opening up new opportunities. In the age of digitization, tour operators face the challenge of reaching their customers on all digital channels and offering them as many individual trips as possible [4]. Equally important are the effects of digitization on customers. In the age of digitization, travelers are always online, networked, mobile and well informed. Travelers not only consume their travel but also evaluate, develop, plan and shape the tourism industry through their consumer behavior. According to a study by Härting et al. (2019), every fourth trip is already booked online on the Internet. For this reason, the business models and their processes of companies are also strongly affected by digitization [4]. In the future, the Internet presence of accommodations, sharing platforms such as Airbnb or others will play an increasingly important role and will greatly change the tourism industry. The digital transformation of existing business models will play an increasingly important role because in the age of digitization, competition between companies does not only consist of processes and products but also takes place at the level of business models [5].

The Sect. 2 describes the research methodology for answering the research question: In this context, Sect. 2.1 deals with the design of the study, while Sect. 2.2 provides insights into data collection, and Sect. 2.3 conducts the analysis of the data. Section 3 shows the implications for the benefits of digital business models in the tourism industry in terms of the mentioned constructs in Fig. 1. The last section, Sect. 4, covers the conclusion of the study, which comprises limitations of the study and future research recommendations.

2 Research Design and Data Collection

2.1 Research Design

The generation of potentials of digital business models in tourism is based on the development of a qualitative study. The research method is based on the specifications of the grounded theory (GT) according to Glaser and Strauss [6]. The results are developed during research through a continuous interaction between analysis and data collection [7]. Within the framework of the GT, the study should address a general research topic rather than a fixed problem. All types of data can be used in the implementation of the GT, both qualitative and quantitative. However, the most common form of data collection is expert interviews, which were also used in this study. Although intensive literature research is not part of the GT, the questions in the interview guide are based on derived or open questions from studies already conducted in the study area. The focus is on the development of categories/influencing factors that are related to each other. The influencing factors form the theoretical basis for deriving recommendations for action and hypotheses for answering the research question [8].

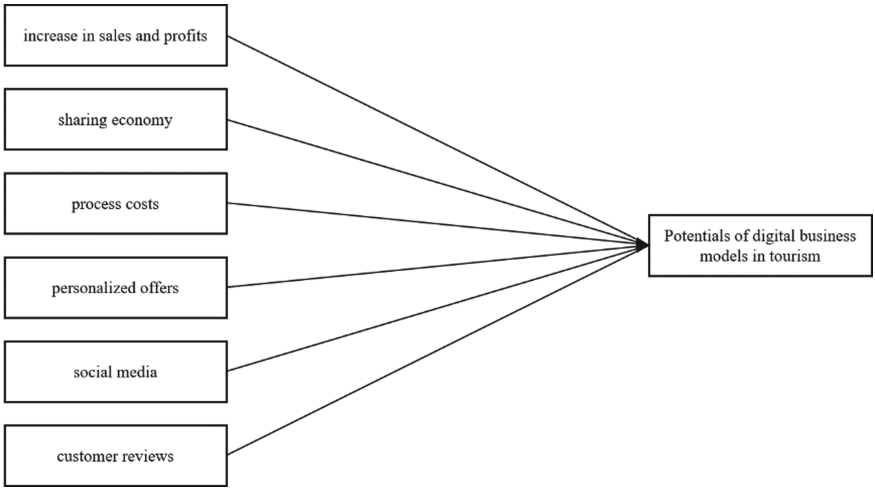


Fig. 1. Hypothesis model

2.2 Data Collection

The experts were selected on the basis of previously defined criteria. It was particularly important that all respondents come from the tourism industry and work in higher positions. A further criterion was that they were familiar with the business models in the industry. For this reason, a calling experience of at least five years in both areas, management responsibility and a company size of at least ten employees are regarded as indispensable. This was followed by the search for what research had been carried out in social networks and in the personal environment. Formal contact was established by e-mail and personal phone calls and through personal visits. In addition, social networks such as Facebook, Xing and LinkedIn were used to contact companies.

In this way, ten experts were found for the interviews (Table 1).

Total of ten experts were interviewed throughout Germany between Jan. 2019 and Feb. 2019. The interviews took place partly by telephone and in discussions directly locally. All interviews were conducted according to the same procedure. To initiate the interview, the study, its objectives and the methodical procedure were first explained.

2.3 Analysis

After completion of the data collection, the interviews were transcribed so that the data could be evaluated. The step of transcription contains a written fixation of the conversation, which contains the literal statements of the interviewees [8]. This was followed by data evaluation based on the grounded theory method according to Glaser and Strauss. The aim of this methodology is to find coherent concepts between the individual interviews and to derive categories [7]. For this purpose, the data were first openly coded, then axially coded and finally selectively coded [9]. The open coding compares the answers of the experts with regard to commonalities. The interview data are compared with each other and examined for their relationships and contexts. This leads

Table 1. Interviewed experts

Pseudonym	Job position of the experts
Anna	Travel agency manager
Franz	Director of operations
Heidi	Hotel manager
Franziska	Travel agency manager
Isabell	Travel agency manager
Frauke	Regional manager
Rainer	Hotel manager
Steffen	Travel agency manager
Herbert	Hotel manager
Hans	Hotel manager

to a first categorization [8]. Using the axial coding method, the authors subcategorized and linked related algorithms with the same property level and dimension [10].

Selective coding was performed to complete the analysis. In this phase, all categories and their indicators are selected and checked for their direct relationship to the core category. This eliminates all categories that are not related to the core category in terms of content [8]. As a result, the model shown in Fig. 1 was developed.

3 Implications for the Benefits of Digital Business Models in the Tourism Industry

The research results are structured according to the representation of the coding paradigm of the GTM. Through the analysis using the GTM, it was again examined whether the following six determinants are regarded as potentials of digital business models. These serve as basis for deriving implications for the exploitation of these potentials in the tourism industry.

The listed examples and quotations are taken from the data material of the interviews carried out. All tested determinants have a positive influence on the business models. One exception, however, is the sharing economy and the reduction of process costs.

3.1 Increase in Sales and Profits

The first positive influencing factor confirmed by the study is that the potential of digital business models leads to an increase in profit and turnover. The experts have made clear statements in this regard that strengthen this focus and about which factors lead to it. Expert Heidi, for example, says about this topic: *"I am sure that things will happen a lot faster because of the willingness of customers to make decisions. I can confirm that this will lead to an increase in sales and therefore profits."* As this example shows, the increase in profit and sales is related to the speed which technologies such as the

Internet provide. Another phenomenon that leads to more sales is that human resources can be leveraged more productively through digitization. Expert Hans stated: *"The staff create more workload than before. It is driven by the support of the new software and systems. Today we have 200 more rooms and we did not have to increase the number of employees. That's why, of course, we are increasing our sales."*

As a result of the results, it can be deduced that the increase in profit and turnover can be seen as a great potential for market growth for the tourism industry.

3.2 Sharing Economy

Sharing economy enables private individuals to offer their apartments on the market. But other sharing possibilities, such as car-sharing or bike-sharing, play also an important role in tourism and offer potential. Above all, however, they are changing the industry, as all the experts surveyed agree. Expert Franz says: *"Yes, I see a change, the change is actually there. Travelers sleep easily and cheaply."* The experts who come from the travel agency see it differently. They see it as new offers for their customers, which is clearly recognizable by the statement of expert Franziska: *"For us it is still a possibility to show the customer another alternative to the usual hotels or vacation homes."* Customer needs have changed a lot over the last few years, which also attracted the attention of experts Hans. *"The needs of customers have changed a lot in recent years. Many people like their holidays to be cozy and intimate. It should feel like home. We can't offer that in a hotel."* Naturally, this possibility of a shared flat gives more choice on the market, which has great potential. Expert Rainer has already recognized this possibility and would like to present it in his hotel. He says: *"In the future, hotels like us will have to follow the trend and see new offers as new opportunities for us. For example, we could set up a bookshelf in our hotel, where every guest can store his old, already read books, and take one he hasn't read yet."*

The following hypothesis, which concentrates on the sharing economy, can be derived: The potential of digital business models in the sharing economy sector leads to more offers in the tourism industry through cheaper pricing.

3.3 Process Costs

The next potential is the reduction of process costs. Less day-to-day business saves a lot of time and enables better communication with customers, as expert Franz confirms. *"We try to let the customers book over the homepage, so we have less work. We also try to communicate clearly via Facebook, with the guests."* Process costs can be saved by implementing many processes digitally, experts even say that personnel can be saved by digital processes. Expert Rainer even says that personnel can be saved at the reception desk. *"Customers can also check in online with us, so in theory we would no longer need a receptionist."* Expert Heidi says on this point: *"On the one hand, you can save costs, but on the other hand, these costs are incurred elsewhere. For example, through system fees for updates, for servers or for the location where the data is managed. For me it's just a shift."* As a result, it can be concluded that process costs can save a lot of time by mapping all processes digitally. The experts see the chance that personnel can be saved, since this creates more through the digital processes than before. The hypothesis in this

area is that the potential of digital business models leads to less personnel requirements due to digital processes.

3.4 Personalized Offers

Personalized offers in the tourism industry have great potential for digital business models. Expert Rainer confirms: *"Personalized offers hold great opportunities to make customers happy and to strengthen their loyalty to us. The trips can be individually customized for each customer according to their needs."* Digital technologies also play an important role in personalized offers because they make everything easier and faster. Suitable offers can be found more quickly and individualized with just a few clicks. Expert Isabell says: *"(…) that digital technologies make it very easy for us to tailor a journey to the person. It only takes a few clicks to individualize the trip for the customer. This saves a lot of time and effort."* With the topic of personalized offers, the experts also associate technologies such as big data, which play a major role. Of course, a lot of time can be saved here as well. Based on individualizing, customers can be better retained. In consequence, following hypothesis can be derived: By the potentials of the personalized offers, customers can be bound better to the enterprise.

3.5 Social Media

The surveyed experts also agree that social media plays an increasingly important role in society. Most experts are already using social media. Only one of the interviewed experts, Herbert, does not use this medium. As a reason for this, he says: *"We are not active in this area at all. Because we simply don't need it."* All other experts already use social media and are active in Facebook, Instagram and WhatsApp to communicate with their customers. In this regard, expert Franz says*: "We are already very active there. We use Facebook for virtually everything: job advertisements, advertising, general communication with guests."* The experts use social media to create offers and distribute them. People can be reached much faster via such channels. A further point that was mentioned in this excerpt is that people can see the offers from anywhere and book them. From these examples, it can be seen that social media offers many possibilities. As a result, it can be stated that platforms such as Facebook, Instagram and WhatsApp have a lot of potential and should definitely be integrated in today's business models. In addition, promotions can be launched to reach many potential customers. Important in the field of social media are influencers on Instagram, who inspire their followers with pictures and thus also advertise the places where the pictures were created. These were briefly addressed by the expert Franz. *"The influencer has also spent the night at our house, she was happy. But we also did everything to make her happy, to be in the best light of course. If something goes wrong, of course you will have a problem."* Finally following hypothesis has been derived: Social media has a positive influence on digital business by helping to spread advertisements worldwide.

3.6 Customer Reviews

All experts share the opinion that customer reviews on rating sections or portals have a great potential and encourage guests to book. Before starting a trip, many customers first

read the reviews on the various review portals, such as Booking.com and TripAdvisor. These reviews significantly influence the purchase decision. Nobody wants to book a hotel that has only bad reviews. The expert Rainer says about this phenomenon: *"You know it, even when you go on vacation, you read through the evaluations and then make your decision whether to book or not."* Another expert Franziska confirms this statement once again: *"I mean, you know the same thing, if you want to book a trip, you first have to look into the evaluations. Something with few stars we would never recommend to customers."*

But the rating platforms also help companies to improve a lot. Because only through criticism, the enterprises can improve something. This is what the expert Frauke says about it: *"You can user some reviews or what unfortunately many people do, sit it out. There are things that can be changed and others that cannot be changed."* The ratings on such platforms are nowadays seen as advertising for the company. This is what expert Heidi says: *"90% of the guests fill out the online reviews. They are treated like advertising now."* However, the experts also say that caution is needed in these reviews. People misrepresent things or, as expert Heidi says, they get paid to write positive and negative reviews themselves.

"A lot of things are misrepresented, then you find out that people who were not even guests in the hotel are writing reviews. That these possibilities are left open, that there are people, who get paid to write positive and negative reviews." Fortunately, most guests take the chance to express their opinion very seriously and give objective criticism. Expert Heidi states: *"There are people who thankfully take this very seriously and fortunately, that is the majority and they describe their experience as it is."* As result in this point, it is to be stated that the evaluation platforms have a large influence on the reservation behavior, since many humans attach importance to the opinion of the others. Booking decisions are made based on the ratings. For hotels, these evaluations serve as means of publicity. The following recommendation can be made here. Companies should integrate the rating platform directly on their homepage and not via external providers. The hypothesis in the area of rating portals is: Potentials of digital business models have a positive influence on the booking behavior.

4 Conclusion

In this study, the potentials of digital business models in tourism were developed on the basis of empirical data from German experts. The following potentials were identified: Sales and profit turnover increase, reduction of process costs, sharing economy, customer rating platforms, personalized offers and social media. Tourist providers should increasingly implement their processes digitally and take advantage of the digital technologies that exist today and profit from the opportunities. With these technologies, everything runs much faster. Communication through rating portals, social media is also much faster and easier than a few years ago. Furthermore, the experts share the opinion that digital technologies can save a lot of time and money. The providers should use social media in any case, to win customers, to find personnel or to communicate with customers. Rating platforms play an important role in today's society. Hotels should integrate this into their Web sites. It is important that companies rethink their existing business models

and transform them into digital ones. Thanks to today's technologies, the corresponding search engines can be used to look for individual trips. Notwithstanding the extensive and beneficial results that have been obtained through this work, limitations must be mentioned as well. First of all, it should be noted that only German experts were interviewed for this research. Furthermore, the risks that are often mentioned in connection with opportunities in the social media sector were not considered in the evaluation. In conclusion, there is new potential in the area of social media and personalized offers, as these serve on the one hand as free advertising media and bind customers to the company in the long term.

References

1. Bundesministerium für Wirtschaft und Energie: https://www.bmwi.de/Redaktion/DE/Dossier/digitalisierung.html. Last accessed 2019/03/16
2. Härting, R., Reichstein, C., Schad, M.: Potentials of digital business models—empirical investigation of data driven impacts in industry. In: Proceedings of the 22nd International Conference, KES-2018, vol. 126, pp. 1495–1506. Elsevier B.V. (2018)
3. Härting, R., Reichstein, C., Härtle, N.: Potentials of digitization in the tourism industry—empirical results from German experts. In: 20th International Conference on Business Information Systems, Lecture Notes, vol. 288, pp. 165–180. Springer, Berlin (2017)
4. Härting, R., Reichstein, C., Haarhoff, R., Härtle, N., Stiefl, J.: Driver to gain from digitization in the tourism industry—insights from South African tourism experts. In: Advances in Intelligent Systems and Computing, Proceedings of ICICT 2018, vol. 3, pp. 293–306. AISC Springer (2019)
5. Härting, R., Reichstein, C., Laemmle, P., Sprengel, A.: Potentials of digital business models in the retail industry—empirical results from German experts. In: Proceedings of the 23rd International Conference, KES-2019, vol. 159, pp. 1053–1062. Elsevier B.V. (2019)
6. Glaser, B., Strauss, A.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Routledge, London (2017)
7. Strauss, A., Corbin, J.: Grounded theory methodology—an overview. In: Handbook of Qualitative Research, Thousand Oaks (1994)
8. Grounded Theory Institute: http://www.groundedtheory.com/what-is-gt.aspx. Last accessed 2019/03/19
9. Stadler Elmer, S.: Mündliche Befragung. In: Empirisches wissenschaftliches Arbeiten: Ein Studienbuch für die Bildungswissenschaften, utb GmbH (2014)
10. Grounded Theory: http://www.fallarchiv.uni-kassel.de/wp-content/uploads/2011/02/huelst_grounded_theory.pdf. Last accessed 2019/03/19

An Efficient and Secure User Authentication Protocol for Smart Card Users

Javed R. Shaikh(✉)

Adama Science and Technology University, Adama, Ethiopia
javedsheikh1987@gmail.com

Abstract. There are so many communication schemes available today for the communication between authorized remote users and servers over an insecure network. These types of authentication schemes generally use the password for the authentication. To have the communication between remote users and servers, many researchers have proposed remote user authentication techniques using the smart card. The advantage of using the smart card is the storage availability and the computation speed. In many proposed protocols, authors argued that their protocol is secure and efficient against any type of attack. Unfortunately, many model fails against the off-line password guessing attack, and with these schemes, wrong password detection is not easy. Also, as many schemes use RSA cryptosystem to offer the authentication, it adds computational overhead to system which is not suitable for the smart card applications. In this paper, efficient smart card-based authentication protocol using elliptic curve cryptography (ECC) is proposed. This proposed scheme has faster computation as compared to the available schemes, and it is secure against variable attacks.

Keywords: ECC · Smart cards · Remote user authentication

1 Introduction

Due to rapid growth of Internet for everyday activities like online banking, electronic commerce (E-commerce) services, etc., and for remote host login, user wants to access these services from remote locations. While accessing the network services from the remote location, user authentication is an essential part of the security requirement for protecting the systems and the network. Many available remote authentication schemes focus on preservation of user anonymity from the eavesdropper. The authentication schemes used in E-commerce require anonymity not only to the authentication server but also to the eavesdropper [1].

Many research scholars have proposed different authentication schemes for the remote user to validate legitimate user and servers. The first novel password-based authentication scheme used for the remote access control has been proposed by the

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_12

Lamport et al. in the year 1981. Then Yang and Sheih [2] proposed a new authentication scheme for the users using the timestamp in 1999, where authors claimed that using this scheme, users can very easily select and alter their password according to their liking. Later, in the year 2003, modified Yang and Sheih scheme was proposed by Shen et al. [3] which is not affected by the forge login attack, and it offers a secure and safe authentication. But according to Liu et al. [4] and Awasthi et al. [5], Shen et al.'s scheme is not secure against forged login attack. To overcome above said problems Liu et al. [4] suggested nonce-based user authentication scheme. Then, in the year 2011, Awasthi et al. [5] presented an improved scheme for authentication where non-storage of data on the server side is taken care.

In recent times, Huang et al.'s [6] indicate that the scheme of Awasthi et al. is not secure if impersonation attack is there, and so, Huang et al. projected a scheme using the timestamp to have secure user authentication with a smart card. In this proposed model no verification information for the users is required by the remote server. However, Jung et al. [7] has performed the cryptanalysis for Huang et al.'s authentication model and found that this scheme is susceptible to the off-line password guessing attack, and during login phase, detection of the wrong password is not that easy task. Jung et al. also pointed that in Huang et al.'s scheme, during the password change phase, it is insecure to change the password of the user.

Here, with this paper, an advanced and enhanced timestamp-based authentication protocol for the remote user is proposed which overcomes all the limitations of Huang et al.'s design. The proposed scheme utilizes the ECC which is suitable for smart card applications as it needs small key size, and computation speed is also high as compared to RSA [8] which is used by the Huang et al.'s scheme.

2 Overview of Huang et al.'s Protocol

In this section, general review of Huang et al.'s authentication protocol with smart card is given. This scheme has four different phases. The details of various phases are explained below.

2.1 Initialization Phase

During this phase of Huang et al.'s scheme, the trusted authority which is responsible to generate universal parameters is the key information center (KIC). During this phase, the following steps are performed by the KIC,

(1) Generates two random large prime numbers p, q and calculates $n = p * q$.
(2) Selects two integers e and d such that $e \cdot d \bmod (p-1)(q-1) = 1$, where e *is* public key and d is private key of the system.

2.2 User Registration Phase

First time registration of user U_i to the server S is performed as explained below,

(1) U_i provides identifier which is IDi and password PW_i to KIC using a secure communication medium.
(2) After getting the IDi and PWi from the user, KIC calculates identifier of the smart card $CID i = f(ID i \oplus d)$, and hidden information $S_i = (CID_i^d \ mod \ n) \oplus f(PW_i)$.
(3) The details *of* $\{n, e, S_i, ID_i\}$ are stored into a smart card by KIC, and then, the card is given to the user U_i using a protected communication medium.

2.3 Authentication and Login Phase

The following steps will carry out when user U_i tries to login and authenticate to the server S.

(1) User U_i inputs password PW_i and finds X_i and Y_i as $X_i = S_i \oplus f\ (PW_i)$ and $Y_i = X_i^{f(ID_i, T_c)} \bmod n$, where T_c denotes user sides present timestamp.
(2) Then U_i conveys its login request to the server S by providing a messages $M = \{ID_i, n, e, T_c, Y_i\}$.

On the server side, once the message M is available at time T_s, the smart card executes the following operations,

(1) Server S checks the ID_i to find out the genuine user and verifies the condition $|T_s - T_c| < \Delta T$ using the timestamp T_s in the received message, where ΔT denotes expected the amount of delay during the transmission. If this condition is true, then the login request of the user is processed, and if the condition fails, then server rejects this request.
(2) Server S calculates $CID_i = f(ID_i \oplus d)$ and finds $Y_i^e = f(ID_i \oplus d)^{f(ID_i, T_c)} \bmod n$. If the condition is true, then the server S completes the request from a user to login, and if the condition fails, then server rejects this request.
(3) Then, server S computes $R = (f(ID_i, T_s'))^d \bmod n$ and convey $M' = \{R, T_s'\}$ to U_i, where T_s' is a server sides present timestamp .
(4) When at the user side reply message M' is received at time T_c', the U_i verifies the timestamp T_s' in the message using condition $|T_s' - T_c'| < \Delta T$, where ΔT is delay expected during transmission. With the validation of condition, the user U_i agrees to the login respond of S. If the condition fails, it ends the whole process.
(5) User U_i computes $R' = R^e \bmod n$ and verifies the condition for $R' = f(ID_i, Ts')$. If the equation is satisfied, then the user U_i accepts the server S. Otherwise, rejects it.

2.4 Password Changing Phase

The user can alter the required password easily without the concern of remote server S. During this phase, the user U_i needs to alter old password PW_i with as new password PW_i' by executing the following steps,

(1) The user U_i needs to choose a new password PW_i'.
(2) U_i calculates $S_i' = S_i \oplus f(PW_i) \oplus f(PW_i')$.
(3) Smart card replaces S_i with S_i', which means that the password change is completed successfully.

3 Security Analysis of Huang et al.'s Protocol

This section highlights the drawbacks of Huang et al.'s user authentication model which are pointed out by the Jung et al. The details of various flaws highlighted by Jung et al. are described as follows.

3.1 Off-Line Password Guessing Attack

In Huang et al.'s scheme, an attacker can get the secrets $\{n, e, S_i, \mathrm{ID}_i\}$ in the smart card if an attacker got the smart card. With this card, an attacker can easily find the request message $\{\mathrm{ID}_i, n, e, T_c, Y_i\}$ for the login between a user and the server. And then, attacker imparts the off-line password guessing attack easily. So, Huang et al.'s design is defenseless if off-line password guessing attack is imparted.

3.2 Slow in Detecting the Wrong Password

During login phase of Huang et al.'s design, when user U_i inputs ID_i and PW_i, the smart card will not validate the user's password by itself. As a result, when user U_i inputs invalid password by mistake, the authentication and login phases are still carried out until server S does the validation of password. This drawback results in the needless wastage of computation and communication costs during login phase.

3.3 Weakness in Password Change Phase

To change the password, an unauthorized user inserts U_i's smart card into its reader and then submits the details of ID_i and PW*, where PW* is unauthorized user's random new password, and finally, it requests to alter the password. During the process of altering the password, the unauthorized user gives random password PW*. The smart card then computes $S_i^* = S_i \oplus \mathrm{f}(\mathrm{PW}_i) \oplus \mathrm{f}(\mathrm{PW}^*)$, which results in $\mathrm{CID}_i\mathrm{d} \oplus \mathrm{f}(\mathrm{PW}^*)$. After these calculations, smart card replaces S_i with S_i^* without any verifying it. Therefore, changing the password is insecure in Huang et al.'s scheme.

4 Proposed Protocol

As Huang et al.'s scheme has drawbacks, a more competent and more secure protocol for the user authentication using ECC is proposed here. The proposed authentication protocol employs one-way hash function instead of a costly cryptosystem. Also, the RSA cryptosystem is replaced by the ECC. For smart cards users, the advantage of using ECC is faster computation operation as ECC has various options to choose elliptic curves which can be used for faster computation operation [9]. Table 1 shows various notations used in the protocol.

Table 1. Notationes used in the proposed protocol

Notations	Meaning
U_A	The user
S	The server
k	The security parameter
q	A large prime number
F_q	A field of prime order q
$E_q(a, b)$	A set of elliptic curve points of order n, where a; b ϵ F_q
Q	A base point of order n over $E_q(a, b)$
d_i: U_i	The private/public key pair of the entity i, where $i = A, S$ where d_i ϵ Z_q and $U_i = d_i * Q$
H ( )	One-way cryptographic hash function (e.g., MD5)
‖	The message concatenation operator
*	The elliptic curve scalar point multiplication
A	The Adversary

4.1 System Initialization Phase

In the system initialization phase, server S initializes system by selecting following parameters.

(1) Select a finite field F_q over $q > 2^{160}$
(2) Select an elliptic curve such as $E_q(a, b)$: $y^2 \bmod q = x^3 + ax + b \bmod q$ having order n over F_q,
(3) Select a point Q having order n over elliptic curve $E_q(a, b)$
(4) Publish $E_q(a, b)$ and value of Q.
(5) User A and server S select their private and public keys as (d_A: U_A) and (d_S: U_s) where $U_A = d_A * Q$ and $U_s = d_S * Q$.

4.2 Registration Phase

In this phase, the following procedure is adopted

(1) User U_A sends ID_A and PW_A to the server S.
(2) Then, server S calculates $H_A = H\ (\mathrm{PW}_A \| d_S)$ and $R_A = H_A * Q$.
(3) Server stores $\{\mathrm{ID}_A, R_A, H(.), E_q, Q\}$ on smart card (Fig. 1).

4.3 Login Phase

In this phase, the following procedure is followed by the card user to login to the server.

(1) User U_A selects current timestamp T_A and inputs identity ID_A and password PW_A.

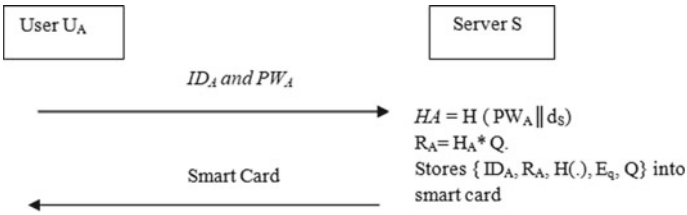


Fig. 1. User registration phase for the proposed protocol

(2) User U_A calculates $K_A = d_A * U_S$.
(3) Then, user U_A computes $C_A = H(T_A \| \mathrm{ID}_A \| K_A \| R_A)$.
(4) User conveys login request with the message M to the server as $M = \{\mathrm{ID}_A, C_A, T_A, R_A\}$.

4.4 Authentication Phase

Once the message having login request is received from the user, the server S performs the following steps to authenticate the user.

(1) Server S selects the current timestamp as T_S and checks for the condition that $(T_S - T_A) \leq \Delta T$, where ΔT is the expected time delay.
(2) Server S computes the value of $K_A = d_S * U_A$.
(3) Then, server S computes $C'_A = H(T_A \| \mathrm{ID}_A \| K_A \| R_A)$ and checks $C_A = ? \ C'_A$. If condition satisfies, then the server sends authentication pass message to the user; otherwise, it rejects the login request.
(4) Then, server S computes $C_S = H(\mathrm{ID}_A \| K_A \| T_S)$ and sends message $M = \{\mathrm{ID}_A, C_S, T_S\}$ to the U_A.

After getting authentication pass message from server, user checks for the authenticity of the server.

(5) User U_A computes $C'_S = H(\mathrm{ID}_A \| K_A \| T_S)$ and checks $C_S = ? \ C'_S$. If condition satisfies, then the server is authenticated, and the session key is established. Session key $K = H(T_S \| \mathrm{ID}_A \| K_A \| R_A)$ (Fig. 2).

To define the efficiency and communication of the proposed scheme, we have used notations to analyze the computational complexity of the proposed protocol compared to other protocols. The notation T_E refers to the time taken for one modular exponentiation, T_M denotes the time required for the computation of one modular multiplication, T_H means the time for executing the hash function and T_{PM} denotes the time complexity for executing the elliptic curve point multiplication. Table 2 shows the comparison of computation times taken by various schemes.

As shown in Table 2, the Huang et al.'s protocol is efficient than the Awasthi et al.'s protocol where each user needs to perform two modular exponentiation (T_E), and two

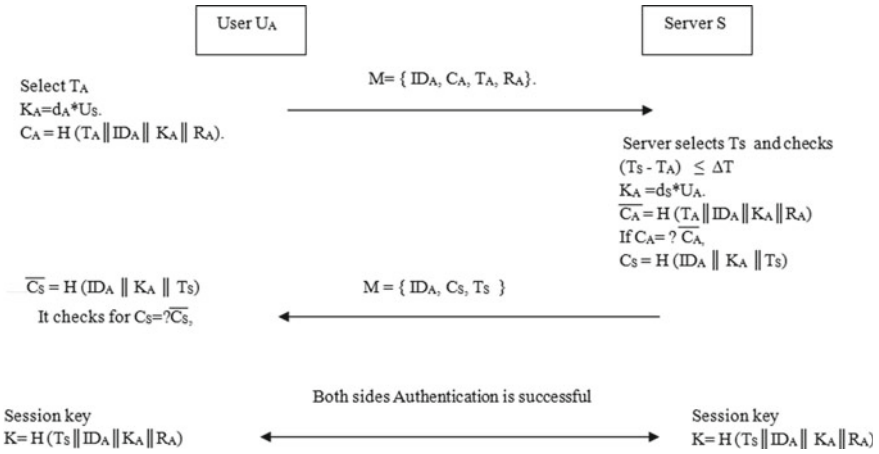


Fig. 2. Login and authentication phase of the proposed protocol

Table 2. Computational cost for various schemes

Schemes	Computations for user to complete authentication	Computations for server to complete authentication
Awasthi et al.'s	$3T_E + 3T_{\mathrm{M}} + 2T_H$	$3T_E + 1T_{\mathrm{M}} + 3T_H$
Huang et al.'s	$2T_E + 2T_H$	$3T_E + 2T_H$
Proposed scheme	$2T_{\mathrm{PM}} + 3T_H$	$3T_{\mathrm{PM}} + 4\mathrm{T_H}$

hash function computation (T_H) for the authentication purpose and for the server authentication Huang et al.'s protocol needs three modular exponentiation (T_E) and two hash function computation (T_H). In the proposed protocol, the user requires two point multiplication (T_{PM}) and three hash function computations (T_H) to achieve its authentication, and on the other side, the server needs three point multiplication (T_{PM}) and four hash function computations (T_H) for its authentication. In the proposed scheme instead of RSA cryptosystem, ECC is used, so this scheme does not require the modular multiplication and modular exponentiation. The time needed for modular multiplication and modular exponentiation is higher than the time needed for point multiplication on an elliptic curve. Therefore, the proposed authentication method is more efficient than the Huang et al.'s protocol.

Here, the values calculated by the authors in paper [10] for the different functions are considered to calculate the approximate times taken by each scheme for different computations. The computational time of a one-way hashing operation (T_H), a symmetric modular multiplication operation (T_M), modular exponentiation operation (T_E) and an elliptic curve point relative multiplication operation (T_{PM}) is 0.00032 s, 0.0056 s, 0.0192 s and 0.0171 s, respectively.

It is very clear from Fig. 3 that the total computational time required by the proposed scheme to complete authentication process is less as compared to the Awasthi et al.'s

and Huang et al.'s scheme. So, the proposed scheme is efficient and lightweight, also as in the proposed protocol ECC is used it is more secure than other schemes.

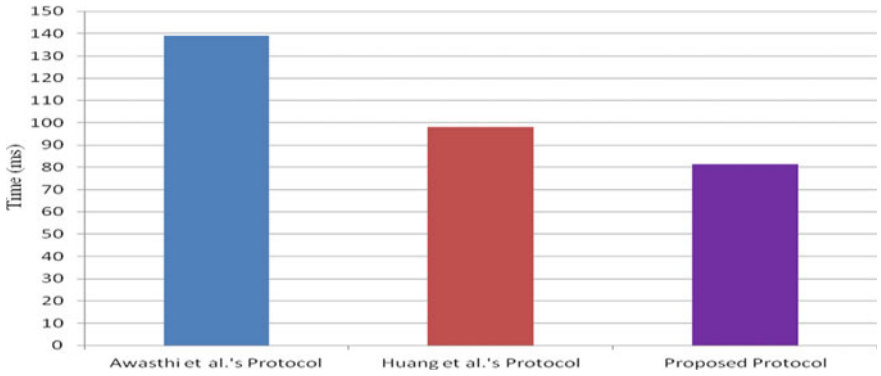


Fig. 3. Total time taken by various protocols to complete authentication process

5 Security Analysis of the Proposed Protocol

The security analysis of given protocol is demonstrated to highlight its security strength concerning various types of security attacks.

As the ECC is used in this proposed protocol, it is difficult for the attacker to find the secret stored at the server side as well as the user side. For an attacker to have access to stored secret, it is necessary to solve the elliptic curve discrete logarithm problem (ECDLP) [11]. The difficulty in solving ECDLP makes it harder to break the system secret. So, we are having the security of the system secrets.

In this authentication scheme, the secret stored on the smart card is R_A which is $R_A = H_A * Q$. Along with R_A, card has the details of ID. If attacker extracts R_A from the smart card, then attacker needs to find H_A; otherwise, it is difficult to find any information about the password. So, due to use of elliptic curve point multiplication, the user can easily achieve the security of stored data on the card.

As in the authentication protocol, only available information for the intruder is $\{ID_A, R_A, T_A, C_A\}$ and $\{ID_A, C_S, T_S\}$. It is very difficult to solve the value of password from this available information. To extract the password, intruder needs to calculate the ECDLP. As ECDLP is hard to break, so the user can secure his/her password with this scheme. Along with ECDLP, one-way hash functions are also used to secure the user password.

6 Conclusions

To provide the user authentication using smart card, Huang et al. have proposed their protocol. However, the protocol proposed by Huang et al. has some vulnerability, and the computation time required is more as it uses the RSA cryptosystem. In this paper,

an efficient and more secure user authentication protocol is proposed which uses the smart card-based user authentication using elliptic curve cryptography. Table 2 shows the computation time taken by each scheme in (ms) to complete authentication at user and server side, and it is very clear that the total computational time required by the proposed scheme to complete authentication process is less as compared to the Awasthi et al.'s and Huang et al.'s scheme. So, the proposed scheme is efficient and lightweight compared to other such schemes. The proposed user authentication design is secure against various online attacks, and it also provides security to the system secrets, stored data on the card and with detection of wrong password.

References

1. Liu, Y., Zhao, Z., Li, H., Luo, Q., Yang, Y.: An efficient remote user authentication scheme with strong anonymity. In: International Conference on Cyberworlds (2008)
2. Yang, W.W., Shieh, S.P.: Password authentication scheme with smart cards. Comput. Secur. **18**(8), 727–733 (1999)
3. Shen, J.J., Lin, C.W., Hwang, M.S.: Security enhancement for the timestamp-based password authentication. Comput. Secur. **22**(7), 591–595 (2003)
4. Liu, J.Y., Zhou, A.M., Gao, M.X.: A new mutual authentication scheme based on nonce and smart cards. Comput. Commun. **31**, 2205–2209 (2008)
5. Awasthi, K., Srivastava, K.S., Mittal, R.C.: An improved timestamp-based remote user authentication scheme. Comput. Electr. Eng. **37**, 869–874 (2011)
6. Huang, H.F., Chang, H.W., Yu, P.K.: Enhancement of timestamp-based user authentication scheme with smart card. Int. J. Netw. Secur. **16**(6), 463–467 (2013)
7. Jung, J., Choi, Y., Lee, D., Kim, J., Won, D.: Security weaknesses of a timestamp-based user authentication scheme with smart card. Int. J. Inf. Educ. Technol. **5**(7), 553 (2015)
8. Shaikh, J.R., Kumar, R., Nenova, M., Iliev, G., Singh, H.: Enhancing e-commerce security using elliptic curve cryptography. Int. J. Curr. Adv. Res. **6**(8), 5338–5342 (2017)
9. Shaikh, J.R., Nenova, M., Iliev, G., Valkova-Jarvis, Z.: Analysis of standard elliptic curves for the implementation of elliptic curve cryptography in resource constraint e-commerce applications. In: Proceedings of IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems, pp. 1–4 (2017)
10. He, D., Kumar, N., Lee, J.H., Sherratt, R.S.: Enhanced three-party security protocol for consumer USB mass storage devices. IEEE Trans. Consum. Electron. **60**(1), 39 (2014)
11. Song, R.: Advanced smart card based password authentication protocol. Comput. Stand. Interf. **32**, 321–325 (2010)

Composite Routing Approach for Vehicular Delay-Tolerant Networks

Sudhakar Pandey(✉), Sanjay Kumar, and Vishakha Chourasia

Department of Information Technology, National Institute of Technology Raipur, Raipur, Chhattisgarh, India
{spandey.it,skumar.it}@nitrr.ac.in,
vishakhachourasia23@gmail.com

Abstract. In the vehicular delay-tolerant networks, (VDTN) association between the sources to the destination is not always achievable at any required period. Consequently, the carrier node saves the message in its intrinsic buffer until an opportunity occurs for forwarding. Fix nodes facilitate in message storage and message relaying. It also helps in improving the performance of VDTN. Considering the mobility of nodes, the bit error rate is high whereas the bit error rate in fixed nodes is comparatively low. In VDTN, bit error rate is not considered in most of the routing schemes. In this article, a composite routing approach is introduced to conquer aforementioned issues. Some features of vehicular ad hoc networks (VANET) are associated to PRoPHET routing protocol for VDTN. The propagation models of VANET are executed for mobile node communication and without it for VDTN. The effect of environmental hindrance is also considered, and this can be either positive or negative. This makes the composite routing approach two-dimensional and much competent. The simulation and performance analysis of the composite approach is done via opportunistic network environment (ONE) simulator. Results show that the composite routing approach outperforms the PRoPHET *wrt*. delivery ratio and average delivery delay.

Keywords: SCF · VDTN · Bit rate

1 Introduction

Internet is the main origination of communication. It facilitates the data transfer and communication through the Transmission Control Protocol/Internet Protocol (TCP/IP) suite [1]. This sort of interprocess communication acquires symmetrical data rate and lower error rate still it cannot sustain fluctuating and long delay [2, 3]. In an intermittently connected network, the terminal path may not be accessible and network is suffered from high error rate, asymmetrical data rate, and high delay. The communication in such challenged network arouse to the fundamental of delay-tolerant networks (DTN) [4].

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_13

In past few years, the increase in handy devices like smart phones and laptops, etc., increased the demand of communication while moving [2, 3]. Vehicular ad hoc network (VANET) is a technology that provides communication among vehicles [5, 6]. VANETs cannot address the problem that arises due to unpredictable nature of mobile nodes. Vehicular delay-tolerant network (VDTN) possesses the properties of VANET and DTN as well. VDTN provides communication between vehicular nodes, the nodes might be sparse, partially and intermittently connected. In VDTN, prior terminal connectivity is not estimated but it is expected that it will be available over some time. Fix nodes, viz. access points (APs) and road side units (RSUs) enhance the performance to a significant extent [7, 8].

In VDTN, three types of communication occurs if network is inducted with fix nodes, i.e., vehicle-to-vehicle, vehicle-to-fixed node, and fixed-to-fixed node. The existing routing schemes in VDTN are one-dimensional. The bit error rate (BER) of fixed nodes and mobile nodes are always different. The existing routing approaches do not consider fixed nodes as fixed nodes because it has the same BER for transmission associating with mobile node and without mobile nodes. However, the transmission associating fixed nodes will have less BER than transmission involving mobile nodes. Also, any of the VDTN approaches do not consider environmental conditions. Situations in network are not always favorable. They may have obstacles like buildings which may restrict the line of sight and hamper the data to be delivered. These factors result in path loss and hence affect the communication, considering all these issues for structuring composite protocol and analyzing its performance through simulation.

In this paper, a composite routing protocol is proposed to conquer these issues. The remainder of the paper is structured as follows. Section 2 provides a thorough overview of the related work. Section 3 presents the proposed approach namely composite routing protocol. In Sect. 4, the performance and effectiveness of proposed approach are discussed and results are analyzed. Section 5 concludes the paper and gives future research directions.

2 Related Work

Many routing approaches have been designed, developed, and simulated by researchers to provide efficient data delivery in VDTN. One of the elementary approaches is to let the source or a relay node to carry the message to the destination. The nodes in VDTN inherit the store-carry-forward (SCF) paradigm [9]. In VDTN, bundle layer is situated over the data link layer and it integrates the incoming IP packets into bundle messages [9]. This bundle in general accumulates IP packets of similar features.

The bundle protocol does not facilitate the routing of data among the vehicular nodes [2, 3]. Routing schemes are used for this purpose, which describes how the data will be routed in the network. The contact time between encountering vehicular nodes remains too short because of frequent change in their acceleration, location, and direction. Therefore, when there appears a contact opportunity advantage should be taken to forward larger bundles of data.

There are a number of routing protocols proposed especially for VDTN. As VDTN incorporate vehicles as nodes, their acceleration, mobility pattern, location, and direction

have to be considered while designing a routing scheme. Vehicular-assisted data delivery (VADD) is one such protocol designed for VDTN [10], where the decision to forward the message is taken by exploiting the information of vehicle and road. Maximum probability (MaxProp) works by flooding the data to all the encountering nodes and once the data is received by the destination; the other replicas of the delivered message will be dropped [11]. Another VDTN routing protocol is probabilistic bundle relaying scheme (PBRS) [12], and this is an infrastructure-assisted protocol that uses road side units (RSUs). Adaptive store-carry-forward (ASCF) also works on the principle of PBRS but it conquers the disruption time problem [13]. This problem can be conquered by transferring the data to the mobile node, which remains in contact with RSU for longer duration. Another protocol called distance-aware routing with copy control (DARCC) forwards the data to the node nearer to the destination node provided the location of destination node is known [3]. A protocol that uses the encounter history of nodes to take the routing decisions is probabilistic routing protocol using history of encounters and transitivity (PRoPHET) [14].

3 Proposed Routing Scheme

Generally, VDTN routing protocols do not assume the existence of stationary nodes, rather they consider stationary node similar to mobile nodes. The data rate of stationary node is less than mobile nodes, still because of assumption data rate of mobile nodes are taken into account. This in turn enhances the bit error rate (BER) and deteriorates the delivery predictability. The situations are not always even in real scenarios. The factors that seriously affect the data rate of communicating nodes are obstacles in the line of sight, distance between communicating nodes, and other interferences. The effect of interference can be either negative or positive both. So, these issues should be considered and conquered while designing an efficient routing protocol.

Conventional routing protocols overlooked the aftermath of obstacles in the line of sight, interferences, and distance between communicating nodes. This work investigates and considered the effect of these factors for effective data delivery. Another illustration of vehicular networks is VANET. The proposed work collaborates some desirable features of VANET with the routing protocol of VDTN. The proposed scheme is then implemented in opportunistic network environment (ONE) simulator to validate its effectiveness. A number of propagation models are available in VANET like Rayleigh model, Nakagami model, log-normal shadowing, etc. [7, 8]. The communication in VDTN can be categorized into vehicle-to-vehicle, vehicle-to-stationary node, and stationary-to-stationary node. The proposed approach used PRoPHET protocol of VDTN and modified some of its features. Among all, two propagation models are implemented in one simulator, i.e., log-normal shadowing and Rayleigh model. To establish communication between stationary nodes, log- normal shadowing model is used while Rayleigh model is used for vehicle-to-vehicle and vehicle-to-stationary nodes.

The typical VDTN routing protocols are one-dimensional and hence consider all the nodes similar to mobile node. This in turn results in high BER (mobile nodes have high BER) and severely affects the overall network performance. The proposed approach uses different propagation model for different cases. These propagation models will check the

type of communicating nodes. If only stationary nodes are involved in communication, the system will switch to log-normal shadowing model. This will result in low BER because of stationary nodes. Also, if mobile nodes are involved, the Rayleigh propagation model will be used.

3.1 Algorithm

The implementation of proposed routing scheme involves the following major steps:

1. First check, whether any nodes are within the bound.
2. If nodes are found in nodes' range, check for the presence of mobile node.
3. Acquired signal power is determined for each node X using propagation model, if it is not beyond the transmission range of another node.
4. Bit energy is calculated using received signal power.
5. BER is determined using modulation technique. Here, BPSK is used.
6. After step 5, it is accumulated with communicated bits to calculate error bits in that communication.
7. Then, arbitrarily chosen bits equal to the value of total error bits are retransmitted.
8. This procedure is imitated whenever two encountering nodes exchange data.

4 Simulation Results and Analysis

The effectiveness of proposed composite routing protocol is validated using one simulator. This is a Java-based simulator having two types of simulation settings. First is simulator-based framework and second involves the exterior parameter values to be set by user, which are needed in simulation. Table 1 depicts the simulator-based settings. The simulation was conducted using Helsinki city scenario. The map of Helsinki city is available in one simulator. The simulation scenario is defined with overall 22 groups. The groups include total 88 nodes in which 78 are mobile nodes and 10 are stationary nodes (RSUs). Each group has some common and some unique features. Buffer size is assumed to be 1 GB. The speed of mobile nodes ranges to 2.8–10 m/s. Four types of mobility models are used, i.e., shortest path map-based movement, bus movement, car movement, and map-based mobility model. Rest of the assumed parameters are shown in Table 1.

The simulation was performed according to the assumed parameters in order to check the outcome of the proposed approach. The simulation results show that the proposed composite routing scheme outperforms the PRoPHET routing protocol. The analysis of results is done on the following metrics, i.e., delivery ratio, average latency, overall network overhead ratio, and average buffer time.

4.1 Delivery Ratio

Figure 1 illustrates the experimental results acquired for delivery ratio with variable message TTL. The results show that the composite routing scheme outperforms the PRoPHET protocol. It is clearly visible that delivery ratio is higher when the message

Table 1. Parameters considered for simulation

Parameters	Values
Simulation time	43200 s
Number of copies	22
Number of nodes	88 (70 cars, 10 RSUs, 8 buses)
Simulation area	4500 m × 3400 m
Routing protocols	PRoPHET and composite protocol
Buffer size	1 GB
Movement range	2.8–10 m/s
Wireless technology	Wi-Fi (IEEE 802.11p)
Mobility pattern	Car movement, bus movement, shortest path map-based movement, and map-based movement
Message lifetime	60, 120, 180, 240, 300, 360 min

TTL is less. The reason behind improved delivery ratio is the decreased BER. The growth in delivery ratio is abandoned to approximately 3 times if message TTL is made twice and by again making it twice TTL will further drop to around 45%. This is because the message engages the space in buffer for lengthy time if its TTL is increased. Therefore, new messages cannot find place in buffer and hence the delivery ratio depletes for larger values of TTL.

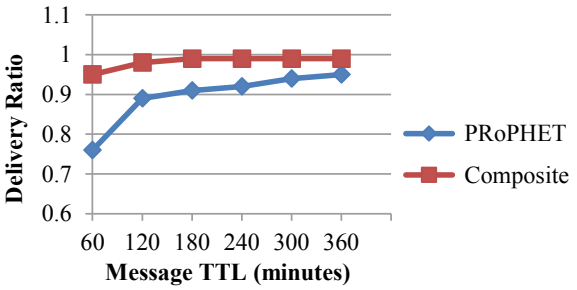


Fig. 1. Comparison of delivery ratio between PRoPHET and composite protocol

4.2 Average Latency

Figure 2 depicts the variation of average delivery latency (delivery time) with respect to message TTL. Delivery latency is basically the average time requisite for messages to get delivered to its destination in the network. In composite routing scheme, messages take lesser time to reach its destination as compared to PRoPHET. From Fig. 3, a fair reduction in the average latency of the network can be noticed. With subsequently increasing TTL,

the average latency of the network decreases. This is because the messages with less value of TTL are discarded before their delivery; on the other hand, the messages with higher TTL are delivered earlier than PRoPHET.

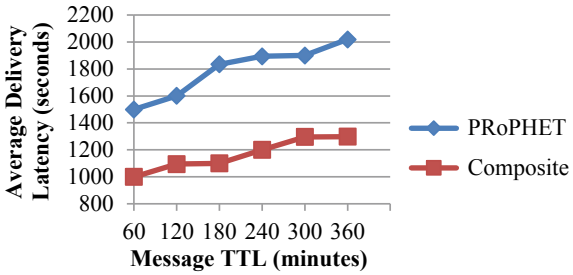


Fig. 2. Comparison of average delivery latency between PRoPHET and composite protocol

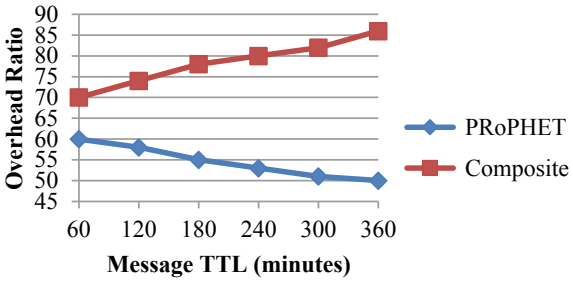


Fig. 3. Comparison of overhead ratio between PRoPHET and composite protocol

4.3 Network Overhead Ratio

As it is clearly visible from Fig. 3, the network overhead ratio of PRoPHET protocol is better than the proposed composite protocol. With the increasing TTL, overhead ratio also increases in composite scheme. This is because composite protocol considered a network with noise and obstructions in the line of sight, which are not considered in PRoPHET. Due to these assumptions, the number of retransmissions increases in the network. The messages are retransmitted due to occurrence of bit error and hence increase the count of messages in the node's buffer, which increases the overhead ratio in the network. Traditional routing approaches did not consider the BER; therefore, no retransmission was required in case of bit error. In these approaches, the messages are deleted from buffer once they reach their destination. The composite routing scheme waits for acknowledgement from receiver before deleting the messages from buffer.

4.4 Average Buffer Time

In the real-world scenario, the buffer is a scarce kind of resource available with nodes and hence it should be utilized carefully. Whenever routing schemes are unable to maintain its

buffer and free it from exaggerate messages, it is said to be not performing up to the mark. One way to utilize the buffer efficiently is to transfer the messages as soon as feasible along the destination node. From Fig. 4, this can be clearly observed that composite protocol outperforms the PRoPHET protocol. The average buffer time decreases by around 50% with the increment in message TTL.

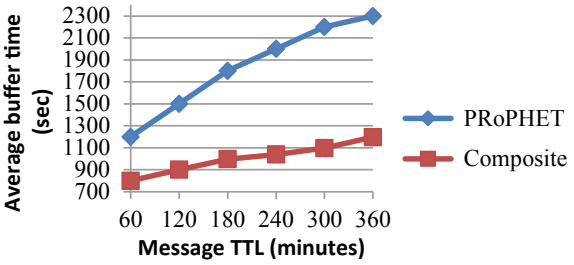


Fig. 4. Comparison of buffer time between PRoPHET and composite protocol

5 Conclusion

In VDTN, the conventional routing approaches assume the ideal network situations. The situations like uniform area and same data rate for each node were considered and do not differentiate between stationary and mobile nodes. The data to be forwarded was considered to be accurate without errors, which is not prudent. This article has proposed a composite routing scheme, in which all these problems are conquered by incorporating two propagation models. This approach uses the two distinct propagation models for transmission associating mobile nodes and stationary nodes. The BER for stationary nodes is less as compared to mobile nodes, hence two-dimensional nature of composite protocol deteriorates the total network BER. Additionally, the environmental obstructions and effect can be appended by involving its sigma and path loss value. Thus, it can be concluded that composite routing protocol outperforms the PRoPHET protocol. The delivery ratio in composite protocol is enhanced significantly along with the average delivery latency and average buffer time.

In future, further work is required to decrease the network overhead ratio. As the availability of resources is limited in the network, energy issues can be taken into consideration as the nodes are available with limited energy.

References

1. Forouzan, B.A., Fegan, S.C.: TCP/IP Protocol Suite. McGraw-Hill Higher Education (2002)
2. Benamar, N., et al.: Routing protocols in vehicular delay tolerant networks: a comprehensive survey. Comput. Commun. **48**, 141–158 (2014)
3. Ahmed, S.H., et al.: Vehicular delay tolerant network (VDTN): routing perspectives. In: 2015 12th Annual IEEE on Consumer Communications and Networking Conference (CCNC), pp. 898–903 (2015)

4. Pereira, P.R., et al.: From delay-tolerant networks to vehicular delay-tolerant networks. IEEE Commun. Surv. Tutor. **14**, 1166–1182 (2012)
5. Al-Sultan, S., et al.: A comprehensive survey on vehicular ad hoc network. J. Network Comput. Appl. **37**, 380–392 (2014)
6. Yousefi, S., et al.: Vehicular ad hoc networks (VANETs): challenges and perspectives. In: 2006 6th International Conference on ITS Telecommunications Proceedings, pp. 761–766 (2006)
7. Khairnar, V.D., Kotecha, K.: Performance of vehicle-to-vehicle communication using IEEE 802.11 p in vehicular ad-hoc network environment, arXiv preprint arXiv:1304.3357 (2013)
8. Khokhar, R.H., et al.: Realistic and efficient radio propagation model for V2X communications. In: KSII Transactions on Internet and Information Systems, pp. 1933–1954 (2013)
9. V. N. Soares*, et al.*, "A layered architecture for vehicular delay-tolerant networks," in *Computers and Communications, 2009. ISCC 2009. IEEE Symposium on*, 2009, pp. 122–127
10. Zhao, J., Cao, G.: VADD: vehicle-assisted data delivery in vehicular ad hoc networks. IEEE Trans. Veh. Technol. **57**, 1910–1922 (2008)
11. Burgess, J., et al.: MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks," in *Infocom*, 2006
12. Khabbaz, M.J., et al.: Probabilistic bundle relaying schemes in two-hop vehicular delay tolerant networks. IEEE Commun. Lett. **15**, 281–283 (2011)
13. Wu, D., et al.: Adaptive carry-store forward scheme in two-hop vehicular delay tolerant networks. IEEE Commun. Lett. **17**, 721–724 (2013)
14. Lindgren, A., et al.: Probabilistic routing in intermittently connected networks. In: International Workshop on Service Assurance with Partial and Intermittent Resources, pp. 239–254 (2004)

Emergency Detection with Environment Sound Using Deep Convolutional Neural Networks

Jivitesh Sharma(✉), Ole-Christoffer Granmo, and Morten Goodwin

Center for Artificial Intelligence Research (CAIR),
University of Agder, Grimstad, Norway
{jivitesh.sharma,ole.granmo,morten.goodwin}@uia.no

Abstract. In this paper, we propose a generic emergency detection system using only the sound produced in the environment. For this task, we employ multiple audio feature extraction techniques like the mel-frequency cepstral coefficients, gammatone frequency cepstral coefficients, constant Q-transform and chromagram. After feature extraction, a deep convolutional neural network (CNN) is used to classify an audio signal as a potential emergency situation or not. The entire model is based on our previous work that sets the new state of the art in the environment sound classification (ESC) task (Our paper is under review in the IEEE/ACM Transactions on Audio, Speech and Language Processing and also available here https://arxiv.org/abs/1908.11219.) We combine the benchmark ESC datasets: UrbanSound8K and ESC-50 (ESC-10 is a subset of ESC-50) and reduce the problem to a binary classification problem. This is done by aggregating sound classes such as sirens, fire crackling, glass breaking, gunshot as the emergency class and others as normal. Even though there are only two classes to distinguish, they are highly imbalanced. To overcome this difficulty, we introduce class weights in calculating the loss while training the model. Our model is able to achieve 99.56% emergency detection accuracy.

Keywords: Emergency management · Emergency detection · Convolutional neural networks · Signal processing · Audio processing · Deep learning

1 Introduction

One of the most crucial stages of an emergency management system is the emergency detection phase. The detection phase consists of constant monitoring of the environment via vision, sound or any other sensor-based system and reporting any abnormal findings immediately. It must be accurate with negligible false

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_14

alarm rates and capable of detecting such events as soon as they occur. This is what makes designing emergency detection systems, a very difficult task. And, its capabilities and weaknesses affect other phases of the emergency management system.

Different types of emergencies require different kinds of sensory detection systems. However, some detection systems are applicable to several types of emergencies. For example, vision-based systems can be used to detect several kinds of emergencies, especially fire [1,2], flood [3], earthquakes [4], droughts [5], avalanches [6], etc. A thorough analysis of vision-based techniques to detect emergency situations can be found in [7].

There are also a few specialist detectors that detect specific types of emergencies that cannot be detected by conventional means. In [8], volatile chemical compounds are identified using a method based on portable gas chromatograph–mass spectrometer. Another specialist detector was proposed in [9] for human activity recognition. It used Bayesian networks and rules-based stochastic context-free grammars to detect abnormal behaviors among monitored humans. Another technique proposed in [10] that monitors human activities was used for fall detection. It uses sensory data like acceleration and Euler angle with monitored parameters such as heart rate and activity of daily lives (ADLs).

A convolutional neural network (CNN) was used in [11] to detect emergency weather situations by processing data from climate datasets. The method demonstrated the ability of deep neural networks to learn complex weather patterns to predict extreme weather conditions. Along with Bayesian hyperparameter optimization, the weather model was able to achieve high performance. A very important work regarding the analysis of reliability of multi-state emergency detection systems was performed in [12]. Wireless sensor network-based approaches to detect fire emergencies have also been popular in recent years [13–15]. The general idea uses a distributed system of nodes, where each node is capable of classifying fire emergencies. Once the fire is detected, the cluster head is alerted and routed to other cluster heads via gateways to alert the firefighters. Each node consists of a fire detection algorithm and sometimes a neural network that has been trained on historical fire data.

There have been many other noteworthy contributions in the field of emergency detection. In this paper, we focus on detecting emergency situations using audio sensory methods. We propose a general-purpose emergency detector that classifies environmental sounds. The system is emergency-type agnostic, so it can be used to detect any kind of crisis using audio signals from the environment only. To the best of our knowledge, this is the first general-purpose emergency sound detection system.

A research work similar to this paper was proposed in [16]. They propose a perception sensor network that detects screaming persons. It also uses audio-visual information for sound source localization. However, unlike our paper, it does not generalize to all major emergency situations and does not employ deep learning. Also, our method uses auditory information only. Also, in [17], a core signal processing approach of using mechanical resonant filters is used to detect

sirens. A siren detection for system ambulance using fast fourier transform was proposed in [18]. Another emergency siren and horn detection system was proposed in [19]. They propose to treat the audio spectrograms as an image and apply semantic segmentation using the U-net [20] to detect and de-noise sirens in traffic. They use another CNN to localize the siren sounds, which makes this system very computationally complex.

Unlike these methods, our proposed system detects not only detects sirens and screams, but also other potential dangerous environmental sounds. We test our system on three benchmark environment sound classification datasets, separately as well as combined, to perform binary classification between emergency and non-emergency sounds. This makes our results credible and tests our method thoroughly. We use multiple audio/signal processing techniques to extract features from audio signals and then use a CNN with separable convolutions to process the time and feature domain information separately. These aspects add up to achieve emergency sound detection accuracy of 99.56%.

The rest of the paper has been organized as follows: Sect. 2 and subsequent sub-sections explain our proposed emergency sound detection system. Section 3 explains our experiments and shows the results of our system on the datasets. Finally, Sect. 4 concludes our work.

2 Proposed Method

For the task of emergency sound detection, we use our previously proposed environmental sound classification (ESC) model that sets the new state-of-the-art results on the all three benchmark ESC datasets[1] [21]. We adjust the model according to the problem at hand. We reduce the size of the model considering the task to solve.

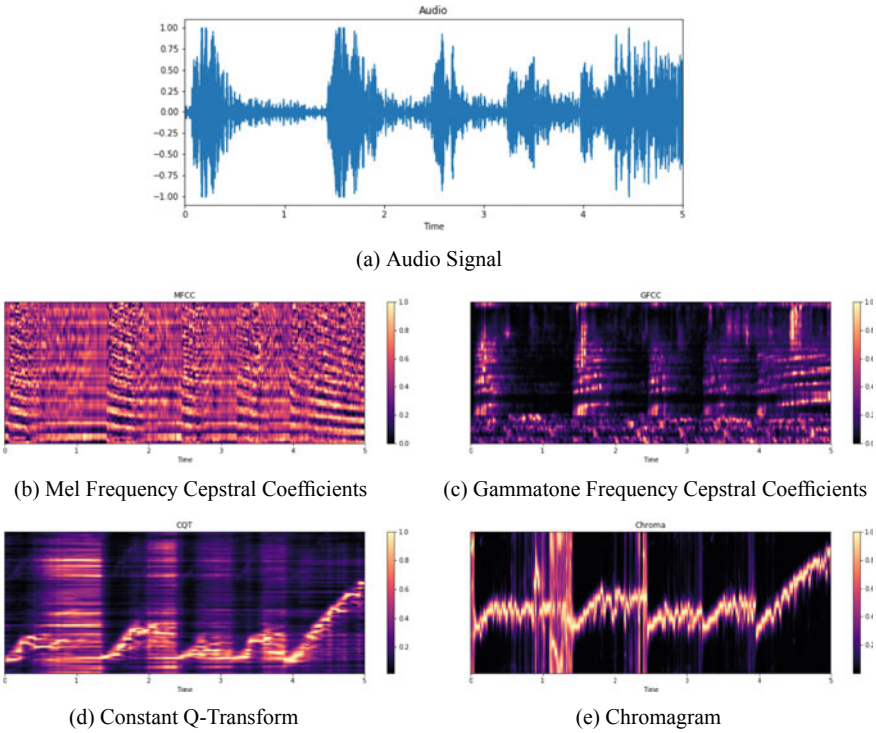
Our model uses multiple feature extraction techniques to create a multi-channel input for the deep CNN. Each feature provides complimentary information to the network. Processing this information is carried out by using separable convolutions that work on the time and feature domain individually. This is advantageous because it reduces redundancy and improves information usage efficiency.

We apply our model to the problem of emergency sound detection. We treat the task as a binary sound classification problem. We explain the feature extraction stage and the classification stage in the subsequent sub-sections.

2.1 Multiple Feature Channels

Using multiple features to process and classify audio signals has been advocated [22–24] to achieve higher performance for both ASR and ESC tasks. However, in our model, unlike the previously proposed feature aggregations, we use four major feature extraction techniques, namely MFCC, GFCC, CQT and chromagram. Incorporating different features provides the deep CNN classifier with

[1] Our paper is under review and also available here https://arxiv.org/abs/1908.11219.



(a) Audio Signal

(b) Mel Frequency Cepstral Coefficients

(c) Gammatone Frequency Cepstral Coefficients

(d) Constant Q-Transform

(e) Chromagram

Fig. 1. Multiple feature channels of an audio file labeled as an emergency sound class (siren)

more distinguishable characteristics and complementary feature representations to accurately classify audio signals. We briefly describe the feature extraction techniques used in our paper:

MFCC: The mel-frequency cepstral coefficient (MFCC) has been a standard feature extraction technique and has been successfully used to benchmark ASR as well as ESC models [25]. The development of MFCC was propelled by human auditory perception. It differs from other cepstral features in the frequency bands which are on the mel-scale. We use 64 bands in the mel-filter bank to produce 64-dimensional features using standard hamming window size of 1024 and hop length of 512. Since MFCC is susceptible to noise, we normalize it between 0 and 1.

GFCC: The gammatone frequency cepstral coefficient (GFCC) has also been a popular choice of feature extraction in ESC and ASR tasks [26]. The gammatone filter is a linear filter that is outlined by an impulse response which is a product of a gamma distribution and sinusoidal tone. It is especially advantageous to use GFCC with MFCC as they complement each other, due to the capability

of GFCC being able to proficiently characterize transient sounds classes such as footsteps and gunshots [24]. We use 64 bands gammatone filters with standard hamming window size of 1024 and hop length of 512 to produce 128-dimensional features.

CQT: The constant Q-transform is a time–frequency analysis technique that is particularly suitable for music audio signals [27,28]. It is essentially a Gabor wavelet transform and according to the results shown in [29], illustrated CQT's ability to capture low-to-mid level frequencies better than MFCC for audio scene classification, which is essentially the same task as ESC. We set the number of bands per octave to 64 and window size of 1024 with a hop length of 512, so we get feature vectors of equal dimensionality as MFCC and GFCC.

Chromagram: Another feature extraction technique that is popular with music audio signals is the chromagram [30]. Chroma features can be used to distinguish among audio signals by assigning them pitch class profiles. This makes chromagrams particularly proficient in audio structure analysis [31]. We use Short-time Fourier Transform (STFT) to compute chroma features. The number of chroma features extracted from the audio waveform is set to 64 with window size of 1024 and hop length of 512.

All these features, 64 dimensional each, are stacked on top of each other to create a four-channel input for the deep CNN classifier. Each feature plays an important role in distinguishing among emergency and non-emergency sounds. An example of the pattern of features extracted by these methods is shown in Fig. 1. The audio signal used for this demonstration belongs to the emergency sound class (Siren) and is shown in Fig. 1a. All the features have been normalized between 0 and 1 for easy visualization. These features behave in a complimentary manner as some have higher values for the spikes in signal amplitude, while some have lower values (magnitude of values can be determined by the color-bar on the right of the figures).

2.2 Deep Convolutional Neural Network

We employ the deep convolutional neural network as the emergency sound classifier for this task. The architecture of our deep CNN for emergency sound classification used in this paper is shown in Table 1. It is a smaller version of our state-of-the-art ESC model [21] and consists of five repetitions of *Conv2D-Conv2D-MaxPool-BatchNorm* with different numbers of kernels and kernel sizes. From Table 1, we can see that almost all convolutional layers are made up of depth-wise separable convolutions. However, unlike depth-wise separable convolutions where an $1 \times m$ kernel is followed by an $m \times 1$ kernel, we use two consecutive $1 \times m$ kernels followed by two $n \times 1$ kernels, where $n \neq m$. All strides are equal to 1.

These types of separable convolutions work really well for audio signals. In images, both dimensions represent the same type of information, i.e., pixels,

Table 1. Deep CNN architecture

Layer type	Kernel size	Pooling size	No. of kernels/neurons
Conv2D	1×3	–	16
Conv2D	1×3	–	16
BatchNorm	–	–	–
MaxPool2D	–	1×2	–
Conv2D	7×1	–	16
Conv2D	7×1	–	16
BatchNorm	–	–	–
MaxPool2D	–	4×1	–
Conv2D	1×3	–	32
Conv2D	1×3	–	32
BatchNorm	–	–	–
MaxPool2D	–	1×2	–
Conv2D	7×1	–	32
Conv2D	7×1	–	32
BatchNorm	–	–	–
MaxPool2D	–	4×1	–
Conv2D	7×3	–	64
Conv2D	7×3	–	64
BatchNorm	–	–	–
MaxPool2D	–	4×2	–
Flatten	–	–	–
Dense	–	–	256
Dense	–	–	2

whereas, in case of audio features, one dimension represents features and the other represents time. So, each type of information much be processed individually. The details of the architecture and working of the separable convolutions can be found in [21]. We use the leaky rectified linear unit (Leaky ReLU) activation function after each layer [32], as given by Eq. 1.

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ 0.01x, & \text{otherwise} \end{cases} \tag{1}$$

The last set of *Conv2D-Conv2D-MaxPool-BatchNorm* consists of $n \times m$ convolutions and pooling operations which is used to aggregate all information from both dimensions. We use batch normalization layers after every couple of convolutional layers. These layers normalize the input to the next layers in order to reduce internal covariate shift which refers to the phenomenon that occurs dur-

ing training when the input distribution of each layer changes due to the changes in the parameters of the previous layers. This requires lowering learning rates which slows down training. Hence, batch normalization is now an indispensable part of a CNN architecture.

The output layer is the standard softmax layer to obtain class probabilities. It consists of two neurons since the emergency sound detection is a binary classification problem. And the binary cross-entropy loss function is used along with the softmax layer. We use the Adam optimizer [33] with Nestorov momentum to train the deep CNN.

In order to avoid overfitting, we use Dropout with ratio 0.5 after each dense layer. To further improve the generalization performance of our model, L_2 regularization was used on the weights of the dense layers with regularization parameter $\lambda = 0.1$. Along with this, we also use class weights during the calculation of loss since the number of instances of emergency sounds is far less than non-emergency sounds. So, we weigh class instances by $\frac{1}{N_C}$, where N_C is the number of instances of class C.

3 Experiments and Results

We use our model, mentioned in the previous section, to detect emergency sounds. To test our model for this task, we consider two benchmark environmental sound datasets, the ESC-50 [34] and UrbanSound8K [35]. We combine these datasets into a single big dataset consisting of 10,732 audio files in total. Then, divide the classes into two major categories: Emergency and Non-emergency sounds. Classes that fall under emergency sounds are sirens, fire crackling, glass breaking, gunshot and thunderstorm. All the other sounds are kept under non-emergency sounds.

We use Tensorflow [36] and Keras [37] to implement our model and use Librosa [38], Essentia [39] and MATLAB Signal Processing Toolbox [40] to extract the audio signal features. For feature and model analysis on the two datasets and the experimental results, please refer to [21]. Here, we use our model with four-channel feature input to detect emergency sounds. The loss is calculated using class weights for each training instance. Class weights are not used for testing. We use k-fold cross-validation, with $k = 5$ and present the average test accuracy of the folds as the final performance metric.

Table 2. Effect of class weights

Model	Class weights	Accuracy
DCNN-5	No	94.09
DCNN-5	Yes	**99.56**

Bold indicate best performance

Table 2 shows the effect of class weights on the performance of the model. Since the dataset is highly imbalanced, due to the task at hand, using class

weights is important. The dataset consists of just 910 sound files that fall under the emergency sounds category, which is just 8.5% of the total audio duration. Due to this, even though the accuracy of the model without class weights is above 90%, still it does not perform well on this task. It overfits on the non-emergency class which has 91.5% of the total instances.

The naming of the model is based on the number of repetitions of the *Conv2D-Conv2D-MaxPool-BatchNorm* layers. Note, however, that the naming of the model in this paper should not be confused with the model in [21] as it consists of more kernels and larger dense layers.

We also test different architectures based on size of the deep CNN as shown in Table 3. All the models in table are trained with class weights. DCNN-6 and DCNN-7 tend to overfit the training data, and DCNN-3 and DCNN-4 do not contain enough layers/parameters to learn the data patterns. Therefore, DCNN-5 is the best performing architecture. Hence, our model with five repetitions of *Conv2D-Conv2D-MaxPool-BatchNorm* and four-channel input consisting of MFCC, GFCC, CQT and chromagram features is able to achieve 99.56% accuracy for the task of emergency sound detection on the combination of ESC-50 and UrbanSound8K datasets.

Table 3. Deep CNN architecture performance

Model	No. of layers	Accuracy
DCNN-3	15	92.49
DCNN-4	19	95.25
DCNN-5	23	**99.56**
DCNN-6	27	99.25
DCNN-7	31	98.45

Bold indicate best performance

4 Conclusion

In this paper, we propose an emergency sound detection system using multi-channel feature input and a deep CNN classifier. We use the reduced version of our model that sets the new state-of-the-art for environment sound classification.[2] We use feature extraction techniques like MFCC, GFCC, CQT and chromagram to create a four-channel input. Each feature contributes to the performance of the model by providing specific information according to its capabilities. We use separable convolutions and pooling operations in our deep CNN network to process feature and time domain information separately. We test our model by combining the ESC-50 and UrbanSound8K benchmark datasets and categorizing emergency sound-related classes in one category and the rest

[2] Our paper is under review and also available here https://arxiv.org/abs/1908.11219

in another category to make a binary classification problem. We also use class weights since the data is quite imbalanced. Our model achieves 99.56% accuracy on the combination of the ESC-50 and UrbanSound8K datasets.

References

1. Sharma, J., Granmo, O.-C., Goodwin, M., Fidje, J.T.: Deep convolutional neural networks for fire detection in images. In: Engineering Applications of Neural Networks, pp. 183–193. Springer, Cham (2017)
2. Sharma, J., Granmo, O.-C., Goodwin, M.: Deep CNN-ELM hybrid models for fire detection in images. In: Artificial Neural Networks and Machine Learning—ICANN 2018, pp. 245–259. Springer, Cham (2018)
3. Narayanan, R., Lekshmy, V.M., Rao, S., Sasidhar, K.: A novel approach to urban flood monitoring using computer vision. In: Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1–7, July 2014
4. Perol, T., Gharbi, M., Denolle, M.: Convolutional neural network for earthquake detection and location. Sci. Adv. **4**(2) (2018)
5. Srivastava, S., Bhugra, S., Lall, B., Chaudhury, S.: Drought stress classification using 3d plant models. CoRR. abs/1709.09496 (2017)
6. Waldeland, A.U., Reksten, J.-H., Salberg., A.-B.: Avalanche detection in SAR images using deep learning. In: IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 2386–2389 (2018)
7. Lopez-Fuentes, L., van de Weijer, J., Hidalgo, M.G., Skinnemoen, H., Bagdanov, A.D.: Review on computer vision techniques in emergency situation. CoRR. abs/1708.07455 (2017)
8. Fair, J.D., Bailey, W.F., Felty, R.A., Gifford, A.E., Shultes, B., Volles, L.-H.: Quantitation by portable gas chromatography: mass spectrometry of VOCs associated with vapor intrusion. Int. J. Anal. Chem. (2010)
9. Mocanu, I., Florea, A.M.: A model for activity recognition and emergency detection in smart environments. In: The First International Conference on Ambient Computing, Applications, Services and Technologies, pp. 23–29 (2011)
10. Mao, A., Ma, X., He, Y., Luo, J.: Highly portable, sensor-based system for human fall monitoring. Sensors **17**(9), 2096 (2017)
11. Liu, Y., Racah, E., Prabhat, Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M.F., Collins, W.D.: Application of deep convolutional neural networks for detecting extreme weather in climate datasets. CoRR. abs/1605.01156 (2016)
12. Nadjafi, M., Farsi, M.A., Jabbari, H.: Reliability analysis of multi-state emergency detection system using simulation approach based on fuzzy failure rate. Int. J. Syst. Assur. Eng. Manage. **8**(3), 532–541 (2017)
13. Dener, M., Ozkok, Y., Bostancioglu, C.: Fire detection systems in wireless sensor networks. World Conference on Technology, Innovation and Entrepreneurship. Procedia—Social and Behavioral Sciences, vol. 195, pp. 1846–1850 (2015)
14. Zhang, J., Li, W., Yin, Z., Liu, S., Guo, X.: Forest fire detection system based on wireless sensor network. In: 2009 4th IEEE Conference on Industrial Electronics and Applications, pp. 520–523, May 2009
15. Khadivi, A., Hasler, M.: Fire detection and localization using wireless sensor networks. In: Komninos, N. (ed.) Sensor Applications, Experimentation, and Logistics, pp. 16–26. Berlin, Heidelberg. Springer, Berlin, Heidelberg (2010)

16. Nguyen, O., Yun, S., Choi, J.: Detection of audio-based emergency situations using perception sensor network. In: 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 763–766, Aug 2016
17. Fragoulis, D.K., Avaritsiotis, J.: A siren detection system based on mechanical resonant filters. Sensors **1**, 09 (2001)
18. Miyazaki, T., Kitazono, Y., Shimakawa, M.: Ambulance siren detector using FFT on DSPIC. In: ICIS 2013 (2013)
19. Marchegiani, L., Newman, P.: Listening for sirens: locating and classifying acoustic alarms in city scenes. CoRR. abs/1810.04989 (2018)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. CoRR. abs/1505.04597 (2015)
21. Sharma, J., Granmo, O.-C., Goodwin, M.: Environment sound classification using multiple feature channels and deep convolutional neural networks. arXiv:1908.11219 (2019)
22. Zhang, Z., Xu, S., Cao, S., Zhang, S.: Deep convolutional neural network with mixup for environmental sound classification. In: Lai, J.-H., Liu, C.-L., Chen, X., Zhou, J., Tan, T., Zheng, N., Zha, H. (eds.) Pattern Recognition and Computer Vision, pp. 356–367. Springer, Cham, (2018)
23. Yu, S., Zhang, K., Wang, J., Madani, K.: Environment sound classification using a two-stream CNN based on decision-level fusion. Sensors **19**(7) (2019)
24. Chachada, S., Kuo, C.J.: Environmental sound recognition: a survey. In: 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, pp. 1–9, Oct 2013
25. Sahidullah, M., Saha, G.: Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. Speech Commun. **54**(4), 543–565 (2012)
26. Shao, Y., Jin, Z., Wang, D., Srinivasan, S.: An auditory-based feature for robust speech recognition. In: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4625–4628, April 2009
27. Schörkhuber, C.: Constant-Q transform toolbox for music processing (2010)
28. Huang, S., Li, Q., Anil, C., Bao, X., Oore, S., Grosse, R.B.: Timbretron: a wavenet (cyclegan (cqt (audio))) pipeline for musical timbre transfer. CoRR. abs/1811.09620 (2018)
29. Lidy, T., Schindler, A.: Cqt-based convolutional neural networks for audio scene classification. In: Proceedings of the Detection and Classification of Acoustic Scenes and Events 2016 Workshop (DCASE2016), vol. 90, pp. 1032–1048. DCASE 2016 Challenge (2016)
30. Shepard, R.N.: Circularity in judgments of relative pitch. J. Acoust. Soc. Am. **36**(12), 2346–2353 (1964)
31. Paulus, J., Müller, M., Klapuri, A.: State of the art report: audio-based music structure analysis. In:International Society for Music Information Retrieval, pp. 625–636, Aug 2010
32. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. CoRR. abs/1505.00853 (2015)
33. Kingma, D.P., Adam, L.J.B.: A method for stochastic optimization. In: ICLR, International Conference on Learning Representations (ICLR), p. 13, San Diego, CA, USA, 7–9 May 2015. Ithaca, NY. arXiv.org
34. Piczak, K.J.: ESC: dataset for environmental sound classification. In: Proceedings of the 23rd Annual ACM Conference on Multimedia, pp. 1015–1018. ACM Press

35. Salamon, J., Jacoby, C., Bello, J.P.: A dataset and taxonomy for urban sound research. In: Proceedings of the 22nd ACM International Conference on Multimedia, MM'14, pp. 1041–1044. ACM, New York, NY, USA (2014)
36. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I, Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from www.tensorflow.org (2015)
37. Chollet, F. et al.: Keras. https://keras.io (2015)
38. McFee, B., Raffel, C., Liang, D., Ellis, D.P.W., McVicar, M., Battenberg, E., Nieto, O.: Librosa: audio and music signal analysis in python (2015)
39. Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J.R., Serra, X.: Essentia: an audio analysis library for music information retrieval. In: International Society for Music Information Retrieval Conference (ISMIR'13), pp. 493–498, Curitiba, Brazil, 04 Nov 2013
40. MATLAB Signal Processing Toolbox 2019. The MathWorks Inc., Natick, MA, USA (2019)

The Application of Finite Element Method to Analyze Passive Failure and Deformation Mechanisms of Soil in Front of Tunnel Face

Nguyen Anh Tuan(✉)

Faculty of Transportation Engineering, Hochiminh City University of Transport, Ho Chi Minh City, Vietnam
tuanna@ut.edu.vn

Abstract. In the last many decades, replying to the urgent needs for the infrastructure construction, shield tunneling has been conducted widely in urban areas. Although the technique has been advanced recently owing to the accumulation of practical experiences and the progress of mechanization, there still remains many unknown problems especially in soft ground tunneling. For tunneling, one must always consider not only the stability of a tunnel itself but also the surface settlement due to deformations of soil around the tunnel. Although the patterns of the surface settlements differ for different soil conditions and methods of tunneling, many field observations and model tests show that the troughs of surface settlement can be approximated by the error functions or Gaussian normal distribution curves. This article concentrates on analyzing passive failure and deformation mechanisms of the soil in front of tunnel face due to tunneling.

Keywords: FEM · Passive failure · Pressure · Deformation · Tunnel face

1 Introduction

Finite element method (FEM) has now been improved and widely used to solve complex elasticity, elasticity–plasticity, and adhesion–plasticity problems. Its advantages due to discontinuity and heterogeneity of stratigraphic structure can handle complex boundary conditions to calculate the stress–strain value and their distribution. Based on that, it is used to analyze the passive failure mechanisms [1].

Once finding out, numerical analysis was a design tool though it was often criticized. However, thanks to the development of information technology, a revolution on the field of underground engineering has taken place so far. From that, tunnel works are calculated with complete numerical analysis.

The convenience of numerical analysis method has been proved. Both material behaviors and boundary conditions have been included in the calculation and parameters study to improve for the tunnel design in order to make it easier.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_15

Currently, there are several available programs for calculating underground structures on the world such as: SAGE CRISP developed by the Crisp Consortium Ltd; PLAXIS developed by PLAXIS BV Ltd.; SIGMA module in GEOSLOPE program developed by GEOSLOPE-International Ltd.; and MISES program according to NATM. These programs all use FEM to calculate the problems of underground works. These tools are quite effective for consultancy agencies to evaluate design and conduct research on underground works as well.

This paper applies FEM, which simulates on PLAXIS program and aims to analyze the passive failure and deformation mechanisms of soil in front of tunnel face during the process of constructing [2–4].

2 Establishing the Numerical Modeling

Table 1 shows the soil parameters used in the simulation by Plaxis 2010 program.

Table 1. Soil parameters

Soil parameters	Unit	First layer sand	Second layer clay
Sand thickness	mm	325	180
Saturated Unit weight γ_{sat}	kN/m^3	20.3	21.1
Elastic modulus E_{50}	kN/m^2	27,000	100,000
Elastic modulus E_{ur}	kN/m^2	81,000	300,000
Elastic modulus E_{oed}	kN/m^2	27,000	100,000
Poisson's ratio v		0.3	0.3
Cohesion c	kN/m^2	1	300
Angle of friction φ	Degree	30	1
Angle of dilation ψ	Degree	0	0
m		0.5	1.0
Failure ratio R_f		0.9	0.9
K_0		0.5	–

Figures 1 and 2 show models for tunnels with $C/D = 1.5$ and $C/D = 3.3$ in saturated sand.

3 Numerical Model Procedures

Each analysis is started by applying an acceleration of 100 g. This is achieved by increasing the soil's specific weight 100 times [5]. The analyses were carried out by assuming that the deformation of the soil increases by g times in one-dimension, the horizontal coefficient of effective vertical stress is K_0. Thus, the initial stress state of soil layers

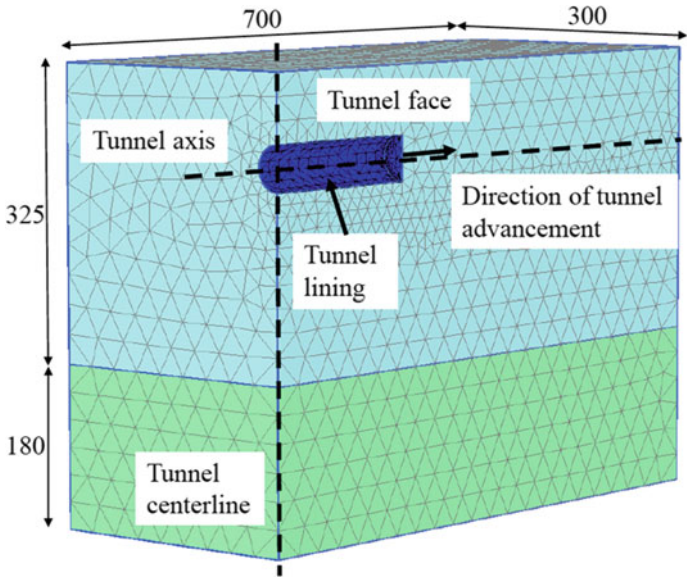


Fig. 1. Tunnel at the position of $C/D = 1.5$

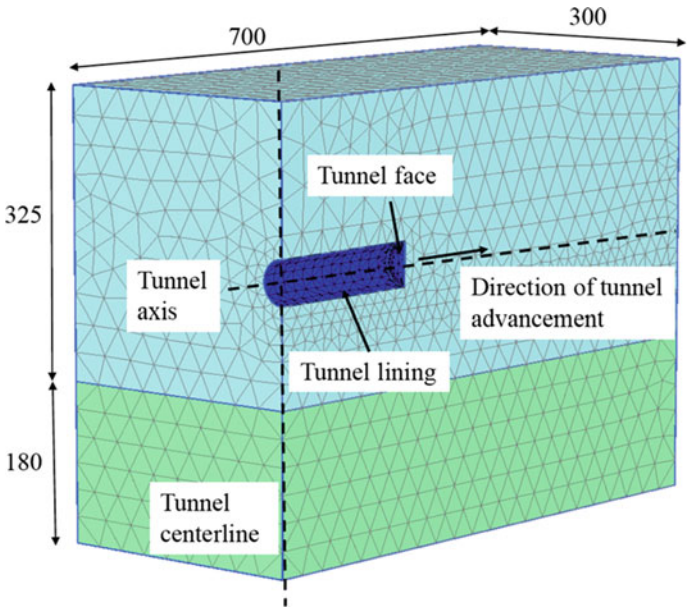


Fig. 2. Tunnel at the position of $C/D = 3.3$

at 100 g acceleration can be determined. For tunnels in sand, this represents a condition when equilibrium is reached. In centrifugal model experiments, the presence of the

model tunnel will redistribute stress around tunnel. Besides, an ideal condition with no side friction is also simulated [6, 7].

The tunnel is assumed to be placed on-site and submerged in water. Excavation is simulated by disabling soil elements in the tunnel area and activating the sheet elements of the tunnel cover in the same time. A constant pressure, which is equal to the static soil pressure at the center of the tunnel, is applied to the tunnel. It should be noted that the tunnel moves away from the soil by g times so K_0 cannot be reached immediately before the passive failure experiment.

A margin to control pressure was placed in the tunnel to determine the passive failure pressure, similar to tunnel construction with shields in practice. For tunnel in sand, drainage is analyzed. In other words, the slow speed is built up to entirely dissipate the excess pore water pressure.

4 Result Analysis

4.1 Failure Mechanism

Figures 3a–d and 4a–d show us the vector of ground displacement in the position of $C/D = 1.5$ and 3.3. From the simulation results, displacement vectors show a funnel-shaped failure mechanism [8]. The mechanism is similar to the localized failure mechanism. The result of simulation shows that the soil elements in front of the tunnel and far away the tunnel affect the soil ground and forms the failure areas.

4.2 Passive Failure Pressure

The passive failure pressure in front of tunnel is calculated by finite element method and synthetized in Table 2.

Compared with results in the centrifugal experiments $T1$ and $T2$ conducted by Tuan et al. [9], Wong [10], the pressure curves calculated by finite element method recognize a similar trend but with the smaller value.

For centrifugal and numerical analysis experiments, we find that in case tunnels located at $C/D = 1.5$, the soil in front of the tunnel face is shifted forward, whereas the soil in regions located further away from the tunnel axis is forced outward. It is observed that the funnel-shaped failure mechanism is narrower than a five-block failure mechanism, which is assumed in existing upper-bound solutions. However, the analyses of passive failure pressures by the upper-bound solutions are reasonably consistent with the measured tunnel pressure. In addition, the displacement pressure curves in finite element method are appropriate with the measured results.

In case the tunnel is placed at C/D ratio is 3.3, soil displacement is limited around the vicinity of the tunnel mirror as it moves forward and a localized failure mechanism related to ground deformation is observed. The tunnel failure pressure in finite element method is smaller than corresponding value (9.4–48.4%) for the C/D ratio is 1.5 and (3.6–25.4%) for the C/D ratio is 3.3. However, it can be seen that there is a correlation of the passive failure pressure in the centrifugal experiment and the finite element method (Fig. 5) [1, 9–11].

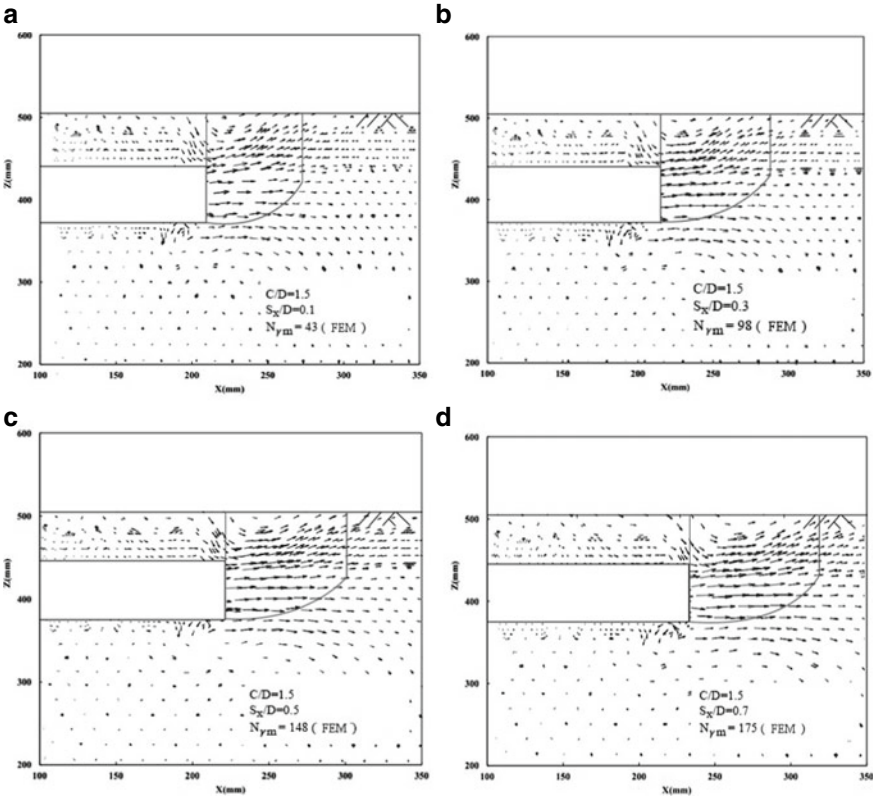


Fig. 3. **a** Displacement vector at $C/D = 1.5$, $S_X/D = 0.1$, **b** displacement vector at $C/D = 1.5$, $S_X/D = 0.3$, **c** displacement vector at $C/D = 1.5$, $S_X/D = 0.5$, **d** displacement vector at $C/D = 1.5$, $S_X/D = 0.7$

4.3 Ground Displacement

The ground displacement in front of tunnel is analyzed by finite element method as shown in Figs. 6 and 7.

In case the tunnel is located at C/D ratio is 1.5, it can be seen that the soil in front of the tunnel face is shifted forward which makes it to be compressed and causes the surface to be exposed. As a result, the funnel-shaped failure is formed.

When the tunnel depth increases ($C/D = 3.3$), the deformation zone is found to be significantly larger and the range of influence is wider accordingly. The stresses in the surrounding site of the tunnel rise, which makes the sand be less pushed up. As a result, it will be compressed in front of the tunnel. The largest displacement is located near the original position of the tunnel, but its value is lower than the ground displacement around the tunnel in which the C/D ratio is 1.5.

The calculation results of soil deformation in finite element method show a similar figure with centrifugal tests, expect for the wider range of influence. Besides, the curve

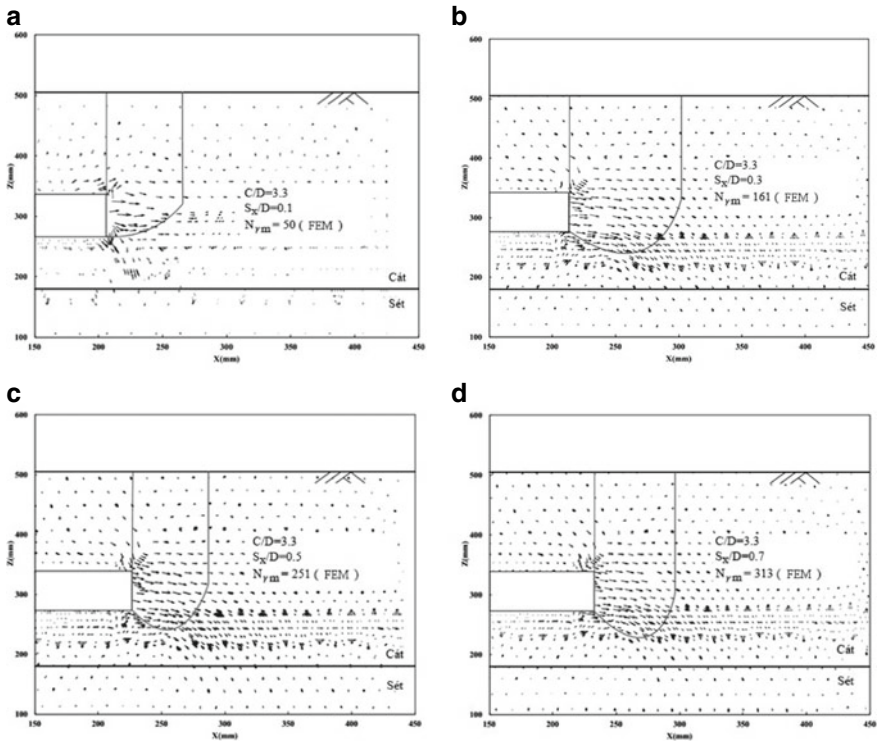


Fig. 4. **a** Displacement vector at $C/D = 3.3$, $S_x/D = 0.1$, **b** displacement vector at $C/D = 3.3$, $S_x/D = 0.3$, **c** displacement vector at $C/D = 3.3$, $S_x/D = 0.5$, **d** displacement vector at $C/D = 3.3$, $S_x/D = 0.7$

Table 2. Calculation results of passive failure pressure by FEM for two placing tunnel cases

S_x/D	$N_{\gamma m}$	
	$C/D = 1.5$	$C/D = 3.3$
0.1	43	50
0.2	64	130
0.3	98	161
0.4	116	190
0.5	148	251
0.6	167	288
0.7	175	313
0.8	183	321

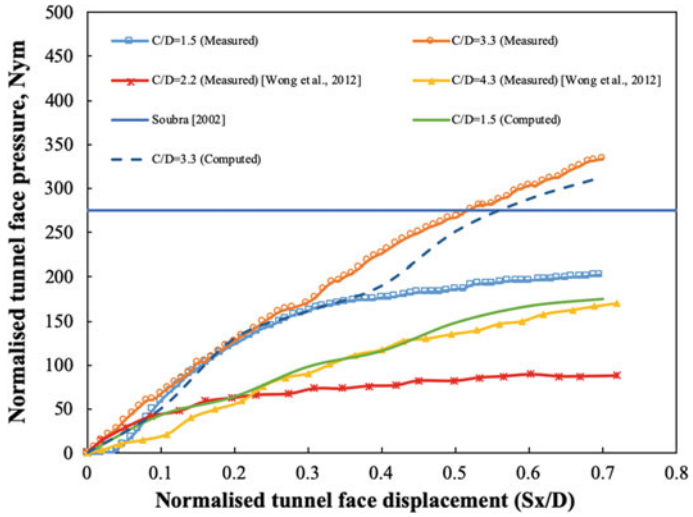


Fig. 5. Relation chart between $N_{\gamma m}$ and S_x/D in FEM and centrifugal test

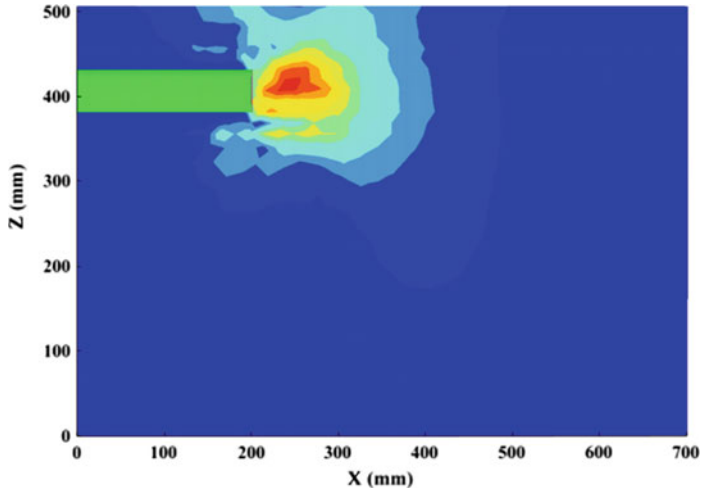


Fig. 6. Ground displacement in front of tunnel at the *C*/*D* ratio is 1.5

graph of ground displacement from all of methods is a similar shape with the Gaussian curve proposed by Peck and Schmidt [12].

5 Conclusions

There is analogy between funnel-shaped failure mechanism and localized failure mechanism soil in front of the tunnel face. The results of the study show that the soil elements

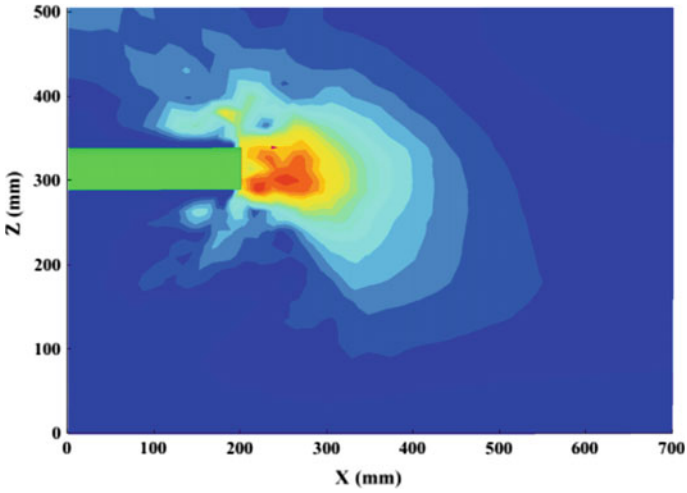


Fig. 7. Ground displacement in front of tunnel at the *C*/*D* ratio is 3.3

in front of the tunnel face and at the far side of the tunnel which effects on the ground surface form a failure area around the tunnel.

The results of failure mechanism in finite element method show a higher figure than in centrifugal tests. One of the reasons is possibly due to the maximum expansion angle which is acceptable when soil elements follow the failure mechanism in numerical simulation.

Compared to the measurement results in centrifugal tests, the curves of passive failure pressure in finite element method show a similar trend, but with a lower value. The reason is that the data in the centrifugal tests was affected by friction between the tunnel and pads, glass or sand. However, this effect is not included in the numerical analysis process.

The value from failure pressure mechanism in finite element method is lower than the measured value (9.4–48.4%) for the *C*/*D* ratio is 1.5 and (3.6–25.4%) for case *C*/*D* 3.3, respectively. However, it can be seen that there is a correlation between the passive failure pressure in centrifugal tests and in finite element method.

As the tunnel moves ahead, the soil in front of the tunnel face is shifted forward accordingly which makes it to be compressed and causes the surface to be exposed. As a result, the funnel-shaped failure is formed. As the depth of tunnel increases, the deformation zone is found to be significantly larger and the range of influence is also wider.

References

1. Soubra, A.H., Dias, D., Emeriault, F., Kastner, R.: Three-Dimensional Face Stability Analysis of Circular Tunnels by a Kinematical Approach, pp. 894–901. Geotechnical Special Publication (2008)
2. Chambon, P., Corte, J.F.: Shallow tunnels in cohesionless soil: stability of tunnel face. J. Geotech. Eng.-ASCE **120**(7), 1148–1165 (1997)

3. Loganathan, N., Poulos, H.G.: Analytical prediction for tuinnelling—induced ground movements in clays. J. Geotech. Geoenviron. Eng. ASCE **124**(9), 846–856 (1998)
4. Oblozinsky, P., Kuwano, J.: Centrifuge experiments on stability of tunnel face. Slovak J. Civ. Eng. **3**, 23–29 (2004)
5. Taylor, R.N.: Geotechnical Centrifuge Technology Geotechnical Centrifuge Technology, 1st edn. Taylor & Francis Group, USA (1995)
6. Grant, R.J.: Movements around a tunnel in two-layer ground. Ph.D. Thesis, City University of London (1998)
7. Mair, R.J.: Centrifugal modelling of tunnel construction in soft clay, Ph.D. Thesis, Cambridge University (1979)
8. Leca, E., Dormieux, L.: Upper and lower bound solutions for the face stability of shallow circular tunnels in frictional material. Geotechnique **40**(4), 581–606 (1990)
9. Tuan, N.A., Chau Ngoc An, Tam, N.M.: Analysing the stability of tunnel face in two layers soil. Vietnam J. Constr. **3**(2017), 182–187 (2017)
10. Wong, K.S.: Passive failure and deformation mechanisms due to tunnelling in sand and clay. Ph.D. Thesis, The Hong Kong University of Science and Technology (2012)
11. Soubra, A.H.: Kinematical approach to the face stability analysis of shallow circular tunnels. In: Proceedings of the Eight International Symposium on Plasticity, pp. 443–445 (2002)
12. Schmidt, B.: Settlements and ground movements associated with tunneling in soil. Ph.D. Thesis, University of Illinois (1969)

The Working Capacity of Vietnamese Local Civil Servants

Ngo Sy Trung[1(✉)], Do Huu Hai[2], Vu Thi Yen Nga[1], and Tran Thi Hanh[1]

[1] Hanoi University of Home Affairs, 36 Xuan La Str., Tay Ho Dist., Ha Noi City, Vietnam
ngosytrung01@yahoo.com

[2] Ho Chi Minh City University of Food Industry, 140 Le Trong Tan Str., Tan Phu Dist., Ho Chi Minh City, Vietnam

Abstract. Local civil servants are those who directly deal with the people's requests, guaranteeing their lawful rights and interests in accordance with the law. In each country, they play a really crucial role in the implementation of state policies and political goals at the grassroots level. The performance of governmental agencies heavily depends on the working capacity of the contingent of local civil servants, which is confirmed by many researchers. Therefore, this research focuses on analyzing the working capacity of Vietnamese local civil servants through such contents as the occupational practice capacity, the Sense and Responsibility for the work, the Attitude to serve the people, and the Additional capacity. On the basis of analyzing the criteria of the civil servant's capacity, the authors designed the questionnaire and conducted a survey of people's opinions on the working capacity of civil servants in governmental agencies representing all three provinces of Vietnam which are Hai Phong, Hue, and Can Tho. The survey was conducted selectively, in which the people answering the questions were the residents who had contacted and worked with civil servants at local government agencies more than 5 times. The research results show that the people assessed the occupational practicing capacity as the factor the most powerfully affecting the working capacity of civil servants, followed by the Attitude to serve the people and the Sense and Responsibility for the work, respectively, and finally is Additional capacity. From the above research results, the authors propose a number of suitable solutions to improve the quality of local civil servants in Vietnam. The solutions include: (1) Strengthening professional retraining for civil servants; (2) Regularly raising awareness of working sense and responsibilities and service attitude of civil servants; (3) Providing civil servants with basic knowledge and skills in foreign languages and information technology to well perform professional tasks.

Keywords: Working capacity · Local civil servants · Vietnam

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_16

1 Overview of the Working Capacity of Vietnamese Local Civil Servants

Capacity is considered as the human ability to work (including his constitution, understanding, and ethical qualities). In many Vietnamese dictionaries, the term "capacity" is explained similarly by many linguists, which is "Capacity is a characteristic of an individual showing his or her proficiency level, which means that s/he can perform one or more kinds of activity properly and competently" [1] or "Capacity is the ability to work well based on the ethical qualities and professional qualifications" [2]. In addition, they also define the term "performing the task" as "Carrying out the assigned work" [3] or "executing an allotted mission" [2]. Thus, it can be seen that capacity to perform an individual's tasks means the ability to fulfil the assignments that s/he is entrusted with, mainly expressed through the results of that work.

For local civil servants, the laws of countries also specify general principles in the performance of their public duties, which are to devotedly serve the people and the state. The Law on Cadres and Civil Servants of Vietnam issued in 2008 specifies the civil servants' duties, which are devotedly serving the people; protecting national interests; properly and fully performing, and taking responsibility for the results of assigned tasks. To well perform these tasks, they must study, acquire, and practice many skills ranging from occupational skills, additional skills for occupational practice, service attitude to sense, and responsibility for the assigned tasks. Since then the civil servants' capacity to perform basic tasks will be formed. This kind of capacity is described by many researchers as well as by the Vietnamese law with specific contents such as the Occupational practicing capacity, the ability to analyze and handle situations, Foreign languages competences, Information technology skills, the Sense and Responsibility for the work and the Attitude to serve the people [4–12].

1.1 The Occupational Practicing Capacity (PPC)

Many researchers refer to the competency of occupational practice as the ability of the subject to master one or a series of actions on the basis of his knowledge or experience to produce the expected results [5, 8]. The above research viewpoint is highly practical when associating with local civil servants, whereby the capacity of occupational practice is determined as the actual ability of civil servants to work (including quality and progress of work performance), ensuring the principle of serving the common interests. This is considered as an important content consistent with the trend of the world's administration. The content reflecting the occupational practicing capacity of local civil servants is also mentioned and analyzed here including:

- Discipline (OPC1): It is a way of working in strict compliance with the provisions of law and the process of performing the tasks in the agency, establishing order and stability in the performance of public duties. However, in many cases, this does not always make people satisfied, especially with the work that needs to be simplified in the settlement to facilitate the people without affecting the operations of governmental agencies.

- Flexibility (OPC2): It is a way of working that complies with the basic rules to ensure discipline in the performance of public duties while simplifying the steps in resolving the people's requests without affecting the collective interests and overall operations of governmental agencies.
- Promptness and accuracy (OPC3): Civil servants use their professional manipulations combined with scientific thinking to solve the requirements of the people in the shortest possible time while ensuring accuracy. Since then the people's satisfaction and trust towards the local government have been created.
- Independently settling the work (OPC4): This is an indispensable condition for each civil servant to well perform the assigned task. However, independently handling the work is usually done for simple tasks associated with a specific stage or just for a pure professional job.
- Coordinating to solve the work (OPC5): Management activities in governmental agencies consist of many complicated tasks that require coordination in the settlement process. Therefore, besides the ability to independently solve simple tasks, civil servants need to have harmonious coordination so that complicated tasks can be solved quickly.

1.2 Sense and Responsibility for the Work (SRW)

The sense and the responsibility for work related to public service morals of civil servants are specified by the laws of Vietnam [6] and other countries, which is expressed in such many aspects as follows:

- Complying with legal regulations in the professional field (SRW1): This is the civil servants' basic working principle ensuring discipline in performing official duties without infringing upon the people's legitimate rights and interests. State management activities are carried out under the principle of respecting the laws and therefore, civil servants must abide by the law in the process of performing their assigned tasks.
- Abiding by regulations on working time (SRW2): Governmental agencies are characterized by regularly serving the people. Thus, in addition to the management of work efficiency, time management needs to be strictly applied. Therefore, civil servants must abide by the working time prescribed to serve the people frequently and continuously.
- Cooperation at work (SRW3): It is the integration, sharing, understanding, and interests of civil servants towards the people who work with the public authorities. Therefore, the civil servants need to grasp a lot of wishes as well as feedback from the people so as to better serve them.
- Eagerly learning and self-improvement at work (SRW4): Social issues are getting more and more complicated and constantly arising associated with the people's demands for the government so that each civil servant himself must continuously self-study with a positive attitude from which can solve the people's work in the best way.
- The dedicated spirit for the collective benefits (SRW5): Public servants are considered as the ones always serving the people. They must always identify themselves as servants and put the collective interests above their personal ones. Therefore, they must

have their professional working manner suitable for each kind of job they assume. They always show their spirits of willingness to work together with their dedication, passion, and serious working attitude.
- Taking responsibility for the results of work settlement (SRW6): This is the standard of competence associated with the civil servants' public service morals. Accordingly, they must be mainly responsible for the results of solving the people's requests, thereby the trust of the people towards the government will be created.

1.3 Attitude to Serve the People (ASP)

The laws of Vietnam and many countries stipulate this content associated with the civil servants' public service ethics. In addition, it is also analyzed in details by many researchers under such many aspects as [7, 8]:

- Respecting the people (ASP1): The local civil servants must put the people's interests first, devotedly serve them, be ready to meet their legitimate requirements as well as guide them in a caring and thoughtful manner.
- Listening to the people's comments and feedback (ASP2): Civil servants need to listen to comments and feedback from the people to grasp the people's inmost feelings, aspirations, and voices towards the authorities. They should become a direct link between the people and the government contributing to promoting the spirit of democracy in state management.
- Being polite in communicating with the people (ASP3): Local civil servants are those who often come in direct contact and handle people's requests. Thus, in the process of handling the people's affairs, they must have courteous attitudes and good communication skills in accordance with the rules and etiquette of contemporary society.
- Fairness in addressing the people's requests (ASP4): In fact, the civil servants serve many subjects at different levels of awareness and backgrounds. Therefore, they must be aware of and act in a fair manner to make the people confident, avoiding unnecessary conflicts to guarantee mutual interests.

1.4 Additional Capacity (AC)

Additional capacity includes a number of complementary capabilities that civil servants require to have to be able to well perform their primary tasks. This capacity has been mentioned by many researchers as follows [5, 8, 9, 12]:

- Foreign language competences (AC1): In addition to the duties which are directly addressing the people's requests, they are also advisers on formulating strategies and implementing local international cooperation activities. In the current context of deep and wide international integration, state agencies need to exchange and learn experiences from foreign countries, from domestic and international organizations. This requires each civil servant to have a certain foreign language proficiency which is considered as the key to integration. In addition to using foreign language fluently, they also need to master foreign languages for specific purposes to meet the

research requirements, to sum up, domestic and foreign practical experiences with an aim to advise their superiors on policymaking related to the implementation of local international cooperation activities.

- The ability to apply information technology (AC2): This is a crucial and necessary content of modern administration. In fact, "e-government" is being widely applied in countries all over the world, so the ability to apply information technology is a basic but highly demanding requirement for civil servants. The performance of governmental agencies' services is assessed through the quickest and best response to the people's requests. In the performance of tasks, if civil servants have information technology skills, it will be an important factor to help the affairs of the agency be favorable.
- The ability to analyze problems (AC3): The civil servants' activities are associated with the implementation of policies and laws by the state. In addition, they have to take responsibility for giving advice to the state in the fields assigned. Therefore, in addition to good occupational practice skills, they must also show their ability to analyze, research and synthesize social issues in order to be able to promptly advise and propose with leaders in policy decision making as a tool for the management and administration of local affairs.
- The ability to handle situations (AC4): In the course of implementing the tasks, many problems that arise are brought to the authorities by the people. There are even unprecedented situations requiring civil servants to have a clever solution to both satisfy the people and not violate the rules and general operations of the agency. This is an issue that requires civil servants to pay great attention through the study of problem-solving skills via their theoretical and practical experiences.

2 Study Model

From the above-mentioned research overview, the authors have developed a study model of factors affecting the working capacity by local civil servants as shown in Fig. 1.

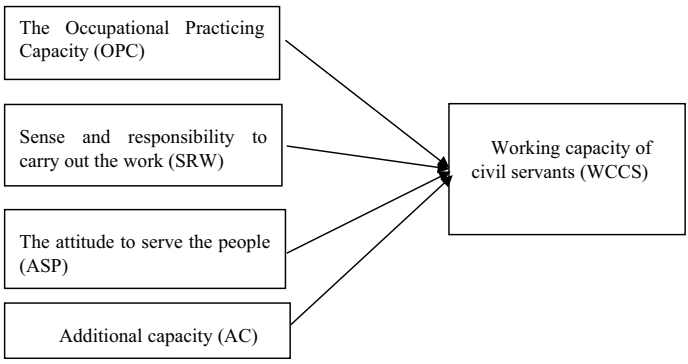


Fig. 1. Research model

In the above-mentioned study model, the independent variables have an influence on the dependent ones (Working capacity of civil servants) at different levels based on the

people's assessments, thereby showing the real working capacity of civil servants. This is the task that the research team focuses on clarifying the relationship between these variables.

Hypothesis:

H1: The capacity of occupational practice is related to (affect) the working capacity of civil servants.

H2: The Sense and Responsibility for the work are related to (affect) the working capacity of civil servants.

H3: The attitude of serving the people is related to (affect) the working capacity of civil servants.

H4: The additional capacity is related to (affect) the working capacity of civil servants.

3 Research Methodology

The research was conducted in two steps which are the preliminary and the formal investigations.

1. Preliminary investigation

The authors used qualitative research conducting in-depth personal interviews combined with group interviews focusing on a number of experts and civil servants working in the local governmental agencies in order to modify the research scale and complete the questionnaires in such a way to be suitable for Vietnam's context.

It is noted that the questionnaire was built based on the research overview and the scales were all tested by previous studies, so the authors only inherited and adjusted the words accordingly. The components in the model are measured with the 5-point Likert scale, in which, 1 is strongly disagree and 5 is strongly agree. The final section contains demographic information including questions about gender, age, education level, income, and the number of years involved in local governance by using the naming/identification scale.

2. Official investigation

The complete research questionnaire was put into the official investigation by quantitative methods in three localities representing all three regions of Vietnam: Hai Phong, Hue, and Can Tho; The survey period was from August to November 2019. The research population is that the local people over 18 years of age, who have worked with and interacted with civil servants at local authorities over 5 times, of all ages, with places of residence, incomes, different knowledge levels.

The measurement model consists of 19 observed variables. According to the principle, there is 1 measurement variable for every five elements [13]. For this reason, the minimum number of samples needed is $19 * 5 = 95$ and implemented in three localities

and thus the minimum sample size is 95 * 3 = 285. However, in this study, the authors selected a sample size of 300, which is larger than the minimum one, in order to ensure the science in selecting the sample. The questionnaires were sent directly to respondents with the non-random sampling method (preliminary interview to know the number of visits to the governmental agencies during the last 5 years by the people surveyed and consented to answer the questionnaire).

The result was 281/300 satisfactory votes, with the response rate of 90%. The characteristics of the sample in Table 1 are 60.5% for females and 39.5% for males with the age from 18 to 35 accounting for 57.2% and over 35 accounting for 42.8%. The times working with governmental agencies are 5–10 ones accounting for 24.3% and more than 10 ones accounting for 75.7%. The data processing method is implemented through the following steps: (1) verifying the scale and reliability of the measurement variables by Cronbach's Alpha coefficient and the value by analyzing the factor of discovery EFA (Exploratory Factor Analysis); (2) analyzing multiple linear regression in order to test the model and research hypotheses.

Table 1. Characteristics of research samples

Sample size ($n = 281$)		Frequency	Ratio (%)
Gender	Male	107	39.5
	Female	164	60.5
Age	From 18 to 25	82	30.2
	From 26 to 30	73	27.0
	Over 35	116	42.8
Number of times working with local government	5–10 times	66	24.3
	Over 10 times	205	75.7
Academic qualifications	High school Diplomas	43	15.8
	Colleges/Professional high Diplomas	35	13.0
	Graduates	118	43.5
	Postgraduates	52	19.2
	Others	23	8.5

4 Research Findings

4.1 Scale Testing

By using SPSS 23.0 software, the authors have drawn the results of the reliability analysis of the scale (Cronbach Alpha) and those of the Exploratory Factor Analysis EFA (Exploratory Factor Analysis) suggesting the type of removal and consolidation of a number of observed variables to help the scales more accurately assess concepts.

The first is to test the scale by using Cronbach's Alpha reliability coefficient. This step is used to eliminate the garbage variables avoiding the case of garbage variables creating fake factors when analyzing the discovery factor EFA [13]. The testing standard is that Cronbach's Alpha coefficient must be greater than 0.6 and the correlation coefficient of the total variable in each scale must be greater than 0.3 [13]. The analysis results in Table 2 show that all scales of the factors meet the criteria. Thus, all scales of factors are reliable and used to analyze the next factor.

Table 2. Summary of reliability and relative minimum variables in the scales

Scales	Number of observed variables	Reliability coefficient (Cronbach Alpha)	The correlation coefficient of the smallest total variables
The occupational practicing capacity (OPC)	5	0.797	0.516
Sense and Responsibility for the work (SRW)	6	0.828	0.480
Attitude to serve the people (ASP)	4	0.696	0.586
Additional capacity (AC)	4	0.690	0.437
Working capacity of civil servants (WCCS)	4	0.739	0.496

The first is to examine Cronbach's Alpha, followed by the Exploratory Factor Analysis (EFA) in order to preliminarily assess unidirectional, convergent, and distinguishable validity of the scale. The EFA discovery factor analysis method is performed with the Principal Component Analysis method and Varimax rotation to subgroup the factors. With a sample size of 281, the factor load coefficient of the observed variables must be greater than 0.5; the variables converge on the same factor and differentiate from other factors [13]. In addition, the KMO coefficient must be between 0.5 and 1.

The analysis results in Table 3 show that all factor load coefficients of all the observed variables are greater than 0.5; Bartlett test with the significance Sig. = 0.000 with KMO coefficient = 0.815. All 23 variables after EFA analysis are divided into four factors with Eigenvalues values greater than 1 and deviation must be greater than 50%. The research model changes into four independent variables, and one dependent variable, which is used for linear regression analysis and subsequent hypothesis testing.

4.2 Regression Analysis and Hypothesis Testing

Firstly, using Pearson correlation analysis to analyze the correlation between quantitative variables. At the significance level of 5%, the correlation coefficients show that

Table 3. The analysis of EFA factors affecting local governance

Observed variables		Factors				
		1	2	3	4	5
The occupational practicing capacity (OPC)	OPC3	0.781				
	OPC4	0.767				
	OPC2	0.754				
	OPC5	0.687				
	OPC1	0.671				
Sense and Responsibility for the work (SRW)	SRW3		0.785			
	SRW5		0.756			
	SRW1		0.735			
	SRW4		0.680			
	SRW6		0.678			
	SRW2		0.661			
Attitude to serve the people (ASP)	ASP2			0.850		
	ASP3			0.837		
	ASP1			0.742		
	ASP4			0.729		
Additional capacity (AC)	AC3				0.756	
	AC2				0.720	
	AC1				0.706	
	AC4				0.690	
Working capacity of civil servants (WCCS)	WCCS3					0.764
	WCCS2					0.762
	WCCS1					0.739
	WCCS4					0.714

the relationship between the dependent and the independent variables are statistically significant (Sig. Value is less than 0.05). The magnitude of the correlation coefficients ensures that there is no multi-collinear phenomenon. Therefore, other statistics can be used to test the relationship between variables.

The next, conducting multiple linear regression analysis on the relationship between the factors belonging to working capacity by local civil servants. As shown in Table 4, the R2 coefficient of 0.863, shows that the multiple linear regression model has been built in accordance with the data set of 86.3%. The regression results show that four independent variables namely OPC, SRCW, ASP, AC (Table 4) with non-normalized Beta coefficient are 0.159; 0.177; 0.153; 0.197, respectively, with significance level less than 0.05. Thus, the hypotheses H1, H2, H3, H4 are accepted.

Table 4. Regression results of model of factors affecting the local governance

Model	Non-normalized regression coefficient		Normalized regression coefficient	Value *t*	Significance (Sig.)	Collinear statistics	
	B	Standard error	Beta			Tolerance	VIF
(Constant)	1.344	0.005		26.911			
OPC	0.159	0.008	0.317	18.861	0.000	0.889	1.125
SRW	0.177	0.012	0.319	14.789	0.000	0.654	1.529
ASP	0.153	0.010	0.348	14.957	0.000	0.678	1.476
AC	0.197	0.010	0.377	20.007	0.000	0.854	1.171

Dependent variable: Working capacity of civil servants (WCCS)
R2 adjusted = 0.863

Based on the coefficient B (Table 4), it can be concluded that Occupational practicing capacity (OPC) has the strongest influence on the Working capacity of civil servants (WCCS), followed by the Attitude to serve the people (ASP) and Sense and Responsibility for the work (SRCW). The attitude capacity (AC) is the least influential factor in the Working capacity of civil servants (WCCS). The linear regression model obtained has the following form:

$$\boldsymbol{WCCS} = 1.344 + 0.159 * \boldsymbol{OPC} + 0.177 * \boldsymbol{SRW} + 0.153 * \boldsymbol{ASP} + 0.197 * \boldsymbol{AC} + e$$

5 Conclusions and Recommendations

The research results have shown that Occupational practicing capacity (OPC) has the strongest influence on the Working capacity of civil servants (WCCS), followed by the Attitude to serve the people (ASP) and the Sense and Responsibility for the work (SRCW); The Attitude capacity (AC) is the least influential factor to the Working capacity of civil servants (WCCS). From the above research findings, the authors have proposed as follows:

1. Annually retraining local civil servants

In Vietnam, there are still many elderly local civil servants who are not well-trained, and are less active in self-study to improving knowledge, skills, and working methods. Besides, the rapid development of Vietnam in recent years has been posing many issues in need of being adjusted not only from the state's policy but also the renovation of civil servants' thoughts and working methods. Therefore, the state needs to enhance annual professional training and retraining so that local civil servants are regularly updated with knowledge and skills to best meet their professional tasks.

If the above-mentioned issue has been well implemented, it will contribute to maintaining the contingent of civil servants with professional qualifications and high capacity meeting state management requirements as well as contribute to the development of high-quality public service personnel to best meet the people's legal requirements.

2. Regularly fostering and improving the Sense and Responsibility for the work and the attitude to serve the people of local civil servants

In the state management activities, if there are good policies, adequate facilities, and techniques but lack of sense, responsibility, attitude to carry out the work, the state management cannot achieve high efficiency. Therefore, local governmental agencies need to have a plan to regularly retrain civil servants on this content to raise their awareness and ideology so that they are always ready to work, ready to serve for the common interests of the people, the locality, and the nation.

Once civil servants are regularly retrained and raised their awareness, working responsibilities, and service attitudes, the people will find themselves be respected, be served enthusiastically and then they will be more confident in the local government. Since then, the local government can mobilize resources from the people in a convenient manner to serve socio-economic development in the area.

3. Equipping local civil servants with basic knowledge and skills on foreign languages and information technology to well perform professional tasks

The research results have shown that the people are less interested in local civil servants' foreign language proficiency and information technology but mainly interested in their occupational practicing capacity, Sense and Responsibility for the work, and service attitude. However, to meet the requirement of the administrative reform and integration, retraining activities need to be planned to ensure that civil servants can afford to equip and update knowledge as well as basic skills in foreign languages and informatics for their professional management work.

Thus, with the development of a specific research theoretical framework on the working capacity of local civil servants, together with the practical survey results in many localities representing the three regions of Vietnam, the authors have pointed out scientific conclusions and proposed a number of state management solutions. Based on the tested research results, the proposed solutions focus on regularly retraining of professional skills and working Sense and Responsibility, service attitude of civil servants. The authors also hope that these research results will provide useful information for managers to make timely and appropriate policy adjustments in order to improve the quality of local civil servants to meet the demand of the people and the socio-economic development at local and national levels.

References

1. Vietnam Encyclopedia: Encyclopedia Publishing House, Episode 3 (2005)
2. Lan. N.: Vietnamese Dictionary and Words. Ho Chi Minh City General Publishing House (2006)
3. Vietnamese Dictionary: Hong Duc Publishing House (2010)
4. Minh, V.B.: The quality of civil servants in state administrative agencies in the northern mountainous provinces in Vietnam—research conducted in Dien Bien province. Ph.D. Thesis, National Economics University (2017)
5. Quynh, N.H.L.: The capacity of public officials in specialized agencies under the People's Committee of Quang Ninh Province, Master Thesis, National Administration Academy (2016)
6. Law on Cadres and Civil Servants issued in 2008 by the Socialist Republic of Vietnam (2008)
7. Can, T.N.: The quality of public service performance—the key issue of administrative reform (2014)
8. Trung, S.N.: Training highly-qualified human resources for civil service. State Management Magazine, No. 9-2016 (2016)
9. Trung, S.N.: Local Governance. Hanoi National University Publishing House (2019)
10. Hai, H.N.: Development of the contingent of public administrative officials at present. J. State Adm. 10-2014 (2014)
11. Tam, T.N.: Some recommendations on building a contingent of cadres and civil servants in line with current administrative reform requirements. State Organ. J. 5-2011 (2011)
12. Thoa, T.T.: Improving the performance of public duties of provincial public servants in Thai Nguyen. Master Thesis, National Academy of Public Administration (2013)
13. Hai, H.D.: Corporate Culture—The Pinnacle of Wisdom. Hanoi Transport Publishing House (2016)
14. Hai, H.D., Ngoc, S.H.: Statistical Probability and Quantitative Methods in Economics, Hanoi Labor Publishing House (2018)

AI-HI…The Technological Transgenders (AI-HI Techno-Trans)

Vijay Singh Rathore(✉)

CS & IT, IIS (Deemed to be University), Jaipur, India
vijaydiamond@gmail.com, vijaysingh.rathore@iisuniv.ac.in

Abstract. The evolution of technology (specifically artificial intelligence and machine learning) in the recent time has made the human life easier, fast, and providing real-time solutions, but at the same time, its negative impacts are also simultaneously affecting the human intelligence and human behavior. This paper discusses some observations on the impact of technology (AI) upon the human intelligence (HI) and human behaviors, and the possible threats as an outcome of the uncontrolled and excess use of the technology enabling loss in the human intelligence, resulting in the emergence of 'AI-HI…The Technological Transgenders' abbreviated as 'AI-HI Techno-Trans.' The paper defines Techno-Trans Children, Techno-Trans Youth, and Techno-Trans Robots, who would collectively forms a Techno-Trans Society enabling the benefits and threats together. The rapid increase in the development of artificial intelligence (AI) technology, along with the simultaneous decline in human intelligence (HI) might be the outcome of the excess use of the technology, and dependency upon it at large toward the formation of the Techno-Trans Society.

Keywords: Artificial intelligence · Human intelligence · Technological transgenders · AI-HI Techno-Trans · Techno-Trans children · Techno-Trans youth · Techno-Trans robots · Techno-Trans society

1 Introduction

AI-HI (Pronounced as 'Aaye-Haaye', a frequently spoken phrase along with clapping by the Street Transgenders in the towns of India) is referred to the impact of technological advancements upon the humans' behaviors and emotions, wherein a 'Machine' is continuously improving its learning through artificial intelligence and machine learning (AI-ML) and evolving toward thinking acting like a 'Human', and on the other hand, the 'Human' is behaving and acting like 'Machines' with the excess use of technological advancements such as mobile phones, IoT-based equipments, social media, e-commerce, and other online/real-time platforms. (Fig. 1).

ArtificialiIntelligence (AI) has been evolving with a greater speed and the day is not far when the machines (Robots) would be doing the things, the same way as humans

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_17

Fig. 1. AI-HI, the technological transgender [1]

do; however, at the same time, simultaneously, most of the 'Human' are losing their intelligence (HI-Human Intelligence) by adopting the technology advancements. The machines would not remain pure machines only, since they will be acting as humans, and the humans are also behaving and acting like machines, so they will also not remain pure humans in the times to come. This, 'Neither Human, nor Machine' transformation can be termed as 'Technological Transgenders,' who are the humans fully dependent upon the technology (i.e., AI). Hence, the term 'AI-HI', which means, the decline of the human intelligence (HI) through the impact of artificial intelligence (AI) enabling the human becoming 'AI-HI…The Technological Transgender' (Techno-Trans) (Fig. 2).

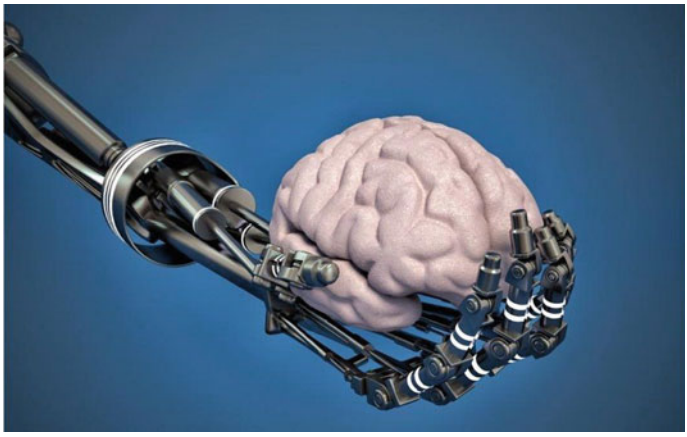


Fig. 2. AI-HI techno-trans [2]

2 Related Study (Review of Literature)

2.1 Artificial Intelligence (AI)

Artificial intelligence (AI), the ability of a digital computer- or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience. Since the development of the digital computer in the 1940s, it has been demonstrated that computers can be programmed to carry out very complex tasks—as, for example, discovering proofs for mathematical theorems or playing chess—with great proficiency. Still, despite continuing advances in computer processing speed and memory capacity, there are as yet no programs that can match human flexibility over wider domains or in tasks requiring much everyday knowledge. On the other hand, some programs have attained the performance levels of human experts and professionals in performing certain specific tasks, so that artificial intelligence in this limited sense is found in applications as diverse as medical diagnosis, computer search engines, and voice or handwriting recognition. [8].

2.2 Human Intelligence (HI)

Human intelligence can be defined as the mental quality that consists of the abilities to learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate one's environment [9].

2.3 Transgender

The dictionary meaning of a Transgender is 'denoting or relating to a person whose sense of personal identity and gender does not correspond with their birth sex.' Transgender people have a gender identity or gender expression that differs from their sex assigned at birth. Transgender is an umbrella term for people whose gender identities, gender expressions, and/or behaviors are different from those culturally associated with the sex to which they were assigned at birth [10, Wikipedia].

3 Research Observations and Findings

3.1 AI-HI…The Technological Transgenders (AI-HI Techno-Trans)

As stated above, Technological Transgender can be defined as being quite similar to the 'Transgender' (whose gender identities, gender expressions, and/or behaviors are different from those culturally associated with the sex to which they were assigned at birth), since by birth if someone is Human, and then, gradually with the excess adoption of the technology, he/she keeps on losing his/her intelligence (i.e., HI-Human Intelligence), becoming solely dependent of technology (i.e., AI-Artificial Intelligence), and behaving like 'machines,' resulting in diverting from the original status of being 'human' to some sort of 'machine' then, he/she may be referred to as 'Technological Transgender' (Techno-Trans).

3.2 Technological Transgender (Human)

Initially, the human is unknown of the fact that the excess use of the technology might lead to negative results. He/she likes the technology (AIML) for various reasons such as the impressive and fast outcomes, ease of work, less efforts, comfort, accuracy, real-time solutions, interesting, time pass, being more socialized and internationalized through Social Media like platforms, and so many more to mention (Fig. 3).

Fig. 3. AI-HI technological transgender (Human)-1 [3]

However, at the same time, simultaneously, the technology keeps imposing and controlling the human mind and makes it habitual of the technology such that after a certain amount of time, the human is helpless without the presence of the technology. The human also, unknowingly, starts forgetting his/her original powers both the mental and physical, and behaves as per the technology driven, like a machine (Fig. 4).

Fig. 4. AI-HI technological transgender (Human)-2 [4]

Now, neither he remains a pure human, nor can he become as a machine, and this intermediate state defines him/her as a Transgender (Neither Human, Nor Machine) because of being habitual of the technology. The human wants to become 'Hi-Fi' (normally referred to a Modern and Up-to-date person), but, unfortunately, it turns 'AI-HI' (AI-HI…The Technological Transgender).

3.3 Technological Transgender (Machine)

On the other hand, the machine is developing their intelligence through the best practices and applications of artificial intelligence and machine learning algorithms, so forth, enabling human capabilities and intelligence inside them (machines). The machines do not have the biological components such as brain, heart, or cells, but still they are learning to act like a human with latest advancement in the technology. Hence, the machines are also not remained machines only, because of doing the human like actions, even though, they are not human, therefore, they are neither machine nor human, so, can be referred to as 'AI-HI Technological Transgender' variant of machine.

4 Research Outcome (Results and Discussions)

The result of this study defines the following Technological Transgenders (AI-HI Techno-Trans).

4.1 AI-HI Techno-Trans Children

In many nuclear families, it can be seen most commonly that the parents have limited time to observe and take care of the child, and to avoid 'unnecessary attention' (which is actually very necessary), they give the child access to either T.V. or smartphone (video games, plays, songs, cartoons, whatsapp, facebook, etc.) and let him involve in the 'virtual digital world.' The child finds it interesting, and keeps playing with the 'digital world,' and forgets everything around. Meanwhile, the parent might feed him, change the dress, or anything else for which earlier it was making resistance, but now, since he is fully involved in the virtual digital world, he does not resist for anything being done by the parent. The parent feels it very comfortable, since the child is silently busy in with the 'digital world,' and the parent now can do their other works without any trouble (Fig. 5).

However, no one is aware that this practice may gradually make the child habitual of the virtual world through the excess use of the technology, and simultaneously starts losing the emotions, tastes, hunger, thrust, immune power, and most of all the emotional attachment with the parent or other relatives, since his all world now, is the 'virtual digital world' provided through the technology. The child evolves himself toward a 'AI-HI Techno Trans Child.' Let us be ready for the future consequences (which will surely not be favorable at all) of developing those 'AI-HI Techno Trans Child.'

Fig. 5. AI-HI techno-trans children [5]

4.2 AI-HI Techo-Trans Youth

The modern youth is also very much habitual of technology, it is almost impossible for them to live or sustain without technology, whether it is Smartphone, digital world, social media platforms (such as facebook, WhatsApp, LinkedIn, Instagram, Online Games, E-Commerce Shopping, They are also loosing emotions, Human Intelligence (HI), and learning capabilities, because, everything is being done through the use of technology, therefore, own intelligence and own emotions are being replaced by the Artificial Intelligence (AI), and machine learning (ML), Data Mining (extraction of valuable information out of the huge data), The youth today is putting minimal efforts to do the things by himself, rather prefers the technology to do the things on behalf on him (Fig. 6).

Fig. 6. AI-HI techo-trans youth [6]

For example, earlier, they used to learn and remember the phone numbers, addresses, paths, routes, land marks (locations), birthdays, anniversaries, concepts, books, authors, knowledge, mathematical tables, calculations, etc., by utilizing the brain capabilities and regular practices, but now, they are dependent on the technology most of the time. Though, the technology provides fast, accurate, and real-time information, but at the same time, the self intelligence (human intelligence) keeps on becoming weaker and weaker by the time, hence the emergence of the 'AI-HI Techno-Trans Youth,' again a big threat to the society in the times to come.

4.3 AI-HI Techno-Trans Robots

On the other side, the technologically created machine, i.e., the 'Robot' has been evolving over the times with the continuous advancements of the technology, artificial intelligence, machine learning, BigData, cloud computing, IoT, etc. The Robots are becoming more powerful and more 'Intelligent,' and are becoming much capable of doing the 'things' in the similar (though, in some situations, much better) way the 'human' does. Those 'AI-HI Techno-Trans Robots' are beneficial to the society if are used in a limited and controlled manner, but, the 'threats' are also waiting if mishandled (Fig. 7).

Fig. 7. AI-HI techno-trans robots [7]

5 Conclusion

5.1 AI-HI Techno-Trans Society

The future of the AI-HI Technological Transgenders community could be imagined wherein these Techno-Trans Children, Techno-Trans Youth, and Techno-Trans Robots would collectively form a Techno-Trans Society enabling the benefits and threats

together. The artificial intelligence (AI) would be increasing rapidly, whereas the human intelligence (HI) might continue to decline as an outcome of the excess use of the technology, and dependency upon it at large. Therefore, the time has come to be aware and awaken. The technology should be used in controlled manner to facilitate the things which are 'essentially' required and need real-time and fast results with better accuracy, and easy approach, however, at the same time, involve the human intelligence while making the decisions, and avoid the AI to control HI, in contrast the HI must lead over the AI, then only a better world of HI society would sustain.

References

1. https://d24fkeqntp1r7r.cloudfront.net/wp-content/uploads/2018/07/08163838/1200.jpg
2. https://ichef.bbci.co.uk/news/976/cpsprodpb/3A40/production/_85321941_robotbrain.jpg
3. https://encryptedtbn0.gstatic.com/images?q=tbn:ANd9GcQSSPK5K30JZcuotOElzEqoN6kEJfGrYRYCidl2ymHmFebpjg0-&s
4. https://www.google.com/search?biw=1440&bih=789&tbm=isch&sa=1&ei=mnDWXZe6Ee-U4-EPurSHyAw&q=negative+impacts+of+excess+use+of+social+media&oq=negative+impacts+of+excess+use+of+social+media&gs_l=img.3…216522.218569..219519…0.0..0.243.1623.0j11j1……0….1..gws-wiz-img.ARILLKsyhEg&ved=0ahUKEwiX77i9lfvlAhVvyjgGHTraAckQ4dUDCAc&uact=5
5. www.google.com
6. https://today.duke.edu/2017/05/more-technology-use-linked-mental-health-issues-risk-adolescents
7. https://ausengcon.com.au/meet-sophia-the-humanoid-robot-that-has-the-world-talking/
8. https://www.britannica.com/technology/artificial-intelligence
9. https://www.britannica.com/science/human-intelligence-psychology
10. Forsyth, C.J., Copes, H.: Encyclopedia of Social Deviance. Sage Publications, p. 740. ISBN 978-1483364698. Archived from the original on 1 Dec 2016. Retrieved 12 Apr 2016 (2014)

Reliability of the Kinovea bidimensional System for Measuring Buccal Movement in a Sample of Italian Adults

Giulia Rossi[1], Giovanni Galeoto[2(✉)], Antonio Amitrano[3], Elisa Leonetti[4], Livia Flammini[4], Marco Tofani[5], and Anna Berardi[2]

[1] Hospital San Giovanni Battista, Rome, Italy
[2] Department of Human Neurosciences, Sapienza University, Rome, Italy
giovanni.galeoto@uniroma1.it
[3] Regional Hospital San Camillo-Forlanini, Rome, Italy
[4] Sapienza Università di Roma, Rome, Italy
[5] Neurorehabilitation Unit, Bambino Gesù Children's Hospital, Rome, Italy

Abstract. The purpose of this study was to assess reliability of Kinovea bidimensional system for measuring buccal movement in frontal or lateral perspectives. This software was used to analyse participants' buccal movements in the anticipatory phase of swallowing. The participants assumed five different types of stimuli diversified on the basis of the volume of the bolus and the size and type of aid (straw or spoon). Thirty-two measurements based on movements of the mouth were recorded using the Kinovea programme. Reliability of 32 measurements was evaluated with Cronbach's alpha and Pearson's correlation coefficient. The results of this study show that the scale has an optimal internal consistency and correlation between items. Thus, using Kinovea software to analyse and study buccal movement can provide clinicians and researchers with important information about the anticipatory phase of swallowing.

Keywords: Assessment · Buccal movement · Kinovea · Technology · Rehabilitation

1 Introduction

The opening of the mouth during the anticipatory phase of swallowing is one of the main components that are activated in the swallowing sequence [1]. In fact, it is hypothesised that the opening movement of the mouth, which occurs in the anticipatory phase, is not only a sensory motor event but also a cognitive event [2, 3]. Before opening the mouth, an individual would make a prediction based on the influence of visual information and memory contents, thus activating, from time to time, the labial movements in a different way based on the instrument with which it is received. Research in recent decades has focused on investigating the hypothesis that there is a relationship between the type of

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_18

stimulus and the degree of mouth opening [4]. This phenomenon can demonstrate the presence of a possible cognitive control behind the swallowing act during the anticipatory phase. In order for these movements to be adequately evaluated, it is necessary to have a valid, simple and easily available technology that can be used by anyone who wants to investigate this task. This study chose the Kinovea software after a careful search for measurement methods that were adequate for its purposes. The researchers chose a two-dimensional evaluation method that used photography or videography to measure the distance between facial points at rest and during movement. On the contrary, three-dimensional evaluations use sophisticated technologies and algorithms and are often more time-intensive, require expensive equipment and are difficult to apply in daily clinical practice.

Kinovea is a free and open-source French software that was created in 2009 as a tool for motion analysis (Kinovea, 0.8.15, Copyright © 2006–2011, Joan Charmant and Contributors). It can be easily found on the Internet or by visiting the manufacturer's website: www.kinovea.org. It is a specialised programme that provides algorithms in the context of computer vision that is for graphic processing that automatically interprets specific elements within digital images or videos. Its functions are aimed at both specialists who study movement and clinical professionals such as physiotherapists, occupational therapists and speech therapists. Once a video has been uploaded to the programme, the software allows the user to measure distances and times manually or via semi-automatic tracking to follow points.

The software was used to analyse buccal movements in a 2015 study aimed at evaluating facial paralysis [4]. The study was related to the two-dimensional movements of the facial musculature aimed at finding a quick and inexpensive method to evaluate facial paralysis. Here, opening was not evaluated between the mouth movements the study examined breath, smile and labial protrusion. This study was useful not only to find a programme suited to the present researchers' purpose, but also because some techniques were extrapolated from it (i.e. marking the points of the face to make the measurements and inserting a reference unit to calibrate the distances). In the present study, however, adhesives were used instead of ink to make the points more visible, of constant size and faster to apply. Furthermore, in order to have reproducible data to make the instrument available at the clinical level, a scale that would allow us the researchers to evaluate the data produced by Kinovea and then extrapolate the results was created [5].

The purpose of the present study is to validate the scale based on the Kinovea programme data. The software is used to analyse the buccal movements of subjects who are not suffering from dysphagia or neurological pathologies in the anticipatory phase of swallowing in relation to the presentation of different amounts of bolus through different tools (straw or spoon).

2 Materials and Methods

The present study was reported by health professionals of Sapienza University of Rome and the Rehabilitation and Outcome Measures Assessment (ROMA) Association [6–19].

For this study, subjects were recruited from universities and hospitals. The inclusion criteria in the reference sample included the absence of dysphagia and neurological deficits aged between 20 and 69 years. All participants were required to sign an informed consent form, by which they authorised the processing of personal data and video recordings for study purposes and denied any allergies to the administered food. Filming was carried out with high-resolution cameras on two smartphones, Samsung A5 and Xiaomi A1 positioned on two tripods in order to guarantee the stability of the video images. Regarding the performance of the test, transparent plastic spoons (filled with varying amounts of jam) and white plastic straws with coloured stripes (to conduct water) were used. Before the start of each film shot, the following was placed on the table in front of the participant: a full glass of water with a straw inside and four spoons of varying sizes filled with different amounts of jam. The amount of jam was kept constant in all the experiments by using measuring spoons: the first spoon (small) contained 1.25 ml of jam, the second spoon (small) contained 5 ml of jam, the third spoon (large) contained 2.5 ml of jam and the fourth spoon (large) contained 10 ml of jam.

Subsequently, stickers were affixed to the participant's face to serve as reference points for subsequent measurements of buccal movements. Two yellow stickers that were 3 cm in length served as the reference sample unit; these were attached in a vertical position: the first was affixed to the left cheek, visible from a lateral perspective; the second was placed on the right cheek, visible from a frontal perspective.

Four circular reflective stickers were also used and placed as reported in Figs. 1 and 2. The first was placed on the left corner of the mouth (point A), and the second on the right corner of the mouth (point B) at the junctions between the upper and lower lips. The third was placed in correspondence to the Cupid's arch of the lip (i.e. above the middle part of the upper lip or point C). The fourth was placed at the opposite point, below the middle part of the lower lip (point D).

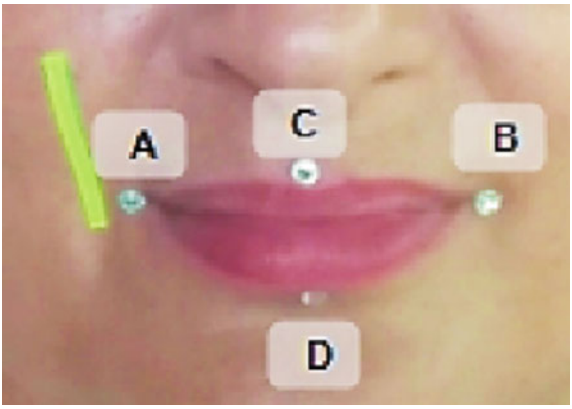


Fig. 1. Frontal positioning of stickers

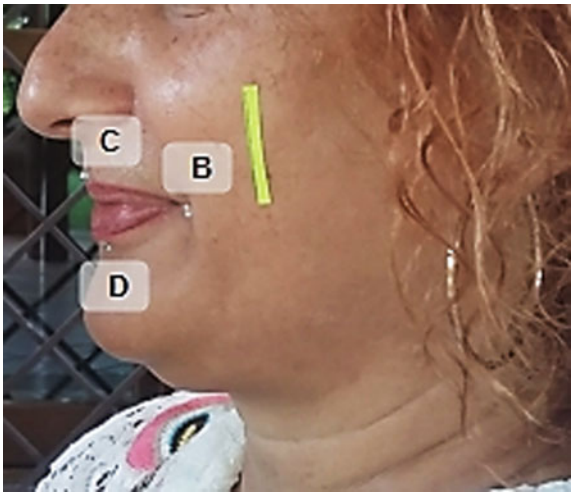


Fig. 2. Lateral positioning of stickers

2.1 Procedures

The subjects were seated at a table with two trestles, one in front of the subject above the table and the other in a lateral position with respect to the subject. The smartphones were placed on the two trestles in order to make video recordings from two different angles. The four reflective stickers applied to the participants' faces were necessary as reference points for the subsequent video analysis. The participants were given precise instructions about how to perform the test before beginning the recordings: drink the water with the straw and then eat the jam contained in the four spoons following the order in which they are positioned on the table (small spoon with less jam, small spoon with more jam, large spoon with less jam, large spoon with more jam).

Moreover, during the bolus intake, the participants were asked to keep their heads in a straight and stable position and to keep the spoon handle in a distal position (a technique designed to avoid covering the mouth during bolus intake). No further indications or explanations were provided with respect to the objectives of the study to avoid conditioning the participants. Video recording began, and the participants were asked to perform the task. The recording concluded when the individual had eaten the jam from all four spoons.

2.2 Kinovea Software

Following the test, the two videos taken from the front and side positions were uploaded to the computer and analysed using the Kinovea software. The programme allows users to analyse video, measuring distances manually or using a semi-automatic tracking system to follow points or trajectories. To carry out the measurements, the video was stopped at moments of initial rest and maximum opening of the mouth and a still image was obtained. Measurements were taken by tracing the image of each line to connect points of interest (indicated by the self-refracting round stickers).

Specifically, the following measurements were made from the frontal video recordings:

- The resting condition of the mouth, the distances AB (horizontal length between the lips), CD (vertical length between the lips), BC and BD.
- The same distances at the time of maximum opening of the mouth during the intake of jam from the four spoons.

The following measurements were made from the lateral video recordings, considering reference points BCD as reported in Fig. 2:

- The distances BC and BD in the resting condition of the mouth
- The same distances at the time of maximum opening of the mouth while drinking water with the straw
- The same distances at the time of maximum opening of the mouth (during jam intake from the four spoons).

2.3 Analysis of Data

Following the video recording and analysis via Kinovea, measurement from both video angles was transcribed in an excel file. The psychometric analysis was performed using the IBM SPSS 25.00 software programme, and the significance level of results was set for p value less than or equal to 0.05. The correlations between the measurements taken at rest and at opening with all the stimuli were analysed via Pearson's correlation coefficient. In addition, the internal consistency of all measurements recorded was evaluated using Cronbach's alpha.

3 Results

A total of 100 people took part in the study, mean age $\pm$ standard deviation was 44 ± 15, and 50% of participants were male. All 100 included individuals signed the informed consent and were evaluated following the predetermined procedures. Specifically, participants assumed five different types of stimuli diversified on the basis of the volume of the bolus and the size and type of aid (straw or spoon). Thirty-two movements of the mouth at rest and at opening with all the stimuli were recorded using the Kinovea programme. An example of measurement is reported on Figs. 3 and 4.

Results shows that, as reported in Table 1 for frontal measurements and Table 2 for lateral measurements, all data recorded by the software are correlated between each other with a $p < 0.01$.

A reliability analysis of the general scale was performed, and Cronbach's alpha resulted with a value of 0.973 for all 32 measurements; moreover, the value of alpha was 0.891 for frontal projection (20 measurements) and 0.941 for measurements on the lateral projection (12 measurements).

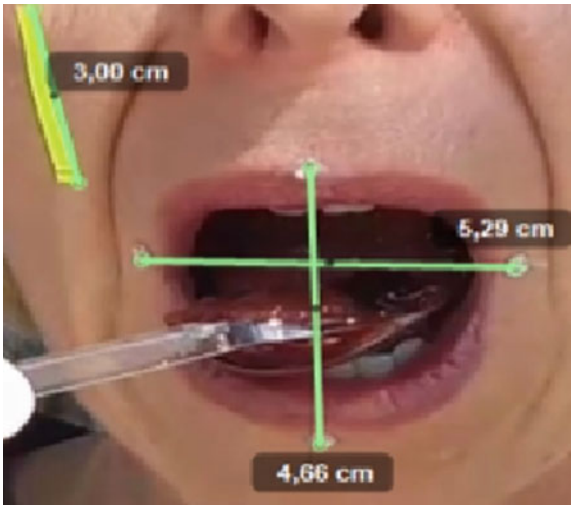


Fig. 3. Lateral measurement

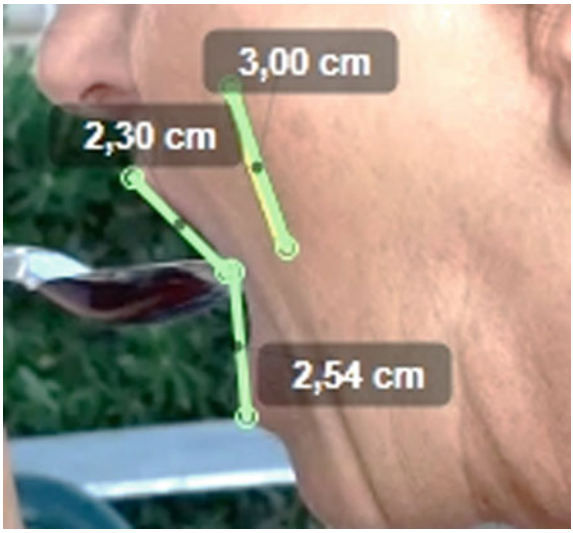


Fig. 4. Frontal measurement

4 Discussion

This study attempted to deepen the analysis of the anticipatory phase of swallowing, verifying whether the variations in the typology and size of the aid and in the volume of the bolus influenced labial movements. Analysing the buccal movements in the anticipatory phase of swallowing is fundamental to demonstrating the possible involvement of a cognitive component. A study was initially carried out on a healthy sample of 100

Table 1. Pearson's correlation coefficient between frontal measurements recorded by Kinovea software

Item	Rest AB (item 1)	Rest CD (item 2)	Rest CB (item 3)	Rest DB (item 4)	Teaspoon 2.5 ml AB (item 5)	Teaspoon 2.5 ml CD (item 6)	Teaspoon 2.5 ml CB (item 7)	Teaspoon 2.5 ml DB (item 8)	Teaspoon 10 ml AB (item 9)	Teaspoon 10 ml CD (item 10)	Teaspoon 10 ml CB (item 11)	Teaspoon 10 ml DB (item 12)	Spoon 2.5 ml AB (item 13)	Spoon 2.5 ml CD (item 14)	Spoon 2.5 ml CB (item 15)	Spoon 2.5 ml DB (item 16)	Spoon 10 ml AB (item 17)	Spoon 10 ml CD (item 18)	Spoon 10 ml CB (item 19)
2	0.36*																		
3	0.82*	0																	
4	0.88*	0.33*	0.78*																
5	0.83*	0.37*	0.70*	0.73*															
6	0.60*	0.51*	0.55*	0.49*	0.69*														
7	0.68*	0.41*	0.75*	0.59*	0.76*	0.76*													
8	0.76*	0.31*	0.61*	0.75*	0.80*	0.74*	0.66*												
9	0.82*	0.37*	0.69*	0.77*	0.95*	0.71*	0.74*	0.80*											
10	0.58*	0.56*	0.52*	0.43*	0.68*	0.90*	0.72*	0.66*	0.73*										
11	0.64*	0.44*	0.65*	0.56*	0.71*	0.73*	0.92*	0.61*	0.72*	0.79*									
12	0.77*	0.40*	0.65*	0.72*	0.83*	0.71*	0.68*	0.85*	0.88*	0.79*	0.70*								
13	0.79*	0.37*	0.69*	0.69*	0.87*	0.61*	0.67*	0.74*	0.91*	0.63*	0.66*	0.80*							
14	0.54*	0.51*	0.51*	0.45*	0.62*	0.86*	0.70*	0.65*	0.65*	0.88*	0.75*	0.74*	0.60*						
15	0.61*	0.45*	0.67*	0.56*	0.69*	0.70*	0.87*	0.62*	0.70*	0.73*	0.91*	0.68*	0.71*	0.80*					
16	0.72*	0.35*	0.63*	0.70*	0.78*	0.70*	0.64*	0.83*	0.82*	0.70*	0.65*	0.88*	0.84*	0.76*	0.72*				
17	0.77*	0.33*	0.64*	0.65*	0.87*	0.63*	0.67*	0.74*	0.90*	0.65*	0.64*	0.77*	0.90*	0.59*	0.65*	0.75*			
18	0.55*	0.60*	0.54*	0.46*	0.62*	0.83*	0.67*	0.63*	0.65*	0.88*	0.73*	0.71*	0.60*	0.92*	0.76*	0.71*	0.60*		
19	0.57*	0.47*	0.63*	0.53*	0.65*	0.67*	0.86*	0.56*	0.68*	0.72*	0.91*	0.64*	0.66*	0.74*	0.92*	0.65*	0.68*	0.75*	
20	0.74*	0.42*	0.64*	0.71*	0.79*	0.72*	0.65*	0.84*	0.83*	0.73*	0.66*	0.90*	0.81*	0.76*	0.69*	0.92*	0.82*	0.78*	0.69*

Item 20 Spoon 10 ml DB; *Significant correlation $p < 0.01$ (two-tailed)

Table 2. Pearson's correlation coefficient between lateral measurements recorded by Kinovea software

Item	Rest BC (item1)	Rest BD (item2)	Straw BC (item3)	Straw BD (item4)	Teaspoon 2.5 ml BC (item5)	Teaspoon 2.5 ml BD (item6)	Teaspoon 10 ml BC (item7)	Teaspoon 10 ml BD (item8)	Spoon 2.5 ml BC(item9)	Spoon 2.5 ml BD (item10)	Spoon 10 ml BC (item11)
2	0.80*										
3	0.62*	0.51*									
4	0.44*	0.59*	0.61*								
5	0.54*	0.38*	0.50*	0.28*							
6	0.43*	0.55*	0.40*	0.55*	0.57*						
7	0.60*	0.44*	0.53*	0.27*	0.90*	0.58*					
8	0.45*	0.50*	0.44*	0.50*	0.61*	0.85*	0.64*				
9	0.57*	0.44*	0.44*	0.19	0.86*	0.53*	0.89*	0.58*			
10	0.40*	0.50*	0.40*	0.54*	0.54*	0.86*	0.56*	0.90*	0.54*		
11	0.51*	0.37*	0.41*	0.20*	0.86*	0.57*	0.90*	0.64*	0.89*	0.59*	
12	0.42*	0.47*	0.37*	0.46*	0.53*	0.79*	0.56*	0.82*	0.54*	0.86*	0.61*

Item 12 Spoon 10 ml BD; *Significant correlation $p < 0.01$ (two-tailed)

subjects, recording video while taking different boluses. Subsequently, the videos were analysed using Kinovea—a software that enables the measurement of the opening levels of the mouth during the anticipatory phase (that is, before the bolus has effective contact with the lips). Finally, a statistical analysis of these measurements was carried out.

In order for these movements to be adequately evaluated, it is necessary to have available a valid, simple and easily available technology that can be used by anyone. The Kinovea programme meets these characteristics and is configured as a reliable tool for experimental investigations. This study created a general scale that included all the items from the data collection; it has a very high reliability value. Indeed, the Cronbach's alpha coefficient, a general index for the verification of internal consistency defining reproducibility over time under the same conditions, was equal to 0.973. The general scale was then divided into two subscales—one that considered all the measurements carried out in the lateral recordings (Cronbach's alpha = 0.941), and one that included those taken from the frontal perspective (Cronbach's alpha =0.978). Both were confirmed to be highly reliable.

The Kinovea programme has already been used for the analysis of facial movements, but not buccal movements, on a sample of ten people. This study, which is based instead on a sample of 100 individuals, presents statistically significant results. It therefore validates the experimental reliability of the software, whose algorithms can be effectively used to evaluate the buccal movements of the anticipatory phase of swallowing.

It is necessary to highlight some functional limits of the study: although it tried to simulate a natural feeding situation, the setting was not completely natural. In fact, participants were given different instructions to reduce their head and arm movements, and the presence of stickers on their faces probably interfered with any spontaneous opening of the mouth. Therefore, although these modifications were necessary for a better performance, they could also have had an impact on the results. Furthermore, the instruction to hold the teaspoons and spoons at the most distal part of the handle, a type of grasping that is not normally evident when eating with such instruments, could partially affect the naturalness of the act. Secondly, the reflective stickers tended to come off during the test, making it necessary to reapply them and begin the experiment again. Finally, despite the indications received, some of the participants in the trial tended to completely cover their mouths when bringing the spoons to their lips, making their mouths not visible in the frontal projection. It was necessary to repeat the experiment in this case.

The Kinovea software was used to evaluate buccal movements in the anticipatory phase of swallowing to determine whether there were differences in oral opening depending on the intake of different amounts of bolus through different instruments of various sizes. The validity of the Kinovea programme for this purpose was investigated based on a very large sample: the results obtained were statistically significant, highlighting the programme's high reproducibility and reliability. Kinovea can be used by any professional working in the field of speech therapy—it is 'user friendly' and does not require specific training to exploit its potential, but it is sophisticated enough to be used in professional research. Hopefully, the present study has paved the way for future investigations on subjects with brain dysfunctions to verify the impact of these aspects on pathology (including through tests with functional magnetic resonance, which were not used in this study due to structural limitations).

References

1. Brodsky, M.B., Verdolini Abbott, K., McNeil, M.R., et al.: Effects of divided attention on swallowing in persons with idiopathic Parkinson's disease. Dysphagia **27**(3), 390–400 (2012)
2. Namasivayam-MacDonald, A.M., Steele, C.M., Keller, H.H.: Perception versus performance of swallow function in residents of long-term care. Am. J. Speech Lang. Pathol. **28**(3), 1198–1205 (2019)
3. Yatabe, N., Takeuchi, K., Izumi, M., et al.: Decreased cognitive function is associated with dysphagia risk in nursing home older residents. Gerodontology **35**(4), 376–381 (2018)
4. Carter, J.A., Hyland, C., Steele, R.E., Collins, E.M.: Dynamics of mouth opening in hydra. Biophys. J. **110**(5), 1191–1201 (2016)
5. Baude, M., Hutin, E., Gracies, J.M.: A bidimensional system of facial movement analysis conception and reliability in adults. Biomed. Res. Int. **2015**, 812961 (2015)
6. Berardi, A., Biondillo, A., Màrquez, M.A., et al.: Validation of the short version of the Van Lieshout Test in an Italian population with cervical spinal cord injuries: a cross-sectional study. Spinal cord **57**(4), 339 (2019)
7. Galeoto, G., Scialpi, A., Grassi, M.L., et al.: General sleep disturbance scale: translation, cultural adaptation, and psychometric properties of the Italian version. Cranio **2019**, 1–9 (2019)
8. Savona, A., Ferralis, L., Saffioti, M., et al.: Evaluation of intra-and inter-rater reliability and concurrent validity of the Italian version of the Jebsen-Taylor hand function test in adults with rheumatoid arthritis. Hand Ther. **24**(2), 48–54 (2019)
9. Berardi, A., Dhrami, L., Tofani, M., et al.: Cross-cultural adaptation and validation in the Italian population of the wolf motor function test in patients with stroke. Funct. Neurol. **33**(4), 229–253 (2018)
10. Dattoli, S., Colucci, M., Soave, M.G., et al.: Evaluation of pelvis postural systems in spinal cord injury patients: Outcome research. J. Spinal Cord Med. **2018**, 1–13 (2018)
11. Galeoto, G., Iori, F., De Santis, R., et al.: The outcome measures for loss of functionality in the activities of daily living of adults after stroke: a systematic review. Top Stroke Rehabil. **26**(3), 236–245 (2019)
12. Ruggieri, M., Palmisano, B., Fratocchi, G., et al.: Validated fall risk assessment tools for use with older adults: a systematic review. Phys. Occup. Ther. Geriatr. **36**(4), 331–353 (2018)
13. Tofani, M., Galeoto, G., Cazzetta, D., Berardi, A., Sansoni, J., Valente, D.: Validation of the pediatric evaluation of disability inventory in an Italian population with autism spectrum disorder: a cross-sectional study. La Clinica Terapeutica **170**(6), e460–e464 (2019)
14. Berardi, A., Galeoto, G., Guarino, D., et al.: Construct validity, test-retest reliability, and the ability to detect change of the Canadian occupational performance measure in a spinal cord injury population. Spinal Cord Ser. Cases **5**(1), 52 (2019)
15. Puledda, F., Viganò, A., Celletti, C., et al.: A study of migraine characteristics in joint hypermobility syndrome aka Ehlers-Danlos syndrome, hypermobility type. Neurol. Sci. **36**(8), 1417–1424 (2015)
16. Galeoto, G., Sansoni, J., Valenti, D., et al.: The effect of physiotherapy on fatigue and physical functioning in chronic fatigue syndrome patients: a systematic review. La Clinica Terapeutica **169**(4), e184–e188 (2018)
17. Nobilia, M., Culicchia, G., Tofani, M., De Santis, R., Savona, A., Guarino, D., et al.: Italian version of the Jebsen–Taylor hand function test for the assessment of hand disorders: a cross-sectional study. Am. J. Occup. Ther. **73**(3), 7303205080p1–7303205080p6 (2019)
18. Tofani, M., Candeloro, C., Sabbadini, M., et al.: A study validating the Italian version of the level of sitting scale in children with cerebral palsy. Clin. Rehabil. **33**(11) (2019)

19. Romagnoli, G., Leone, A., Sansoni, J., Tofani, M., De Santis, R., Valente, D., et al.: Occupational therapy's efficacy in children with Asperger's syndrome: a systematic review of randomized controlled trials. La Clinica Terapeutica **170**(5), e382–e387 (2019)

A Technique to Analyze a Cassandra Tombstone for Identification of Operations

Vinod Pachghare(✉) and Rupali Chopade

College of Engineering Pune, Savitribai Phule Pune University (SPPU), Wellesely Rd, Shivajinagar, Pune, Maharashtra 411005, India
{vkp.comp,rmc18.comp}@coep.ac.in

Abstract. High scalability, flexible storage of data, linear performance, and many other strong features made Cassandra database a popular database. The Cassandra is one of the column-based NoSQL databases. Cassandra performs faster write operations with an additional overhead of tombstone. Whenever logical delete operation is performed, it will result in a tombstone. This tombstone value is useful to update the other replica about the deletion. The major problem with tombstone is that it is not only generated with delete operation but there are other factors also which generate this tombstone. This paper is an effort to identify the key of deleted data and other reasons for tombstone generation. The technique is proposed for the same. This paper is the first attempt to work on Cassandra's tombstone.

Keywords: Cassandra · Tombstone · Delete · Insert · Update · Null

1 Introduction

NoSQL databases are needed for Internet-dependent age and these databases are widely adopted by top organizations [1]. Due to their popularity, many universities have introduced NoSQL databases as part of their curriculum [2]. The popularity of anything introduces security challenges. As sensitive and private information is stored in the databases, it is important to secure the database. The idea presented through this paper will be helpful to identify which record is deleted. As per DB-Engines ranking [3], Cassandra stands at third position among NoSQL databases and at tenth position among overall databases. The ranking of the Cassandra database and this kind of research work not performed yet on the mentioned database are the main motivational factors for selecting this database. Cassandra is a distributed database that processes an enormous amount of data very rapidly due to its time-stamping functionality [4]. Cassandra is a column-based database, which stores related data in the form of columns called a column family. Cassandra database is suitable for applications when schema frequently changes; a large number of columns are available and data is in the scattered form [1]. Through this

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_19

paper, we are focusing on one of the important write operations in the database which is 'delete' operation. Whenever logical deletion is performed, it will create a tombstone. Tombstone is a special value used as a marker to indicate that this record is now deleted from the database [5]. The proposed technique is the first attempt to work on Cassandra's tombstone to correctly identify the operations performed.

The rest of the paper is organized as follows. Section 2 throws a light on the related work. Cassandra's internal architecture working for a write operation is explained in Sect. 3. The proposed technique working algorithm, evaluation, and performance analysis are presented in Sect. 4. Finally, we conclude the paper.

2 Related Work

Any specific research work is not happened on Cassandra's tombstone as per Scopus' results [6]. This search is refined with Cassandra and data recovery, then hardly five documents are searched. So this section specifies the work done related to the Cassandra database and not exactly related to Cassandra's tombstone.

Lakshman et al. explained [7] the storage structure of the Cassandra database. This database data model stores the data in columnar format. Columns are arranged together in column families. Cassandra architecture is clarified by covering the partitioning technique, replication, bootstrapping, membership, persistence, and implementation details. The use of Cassandra in the Facebook application has benefited a lot.

Rendezvous hashing is one of the partitioning techniques used for Cassandra. Data partitioning is the critical task for NoSQL databases as compared to relational databases. Sally et al. proposed Map Reduce Rendezvous hashing [8] which improves the performance by a hierarchical division of a system into the Rendezvous region, load balancing and monitoring, and enhancement in the timing of hashing process.

For large-scale database applications, multidimensional range queries are a basic requirement and it is available mostly for distributed ordered tables. Cassandra is also a distributed ordered table. Chen Feng et al. proposed complemental clustering index (CCI) for range queries [9] which gives better performance for distributed ordered tables and also shows that Cassandra database gives maximum efficiency with hash tables than distributed ordered tables for the range queries.

Roy Friedman et al. proposed [10] a protocol for the Cassandra database against Byzantine failure. Byzantine fault is generally observed in distributed systems, in which observers find different reasons for failure. The proposed protocol is a combination of MAC and elliptic curve digital signature algorithm. The performance is benchmarked using Yahoo! Cloud Serving Benchmark (YCSB).

Rahman et al. proposed a mechanism [11] for the recovery of PostgreSQL and Cassandra applications. It works for centralized and distributed versions of the Cassandra database. The proposed mechanism is experimented on crash-consistent snapshots of virtual machines (VMs), first with a single VM and then with two VMs. Cassandra commit log consists of write-ahead log (WAL) entries, which is useful for the same.

3 Cassandra Write Architecture and Working

Whenever Cassandra write is executed [12], it stores the data in memtable, which is a memory-resident table. Along with memtable, it also stores this data in the commit log, which is available on disk. The data is stored in the sorted form in memtable. Data is available in memtable until a certain limit called a threshold. When the threshold is reached, data is flushed to SSTable. The data is automatically flushed from memtable to SSTable on disk. To manually flush the data, a nodetool command is used. To fetch the data from sstabledump utility, it is necessary to flush the data. The sstabledump utility works by fetching the data from sstable and displays the contents in the JSON form. The memtable and SSTable are available per table while the commit log is shared among tables. The overall working is shown in Fig. 1.

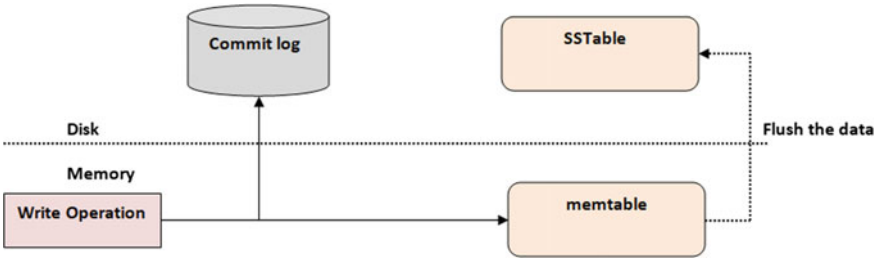


Fig. 1. Cassandra write data operation flow

The SSTable creates the files as shown in Fig. 2, which are generally available under the path 'C:\apache-cassandra-3.11.4\data\data\mydb\books-8a9eda300f5311ea8764d77b1200b03f'. Here, mydb is a keyspace name, books is a table name followed by a 32 digit number generated as a unique identifier for the table.

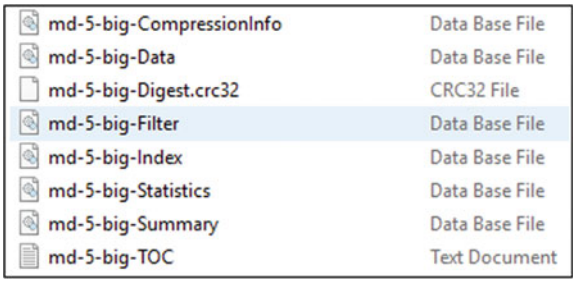


Fig. 2. SSTable structure

Table 1 gives an idea about all files from Fig. 2.

The actual SSTable data is available under Data.db file. Tombstone is designed basically for the delete operation. The tombstone is available for a certain period of time called gc_grace_seconds [13]. The default value of the same is 864,000, which is

Table 1. SSTable file contents

File Name	Functionality
Data.db	The actual SSTable data
CompressionInfo.db	Compression information
Digest-crc32	The checksum of a data file
Filter.db	Data is first checked in memory before checking it on disk.
Index.db	Index to a key which points to the data position
Statistics.db	Statistical information about SSTable contents
Summary.db	Details about partition index
TOC.db	List of all components information (table of contents)

near about ten days. After the time is expired, a tombstone will get deleted. Related to tombstone, there are other operations also, which generate tombstones, as listed below [14].

Case 1—When delete operation is performed
Case 2—When a null value is inserted using insert/update operation
Case 3—When an update operation is performed
Case 4—Use of materialized view
Case 5—Data exploration with time-to-live (TTL).

Among the list, the proposed technique is able to correctly identify the first three cases and these are explained in Sect. 4.

4 Proposed Technique

The proposed technique is designed using Python and experimented on Windows 10 64-bit OS, 32 GB RAM with Intel Xeon 3.60 GHz processor. Cassandra database version used is 3.11.4. The working of the technique is shown in algorithm 1. As per algorithm, the sstabledump file is used to load the contents of a tombstone from database file Data.db. The path is traversed to locate the Data.db files, available under the data folder of Cassandra folder. Whenever a keyspace is created, it will create a folder named with keyspace. This folder will create other folders with the names of tables as per their creation, followed by a unique identification number for the table. The algorithm works by extracting a specific keyword from Data.db files. These details are initially stored in a CSV file named 'cassandra_tomb.csv'. Then a specific keyword is searched through row number and specific operation is identified.

4.1 Evaluation

The evaluation of the proposed technique is given with the cases mentioned in Sect. 3. This technique will display the list of all available keyspaces.

Algorithm 1: Identification of cassandra tombstone operation types

```
Input: Cassandra Keyspace and Table name
Output: Distinguished results for deleted record, updated record and
        inserted record with null value
Change path directory to location of cassandra keyspace
Print all directories from above specified path /* Which is actually
  keyspace names                                                  */
ipt_ks=input('Enter keyspace to check:') /* Ask user to input Keyspace
  */
Change directory path to user selected keyspace
Print all directories from above specified path /* Which is actually table
  names                                                           */
ipt_tab=input('Enter table to check (Copy the entire table name from
above list):')
          // Ask user to input Table name to observe for changes, if any
Change directory path to user selected Table name
Open CSV file to store the above table dump data
for file in glob.glob("*-Data.db") do
                                  // Which actually contains the SSTable data
    Change directory path to sstabledump file
    myCmd = os.popen('sstabledump'+' ' +path1).read()
     // Execute the sstabledump command with selected table's Data.db file
    for line in myCmd do
        Write the result of above command to CSV file
Close the CSV file
print('Keyspace and Table is", ipt_ks+"."+ipt_tab[:-33])
Open the CSV file with reader
for num,row in enumerate(reader) do
    if key in row[0] then
        new_key=(str(row))
    if ('deletion_info' and 'marked_deleted') in row[0] then
        new_info=str(row)
        print('This record is deleted from above table on -
        ',new_info[47:73],'for -',new_key[8:25])
    if ('deletion_info' and 'marked_deleted') in row[1] then
        new_info=str(row)
        print('This record is deleted from above table on -
        ',new_info[47:73],'for -',new_key[8:25])
    if 'deletion_info' in row[1] and 'marked_deleted' not in row[0]
     then
        new_info1=str(row)
        print('This record is inserted with null in above table with -
        ',new_info1[15:100],'for -',new_key[8:25])
    if 'tstamp' in row[2] then
        up_value=(str(row[1]))
        up_time=(str(row[2]))
        print('This record is updated in above table -
        ','with',up_value,'for ',new_key[8:25],'on',up_time)
```

Then, the user will select the keyspace. After selecting the keyspace name, the list of all tables available under selected keyspace will get displayed. The system will ask the user to select the table name. After selecting the table name, automatically all Data.db

files will be extracted, which contain the JSON data generated from sstabledump utility. Then the system will correctly identify which operation is performed. The different operations identified by the system are evaluated as different cases and are given here.

Case 1: When delete operation is performed
This case will identify which record is deleted from a selected table corresponding to the primary key, along with the timestamp of operation.

```
['coep', 'ks1', 'mydb', 'system', 'system_auth', 'system_distributed', 'system_schema', 'system_traces', 'testkeyspace']

Enter keyspace to check:mydb
mydb
['books-8a9eda300f5311ea8764d77b1200b03f', 'books_info-b12123b00f5411ea8764d77b1200b03f', 'employee-f33760400b6011ea81c2db82293a2e0d']

Enter table to check (Copy the entire table name from above list):books_info-b12123b00f5411ea8764d77b1200b03f
books_info-b12123b00f5411ea8764d77b1200b03f
md-1-big-Data.db

Keyspace and Table is mydb.books_info

This record is deleted from above table on -  "2019-11-25T07:28:04.729Z" for  - "key" : [ "987" ]
```

Case 2: When insert/update operation is performed with a null value
When insert/update operation is performed by inserting a null value, this case will find out for which field null value is inserted, along with primary key value and timestamp of operation.

```
['coep', 'ks1', 'mydb', 'system', 'system_auth', 'system_distributed', 'system_schema', 'system_traces', 'testkeyspace']

Enter keyspace to check:mydb
mydb
['books-8a9eda300f5311ea8764d77b1200b03f', 'books_info-b12123b00f5411ea8764d77b1200b03f', 'employee-f33760400b6011ea81c2db82293a2e0d']

Enter table to check (Copy the entire table name from above list):books-8a9eda300f5311ea8764d77b1200b03f
books-8a9eda300f5311ea8764d77b1200b03f
md-1-big-Data.db

Keyspace and Table is mydb.books

This record is inserted with null in above table with -  name" : "price"', ' "deletion_info" : { "local_delete_time" : "2019-11-25T07:17:43Z"  for  - "key" : [ "124" ]
```

Case 3: When an update operation is performed
This case will identify what value is updated, along with the primary key value and timestamp of operation for an update.

```
['coep', 'ks1', 'mydb', 'system', 'system_auth', 'system_distributed', 'system_schema', 'system_traces', 'testkeyspace']

Enter keyspace to check:Ks1
['employee-252e63a0e42911e9a6ff3f69359e5473', 'my_table-d093eb900cf411eab8610d4e631b3776', 'student-7120cac00f5611eaa3c70584ee8b2f65']

Enter table to check (Copy the entire table name from above list):student-7120cac00f5611eaa3c70584ee8b2f65
student-7120cac00f5611eaa3c70584ee8b2f65
md-1-big-Data.db

Keyspace and Table is Ks1.student

This record is updated in above table -  with  "value" : "SE IT" for  "key" : [ "72" ]' on  "tstamp" : "2019-11-25T07:39:27.634Z" }
```

Case 4: When Case 1 to Case 3, all cases are available
This case will differentiate Case 1 to Case 3 correctly.

```
['coep', 'ks1', 'mydb', 'system', 'system_auth', 'system_distributed', 'system_schema', 'system_traces', 'testkeyspace']

Enter keyspace to check:coep
coep
['emp-339666000c4611ea9d01c5e5855bfd3d', 'employee-cf71d400ea6911e986ff5de7fef2e3ba', 'encrypted_keys-af44bb000b8311eaa6abcb205041a7a2']

Enter table to check (Copy the entire table name from above list):employee-cf71d400ea6911e986ff5de7fef2e3ba
employee-cf71d400ea6911e986ff5de7fef2e3ba
md-5-big-Data.db
md-6-big-Data.db
md-7-big-Data.db

Keyspace and Table is coep.employee

This record is deleted from above table on -  "2019-11-21T06:22:07.425Z" for  - "key" : [ "261" ]

This record is inserted with null in above table with -  name" : "ephone"', ' "deletion_info" : { "local_delete_time" : "2019-11-21T07:20:17Z" for  - "key" : [ "111" ]

This record is inserted with null in above table with -  name" : "ephone"', ' "deletion_info" : { "local_delete_time" : "2019-11-21T07:32:27Z" for  - "key" : [ "222" ]

This record is deleted from above table on -  "2019-11-21T06:09:20.480Z" for  - "key" : [ "121" ]

This record is deleted from above table on -  "2019-11-21T07:44:19.592Z" for  - "key" : [ "111" ]

This record is updated in above table -  with  "value" : "ABC" for  "key" : [ "737" ] on  "tstamp" : "2019-11-22T10:22:39.852Z" }
```

4.2 Performance Analysis

The execution time analysis to identify the type of operation is shown in Fig. 3 with a graph.

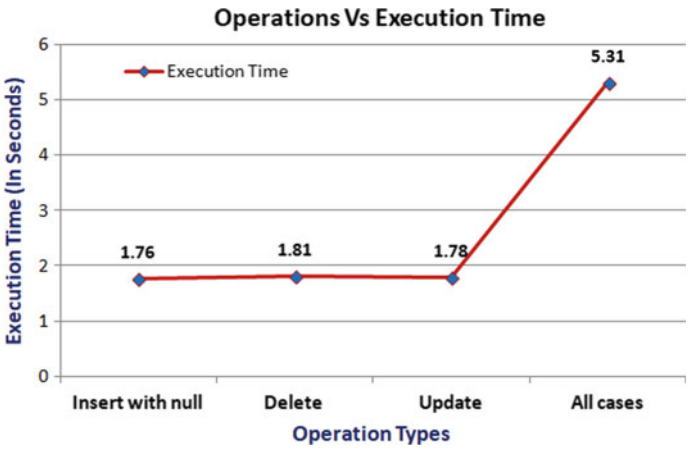


Fig. 3. Execution time analysis

It can be observed from the graph that, the execution time required in identifying insert, updating, and delete is almost the same. It means when only one type of case is present in the specified table, then the time required is less. The time required is more when the combination of all operations or more than one case is available. The execution time is recorded by taking five readings for each case and the average time is calculated. The time and space complexity of algorithm1 are O(n) and constant, respectively.

5 Conclusion

To successfully recover the data from any database, it is important to first find out which data is deleted. Finding deleted data is also a critical task. In the Cassandra database, the tombstone is a special feature useful for the delete operation. The tombstone will get generated for many other operations also (explained in Sect. 3). As more numbers of tombstones are generated, it may affect the overall performance. To deal with performance issues, autocompaction of a tombstone can be useful. Through this paper, we have tried to differentiate operations for which tombstone is generated. The limitation of this technique is that it does not work for materialized view and data exploration. The future work of this study is to focus on the limitation as well as to restore the deleted data with the help of a tombstone.

References

1. Kaur, K., Rani, R.: Modeling and querying data in NoSQL databases. In: 2013 IEEE International Conference on Big Data, pp. 1–7 (2013)
2. Grossniklaus, M., Maier, D.: The curriculum forecast for Portland: cloudy with a chance of data. ACM Sigmod Rec. **41**, 74–77 (2012)
3. DB-Engines Ranking—popularity ranking of database management systems. https://db-engines.com/en/ranking
4. Apache Cassandra. http://cassandra.apache.org/
5. About Deletes and Tombstones in Cassandra. https://thelastpickle.com/blog/2016/07/27/about-deletes-and-tombstones.html
6. Scopus preview—Scopus—Welcome to Scopus. https://www.scopus.com/home.uri
7. Lakshman, A., Malik, P.: Cassandra: a decentralized structured storage system. ACM SIGOPS Oper. Syst. Rev. **44**, 35–40 (2010)
8. Elghamrawy, S.M., Hassanien, A.E.: A partitioning framework for Cassandra NoSQL database using Rendezvous hashing. J. Supercomput. **73**, 4444–4465 (2017)
9. Feng, C., Zou, Y., Xu, Z.: Ccindex for cassandra: a novel scheme for multi-dimensional range queries in cassandra. In: 2011 Seventh International Conference on Semantics, Knowledge and Grids, pp. 130–136 (2011)
10. Friedman, R., Licher, R.: Hardening cassandra against byzantine failures. arXiv Prepr arXiv161002885 (2016)
11. Rahman, M.R., Piduri, S., Languev, I., et al.: Software-defined consistency group abstractions for virtual machines. In: Proceedings of the 4th Workshop on Distributed Cloud Computing, p. 3 (2016)
12. How is data written? | Apache Cassandra™ 3.0. https://docs.datastax.com/en/cassandra/3.0/cassandra/dml/dmlHowDataWritten.html
13. How is data deleted? | Apache Cassandra™ 3.0. https://docs.datastax.com/en/cassandra/3.0/cassandra/dml/dmlAboutDeletes.html
14. What are tombstones? | DSE 6.0 Arch guide. https://docs.datastax.com/en/dse/6.0/dse-arch/datastax_enterprise/dbInternals/archTombstones.html

Computer-Assisted Self-Training for Kyudo Posture Rectification Using Computer Vision Methods

Wardah Farrukh and Dustin van der Haar(✉)

Academy of Computer Science and Software Engineering, University of Johannesburg, Cnr University Road and Kingsway Avenue, APK Campus, Johannesburg, South Africa
wardah_farrukh@yahoo.com, dvanderhaar@uj.ac.za

Abstract. To some individuals, particularly archery students, perfecting the art of Kyudo is of utmost importance. These devoted students are always trying to correct their posture because it plays a significant role in effectively shooting at the target. However, due to the lack of attention from instructors, students are often forced to train on their own without any guidance. It is difficult for students to analyze their own faults because the shoulders, hips, and feet should be in line with another, parallel to the floor and straight to the target. The proposed solution is, therefore, a system that aims to assist students in correcting their posture. The system will classify the technique presented by the user and using PoseNet, the system will output coordinates and draw a skeleton structure of the user's technique along with the instructor's technique. The coordinates will then be measured for similarity and appropriate feedback is provided to the user. The results for classification, using CNN and SVM showed an accuracy of 81.25% and 80.2%, respectively. The results indicate the feasibility of the approach, however, improvement is required in certain areas. Recommendations for improving the approach are discussed.

Keywords: Computer vision · Convolutional Neural Network · Support vector machine · PoseNet

1 Introduction

Kyudo, which means "the way of the bow", is an ancient Japanese sport which gained its popularity in 1949 [1]. There are eight techniques in Kyudo [2] which are: (1) AshiBumi-positioning of the foot, (2) Dozukuri-positioning of the torso, (3) Yugamae-readying the body, (4) Uchiokoshi-lifting the bow, (5) Hikiwake-drawing the bow, (6) Kai-full draw, (7) Hanare-releasing the bow and (8) Zanshin-maintaining the form. Each of these techniques requires the correct posture. In his research article on Japanese Archery, Cohen clearly states that Kyudo requires aligning the hands and feet in the correct

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_20

position before aiming at the target [3]. He also mentions that the body should be so stable that it becomes an anchor [3]. According to the Kyudo International Federation, the archer is not only judged by hitting or missing the target but how the techniques and the process of shooting are carried out. Hence, the proposed system will approach this problem with the intention of facilitating and improving the training sessions of Kyudo archers. Due to limited research and work done in the field of Kyudo, this paper will only focus on two important techniques in Kyudo; Kai and Hanare.

Self-training plays a significant role in Kyudo. Generally, many students require extra time to exercise and train on their own. Usually, a skilled instructor or coach is needed in these training sessions. However, employing an instructor can be troublesome and costly. Therefore, due to the lack of feedback from an instructor, students often do not realize that their stance of a particular technique is skewed. With the rapid advancement in technology and computer vision, students desire systems that will help them improve their performance. Thus, the development of a computer-assisted self-training system has become an emerging research topic. Similar systems developed and researched in this domain include gym fitness using real-time guidance [4] and a virtual environment for rugby skills [5]. There is a need for self-training that provides effective training for students without incurring a great deal of cost. Many self-training systems have been researched and developed, however, not Kyudo. Therefore, in this paper, we take Kyudo, also known as Japanese Archery, as the research topic and develop a Kyudo posture rectification system for assisting the archer in perfecting their posture.

The paper is organized as follows: In Sect. 2 a literature review is conducted. Section 2.1 describes the problem background. Section 2.2 investigates similar work that has been researched on self-training systems. Section 3 discusses the experimental setup of the model. The preparation of the dataset is discussed in Sect. 3.1 and the different pipelines are outlined in Sect. 3.2. Section 4 discusses the results that were obtained, and Sect. 5 concludes the paper with recommendations for improving the model.

2 Literature Review

2.1 Problem Background

Kyudo is a traditional archery that originated in Japan in the twelfth century. However, it only gained its popularity during the sixteenth century. It has now become an international sport in many parts of the world with many Kyudo training centers (Kyudojo) established. Currently, there are two ways to practice and learn Kyudo. The first method involves learning Kyudo in the Kyudojo with an instructor whereby the instructor demonstrates the techniques and movements required to execute and the student is expected to follow the instructor [3]. According to Chan et al. [6], one of the problems that students face is the lack of time to attend training sessions and at times instructors are unavailable or cannot give their full attention to each student. This form of training usually has many students in one class and therefore not all of them are able to follow the demonstrations by instructors due to obscurity [7]. The second method involves using materials such as books, magazines, and online videos (YouTube) [7]. However, these methods are ineffective and unsophisticated because feedback is not provided when attempting to perform a technique. Hence, it has become necessary to develop a one-on-one training

system that will improve the quality and performance of the students. The subsection to follow will outline research in the area of computer-assisted training.

2.2 Similar Work

Currently, limited research exists in the field of Kyudo posture rectification. However, there are many research articles on computer-assisted self-training systems for various other sports such as cricket, tennis, karate, dance, and exercises such as Yoga. The work that is related to the proposed system will be reviewed in this section.

Yoga Self-Training System
Chen et al. [8] propose the system "YogaST" which uses Feature Axis Extraction to detect the differences between the posture of the practitioner and expert. A Kinect camera was used to capture the user's body and extract the body contour. The system was able to identify the tree and dog pose of the user correctly. However, the warrior poses received a lower accuracy. The errors in feature extraction were mainly caused by incorrect segmentation of the body maps. Although the system was able to identify the poses, the feedback that was recommended were only online YouTube videos.

Karate Self-Training System
Hachaj et al. [9] propose a karate technique recognition system that was based on the Hidden Markov Model (HMM). The data was captured using a Microsoft Kinect camera. Thereafter, features based on angles were calculated from certain body joints. The data was then segmented using a Gesture Description Language (GDL) classifier to receive signals from a recording. The signal that was identified by the GDL classifier was processed by the HMM classifier. Based on the results the proposed system recognized karate techniques satisfactorily. Karate movements are difficult to classify because of its speed and agility. Therefore, the main errors were mainly due to misclassifications.

Virtual Reality Dance Training System
Chan et al. [6] propose a VR dance training system integrated with motion capture. The proposed system requires the user to wear a motion capture suit and follow the motion of the virtual dance teacher. The proposed system also provides feedback on how the dance moves can be improved. It collects data and evaluates the difference between the user and the virtual teacher. According to the results, the system assisted users in rectifying their movements.

Cricket Self-Training System
A system proposed by Kelly et al. [10] captures the motion of the batsman while playing a stroke. The stroke of the batsman is compared to known strokes which are classified using Fuzzy Sets, based on the body position and the ball or bat position. Appropriate feedback is provided to the user. The feedback outlines how well the stroke was performed and suggestions are made for perfecting that stroke.

Humanoid Robot for Kyudo
In 2010, a humanoid robot "iCub" was developed by Kormushev et al. [11] to learn

the skill of Kyudo. By employing certain mechanisms, the robot acquired the skill of holding the bow and releasing the arrow. Thereafter, the robot learnt how to shoot the arrow at the target by itself. The proposed system employed a local regression algorithm called ARCHER. Despite certain limitations such as the narrow joint angles of the robot and its inability to grasp, pull, and release the rope the results confirmed the feasibility of the approach.

Computer vision plays a key role in sports. According to Thomas et al. [12] there are well-established applications that use modern technologies such as multi-camera ball tracking, assisting referees, and providing viewers with analysis. Modern technologies used for sports analysis include image sensors, electromyogram sensors, and wearable technology. Despite recent developments in computer vision applications, the world of sports introduces many difficulties such as occlusions, abrupt motion, and similar appearances [12]. Therefore, a distinct and clear direction for computer vision in sports is needed to overcome these difficulties.

3 Experimental Setup

The proposed system aims to identify the type of technique presented by the user and provide appropriate feedback based on the score calculated using computer vision methods. Currently, the research that exists in the field of Kyudo posture rectification is limited. Thus, no existing datasets associated with Kyudo were found. According to Roh et al. [13], a dataset can be generated if there are no existing datasets. The dataset was therefore manually generated using different images of the Kai and Hanare technique.

3.1 Dataset Preparation

The dataset was collected by gathering various images of two important techniques in Kyudo; Kai and Hanare. A total of 320 images were collected. Before the images were placed into their respective training and testing sets each image was cropped to bounding box coordinates to discard everything in the image but the person. Once the images were cropped the next step was to label each image. A.csv (Comma-separated values) file was created which contained names of the images as well as their corresponding ground truth labels, which was either Kai or Hanare. The training and testing split used for the model was 80% for training and 20% for testing. Before training the classifiers, the images were resized by 100×100 and the pixel values of each image were converted to avoid feature scaling issues [14].

3.2 Implementation Details

The Keras framework was used to implement the proposed model. Three pipelines are implemented which consist of the following steps: The first pipeline uses Convolutional Neural Networks and Cosine Similarity to complete the task. The second pipeline consists of Support Vector Machines and Euclidean Distance. The first two pipelines use PoseNet for detecting the users' and instructors' coordinates. The third pipeline makes use of Convolutional Neural Networks, the ORB algorithm, and keypoint matching to

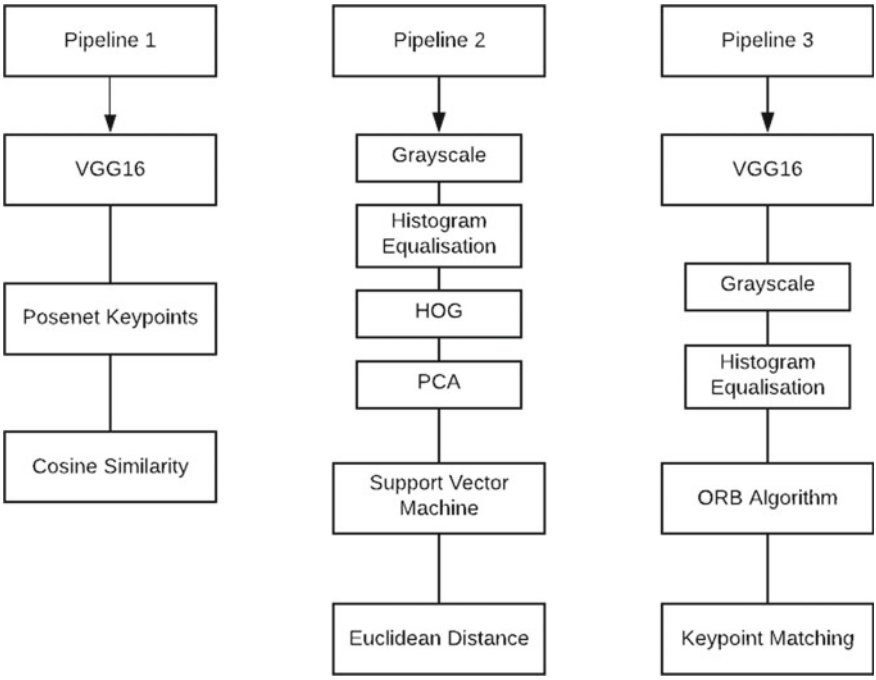


Fig. 1. Pipelines of model

provide feedback to the user. Figure 1 illustrates the different pipelines and the algorithms used in the model.

Convolutional Neural Network (Pipeline 1)
A Convolutional Neural Network is a deep learning algorithm that takes an image as its input and determines the importance of each object in the image [15]. The architecture constructed for the classification task is the VGG16 network [16]. The VGG16 network is a deep neural network with a 16-layer operation [16] and is used because of its simplicity. Thus, a sequential model, which stacks layers in a logical order is implemented [13]. The filter size of the input layer is set to 32, the kernel size of the convolutional window is set to 3×3 and the shape of the input image is set as (100, 100, 3). The loss optimizer used is categorical cross-entropy. The last layer has an output of 2 and uses sigmoid as its activation function [14] to determine if the image presented by the user is Kai or Hanare.

Support Vector Machine (Pipeline 2)
Before the dataset can be trained, certain features are extracted from the images in the dataset. The first step is to grayscale the images. The next step is to use histogram equalization to adjust the contrast in the images [17]. Thereafter, HOG (Histogram of Gradients) is used to detect the body of the instructor in the image. Patwary and Parvin [18] state that this is an efficient algorithm for detecting the human body. After HOG is completed, PCA (Principal Component Analysis), a dimensionality reduction technique

is used to transform the features into a lower-dimensional space. The dataset is then trained using SVM. The kernel type is set to linear and the penalty parameter is set to 1.

PoseNet (Pipeline 1 and 2)
PoseNet is a machine learning model which makes it possible to estimate human pose [19]. PoseNet estimates either a single pose or multiple poses from an image. For this model, a single pose is estimated. The PoseNet architecture used for the proposed model is MobileNet v1. MobileNet is faster and smaller, however, it has a lower accuracy than the ResNet50 architecture. Default values of the parameters (output stride, input resolution, multiplier) were used for the model.

Oriented FAST and Rotated Brief Algorithm (Pipeline 3)
ORB is a local feature detector introduced by Ethan Rublee in 2011 [20]. The algorithm is a modification of the FAST keypoints detector algorithm and BRIEF descriptor [20]. Initially, the algorithm uses FAST to detect keypoints and Harris corner measures to find the top n points [21]. Before the ORB algorithm is applied, the technique presented by the user and the instructor's technique is grayscaled and the contrast is adjusted using histogram equalization. Thereafter, the ORB algorithm is used to detect features in the image. The brute-force matcher is used for matching features. The descriptor of one feature in the first image is matched with the other features in the second image using distance estimation.

4 Results

The results of the proposed system are an indication of how well the model can identify the technique presented by the user and its ability to give constructive feedback. This section will discuss the results of the different pipelines. Goutte and Gaussier [22] state that a complete view of the system's performance is given by its precision, recall, accuracy, and F1 scores. Generally, the results indicate that there is room for improvement.

4.1 Convolutional Neural Network

The Convolutional Neural Network which was trained over 1000 epochs received an accuracy score of 81.25%. However, Sokolova and Lapalme [23] state that other performance measures are needed to determine the performance of the model. The accuracy and loss graph are illustrated in Figs. 2 and 3. The accuracy curve depicts that the model is fitting the training set better but losing its ability to predict new data. The loss curve depicts that the training curve resembles a good fit. However, the validation curve increases after every epoch, and the model is not generalizing adequately on the validation set. The results of the dataset are summarized in Table 1.

The ROC curve is used to visualize the performance of the classification problem. It is known to be one of the most important evaluation metrics for determining the performance of a model [23]. Figure 4 is an illustration of the ROC curve.

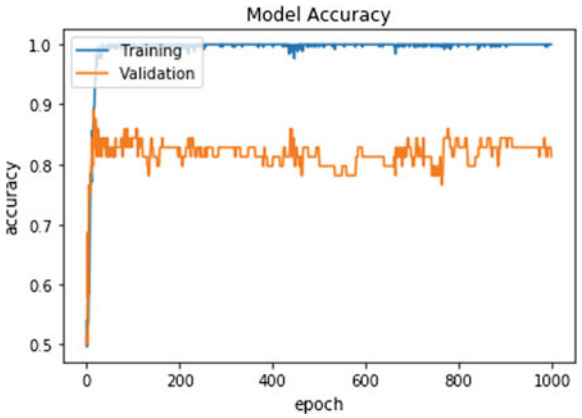


Fig. 2. Model loss

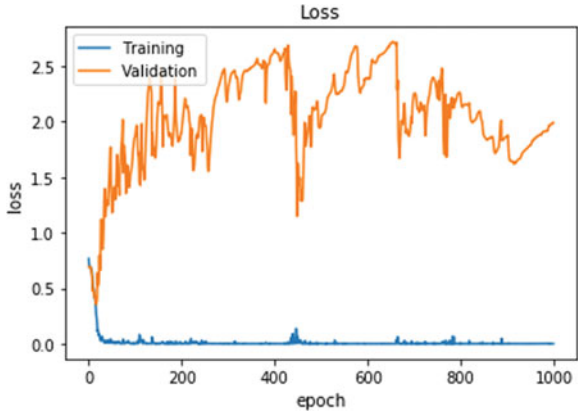


Fig. 3. Model accuracy

Table 1. Results of dataset

Accuracy		0.8125	
	Precision	Recall	F1 Score
Kai	0.78	0.88	0.82
Hanare	0.86	0.75	0.80

4.2 Support Vector Machine

Before the dataset can be trained, the number of components that should ideally be used in PCA for this model is determined. PCA is used to reduce the number of dimensions in the dataset. A Variance plot is generated to determine the number of components that should be taken into consideration. Figure 5 illustrates this graph. The graph depicts

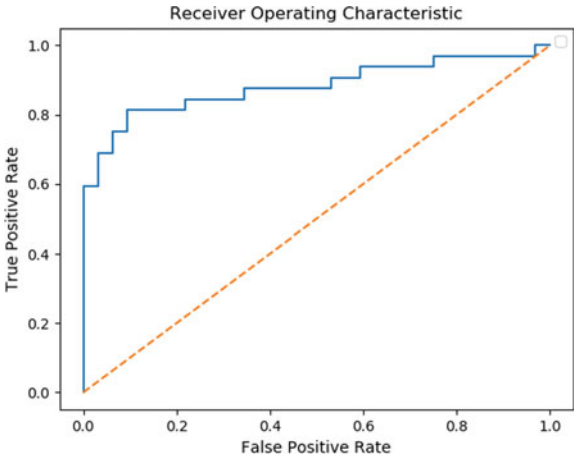


Fig. 4. ROC curve for dataset

that by picking 100 components we will preserve 98% of the total variance of the data. Thereafter, the dataset is trained. The performance measures of the model using SVM classification are summarized in Table 2.

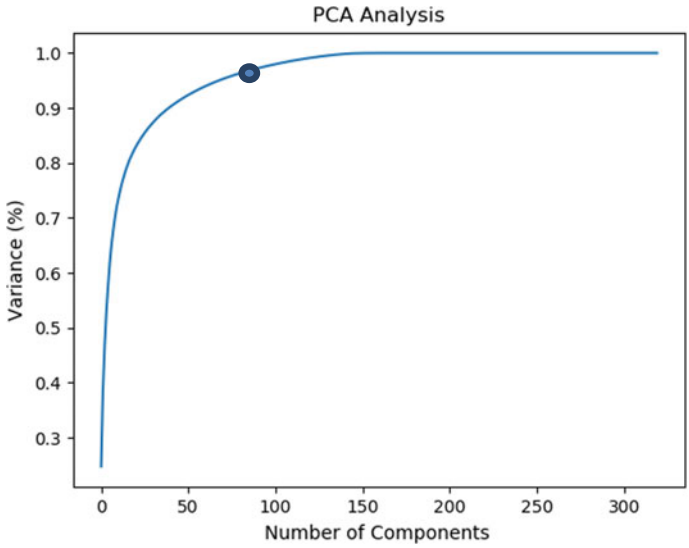


Fig. 5. Variance plot

The Support Vector Machine received an accuracy score of 80.2% which is a slightly lower accuracy from the previous result. The precision and recall rate also performed slightly worse than the previous classifier and an average F1 score of 79% was received.

Table 2. Results of dataset

Accuracy		0.802	
	Precision	Recall	F1 Score
Kai	0.73	1.00	0.85
Hanare	1.00	0.57	0.72

The low recall rate of Hanare indicates that some of the Hanare content remained unidentified and has a low sensitivity to identifying positive occurrences (Fig. 6).

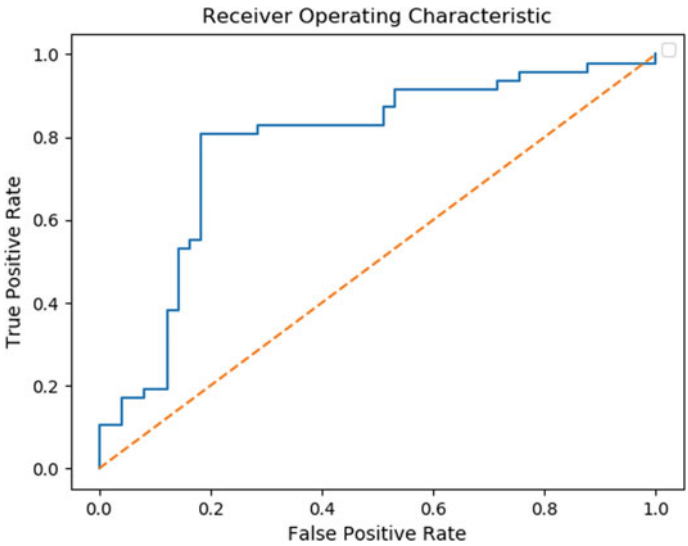
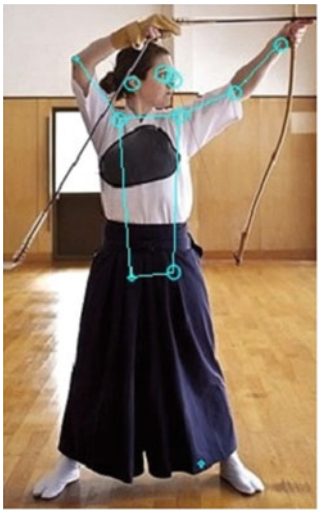


Fig. 6. ROC curve of dataset

4.3 PoseNet

After recognition of the technique presented by the user, the PoseNet model is used to determine the [x, y] coordinates of the instructor and the user. Currently, PoseNet detects 17 keypoints. Of these, five are facial key points which are irrelevant and are therefore not included. The keypoints are then converted to vectors and normalized using L2 normalization. After, normalization, Euclidean distance and Cosine similarity are used to determine the similarity between the two vectors. It was found that using cosine similarity between the vectors is more efficient than Euclidean distance because it determines how closely the vectors point in the same direction. The direction of the arrows indicates how similar the poses are. Vectors that are in the same direction are similar while vectors in the opposite direction are different.

The model also draws or plots the key points on the images. However, due to obscurity in the image, the model is unable to draw and determine the key points of the instructor's feet and will thus have an incorrect [x, y] coordinate. The results determine that this needs to be improved in order to provide the correct feedback to the user. Figure 7 represents the key points of the instructor and their corresponding coordinates. After a similarity score is obtained using Euclidean distance and cosine similarity, it is converted to a percentage. Thereafter, some feedback is provided to the user.



Keypoint leftShoulder, coord = [690.32061139 484.87737081]
Keypoint rightShoulder, coord = [700.38573595 373.27887143]
Keypoint leftElbow, coord = [683.28965064 607.13693545]
Keypoint rightElbow, coord = [714.88721996 226.71521065]
Keypoint leftWrist, coord = [647.86098958 708.61919318]
Keypoint rightWrist, coord = [642.74800154 270.4623985]
Keypoint leftHip, coord = [929.10756852 480.15077092]
Keypoint rightHip, coord = [941.6815156 391.04189898]
Keypoint leftKnee, coord = [1143.24548198 483.26913525]
Keypoint rightKnee, coord = [1140.81382127 370.09321926]
Keypoint leftAnkle, coord = [1315.07705077 500.05567827]
Keypoint rightAnkle, coord = [1318.55881016 375.63203272]

Fig. 7. Keypoints of instructor's Kai technique

4.4 Feature Extraction Using ORB

After the ORB algorithm is applied to the images, the number of matches between the two images was determined and a match score (%) is calculated to determine the similarity between the two techniques. Rublee et al. [21] state that even though ORB is the fastest algorithm SIFT performs the best. However, SIFT poses a computational burden [24]. Hence, it was decided that the ORB algorithm would be used.

5 Conclusion

By using computer vision methods, the proposed model can effectively distinguish between the techniques presented by the user, compare vector coordinates, and provide appropriate feedback based on calculations. The CNN classifier received an accuracy score which is slightly greater than the SVM classifier. The results indicate the feasibility of the approach, however, work for improvement is required in certain areas. In future work, a larger dataset can be used to increase the performance of the classifiers, SURF features can be extracted from images in the dataset before being trained and

clear images of instructors' techniques should be used to obtain all the coordinates for effective feedback. Another recommendation or improvement can be to provide users with visual feedback. Virtual reality has been introduced to many sports, however, not to Kyudo. Once improvements have been made this work can effectively assist Kyudo students.

References

1. Shoji, Y.: The myth of Zen in the art of archery. Jpn. J. Religious Stud. **28**(2), 2–30 (2001)
2. Genzini, L.: Kyudo—the way of the bow: the art of shooting the traditional Japanese bow according to the Heki Insai Ha School: technical manual, Luigi Genzini, Rom
3. Cohen, E.: Kyudo—resonance involuted and the folding of time in Japanese archery. Anthropos Int. Rev. Anthropol. Linguist. **2**(109), 525–537 (2014)
4. Jin, X., Yao, Y., Jiang, Q., Huang, X., Zhang, J., Zhang, X., Zhang, K.: Virtual personal trainer via the kinect sensor. In: 2015 IEEE 16th International Conference on Communication Technology, pp. 460–463 (2015)
5. Miles, H.C., Pop, S.R., Watt, S.J., Lawrence, G.P., John, N.W., Perrot, V., Mallet, P., Mestre, D.R.: Investigation of a virtual environment for rugby skills training. In: 2013 International Conference on Cyberworlds, Yokohama, Japan (2013)
6. Chan, J.C., Leung, H., Tang, J.K., Komura, T.: A virtual reality dance training system using motion capture technology. IEEE Trans. Learn. Technol. **4**(2), 187–195 (2011)
7. Ain Mohd Jelani, N., Zulkifli, A.N., Ismail, S., Yusoff, M.: Taekwondo trainees' satisfaction towards using the virtual taekwondo training environment prototype. In: AIP Conference Proceedings, vol. 1891, p. 020099 (2017)
8. Chen, H.T., He, Y.Z., Hsu, C.C., Chou, C.L.: Yoga posture recognition for self-training. In: International Conference on Multimedia Modeling, pp. 496–505. Springer, Cham (2014)
9. Hachaj, T., Ogiela, M., Koptyra, K.: Application of assistive computer vision methods to Oyama Karate techniques recognition. Symmetry **7**, 1670–1698 (2015)
10. Kelly, M., Curtis, K., Craven, M.: Fuzzy Sets and Cricket Batting (2002)
11. Kormushev, P., Calinon, S., Saegusa, R., Metta, G.: Learning the skill of archery by a humanoid robot iCub. In: Proceedings IEEE International Conference on Humanoid Robots (Humanoids), pp. 417–423. IEEE, Nashville (2011)
12. Thomas, G., Gade, R., Moeslund, T.B., Carr, P.: Computer vision for sports: Current applications and research topics. In: Computer Vision and Image Understanding, p. 159 (2017)
13. Roh, Y., Heo, G., Whang, S.E.: A survey on data collection for machine learning: a big data—AI integration perspective. IEEE Trans. Knowl. Data Eng. (2019)
14. Vallet, A., Sakamoto, H.: A multi-label convolutional neural network for automatic image annotation. J. Inf. Process. **26**(6), 767–775 (2015)
15. Chen, F.Y.L., Yang, H.: Image-based product recommendation system with convolutional neural networks, Stanford (2017)
16. Mateen, M., Wen, J., Nasrullah, D., Song, S., Huang, Z.: Fundus image classification using VGG-19 architecture with PCA and SVD. Symmetry **11**(1) (2018)
17. Dorothy, R., Joany, R.M., Rathish, J., Prabha, S., Rajendran, S., Joseph, S.: Image enhancement by histogram equalization. Int. J. Nano Corros. Sci. Eng. **2**, 21–30 (2015)
18. Patwary, M.J.A., Parvin, S.: Significant HOG-histogram of oriented gradient feature selection for human detection. Int. J. Comput. Appl. **132**(17), 20–24 (2015)
19. Garg, P.: Comparison of human poses with PoseNet. Medium, 12 May 2019. [Online]. Available: https://medium.com/@priyaanka.garg/comparison-of-human-poses-with-posenet-e9ffc36b7427. Accessed 24 Oct 2019

20. Abdulmajeed, M., Seyfi, L.: Object recognition system based on oriented FAST and rotated BRIEF. In: ISAS 2018-Winter 2nd International Symposium on Innovative Approaches in Scientific Studies, pp. 179–181 (2018)
21. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2564–2571 (2011)
22. Goutte, C., Gaussier, E.: A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In: Lecture Notes in Computer Science, vol. 3408, pp. 345–359 (2005)
23. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Inf. Process. Manage. **45**, 427–437 (2009)
24. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: 2011 International Conference on Computer Vision. In: ORB: An Efficient Alternative to SIFT or SURF. IEEE (2012)

Accident Detection Using Time-Distributed Model in Videos

Garima Gupta(✉), Ritwik Singh, Ashish Singh Patel, and Muneendra Ojha

Department of Computer Science and Engineering, International Institute of Information Technology, Naya Raipur, India
garima17100@iiitnr.edu.in, ritwik17100@iiitnr.edu.in, ashish@iiitnr.edu.in, muneendra@iiitnr.edu.in

Abstract. CCTV surveillance cameras are installed in the majority of roads and highways these days; therefore, it generates millions and millions of hours of data, thus captures a variety of real-world incidents. Road accidents are one of the most severe and fatal incidents, which disrupt the smooth flow of traffic as well leading to wastage of time and resources. Detection of accidents not only helps us to save the life of victims, but also helps in reducing traffic congestion. In this, we have proposed a framework for accident detection based on hierarchical recurrent neural network. The framework localizes and identifies the presence of road accidents in the captured video. The framework contains a time-distributed model which learns both the temporal and spatial features of the video, making the framework more efficient. The proposed approach is evaluated on the dataset, built by obtaining recorded road accident videos from YouTube. Results demonstrate the applicability of our approach performs, accident detection, and localization effectively.

Keywords: Accident · Traffic · Time-distributed · Deep learning · Accident detection · Long short-term memory networks · Neural networks

1 Introduction

Nowadays, industrialization and technological advancement lead to an exponential increase in transportation. Hence, there is an increase in usage of the automobile which has lead to lots of accidents in previous years. They are the prominent reasons for deaths, injuries, and property damage every year. According to the stats stated by eSocialScience, India had about 5 lakh road causalities. The road causalities killed approximate of 1.5 lakh people in 2015. India signed the Brasilia declaration, and it expects to cut down the road causalities and traffic mortality by 50% before 2022. It seeks to focus on concerns related to road causalities, third-party insurance, and road safety measures.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_21

One of the most important and challenging problems that could be solved with the data generated from these cameras is accident detection. Generally, accidents occur rarely in comparison to ordinary activities. Hence, to mitigate the waste of time and human efforts, there comes a pressing need to develop a system that automatically detects accidents in videos. The main objective of the accident detection system is to detect an activity that diverge from ordinary patterns and classify the time window of the occurring accident. Thus, accident detection can be treated as coarse level of video understanding, that filters out accidents from regular ordinary patterns.

A video incorporates both temporal as well as spatial features. Each frame in the video holds the current context of the video, i.e, spatial information and also the context of the frames before it in time, i.e., temporal feature [2,3]. Hence, video analytics helps us elucidate data more efficiently thereby has a wide variety of applications soccer score generator, violence detector, and traffic accident detector. However, the implementation of each of these applications varies at vivid levels.

Accident detection in real time could end up being more complex and diverse. However, a high proportion of approaches has been made on image datasets. Some of the existing approaches are based on probabilistic and deterministic models that use handcrafted features which further requires to employ feature engineering techniques and so the deep learning models prove to be of better use due to their automatic feature learning capability. Most of them either work on a single frame or they use annotations for training their model. In this work, we propose a time-distributed long short-term memory (LSTM) network to learn accidents by treating normal and accident videos as negative and positive samples during training. During experimenting, a long untrimmed video is divided into segments and the segments are converted to frames and then fed into our deep learning model. Also, we present tentative results on our new dataset.

The remainder of this paper is organized as follows: Sect. 2 describes related work that has been done considering accident detection; Sect. 3 describes the methodology of the proposed solution; Sect. 4 empirically evaluates the experiments done on dataset and modeling of the solution and the results achieved.

2 Related Work

Accident detection problems have been approached with wide category of solutions and have been a constant interest of researchers in various fields of science. However, a high proportion of approaches has been made on image datasets also, on probabilistic and deterministic models that use handcrafted features which further requires to employ feature engineering techniques and so the deep learning models prove to be of better use due to their automatic feature learning capability. The paper [11] summarizes the different road accident detection techniques of this type. These methods suffer from design flaws such as easy tampering, failure, and damage caused to the hardware by third parties. Moving to designs that incorporate image processing and computer vision, approaches

that use image masking(car or foreground) have been taken [8]. These methods suffer from not being robust as they can easily be attacked and also they may not be able to detect complex or subtle accidents. Du et al. [10] have also proposed systems that detect each vehicle and track the behavior of each vehicle and recognize situations and events that are likely to result from the chain of behavior. The algorithm works successfully on images but requires preprocessing the images to grayscale. More advanced approaches have been developed with classical machine learning methods like SVM dynamic saliency map in [16], random forest [4], and using big data processing and computational intelligence [13]. Many deep learning models and algorithms have also been adopted for this task. Dynamic-spatial-attention (DSA) [1], recurrent neural network (RNN) [14] has been extensively used for this task. LSTMs are enhanced recurrent neural networks that could solve more complicated, artificial long-time lag tasks that could not be solved by recurrent neural networks due to vanishing and exploding gradients in them. When it comes to videos, actions or text that contain temporal variables, the hierarchical recurrent neural network has been prevalent lately. It has found use in skeleton-based action recognition [6]. A hierarchical neural autoencoder has been used for paragraphs and documents [12]. Rachel Wagner Kaiser used the hierarchical recurrent neural network for dashcam videos [9]. In this paper, we use the hierarchical recurrent neural network (HRNN) for classifying CCTV videos. Frames are extracted from the videos and are of 99 segments. Then features are extracted through two different LSTMs, and extracted features are passed through fully connected layer for accident detection. Please see Fig. 1.

3 Methodology

Long short-term memory network is an improvised technique under the umbrella of recurrent neural networks that can learn long-period dependencies. It was proposed by Hochreiter and Schmidhuber [7]. They can memorize information for long term. They are capable of learning the information between each frame of videos with ease. We introduce a time-distributed RNN-based model that utilizes frames of videos to predict whether an accident has taken place or not in the video. In Sect. 3.1, we describe the data processing part that is used to train our presented model. Section 3.2 describes our time-distributed RNN model architecture.

3.1 Data Preprocessing

We have extracted frames from videos using OpenCV. We have extracted 99 frames for each video and have also applied data augmentation for increasing our dataset size. Then, we have resized each frame to the size of (144×256) and have also normalized the frames. For this, we have used a skimage library as well as Keras preprocessing library.

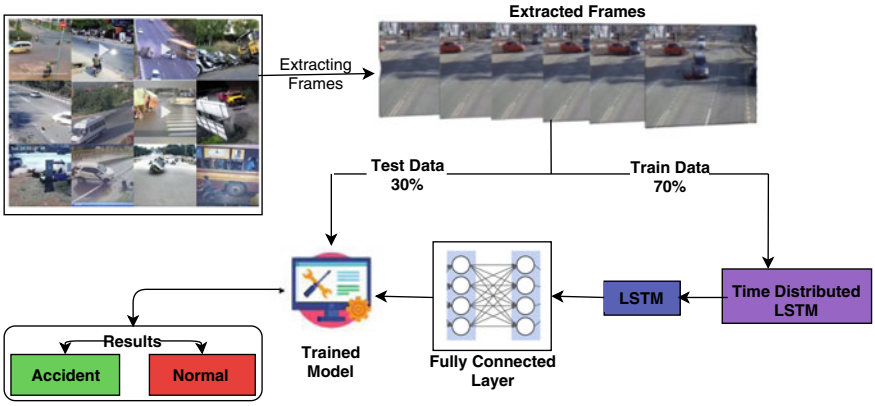


Fig. 1. Proposed time-distributed model architecture

3.2 Model Architecture

The model architecture is achieved by combining one time-distributed LSTM, one LSTM network, and one dense layer which is used to carry out the main binary classification task. The LSTM is one of the modern methods which has been used in image captioning, neural machine translation, video description, etc., as it can model long-term sequential dependencies as it uses additive updates to its cell state and therefore avoids vanishing gradients. LSTMs deal with the memorization of patterns observed in current time t and also of other states. It is also capable of recollecting and forgetting the correlations throughout the patterns based on all the states. With hidden states ht , internal memory cell state ct and three gates i_t ,o_t ,f_t ,g_t is a candidate memory cell state from the present input and the past hidden the following LSTM equations were used to calculate the output for the next block and store the memory in the current block.

$$i_t = \sigma(W^i x_t \oplus w_t + U^i h_{t-1} + b_i) \tag{1}$$

$$o_t = \sigma(W^o x_t \oplus w_t + U^o h_{t-1} + b_o) \tag{2}$$

$$f_t = \sigma(W^f x_t \oplus w_t + U^f h_{t-1} + b_f) \tag{3}$$

$$g_t = \tanh(W^g x_t \oplus w_t + U^g h_{t-1} + b_g) \tag{4}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot g_t \tag{5}$$

$$h_t = o_t \cdot \tanh c_{t-1} \tag{6}$$

where $\otimes$ is a concatenation operator of the vector and is an element-wise multiplication of two vectors, W is weight matrix from input to hidden state, and U is weight matrix from hidden to hidden. Time-distributed LSTM : The LSTM takes inputs for each video sequence for T time steps.Let $X_1, X_2, X_3, X_4, \ldots, X_{t-1}, X_t$ be the image sequences (Fig. 2).

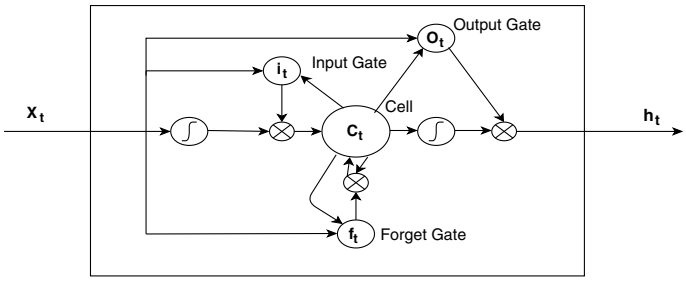


Fig. 2. A LSTM unit with input, output, forget gates

The time-distributed LSTM learns features from the input sequence. Time-distributed LSTM is able to learn temporal and spatial features in a hierarchical structure. Then, the extracted features are passed through a fully connected dense layer for classifying videos using softmax activation.

Network architecture: We extract frames of videos and then normalize each frame and then also resize each image to the size of (144×256) and store them in an array. The training and testing set of data are generated by dividing 376 videos to each set in the ratio of 7:3. The training set is utilized for training the model and testing set for evaluating it. Now, the X_train is passed through the layer of a time-distributed LSTM, which endeavors to learn about the temporal features between each frame. There are 128 hidden states in time-distributed LSTM that memorizes the features. Similarly, we pass the above-extracted feature through an LSTM network which also has 128 hidden states. The feature learned through the LSTM layers is passed through a dense layer for classification which uses softmax activation calculates the probability distribution of the accident over different events (Fig. 3).

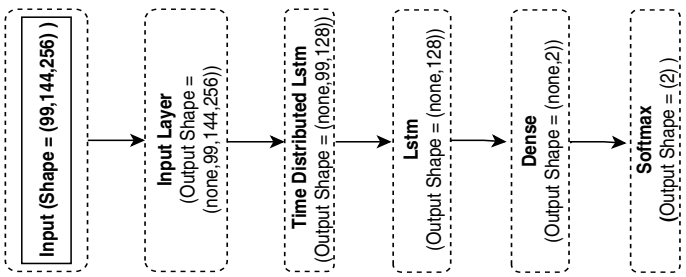


Fig. 3. Network architecture of the model

3.3 Performance Evaluation

The performance of the proposed model can be defined as,

$$\text{loss} = -(y \log(p) + (1 - y) \log(1 - p))$$

where p is the predicted probability observation of a class, y is the binary indicator that indicates 0 or 1 depending on the predicted class label.

4 Experiment

4.1 Dataset

We have evaluated the proposed model on DETRAC dataset and accident videos taken online source (YouTube) to demonstrate the flexibility of our machine learning model. Annotation for person-level activities was not used in this research. Instead, the model is systematically learned from the internal representation of individual activities. The dataset is composed of 376 videos of negative and positive samples. The positive data comprises of accident videos that had 94 segments originally and on applying data augmentation techniques, the videos were flipped and we then had 188 positive segments. The negative samples were 188 normal incidents segments from DETRAC dataset [15].

DETRAC dataset: The dataset comprises 10 h of videos that were recorded in Beijing and Tianjin. The videos were recorded at 25 fps (frames-per-second). The dataset contains more than 140 thousand frames.

4.2 Metrics

Classification accuracy is used for evaluating the performance of the model. It is what we usually refer to, when we use the term accuracy. In the context of the research, it is the ratio of number of detected accidents to the total number of accidents. Mathematically, it translates to :

$$\text{Accuracy} = \frac{\text{Number of accidents detected}}{\text{Total number of accidents}}$$

Precision tells us how pertinent are the positive predictions and in the context of this research, it is the ability of the classifier not to label the case of accident as case that is not an accident. It is the ratio of true positives to the total number of positive predictions done by the classifier. Mathematically, it translates to :

$$\text{Precision} = \frac{\text{Number of correct accidents detected}}{\text{Number of accidents detected}}$$

Recall tells us what percentage of actual positives are identified as positive, and in the context of this research, it is the ability of the classifier to find all

cases having accident. It is the ratio of true positives to the actual number of positives cases in the testing data. Mathematically, it translates to :

$$\text{Recall} = \frac{\text{Number of correct accidents detected}}{\text{Actual number of accident detected}}$$

$F1$ score measures the test's accuracy. This is the harmonic mean between precision and recall. It tells how precise our classifier is as well as how robust it is. A greater $F1$ score signifies that the performance of our model is better. The concept of $F1$ score is used to establish a balance between precision and recall. Mathematically, it translates to:

$$F1 = 2 * \frac{1}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}$$

Support is the frequency of each class in "y_true," where "y_true" is the ground truth.

4.3 Training

For training, we have *NAdam* that accelerates the learning process by adding up the exponential decay of the moving averages for the penultimate and current gradient as our model optimizer in [5]and categorical cross-entropy as the loss function.*NAdam* incorporates Adam and Nesterov accelerated gradient(NAG) which helps in updating parameters with momentum step before computing the gradient. We have trained the model of time-distributed LSTM with mini-batch of 15 examples for 40 epochs. The learning rate was set as $\lambda = 0.001$ for training. Keras and TensorFlow frameworks were used to implement our model.

4.4 Results and Discussion

In our proposed time-distributed model, the network has been trained on normal videos using the temporal window of 99 frames and used the accuracy to measure the performance of the suggested model. Table 1 presents the evaluations for the proposed model.

With the above architecture trained for 700 mini-batches with 15 epochs and batch size of 25, we were able to get model with the accuracy and training curves, as shown in Fig. 4.

We have tested the model for 48 videos and plotted the confusion matrix and calculated its precision, recall, $F1$ score and support for various cases. The information calculated can be seen in Tables 2 and 3, respectively. The model made a total of 48 predictions (i.e., it was given an input of 24 positive videos and 24 negative videos for detecting whether it contains accident or not). Out of those 48 predictions, model predicted 19 times "YES" and 29 times "NO." Hence, model predicted 7 wrong predictions whereas other predictions were correct.

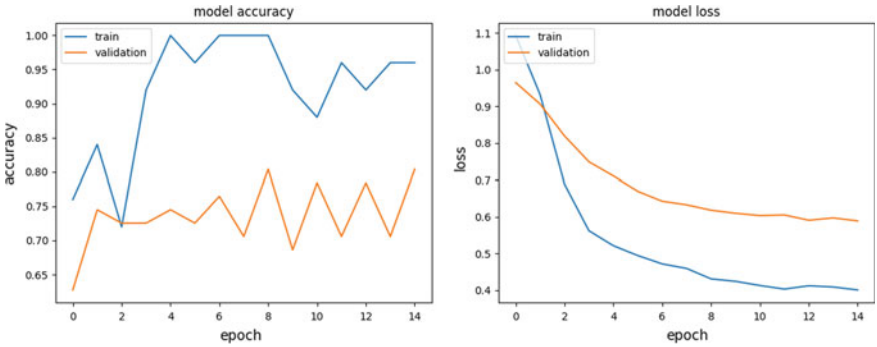
The model gave us precision of 0.95 for accident prediction and 0.79 for non-accident videos. And also, the $F1$ score of the model was 0.87 for negative

samples and 0.84 for positive samples. The model performed quite good with $F1$ score of 0.85 for weighted results and has not underfitted or overfitted.

The model gave good results on testing and was able to classify videos without any manual annotations.

Table 1. Accuracy metrics evaluation results of the proposed model

Execution phase	Accuracy(%)
Training	94.33
Testing	85.41



(a) Training and validation curve for model accuracy

(b) Training and validation curve for model loss

Fig. 4. Training and validation curve for model accuracy and loss

Table 2. Confusion matrix obtained while testing 48 videos

Actual/predicted	No	Yes
No	23	1
Yes	6	18

Table 3. Report obtained using confusion matrix while testing 48 videos

Prediction class	Precision	Recall	$F1$ score	Support (%)
0	0.79	0.96	0.87	24
1	0.95	0.75	0.84	24

5 Conclusion

The developed work proposes an approach to detect accidents from CCTV surveillance using a deep learning technique. Often a lot of complexities are involved in video classification. Therefore, an attempt has been made to exploit both temporal and spatial features in the proposed model. We have implemented the hierarchical recurrent neural network to tackle the complex task of classifying video footage. The HRNN incorporates a time-dependent aspect of the frames within each video to predict how likely it is, and a new video contains a car accident. To evaluate this model, the presented work involves a compilation of a dataset that contains real-world accident scenarios in India. The experimental results on this dataset prove that our model works better even without a manually annotated dataset. Although the model only works on trimmed set of videos and have not been trained on dataset containing distinct scenarios.

References

1. Chan, F.H., Chen, Y.T., Xiang, Y., Sun, M.: Anticipating accidents in dashcam videos. In: Asian Conference on Computer Vision, pp. 136–153. Springer, Berlin (2016)
2. Dey, N., Ashour, A., Patra, P.K.: Feature detectors and motion detection in video processing. IGI Global (2016)
3. Dey, N., Dey, N., Ashour, A., Acharjee, S.: Applied Video Processing in Surveillance and Monitoring Systems. IGI Global (2016)
4. Dogru, N., Subasi, A.: Traffic accident detection using random forest classifier. In: 2018 15th Learning and Technology Conference (L&T), pp. 40–45. IEEE, Berlin (2018)
5. Dozat, T.: Incorporating Nesterov momentum into Adam (2016)
6. Du, Y., Wang, W., Wang, L.: Hierarchical recurrent neural network for skeleton based action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1110–1118 (2015)
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
8. Jeong, W.J., Lee, J.M., Park, K.T., Moon, Y.S., et al.: Car accident detection in parking lots. In: 2014 International Conference on Electronics, Information and Communications (ICEIC), pp. 1–2. IEEE, New York (2014)
9. Kaiser, R.W.: Crash Catcher (2018), https://github.com/rwk506/CrashCatcher
10. Kamijo, S., Matsushita, Y., Ikeuchi, K., Sakauchi, M.: Traffic monitoring and accident detection at intersections. IEEE Trans. Intell. Transp. Syst. **1**(2), 108–118 (2000)

11. Khalil, U., Javid, T., Nasir, A.: Automatic road accident detection techniques: A brief survey. In: 2017 International Symposium on Wireless Systems and Networks (ISWSN), pp. 1–6. IEEE, New York (2017)
12. Li, J., Luong, M.T., Jurafsky, D.: A hierarchical neural autoencoder for paragraphs and documents. arXiv preprint arXiv:1506.01057 (2015)
13. Ozbayoglu, M., Kucukayan, G., Dogdu, E.: A real-time autonomous highway accident detection model based on big data processing and computational intelligence. In: 2016 IEEE International Conference on Big Data (Big Data), pp. 1807–1813. IEEE, New York (2016)
14. Shah, A.P., Lamare, J.B., Nguyen-Anh, T., Hauptmann, A.: CADP: A novel dataset for CCTV traffic camera based accident analysis. In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–9. IEEE, New York (2018)
15. Wen, L., Du, D., Cai, Z., Lei, Z., Chang, M.C., Qi, H., Lim, J., Yang, M.H., Lyu, S.: UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking. arXiv preprint arXiv:1511.04136 (2015)
16. Woo, J.W., Lee, W., Lee, M.: A traffic surveillance system using dynamic saliency map and SVM boosting. Int. J. Control Autom. Syst. **8**(5), 948–956 (2010)

Analysis of Traditional Computer Vision Techniques Used for Hemp Leaf Water Stress Detection and Classification

Waseem Shaikjee and Dustin van der Haar(✉)

Academy of Computer Science and Software Engineering, University of Johannesburg, Cnr University Road and Kingsway Avenue, APK Campus, Johannesburg, South Africa
waseem.shaikjee@gmail.com, dvanderhaar@uj.ac.za

Abstract. *Cannabis sativa* L. has risen in popularity due to its large variety of uses and environmentally friendly impact. *C. sativa* L. is extremely sensitive and displays phenotypic responses to water stress in its leaf and stem structure. Optimizing the use of water in the agricultural process of cultivating hemp requires the determining of the water potential in the hemp plant. Computer Vision techniques to determine water potential can be used as opposed to traditional destructive and complex to implement techniques. The goal of this study is to prove that water stress detection in hemp leaves can be achieved using computer vision as well to create a model and compare computer vision techniques. This study used a dataset pooling technique to create the dataset of hemp leaves. The dataset is split randomly at an 80–20% ratio of training data and testing data, respectively. Two derivatives of the traditional pattern recognition pipelining model were used. The first pipeline employed traditional computer vision techniques such as Canny Edge Detection, Contour Analysis, SIFT, and SVM Classification. The second pipeline embraced an object detection approach by implementing Haar Cascades. The results of the study vary greatly leading to researchers to believe that more work needs to be done to improve performance.

Keywords: Computer vision · Hemp · Water stress · SVM · SIFT

1 Introduction

Cannabis sativa L., a plant once indigenous to the eastern parts of Asia, along with being one of the oldest domesticated crops in human history, is now a commonly cultivated crop that lacks scientific study [1–3]. The word "Hemp" was a term used to label any plant belonging to the Cannabis Genus but in recent times it is used to refer to Cannabis cultivated for non-narcotic use. The umbrella term hemp covers many strains of the *C. sativa* species that contain low amounts of tetrahydrocannabinol (THC) [4]. THC is the psychoactive cannabinoid (biomolecule) found in Cannabis plants. Cannabinoids are

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_22

primarily responsible for the psychedelic effects on the brain [4]. The high levels of THC found in Cannabis have caused the plant to become a prohibited narcotic in almost all countries. However, recent interest in the medicinal effects of Cannabis has led to the loosening of regulations and laws around Cannabis cultivation [1]. This is leading to the greater acceptance of cultivating Cannabis with low THC levels for industrial uses. With research restricted by the once illicit status of hemp [1], there lacks research on the understanding of the phenotypic effects of water stress on hemp plants [5]. A better understanding of how hemp reacts to water stress can provide insight into how technological solutions can be used to solve present-day agricultural issues.

Water plays an important role in the rapid growth of hemp. The yield amount of hemp cultivation is reliant on the access of water. High water demands for hemp cultivation has negative effects on the environment. Correct use of water in the agricultural sector can help minimize the impact of agriculture on the environment and maintain better use of resources.

Hemp is an excellent candidate for water stress research as it is known to have high sensitivity towards water stress [5]. Water stress in the context of agriculture refers to a limited water supply to the plant or an overabundance of water supply [6]. Water stress in hemp plants induces phenotypic responses in the plant leaves [6]. Dehydration leads to the reduction of the hemp leaves the surface area, a defence mechanism which is indicated by the curling of the leaf's edges inwards. Hemp plant overhydration is indicated by the increase in leaf surface area and weakening in the stem's cellular structure, weakening the stem's ability to uphold the hemp leaf. By measuring the water potential in a leaf, is it possible to create a modern computing system that can optimize water usage in cultivation to produce the optimum yield in an inexpensive and easy to use approach?

Traditional methodologies used to determine water stress and water potential in plants include soil water potential measurement techniques, and leaf water potential [7]. Unfortunately, soil water potential measurement does not give a true indication of the water potential in the plant [7]. Thereby making, leaf water potential is the most accurate measurement. Traditional water potential measurement techniques rely on either expensive sensors or destructive methods as well as skilled professionals to apply these techniques [7]. Computer vision can be applied to measure the phenotypic responses of hemp leaves to give an approximation of water potential in the leaf.

The research in this paper focuses on computer vision techniques and how best it is applied in the context of determining the water stress levels in hemp leaves. Techniques are compared in terms of accuracy and ability to implemented in real-world scenarios. The computer vision techniques used will be broken up into three separate program pipelines, comparing traditional computer vision methodology with well researched and applied methods as well as more modern techniques.

This research paper covers the reason for water stress detection in hemp plants using computer vision in the problem background. Similar studies are discussed explaining approaches in water stress detection that have been already researched. This paper then proposes a model of how the experiment was set up and the reasons for following the selected approaches. The results of the computer vision approach used are discussed in the results and critique section.

2 Problem Background

The severity of global warming is increasing, bringing about extreme weather conditions that lay on the opposite ends of the weather spectrum [5]. The effects of global warming can be seen by the increase in prolonged droughts and intense floods being experienced all over the world [5]. Extreme weather affects the development and growth of vital crops needed to sustain humanity [5]. A new way of perceiving how we grow and use crops is needed to minimize the effects of global warming and climate change. Industrial hemp may offer a more environmentally friendly approach to the cultivation of essential crops as well as the manufacturing of environmentally friendly products [8].

Hemp is one of the oldest domesticated crops in human history [1]. While mainly being a source of fibre, hemp also provides oils and seeds as raw materials [5]. These raw materials can be used to make, food, medicine, bioplastics, bio-composites, biofuels, etc. [5]. Hemp is not only a source for raw materials, but the growth cycle of hemp can be used to purify the soil, water sources and deter from herbicide usage [9]. This makes hemp an appealing crop to cultivate and environmentally friendly. However, large-scale agriculture has an impact on the environment and the ecosystems that exist within that environment.

Growing populations create larger demands from the agricultural supply chains. This forces producers to employ methodologies that ensure demands are met. These methodologies can have negative impacts on the environment. In the context of industrial agriculture, chemical control methods are applied to protect the crop and the final crop yield. Chemical controls such as pesticides and fertilizers are known to damage ecosystems [10]. Not only are the by-products of large-scale agriculture affecting the environment but furthermore put a strain on resources in an environment due to high demands [10]. In the hemp cultivation process, water is one of the most important resources in growth and development. However, substantially hot, dry or wet environments make it extremely challenging to cultivate hemp crops [5].

Hemp is very sensitive to water stress, both drought and flood water stress can impair plant growth and reduce the quality of fibres, oil and the overall yield of the crop [5]. The hemp growth lifecycle is resource-intensive and requires large amounts of water to produce the highest crop yield [8]. It is human nature to use more than what is needed, in the hopes to achieve the best possible outcome. In this case, it would be the notion of keeping the hemp crop well irrigated, in the hope to achieve the best possible crop yield. This is sometimes done not knowing the water status of the plant. In order to reduce the effects of global warming and minimize environmental impact, the growth and development processes hemp needs to be optimized [7].

The goal of environmentally friendly hemp cultivation is to increase the water use efficiency to such a level where the least amount of water is used to produce the same crop yield as a crop watered in abundance [7]. Water use efficiency is used to describe the relationship between crop growth and the amount of water used to irrigate the crop [7]. Determining when water stress transpires, provides vital information for maximizing crop production and water use efficiency [11].

Water stress describes the period in which the rate of water loss declines below that of a well-watered crop found in the same locality or environment [6]. Plants exhibit symptoms of water stress when there is a deficit or excess of water in the plant. The

stress becomes so unfavourable that it affects the plant's critical processes [6]. Different plant processes are affected by water stress in different ways [6]. Several morphological and physiological changes in a plant occur when a plant is stressed if monitored, this can provide specific indicators to specific types of stress [6].

Many Existing systems and approaches to measure plant water stress indicators include stomatal regulation measuring [11], leaf water potential measurements, and sap flow measurements [12]. The problem with these types of approaches is that they are intrusive and require contact with the plant, thus making it difficult to capture data easily and quickly [12]. Using the previous approaches also require manual effort and a considerable amount of knowledge and skill to be able to employ such approaches correctly [12]. These approaches are not easily scalable and are expensive to implement, as well as to be destructive to the plant itself [12]. These methods of measuring the water status in plants are scientifically sound and have been used by botanists as proven methodologies. However, they are difficult to implement on a large-scale in commercial environments [12]. A solution to this problem is the use of a non-contact approach to measuring the key morphological indicators or water stress in a plant.

In large agricultural scenarios, where land used for cultivation can span several hundred acres. It is expensive, time-consuming and ineffective to manually check the water stress levels of the crop using contact or destructive approaches. By employing a computer vision solution to identify and recognize water stress symptoms in hemp plants via morphological indicators expressed in the leaf, can lead to better water use efficiency. This directly decreases the cost of growth as well as the environmental impact of hemp cultivation and increasing the quality and yield of the crop.

3 Similar Work

A study focused on determining water stress in plants by monitoring leaf movement, used computer vision to create a non-contact system that focuses on the early detection of plant water stress. The research stated that water is extremely necessary for plant growth and development [12]. Plants use water for photosynthesis as well as to maintain the turgidity of the plant [12]. A plant must retain enough water to keep the plant turgid if there is not enough water the plant will begin to wilt as a result [12]. With low levels of water available to the plant, the next symptom of water stress is the slowing down of the photosynthesis process [12]. This is a critical process to the plant's survival and if not maintained is detrimental to the plant [12]. Thus, an indicator to detect water stress early enough in a plant before the effects of water stress are irreversible is needed [5].

The computer vision methodology used to detect water stress proposed in the study used plant movement and top-projected canopy area as a water stress indicator [12]. Images were captured directly horizontal to the topsoil of the plant, exactly one meter away from the bottom of the plant [12]. Images were captured at a pixel resolution of 640 $\times$ 480, every 15 min during light periods of the day (daylight) [12]. Image processing and feature extraction techniques were applied.

The image processing first applied grey level filters to the image to remove noise or unwanted features in the image. The region of interest is isolated from the image since all plants were in the same position region of interest isolation was simple blob extraction

form the image [12]. A grey level histogram is used to determine the threshold level used later in the image binarization process [12]. A binary image is calculated using the threshold determined from the grey level histogram [12]. Pixel density was hereby used as the indicator for top-projected canopy area (TPCA) [12].

Classification of water-stressed plants was determined by a mean and standard deviation from the well-watered plants in this experiment [12]. The resulting conclusion from this research is that this experiment was able to detect water stress in plants using a top-projected canopy area with low water demand using a non-contact computer vision method [12].

4 Experiment Setup

This study built a hemp leaf dataset using image scraping and dataset pooling techniques. 1200 Images were pooled from various hemp datasets into a single dataset used for this study. Using image annotation ground truth labels of the hemp leaves found in the image are set with their respective locations. Images are then loaded into the program where it is converted from its respective encoding to objects containing the pixel matrix and relevant metadata.

The objects are then separated into a training dataset and a testing dataset. The training dataset is made up of 80% of the hemp leaf objects and the testing dataset is made up of the remaining 20%. Training and testing datasets are selected at random from the population. This form of data sampling allows for greater accuracy in testing the computer vision system. Random data sampling allows for the ability to decrease the size of the dataset used but still retain accuracy, thus reducing computational time.

Random data sampling decreases the size of the dataset used, thus reducing computational time. Using dataset pooling and random data sampling generates limitations and inconsistencies in the end results of the system.

5 Model

The computer vision techniques applied in this study inherits techniques from pattern recognition in computer science and as such will use the established pattern recognition framework. Pattern recognition employs a data processing pipeline framework where data is passed through several stages and each stage performs a set of processing objectives on the data. Computer vision has derived its own pipeline framework from the pattern recognition framework, which this study is based on (Fig. 1).

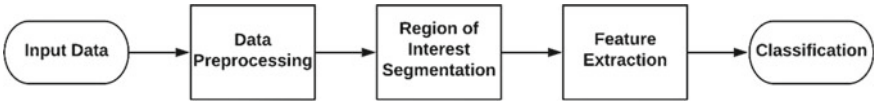


Fig. 1. Computer vision pipeline

This study employs two separate pipelines used to compare different computer vision techniques capable of determining water stress in hemp leaves. The first pipeline

implements traditional computer vision techniques that require explicit implementation and guidance on how to process the data. The second pipeline uses semi-intelligent algorithms to process the image data.

5.1 Pipeline 1

Image Pre-processing
The pre-processing stage focuses on preparing the images for the computer vision process. The preparation includes noise reduction, image normalization and ensuring that only the relevant features are extracted from the image. This study employed several pre-processing techniques, of which each will discussed.

Greyscale Transformation

Greyscale Transformation is an image matrix transformation algorithm that takes an image matrix of several channels and converts it to an image matrix of weighted intensity. This process takes a single pixel and averages out the channels found in that pixel, mainly the Red, Green and Blue Channels, and is converted into a weight between 0 and 255. Greyscale transformation reduces the amounts of channels in an image, reducing the amount of data being processed. This process removes the effect colour will have on the feature extraction process, hereby creating consistency between the images.

Gaussian Blurring

Gaussian Blurring distributes the average weight within a specific kernel space in an image based on a precalculated kernel of a specific size and standard deviation value. The Gaussian kernel is calculated using the Gaussian function, depicted in Eq. 1, which accepts two parameters. The first parameter is the size of the Gaussian kernel. The Gaussian kernel must use kernel sizes that allow for centrosymmetric matrices. The second parameter is the standard deviation represented by σ.

$$G(x, y) = \frac{1}{2\pi\sigma^2} \mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

The calculated kernel is then convoluted on to the image, blurring the image and creating smoother transitions between sharp edges in the image. Gaussian blurring reduces weak edges in the image, thus allowing for only the strongest edges to be detected. Strong edges are important for extracting the shape of the hemp leaf from the image.

Erosion and Dilation

Erosion and Dilation is a two-part technique that aims to remove unnecessary features from an image. Erosion works by removing pixels from the boundaries of objects within an image. This is done by calculating a local minimum within a set window of the image and continuing the calculation for the entire image.

Dilation is the opposite of erosion where the local maximum of a specific window is calculated. The dilation technique adds pixels to the boundaries of the objects in the image. To achieve noise reduction with Erosion and Dilation, the window size for erosion

must be greater than the window size for dilation. This is done to completely remove insignificant objects and blobs of pixels in an image.

OTSU Thresholding

OTSU Thresholding is used to convert an image of varying intensities into a binary image where images are grouped as either foreground or background. OTSU Thresholding uses thresholding variance and standard deviation to select the best threshold for a window within the image. This ensures that each window receives the optimal threshold calculation. Thresholding is used in this study to distinguish distinct edges of the hemp leaves in a given image.

Localization/Region of Interest Segmentation

Localization is a range of algorithms that determine where objects are in an image. The objects still need to be identified in this stage. Region of Interest segmentation is a process of determining if a detected object is the desired object and extracting that object for further processing. This study required that all objects in an image are detected and localization techniques applied to determine where in the image does that object exist. The region of interest segmentation techniques was used to determine if the detected object was a hemp leaf and if positive extract the leaf for water stress detection.

Canny Edge Detection

Canny Edge Detection calculates the gradients of pixel intensities in an image. A drastic change from one gradient to another indicates the presence of an edge. To determine when a gradient change indicates an edge a threshold is set. The Canny Threshold affects the intensities in the image that dictate an edge of an object. This Thresholding is mitigated using the OTSU Thresholding algorithm allowing for better edge detection and a decrease in Edge detection variance between different images.

Contour Analysis

Contour Analysis extracts the shape of an object in an image. A contour is a vector of connected pixels and their gradients that form the boundaries of a shape. From the Canny Edge Detection process where the boundaries of the objects in an image where defined, the Contour Analysis is used to extract the shape of the boundaries. Using the shape of the boundaries, analysis of the shape will help identify if the extracted shape is a hemp leaf or not.

Convex Hull

Convex Hull calculates the furthest points bounding a shape. The convex hull technique is used to identify the tips of a hemp leaf. The hemp leaf is characteristically made up of five to seven appendages each with an acute apex (tip of the leaf). Thus, by counting the number of acute apexes, a weak classifier can be used to determine if an object is a hemp leaf.

Feature Extraction
SIFT

Scale Invariant Feature Transform is a technique used to detect distinct features in an image that is resistant to scale, changes in size, changes in rotation and acute distortion. SIFT calculates key points in an image that are local maxima and minima using the Difference of Gaussian function. The Gaussian function is calculated in different scale-spaces to ensure only the most points are selected despite the scale.

Key points derived from the SIFT algorithm identify the local Maxima or Minima as well as the orientation of the key point. Each key point is used to generate a descriptor of the key point, which is a histogram made up of 128 bins. Each bin is an orientation or gradient of the surrounding pixels. The key point descriptor is used to classify the features on the hemp leaf.

K-Means Clustering

K-Means Clustering is a means of separating data into a specified number of clusters. The *K*-Means algorithm calculates centres based on the number of specified classes. New data is tested against the centres and the data will be assigned to the cluster with the closet centre. For this study, four centres were chosen to represent each expected classification outcome; Dehydrated, Optimal Hydration levels, Over hydrated and Unclassified.

Classification
Support Vector Machine

Due to its popularity, a Support Vector Machine classification technique was used in this study. An SVM aims to separate data into the most effective way possible by calculating a hyperplane that best achieves that. This study implemented a multiclass SVM using both a linear kernel and a Radial Base Function kernel.

5.2 Pipeline 2

The second pipeline implements the same pre-processing, feature extraction and classification as the first pipeline. The distinguishing factor is the localization and Region of Interest Segmentation where and object detection approach was applied. The object detection algorithm will be used to detect the hemp leaves in the image as well as extract the leaves from the image.

Localization and Region of Interest Segmentation
Haar Cascades

Haar Cascading is an object detection algorithm that can be trained to detect specific objects. Haar Cascades calculates pixel intensities in a window, sums up the intensities in a region, and then calculates the difference of neighbouring intensities which creates weak features. The weak features are cascaded, and the calculations are repeated, increasing the feature strength.

6 Results, Critique and Analysis

This study uses a computer vision framework for the results which focuses on a quantitative approach. The focus is on determining the accuracy of the pipeline used as well as comparing the results with the various pipelines. The results are separated by sections. Section 6.1 deals with the accuracy of the parameters chosen for the SVM classifier. The parameters are used in the training of the SVM classifier. Section 6.1 also compares the classifier metrics generated by the system.

6.1 Grid Search and Best Parameter Calculations

Grid search is an algorithm that uses the extracted features from the dataset and calculates the best parameters to be used for the SVM classifier. The grid search uses a trial and error approach to calculate the best parameters. In this study, the grid search is performed on subsets of the dataset of hemp features extracted from the computer vision process.

The evaluation of the computer vision techniques is completed by using the ground truth labels set on each image in the data send and comparing it to the predicted output. The results are inserted into a confusion matrix which is then used to calculate the given matrix that evaluates the chosen pipeline.

Performance measurement of the machine learning techniques employed by this study is illustrated using a Receiver Operating Characteristics curve. Figures 2 and 3, as well as Tables 1 and 2, show the performance of the SVM model with a linear and radial basis function kernel for each of the classes used to classify hemp leaves.

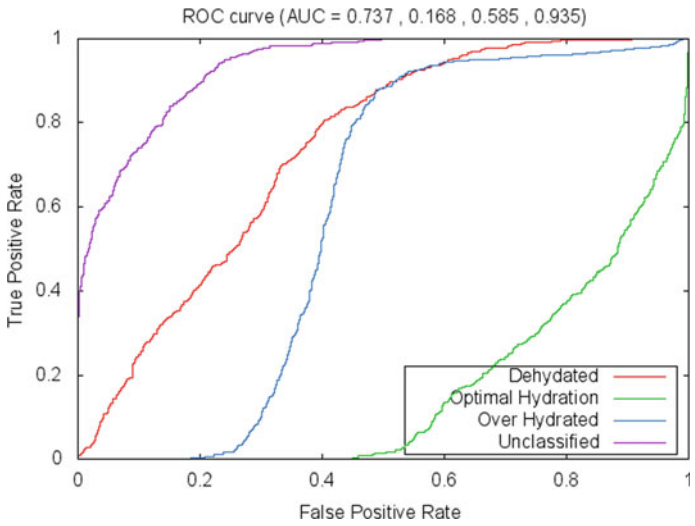


Fig. 2. ROC function for linear SVM model

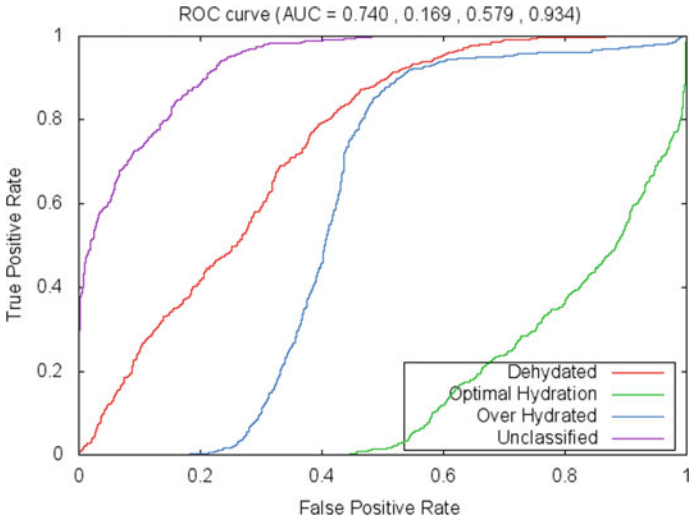


Fig. 3. ROC curve for RBF SVM model

Table 1. Metrics for linear SVM model

Linear SVM kernel				
	Dehydrated	Optimal hydration	Over hydrated	Unclassified
Precision (%)	74.5865	41.6297	50.2219	86.7689
Recall/True positive rate	0.4494	0.5753	0.47245	0.3996
Specificity/True negative rate	0.3139	0.7565	0.4429	0.1653

Table 2. Metrics for RBF SVM model

RBF SVM kernel				
	Dehydrated	Optimal hydration	Over hydrated	Unclassified
Precision (%)	74.5865	41.67	50.3832	86.6882
Recall/True positive rate	0.4489	0.57532	0.4727	0.3996
Specificity/True negative rate	0.31365	0.7566	0.4435	0.1653

The use of OTSU Thresholding drastically increased the consistency in the canny edge detections ability to detect object edges, which allowed for consistent edge detection despite the variability in the image dataset. Using the number of edge points calculated by the convex hull function as an indicator of hemp leaf appendages works well when

dealing with the best-case scenario image. Detection was unsuccessful with images where the tips of the hemp leaf appendages were overlapped or obscured. SIFT features are good for dealing with variability in the image dataset and detecting the exact same image.

7 Conclusion

Cannabis sativa L. known by the colloquial term hemp displays various phenotypic reactions to water stress. Using computer vision techniques, hemp leaf water potential can be determined by analyzing the phenotypic shape of the leaf. This study focused on the computer vision techniques that can be used to extract the water potential of the hemp leaf. The study used two pipelines, with the first focusing on traditional computer vision algorithms and the second pipeline using a more object detection-oriented approach.

The first pipelines architecture consisted of using Greyscale Transformation, Gaussian Blurring, Erosion and Dilation, and OTSU Thresholding for the pre-processing of the hemp images. The localization and Region of Interest Segmentation used a combination of Canny Edge Detection, Contour analysis, and Masking. The Scale Invariant Feature transform technique was the main feature extraction technique used in this study due to its accuracy over different scales, rotations, and sizes of the image. The classification used was an SVM classifier that used both a Linear Kernel and Radial Base Function kernel. The second pipeline focused mainly on using Haar Cascades for its object detection, using it to detect the hemp leaf in the image.

The results gained from this study showed that the system displayed poor accuracy and overall precision. This is caused by many problems in the design and implementation of the system. The main issue that affects the accuracy of the system is poor problem space decomposition and poor designing of the system. By not having a publicly available comprehensive dataset, it affects the overall training of the machine learning algorithms as well as the accuracy of the evaluation metrics. This study showed that there is great potential for determining the water stress in hemp leaves using computer vision as well as highlighting issues that can affect the accuracy of such a computer vision system. There were certain shortcomings such as inconsistent dataset and pipeline design flaws. This study, however, proves that this is a viable concept that can be further explored.

References

1. Amaducci, S., Gusovius, H.-J.: Hemp—cultivation, extraction and processing. In: Industrial Applications of Natural Fibres, pp. 109–134 (2010)
2. Callaway, J.C.: A more reliable evaluation of hemp THC levels is necessary and possible. J. Ind. Hemp **13**, 117–144 (2008)
3. Cosentino, S.L., Riggi, E., Testa, G., Scordia, D., Copani, V.: Evaluation of European developed fibre hemp genotypes (*Cannabis sativa* L.) in semi-arid Mediterranean environment. Ind. Crops Prod. **50**, 312–324 (2013)
4. Florian, M.-L.E., Kronkright, D.P., Norton, R.E.: The Conservation of Artifacts Made from Plant Materials. Getty Conservation Institute (2002)

5. Gao, C., Cheng, C., Zhao, L., Yu, Y., Tang, Q., Xin, P., et al.: Genome-wide expression profiles of hemp (*Cannabis sativa* L.) in response to drought stress. Int. J. Genomics **2018**, 1–13 (2018)
6. Kacira, M., Ling, P.P., Short, T.H.: Machine vision extracted plant movement for early detection of plant water stress. Trans. ASAE **45** (2002)
7. Meyer, W.S., Green, G.C.: Plant indicators of wheat and soybean crop water stress. Irrig. Sci. **2** (1981)
8. Oosterhuis, D.M., Walker, S.: Stomatal resistance measurement as an indicator of water deficit stress in wheat and soybeans. S. Afr. J. Plant Soil **4**, 113–120 (1987)
9. Reisinger, P., Lehoczky, É., & Komives, T.: Competitiveness and precision management of the noxious weed *Cannabis sativa* L. in winter wheat. Commun. Soil Sci. Plant Anal. **36**, 629–634 (2005)
10. Sawler, J., Stout, J.M., Gardner, K.M., Hudson, D., Vidmar, J., Butler, L., et al.: The genetic structure of marijuana and hemp. PLoS ONE **10**, e0133292 (2015)
11. Tang, K., Fracasso, A., Struik, P.C., Yin, X., Amaducci, S.: Water- and nitrogen-use efficiencies of hemp (*Cannabis sativa* L.) based on whole-canopy measurements and modeling. Front. Plant Sci. **9** (2018)
12. Tourangeau, W.: Re-defining environmental harms: green criminology and the state of Canada's hemp industry. Can. J. Criminol. Crim. Justice **57**, 528–554 (2015)

Active Surface Calculation Method Using Kinetic-Hydrodynamic Model

Yury Nikitchenko, Sergey Popov, and Alena Tikhonovets(✉)

Moscow Aviation Institute (National Research University), Moscow, Russia
nikitchenko7@yandex.ru, flowmech@mail.ru, tikhalena@gmail.com

Abstract. The problem of flow over a thin plate of infinite magnitude installed across the flow is considered. The frontal surface of the plate absorbs gas. For calculations, we used a mathematical flow model containing a combination of the Navier-Stokes-Fourier model and the model kinetic equation of polyatomic gases. The features of the mathematical implementation of the combined model are described. The calculations were performed for a supersonic flow with a Mach number of 2.31 for a Knudsen number of 0.1…0.001 and a plate surface absorption coefficient from 0 to 1. The obtained flow fields were compared with solutions of the model kinetic equation of polyatomic gases. The drag coefficient of the plate was compared with known experimental data. For all considered flow parameters, a satisfactory agreement with the known data is obtained. It is shown that there are no gaps in the derivatives of gas-dynamic parameters in the joining region of the kinetic and hydrodynamic components of the model. The increase in the computational efficiency of the model with respect to the solutions of model kinetic equations is estimated. The conclusion is drawn on the suitability of the considered kinetic-hydrodynamic model for describing highly nonequilibrium flows.

Keywords: Polyatomic gases · Navier-Stokes-Fourier model · Model kinetic equation · Combined model · Dynamic nonequilibrium · Absorbing surface

1 Introduction

Models of continuum mechanics (hydrodynamic models) do not allow a physically adequate description of the interaction of a gas medium with an active surface. Some approximate relations for this process can be found in [1, 2]. Models of molecular kinetic theory (kinetic models) solve this problem relatively simply [2]. The flows of mass, momentum, and energy to the surface is determined from the solution of the kinetic equation. Using these parameters and taking into account the properties of the streamlined surface, the weight function (the distribution function of the molecules over the velocities) of the molecules reflected from the surface is restored. The set of weight functions of the molecules incident and reflected by the surface allows one to determine any parameter of the gas medium at the boundary point.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_23

At the same time, kinetic models are not of practical interest for describing dense (Kn $<$ 0.1) gas flows, due to the huge volume of computational operations and the RAM of a computing device. To date, a number of hybrid and combined models have been developed that use the kinetic description of the flow in highly nonequilibrium (highly gradient) regions and the hydrodynamic description in other regions. A brief review of such models is presented in [3]. In this work, as well as in [4], a kinetic-hydrodynamic model (KHM) is presented, which allows one to obtain smooth solutions under conditions of high nonequilibrium. When calculating the profiles of shock waves and near-wall flows, the KHM yielded results that are in satisfactory agreement with the experimental and calculated data.

In [5], the problem of supersonic flow around a thin plate mounted across a flow was considered. The frontal surface of the plate absorbed gas. The problem was solved on the basis of the model kinetic equation (MKE). The effect of a qualitative change in the dependence of the drag of the plate on the absorption coefficient of the surface with a change in the Knudsen number was shown. In the present work, this problem is solved using the KHM. The results are compared with the solution of the MKE. The main attention is paid to increasing the computational efficiency of KHM in comparison to MKE. For this purpose, a special algorithm for the numerical implementation of the kinetic component of the KHM has been developed. The accuracy of determining the drag coefficient is estimated.

2 Basic Assumptions and Notations

The flows of monocomponent perfect gases are considered. All expressions are written for polyatomic gases. In the case of monatomic gases, the expressions remain valid after obvious transformations (Table 1).

Table 1. Notations used

t, x_i	Time and geometric coordinate
ξ_i, ε	Molecular velocity and rotational energy
L, λ_∞	Plate height and mean free path of the molecule in an undisturbed flow
m_0, n, $\rho = m_0 n$	Molecular mass, molecular concentration, and gas density
u_i, $c_i = \xi_i - u_i$	Group (macroscopic) and thermal molecular velocities
f	Weight function
c_x, α	Drag coefficient and absorption factor of the plate
M, Kn, Pr	Mach, Knudsen, and Prandtl numbers

The velocity space integral is denoted by $\int \ldots \mathrm{d}\mathbf{c} \equiv \int_{-\infty}^{+\infty} \mathrm{d}c_1 \int_{-\infty}^{+\infty} \mathrm{d}c_2 \int_{-\infty}^{+\infty} \ldots \mathrm{d}c_3$. In considered models, the following is accepted: $\mathrm{Pr} = 4\gamma/(9\gamma - 5)$.

3 The Problem of Flow Over a Plate with an Absorbing Surface

We consider a plane flow in rectangular Cartesian coordinates $0XY$. The undisturbed flow moved in the direction of the $0X$-axis through the left boundary of the computational domain. The plate is located spanwise. The problem is symmetric about the plane $y = 0$. The scheme of the problem is shown in Fig. 1.

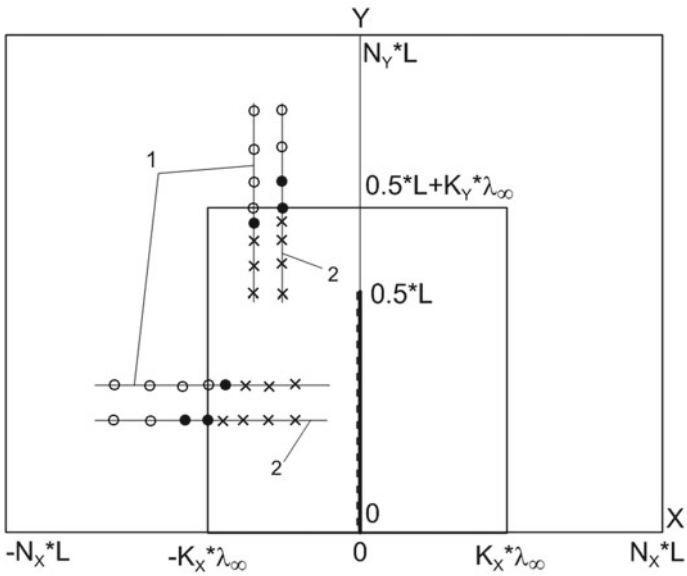


Fig. 1. Scheme of the computational domain. Bold line is a plate, the absorbing surface is indicated by a dotted line; K_X is the kinetic half-region width in the mean free path of the molecule; N_X is the width of the hydrodynamic half-region in units of plate height; N_Y is the height of the hydrodynamic semi-region in units of plate height; 1 is a scheme of an arrangement of joining nodes for the hydrodynamic region; 2 is a scheme of the joining nodes for the kinetic region; Black circle—joining nodes; White circle—nodes of the hydrodynamic grid; x—nodes of the kinetic grid

The size of the entire computational domain, which is also the external size of the hydrodynamic region, was: $N_x = 3 \ldots 4$, $N_y = 2 \ldots 3$. The size of the kinetic region, which is also the internal size of the hydrodynamic region, was taken $K_x = 10 \ldots 20$, $K_y = 10 \ldots 15$.

The boundary conditions at the external boundaries of the computational domain are set as follows: in the section $x = -N_xL$, the parameters of the undisturbed flow are accepted, in the section $y = N_yL$ $\partial^2/\partial y^2\big(\rho, u_x, u_y, T\big) = 0$ is accepted, in the section $x = N_xL$ $\partial^2/\partial x^2\big(\rho, u_x, u_y, T\big) = 0$ is accepted.

On the frontal surface of the plate (dotted line in Fig. 1), the surface absorption coefficient was set

$$\alpha = (J_{\text{down}} - J_{\text{reflect}})/J_{\text{down}} \tag{1}$$

where

$$J_{\text{down}} = \int_{-\infty}^{+\infty} \mathrm{d}c_y \int_{-\infty}^{+\infty} \mathrm{d}c_z \int_{0}^{\infty} \mathrm{d}\varepsilon \int_{0}^{+\infty} \xi_x f \, \mathrm{d}c_x,$$

$$J_{\text{reflect}} = -\int_{-\infty}^{+\infty} \mathrm{d}c_y \int_{-\infty}^{+\infty} \mathrm{d}c_z \int_{0}^{\infty} \mathrm{d}\varepsilon \int_{-\infty}^{0} \xi_x f \, \mathrm{d}c_x \tag{2}$$

are the densities of the incident and reflected molecular flow.

The following relation is considered as the drag coefficient of the plate:

$$c_x = \left(\int_{F_1} \left(P_{xx} + \rho u_x^2 \right) \mathrm{d}F - \int_{F_2} P_{xx} \mathrm{d}F \right) / \left(p_\infty + \rho_\infty u_\infty^2 \right) S \tag{3}$$

where F_1, F_2—frontal and bottom surfaces of the plate; $S = L \times 1$ is the area of the plate unit span. In contrast to the drag coefficient in the traditional sense, this dependence uses the impulse flow of an undisturbed flow as a characteristic quantity. Such a universal form of drag coefficient allows, for example, to consider the problem of flow past absorbing surfaces located in a stationary gas. This positive feature of the coefficient determined its choice.

In the present work, the temperature of the entire plate is taken equal to the stagnation temperature of the gas.

4 The Method of Numerical Implementation of the Kinetic Component of the Combined Model

Most of the memory consumed by the KHM is occupied by the weight function f, since the dimension of the phase space, even after formal integration over the coordinates ξ_z and ε [5], is twice the dimension of the geometric space $0XY$. However, when solving stationary problems, there is no need to store all the values of f in the computer's memory. In this problem, for the selected half-spaces of molecular velocities ξ_x and ξ_y, it is enough to consider one quadrant of the velocity space. Figure 2 shows such a quadrant for the half-spaces $\xi_x \geq 0$ and $\xi_y \geq 0$.

Figure 3 shows the lines used to calculate f on the n and $(n+1)$ lines. Advancing along Y is performed from bottom to top, i.e., $\xi_y \geq 0$. The function f is written into a four-dimensional array in which the Y-coordinate has three values: 0, 1, and 2. The values of the function are calculated in the section of the array $y = 0$ from the values in sections $y = 1$ and $y = 2$.

After determining f at all nodes of the selected quadrant (velocity space) and at all nodes of the section $y = 0$ (geometric space), incomplete moments of the weight function are calculated and written into memory. Incomplete moments are calculated for molecular velocities ξ_i.

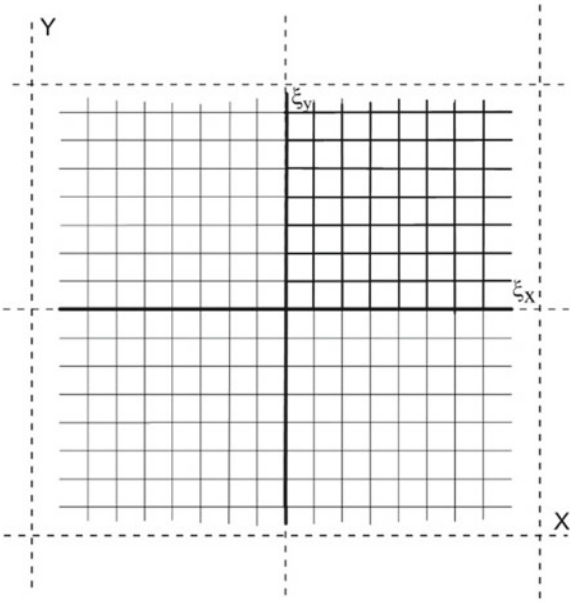


Fig. 2. Computational grid in the velocity space. Dashed lines—geometric grid, solid lines—velocity grid, thickened solid lines—calculated quadrant of the velocity grid ($\xi_x \geq 0$, $\xi_y \geq 0$)

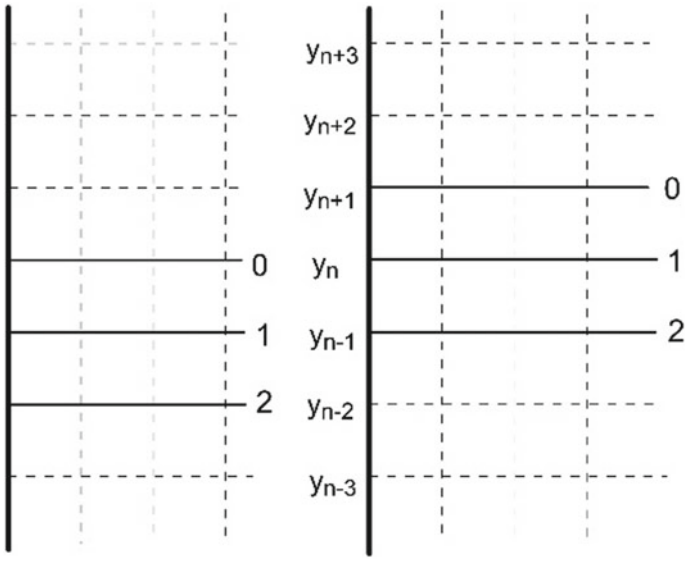


Fig. 3. Algorithm for calculating the weight function at the n and $(n+1)$ values of the Y-coordinate. The bold line is the plate

When moving to the next $(n + 1)$ row of the main geometric grid, the section of the array $y = 0$ is renumbered (but not overwritten) as $y = 1$, the section $y = 1$ as $y = 2$. The calculation of the section $y = 0$ continues according to the scheme described above.

After calculating the incomplete moments, the macroparameters of the gas medium are determined in all quadrants of all geometric nodes.

When solving the problem of flow over a plate with $\mathrm{Kn} = 0.01$, the described technique made it possible to reduce the memory size by approximately two orders of magnitude.

5 Calculation Results

The calculations were performed for a diatomic gas at $M = 2.31$ and $\mathrm{Kn} = 0.1 \ldots 0.01$. For the selected Mach number, there is a sufficiently large set of experimental and calculated data in the transition region of the flow, i.e., for the indicated Knudsen numbers. A selection of data for the drag coefficient of plates c_x can be found in [6]. The calculation results, including the flow fields, were compared with calculations according to the MKE [7, 8, 9], which are in satisfactory agreement with the data of [6].

Figure 4 shows the dependence $c_x(\alpha)$. At $\mathrm{Kn} = 0.1$, the calculations according to the MKE and KHM coincide, since at these Knudsen numbers, almost the entire computational domain was in the kinetic component of the KHM, and its hydrodynamic component served as boundary conditions at the outer boundary of the computational domain.

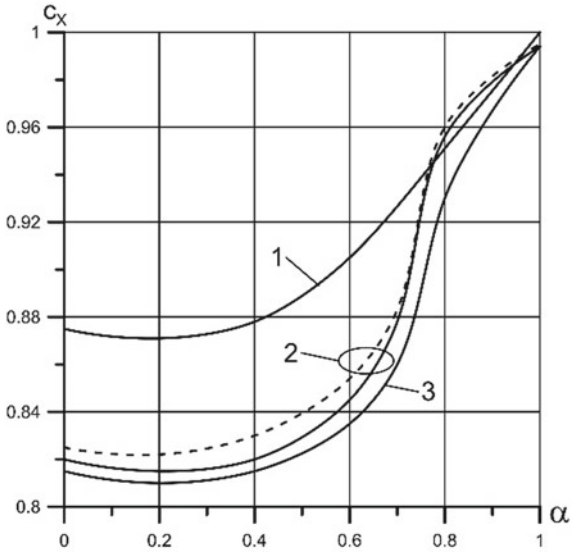


Fig. 4. Dependence of the drag coefficient of the plate on the absorption coefficient of its frontal surface. 1—MKE [7] and KHM, $\mathrm{Kn} = 0.1$; 2—dotted line MKE, $\mathrm{Kn} = 0.1$, 2—solid KHM

For $\mathrm{Kn} = 0.01$ and $\alpha < 0.7$, there is some discrepancy between the models used. As follows from the graphs below, at these values of the parameters, the entire highly

gradient region is in the hydrodynamic component of the KHM. The methods of continuum mechanics are significantly inferior in the accuracy of calculating high-gradient regions to the methods of molecular kinetic theory. This, apparently, is the reason for the discrepancy between the calculation results.

The following graphs show the density distributions, projections of the velocity u_x, and temperature along the X-axis at Kn $= 0.01$. The flow region in front of the frontal surface of the plate ($x < 0$) in the plane of symmetry ($y = 0$). In this region, the greatest flow disequilibrium takes place.

In Fig. 5a, $\alpha = 0.6$, the entire highly gradient flow region is in the hydrodynamic component of the KHM (to the left of the joining region). The density, velocity, and temperature profiles are shifted downstream relative to the MKE profiles. A significant difference is observed in the density profiles.

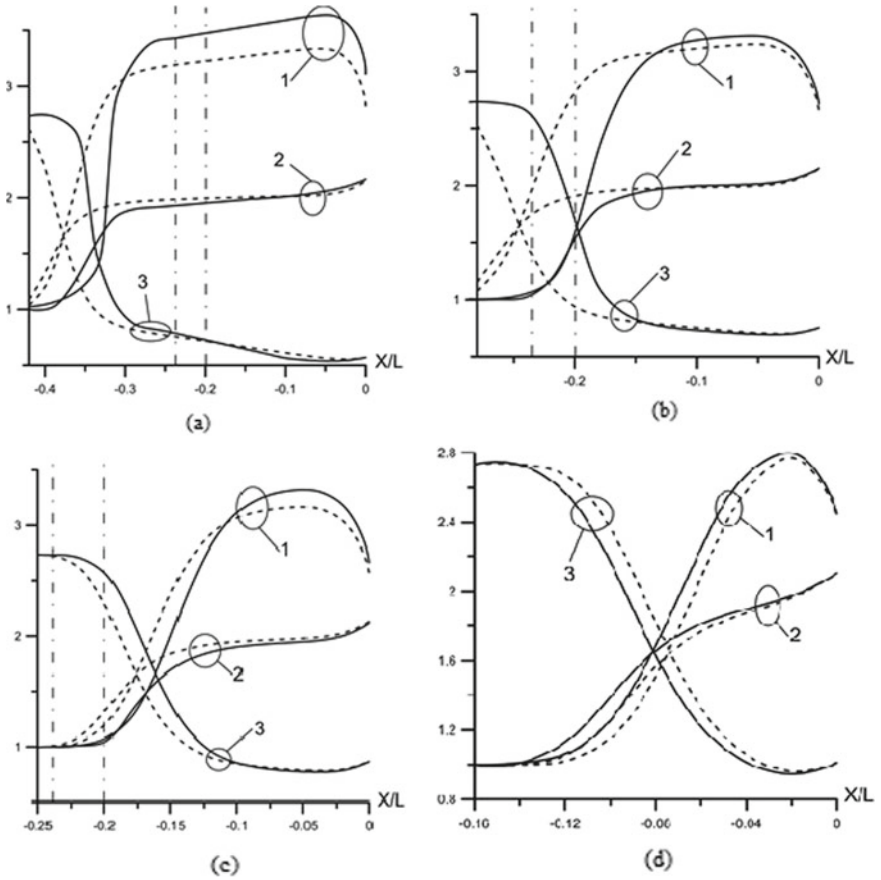


Fig. 5. Distribution of density, velocity, u_x and temperature in front of the frontal surface. Kn $=$ 0.01. **a** $\alpha = 0.6$; **b** $\alpha = 0.7$; **c** $\alpha = 0.75$; **d** $\alpha = 0.8$. 1—density; 2—temperature; 3—u_x. Dotted lines—MKE; solid lines—KHM; vertical dash-dotted lines are the boundaries of the joining region

As the absorption coefficient increases, the high-gradient flow region shifts more and more to the kinetic subdomain (to the right of the joining region). The profiles of MKE and KHM are close in both location and parameter values, Fig. 5b, c. Particular attention should be paid to the fact that in the joining region, KHM gives smooth (without derivative discontinuities) solutions. This positive feature of the model developed by the authors is analyzed in detail in [3, 4, 10].

At $\alpha = 0.8$, Fig. 5d, the high-gradient region is completely in the kinetic component of the KHM. The right border of the stitching region has the coordinate $x = -0.2$ and does not fall into the figure field. The profiles of KHM and MKE almost coincide. A relatively small difference is apparently due to the fact that for calculations at $\mathrm{Kn} = 0.01$ according to the MKE a geometric grid with a step of $2\lambda_\infty$ was used, and calculations in the kinetic component of the KHM were performed with a half-step. The need to overestimate the step in the MKE model was associated with the technical capabilities of the used computing tools.

6 Conclusion

The calculation results show that the combined kinetic-hydrodynamic model allows physically adequate description of the processes occurring in the transition region of the gas medium flow. In the field of joining of the kinetic and hydrodynamic components of the model, there are no gaps in the derivatives of the gas-dynamic parameters.

The KHM model allows one to set the boundary conditions on absorbing surfaces. The values of such an integral characteristic as $c_x(\alpha)$ calculated by the KHM are in satisfactory agreement with the results of calculations according to the MKE.

When calculating relatively dense gases ($\mathrm{Kn} = 0.01$), the method described above for the numerical implementation of the kinetic component of the KHM allows us to reduce the memory consumption of the computing device by about two orders of magnitude compared to the traditional use of combined or hybrid models. Note that the use of KHM to solve this problem at $\mathrm{Kn} = 0.01$, even with its traditional use, reduces the memory and processor time by about two orders of magnitude compared to MKE.

The developed KHM model can be used in a wide range of Knudsen numbers.

In the future, it is planned to conduct a study of the KHM in the region of sub- and hypersonic flows, that is, in a wide range of Mach numbers.

This work was financially supported by the Ministry of Education and Science of the Russian Federation, project No. FSFF-2020-0013.

References

1. Salbreux, G., Jülicher, F.: Mechanics of active surfaces. Phys. Rev. E **96**, 032404 (2017)
2. Rioux, R.W.: The rate of fluid absorption in porous media. Electronic Theses and Dissertations, 234 (2003)
3. Nikitchenko, J.A., Popov, S.A., Tikhonovets, A.V.: Composed kinetic-hydrodynamic model of polyatomic gas flow. Matem. Modelirovanie **31**(2), 18–32 (2019)
4. Nikitchenko, Y., Popov, S., Tikhonovets, A.: Special aspects of hybrid kinetic-hydrodynamic model when describing the shape of shockwaves. In: Rodrigues, J., et al. (eds.) Computational Science—ICCS 2019. Lecture Notes in Computer Science, vol. 11539. Springer, Cham (2019)

5. Glinkina, V.S., Nikitchenko, Y.A., Popov, S.A., Ryzhov, Y.A.: Drag coefficient of an absorbing plate set transverse to a flow. Fluid Dyn. **51**(6), 791–798 (2016)
6. Bondarev, E.N., Dubasov, V.T., Ryzhov, J.A., Svirshhevskij, S.B., Semenchikov, N.V.: Ajerogidromehanika. M. Mashinostroenie 608 s (1993)
7. Nikitchenko, Y.A.: Model kinetic equation for polyatomic gases. Comput. Math. Math. Phys. **57**(11), 1843–1855 (2017)
8. Thompson, M.J.:. An Introduction to Astrophysical Fluid Dynamics. Imperial College Press, London (2006)
9. Rovenskaya, O.I., Croce, G.: Numerical simulation of gas flow in rough micro channels: hybrid kinetic–continuum approach versus Navier–Stokes. Microfluid. Nanofluid. **20**, 81 (2016)
10. Berezko, M.E., Nikitchenko, Y.A., Tikhonovets, A.V.: Joint of kinetic and hydrodynamic models by the example of the Couette flow. Tr. MAI, No. 94 (2017). http://trudymai.ru/published.php?ID=80922

A Three-Layer Architecture for Intelligent Intrusion Detection Using Deep Learning

Liu Zhiqiang, Lin Zhijun, Gong Ting, Shi Yucheng, and Mohi-Ud-Din Ghulam(✉)

Northwestern Polytechnical University, Xi'an, China
gmohiudin@hotmail.com

Abstract. Recently, the increasing number of machine learning algorithms has been used in network intrusion detection system (NIDS) to detect abnormal behaviors in the network. Many available datasets were created to evaluate the performance of the model, such as KDD CUP99 and NSL-KDD. However, with the increasing scale of data and the emergence of advanced attacks, conventional machine learning algorithms can hardly perform well. Fortunately, the development of deep learning provides new direction for solving these problems. In this paper, in order to detect novel attacks in a network and improve detection efficiency, we proposed a flexible framework based on deep neural network (DNN). In our framework, we apply different feature reduction methods and activation functions to get the best performance. Moreover, through changing hyper-parameter of the model, we select better network structure. To evaluate our framework, we select ISCX 2012 and CICIDS 2017 as a benchmark and apply the proposed framework to these datasets. As a result, we observe high accuracy rate and low FAR for both binary and multi-class classifications. Overall, our proposed framework is universal and useful for detecting zero-day attacks.

Keywords: NIDS · DNN · ISCX 2012 · CICIDS 2017

1 Introduction

In today's large-scale Internet, account of traffic flow will be generated every day. Some of them that are different from normal behavior are identified as malicious attacks. With consistent generation and evolution of malicious attacks in networks, our network environment is facing serious threats. For example, data leaks cost 350 million dollars per year for Yahoo because of bitcoin [1]. Therefore, cybersecurity is extremely challenging for governments and individuals. In the past years, the intrusion detection system(IDS) was proposed to detect attacks on time for cybersecurity.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_24

IDS is an actively automatic detection method in network level. Essential aspects of IDS are to classify and prevent intrusions.IDS can be divided into misuse detection and anomaly detection-based intrusion detection systems. Misuse detection uses pre-defined tags and filter-based attack detection. This kind of method relies heavily on manual tag dataset, which will cause a lot of waste and threat of human and material resources. Moreover, this kind of method is only limited to identify known attacks. However, when facing the complex network attack types, only based on anomaly detection method can identify the network attacks which did not appear. However, because the false positive rate is based on anomaly detection is very high, only using this kind of method will lead to large number of false alarms affecting the efficiency and accuracy of the IDS. Therefore, in most common group application scenarios, intrusion detection system based on anomaly detection and misuse detection will be used at the same time.

In recent years, many conventional machine learning algorithms have been used in IDS to detect attacks in network. However, with the increasing scale of network and emergence of advanced attacks, these models initiate many problems, such as high false positive rate [2] and broader attack range [3]. Fortunately, with the rapid development of deep learning, deep learning has achieved excellent results in many domains. It also leads to a new research direction for IDS.

Due to dynamic characteristics and consistent changing of attacks, using deep learning technology to detect intrusions is a trend to the modern intrusion detection system. The deep learning model is helpful to develop a flexible and effective intrusion detection system and classify the unpredictable and complex network situations. Overall, deep learning improves the efficiency of IDS. So based on deep neural network (DNN), which is one kind of models in deep learning, we propose creative framework to improve the performance of IDS.

2 Related Work

2.1 Review Relevant Literature

In [4], DNN was proved that it was reliable and useful to detect different types of attack in a large number of data. The primary evaluation of this paper is based on different types of processors and cores. However, NSL-KDD dataset used in this paper does not contain new types of attack in current network. Moreover, based on new processor we used, it did not achieve the expected effect. In [5], authors used shallow learning rate. They ran model for 1000 number of epochs and compared the layers of DNN from one to five. Finally, they decided to choose the three layers as the superior learning algorithms layers. The drawback of this method is requirement for more adversarial environments. Moreover, the KDD CUP is too old to represent current traffic flow.

In [6], the paper presented a new method to identify attacks by combining with the feed-forward neural network and convolutional neural network. In [7], authors utilized three different DNN structures, including convolutional neural networks, autoencoders, and recurrent neural networks, to identify attacks by

using NSL-KDD. However, the accuracy rate of this method did not arrive at 90% so that we do not consider them as powerful methods.

In [8], the paper explored and developed an effective intrusion detection system based on deep learning model. Authors' comprehensive evaluation of experiments of DNNs and other classical machine learning classifiers is shown on various publicly available benchmark malware datasets. The optimal network parameters and network topologies for DNNs are chosen through the following hyper-parameter selection methods with KDDCup 99 dataset. The DNN model which performed well on KDDCup 99 is applied to other datasets, such as NSL-KDD, UNSW-NB15, Kyoto, WSN-DS, and CICIDS 2017, to conduct the benchmark. However, this paper did not mention FAR which is one significant metric in IDS. Moreover, the different number of features between distinct datasets will influence the performance. However, this paper did not consider the feature reduction method.

3 Proposed Method

Dataset
While the usage of network, more researches are working on network security. According to the researches in the past ten years, it is becoming common to combine machine learning technology with intrusion detection system. Therefore, various datasets used for IDS were created by researches.

3.1 ISCX 2012

In the past decade, more and more intrusion detection systems have been proposed. In order to evaluate the effectiveness of intrusion detection system, Canadian Institute of Network Security provides a new dataset named ISCX 2012, which includes the latest security threats and attack functions.

3.2 CICIDS2017

CICIDS 2017 dataset contains benign and the most up-to-date common attacks, which resembles the true real-world data. It also includes the results of the network traffic analysis using CICFlowMeter with labeled flows based on the time stamp, source and destination IPs, source and destination ports, protocols and attack. Also, available is the extracted features definition.

FRAMEWORK
With the emergence of advanced attacks and increasing scale of network traffic, how to detect attacks effectively is becoming a challenge. Recently, most researches combine artificial intelligence and intrusion detection system with improving detection efficiency. However, most of them used old datasets so that they could not consider advanced attacks. Moreover, high FAR is also a significant problem in these researches.

In this section, we propose a three-layer architecture to identify attacks more effectively and efficiently. Our proposed framework is based on DNN. We consider the influence of feature reduction methods and activation functions, so we apply different parameters on models. By comparing capability of different models, we select the better performance model. Also, we change hyper-parameters to obtain a better model structure.

1. We select principal components analysis (PCA) and autoencoder (AE) to construct new datasets with different numbers of features. In this step, we select 86%, 67% and 51% number of features, respectively.
2. We apply different activation functions on DNN models to identify the most suitable one. Activation functions we used are "ReLU," "Sigmoid," and "tanh," which are widely used in researches.
3. Combining both first and second steps, we make a comparison of the performance of models with different activation functions and number of features. As a result, we choose the better performance model, which uses best feature reduction method, number of features, and activation function.
4. We consider hidden layers and neurons as important hyper-parameters for DNN. Based on the output from third step, we change hidden layers and neurons to identify better hyper-parameter for our model.

4 Results and Evaluation

In this section, we use ISCX 2012 and CICIDS 2017 as a benchmark. We divided each dataset into a training set and a testing set in a ratio of three to two. The training set is used to train the model, and the testing set is used for evaluation.

We consider the binary and multi-class classifications on ISCX 2012 and CICIDS 2017. Specifically, we apply the proposed framework for each dataset and classification. By comparing the accuracy of different models, we select the best parameter, that is best activation function, feature reduction method, and number of features. Moreover, we establish completed DNN model using best parameter and change hyper-parameters to gain better results. In the end, We present different metrics to evaluate the capability of the model. We analyze the differences between the two datasets by comparing metrics. Moreover, we display the accuracy of each attack for multi-class classification.

4.1 Binary Classification

Binary classification is a common problem in intrusion detection systems. Traffic flow is defined as two types. One is normal behavior and another is abnormal behavior. In this part, we build binary classifiers for ISCX 2017 and CICIDS 2017 to detect anomalies.

ISCX 2012 Figure 1 displays the accuracy of binary classification for ISCX 2012. On the left side, we use PCA as the feature reduction method. On the right side, we utilize AE to reduce number of features.

a. Feature Reduction and Finding best activation function Using PCA
When we use PCA as a feature reduction method, we observe that activation function "Sigmoid" performs best in 86% number of features. Therefore, we choose "sigmoid" and 86% number of features as the best parameter in this step.
b. Feature Reduction and Finding best activation function Using AE
We find out that the performance of activation function "Sigmoid" becomes better with increasing of the number of features. In 86% number of features, the accuracy is highest in all cases. We also select "sigmoid" and 86% number of features as the best parameter in this step.
c. Selecting a feature reduction method between PCA and AE
We compare the outputs from step a and step b. We gain that accuracy of the classifiers using PCA is higher than this using AE, so we consider PCA is better feature reduction method.

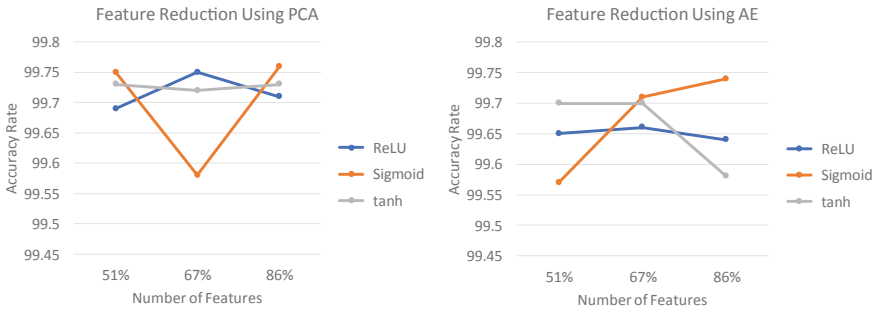


Fig. 1. ISCX 2012 binary classification using PCA and AE

CICIDS 2017 In this part, we consider binary classification of CICIDS 2017. On the left side of Fig. 2, we use PCA and display the accuracy of different parameters. On the right part, we consider AE as the feature reduction method.

a. Feature Reduction and Finding best activation function Using PCA
We observe that activation function "tanh" plotted with gray line generally performs better than others. In 67% number of features, the accuracy of "tanh" arrives peak which is also the highest value among all lines. As a result, we select "tanh" and 67% number of features as our best parameter by using PCA.

b. Feature Reduction and Finding best activation function Using AE
We find that activation function "ReLU" maintains stable performance under different conditions. "Sigmoid" performs the worst. In 67% of the feature numbers, "tanh" is best performance and gains 98.5% accuracy rate. In conclusion, we choose "tanh" and 67% number of features in this step.
c. Selecting a feature reduction method between PCA and AE
Comparing accuracy between PCA and AE using best parameters gained in step a and b, respectively, we select PCA as better the feature reduction method.

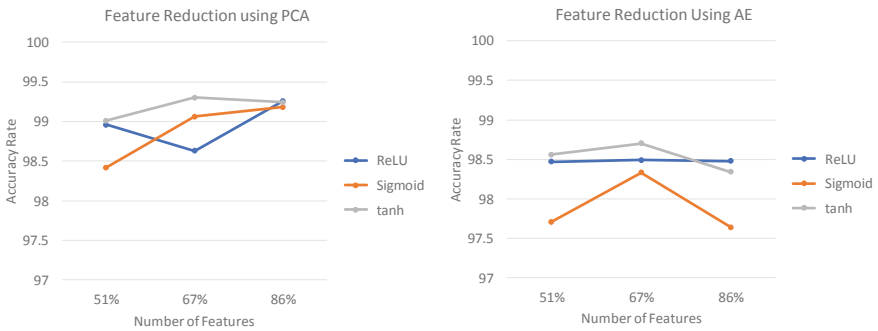


Fig. 2. CICIDS 2017 binary classification using PCA and AE

4.2 Multi-class Classification

In multi-class classification, the target label contains multiple categories. Multi-class classifier predicts a specific category for each type of attacks. Multi-class classification is more able to reflect situation in real networks because of various attacks.

To identify specific types of attack, we present the multi-class classification of CICIDS2017 and identify different types of attack in the dataset. For ISCX 2012, because of lacking multi-class labels, we do not consider multi-class classification.

CICIDS 2017

a. Feature Reduction and Finding best activation function Using PCA
On the left side of Fig. 3, "ReLU" performs better than others in 67% and 86% number of features. Moreover, the accuracy of "ReLU" is highest in 86% number of features. We select "ReLU" and 86% number of features in this step.

b. Feature Reduction and Finding best activation function Using AE
We analyze the right side of Fig. 3 and observe that activation function "tanh" performs best in 51% number of features; this is 98.5%. It is also the best performance by using AE. Therefore, we choose "tanh" and 51% as the best parameter.
c. Selecting a feature reduction method between PCA and AE
We compare the best accuracy of step a and step b. As a result, we consider PCA is better than AE.

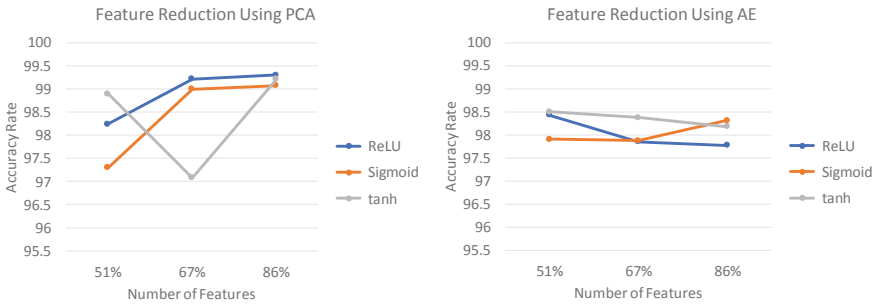


Fig. 3. CICIDS 2017 multi-class classification using PCA and AE

4.3 Hyper-parameter

We select three different hyper-parameters in this part. In hyper-parameter1, we select three hidden layers and one hundred and twenty neurons. In hyper-parameter2, we design the DNN model which includes three hidden layers and one hundred and eighty neurons. In hyper-parameter3, we utilize five hidden layers and one hundred and eighty.

ISCX 2012 In Fig. 4, we show the effect of hyper-parameters for binary classification for ISCX 2012. With the change of hyper-parameters, the prediction results of our model stay between 99.6 and 99.8%. When we use hyper-parameter1, the accuracy rate is highest so that we consider hyper-parameter1 is the better option of binary classification for ISCX 2012.

CICIDS 2017 In Fig. 5, we present the performance of binary and multi-class classifications for CICIDS 2017. For binary classification, the accuracy rate is lowest in hyper-parameter2 and becomes best in hyper-parameter3. For multi-class classification, the model performs best by using hyper-parameter1.

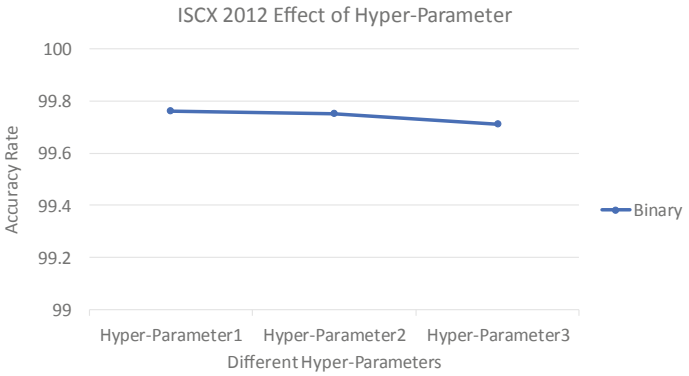


Fig. 4. ISCX 2012 using different hyper-parameters

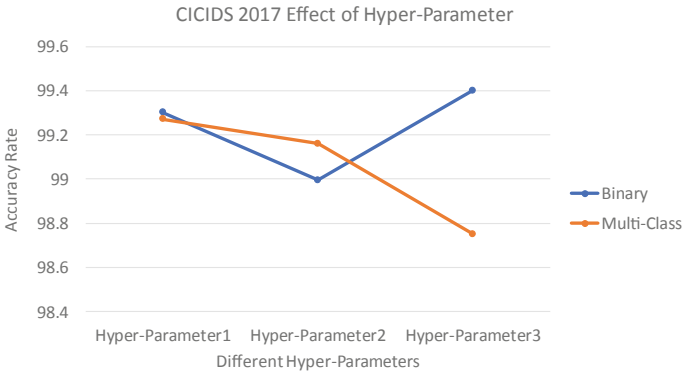


Fig. 5. CICIDS 2017 using different hyper-parameters

4.4 Comparison Between ISCX 2012 and CICIDS 2017

In this part, we use distinct performance metrics, such as precision, recall, and FAR(False Alarm Rate), to evaluate our models for binary classification. Moreover, we make a comparison between ISCX 2012 and CICIDS 2017.

From Fig. 6, in general, we observer that our method shows excellent performance, such as high accuracy and F1-score, that is over 99%, and low FAR, that is about 0.2%, for ISCX 2012 and CICIDS 2017. By comparison, the accuracy and FAR of ISCX 2012 perform better than CICIDS 2017. However, the AUC of CICIDS 2017 is higher than ISCX 2012.

The ROC curve is one of the important aspects for evaluating the quality of classifiers. The larger the area under the ROC curve, the better the performance of model. In Fig. 7, we observe our method performs well because the area under the ROC curve almost arrives one which is the maximum value of the area. Moreover, we find the ROC curve of CICIDS 2017 is more smooth and the

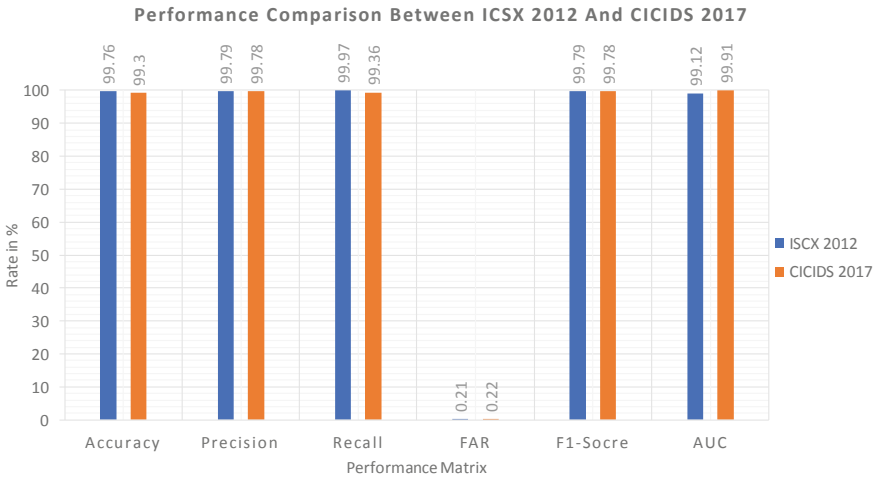


Fig. 6. Performance comparison between ISCX 2012 and CICIDS 2017

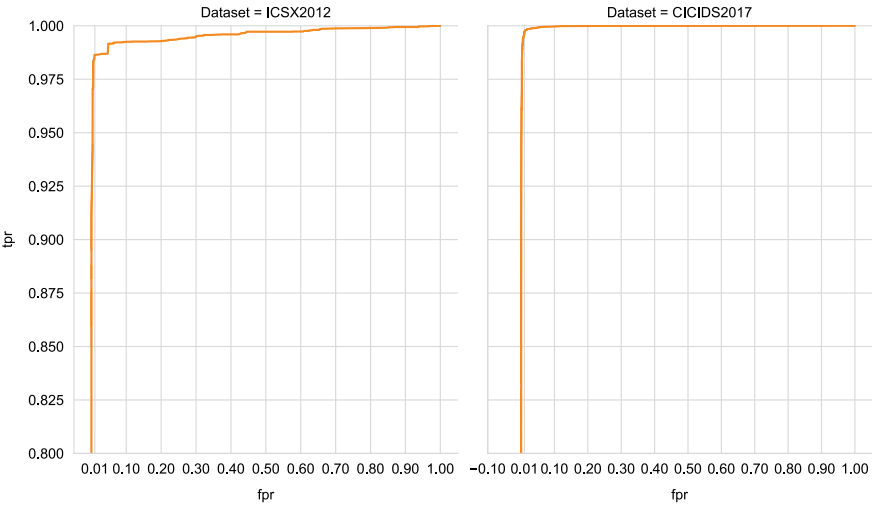


Fig. 7. ROC curve of ISCX 2012 And CICIDS 2017

area under curve is larger than the ROC curve of ISCX 2012. We consider that classifier of CICIDS 2017 is better than this of ICSX 2012.

In Fig. 8, we present the accuracy of each type of attacks in CICIDS 2017. For some attacks, such as DDoS and PortScan, our proposed framework performs well and gains high accuracy, that is over 90%. However, for some other attacks, such as Bot and XSS, our method can hardly detect specific types of attack due to the imbalance of records in the dataset. These types of attack are rare in the real world and occupy small parts of the dataset so could not be identified well.

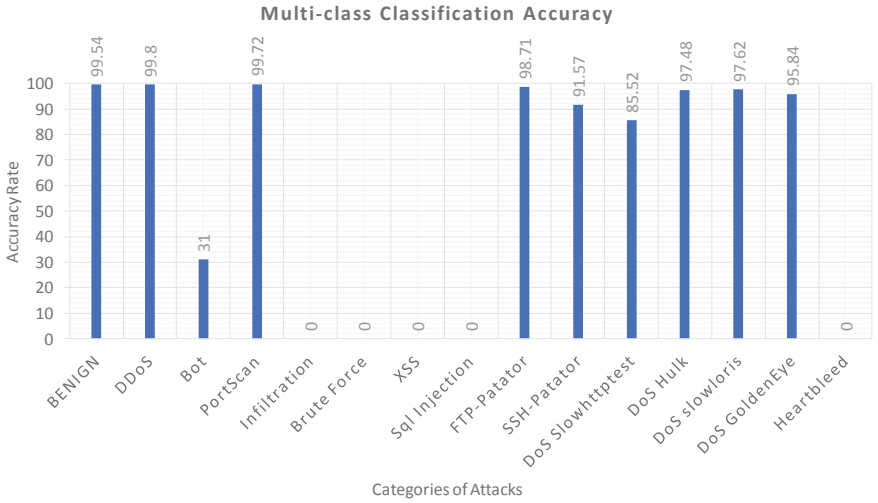


Fig. 8. Accuracy of different types of attack

5 Conclusion and Future Work

In this paper, we proposed a highly flexible framework with on DNN for handling and analyzing novel attacks and large-scale data in networks. In the framework, parameters can be adjusted according to the specific network conditions, so that we can obtain the best model. We apply our framework on ISCX 2012 and CICIDS 2017 and evaluate model performance. For binary classification, our method shows a powerful ability to identify abnormal behaviors. In multi-class classification, We observe high accuracy for common attacks. However, our framework could not identify rare attacks in datasets. Overall, the proposed method improves performance for detecting attacks. Our future work is processing the imbalanced data in datasets to gain better performance for multi-class classification.

References

1. Larson, D.: Distributed denial of service attacks-holding back the flood. Netw. Secur. **2016**(3), 5–7 (2016)
2. Staudemeyer, R.C.: Applying long short-term memory recurrent neural networks to intrusion detection. South Afr. Comput. J. **56**(1), 136–154 (2015)
3. Mishra, P., Varadharajan, V., Tupakula, U., Pilli, E.S.: A detailed investigation and analysis of using machine learning techniques for intrusion detection. IEEE Commun. Surveys Tuts. https://doi.org/10.1109/comst.2018.2847722 (to be published)
4. Potluri, Sasanka: Diedrich, Christian: Accelerated Deep Neural Networks for Enhanced Intrusion Detection System. IEEE Press, New York (2016)

5. Rahul Vigneswaran, K., et al.: Evaluating Shallow and Deep Neural Networks for Network Intrusion Detection Systems in Cyber Security. IEEE Press, India (2018)
6. Zhu, M., Ye, K., Xu, C.-Z.: Network anomaly detection and identification based on deep learning methods, pp. 219–234 (2018). https://doi.org/10.1007/978-3-319-94295-7_15
7. Naseer, Sheraz, et al.: Enhanced network anomaly detection based on deep neural networks. IEEE Access **6**, 48231–48246 (2018)
8. Vinayakumar, R.: Deep learning approach for intelligent intrusion detection system, vol. 2019, pp. 41525–41550. https://doi.org/10.1109/ACCESS.2019.2895334
9. Ozgur, A., Erdem, H.: A review of KDD99 dataset usage in intrusion detection and machine learning between 2010 and 2015. PeerJ PrePrints **4**(2016), Art. no. e1954
10. Sabhnani, M., Serpen, G.: Why machine learning algorithms fail in misuse detection on KDD intrusion detection data set. Intell. Data Anal. **8**(4), 403–415 (2004)
11. Tang, M., Alazab, M., Luo, Y., Donlon, M.: Disclosure of cyber security vulnerabilities: time series modelling. Int. J. Electron. Secur. Digit. Forensics **10**(3), 255–275 (2018)
12. Paxson, V.: Bro: A system for detecting network intruders in realtime. Comput. Netw. **31**(23–24), 2435–2463 (1999). https://doi.org/10.1016/S1389-1286(99)00112-7
13. NSL-KDD dataset (2018). http://www.unb.ca/cic/datasets/nsl.html
14. Abadi, B., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al.: Tensorflow: a system for large-scale machine learning. OSDI **16**, 265–283 (2016)
15. Koch, R., Golling, M.G., Rodosek, G.D.: Towards comparability of intrusion detection systems: new data sets. In: Proceedings of the TERENA Networking Conference, p. 7 (2017)
16. Sato, M., Yamaki, H., Takakura, H.: Unknown attacks detection using feature extraction from anomaly-based ids alerts. In: 2012 IEEE/IPSJ 12th International Symposium on Applications and the Internet (SAINT), pp. 273–277 (2012)
17. McHugh, J.: Testing intrusion detection systems: a critique of the 1998 and 1999 Darpa intrusion detection system evaluations as performed by Lincoln laboratory. ACM Trans. Inf. Syst. Secur. **3**(4), 262–294 (2000)

Program Execution Comprehension Modelling for Algorithmic Languages Learning Using Ontology-Based Techniques

Mikhail Denisov[1], Anton Anikin[1,2(✉)], Oleg Sychev[1], and Alexander Katyshev[1]

[1] Volgograd State Technical University, Volgograd, Russia
[2] Software Engineering School, Volgograd, Russia
Anton@Anikin.name

Abstract. In this paper, we propose an ontology-based approach to model a program execution comprehension so to be able to explain to the novice programmer the essence of his/her error. We have studied the algorithmic languages model operating with actions and basic control structures ("sequence," "branching," and "looping") and designed the rules to capture any deviation from the permissible.

Keywords: e-learning · Ontology · Comprehension · Explanation · Algorithmic language

1 Introduction

Understanding is the most important part of the learning process, in which the learner comprehends and summarizes external facts, acquiring knowledge and skills. Usually, this requires a demonstration of someone else's understanding (e.g., the teacher), through which the student forms his understanding. In the absence of someone to explain, the assimilation is difficult and heavily depends on the intellectual abilities of the student. In this case, one is to look for patterns in plenty of facts to gain understanding through insight.

The facts summarization step requires intellectual work, but once completed, one got an understanding skill, which then works for him/her. We think of that as a developing of an inner cognitive model. With time, the model goes to static and in case of incorrectness can cause prejudice, ex. there was something wrong accepted at the early stages.

Since getting a comprehension, one can apply it to solving various tasks covered by the corresponding knowledge domain. Even those tasks that have not encountered before can be dealt with or at least explained unless exceeding the acquired range of understanding. So, the complete comprehension can be

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_25

defined as an ability to explain anything and answer any question within the covered domain, i.e., if somebody or something endowed with the ability to explain everything about X, it can be thought that it understands X.

The definition above implies that such a "smart" entity is not required to be a human. In the sense of this definition, it is theoretically possible to build an artificial system that understands some subject domain. The system should employ a model imitating intellectual activity. Such systems are commonly classified as decision support expert systems or educating expert systems. An important property of them is the above-mentioned ability to explain to the user what they output. As of our knowledge, there are no known systems applying ontologies to explanations generation in programming language learning.

Undoubtedly, one cannot create an explainer for history, philosophy, or other domains in which it is impossible to have an unambiguous opinion about all. Even the strict sciences are not devoid of ambiguity. Also, a model of a huge domain that should be fine detailed too does supposedly require giant resources to store and process. These practical considerations outline the limitations which are to keep in mind.

The authors' current contribution includes an outline of an ontology schema that represents the algorithm and trace, developed from an elaborate set of rules and constraints on the procedural program model. The few experiments show the realizability of constraints as ontology rules. The implementation of the test prototype has not yet been carried out.

The paper is organized as follows. The next "Problem Formulation" section states the origins of the problem, followed by a brief overview of related works on two domains related to ontologies application: explanation generation and source code processing. "The Methodology Basis" section describes the algorithm and trace models, focusing on the structure and semantics of in-ontology representation. "The Implementation Issues" section notes some details of the implementation, in particular, the standards and tools used.

2 Problem Formulation

To test our ideas and assumptions, we decided to design a complete workflow for the small and easily formalized task of explaining in teaching. The task taken from our University practice is to help the students understand how the program in the C language works.

The process of programming learning starts with understanding the basic concept of programming, which is a stumbling block for many junior students. The difficulty is due rather large abstractness of programming concepts and the dependencies between them, which are not clear to an inexperienced learner dealing with the syntax of a programming language.

Basic programming concepts contain more than one model to understand. For example: the operation priorities in an expression; variables usage; control structures ("sequence,", "branching," "looping"); subprograms (functions); the variables visibility scope; memory usage; call stack. All of the above are the

abstract models clear to an experienced programmer. Many of them can be studied separately from others. A subset that we have chosen is the dynamic properties of the control structures—the mapping of an algorithm to its linear sequential execution (we call it "trace"). See an example in Fig. 1.

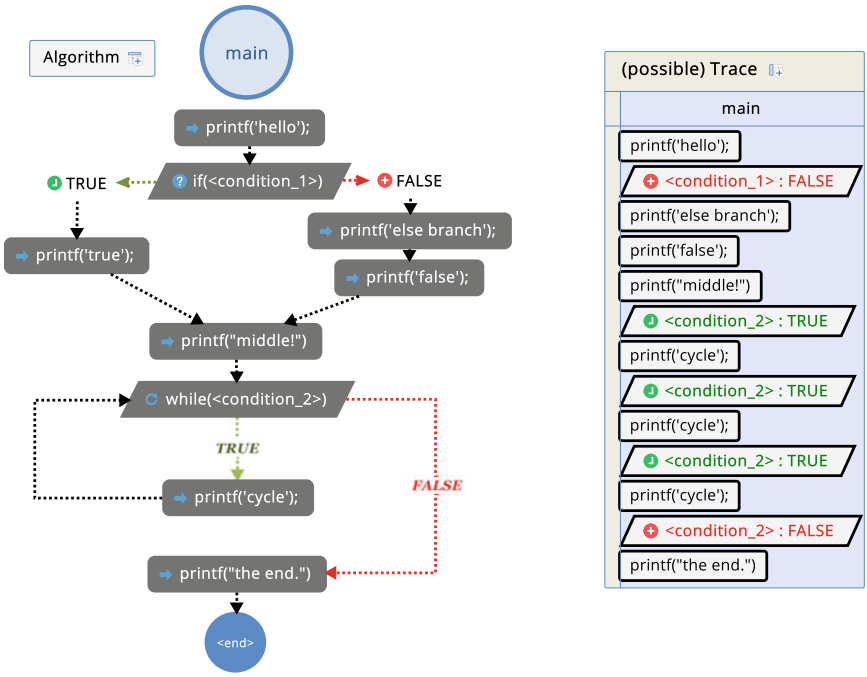


Fig. 1. Algorithm and trace example

A possible goal is to teach the student to understand what actions a program should perform when processing a looping, and conversely, how to make a looping, so that it produces the desired sequence of commands and to avoid an infinite loop execution. One may need some illustrative examples. A good option is to let students play with them by answering some questions to understand the rules by which it works. With any mistake, the need is to highlight to a violation of a particular rule. For example, a looping continues with next iteration if and only if the loop condition evaluates to true, otherwise a mistake is reported. This task is more for a simulation program than for a human expert.

The teacher could find useful automatic tasks generation by templates or properties. It provides a large variety of task forms and prevents students from cheating on earlier issued tasks. See the system structure brief at Fig. 2.

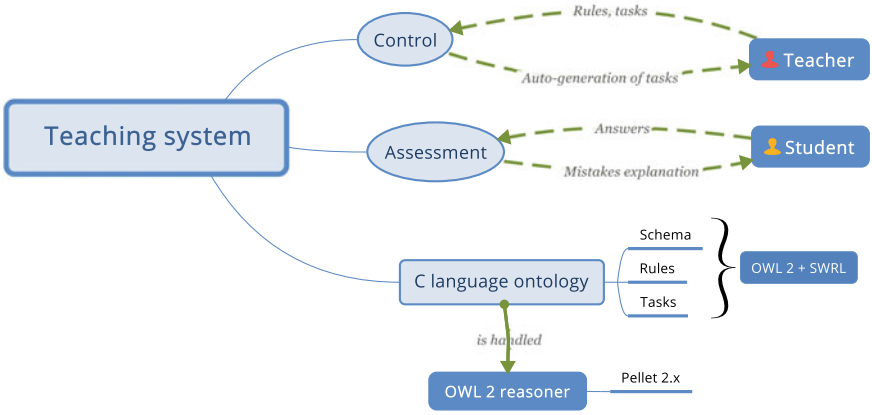


Fig. 2. The system structure brief

To solve the problem of automatically checking answers to programming tasks with a human-understandable explanation, we design software that consists of the formal model of programming, the explanation mechanisms, the task, and the answer, and in other words, prior domain knowledge, factual knowledge, and inference rules. Creating a knowledge-based system is inefficient if using traditional procedural programming tools. Hence, different approaches were invented (which, however, were not previously used for programming teaching), like the semantic Web technology, based on ontologies, which store knowledge as a graph.

The model of the selected subset of programming principles, expressed by the domain schema and a set of constraints, is represented in the form of an OWL ontology. Then with facts of a particular task and answer, the system can detect errors and generate explanations by use of a reasoner.

Our ontology scheme consists of classes and properties (links) necessary to represent the two hypostases of a procedural program—the algorithm as a composition of control structures and the trace as a linear sequence of executable atomic instructions.

3 Related Work

Ontology is usually defined as the form and methodology dedicated to a subject area description. The reasoner program can automatically obtain useful inferences. Most researchers stating the use of ontology in education, specifically e-learning, aim to describe the whole of particular subject domains [12] or do work with such models like in [2]. Some do not try to cover the entire world's diversity but concentrate on smaller and decidable tasks (check [5]). We think that keeping the methods within the framework of mathematics supplies hope for practical feasibility.

Applications of ontology go far beyond educational tasks. It is popular to apply ontology to AI, NLP, BI, enterprise modelling, etc. For example, [6] contribute to self-educated systems.

We have selected several articles that are highly relevant to our topic and are worth mentioning in the review.

3.1 Ontology in Comprehension and Question Generation

A system [9] based on the TREPAN algorithm is to explain artificial neural networks to end-users using decision trees. The ontology models as a domain knowledge were incorporated into the system to improve the quality of generated explanations. Applied to finance and medicine domains, the understandability of decision trees was measured by a user study.

Paper [3] discusses the e-learning task on the comprehension level according to the taxonomy of educational objectives by Bloom (1956). It analyzes approaches to support student's comprehension using means of e-learning driven by subject-domain ontologies. In particular, the explanations generation and the adaptiveness of personal education paths are considered.

Paper [1] has dedicated to the multiple choice question (MCQ) generation based on ontology. An experimental comparison to an ontology-based MCQ item generation system known as OntoQue was performed. As an improvement that incorporates learning content, learning objectives, lexical knowledge, and scenarios into a single cohesive framework, a novel, holistic view of the problem was suggested.

3.2 Ontology in Code Analysis

Modern compilers (the programs that convert source code into executable machine code) allow for program optimization, parallelization, error detection, and performance degradation, thanks to analyzer incorporated. The authors of [13] offer a "declarative" approach for compiler analyzer development using logical rules (namely written in SWI-Prolog). This significantly reduces the amount of code (by times) and noticeably improves the extensibility and supportability comparing to the alternative "imperative" approach. The increased level of abstraction allows other potential developers to understand and reuse the analyzer.

The implementation of the approach involves converting the C code into an ontology form, then imports each fact into the Prolog fact store. This feature was very promising for us as we deal with C language too until we fail to run their 16-bit utility. It also seems difficult for us to rebuild the tool from sources.

The authors of paper [10] (2011) discuss the problem of searching and reusing of clean and debugged code. Applying meaningful search patterns to ontology-represented code, they probably were first to go beyond the programming language syntax. The proposed solution extracts the metadata about the structure of all completed code, saves it into ontology, and also moves the linked code to a

distributed repository. This simplifies search and reuse of code snippets. In our opinion, designing of proper ontology schema is the most challenging task here.

The related task for the Java language is considered in [4]. The ontology scheme, the Java-to-RDF parser, and huge dataset collections are freely available at the "CodeOntology" project Web site [8].

4 The Methodology Basis

A model for algorithmic programming languages provides formal descriptions of the algorithm and the trace. The description of the algorithm (static part) is gathered from the syntactic language constructions. This task has many implementations, as listed above. Our C language schema is based on an open schema from [8] as the Java has a lot in common with C.

Trace description elements (dynamic part) are linked to static elements (the algorithm). The trace scheme takes into account the possible repetitions and unfulfilled branches of the algorithm. It also allows introducing the information missing in the algorithm, but necessary for the trace construction: it is the results of control conditions calculation. These conditions affect the trace structure and distinguish different traces.

Technically, both schemes are two OWL 2 [7] class hierarchies. Relations (OWL properties) structure between these classes reflects the dominance of the algorithm, so there are no dependencies directed from the algorithm to the trace. Thus, there may be several different execution traces for a single algorithm, even within the same ontology.

4.1 Algorithm Model

The algorithm scheme we compiled so far can express the composition of algorithmic structures and atomic statements. A sequence (block statement), a branching (if-else statement), and a looping (for and while loops) are supported. An empty block is to be substituted with a special "no-op" statement. The conditions controlling the decisions in "if-else" and looping statements can be denoted as `condition_1`, `condition_2` and like that. The atomic statements can be denoted as `stmt_1`, `stmt_2` and like that.

No further detalization is provided yet. This means that the model does not consider what each action does. It considers only the fact of an action's appearance after/before another act. If such a truncated model could work, then we extend the model with more practically useful programming principles.

The scheme of the algorithm in Fig. 3 is adapted from a similar scheme for Java and contains classes not used yet. In particular, `Type`, `Variable`, and `Function` classes are left for the future moment when new functionality is added.

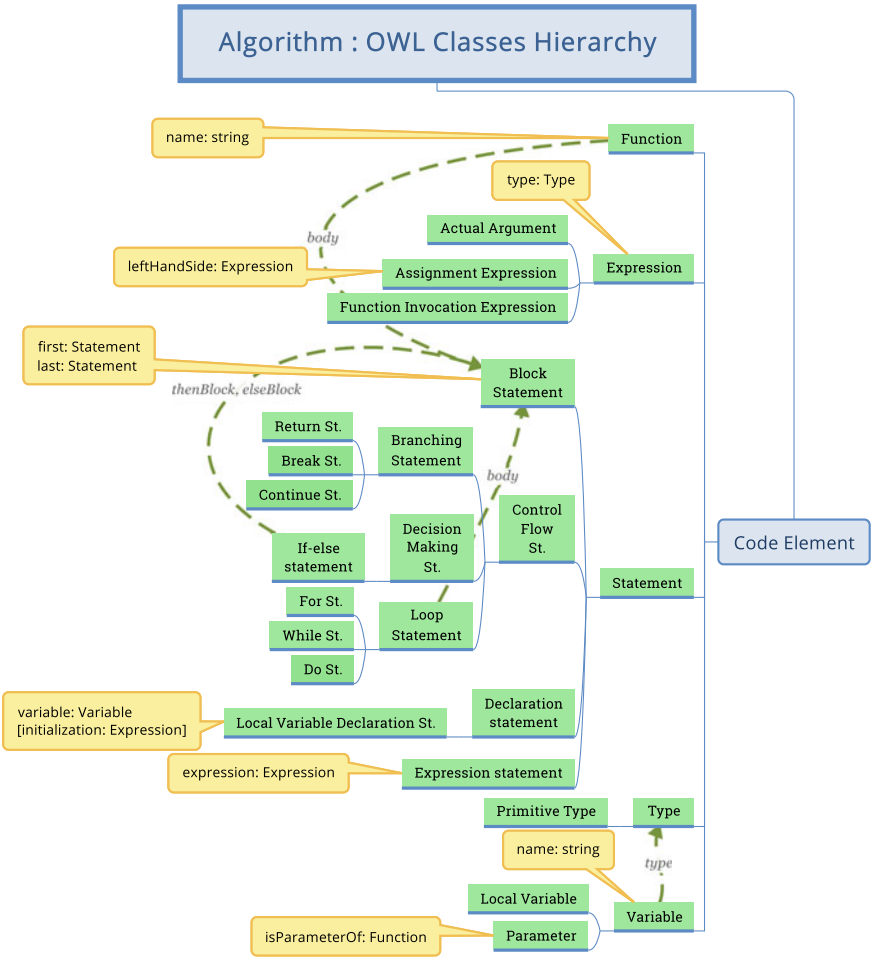


Fig. 3. Algorithm scheme: OWL classes hierarchy

The hierarchy (drawn from right to left) of classes represents the instances of `rdfs:subClassOf` relation between classes, and dashed arrows represent some of the domain-specific relations between class instances. For example, `Function` and `Loop statement` instances should have a `'body'` property connected to an instance of `Block statement`.

The example source code is shown in listing at Fig. 4. Note that `'printf();'` function calls are used as statements just for clarity. Due to model limitations, we could use atomic names like `'print_statement_1'` too. We use the less detailed notation for conditions (like `'cond_1'`) as the model cannot calculate any run-time values, thus receiving these values from an external source, ex. from the teacher compiling a task.

```
void main()
{
        printf("hello");
        if (<cond_1>)
        {
                printf("true");
        }
        else
        {
                printf("else block");
                printf("false");
        }
        printf("middle!");
        while (<cond_2>)
        {
                printf("cycle");
        }
        printf("the end.");
}
```

Fig. 4. Algorithm source code in C (with placeholders as conditions)

Figure 5 shows an in-ontology structure of the algorithm example which is drawn at left on Fig. 1 and is explained below.

- The hierarchy (left to right) of outlined nodes represents the nesting relation.
- The filled rectangle nodes are the atomic statements which always occur explicitly in the algorithm and could form an exact trace.
- The roundrect nodes are the composite control structures whose are usually represented as programming language syntax constructs.
- The underlined nodes are designed as `‘‘.property:’’` to remind a predicate from the known RDF triple structure “*Subject - Predicate - Object*” (the leading dot can be thought here as member accessor in OOP). A node at left is the owner (*subject*), a node at right is a value of property (*object*), and the underlined node itself names the property (*predicate*).
- The dashed arrows represent additional specific relations between statements within a sequence (shown for one block only). A text by the arrow corresponds to the property name in a triple which connects potential instances (of classes connected by the arrow). That is, `‘first’` and `‘last’` properties connect a block with two boundary statements. The order of sequence is provided by the `‘next’` property is linking statements of the same nesting.

Currently, examples of algorithms are crafted manually, which is very tedious even for very small C programs. In the future, we plan to develop a utility to convert C code into compatible ontology.

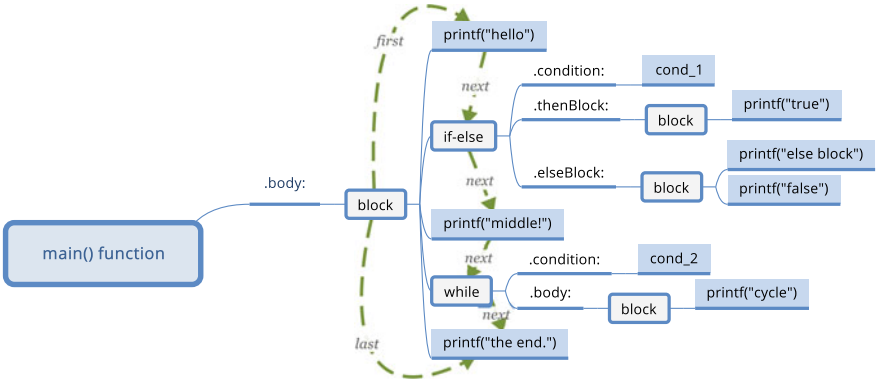


Fig. 5. Algorithm example: in-ontology structure

4.2 Trace Model

The trace scheme design is based on the concept of execution context (class 'context'). Each statement (both atomic and complex) is executed within a control structure context (all but the `'main()'` function). For example, the act `'printf(''else branch'');#1'` (originated by `'printf(''else branch'');'` statement) has path of contexts which can be written like

```
main.body > if_stmt#1.thenBlock > printf(''else branch'');#1
```

where `#n` is a counter of execution.

The trace scheme (see Fig. 6) is shown as the hierarchy of the trace elements representing the behavior of code elements. So each trace element must be connected to specific code element via `'origin'` property. Of course, multiple acts can be connected to the same code element.

The classes hierarchy (top to bottom) is translated into `rdfs:subClassOf` relations again.

The `Act` and `ConditionAct` classes represent atomic statements and control conditions, respectively. `ConditionAct` instance must have `'evalsTo'` property assigned to either `TRUE` or `FALSE`.

The `Context` class and its descendants represent the execution contexts provided by complex code structures. For example, `Decision` (a base of `If_else` context) instance has a `Block` instance as sub-context (pointed by arrow).

The `Iteration` context class is a notable addition to the trace scheme as it does not correspond to any algorithm element. The class is introduced to model each repetition of looping and has `'index'` property to store the iteration sequence number.

Again, to describe the order in a sequence the `'next'` property is used. Besides, the `'before'` property as the transitive closure of over `'next'` is introduced. We use it to search for misplaced acts. The `'first'` and `'last'` properties link a `Block` context to a sequence of acts directly nested into it.

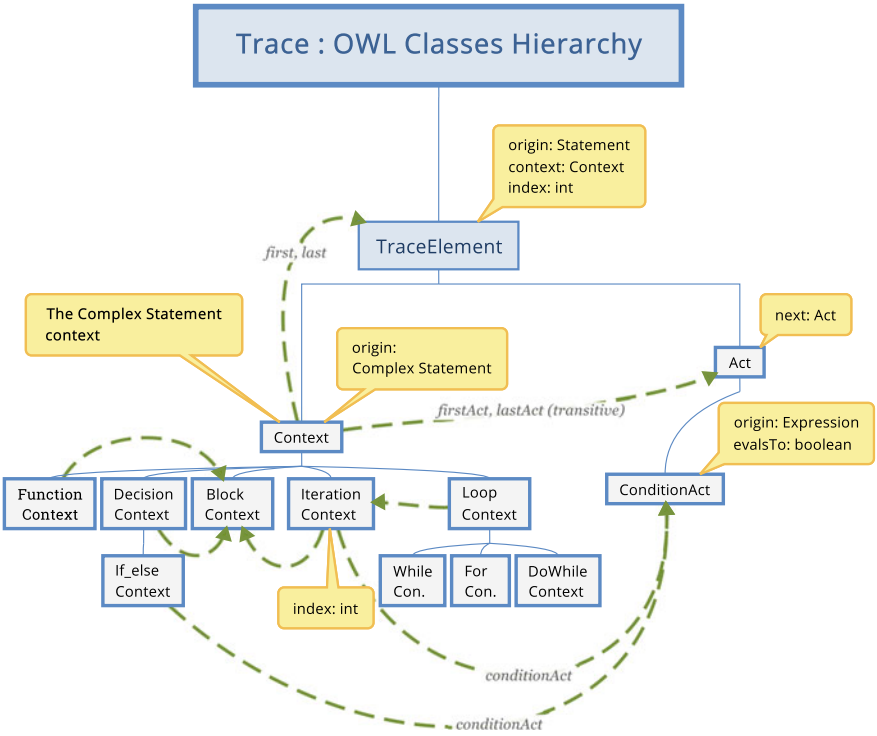


Fig. 6. Trace scheme: OWL classes hierarchy

Alongside with `'first'` and `'last'` properties, we had to introduced `'firstAct'` and `'lastAct'` properties. These properties point to "terminal" acts (`Act` or `Condition Act` instances) only and are expanded up along the hierarchy of nested contexts. That means if, for example, a function starts with a do-while loop which starts with `IF` statement containing first act `F`, then all of the complex statements listed above start with act `F`. The `'lastAct'` property applies similarly.

So, `'firstAct'` and `'lastAct'` properties are transitive. The trick is required to check the correctness of any sequential pair of acts. Consider an example shown in Figs. 1, 5, 7, and 4 (the listing). The statement `'printf(''false'');'` must be followed by `'printf(''middle!'');'` because the last statement of `ELSE` block ends the `IF_ELSE` statement and requires the next statement to be invoked. We do understand that the next statement to execute is `'printf(''middle!'');'` as it is located right after the `IF_ELSE` statement. In general, the nesting depth is unknown, so the check requires observing two paths of arbitrary length.

The SWRL language (used for inference rules declaration) supports direct matching of graph patterns of fixed size only. So, the matching of arbitrary

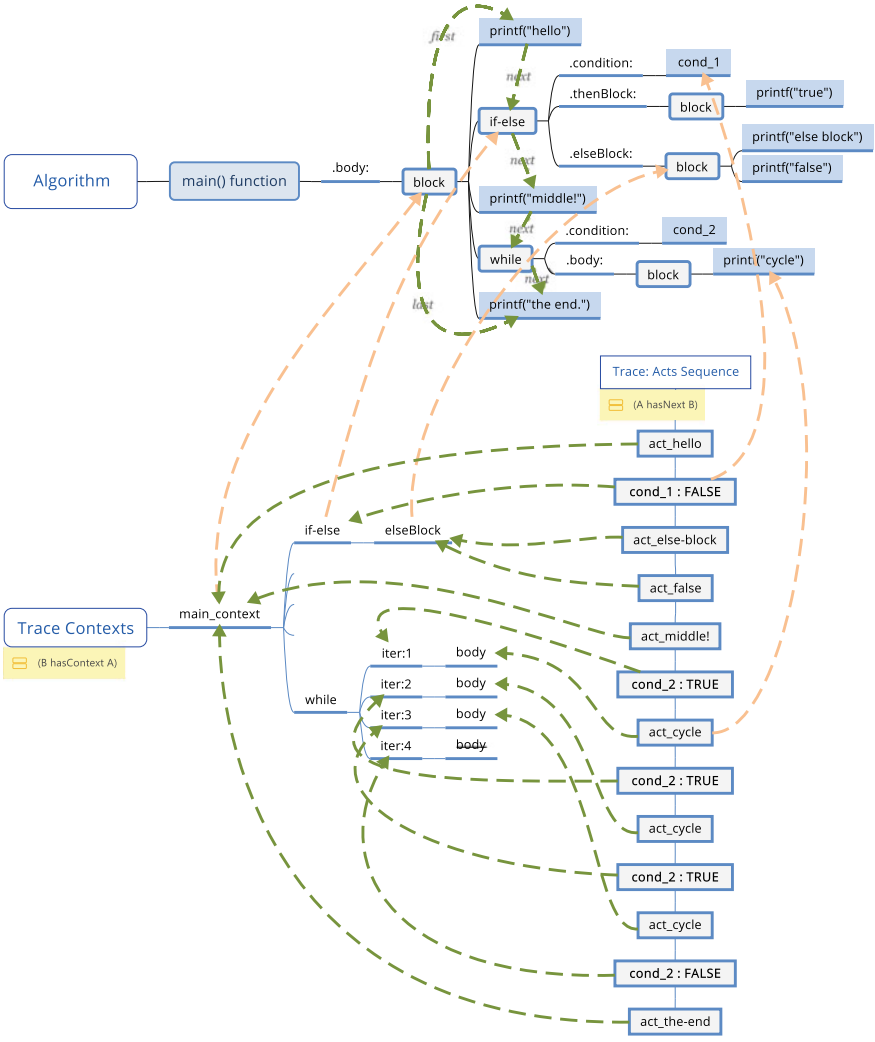


Fig. 7. Trace example (bottom part) and its algorithm (above)

paths requires reducing them into constructs of fixed size. This is what is done by use of ‘`firstAct`’ and ‘`lastAct`’ transitive properties.

Figure 7 shows an in-ontology structure of the trace (a vertical top-to-bottom sequence) connected to its contexts tree (at left) and algorithm (above) which is the same as drawn on Fig. 5. The explanations are below.

- All the acts in trace sequence are connected one-to-another with ‘`next`’ property. Each act and context in the tree connects to its left neighbor via ‘`hasContext`’ property (both the arrows from trace-to-contexts’ tree and

the child-parent relations in the tree). So any trace element is within the root context named as `'main_context'` (the name is chosen arbitrarily).
- The algorithm was discussed above at the description of Fig. 5.
- All the arrows connected to the algorithm express the `'origin'` link. Just a few of them are shown to maintain readability, but note that each act and each context must have exactly one origin element in an algorithm tree.

Figure 7 illustrates some features of developed model. The most interesting is about the structure of acts' contexts.

The context of `IF_ELSE` statement (`'if-else'` underlined node) is distinguished from the context of `ELSE` block. The `'While'` context and `Iteration` and `body` are distinguished too.

The looping is thought as the most complicated abstraction. `Loop`'s trace divides into iterations each of whose starts with condition checking (the exception is the start of `DO-WHILE` loop) and can be modified by `'continue'` or `'break'` occurrence. So, the whole looping context consists of iterations (plus a single initialization act in case of `FOR` loop). The iteration context can have multiple instances within a loop context, thus enumerated (by setting `'index'` property). The iteration context consists of `'conditionAct'` instance and the body block if the condition has been evaluated to `TRUE`. No `body` is attached in case of `FALSE` (see the `''body''` crossed out for `''iter:4''` of `''while''` node).

5 The Implementation Issues

The currently popular so-called semantic Web is standardized by the World Wide Web Consortium (W3C). The key concept of the semantic Web is ontology. Ontology is also often referred to as a knowledge base. Similar to a database, an ontology stores the data as linked structure (specifically, a directed graph or a set of *triples*). The basic schema of ontology ships with predefined semantic meaning, so the data treated as knowledge becomes more intelligent.

To "revive" the knowledge in an ontology, the *facts* and *axioms* must be represented in a strict formal form. Common is the resource description framework (RDF), RDF schema (RDFS), the OWL (Web ontology language), and the use of a logical solver (reasoner) to obtain new logically inferred facts.

In the implementation, we use OWL 2 language [7] to represent knowledge and facts and relationships, SWRL language to declare rules, as it is more expressive than OWL only. Working tools include Protege 5 ontology editor, Pellet 2 (ontological reasoner), and Owlready library [11].

As using OWL 2 and SWRL rules, we need suitable reasoner supporting SWRL. The Java API is also required to work with Protege 5. Choosing between Hermit 1.3.8 (last released in 2013) which declares full OWL 2 compatibility, and Pellet 2.3.1 (last released in 2017), we found out that Hermit cannot deal with relatively simple SWRL rules, while Pellet can. So, we use Pellet 2.3.1. Probably, we will move to Stardog 7 + Pellet 3.x in future.

6 Conclusion

In this paper, we discussed an approach and implementation details of an educational system such to be able to explain the essence of mistake in algorithmic programming languages learning using the ontology-based techniques. Our study considers rather simple abstractions that are limited to be formalizable, unambiguous, and suitable for automatic reasoning. The model of procedural programming satisfies the requirements and, fortunately, can be completely covered with a not very large set of rules.

The first developed version can check the correctness of trace by an algorithm given. We plan to describe other models of procedural programming, in particular, the calculation of expressions with operation priorities, the use of variables, dynamic state of the program, etc. We will also develop an automatic generation of tasks for types of errors (with insignificant random variability).

Acknowledgements. This paper presents the results of research carried out under the RFBR grants 18-07-00032 "Intelligent support of decision making of knowledge management for learning and scientific research based on the collaborative creation and reuse of the domain information space and ontology knowledge representation model" and 20-07-00764 "Conceptual modeling of the knowledge domain on the comprehension level for intelligent decision-making systems in the learning."

References

1. Al-Yahya, M.: Ontology-based multiple choice question generation. Sci. World J. **2014** (2014)
2. Anikin, A., Litovkin, D., Kultsova, M., Sarkisova, E., Petrova, T.: Ontology visualization: approaches and software tools for visual representation of large ontologies in learning. Commun. Comput. Inf. Sci. **754**, 133–149 (2017)
3. Anikin, A., Sychev, O.: Ontology-based modelling for learning on bloom's taxonomy comprehension level. In: Samsonovich, A.V. (ed.) Biologically Inspired Cognitive Architectures 2019, pp. 22–27. Springer International Publishing, Cham (2020)
4. Atzeni, M., Atzori, M.: CodeOntology: RDF-ization of source code. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (2017)
5. Atzeni, M., Atzori, M.: What is the cube root of 27? Question answering over codeontology. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (2018)
6. Babli, M., Onaindia, E.: A context-aware knowledge acquisition for planning applications using ontologies (04 2019)
7. Bechhofer, S.: OWL: Web ontology language. In: Encyclopedia of Database Systems, pp. 1–2 (2016)
8. Codeontology project website. http://codeontology.org. Accessed 13 Nov 2019
9. Confalonieri, R., Weyde, T., Besold, T.R., del Prado Martín, F.M.: Trepan reloaded: a knowledge-driven approach to explaining artificial neural networks (2019)
10. Gopinath, G., Sagayaraj, S.: To generate the ontology from java source code. Int. J. Adv. Comput. Sci. Appl. 2 (03 2011)

11. Lamy, J.B.: Owlready: ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. Artif. Intell. Med. **80** (2017)
12. Liu, K., Zhao, J., He, S., Zhang, Y.: Question answering over knowledge bases. IEEE Intell. Syst. **30**(5), 26–35 (2015)
13. Zhao, Y., Chen, G., Liao, C., Shen, X.: Towards ontology-based program analysis. In: Krishnamurthi, S., Lerner, B.S. (eds.) 30th European Conference on Object-Oriented Programming (ECOOP 2016). Leibniz International Proceedings in Informatics (LIPIcs), vol. 56, pp. 26:1–26:25. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany (2016) http://drops.dagstuhl.de/opus/volltexte/2016/6120

Multipathing Communication in Heterogeneous Ad Hoc Network

Chaimae Benjbara(✉), Nada Mouchfiq, and Ahmed Habbani

National School of Computer Science and Systems Analysis "ENSIAS", Mohammed V University - SOUISSI, Rabat, Morocco
c.benjbara@um5s.net.ma
{nada_mouchfiq,ahmed.habbani}@um5.ac.ma

Abstract. Multipath routing is a solution that meets the requirements of existing systems (security, bandwidth, network lifetime, etc.). Currently, several studies are trying to apply and adapt the principle of multipath in several types of networks for benefiting from its advantages. The wireless network without infrastructure (ad hoc network) is more flexible and less expensive compared to other networks and uses in its turn the multipath routing in the communication between the devices to enhance its strengths. The ad hoc network regroups several categories of network, and each one has its own node characteristics and communication protocol. In most application areas based on new technologies, such as smart cities, precision agriculture, medicine, and others, we can find different types of ad hoc networks in the same architecture constituting a heterogeneous network. So, the heterogeneous ad hoc network (HAdN) appears and brings with it a new line of research concerning the common and compatible routing protocols inside this type of network. This article proposes a new multipath routing protocol version based on the OLSR standard ensuring the direct communication between the various HAdN components. The new protocol will initially be implemented in a specific network architecture consisting of MANET, VANET, and FANET.

Keywords: Heterogeneous · Ad hoc · Multipath · Routing · Communication · Identification · Classification

1 Introduction

The wireless ad hoc network is a spontaneous and decentralized network, where the nodes switch between the router and the host roles according to their position in the communication. It groups various type of ad hoc network characterized by a unique communication concept between its specific components, such as

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_26

wireless mesh network "WSN," wireless sensor network "WSN," mobile ad hoc network "MANET," vehicular ad hoc network "VANET," flying ad hoc network "FANET," and body ad hoc network "BANET." The presence of at least two types of these networks creates a new network form, named heterogeneous ad hoc network, which has become a challenge for researchers in improving its performances. The heterogeneity according to the studies in progress divided into two visions: the communication channel and the nature of the nodes, which makes the diversified study of this network in several aspects: routing protocols, self-organization, energy-saving, security mechanism, etc.

Routing protocols present the dynamic of the network, used by each node to discover the topology of its environment and control the exchange of the packets. There is a wide choice unicast and multicast routing protocol in ad hoc network belonging to different families [1]:

- Proactive: keeps fresh the lists of destinations and their routes by periodically distributing routing tables around the network,
- Reactive: finds a route on-demand by swamping the network with route request packets,
- Hybrid: combines the advantages of proactive and of reactive routing,
- Back-pressure routing: chooses next hops dynamically as a packet is in progress toward its destination. These decisions are based on congestion gradients of neighbor nodes.

The multipath routing is the routing protocol tendency used by many researchers. It improves the communication quality by affecting diverse aspects: security, bandwidth, load balancing, and transmission delay.

In this article, we choose a specific heterogeneous ad hoc network composed of three different environments MANET, VANET, and FANET. The communication inside this network ensured by a new multipath protocol described in the following sections. Section 2 treats the current method of communication within a heterogeneous network. Section 3 presents the general idea of the patented multipath routing protocol. Then, part Sect. 4 illustrates the application of two steps in our multipath routing inside the chosen heterogeneous network. Finally, in the last section, we conclude our work.

2 Related Work

In [2], the authors chose to study a heterogeneous ad hoc network composed by MANET, VANET, and Mesh. The communication inside this heterogeneous network is impossible because there is a lack of interoperability between the different routing protocols used in each network. So, t proposed an ad hoc traversal routing "ATR" like solution resolving this obstacle of communication between nodes. The ATR is a mechanism implemented in some nodes in the network, named ATR nodes. These nodes represent gateways that convert control messages from one network to another according to the type of routing protocol used:

- Reactive protocol (AODV): The ATR node is a transmitter or receiver within its network,and it exchanges the route request messages (RR) with its neighbors. If it is a transmitter gateway, it converts the RR message into an ATR route request message (ATR-RR) sent to its ATR node neighbors. If the ATR node is the receiver gateway, it wears two hats: a router transmitting the ATR-RR messages to the next ATR node or converter transforming the ATR-RR to an RR message and sends it to its nodes neighbors in the same network.
- Proactive protocol (DSDV): The ATR nodes collect the address information of the nodes in the network to which they belong, and each ATR node shares this collected information with its neighboring ATR nodes. Then, the proactive protocol can create the route entry for nodes of different networks in the routing table.

3 Our Contribution

3.1 Introduction

The power of the ad hoc network has embodied in its characteristics: s-configuration, self-organization, flexibility, and ease of using. This type of network partitioned into several categories: wireless mesh network "WMN" [3], wireless sensor network "WSN" [4], mobile ad hoc network "MANET," vehicular ad hoc network "VANET" [5], and flying ad hoc network "FANET" [6]. Each one regroups a specific nature of nodes communicating via a well-defined routing protocol inside a precise architecture. This wealth of ad hoc network allows critical areas such as the military, health, road safety, agriculture, and others to adopt one or more types of this network within their architectures (Table 1).

Table 1. Comparison table of the different ad hoc network [7]

Criteria	Ad hoc network types				
	WMN	*WSN*	*MANET*	*VANET*	*FANET*
Node mobility	Low	Low	Low	Medium	High
Node density	High	High	High	Medium	Low
Topology change	Slow steady	Slow steady	Slow steady	Average speed	Rapid speedy
Energy	Need	Need	Need	Not needed	Needed for small UAVs
Power	Limited	Limited	Limited	Average	Very big
Localization	GPS	GPS	GPS	GPS, AGPS, DGPS	GPS, AGPS, DGPS, IMU
Protocol	AODV	AODV	OLSR	QoS-OLSR	P-OLSR

3.2 Patented Solution

The study made in the previous section shows a big problem of the direct exchange of information between the various type of devices inside a heterogeneous ad hoc network. To fix this problem, we propose a protocol that ensures

the communication with the same jargon within this network, regrouping homogeneous nodes while maintaining the characteristics and properties of each node. The heterogeneous multipath protocol has a precise architecture made of several steps linked to assure guaranteed data transmission (Fig. 1).

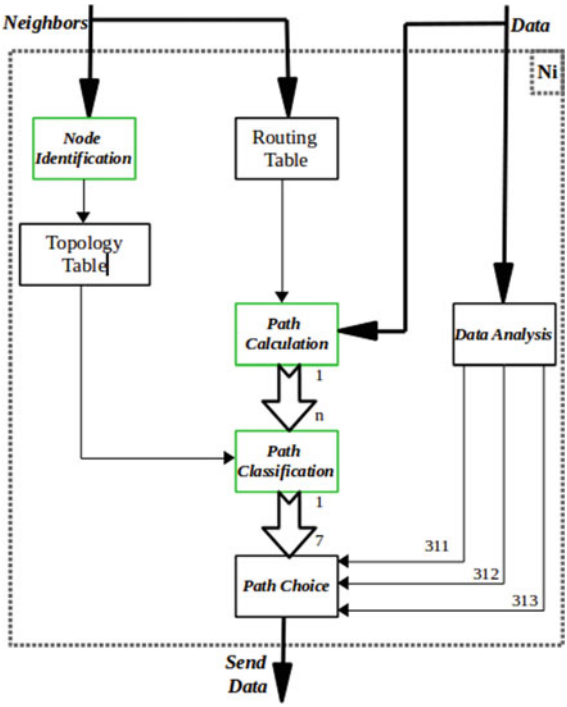


Fig. 1. Node communication architecture [8]

4 Solution Description

According to the comparison table, the MANET, VANET, and FANET networks belong to the same family and use different versions of the same routing protocol "OLSR." So, we can study a heterogeneous network constituted by these three networks (Fig. 2).

The multipath routing improves overall network performance by sharing available network resources (bandwidth, load paths, etc.). The OLSR protocol has several multipath versions to benefit from the advantages of multipathing [9]. In this work, we propose a new version of multipath OLSR named multipath heterogeneous ad hoc routing OLSR "MHAR-OLSR" by adding many parameters. The next sections describe two steps of the new protocol: identification of the nodes and the classification path.

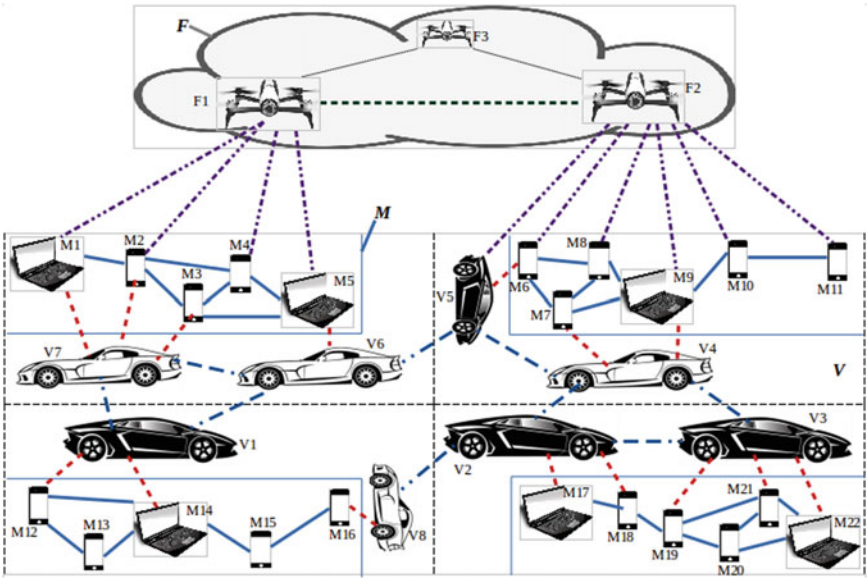


Fig. 2. Heterogeneous ad hoc network MANET VANET FANET "HAdN-MVF" [7]

4.1 Identification Nodes

The purpose of the identification is to specify the nature of the neighbors of each node inside the HAdN-MVF. The identification of the nodes presents the root of the next steps. OLSR standard allows each device to discover the topology of its environment. In a heterogeneous network, each node must have a clear and precise vision of its neighbors for knowing its nature. Range and proportional speed are the key criteria on which each node bases itself to assign their neighbors an identifier "ID-Node" indicating their type of network. The ID takes three values 1, 2, and 3, respectively, representing MANET, VANET, and FANET as shown in the flowchart below (Fig. 3).

The city bus fleet range of the metropolitan area of Seattle Washington [10] reaches 3 km, ensuring service to the passengers into the regular routes of buses throughout the city. The 20 km/h is the best average speed record achieved by Wilson Kiprotich in marathon 2013. So, it can represent the maximum speed in the MANET network.

- **New OLSR messages structure**:

In OLSR standard, the exchange of the Hello messages allows the nodes to fill its neighborhood table with the neighbors of the first and second hops. The identification needs to insert important information in the Hello message, enabling to specify the nature of the neighbor node (Fig. 4): latitude and longitude occupy 4 bytes each as floating-point numbers, also altitude and range represented in a fixed-point number of 16 bits. So, the 12 bytes addition into a standard Hello

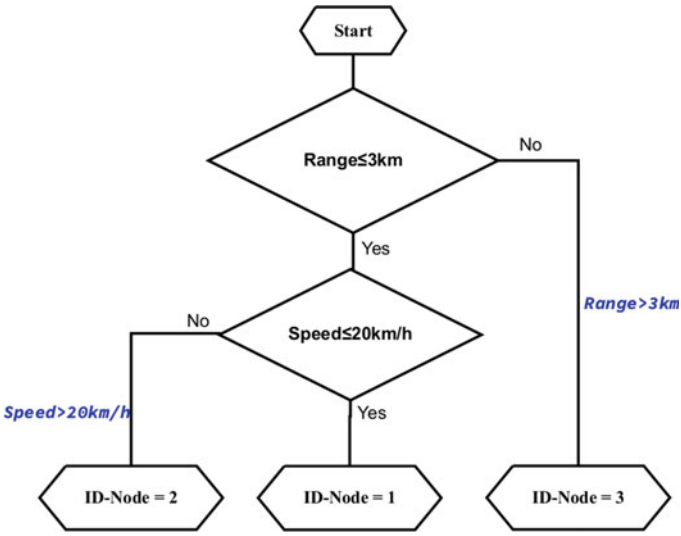


Fig. 3. Identification organizational chart [7,8]

Byte 0 0 1 2 3 4 5 6 7	Byte 1 0 1 2 3 4 5 6 7	Byte 2 0 1 2 3 4 5 6 7	Byte 3 0 1 2 3 4 5 6 7
Reserved		Htime	Willingness
Altitude		Range	
Latitude			
Longitude			
Link Code	Reserved	Link Message Size	
Neighbor Interface Address			
Neighbor Interface Address			
...			

Fig. 4. New Hello message format [7,8]

message will not affect the size of the 802.11 frames in medium and large networks.

The TC message ensures nodes to discover the network topology. In turn, it will benefit from modification by adding the average instantaneous speed formatted on a fixed-point number occupying a block of 16-bit, and dynamo of the new protocol ID-node reserved a fixed-point number of 8 bits (Fig. 5).

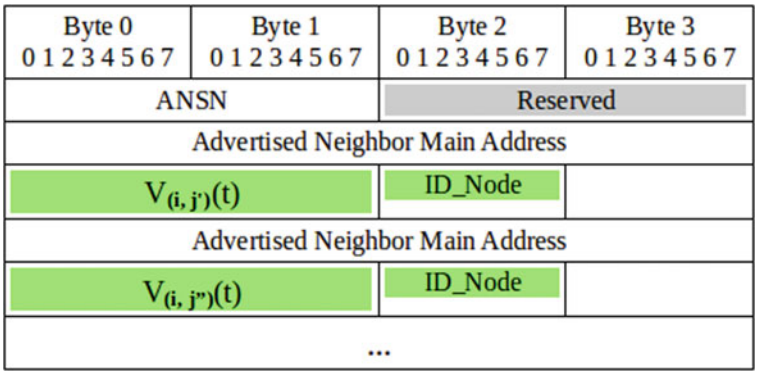


Fig. 5. New TC message format [7,8]

To calculate the average instantaneous speed:

$$\begin{cases} V_{(i,j)}(t) = \tilde{V}_{(i,j)}(t) + V_{(i,j)}(t-1) \\ V_{(i,j)}(0) = 0 \end{cases} \tag{1}$$

We need the instantaneous relative velocity represented as follows:

$$\tilde{V}_{(i,j)}(t) = \frac{d_{(i,j)}(t) - d_{(i,j)}(t-1)}{\delta t} \tag{2}$$

where:

* $d_{(i,j)}(t)$ and $d_{(i,j)}(t-1)$ are the distances between two neighboring nodes at two successive instants.

* δt : is the difference between the arrival time of the first and last message Hello.

For having a more precise speed value, we add an exponential average. Then, the instantaneous speed takes this new form:

$$\begin{cases} V_{(i,j)}(t) = \alpha \tilde{V}_{(i,j)}(t) + (1-\alpha) V_{(i,j)}(t-1), \ 0 \leq \alpha \leq 1 \\ V_{(i,j)}(0) = 0 \end{cases} \tag{3}$$

4.2 Path Classification

The MHAR-OLSR protocol computes the multipaths like the other OLSR versions using a multipath Dijkstra algorithm but basing on the topology table created by the new TC message. The routes in ordinary multipath protocols are classified based on the number of hops. But in MHAR-OLSR, the identifier of the node added into the topology table will be taken in the classification.

The source node "S" calculates the routes that connect it with its destination. After, in the first step, S arranges them from the shortest to the longest based on the number of hops. Then, S classifies these ordered paths according to the nature of their components and assigns each path an identifier "Path-ID" indicating its

importance. It should note that S is the only device responsible for calculating and classifying the paths that connect it with its destination. The following flowchart gives a clearer view of how to classify the detected paths (Fig. 6).

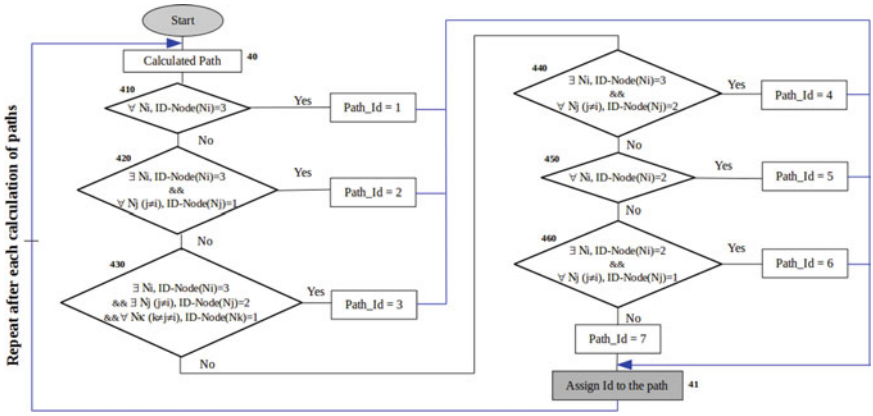


Fig. 6. Classification principle

The path identifier varies between 1 and 7 showing the quality and efficiency of this path as well as its availability for sending data. Identifier 1 represents a reliable and fast path, against the identifier 7 has the longest and least reliable path compared to other routes. We can have several paths with the same identifier, and in this case, we increase the availability of routes to the destination. To evaluate the functioning and efficiency of our protocol and know its limits, it is currently implemented in NS3. Our work aims to improve the daily life of the students at our university at a low cost. For that, we will propose to them a free protocol that will ensure access to important information (frame schedule, availability of means of transport (Taxis, Bus)), knowing that no company or application currently provides this service.

5 Conclusion

In this article, we gave an overview of the different ad hoc networks, as revealed the problem of communication between them. In the rest of the article, we proposed a heterogeneous ad hoc network grouping the MANET, VANET, and FANET networks, called HAdN-MVF. A new protocol "MHAR-OLSR" based on the OLSR standard has designed for this type of network, whose purpose is to ensure communication between the heterogeneous components of the HAdN-MVF network while respecting their nature.

MHAR-OLSR is a multipath protocol specifying the nature of the nodes that make up the HAdN-MVF network by assigning them an identifier that will use

in the classification of calculated paths connecting the source to the destination. Each path will take an indicator that will specify its reliability and availability.

In the next step, after implementing the first version of the MHAR-OLSR protocol and testing its performances, we will move on to the last phase of our patent which deals with the choice of path based on the nature of the data to be sent.

References

1. Rani, V., Dhir, R., Ambedkar, B.R.: A study of Ad-Hoc network: a review. Int. J. Adv. Res. Comput. Sci. Software Eng. **3**(3) (2013)
2. Fujiwara, S., Ohta, T., Kakuda, Y.: An inter-domain routing for heterogeneous mobile Ad Hoc networks using packet conversion and address sharing. In: 32nd International Conference on Distributed Computing Systems Workshops, pp. 349–355, June 2012
3. Piechowiak, M., Zwierzykowski, P., Owczarek, P., Wasowicz, M.: Comparative analysis of routing protocols for wireless mesh networks. In: 10th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), pp. 1–5, July 2016
4. Kashyap, V.K., Astya, R., Nand, P., Pandey, G.: Comparative study of AODV and DSR routing protocols in wireless sensor network using NS-2 simulator. In: International Conference on Computing, Communication and Automation (ICCCA), pp. 687–690, May 2017
5. Hari Krishna, B.: Study of Ad hoc networks with reference to MANET , VANET , FANET. In: Conference Proceedings (2017)
6. Rosati, S., Kruelecki, K., Heitz, G., Floreano, D., Rimoldi, B.: Dynamic routing for flying Ad Hoc networks. IEEE Trans. Veh. Technol. **65**(3), 1690–1700 (2016)
7. Benjbara, C., Habbani, A.: Communication in a heterogeneous Ad Hoc network. Int. J. Comput. Inf. Eng. **13**(10), 524–529 (2019)
8. Benjbara, C., Habbani, A., El Mahdi, F.: Procédé de communication basé sur la classification des chemins multiples. Patent, IPC H04W 4/00, MA42342, November 2019
9. Benjbara, C., Habbani, A., El Mahdi, F., Essaid, B.: Multi-path routing protocol in the smart digital environment. In: Proceedings of the 2017 International Conference on Smart Digital Environment, pp. 14–18, 2017
10. Jetcheva, J.G., Hu, Y.C., PalChaudhuri, S., Saha, A.K., Johnson, D.B.: Design and evaluation of a metropolitan area multitier wireless ad hoc network architecture. In: Proceedings Fifth IEEE Workshop on Mobile Computing Systems and Applications, pp. 32–43, October 2003

Integration of Web-Scraped Data in CPM Tools: The Case of Project Sibilla

Alessio Bechini[1], Beatrice Lazzerini[1], Francesco Marcelloni[1], and Alessandro Renda[1,2(✉)]

[1] Department of Information Engineering, University of Pisa, via G. Caruso, 56122 Pisa, Italy
{alessio.bechini,beatrice.lazzerini,francesco.marcelloni}@unipi.it
[2] University of Florence, Florence , Italy
alessandro.renda@unifi.it

Abstract. Modern corporate performance management (CPM) systems are crucial tools for enterprises, but they typically lack a seamless integration with solutions in the Industry 4.0 domain for the exploitation of large amounts of data originated outside the enterprise boundaries. In this paper, we propose a solution to this problem, according to lessons learned in the development of project "Sibilla," aimed at devising innovative tools in the business intelligence area. A proper software module is introduced with the purpose of enriching existing predictive analysis models with knowledge extracted from the Web and social networks. In particular, we describe how to support two functionalities: identification of planned real-world events and monitoring of public opinion on topics of interest to the company. The effectiveness of the proposed solution has been evaluated by means of a long-term experimental campaign.

Keywords: Web intelligence · Business intelligence · Social sensing · Event detection · Opinion mining

1 Introduction

Nowadays, business intelligence (BI) and corporate performance management (CPM) systems are crucial tools for many different enterprises. Anyway, such systems still lack seamless integration with solutions and technologies in the Industry 4.0 (I4.0) domain, such as support to data mining and machine learning functionalities to properly operate on large amounts of data that can be collected from the Web, or acquired through IoT infrastructures. Overcoming such a limitation of current CPM/BI systems means making available tools for predictive analysis and business processes optimization, which so far have often been exclusive prerogative of large companies: thus, it may represent an important growing opportunity for small- and medium-sized enterprises. This is the

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_27

goal of the Italian project SIBILLA (acronym, in Italian, for *design and development of a business intelligence system for Industry 4.0*), which aims to develop a BI system that supports typical I4.0 technologies and is able to exploit Big Data from heterogeneous sources.

This work stems from activities in the SIBILLA project and focuses on the integrated solutions devised for automatic knowledge extraction from the Web and from social networks. In particular, we describe and evaluate how the developed system selectively collects data regarding planned real-world events and captures opinion polarity toward company-related topics. Such information can subsequently be used to derive more accurate BI models for predictive analysis.

In Sect. 2, a brief insight on the background about BI, CPM and Web mining is provided. After a description of the system module for Web crawling and data mining module in Sect. 3, the setup of a preliminary experimental campaign is detailed in Sect. 4. Section 5 reports experimental results while Sect. 6 contains concluding remarks.

2 Background: BI, CPM, Web Mining

BI solutions are software tools aimed to efficiently manipulate data and extract information useful to define enterprise strategies. Traditionally, such tools deal only with company-wide economic data and analyze business performance in economic and financial terms, without identifying countermeasures to improve performance. Instead, CPM systems perform analysis of enterprise processes and key performance indicators (KPIs), leveraging data from several information sources. Thus, a sharper picture of the actual company framework is provided, as a means for making decisions and predicting their consequences.

In this context, data automatically collected from the Web [2], although often unstructured and noisy, may represent an invaluable source of knowledge for companies. Several applications that exploit algorithms in the field of text mining and natural language processing have been recently proposed. It has been recognized that social media streams can offer an insight on the brand reputation and quality perception [3]. In particular, Twitter, as one of the most popular social networks, has been successfully exploited for performing sentiment analysis and opinion mining [7,12]. Moreover, news media streams have been exploited to automatically detect and track new events [8,11] and have been integrated in frameworks for custom application such as city areas profiling [5]. However, to the best of our knowledge, none of the mentioned works conceives *topic monitoring* and *event detection* modules as an integral part of a CPM tool.

3 A System Module for Web Crawling and Data Mining

According to the experience gained in project SIBILLA, we propose a specific software module (named "WCDM") to support the functionalities of Web crawling and data mining. Its internal architecture is depicted in Fig. 1 and consists of the following components: (i) a *social and Web crawling sub-module* to collect

raw data from Web and social networks; (ii) a *data mining sub-module* to capture events from Web ("Event Detection") and uncover the opinion of social networks users toward a specific topic ("Topic Monitoring"); (iii) an *interface sub-module*, to support interaction with other components of the overall CPM tool and configuration/supervision by the administrator. Furthermore, a dedicated database takes care of dealing with required persistence of data, metadata, documents, and indices. The main features of the three sub-modules are described hereafter.

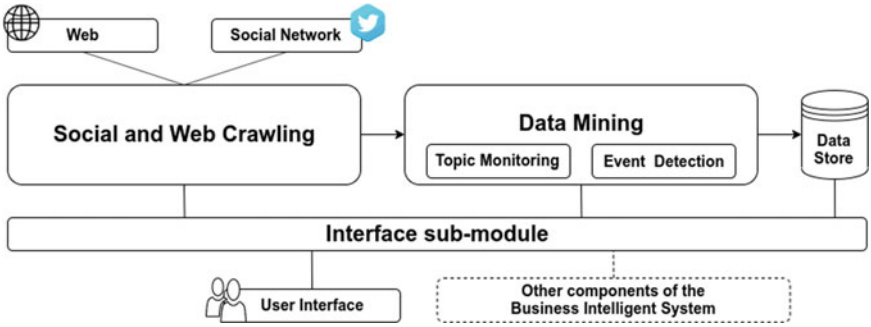


Fig. 1. Architectural scheme of the WCDM software module

3.1 Event Detection Sub-module

The event detection sub-module aims to extract from the Web useful information that characterizes real-world events, both occurred and scheduled ones. In particular, events of interests are those that influence presence (or absence) of people in a given place on a given date and, as such, may have a positive (or negative) impact on the company sales and marginality. For example, the presence of a *festival* causes the gathering of people and therefore can represent a possible favorable business situation. On the other hand, the occurrence of a *flood* or an *earthquake* can explain the negative sales trend in a given region and on a given date. To narrow the search space, the tool user has to configure parameters over three dimensions: the language of the analysis (Italian or English), the list of seed URLs for guiding the crawling process (mainly newswires and newspapers Web sites), and the classes of events of interest (identified by *tag-words*, e.g., "concert," "festival," or "earthquake").

The required data have to be collected *selectively* [10], and the developed solution has been based on Scrapy, a Python open-source scraping framework. The crawler is run periodically, in order to capture newly published events almost in real-time: the list of URLs to visit is dynamically updated by exploiting the RSS feed mechanism, often offered by newspaper Web sites, thus including only

previously unseen links. Text and metadata of the article are extracted on the basis of known structure of the document object model (DOM) or exploiting ad hoc Python libraries.

The characterization of an event instance requires the identification, out of the scraped raw text, of the following attributes: date, place, and class of event. This task is accomplished by means of several tools. Named e recognition (NER) [13] lets us find and categorize entities in the text, and it is particularly suited for extracting *location* information. *Dates* are spotted out by means of regular expressions, adequately designed to discover both single dates (e.g., "on Friday, 22nd November") and date ranges (e.g, "from Friday 6 to Saturday 7 November"). Finally, the most relevant keywords are identified considering the lemmas with the highest number of occurrences in the text, title, and article tags. The article is assigned to one or more classes of events and saved into the database if a *tag-word* matches a lemma in the keyword list (i.e., the word *concert*, which identifies a class of events of interest, belongs to the article keywords).

A news source may publish many updates or reminders of an event that had been published in advance; indeed, after the identification of an event instance, we check if it represents a *new* event or a duplicate of an already discovered event. To avoid storing duplicates, we save a newly captured event Ev only if it satisfies the following condition:

$$Sim(Ev, Ev_i) < TH_{sim} \tag{1}$$

for all the discovered events Ev_i having the same location and date of the new event. $Sim(Ev, Ev_i)$ is a similarity measure between two events (e.g., Jaccard index on the two sets of keywords), and TH_{sim} is a threshold value in $[0, 1]$.

3.2 Topic Monitoring Sub-module

The topic monitoring sub-module aims at keeping track of the people opinion expressed about a predefined topic by the users on social networks. The sub-module requires the tool user to specify a query containing information about the object to be monitored (i.e., input keywords), the language of the analysis, and the time window of the monitoring campaign.

Data are collected from Twitter, the popular social network where users share thoughts and opinions as *tweets*. The Tweepy library has been used in order to access the Twitter streaming API. Whenever a new tweet containing the input keyword is captured, it is saved into the adopted data store. Text of tweets is preprocessed and elaborated using standard text mining techniques (i.e., tokenization, lemmatization, TF-IDF) in order to be converted in numeric features and fed as input to a classification model.

As soon as tweets are collected, a tool user can label them (as positive, negative, or neutral) in order to build a *training set*. It will serve as learning set for a machine learning model that will be used to automatically infer the class of tweets collected afterwards. Before deploying the classifier on the data stream, its performance is evaluated using k-fold cross-validation over the training set:

different configurations of a classification algorithm are compared, and the user can select the most appropriate one on the basis of several metrics. The basic classification model is linear support vector machine (L-SVM), which has proven to be effective for stance analysis on tweets [7,12]. To gain flexibility in different real-world scenarios, two variants of the learning paradigm, particularly suited for dealing with class imbalance, are also provided: L-SVM with automatic class balancing, in which the adjustments on class weights are inversely proportional to class frequencies in the input data (LSVM-b), and L-SVM with oversampling, where the SMOTE algorithm [6] is used to properly balance the dataset.

At the beginning of the monitoring campaign, however, the number of collected and labeled tweets may not be sufficient to train a model in a supervised setting. Even after labelling, the performance of the classification models may be considered unsatisfactory by the user. To deal with such *cold-start*, a class is assigned to any collected tweet even when no trained model is available. The prediction is performed with a lexicon-based classifier (i.e., relying on an external lexical resource) on the basis of the polarity of the words which compose a tweet. Lexicon resources tailored for the sentiment analysis task have been made available for several languages, including Italian [1] and English [9].

As a further analysis, collected tweets are organized into thematic areas using latent Dirichlet allocation (LDA) [4], which has successfully been adopted in [14] to provide a partitioning on tweets as well. A tool user may adopt LDA for two main reasons: (i) to obtain a fine-grained analysis of the sentiment over the discovered sub-topics, and (ii) to identify the keywords that characterize each sub-topic. Potentially, it could help spotting out off-topics and avoid including tweets that contain off-topic keywords with a blacklist mechanism.

3.3 Interface Sub-module

The interface sub-module meets requirements of flexibility, configurability, and accessibility by providing a collection of REST APIs. Examples are `find_events` and `add_twitter_search` routines, which report information about captured events and start a topic monitoring job on Twitter, respectively. Furthermore, the sub-module handles the interaction with other components of the BI system, implementing an asynchronous notification mechanism for effectively handling potentially harmful events in a timely manner.

4 Experimental Setup

The most relevant results have been obtained for the topic monitoring sub-module. In this section, we describe experiments carried out in such context. The system prototype has been tested for uncovering the perception of users about the Gucci tailoring pre-fall 2019 advertising campaign, conducted with singer Harry Styles in the role of testimonial.

The system has been configured to collect tweets containing the words “harry” and “gucci” in the time span between May and October 2019, expressed

in Italian language. During the first month of the monitoring campaign (from May 1st to May 30th), 922 tweets were collected and assigned to one of the three classes (neutral, positive, negative opinion) using unsupervised lexicon-based approach. Although useful to cope with cold-start, the lexicon-based method cannot guarantee good performance in the noisy, informal, and irregular environment such as that of Twitter: Few examples of misclassified tweets are provided in Table 1. Indeed, we randomly sampled 400 tweets for manual labelling for the purpose of training a classification model. The 400 labelled tweets were used to compare the performance of different classification models using tenfold stratified cross-validation. The support of each class is 121 (neutral), 242 (positive), and 37 (negative).

Table 1. Examples of mis-classified tweets by lexicon-based approach

Confusing factor	Original text (*English translation*)	Lexicon prediction	Actual class
Irony or sarcasm	Buongiorno a Harry con gli occhiali e la bandana per la campagna di Gucci Pre-Fall 2019 Voglio piangere https://t.co/l50Z43lAuS *(Good morning to Harry with glasses and bandana for Gucci's Pre-Fall 2019 campaign I want to cry* https://t.co/l50Z43lAuS*)*	Negative	Positive
Irony or sarcasm	UNO NON PUó SVEGLIARSI E TROVARSI FOTO DEL GENERE @Harry_Styles @gucci IO VI DENUNCIO *(ONE CANNOT WAKE UP AND FIND SUCH PICTURES @Harry_Styles @gucci I'LL REPORT YOU)*	Negative	Positive
Ambiguity	Per un Harry Styles che à riluttante ad usare i social, fortuna che alla fine, ci pensa sempre Gucci a fornirci del suo materiale con foto e video inediti *For a Harry Styles who is reluctant to use social media, luckily, in the end, Gucci always provides us with his material with unpublished photos and videos*	Neutral	Positive

5 Results and Discussion

For the sake of brevity, in Table 2, we just report the results of the following methods: lexicon-based, linear SVM (L-SVM), and linear SVM with class balancing (L-SVMb).

Table 2. Results obtained by using different classification methods

Classifier	Class	Precision	Recall	F-score	Accuracy
Lexicon	Neutral	0.60	0.43	0.50	0.54
	Positive	0.76	0.57	0.66	
	Negative	0.17	0.62	0.27	
L-SVM	Neutral	0.70	0.89	0.78	0.70
	Positive	0.72	0.52	0.60	
	Negative	0.33	0.03	0.05	
L-SVMb	Neutral	0.72	0.77	0.74	0.68
	Positive	0.67	0.60	0.61	
	Negative	0.21	0.11	0.14	

Unsurprisingly, supervised approaches (L-SVM and L-SVMb) achieve higher performance than lexicon-based method in terms of accuracy. On the other hand, class imbalance strongly affects the performance of supervised approaches in detecting the negative minority class.

A user can retrieve such performance information using ad hoc REST API and decide whether to label other tweets or to choose the most suitable classification model. For the purpose of our test, we chose the L-SVMb model, which guarantees a high global accuracy and better, albeit poor, performance on the minority class compared to basic L-SVM model. The selected classification model has been used to infer the class of the messages collected during the remaining months of the campaign, i.e., from June to October 2019. We were able to collect 1312 tweets, with a daily volume as depicted in Fig. 2: A visual analysis suggests that the average daily tweet volume is rather limited (8.34 tweets per day), but it shows two spikes in correspondence of advertising events that have triggered the discussion on the social network. In particular, the peak on August 1st can be associated with the news of the leading role of Harry Styles in the advertising of the new unisex fragrance of Gucci (detected polarity distribution: 80% *pos.*, 18% *neut.*, 2% *neg.*). The peak on September 8th may be associated with the broadcasting of the spot on Italian TV (detected polarity distribution: 85% *pos.*, 13% *neut.*, 2% *neg.*). Indeed, thanks to the WCDM module, the CPM tool can be enriched with the valuable information that the advertising campaign is positively perceived by social media users.

To evaluate the performance of our model over time, i.e., on new instances of the data stream, we manually annotated 100 tweets from each event. The

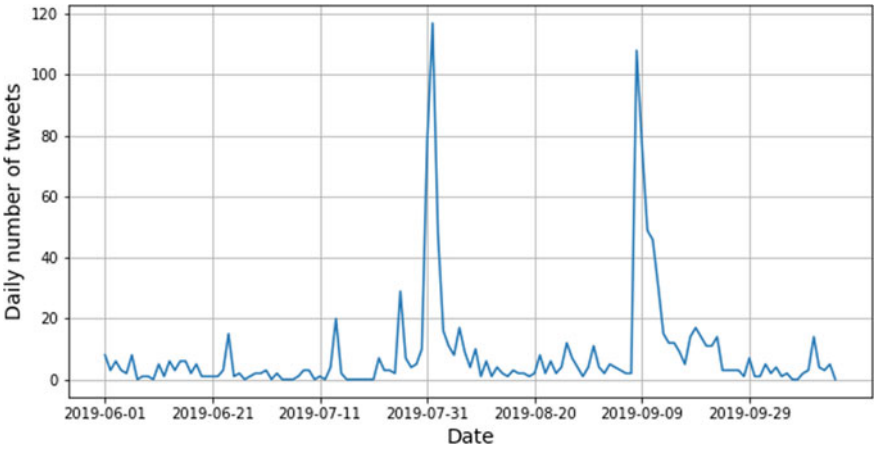


Fig. 2. Number of tweets per day (from 1 June, 2019, to 31 October, 2019

accuracy values obtained (0.70 for the first event, 0.84 for the second one) suggest that our system does not suffer from concept drift, i.e., change in the data generating process which would possibly deteriorate classification performance.

6 Conclusion

In this paper, we have proposed a solution for the enhancement of analysis functionalities of BI/CMP systems by integrating a software module dedicated to deal with data from Web and social networks. In particular, the additional knowledge is obtained by leveraging several NLP and machine learning techniques in the event detection and topic monitoring sub-modules: the first aims at capturing information from the Web about real-world events that may affect company sales; the latter aims at uncovering trends in the public opinion about company-related topics. In a preliminary study, we have shown the effectiveness of the software prototype in performing the topic monitoring functionality.

Acknowledgements. This work was partially supported by Regione Toscana, Italy, in the context of the project SIBILLA, "POR-FESR 2014–2020 - Bando 2" and by the Italian Ministry of Education and Research (MIUR) in the framework of the Cross-Lab project (Departments of Excellence). Authors wish to thank Daniele Agosti for the implementation of some parts of the software, and all partners of the project: FabricaLab srl, BSD Design srl, and TomorrowData srl.

References

1. Basile, V., Nissim, M.: Sentiment analysis on Italian tweets. In: 4th WS on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 100–107. ACL (2013)

2. Bechini, A., Gazzà, D., Marchetti, A., Tesconi, M.: Towards a general architecture for social media data capture from a multi-domain perspective. In: IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), pp. 1093–1100 (2016)
3. Benedetto, F., Tedeschi, A.: Big data sentiment analysis for brand monitoring in social media streams by cloud computing. In: Sentiment Analysis and Ontology Engineering, pp. 341–377. Springer, Berlin (2016)
4. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
5. Cascone, L., Ducange, P., Marcelloni, F.: Exploiting online newspaper articles metadata for profiling city areas. In: Intelligent Data Engineering and Automated Learning IDEAL 2019, pp. 203–215. Springer, Berlin (2019)
6. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
7. D'Andrea, E., Ducange, P., Bechini, A., Renda, A., Marcelloni, F.: Monitoring the public opinion about the vaccination topic from tweets analysis. Exp. Syst. Appl. **116**, 209–226 (2019)
8. Fu, Y., Zhou, M.Q., Wang, X.S., Luan, H.: On-line event detection from web news stream. In: 5th International Conference on Pervasive Computing and Applications, pp. 105–110. IEEE, New York (2010)
9. Hutto, C.J., Gilbert, E.: VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: 8th International AAAI Conference on Weblogs and Social Media (2014)
10. Kumar, M., Bhatia, R., Rattan, D.: A survey of web crawlers for information retrieval. Wiley Interdis. Rev.: Data Mining Knowl. Dis. **7**(6), e1218 (2017)
11. Kumaran, G., Allan, J.: Text classification and named entities for new event detection. In: 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 297–304. ACM, New York (2004)
12. Mohammad, S.M., Sobhani, P., Kiritchenko, S.: Stance and sentiment in tweets. ACM TOIT **17**(3), 26:1–26:23 (2017)
13. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Invest. **30**(1), 3–26 (2007)
14. Zou, L., Song, W.W.: LDA-TM: A two-step approach to Twitter topic data clustering. In: 2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), pp. 342–347 (2016)

Security Issues in MANETs: A Survey

Nada Mouchfiq(✉), Ahmed Habbani, and Chaimae Benjbara

ENSIAS, Mohammed V University - SOUISSI, Rabat, Morocco
mouchfiq.nada@gmail.com, ahmed.habbani@um5.ac.ma, c.benjbara@um5s.net.ma

Abstract. Recently, the notion of security has become increasingly important to research on wireless networks, later the Internet of things (IoT) and ad hoc networks that are a part of IoT environment. The various changes in the digital environment, as it must identify the different changes in the digital environment, are namely the protection of the personal and professional data of this intelligent world. In the work already done, there are several methods and levels of security. In this study, we identify that security aggression MANETs are facing with the security services to provide in each layer. In order to reach our target, we did a literature search gathering information on different types of attacks and solutions, as well as the latest security work. Then, we discuss the proposed solutions to ensure a high level of security in networks by recalling the latest work of our team. Finally, we propose a method of security better adapted to our needs based on blockchain technology.

Keywords: Attacks · MANETs · Security

1 Introduction

Nowadays, the phenomenon of information and communications technology infrastructures and vast flows of information have become essential in defining of modern life. Each challenge represents an ample and discreet notion where know-how and leadership in social work can be focused on new bold ideas, scientific discoveries, and surprising innovations [1]. In the coming years, IoT will make the link between several intelligent devices.

The Internet of things refers to a type of network to connect anything with the Internet [2]. This environment is composed of different networks (cloud, big data, Industry 4.0,..) among them, we find ad hoc networks that are several types: MANETs, VANETs, FANETs ... Each network has his own challenges, and ad hoc networks can face several challenges [3] including: routing and multicasting, quality of service (QoS), energy, security, and mobility.

MANETs are a system of mobile nodes connected with each other via wireless links without infrastructure support [4]. The figure below shows the exact classification of the MANETS networks architecture [1] based on the OSI model (Fig. 1).

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_28

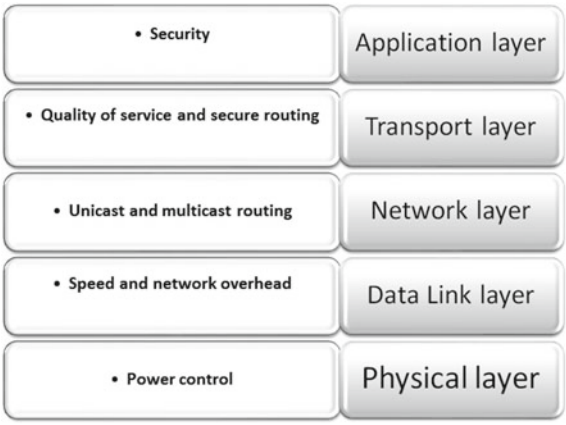


Fig. 1. MANETs layers

Several studies have proposed solutions to solve safety problems that harm MANETs [5]:

- Cluster-based intrusion detection technique.
- Misbehavior detection through cross-layer analysis.
- Defense method against wormhole attacks.
- Defense mechanism against rushing attacks.

The content of the paper is systematized as follows. Section 2: MANETs, Sect. 3: Related Work, Sect. 4: Discussion, and we will finish with the conclusion.

2 MANETs

2.1 MANETs Applications

The decentralized nature of MANETs [6] and the fact that they allow users to connect and trade information regardless of their geographical location or adjacency to the infrastructure makes them more flexible and robust and broadens their domains' use (Fig. 2).

2.2 MANETs Challenges

MANETs are defined as stand-alone systems that are configured using different devices without infrastructure and who or access point such as mobile phones [7]. On the other hand, MANET networks can face several challenges including:

- Energy constraint.
- Constraint of absence of fixed limits.
- Lack of protection against external signals.

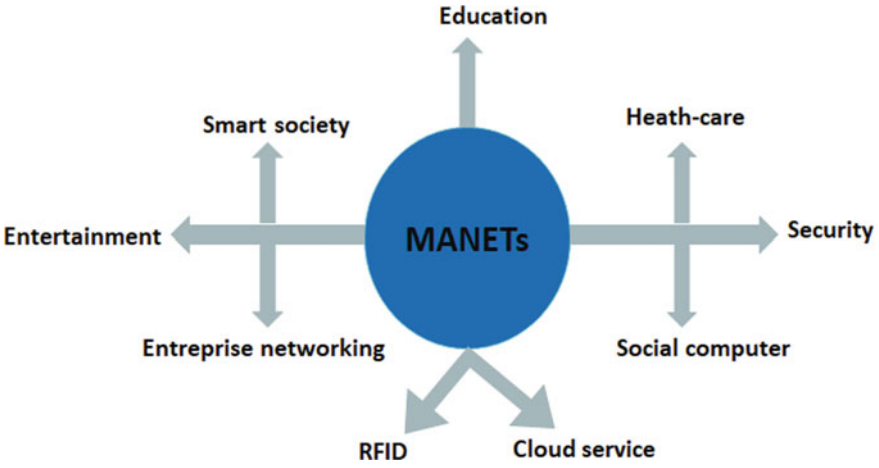


Fig. 2. MANETs applications

- Routing constraint with the dynamics of the network topology.
- Constraint related to variation of link and device capabilities.

MANETs are essentially identified by their active behavior, restricted frequency range, and the fact that they use low-power equipment, restricted processors, involved protocols, and restricted security [8]. It has become necessary to improve the security side of MANET networks.

2.3 MANETs Security

The security solution must give full insurance spanning the whole protocol pile [9]. There is no sole system that can arrange all the security sides in MANETs.

- Confidentiality (C): Ensure that data packets and configuration parameter information are not accessed or appropriated to an attacker or disclosed to unknown entities.
- Integrity (I): ensure that the packets exchanged (or stored) have not been modified in an unauthorized manner (ensure that the data is not modified by unauthorized agents).
- Disponibility (D): Ensure that the packets exchanged are always available and that devices and agents are not prevented from having access to the information.

The dynamic appearance of MANETs makes them vulnerable to multiple attacks, and so many threats can harm the MANETs networks. In the following table, we can find various attacks in each of the five layers in the MANETs [10] (Table 1).

Security should be taken into account at the early stage of design of basic networking mechanisms [11]. In our study, we have identified the security threats

Table 1. MANETs threats

Layer	Attacks
Application layer	Repudiation
	Malicious code
Transport layer	Session Hijacking
	SYN flooding
Network layer	Blackhole
	Grayhole
	Wormhole
	Information disclosure
	Message altering
	Sending data to node out
	Transmission range
	Routing attacks
Data link layer	Jamming
	Selfish misbehavior of nodes
	Malicious behavior of nodes
	Traffic analysis
Physical layer	Eavesdropping
	Active interference
	Jamming

in each layer and corresponding countermeasures [12]. The following list summarizes some potential security issues for MANETs.

- Protecting forwarding protocols
- Preventing signal jamming denial-of-service attacks
- Authentication
- Data encryption
- Providing link
- layer security support.

3 Related Work

3.1 Detection of Node-Misbehavior Using Overhearing and Autonomous Agents in Wireless Ad Hoc Networks

In order to detect a malicious knot, the authors of [13] used two techniques. The first method is based on the principle of warming up, they have named it neighbor misconduct misbehavior (OMD), and it is dedicated to listening nodes and calculating packet transmission ratio (PFR) for each node and the second method focuses on the detection of abnormal behaviors within the Autonomous Agent-based Misbehavior Detection (AAMD).

3.2 SDN Security and Energy Efficiency

In this work, [14] they proposed an architecture for the SDN by shedding light on the operation of each of its components. They give strategies for achieving energy performance in the networks by way of the SDN since the SDN can adapt its parameters according to the size of the information to be sent.

3.3 SDN Framework for Securing IoT Networks

The principle of this model [15] allows the data of the IoT layer to be transmitted in an elitist way via the main framework of the SDN layer, and in which the essential functions of analysis of the security (control of access ... etc.) are put in place, which makes it possible to bridge the separation which could exist between the IoT networks and the usual computer networks.

3.4 Inter-Blockchain Report Model (Connection)

After carrying out the transaction operation [16], the routers as well as the routing tables are updated, and the blockchain system of the router keeps all the addresses of each blockchain. This proposal advances an interactive structure of the blockchain dedicated to the exchange of information from an arbitrary blockchain system.

3.5 Establishing Secure Routing Path Using Trust to Enhance Security in MANET

In this approach [17], a protected direction is put in place in pernicious situations. Given the introduced scheme, thus, more accurate confidence is obtained by considering around the perception of diverted perception of single-rebound bystanders and other detracting factors, such as support of configurations and association circumstances distant, which can cause the fall of parcels. The proposed AODV model significantly improves throughput and packet delivery report (PDR), with a slight expansion in variable end-to-end delay and message overload.

4 Discussion

According to our research and the literature review we conducted in the previous section, it has been found that there are new security approaches applied in other networks such as the blockchain in IoT we have mentioned in our article (Table 2). According to the literature review that we conducted, we found that the proposed solution based on the blockchain principle ensures the two security criteria namely: integrity and confidentiality and remains the availability which can be easily ensured by IDS systems ; on the other hand, the other proposals that we have dealt with only provide one of the security criteria. From

Table 2. Summary of the related works

Authors	Proposed solutions	Results and security criteria verified
[13]	Neighbor misconduct misbehavior	Listening nodes calculating PTR (C)
[13]	Autonomous agent-based misbehavior detection	Detection of abnormal behaviors (D)
[14]	New architecture for the SDN security	Improving energy efficiency (C)
[15]	SDN framework for securing IoT networks	Controlling access (I)
[16]	Inter-blockchain report model	Ensuring the consistency of messages (C) (I)
[17]	Secure routing path using Trust	Improving throughput and PDR (C)

what we have already seen, it is remarkable that the fact of applying blockchain technology to these networks makes them more secure.

A blockchain is a chain of blocks which contain specific information (database). In the case of the assigned network of blockchain [18], every contributor in the network manages, approbates, and restores new accesses. The figure above summarizes the principle of blockchain (Fig. 3).

All affiliates of the network have almost the same statures and stock the similar copy of blockchain. Due to the large security and accuracy, blockchain has been spreaded in diverse functions scenarios and is acknowledged one of the decisive approaches to develop the evolution of the world.

We can summarize the whole procedure on two levels namely: the affirmation of the block and the choice of the chain which is more expanded. By doing so, we ensure that only the appropriate blocks are communicated.

There are major elements counting:

- Checking if the header hash conforms the fixed dilemma
- Block size in predicted assignments
- Block structure
- Checking the timestamp
- Verification of all transactions.

The blockchain can offer many explanations and advantages to increase and boost security such as: elimination of trust theaters, reduction of costs through, automation and simplification, dematerialization of a number of processes, decentralization and dematerialization of transactions and assets.

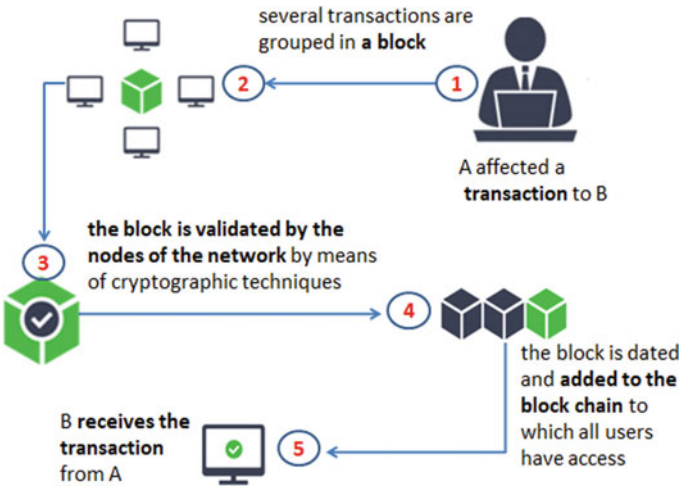


Fig. 3. Blockchain principle

5 Conclusion

In this article, we gave an overview of the MANETs networks and their architecture and infrastructure, then we discussed the security of these networks by citing the security criteria and talking about the offensive intrusions that can damage these structures. We also presented a literature review of the work already done in the security of IoT and ad hoc networks. And based on this study, we found that applying blockchain technology makes the network more secure.

And then as a planned task and in order to ensure the continuity of the team work and especially to ensure security for the network using the improved OLSR protocols developed in our team, we will propose an approach of security more adapted to our needs, this method will be based on the principle of the blockchain.

References

1. Srivastava, V., Motani, M.: Cross-layer design: a survey and the road ahead. In: IEEE Communications Magazine, pp. 112–119, December 2005
2. CONDENSE: A Reconfigurable Knowledge Acquisition Architecture for Future 5G IoT. IEEE Journals & Magazine. CONDENSE: A Reconfigurable Knowledge Acquisition Architecture for Future 5G IoT. IEEE Journals Magazine
3. Yang, H., Luo, H., Ye, F., Lu, S., Zhang, L.: Security in mobile ad hoc networks: challenges and solutions. IEEE Wireless Commun. **11**(1), 38–47 (2004)
4. Shao, M.-H., Lin, J.-B., Lee, Y.-P.: Cluster-based cooperative back propagation network approach for intrusion detection in MANET. In: IEEE10th International Conference on Computer and Information Technology (CIT), pp. 1627–1632, July 2010

5. Goriya, N., Rajput, I.J., Mehta, M.: Low control overhead for cluster maintenance in wireless network for DSR protocol. COMPUSOFT Int. J. Adv. Comput. Technol. **4**(5), 1736 (2015)
6. Quy, V., Ban, N.T., Han, N.D.: A high-performance routing protocol for multimedia applications in MANETs. J. Commun. **14**(4) (2019)
7. Menon, V.G.: A comprehensive survey on opportunistic routing protocols for MANETs: issues, challenges and future directions (2019)
8. Saudi, N.A.M., Arshad, M.A., Buja, A.G., et al.: Mobile Ad-Hoc Network (MANET) routing protocols: a performance assessment. In: Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017), pp. 53–59. Springer, Singapore (2019)
9. Mohammad, S.N., Singh, R.P., Dey, A., et al.: ESMBCRT: enhance security to MANETs against black hole attack using MCR technique. In: Innovations in Electronics and Communication Engineering, pp. 319–326. Springer, Singapore (2019)
10. Khanna, N., Sachdeva, M.: A comprehensive taxonomy of schemes to detect and mitigate blackhole attack and its variants in MANETs. Comput. Sci. Rev. **32**, 24–44 (2019)
11. Zakki Ul Rehman, K., Sharma, A.: Security aspects of MANETs: a review (2019)
12. Kunz, T.: Approaches to Securing P2PSIP in MANETs. In: E-Business and Telecommunications: 14th International Joint Conference, ICETE 2017, Madrid, Spain, July 24–26, 2017, Revised Selected Paper. Springer, 2019, p. 38
13. Agarwal, D., Rout, R.R., Ravichandra, S.: Detection of node-misbehavior using overhearing and autonomous agents in wireless Ad-Hoc networks. In: Applications and Innovations in Mobile Computing (AIMoC), vol. 12, pp. 152–157. IEEE, New York (2015)
14. Danda, B.R., Reddy, S.R.: Software defined networking architecture, security and energy efficiency: a survey **19**(1) (FIRST QUARTER 2017)
15. Krishnan, P., Najeem, J.S., Achuthan, K.: SDN framework for securing IoT networks. In: Ubiquitous Communications and Network Computing, pp. 116–129 (2017)
16. Qasse, I.A., Abu Talib, M., Nasir, Q.: Inter blockchain communication: a survey. In: Proceedings of the ArabWIC 6th Annual International Conference Research Track, p. 2. ACM, New York (2019)
17. Sarbhukan, V.V., Ragha, L.: Establishing secure routing path using trust to enhance security in MANET. Wireless Pers. Commun. pp. 1–11 (2019)
18. Wessling, F., Gruhn, V.: Engineering software architectures of blockchain-oriented applications. In: 2018 IEEE International Conference on Software Architecture Companion (ICSA-C) (2018)

Depression Detection Using Audio-Visual Data and Artificial Intelligence: A Systematic Mapping Study

José Balbuena, Hilda Samamé(✉), Silvana Almeyda, Juan Mendoza, and José Antonio Pow-Sang

Graduate School, Pontificia Universidad Católica Del Perú, Lima, Peru
{jose.balbuena,hsamame,silvana.almeyda,mendoza.jm, japowsang}@pucp.edu.pe

Abstract. Major depression disorder is a mental issue that has been increasing in the last decade, in consequence, prediction or detection of this mental disorder in early stages is necessary. Artificial intelligence techniques have been developed in order to ease the diagnosis of different illnesses, including depression, using audio-visual information such as voice or video recordings and medical images. This research field is growing, and some organizations and descriptions are required. In the present work, a systematic mapping study was conducted in order to summarize the factors involved in depression detection such as artificial intelligence techniques, source of information, and depression scales.

Keywords: Systematic mapping study · Depression detection · Artificial intelligence · Deep learning · Machine learning · Voice · Video

1 Introduction

Rates of depression disorders such as major depression disorder (MDD), usually called depression, have been increasing [1] since the last decade. According to the World Health Organization (WHO) [2], depression is a common mental disorder, and 300 million people are estimated to suffer it, approximately 4.4% of the global population. Due to this fact, a fast method for detecting depression is necessary, and it is a topic of research for many fields such as psychology, medicine, and computer science.

The advances in artificial intelligence (AI) have a significant impact on medical diagnosis, for example, using brain magnetic resonance imaging (MRI) and deep learning algorithms neuro-scientist can predict the diagnosis of autism with a positive rate of 81% [3]. Furthermore, psychologists have been using AI for mental disorder diagnoses such as depression through MRI, biomarkers, and audio-visual modalities [4]. Depression detection using audio-visual information has been a focus of attention, and it is shown

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_29

with the emerging challenges such as AVEC-2016 where video recordings of interviews were used for prediction [5].

Diagnosis of depression disorders using AI is a growing and trending research field due to the effectiveness shown in medical cases. An organization and description are necessary in order to help the development of knowledge in this field [6]. A systematic mapping study (SMS) about the detection of depression through audio-visual modalities and AI is proposed, to get information about trending algorithms, datasets, evaluation, and testing.

This paper is organized as follows: Sect. 2 describes an overall background of depression and artificial intelligence. In Sect. 3, the research methodology and its process are presented. The results of the research are discussed and analyzed in Sect. 4. Finally, the conclusions of the present work are presented in Sect. 5.

2 Background and Related Work

Depression is characterized by persistent sadness and a loss of interest in activities that people enjoy, accompanied by an inability to carry out daily routines, for at least two weeks. Depression symptoms are different for each person [7], some are slowly speaking, movements and thinking, fatigue and tiredness, irritability, and others. Because of the variety of symptoms, different types of tests or questionnaires are used for screening depression [8]. Clinicians choose their preferred method that usually consists of a multiple-choice report, and the most commonly used are listed:

- *Patient Health Questionnaire (PHQ)*: this questionnaire can be applied for different mental disorders, and the ones for depression are PHQ-8 y PHQ-9. The difference is that the PHQ-9 measures suicidal intention.
- *Beck Depression Inventory (BDI)*: is used to measure the depression severity; there are three (03) types and the most used is BDI-II.
- *Hamilton Depression Rating Scale (HDRS)*: is used to determinate depression and to evaluate the recovery of the patients [9].

In the field of AI, according to [10], AI could be understood using four (04) different approaches or definitions, but the one that has been studied lately is the acting humanly. In this, AI approach included techniques such as machine learning and computer vision; both capabilities are needed for diagnosing depression like a field specialist. Machine learning (ML) has techniques for classifying depression disorder or even predicting the level of depression. Some ML techniques allow binary or multiclass classification, such as support vector machine (SVM), decision trees, and probabilistic models. Also, regression techniques such as linear regression (LR) or artificial neural networks (ANN) allow getting a continuous value. The previous algorithms need information in order to learn how to classify or predict depression; for this reason, a dataset of relevant features is needed to train it.

Audio-visual data such as video interviews, speech recordings, medical images, and photographs could be used to understand people's behaviors; also, this information might be the input dataset to train ML models. Audio-visual data are complex for getting

features. However, the advances in deep learning have improved the feature extraction [11] through the use of convolutional neural networks (CNN) or recursive neural networks (RNN). In consequence, audio-visual data have become an essential source of information for ML and automated depression diagnosis.

In the present investigation, a review related to the present work was found, and this paper has reviewed various ML techniques to diagnose MDD and some attempts to provide an effective treatment; in addition, the authors reviewed models that use a neuroimaging dataset to predict the response to the antidepressant [4].

3 Mapping Study Process

A systematic mapping study (SMS) is used to provide an overview of a research area, identifying the quantity and type of research and results available related to the field []. The current SMS was conducted using the guidelines proposed by Petersen in [12] for SMS in the software engineering research area, which is closely related to computer science. The SMS was divided into the three (03) following components:

- Planning: refers to the pre-review activities, starting with the selection of the review methodology; then with identification of the research questions and search terms. The next step is the construction of inclusion and exclusion criteria, to continue with the selection of bibliographic databases; and finally, the definition of search strings for every source.
- Conducting: In this step, a database was generated using the retrieved information from the bibliographic databases. The inclusion and exclusion criteria were applied to the previous database to reduce the researches. Finally, the data were summarized using graphics and tables for a better understanding.
- Reporting: The last step was discussing the results to answer the research question formulated in the first step.

3.1 Research Questions

Research questions elaboration was made using the PICo methodology [13] that allows identifying the searching population, intervention, and context. Some keywords are related to each concept of the PICo method, and these are described in Table 1.

Table 1. Identification of research concepts

Criteria	Field of research
Population	Machine learning techniques
Intervention	Prediction or detection of depression on people based on audio-visual data
Context	Academic or professional works related

After the research concepts were defined, the research questions (RQ) were constructed to organize the insights that were extracted about the use of AI algorithms for

the prediction of depression through audio-visual information. A total of five (05) RQ were proposed and are explained below.

- RQ1: What are the types of depression scales used for detection?
- RQ2: What are the datasets used to train the algorithm?
- RQ3: What are the machine learning or deep learning techniques used?
  - RQ3.1: What are the approaches used on ML techniques in publications?
- RQ4: What are the means of publishing research?
- RQ5: Which are the years and countries with the most significant dissemination of this type of research?

The RQ1 highlights the psychological knowledge (scales or indicators) used in order to detect depression in people, while the RQ2 investigates the datasets that were used in the application of AI algorithms, and the type of multimedia data (images, videos, or sound records).

The RQ3 highlights the main techniques currently used; this information is useful for AI researches and practitioners who are interested to know what techniques have been developing and improving lately. The RQ3.1 allows knowing the approach of the author or authors if it is a proposal, comparison, or validation of the applied technique. The RQ4 investigates the different sources of publication of the studies accomplished. The RQ5 provides insight into the development of such studies around the world and through the years since 2016.

3.2 Database and Search Strategies

To define the search strategy, we made a selection of keywords based on the concepts defined in Table 1, to develop the search strings; search terms are shown in Table 2.

Table 2. Identification of research concepts

Criteria	Field of research
Population	Machine learning techniques
Intervention	Prediction or detection of depression on people based on audio-visual data
Context	Academic or professional works related

For this study, the digital libraries that cover the most substantial amount of literature on machine learning were consulted. For this reason, Scopus, IEEE Xplore, and Web of Science were chosen as they have a significant number of journals and articles related to the ML area. The previously determined keywords are essential to conducting searches in electronic databases; at the same time, the search fields or options offered by each of them must be taken into account.

As a result of searching from the search string defined, a total of 741 publications were returned in Scopus, a total of 68 publications were returned in IEEE Xplore, and a total of 428 publications were returned in Web of Science.

3.3 Paper Selection

The selection criteria were organized in two inclusion (IC) and three exclusion criteria (EC). As inclusion criteria, we defined (IC1) as the article was published since 2016, and (IC2) as research is from journal articles/conference papers. Then, for exclusion criteria, we defined the article does not present a proposal for a method/algorithm/model of detection/classification of depression in people or comparison of ML techniques (EC1). The publication is not related to the machine learning/deep learning field (EC2). Finally, (EC3) as the detection method is not based on voice/images/both data.

Then, in the selection process, a duplicate removal procedure (i.e., papers being present in more than one database) was executed, which resulted in the removal of 327 items. Therefore, with a total of 910 articles, the next stage (Stage 1) continued with the application of the selection criteria. Then, the IC1 was applied, and 683 studies published since 2016 were accepted.

Subsequently, in the next stage (Stage 2), the articles were divided randomly into four groups, so each member made the application of the remaining inclusion and exclusion criteria. Then, to reduce bias, each member reviewed the selection made by another author of this study in order to validate the selection or exclusion of each publication.

As shown in Table 3, the initial search had 1237 results, after excluding duplicate studies, 910 papers were obtained. Following the first stage, 683 papers were retrieved and after the second stage, 653 papers were removed and 30 papers were selected, which are part of the analysis of this research. Table 4 shows the list of the 30 selected papers (in the following link https://drive.google.com/open?id=14y9GFBbyg3Vn7bUDVWlSw6PujdiCKLKs, all the references on Table 4 are listed).

Table 3. Search results

Database	Initial	No duplicates	Stage 1	Stage 2
Scopus	741	741	552	30
IEEE Xplore	68	5	4	0
Web of Science	429	164	127	0
Total	1238	910	683	30

Table 4. List of selected papers

ID	Author	Title	Year
S01	Kliper et al.	Prosodic analysis of speech and the underlying mental state	2016
S02	Ramasubbu et al.	Accuracy of automated classification of major depressive disorder as a function of symptom severity	2016
S03	Cummins et al.	Enhancing speech-based depression detection through gender dependent vowel-level formant features	2017
S04	Kang et al.	Deep transformation learning for depression diagnosis from facial images	2017
S05	Pampouchidou et al.	Quantitative comparison of motion history image variants for video-based depression assessment	2017
S06	Reece et al.	Instagram photographs reveal predictive markers of depression	2017
S07	Simantiraki et al.	Glottal source features for automatic speech-based depression assessment	2017
S08	Syed et al.	Depression severity prediction based on biomarkers of psychomotor retardation	2017
S09	Roniotis et al.	Detection and management of depression in cancer patients using augmented reality technologies, multimodal signal processing, and persuasive interfaces	2017
S10	Yoshida et al.	Prediction of clinical depression scores and detection of changes in whole brain using resting-state functional MRI data with partial least squares regression	2017
S11	Cholet et al.	Diagnostic automatique de l'état dépressif (Classification of depressive moods	2018
S12	Eigbe et al.	Toward visual behavior markers of suicidal ideation	2018
S13	Jan et al.	Artificial intelligent system for automatic depression level analysis through visual and vocal expressions	2018
S14	Cholet et al.	Prototype-based classifier for automatic diagnosis of depressive mood	2018
S15	Koutsouleris et al.	Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: A multimodal, multisite machine learning analysis	2018
S16	X. Lu et al.	A novel method design for diagnosis of psychological symptoms of depression using speech analysis	2018
S17	Mousavian et al.	Feature selection and imbalanced data handling for depression detection	2018

(*continued*)

Table 4. (*continued*)

ID	Author	Title	Year
S18	Roniotis et al.	Detecting depression using voice signal extracted by Chatbots: A feasibility study	2018
S19	Samareh et al.	Predicting depression severity by multimodal feature engineering and fusion	2018
S20	Stepanov et al.	Depression severity estimation from multiple modalities	2018
S21	Trancoso et al.	Speech analytics for medical applications	2018
S22	Al-gawwam et al.	Depression detection from eye blink features	2019
S23	Cash et al.	A multivariate neuroimaging biomarker of individual outcome to transcranial magnetic stimulation in depressions	2019
S24	Ding et al.	Classifying major depression patients and healthy controls using EEG, eye tracking and galvanic skin response data	2019
S25	Lam et al.	Context-aware deep learning for multimodal depression detection	2019
S26	Vincent et al.	Effective classification of major depressive disorder patients using machine learning techniques	2019
S27	Liana et al.	Predicting anxiety from whole brain activity patterns to emotional faces in young adults: a machine learning approach	2019
S28	Tasnim et al.	Detecting depression from voice	2019
S29	Ezekiel et al.	Detecting depression using a framework combining deep multimodal neural networks with a purpose-built automated evaluation	2019
S30	Nan Zhao et al.	See your mental state from your walk: Recognizing anxiety and depression through Kinect-recorded gait data	2019

3.4 Data Extraction

At this step, relevant information was extracted from papers, especially the ones related to the RQs, so after analyzing each publication, new categories were added to our general table. The main information was about categorization of depression, ML method, depression level, dataset, and country.

The data collected from the mentioned categories were mainly obtained by reviewing the abstract, methodology, and conclusions of each one of them. However, in the case of knowing the ML technique used, the methodology section was reviewed to get further information about it. Then, these data were used to analyze the frequencies and design the graphics that will answer the research questions.

4 Results

In this section, the results of the mapping study are discussed, and some tables are shown to present the synthesis of the data.

For RQ1, Table 5 shows the distribution of diagnostic methods or types of depression scale used for detection. Eleven (11) different methods were found among all classified papers; also, it should be noted that in some articles more than one diagnostic method was found. The most used was the Patient Health Questionnaire (PHQ) and Beck Depression Inventory (BDI) with a presence in 10 and 12 papers, respectively.

Table 5. Depression diagnosis methods

Depression diagnosis method	Frequency
Beck Depression Inventory (BDI)	12
Patient Health Questionnarie (PHQ)	10
Hamiltong Rating Scale of Depression (HRSD)	3
Montgomery and Absberg Depression Rating Scale (MADRS)	2
Positive and Negative Affect Schedule (PANAS)	2
Center of Epidemiological Studies Depression Scale (CES-D)	1
Scale for the Assessment of Negative Symptoms (SANS)	1
Snaith--Hamilton Pleasure Scale (SHAPS)	1
Mood and Anxiety Symptom Questionnaire (MASQ-D)	1
Mini-international neuropsychiatric interview (MINI)	1
Self-Rating Depression Scale (SDS)	1
No information	3

The results for RQ2 are shown in Table 6, and it was determined that if the investigation indicated that the dataset was developed in-house, then the dataset was considered to be particular. According to this, it was possible to identify that 11 papers worked with a particular dataset, while 15 papers used the AVEC dataset distributed in AVEC 2013/2014 and AVEC 2016/2017 with 8 and 7, respectively. Besides, no information about the data used was found in 2 papers. Figure 1 shows the distribution of the datasets recognized.

In RQ3, to perform the analysis of the ML techniques used in the articles, we made emphasis on the research approach, depending on this, it was taken into account if the author(s) proposed the employment of a single technique, or if they performed the comparison of different techniques. In these cases, each algorithm was considered as one. However, there were publications where they used various techniques to generate an ensemble method, for example, in S16, the use of neural network, support vector machine, Markov model and Gauss mixture techniques was adopted together as an ensemble method. Moreover, the detailed list of ML techniques used at least once in the included studies is given in Table 6.

Table 6. List of machine learning techniques

Techniques	References	Frequency
Support vector machine	S01, S02, S05, S12, S15, S17, S20, S23, S24, S26, S28	11
Random forest	S06, S17, S18, S19, S24, S28	6
Proposed method	S04, S07, S09, S22, S29	5
Neural network	S05, S11, S14, S21	4
Ensemble	S03, S13, S16	3
Partial least squares regression	S08, S10	2
Gaussian method	S27,S30	2
Logistic regression	S24	1
Support vector regression	S08	1
Bayesian logistic regression	S06	1
Decision tree	S26	1
Linear regression	S30	1
Naive Bayes	S26	1
Convolutional neural network	S25	1
Gradient boosting tree	S28	1
Deep neural network	S28	1

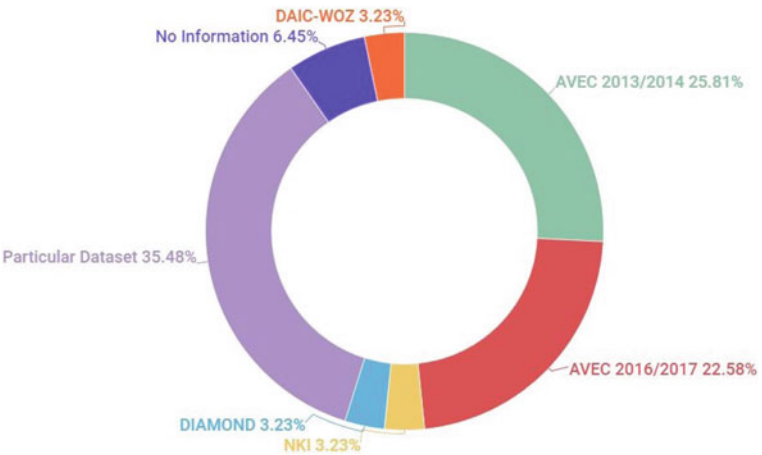


Fig. 1. Distribution of the datasets used in the ML algorithms

For RQ3.1, The approaches of the selected studies were also analyzed; for this, it was divided into three groups: validation, evaluation, and comparison. In the case of comparison, those papers that showed the comparison of the performance of various techniques

were considered. On the other hand, those studies that carried out the application of the techniques on datasets already created were considered as evaluation. Finally, those studies in which they applied the techniques on individuals (real cases) in order to see if it is feasible to use in the detection of depression, were considered as validation. The distribution of the 30 studies gives a result of 16.7%, 33.3% and 50% in comparison, validation and evaluation, respectively.

For RQ4, the media found more results when conducting the investigation were Conference Paper and Article with 18 and 12 papers, respectively.

For RQ5, to provide an overview of the efforts in the machine learning area, the distribution of the 30 studies by country is shown in Fig. 2. The most relevant and widespread documents of this type of research were found in the USA, UK, India, and China with 8, 5, 4, and 4 documents, respectively. The interest of researchers in recent years has increased moderately since 2016, where 6.67% was obtained, with a significant increase in 2018 (36.67%) and 2019 (30%) of the studies carried out. It should be noted that at the time of the search, the year 2019 has not ended, so it is expected that this year may exceed the number of 2018.

Fig. 2. Treemap of the researchers per countries

The results obtained in this systematic mapping show that there is an increase in research on the detection of depression in people through the application of ML techniques. However, studies that have the purpose of proposing an algorithm or technique must be validated to have results that are more in line with reality.

5 Conclusions

The main objective of this study is to analyze and classify the current state of the art (since 2016) in the area of machine learning and the detection of depression in people through multimedia data. This research provides a catalogue of 30 studies found in this topic, which can be helpful to professionals in the areas of computer science, psychology, and medicine.

Our study shows that in the last two years, there is an increase of 50% in the number of publications, which means that more research is being done in this field. Several ML techniques are proposed based on different approaches. We found that all ML techniques used, correspond to the supervised learning algorithms and that the most adopted technique is support vector machine (about 26% of the total), followed by random forest with a 15% and proposed methods with 12%. Upon further investigation, it was discovered that many of the papers had developed their dataset, based on real patients. Though, the second dataset most used is the AVEC dataset.

As future work, it is considered to focus on the study of those publications that perform the validation of the techniques in real scenarios, in order to find the insights of the validation of these types of techniques with more realistic results, as well as the applicability of machine learning in that type of context.

References

1. Hidaka, B.H.: Depression as a disease of modernity: explanations for increasing prevalence. J. Affect. Disord. **140**(3), 205–214 (2012)
2. World Health Organization: Depression and other common mental disorders: global health estimates, p. 24 (2017)
3. Loh, E.: Medicine and the rise of the robots: a qualitative review of recent advances of artificial intelligence in health. BMJ Leader **2**(2), 59–63 (2018)
4. Kharel, P.: Early detection of depression and treatment response prediction using machine learning: a review. In: Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), pp. 1–7 (2019)
5. Valstar, M., Gratch, J., Schuller, B., Ringevaly, F., Lalanne, D., Torres, M. T., Scherer, S., Stratou, G., Cowie, R., Panticz, M.: AVEC 2016—depression, mood, and emotion recognition workshop and challenge. In: AVEC 2016—Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, co-located with ACM Multimedia 2016, pp. 3–10 (2016)
6. Usman, M., Britto, R., Börstler, J., Mendes, E.: Taxonomies in software engineering: a Systematic mapping study and a revised taxonomy development method (2017)
7. Bhowmik, D., Sampath Kumar, K.P., Srivastava, S., Paswan, S., Dutta, A.S.: Depression-symptoms, causes, medications and therapies. Pharma Innov. **1**(3) (2012)
8. Berg, A.O.: Screening for depression: recommendations and rationale. Am. Fam. Physician **66**(4), 647–650 (2002)
9. Guilt,F.O.F.: Hamilton depression rating scale (HAM-D) also HDRS or HRSD, vol. 2, pp. 2–5 (2010)
10. Russel, S., Norvig, P.: Artificial Intelligence: A Modern Approach, 3rd edn. New Jersey: Pearson Education, Inc. (2010)
11. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
12. Petersen, K., Feldt, R., Mujtaba, S., Mattson, M.: Systematic mapping studies in software engineering. in: 12th International Conference on Evaluation and Assessment in Software Engineering, EASE 2008, pp. 1–10 (2008)
13. Murdoch.edu.au, Using PICO or PICo—Systematic Reviews—Research Guide—Subject Guides at Murdoch University

IT&C System Solution for Visually Impaired Romanian Teenagers

Isabela Todiriţe, Corina Radu Frenţ, and Mihaiela Iliescu(✉)

Institute of Solid Mechanics of the Romanian Academy, 15 Constantin Mille, 030167 Bucharest, Romania
{todirite.isabela,mihaiela.iliescu}@imsar.ro,
corinagabrielaradu@gmail.com

Abstract. The system presented in this paper is mainly aimed for improving the quality of education. It is intended to help people with visual, reading–writing disabilities, learning disabilities, and moderate autism spectrum disorders. The system would assist them in social integration as it is well known that such people are isolated and hence other health problems arise over time. In Romania, there is high speed and fast Internet connection, enabling user's access from almost any place of the country. This represents the background for access to education that could start with discovering the keyboard with voice assistance, browse the Internet with voice assistance, also they could learn to use different voice-assisted programs in their native language or choose any other language, thus opening new horizons for education, knowledge, and information. The solution is based on the Vinux-GNU/Linux operating system, which has undergone multiple adaptations, designed to facilitate user access to information even when it has an older generation computer, so no new financial investments are needed.

Keywords: IT&C system · Visually impaired · Romanian teenagers · E-learning · Education

1 Introduction

A useful starting point may be to recognize that visually impaired people need all the information that fully sighted people need: they need to be sufficiently well informed to be able to participate fully as citizens; they need to know about their rights and entitlements; they need information that will enable them to make rational choices as consumers; and they need information to support them in their work, their learning, and their leisure. What differentiates visually impaired from other people is the fact that they may need to receive all this information in accessible formats. In addition to the information needs that they share with everyone else, visually impaired people need information that relates to their position as disabled people. Like wheelchair users or people with hearing difficulties, they need information about specific rights and entitlements open to disabled people. One way to conceive the information needs of visually impaired people, therefore, is to

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_30

identify the core of needs that is concerned specifically with visual impairment. This is surrounded by a related set of needs that are common to all disabled people, which, in turn, is surrounded by a periphery of information needs that are shared by all members of society [1]. Overlaying this is the need for information to be made available in accessible formats.

2 Methodology

The integrated system is money free, with voice support in Romanian starting from installation, an option valid for many languages, which gives people with visual impairments a high degree of independence, getting to install their operating system alone. The installation process involves various stages structured by authors in the diagram—see Fig. 1.

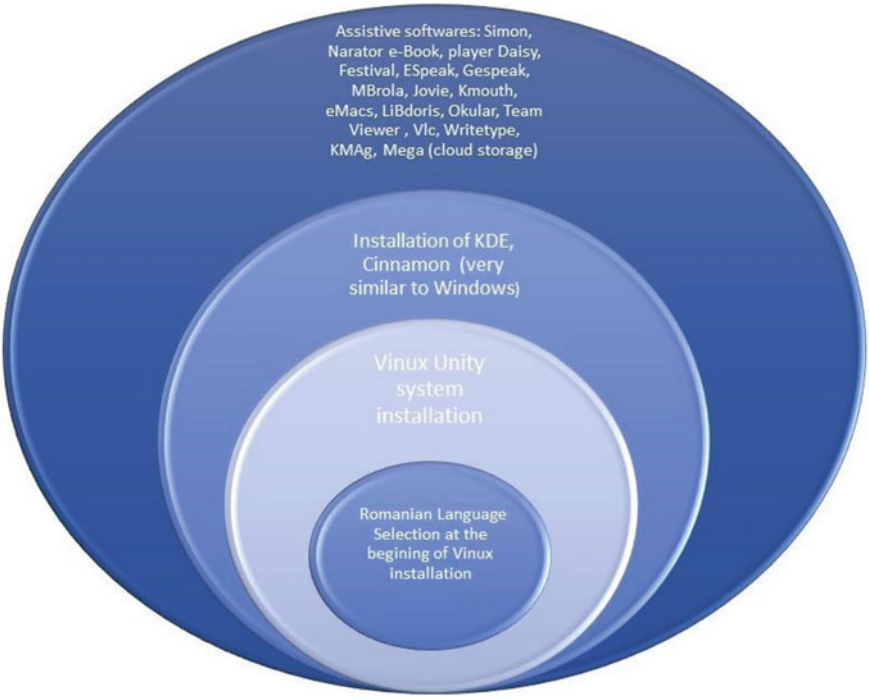


Fig. 1. Diagram that highlights the components integrated into the system

The installation process is assisted vocally, for the first screen the voice speaks in English and the requirements are not understandable, but from the second screen when choosing the Romanian language everything flows in this language. On the first screen, the installation option will be chosen, as shown in Fig. 2.

Even from the beginning, this system has incorporated into the desktop environment, guidance on how to work with the keyboard—see Fig. 3. This element is very important

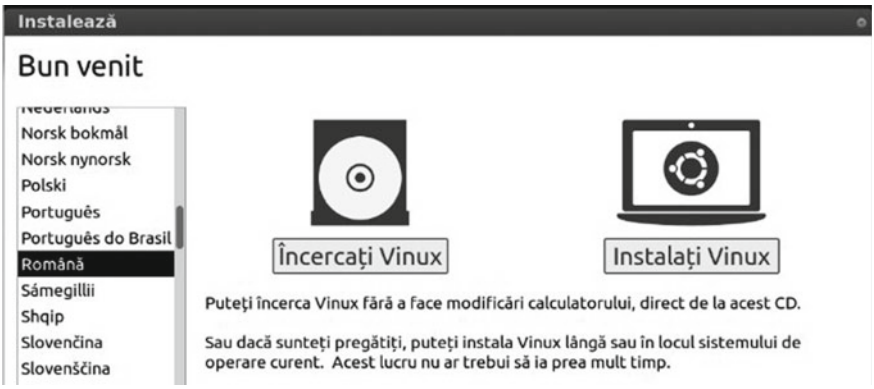


Fig. 2. Screenshot for language choice and its effect (in Romanian language), while installation in the framework of this research

for the visually impaired because without it, there would be required additional effort to search the Internet for tutorials in order to work with the keyboard and the shortcuts defined by this operating system are read by the synthetic voice. This is also useful for people with learning disabilities who are having a harder time memorizing.

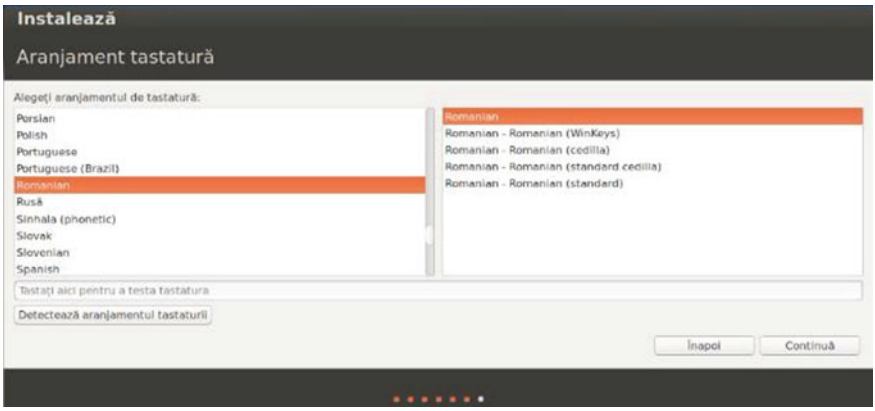


Fig. 3. Screenshot for the keyboard-specific language setting, while installation in the framework of this research

The base operating system contains the LibreOffice suite, Mozilla Firefox Internet browser, email client, and media player, so no further installation is required after the operating system installation is complete.

A suite of programs is needed to help people with visual and/or dyslexic disabilities and uses the computer, namely: Simon, e-Book Narrator, Daisy Player, Festival, eSpeak, Gespeaker, MBrola, Jovie, Kmouth, eMacs, and Libdoris [2]. The next step involves installing the KDE [2] desktop environment that contains the tools needed to read PDF

documents, etc. The desktop environment Cinnamon is also installed due to its similitude with Windows and easy of utilization.

For playing the audio files, we have installed the VLC program, and for the remote assistance, TeamViewer, thus enabling the users when they are in difficulty to be assisted from a distance—see Fig. 4.

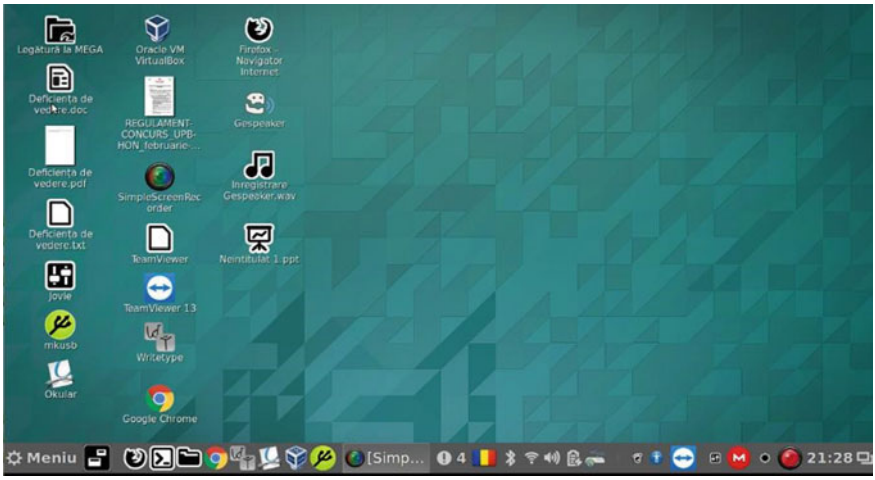


Fig. 4. Screenshot of the completion process of the installation with assistive software—in Romanian language, at the end of installation in the framework of this research

While designing the system it was considered the fact that people with visual disabilities need additional software tools compared to the other people and here we refer to software tools that offer to increase the size of the characters on the screen, namely Kmag (magnifying glass) [2] and voice assistance in computer usage (in TTS—text-to-speech format). The operating system is equipped with Orca Screen Reader and Orca Teacher Application [2]. Orca is the software suite that tells the user what to do, describing vocally what is on the screen and giving directions for certain activities. These facilities are also useful for people with reading and/or writing disabilities and for people with average learning disabilities. We mention another very important facility, namely the spelling support in Romanian, which is of great help to the persons with the mentioned disabilities. The system warns of the misspelling of a word by the message "ortografiere greşită (misspelling)." This research project offers people with visual impairments the opportunity to use cloud technologies: cloud storage (e.g., MEGA [3]), knowing that cloud storage solutions allow access to data from any device (including mobile phone), most cloud storage solutions also provide antivirus protection.

Comparing the Windows platform, there are application packages, which are not free, as shown in the example below.

- The Windows 10 Home operating system in Romanian language costs: 544.99 RON/118 euro (according to [4]).

- MS Suite: Office 2016-Pro + Retail in Romanian language costs 1195 RON/260 euro (according to [4]).
- JAWS [5] in Romanian costs from 260 RON/57 euro (for individuals) to 3373 RON/733 euro for companies, NGOs, institutions, universities, etc. (according to [6]).
- Apple OS also provides assistive solutions for visually impaired, but the price for Apple systems is the highest on market. The cheapest Apple laptop is at [6] 4120 RON (approximately 900 euro).

Also, Google offers some vocal assisting on Android smartphones. There are also other GNU-Linux solutions for visually impaired persons, but not so complete as this one. Further case studies refer to this GNU/Linux operating system, based on Vinux, which gives people with visual impairments a high degree of independence, getting to install their operating system by themselves [7].

3 Case Studies and Results Analysis

Within this research we analyzed the adaptation of the subjects to new and unknown contexts, new type software, and software installation. The subjects are persons with visual disabilities and who also have other disabilities such as learning deficiencies, dyslexia, acalculia, motor disabilities in the lower limbs of different degrees and in the upper limbs the disability to use of at least three fingers.

The subjects are between the ages of 15 and 17 years and with a wide range of manifestations because of the diseases they suffer. The subjects were chosen according to the age criterion and it was desired that those who would take part in the research be newcomers to the special education institution, registered in the first year of vocational school, when the research for them began. The research was planned during one school year. After one year the information collected was analyzed, and for each case, depending on the results, it was decided whether the research would be extended. The maximum period has been considered with extension possibilities up to the whole training period, as long as the student follows the cycle of education which he/she was enrolled to. All target groups consist of 5–10 subjects, due to their special situation/needs.

The target group below was the most varied from the point of view of group members (subjects) manifestations and is worth mentioning, see results in Table 1.

3.1 Target Group C1

In the research, it was taken into account that out of the nine subjects only two have a home computer (S1 C1 and S3 C1, and S3 C1 has a Playstation that uses 90% of the time spent in the family). Of the nine subjects, eight use the smartphone (S4 C1 does not use a smartphone because the parents decided so, and S6 C1 only received the smartphone when it was registered with this institution). When designing the activities, it was taken into account the fact that most students have learning disabilities.

The research began with game-type activities where they were asked to draw objects of nature of their own choice, objects that they can reach, the second activity: to verbally

Table 1. Target group C1 criteria

Subject (S) Category (C)	Age	Sex	Partial/total blindness and other health problems	Institutionalized/in the family (mentions for special situations)	Residence on the campus of the institution/in the family	Urban/rural
S1 (C1)	15	M	20% blindness (in both eyes), motor disability in the lower limbs, sometimes with pain, average learning deficiency	In family	In family	Urban
S2 (C1)	15	F	80% blindness (both eyes)	Institutionalized	In campus	Rural
S3 (C1)	16	M	98% blindness (both eyes), from birth	In family	In family	Urban
S4 (C1)	17	M	10% blindness (both eyes), autism, medium learning deficiency	In family	In family	Rural
S5 (C1)	17	F	Blindness 30% in the left eye, blindness 20% in the right eye, motor disability in the lower limbs, deficiency of average learning	In family	In family	Urban
S6 (C1)	17	M	Total blindness, from birth, acalculia, at 10 years deafened (has a hearing aid), at 13 years he has epilepsy	In family	In family	Rural

(*continued*)

Table 1. (*continued*)

Subject (S) Category (C)	Age	Sex	Partial/total blindness and other health problems	Institutionalized/in the family (mentions for special situations)	Residence on the campus of the institution/in the family	Urban/rural
S7 (C1)	16	F	60% blindness, average learning deficiency	In family, two twin sisters with ADHD and one brother with no problems	In campus	Urban
S8 (C1)	15	F	40% blindness, medium learning deficiency, trichotillomania. She is under psychiatric supervision	In family	In family	Urban
S9 (C1)	17	M	20% blindness, Down syndrome, manifestations in the autism spectrum	In family	In family	Urban

express emotional states related to certain natural phenomena and to draw lightning. Subjects drew an object of their choice, e.g., S1 C1 drew a speedway inspired by the games on the Playstation he plays; S2 C1 because he likes to draw portraits, drew using a magnifying glass a princess head; S3 C1 because he likes to beat the drums he drew two drums; S4 C1 did not manage to express himself by drawing for the activity of his choice; S5 C1 had not long started drawing a little each day and drew a summer landscape; S6 C1 did not draw anything; S7 C1 drew a leaf; S8 C1 drew a flower; S9 C1 drew curved lines unable to say what it represented. The second activity being a concrete one was easier for many of them to express themselves through the drawing, one of them said the lightning scares him and the others did not associate with any emotion. It is interesting to analyze this because the lightning is not a tangible object and S6 C1 being blind at birth stated that he does not know what the lightning looks like, but he can express the emotion that he is trying at the moment, as fear. This student likes to listen stories, fairy tales, plays and he likes to listen them for a long time until he learns and then recites them. He recites very beautifully and feels that he lives what he recites; emotion is a way for him to understand the reality.

Subject S3 C1 has vision problems since he was born, he sees only shadows, he drew lightning as an exclamation mark -!- explained that he perceives lightning as a scratch in the sky and at the base, when it is strong he sees a slightly brighter point. The themes were chosen to see if the subjects in different situations can make a mental representation of the objects, phenomena or if they code in a certain way.

The purpose of this activity was to represent on paper and then make it as a mental projection in order to further discuss standard symbolization, symbols inscribed on various electrical devices. Within the next two months of activities, the subjects had to discover the world around them and there were integrated games as activities so that to accommodate them with the system interface.

Other activities aimed at displaying folders, files, etc., on the screen. This is how the research went and at the end of the school year, the conclusions S8 C1 and S9 C1 only managed to access the Internet and with help, to open applications for drawing, writing, and the others understood the role of the screen reader, how to browse the menus and they made it much easier when they had to surf the Internet, to look for information.

3.2 Target Group C2

Another target group worth mentioning is the one made up of five students with only visual problems, no intellectual problems. Three of them have total blindness and, despite this, in the second semester of using the system managed to install the system by themselves. This group utilized most of application incorporate of the final operating system.

3.3 Target Groups C3 and C4

The target groups C3 and C4 are subject only with visual disabilities (five subjects each), no intellectual problems, but who are not passionate about IT&C. They live in the family and are from the urban environment. They have a minimum knowledge of IT&C from the middle school and have worked under the Windows operating system. They managed to

work with this system quite well after a semester and a half because, as in all cases, only worked at school 2 h/week. They lack enthusiasm when it comes to school. They have home computer and smartphone. They know how to browse Internet, write documents, utilization of media player.

4 Conclusions

The system can be used by people with disabilities, in Romanian language. Depending on their health problems and enthusiasm, they can get to use it the best they can.

Target groups C1 that had subjects with multiple disabilities used a smaller group of software packages (e.g., LibreOffice, Typewriter, Internet browsers, VLC, and Orca suite).

People with learning disabilities, dyslexics, etc., have had a longer learning period than people who are only blind or visually impaired, and therefore, the learning period must be personalized in order to be able to use this integrated operating system.

The students C2 are passionate about IT&C have bought their own laptops, and before they have tried to install themselves several times the Windows operating system, attempts that often failed, the only successful installations were those who have received help from teachers or friends.

This target group C2 was the one with the best results and for them the research has ceased, but an activity of deepening the operating system has started.

These target groups C3 and C4 are not passionate about IT&C, so they obtained average results.

The integrated system is money-free, with voice support in Romanian starting from installation, an option valid for many languages, which gives people with visual impairments a high degree of independence, getting to install their operating system by themselves.

References

1. Moore, N.: The Information needs of visually impaired people. A review of research for the RNIB (2000). https://disability-studies.leeds.ac.uk/wp-content/uploads/sites/40/library/moore-nick-rnib-report.pdf. RNIB. Last accessed 10 Nov 2019
2. Open source GNU/Linux software: https://linux.softpedia.com/dyn-search.php?search_term=vinux. Last accessed on 14 Nov 2019
3. https://www.mega.nz. Accessed on 7 Apr 2019
4. https://www.emag.ro. Last accessed on 10 Oct 2018
5. Journal on Technology and Persons with Disabilities, Volume I, June 2014, ISSN 2330-4219, 28th Annual International Technology and Persons with Disabilities Conference. Homepage URL http://scholarworks.csun.edu/bitstream/handle/10211.3/121930/JTPD201406.pdf?sequence=1. Last accessed on 11 June 2018
6. https://www.anjo.ro. Last accessed on 3 Dec 2018
7. https://opensource.com/life/15/4/6-reasons-people-disabilities-should-use-linux. Last accessed on 21 Dec 2019

A Survey on Named Entity Recognition Solutions Applied for Cybersecurity-Related Text Processing

Tiberiu-Marian Georgescu[1(✉)], Bogdan Iancu[1], Alin Zamfiroiu[1,2], Mihai Doinea[1], Catalin Emilian Boja[1], and Cosmin Cartas[1]

[1] The Bucharest University of Economic Studies, Bucharest, Romania
tiberiugeorgescu@ase.ro
[2] National Institute for Research and Development in Informatics, Bucharest, Romania

Abstract. Named entity recognition (NER) is one of the most common Natural Language Processing (NLP) tasks. As nowadays large quantities of unstructured data are produced, the organizations have begun to be more interested in NER solutions. In the first part of this article, we describe the evolution of NER, and we discuss the most common NER approaches. Later, we address the state-of-the-art NER machine learning solutions. We focus both on open-source and commercial solutions. The most important solutions are identified and compared based on a methodology proposed by the authors. Since the authors are involved in using NER on cybersecurity-related text, the study focuses mainly on NER aspects related to cybersecurity domain. Nevertheless, this survey has a general nature, and therefore, our conclusions can be useful as well for those interested in using NER solutions in other domains.

Keywords: Named entity recognition · Natural language processing · Text processing · Information extraction · Cybersecurity · Machine learning

1 Introduction

1.1 Motivation of Our Work

Our team's research work focuses on domain cognitive analysis and on developing semantic indexing solutions. During this process, we identified NER as an essential task in our projects. Currently, we are involved in developing NER solutions for cybersecurity field, and therefore, the study of NER in this article is conducted mainly for implementing it in the cybersecurity field. Nevertheless, from a technical point of view, the NER solutions are usually very similar regarding the domain of implementation. This study can be useful for anyone interested in implementing NER domain-based solutions.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_31

Article [1] describes a semantic indexing solution of existing vulnerabilities in the Internet of Things (IoT) systems and is developed by two of the authors of this paper (B.I. and T.-M.G.). A core component of the indexing solution consisted of a NER model developed for IoT security. Besides this work, one of the authors (T.-M.G.) developed a cybersecurity cognitive analysis solution available on [2]. Both projects are based on NER models implemented using platform as a service (PaaS)-based solutions. Although the results were satisfactory, we consider using different types of NER solutions in the future, such as open-source and on-premise.

This paper explores the main NER solutions available. Section one describes NER's importance as a part of NLP. Section 2 describes the evolution of NER, as well as the characteristics of the main NER approaches: rule-based, machine learning (ML), and hybrid. A comparison between rule-based and ML based solutions is conducted. Section 3 outlines the NERs based on ML evolution and its main approaches. Section 4 presents a detailed analysis of the main NER solutions identified, both open-source and commercial. A methodology was developed and used to compare the identified NER solutions.

1.2 NER as Part of NLP

The purpose of NLP is to learn, understand, and produce content in natural language by employing computational techniques [3]. NLP represents the computerized approach to the analysis and representation of human language and can be used to automatically extract relevant information from the text. The NLP comprises a wide range of computational techniques for analyzing and representing texts that appear in natural language at various levels of linguistic analysis, to perform human language processing various tasks or applications. Currently, there is an increased interest in the automation of services in NLP [4].

In the context of NLP, information extraction (IE) is the process of automatically obtaining relevant information from a text or group of texts. Such tools facilitate the extraction of relevant information from text documents, databases, websites, and other sources. IE techniques can extract information from unstructured, semi-structured, or structured documents, which can be read by the machine [5]. The task of extracting information is to obtain structured information from unstructured or semi-structured data. There are several subtasks of IE required to obtain structured information, such as NER, relation extraction, part-of-speech tagging (POST), coreference resolution, or parsing.

IE depends primarily on NER, which is used to recognize the targeted information to extract. NER identifies relevant entities and categorizes them on classes and subclasses. One entity can consist of one or more words, called tokens.

The NER process has a lot of approaches and usages in the last decades. In [6], the applications of named entity recognition are presented:

- Information extraction systems;
- Question-answering systems;
- Machine translation systems;
- Automatic summarization systems;

- Semantic annotation.

2 Named Entity Recognition Approaches

According to Jamil and Zafar [7], there are currently three main NER approaches: rule-based, machine learning-based, or statistical approach, and hybrid.

Studying the state of the art, we identified that most of the articles that are comparing rule-based approaches and ML approaches are focused on specific areas and are not general. Most of the studies discuss the differences between rule-based and ML NLP approaches, such as [8], [9] or [10]. Fewer articles discuss the two approaches only from the NER perspective, such as [11, 12].

Table 1 illustrates a comparison between rule-based and ML based NER systems. Further, the results are discussed from our perspective.

Table 1. Comparison of rule-based and ML based NER systems

	NER systems	
Criteria	Rule-based	ML-based
Flexibility	High	Low/Average
Scalability	Low	High
Troubleshooting	Easy	Difficult
Development speed	Average	Low/Average/High
Training data	Small volumes of data are required	Large volumes of data are required
Precision	High	Low/Average
Recall	Low	High
Independence from the domain	Domain dependent	Domain independent
Understanding the la n-guage domain	Greatly needed	It is not compulsory/it is necessary to a small extent
The complexity of parser development	High	Medium / Low
Personnel background	Developers and linguists	Developers and annotators
Programming paradigm	Conventional / explicit	Based on Artificial Intelligence
Computational power needed	Low	High
Extens ibility	Hard	Easy

1. **Flexibility**: Rule-based NER models are very flexible. Modifying the model implies editing existing rules, adding new ones or enriching dictionaries. Such examples can be adding new functions, data types, translation rules, etc. Usually, these upgrades can be implemented without significant improvements to the core system. On the contrary, one ML NER model that is built and trained is much more difficult to modify. Usually, the data corpus is large, and therefore, modifying the ground truth implies a large volume of new training data.
2. **Scalability**: Rule-based NLP, as well as rule-based NER projects, are generally not very scalable. The bigger a rule-based system is, the more complex it tends to be, therefore more rules are needed. A high number of interacting rules leads to poor results. ML NER can be very scalable, and once the model is well built, it is just a matter of training process. There are many huge NER based on ML projects available.
3. **Troubleshooting**: Rule-based NER modeling evolution can easily be observed and readjusted ongoing. On the other hand, ML NER models, once trained, they are very difficult to troubleshoot. They rely on statistical and probabilistic results, which cannot be fully observed. Sometimes, a small amount of training data can have a big impact on the model.
4. **Development speed**: The speed can vary very much for ML based NER, depending on the approach. Besides building the model, the training time can vary very much depending on the volume of corpus and the type of ML: supervised, unsupervised, or semi-supervised. Supervised or semi-supervised solutions can require much time, depending on the complexity of the model.
5. **Training data**: Machine learning systems require massive training data, compared to rule-based NER for which is necessary small amounts of data or not even any data.
6. **Precision**: If precision is important for a project, then rule-based seems to be the better choice. It is very difficult for an average or big project to have a precision over 85% using ML NER.
7. **Recall**: When a high coverage is needed, then ML is the better choice, especially for medium to large projects, where the limited number of rules cannot cover as much as the probabilistic approaches.
8. **Independence from the domain**: Rule-based NER models are very dependent on the domain, compared to unsupervised approaches, for example.
9. **Understanding the language domain**: Rule-based NER models can hardly be adapted for a new language, compared to ML models where usually it is possible through new data corpus and specific adjustments.
10. **Parser complexity**: Rule-based NER models quality depends very much on the parser's performance. The parser is responsible for handling various tasks such as fixing the syntax or eliminating the typos.
11. **Personnel background**: In rule-based approach, linguistic or knowledge engineers are necessary to model the rules, compared to the ML approach, where annotators are required.
12. **Programming paradigm**: Rule-based uses a classical programming approach, where ML implies an artificial intelligence-oriented approach.

13. **Computational power**: ML approach requires more computational resources than rule-based, especially for big NER projects.
14. **Extensibility**: An ML NER system can be extended by changing or adding more training data. Rule-based NER approaches are more difficult to extend when it comes to adding more rules because this can increase the complexity and worsen the performances [13].

Many modern NER solutions use a hybrid approach, combining the advantages of both rule-based and ML based systems.

3 NER Based on ML Approaches/Evolution

Named entity recognition, NER, is an important process of Natural Language Processing domain, consisting of decoding and extracting the meaning of text entities, part of a more general context. On its freshly start [14], NER has been a component that mostly relied on deterministic rules that were able to extract named entities from text structures, based on different patterns that seemed to apply to specific elements. But this approach was not as efficient as needed, and therefore, other directions were researched. The most notable one is with the help of machine learning algorithms that managed to give better results using a probabilistic approach of the problem [15]. A series of supervised and unsupervised machine learning models were evaluated in the field of NER systems, some of them performing better than others, but mostly in a single language [16, 17] or single domain setup. According to Yadav and Bethard [18], there are no NER systems with good performance in both multi-lingual and multi-domain.

Based on the prerequisites that a NER system needs to be able to identify named entities from contextual text structures, a thorough study [18] suggested the following classification:

- Knowledge-based systems refer to NER systems that use language-specific lexicons and domain-specific knowledge; such systems have quite an outstanding performance when dealing with homogeneous texts where the named entities follow the same set of rules; whenever the domain changes, such systems can be influenced by a level of noise [19] coming from the acquisition process or even from different terms pertaining to different domains [20];
- Unsupervised and bootstrapped systems—this approach is based on exploiting the patterns that are found in different named entities by using key terms or seeds, from which named entities are composed of; these patterns have a clear semantic, and they are called extractors, easy to be distinguished, and extracted from the context just by searching for these semi-fixed cue phrases such as: “The paper proposes to present ….” or “The solution is emphasizing the advantages of …” [21];
- Feature-engineered supervised systems represent NER systems that are based on domain knowledge of data in order to identify named entities; supervised machine learning algorithms are used, so that residual knowledge that exists in the training corpus is to be efficiently harnessed; the performance of algorithms and models like support vector machines, hidden Markov model, conditional random field, or the maximum entropy was measured, with compared results in [22];

- Feature-inferring neural network systems—such systems make use of the inference process allowing the NER model to be applied onto a totally new set of features, learning to identify new patterns as they ingest new data; this approach can be applied on different levels when in the presence of NER systems: word-level based [23], character-level based, or combined architecture.

Machine learning algorithms brought remarkable results in the field of named entity recognition systems by allowing researchers to create artificial intelligence models that are based on different combinations of classification algorithms. Paper [24] discusses the advances in ML based NER in more detail. The key aspect is that there is not only one machine learning algorithm that needs to be used to get good results but multiple models that need to be connected for a NER system to have remarkable results.

4 Named Entity Recognition Solutions: Open Source Versus Commercial

The named entity recognition process was implemented in the last decades by taking different approaches, as presented in the previous sections. To choose the right solution for processing text related to the cybersecurity domain, we split the existing systems into two categories: open source and commercial. Our goal is to decide what is the best solution for our project from each category. Based on the obtained result, we can do a cost/benefits analysis that will help us better choose the optimal solution.

In [6] and [25] different systems for named entity recognition are presented. All the systems are grouped by year, programming language, techniques, corpus, and obtained results. We have selected four open-source solutions from this analysis: TagMe, TextRazor, Stanford NER, and Open Calais.

In [26], a comparative analysis of SaaS algorithms for NER is conducted. The solutions presented in the analysis are DBPedia Spotlight, Google cloud natural language, Azure Machine Learning, and IBM Watson Natural Language Understanding. Some of them are also presented in [25] and [27].

Starting from all the identified solutions, we have chosen five relevant attributes that can be used to create a metric that computes the popularity/impact of each solution. To compute the values for these attributes, we have used the GitHub platform to get the number of repositories (1), the number of commits for each solution (2), the number of issues for each solution (3). From the StackOverflow platform, we took the number of tags per solution (4) and the number of current contributors on OpenHub (5). Table 2 presents the computed values.

We can easily observe that the SaaS solutions do not have contributors to the OpenHub platform that was expected because these solutions are not open source.

For the five criteria presented, we determined the weights based on a survey fulfilled by 50 cybersecurity specialists who rank the importance of each criterion. The survey group was composed from specialists that have experience and knowledge in the cybersecurity field and work in software development. The final weights are presented in Table 3.

Table 2. Computed metrics for the identified solutions

	Solution	Number of repositories on GitHub (1)	Number of commits on GitHub (2)	Number of issues on GitHub (3)	Number of tags on StackOverflow (4)	Number of contributors on OpenHub (5)
Open-source solutions	TagMe	132	541	286	0	0
	TextRazor	32	187	64	0	0
	Stanford NER	160	933	1166	2999	17
	Open Calais	42	132	119	27	0
SaaS based solutions	DBPedia Spotlight	70	429	503	1244	0
	Google cloud natural language	91	671	3619	59	0
	Azure Machine Learning	601	3616	3318	809	0
	IBM Watson Natural Language Understanding	35	32	148	1896	0

Table 3. Weights for each criteria

Criteria	NRGH(1)	NCGH (2)	NIGH (3)	NTSO (4)	NCOH(5)
Weight	0.2	0.21	0.21	0.19	0.19

The metric that determines the impact for each solution is determined by the formula:

$$\begin{aligned} \text{DVIS} = {} & \text{NRGH} * W_{\text{NRGH}} + \text{NCGH} * W_{\text{NCGH}} + \text{NIGH} * W_{\text{NIGH}} \\ & + \text{NTSO} * W_{\text{NTSO}} + \text{NCOH} * W_{\text{NCOH}} \end{aligned}$$

where:

DVI Development velocity and impact score;
NRGH Normalized value for “Number of repositories on GitHub”;
W_{NRGH} Weight for the “Number of repositories on GitHub” criterion;
NCGH Normalized value for “Number of commits on GitHub”;
W_{NCGH} Weight for the “Number of commits on GitHub” criterion;

NIGH Normalized value for "Number of issues on GitHub";
W_{NIGH} Weight for the "Number of issues on GitHub" criterion;
NTSO Normalized value for "Number of tags on StackOverflow";
W_{NTSO} Weight for the "Number of tags on StackOverflow" criterion;
NCOH Normalized value for "Number of contributors on OpenHub";
W_{NCOH} Weight for the "Number of contributors on OpenHub" criterion.

The proposed metric was computed for the presented data. The obtained values were normalized using the Von Neumann–Morgenstern method. Table 4 describes the obtained values.

Table 4. Normalized values for our metric

	Solution	Number of repositories on GitHub	Number of commits on GitHub	Number of issues on GitHub	Number of tags on StackOverflow	Number of contributors on OpenHub	DVIS
Open-source solutions	TagMe	0.22	0.15	0.92	0.00	0	0.27
	TextRazor	0.05	0.05	0.98	0.00	0	0.23
	Stanford NER	0.27	0.26	0.68	1.00	1	0.63
	Open Calais	0.07	0.04	0.97	0.01	0	0.23
SaaS based solutions	DBPedia Spotlight	0.12	0.12	0.86	0.41	0	0.31
	Google cloud natural language	0.15	0.19	0	0.02	0	0.07
	Azure Machine Learning	1.00	1.00	0.08	0.27	0	0.48
	IBM Watson Natural Language Understanding	0.06	0.01	0.96	0.63	0	0.34

We can conclude, based on our metric, that the Stanford NER solution is the most popular when it comes to performing named entity recognition. The cybersecurity specialists who answered our survey said that they prefer open-source solutions (80%) to commercial ones (20%). One reason could be the large number of plugins created by the community for these solutions.

It is important to be aware that the obtained results are influenced by the chosen solutions and criteria. If other solutions or criteria would have been chosen, the result might be different. Better analysis can be made only with a big number of solutions and by taking into consideration a large number of criteria.

5 Conclusion

This article discusses the main NER approaches and illustrates a comparison between the most significant NER solutions based on a methodology we proposed. Our motivation is to find the best NER approaches, methods, and instruments in order to apply them for cybersecurity-related text processing.

Real life has proven that the cybersecurity field is too complex human professionals to be able to evaluate and monitor, in real time, the multitude angles and perspectives of the attacker, the defender, and the software developer. The attacker will always try to exploit either vulnerabilities or blind spots created during the development and testing phases, or to find different vectors of attack, that sometimes may sound impossible or very difficult to be implemented. The defender must outthink and never underrate the attacker, must verify all the nodes in its chain of tools and solutions, and must be aware of all the security vulnerabilities that are found on a daily basis [28]. All these perspectives are important and the last one could be difficult to cope with as the volume of information is increasing every day. Therefore, we think that cybersecurity needs components that will automatically collect, identify, or classify vulnerabilities based on public or private sources of information that sometimes start with a simple blog article or a post in a cybersecurity forum.

Funding. This paper presents results obtained within the PN-III-P1-1.2-PCCDI-2017-0272 ATLAS project ("Hub inovativ pentru tehnologii avansate de securitate cibernetică/Innovative Hub for Advanced Cyber Security Technologies"), financed by UEFISCDI through the PN III—"Dezvoltarea sistemului national de cercetare-dezvoltare," PN-III-P1-1.2-PCCDI-2017-1 program.

References

1. Georgescu, T.-M., Iancu, B., Zurini, M.: Named-entity-recognition-based automated system for diagnosing cybersecurity situations in IoT networks. Sensors **19**(15), 3380 (2019)
2. Georgescu T.-M.: Cybersecurity analyzer. [Online]. http://www.cybersecurityanalyzer.com/ (2019)
3. Hirschberg, J., Manning, C.D.: Advances in natural language processing. Science **349**(6245), 261–266 (2015)
4. Fritzner, J.E.H.: Automated information extraction in natural language. MS thesis, NTNU (2017)
5. Rouse, M.: Information extraction (IE). [Online]. https://whatis.techtarget.com/definition/information-extraction-IE (2018)
6. Goyal, A., Kumar, M., Gupta, V.: Named entity recognition: applications, approaches and challenges. Int. J. Adv. Res. Sci. Eng. **6**(10) (2017)
7. Jamil, Q., & Zafar, M.R.: Big data and named entity recognition approaches for Urdu language. EAI Endorsed Trans. Scalable Inf. Syst. **5**(16), e3 (2018)
8. Cronin, R.M., Fabbri, D., Denny, J.C., Rosenbloom, S.T., Jackson, G.P.: A comparison of rule-based and machine learning approaches for classifying patient portal messages. Int. J. Med. Inf. **105**, 110–120 (2017)
9. Khan, W., Daud, A., Nasir, J.A., Amjad, T.: A survey on the state-of-the-art machine learning models in the context of NLP. Kuwait J. Sci. **43**(4) (2016)

10. Dorash, M.: Machine learning vs. rule based systems in NLP. https://medium.com/friendly-data/machine-learning-vs-rule-based-systems-in-nlp-5476de53c3b8 (2017)
11. Gorinski, P.J., Wu, H., Grover, C., Tobin, R., Talbot, C., Whalley, H., Alex, B.: Named Entity Recognition for Electronic Health Records: A Comparison of Rule-based and Machine Learning Approaches. arXiv preprint arXiv:1903.03985 (2019)
12. Derczynski, L., Maynard, D., Rizzo, G., Van Erp, M., Gorrell, G., Troncy, R., Bontcheva, K.: Analysis of named entity recognition and linking for tweets. Inf. Process. Manag. **51**(2), 32–49 (2015)
13. Toral, A.: Machine learning in natural language processing. In: Guest Lecture for the CA 684 Machine Learning Module. https://www.semanticscholar.org/paper/Machine-Learning-in-Natural-Language-Processing-for-Toral/8901d400868e17606e6641001638f44323b93430 (2015)
14. Grishman, R., Sundheim, B.: Message understanding conference-6: a brief history. In: COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics (1996)
15. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Investigationes **30**(1), 3–26 (2007)
16. Leaman, R., Gonzalez, G.: Banner: an executable survey of advances in biomedical named entity recognition. In: Biocomputing 2008, World Scientific, pp. 652–663 (2008)
17. Patil, A.S., Pawar Nita Patil, B.V.: Survey of named entity recognition systems with respect to indian and foreign languages. Int. J. Comput. Appl. **134**(16) (2016)
18. Yadav, V., Bethard, S.: A survey on recent advances in named entity recognition from deep learning models. arXiv preprint arXiv:1910.11470 (2019)
19. Colavizza, G., Rochat, Y., Ehrmann, F.K.M.: Diachronic evaluation of NER systems on old newspapers. In: Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016), pp. 97–107 (2016)
20. Vilain, M., Su, J., Lubar, S.: Entity extraction is a boring solved problem: or is it? In: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics, vol. Companion Volume, Short (2007)
21. Abdalla, R.M., Teufel, S.: A bootstrapping approach to unsupervised detection of cue phrase variants. In: Proceedings of ACL/COLING (2006)
22. Putthividhya, D., Hu, J.: Bootstrapped named entity recognition for product attribute extraction. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics (2011)
23. Chalapathy, R., Borzeshi, E.Z., Piccardi, M.: An investigation of recurrent neural architectures for drug name recognition. In: Proceedings of the Seventh International Workshop on Health Text Mining and Information Analysis (2016)
24. Li, J., Sun, A., Han, J., Li, C.: A Survey on Deep Learning for Named Entity Recognition. arXiv preprint arXiv:1812.09449 (2018)
25. Abdallah, Z.S., Carman, M., Haffari, G.: "Multi-domain evaluation framework for named entity recognition tools. Comput. Speech Lang. **43**(43), 34–55 (2017)
26. Iancu, B.: Comparative analysis of the main SaaS algorithms for named entity recognition applied for romanian language. Rom. J. Inf. Technol. Autom. Control. Revista Romana de Informatica si Automatica **28**(1), 25–34 (2018)
27. Heuss, T., Humm, B., Henninger, C. Rippl, T.: A comparison of NER tools wrt a domain-specific vocabulary. In: Proceedings of the 10th International Conference on Semantic Systems, pp. 100–107 (2014)
28. Andrade, R.O., Yoo, S.G.: Cognitive security: a comprehensive study of cognitive science in cybersecurity. J. Inf. Secur. Appl. **48**(102352) (2019)

URL's Folder Name Length as a Phishing Detection Feature

Melad Mohamed Al-Daeef[1], Nurlida Basir[2(✉)], and Madihah Mohd Saudi[2]

[1] Department of Data Analysis and Computer, Faculty of Economics, Elmergib University, Al-Khoms, Libya
mmaldaeef@elmergib.edu.ly

[2] Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Nilai, Malaysia
{nurlida,madihah}@usim.edu.my

Abstract. Phishing is a cybercrime in which, the phishers try to control users' credentials. The phishers usually construct fake URLs that take to phishing websites where the users might disclose, and loss their credentials. This paper introduces a new feature to detect phishing URLs. The length of folder name of URL's path is utilized in this work as a phishing detection feature, namely FldrNmLnth. Results from analysis of two email datasets show that many of phishing URLs are constructed using upto 230 characters as names of their folders' names. The length of folders' names of legitimate URLs, on the other side, has not exceeded 30 characters. This length variance motivates the use of folder name length as a feature to detect phishing URLs.

Keywords: Phishing URL · URL path · Folder name

1 Introduction

Numerous of transactions can be performed throughout the Internet. Unfortunately, these transactions are at the risk of phishing attacks when the phishers try to steal users' personal information for improper use. In 2016, a report by Anti-Phishing Working Group show that the retail and financial industries are the most targeted sectors by the phishers [1]. Phishing activities, however, might target other industries and Internet-based services. Phishing attacks usually start by sending phishing emails with fake links (URLs) that take the victims to phishing websites that previously designed to look like the original ones. At phishing websites, the phishers can collect victims' credentials for improper use. Since most of phishing emails contain fake URLs, the structure of these URLs, therefore, can be analyzed to detect the attacks.

URL is an abbreviation for Uniform Resource Locator which is used to locate the resources on the Internet [2]. Figure 1 shows the general structure of the URL which generally consists of three parts: the protocol, the domain name, and path. The URL

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_32

path supplies the details of how a resource can be accessed, the path constructed from the folder and/or sub-folder, and file names. The FldrNmLnth in this paper refers to the length of the folder and/or sub-folder names of a given URL's path.

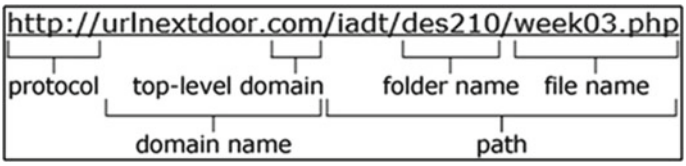


Fig. 1. General structure of the URL

The phishers can modify the folder name part of URL path to construct fake URLs [3]. Researchers in [4] have shown an example of how the images sub-folder was used to abnormally hold a sub-folder named www.chase.com whereas it should hold the image sources of the entire website. The following is the phishing URL; http://hitechsense.com/images/www.chase.com/update.php?. By modifying the folder name part, the phishers can construct fake URLs using very long path names. Thus, they can hide some parts, especially the end part, of the phishing URLs since the address bar of Internet browsers is usually limited to display a finite number of characters. The phishers can place the active going-to destination of the phishing URL as a name of one of its folders or sub-folders to conceal it. The long URLs, in addition, are difficult to be memorized by a human, the Internet users; therefore, will find it is more convenient to click on the URLs in phishing emails instead of memorizing and rewriting them into their browsers [4–7].

The analysis of URLs from two legitimate and phishing email datasets shows that the phishing URLs are constructed using a noticeable longer folder names than the folder names of legitimate URLs. The analyzed URLs were extracted from more than 13,000 legitimate and phishing emails from two datasets. The experimental results in this paper show how to use the folder name length to detect the phishing URLs. This method presents a new phishing URL detection feature that targets a subset of phishing attacks based on the length of URL's folder name. This feature, however, cannot be relied upon as a solo feature to detect the attacks, and many of phishing URLs might use folder names with normal length. Achieved results will be more reliable if the FldrNmLnth is utilized to work with other URL-based features.

This paper is constructed as the following; Sect. 2 discusses the related work that use the URL-based phishing detection features in general and the URL's path and folder name in special. Section 3 presents the experiments and the results from utilizing this newly introduced phishing detection feature. Section 4 lastly concludes the paper.

2 Related Work

Since the analysis of URLs can lead to detect more phishing attacks, the researchers in many studies, therefore, have utilized a numerous of URL-based phishing detection features. URLs' lexical properties, for example, were used in many studies to identify the phishing URLs [5, 8, 9]. Researchers in [10] have treated the common substrings and

unique URLs folders' names as a lexical features to detect malicious URLs that might be generated by malicious programs. Other researchers have used the delimiters of URL's path string as a binary feature to identify malicious URLs, such delimiters include (/, ?, ., = , - and _) [8]. Since phishing URLs often have unusual length, researchers in other studies have implemented the length of URLs and subdomains as a feature to detect phishing URLs [11, 12]. Other researchers have used the length of URL domain name part to detect phishing URLs [13]. Researchers in [14] have adopted the idea of the folder structure similarity to generate, or predict, a blacklist of new suspicious URLs from existed phishing URLs that share the same common folder structure. The literature shows that, a variety of URL-based features were used by the researchers, and these features include but not limited to URL IPAddress, URL noIPAddresses, URL @Symbol, URL noLinks, URL noIntLinks, URL noExtLinks, URL no-ImgLinks, URL noDomains, URL maxNoPeriods, URL linkText, URL nonModalHereLinks, URL ports, URL noPorts, a description on these features can be found in [15].

Although many researchers have studied the URL characteristics and they have utilized a numerous of URL-based phishing detection features, no work, however, has previously considered the length of URL's folder name as a phishing detection feature. In this paper, the folder name length of URL's path is used as a feature to determine whether a given URL is a legitimate or phishing instance. Figure 2 shows legitimate URL examples, whereas Fig. 3 shows phishing URLs that use extremely long folders' names.

```
http://www.calpx.com/regulatory/marketcompliance/annual
report.html

http://www.verisign.com/developer/rsc/qd/authent
icode/index.html
```

Fig. 2. Examples of legitimate URLs

```
http://ibank.internationalbanking.barclays.co.uk.vectorpublishin
q.ru/OFSTWljaGFlbCBTb3V0aHdhcmQ7NjAqRGVubWFyayBSYW9kLCBDb3R0ZW5o
YW07Q2FtYnJpZGdlO0NCNCA4UVM7TWljaGFlbCBTb3V0aHdhcmQ7NjAqRGVubWFy
ayBSYW9kLCBDb3R0ZW5oYW07Q2FtYnJpZGdlO0NCNCA4UVM7/LoginMember.do.
php

http://202.85.146.119:8081/citibusinessonline.daus.citibank.com.
cgibin.refresh.q8913bu99912908662ba.update.db.ls351srv0001293838
3199867.bin/cbusol/signon.do.htm
```

Fig. 3. Examples of phishing URLs that use long folder names

3 Experiments

In this paper, a URL is identified as a phishing instance if one of its folders' and/or sub-folders' names is longer than 30 characters. This length, 30, was decided based on

the experimental results of analyzing two of legitimate and phishing email datasets. The URLs that extracted from the legitimate email dataset were found using less than 30 characters for their folder names. On the other side, the URLs that extracted from the phishing email dataset were found using more than 30 characters for their folder names.

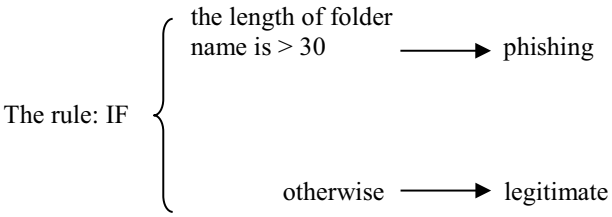
The legitimate email dataset is comprised of 10,000 emails collected by Shetty and Adibi [16], Cormack and Lynam [17], and Spamassassin public corpus [18]. The phishing email dataset is comprised of 3240 emails collected by J. Nazario [19]. Analyzing process shows that the folders' names that are extracted from legitimate URLs are extremely shorter than the folders' names of phishing URLs. This length variance has motivated us to use this criterion as a feature to detect phishing URLs. The analyzing results show that most of the URLs in phishing emails use very long folders' names for their paths; many of them use up to 230 characters. For the purpose of simple analysis observation and counting process, the length of folders' names is divided to 23 sub-ranges of 10 characters as shown in Table 2. More clearly, the sub-range 1 in Table 2 is to count the legitimate and/or phishing instances that not exceed 10 characters for their folder length, whereas the sub-range 2 is to count the legitimate and/or phishing instances that not exceed 20 characters for their folder length.

The performance of FldrNmLnth feature is evaluated based on its contribution in detecting phishing URLs. The results from our previous experiment in [20] are used here to highlight the performance of this feature. In [20], the FldrNmLnth was used along with a set of other URL-based features to participate in detecting phishing URLs. Although that this feature was utilized in our previous work, however, its performance details and experimental results are first time presented in this paper. The experiment was set up to check all extracted URLs against this feature. The free version of EditPad Pro text editor [21] was used to extract all URLs from involved legitimate and phishing emails. EditPad Pro supports regular expressions, or Regexps for short. Regexps is a sequence of special characters that define a search pattern of string searching algorithms, and this technique can simplify many text processing tasks such as performing "find" or "find and replace" operations on analyzed strings—emails' contents in this work. Table 1 shows the Regexps that used to extract all URLs from involved legitimate and phishing emails.

Table 1. Regexps that used to extract URL' folders names

No.	Regexp
1	(http://|https://|www[.])((. + ?[“‘ > <??])|(. + ([.][a–z]{2,5})$))
2	(http://|https://|www). + (([.]gif)|([.]jpg)|([.]jpeg))
3	(http://|https://|www). + ?([?])
4	(http://|https://|www). + (? = [?])
5	(http://|https://|www). + (([.])([a–z]{2,5})$)
6	(http://|https://|www[.]).*?([/])
7	[^a–zA–Z0–9/]{1,7}$

Only URLs that match the general URL structure as shown in Fig. 1 were extracted, in other words, only the URLs with protocol, domain name, and path are involved in this experiment. All URLs that link to image sources are excluded from evaluation experiment because phishers usually "borrow" such URLs from the original---being spoofed---sites to trick the victims by showing them such images [22–24]. The following rule is then used to implement the FldrNmLnth feature;



Results in Table 2 show that the length of some folders' names that extracted from phishing emails was up to 225 characters; on the other side, the longest folder name of URLs that is extracted from legitimate emails has not exceeded 30 characters. This extreme length variation motivated us to use the FldrNmLnth as a feature to detect phishing URLs. Table 2 shows that almost the same number of URLs with short folders' names that not exceeded 10 characters was extracted from legitimate and phishing emails, 14,433 and 14,384, respectively. The results also show that a small number of folders' names that range from 11 to 20 characters were extracted from legitimate URLs compared with the higher number of folders' names in the same range that extracted from phishing URLs, 881 and 3439, respectively. For the folders' names that range from 21 to 30 characters, we can compare the 28 with 382.

From the sub-range 4 onward, it can be seen that no legitimate URLs with folders' names that longer than 30 characters. At the last sub-range it can be seen that no legitimate URLs were constructed using folder names that ranges from 221 to 230 characters, while a number of 5 URLs were captured. This makes evidence that the phishers usually tend to use very long folders' names when they construct phishing URLs. Results from our previous work are used here to calculate the True Positive and False Positive rates of the FldrNmLnth feature to highlight its impact among other URL-based features' set as in our previous in [20]. The TP and FP are calculated using the Eqs. (1) and (2).

$$\text{TP rate} = \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

$$\text{FP rate} = \frac{\text{FP}}{\text{N}} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{2}$$

whereas:

- TP—True Positive which stands for the phishing instances that correctly identified as phishing.
- TN—True Negative which stands for the legitimate instances that correctly identified as legitimate.

Table 2. Number of folders in each sub-range of folder name length

Sub-ranges of FldrNmLnth		Number of folders in legitimate URLs	Number of folders in phishing URLs
No.	Length in characters		
1	1–10	14,433	14,384
2	11–20	881	3439
3	21–30	28	382
4	31–40	0	62
5	41–50	0	52
6	51–60	0	6
7	61–70	0	10
8	71–80	0	9
9	81–90	0	5
10	91–100	0	56
11	101–110	0	92
12	111–120	0	3
13	121–130	0	4
14	131–140	0	0
15	141–150	0	45
16	151–160	0	0
17	161–170	0	0
18	171–180	0	8
19	181–190	0	0
20	191–200	0	0
21	201–210	0	1
22	211–220	0	10
23	221–230	0	5

- FP—False Positive which stands for the legitimate instances that incorrectly identified as phishing.
- FN—False Negative which stands for the phishing instances that incorrectly identified as legitimate.

Based on that, the TP rate $= \frac{247}{247+2993} = \frac{247}{3240} = 0.08 = 8\%$, whereas the 247s represent the number of phishing emails that correctly identified as phishing by the FldrNmLnth feature, and the 3240 represents the number of emails that comprise the analyzed phishing email dataset.

The FP rate $= \frac{27}{27+9973} = \frac{27}{10000} = 0.003 = 0.3\%$, whereas the 27 represents the number of legitimate emails that incorrectly identified as phishing by the FldrNmLnth

feature, and the 10,000 represents the number of emails that comprise the analyzed legitimate email dataset.

4 Conclusion

The FldrNmLnth feature is introduced in this paper to detect phishing URLs that use very long folders' names in their paths. Experimental results show that, a considerable number of phishing URLs were constructed using folders' names that are longer than 30 characters, many of them have been using up to 230 characters. Phishers usually tend to use long URLs to hide the actual going-to destination since the address bar of Internet browsers is limited to show a finite number of characters. In terms of TP and FP rates, the implementation of the FldrNmLnth feature has achieved 8% TP rate and 0.3% FP rate when it was utilized to work with other URL-based features as in our previous experiment in [20]. For the future work, the performance of this feature can be further evaluated by using URLs from different legitimate and phishing datasets. The achieved results from different experiments can be compared with our results to improve the performance of this feature.

Acknowledgements. This work has been supported by Intelligent Spam Detection Model Based on Immunological Memory of Adaptive Immune System (USIM/FRGS/FST/50116-50).

References

1. Anti-Phishing Working Group: Phishing Activity Trends Report Q4 2016 (2016)
2. Nguyen, L.A.T., To, B.L., Nguyen, H.K., Nguyen, M.H.: A novel approach for phishing detection using URL-based heuristic. In: 2014 International Conference on Computing, Management and Telecommunications (ComManTel 2014), pp. 298–303 (2014)
3. Marchal, S., Armano, G., Grondahl, T., Saari, K., Singh, N., Asokan, N.: Off-the-hook: An efficient and usable client-side phishing prevention application. IEEE Trans. Comput. **66**(10), 1717–1733 (2017)
4. Zhang, J., Wang, Y.: A real-time automatic detection of phishing URLs. In: Proceedings of 2nd International Conference on Computer Science and Network Technology, *ICCSNT 2012* (2012)
5. Whittaker, C., Ryner, B., Nazif, M.: Large-scale automatic classification of phishing pages. In *NDSS Symposium 2010* (2010)
6. Suriya, R., Saravanan, K., Thangavelu, A.: An integrated approach to detect phishing mail attacks. In: Proceedings of the 2nd International Conference on Security of Information and Networks—SIN '09, p. 193 (2009)
7. Wang, Z., Peng, X., Wang, Y., Luo, Y., Huang, L., Cai, B.: Analysis on the characteristics of URL spam. In: Proceedings—2012 International Conference on Computer Science and Electronics Engineering, ICCSEE 2012, vol. 1, pp. 246–249 (2012)
8. Ma, J., Saul, L.K., Savage, S., Voelker, G.M.: Beyond blacklists : learning to detect malicious web sites from suspicious URLs. In: Kdd'09, pp. 1245–1253 (2009)
9. Basnet, R., Mukkamala, S., Sung, A.H.: Detection of phishing attacks: a machine learning approach. Stud. Fuzziness Soft Comput. (2008)

10. Huang, D., Xu, K., Pei, J.: Malicious URL detection by dynamically mining patterns without pre-defined elements. World Wide Web **17**(6), 1375–1394 (2013)
11. Lee, C.H., Kim, D.H., Lee, J.L.: Heuristic based Approach for phishing site detection using URL features. In: *Third International Conference on Advances in Computing, Electronics and Electrical Technology—CEET 2015* (2015)
12. Basnet, R.B., Sung, A.H.. Liu, Q.: Learning to detect phishing URLs. Int. J. Res. Eng. Technol. **03**(06), 11–24 (2014)
13. McGrath, D.K., Gupta, M.: Behind phishing: an examination of phisher modi operandi. In: LEET'08 Proceedings of the 1st Usenix Workshop on Large-Scale Exploits and Emergent Threats (2008)
14. Prakash, P., Kumar, M., Rao Kompella, R., Gupta, M.: PhishNet: predictive blacklisting to detect phishing attacks. In: Proceedings—IEEE INFOCOM (2010)
15. Toolan, F., Carthy, J.: Feature selection for spam and phishing detection. In: General Members Meeting and eCrime Researchers Summit, eCrime 2010 (2010)
16. Shetty, J., Adibi, J.: The Enron email dataset database schema and brief statistical report. Distribution (2004)
17. Cormack, G., Lynam, T.: TREC 2005 spam track overview. In: Sixteenth Text Retrieval Conference (TREC 2007), no. Trec. pp. 1–9 (2005)
18. Spamassassin: Public corpus. http://spamassassin.apache.org/publiccorpus (2006)
19. Nazario, J.: Phishing corpus, p. 0 (2007)
20. Al-Daeef, M.M., Basir, N., Saudi, M.M.: Evaluation of phishing email classification features: reliability ratio measure. In: Lecture Notes in Engineering and Computer Science, vol. 2229 (2017)
21. Goyvaerts, J.: EditPad Pro. Just Great Software Co. Ltd. (2011)
22. Ludl, C., McAllister, S., Kirda, E., Kruegel, C.: On the Effectiveness of Techniques to Detect Phishing Sites, pp. 20–39 (2007)
23. Bergholz, A., De Beer, J., Glahn, S., Moens, M.F., Paaß, G., Strobel, S.: New filtering approaches for phishing email. J. Comput. Secur. **18**(1), 7–35 (2010)
24. Waziri, I.: Website forgery: understanding phishing attacks and nontechnical countermeasures. In: IEEE 2nd International conference on cyber security and cloud computing. CSCloud 2015—IEEE International Symposium of Smart Cloud, pp. 445–450 (2016)

Future of Governance for Digital Platform Ecosystems

Meelis Kitsing[1(✉)] and Johanna Vallistu[2]

[1] Estonian Business School and Foresight Centre, Tallinn, Estonia
meelis.kitsing@ebs.ee
[2] TalTech and Foresight Centre, Tallinn, Estonia

Abstract. The future of governance and regulation of digital platforms is uncertain. The development of digital platforms depends on a number of political and social factors in addition to economic and technological drivers. Particularly, the role of governance and institutions are crucial for structuring platform ecosystems. Therefore, this paper explores alternative futures of digital platform ecosystem governance through the scenario planning approach. This approach allows considering alternative future trajectories rather than rely on extrapolation of current trends. The paper discusses different scenarios developed by international and national organizations. Three alternative futures for the governance of digital platform ecosystems emerge resulting from scenario matching: private platform ecosystems, government platform ecosystems, and decentralized platform ecosystems. The conclusion highlights the key implications of different meta-scenarios.

Keywords: Digital platforms · Scenario planning · Governance

1 Introduction

The digital platforms are increasingly structuring our economic, political, and social life. It is not only that some large digital platforms have organized many economic sectors such as shopping and entertainment but they have wide-reaching macroeconomic impacts as well as the ability to challenge the essence of democratic politics.

Digital platforms have been around in one form or another since the mid-1990s. Engineering literature on platforms dates back to the 1980s. Economics and management scholars started to tackle platforms in the late 1990s and in the 2000s. Scholars have explored various fundamental aspects of platforms such as lock-ins, network effects, winner take all nature, two-sided markets, and multisided platform markets as well as long-tail markets [1–3]. However, overtime the importance of digital platforms has grown. This has led also to additional emphasis in academic literature on the culture connectivity, boundary resources, and platform-dependent entrepreneurship [4]. Digital

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_33

platforms are not just facilitators in business transactions but increasingly both participants, gate-keepers, and rule-makers in two- or multisided markets. This exploitation of information asymmetries and reliance on conflicts of interests creates opportunities for abuse, which may not only undermine specific market-places but trust in capitalism and democratic governance. At the same time, digital platforms have become so embedded in our daily life that it is difficult to go on with daily activities or even imagine life without them. This is so particularly in the case of large digital platforms such as Google, Amazon, Facebook, Apple, Tencent, Baidu, and Alibaba. Hence, the paper uses the concept of digital platform ecosystems that have far-reaching economic, political, cultural, and social consequences [5]. The use of ecosystems allows avoiding economic determinism where "economy" determines developments in other areas. Following rich literature operate in a vacuum but are interdependent with politics, culture, and society [6]. There is a rigorous debate on how this interaction is structured. However, it does not have to be elaborated here. Most importantly, the economy cannot be decoupled of politics, culture, and social issues. Particularly so as platform businesses have a direct impact on and politics has an impact on platform businesses.

Furthermore, digital platforms face uncertain future—particularly in the long run. It is not clear at all what kind of platform ecosystem model will dominate in the future or whether there will be one universal single model. Following literature in comparative political economy, it is likely that different approaches will emerge as a result of interaction of politics, economics, and social issues. Hence, the governance and institutional frameworks are fundamental for structuring the platform ecosystems. These ecosystems are not technologically pre-determined.

This paper will explore alternative futures for digital platform ecosystems by relying on scenario planning instead of forecasting and prediction. Instead of emphasizing one prediction or forecast on the basis of previous developments and current trends, it would be wise to think about it in terms of alternative scenarios. These alternative scenarios allow us to break linear logic and by asking "what…if" questions expand the range of future alternatives [7, 8].

Particularly so as the future of digital platforms is certainly characterized by high complexity and high uncertainty. Our available knowledge about different factors impacting the future is limited. At the same time, variables to be considered are large. In economics and social science scholarship, at least four different perspectives emerge about technology and its social consequences [6]. Hence, many foresight teams have created alternative scenarios, which among other developments tackle the potential futures of digital platforms. The next section will discuss the scenarios of various international and national organizations by highlighting common denominators and differences. The paper concludes by highlighting implications.

2 Alternative Scenarios for Digital Platform Ecosystems

A number of foresight teams at the international and national organizations have developed alternative scenarios which in one way or another tackle the future of digital platform ecosystems. We will discuss recent scenarios developed by the Organisation for Economic Cooperation and Development (OECD) on digital transformation, government scenarios of Joint Research Centre of European Commission and Institute of

Policy Studies at National University of Singapore (IPS), future of work scenarios by the World Economic Forum (WEF) as well as future of work and governance scenarios of the Foresight Centre (FC) at the Estonian Parliament (in the case of latter two, both authors of this paper have been the key contributors to the scenarios). All these scenarios are sufficiently generic and can be applied in a different economic, political, and social context. However, as scenarios are ideal types, then some of them seem certainly more utopian than others in a specific context. Our current understanding may indicate that some of these scenarios are more or less likely depending on the specific economic, social, and political contexts. However, these scenarios do allow breaking-up linear logic in thinking about the future and widening the view of potential futures of the ecosystems.

The OECD digital transformation scenarios address the future of digital platform ecosystems most directly [9]. Their scenario "Corporate Connectors" is probably one of the most realistic from today's perspective as it foresees the increasing dominance of large private digital platforms. The scenario "Platform Governments" foresees increasing importance of government or government-supported platforms, which is more likely in some parts of the world than others. The scenario "iChoose" emphasizes the importance of privacy and individual rights to data control. However, the least likely scenario is "Artificial Invisible Hands" which represents radical decentralization of governance where nobody controls the data.

Four scenarios on the future government published by the European Commission's Joint Research Centre (JRC) have placed an important emphasis on digital platforms [10]. These scenarios to some extent overlap with the OECD's scenarios on digital transformation as digitalization is a fundamental factor. Their scenario "DIY Democracy" entails limited availability of public services, which is substituted by strong co-creation of services by citizens. Bottom-up digital platforms facilitate grassroots initiatives but offline engagement at a local level remains important as well. Their scenario "Private Algocracy" is characterized by dominance of large private digital companies where citizens' interests are derived from their data profile. The scenario "Super Collaborative Government" combines the rise of artificial intelligence (AI) with a citizen-centric government. Citizens can engage seamlessly in decision-making through digital platforms. The scenario "Over-Regulatocracy" visions nationalization of leading digital platforms under democratic governments. However, citizens have difficulties obtaining rights and accessing good services because of bureaucratic overreach. Technology-centricity in the JRC scenarios is certainly a limitation as institutional constraints and enablers are not fully explored. As was emphasized in the first section of this paper, digital platform ecosystems do not operate in a vacuum and they interact with economic, political, and social context. These scenarios certainly stem from a technology optimistic perspective and some of them sound utopian. Even if some OECD and JRC scenarios may seem from our current perspective utopian or dystopian, we cannot dismiss them as impossible because the future remains uncertain. Obviously, utopias not only exist in scenarios but can be found everywhere (for instance, in government strategy documents) as they often serve as mental short-cuts for decision-makers.

At the Foresight Centre, we created five public sector governance scenarios for Estonia which also paint different pictures of the role of digital platform ecosystems [11]. The scenario "Ad Hoc Governance" sees rapid digitalization in some areas as a

priority while other areas are left behind because of government budget constraints. This implies that digital platform ecosystems suffer from inconsistent developments. The scenario "Night-watchman State" is concerned about privacy and security concerns of excessive government digitalization while trying to use a standardized approach for efficiency gains. This scenario is favorable for global private platforms. The scenario "Entrepreneurial State" is about building up a highly digitalized state for entrepreneurial purposes and enhancing digital government platforms globally by relying heavily on public-private collaboration. The scenario "Caretaker State" is about the massive use of digital technologies in preventing the spread of social ills and intervention for the benefit of citizens' well-being. The government does not have global ambitions and aims to develop sophisticated platforms for serving citizens domestically. The scenario "Networked Governance" pictures decentralized public sector governance with high degree of co-creation and use of digital platforms by different actors. As a result, platform ecosystems are diverse and uneven where governmental, business, civil society, and local community platforms operate. These five Estonian governance scenarios have many common elements with three scenarios developed for Singapore by the Institute for Policy Studies (IPS) at National University of Singapore. Even though these scenarios do not specifically discuss digital platforms, it is possible to derive implications for digital platform ecosystems. Their scenarios "SingaStore.com" emphasizes the importance of economic development and private businesses at the expense of social cohesion. It would imply the dominance of private platforms as in the Estonian scenario "Night-watchman State"—even though Singaporean scenarios see a somewhat more important role for the government [12]. The scenario "SingaCity.gov" emphasizes the role of government in promoting egalitarian social values and human development (like Estonian scenarios "Caretaker State). This implies the development of dominant government platforms, which structure the ecosystem. The scenario "WikiCity.org" is essentially a decentralized governance scenario where no central authority dominates—it is a self-activating and self-correcting community (like Estonian scenario "Networked Governance"). This implies diverse and pluralist platform ecosystem with many different players from businesses and communities.

2.1 Future of Work Scenarios

It is worthwhile to look at two sets of future of work scenarios, which elaborate on the issues of digital platform ecosystems from the perspectives of global work platforms. World Economic Forum's (WEF) eight scenarios on the future of work offer a take on the interaction of technological change, the learning evolution, and the talent mobility in the context of long-term employment patterns in the world [13].

We analyze the scenarios from the prism of the adaptation with the technological change, which can be either steady or accelerated. The scenarios where technological change and diffusion accelerate significantly are the ones where skills and mobility determine whether opportunities posed by technological development are captured. However, these scenarios also entail the risk of increasing polarization.

"Robot Replacement" is a world with low talent mobility and slow learning evolution. The outcome is the hollowing out of the labor markets. This results in protective

governments, tight border control, and a looming threat of social disruptions. In a "Polarized World" the learning evolution is still low but there is a high level of talent mobility. As a result, the highly skilled emigrate from regions with slow development and enclaves of global super economies form. Both scenarios suggest development trajectories toward government platforms. In a scenario called "Productive Locals" the adjustment to fast technological development has been matched with life-long learning and vast retraining programmes. Talent mobility, however, is low to make sure that the investment in human capital makes sense regionally. Online labor markets on digital platforms make up for talent shortage in some regions but this is not sufficient to ensure the skills' externalities and the spread of innovative ideas.

"Agile Adapters" combines accelerated technological change with fast learning evolution yielding a hyper-agile world where people combine physical and virtual mobility. There is widespread use of private online digital platforms for work. In the worlds with steady technological change and diffusion the growth potential is smaller but is mostly more inclusive towards a larger group of people. "Workforce Autarkies" combine steady technological change with low learning evolution and low mobility results in losing the competitiveness of local labor markets due to talent shortage.

In "Mass Movement" the learning evolution is slow and regions mostly experience slow growth making their talent looking for alternative options. The online work platforms are a marginal phenomenon in the scenario resulting in high migration flows of talent to where the highest growth and most innovative solutions are offered. "Empowered Entrepreneurs" is a world where amid slow adoption of technology the investment in human capital has been high and thus governments are limiting mobility not to lose their talent. Workers are empowered to create opportunities for themselves and often become entrepreneurs, often selling their services globally using online platforms. Both learning evolution and mobility are high in the scenario of "Skilled Flows". Online platform work is still a marginal phenomenon in the scenario. This results in high inequality between countries and regions as some have better access to technology and thus migrating talent than others.

Furthermore, four future of work scenarios were developed by the Foresight Centre [14]. The scenarios envision the impact of technological change and attitudes towards mobility as the main shapers of Estonian employment market outcomes in the future. Two of the future work scenarios describe worlds where technological development enables to create highly skilled jobs offsetting the effect of automation in low-to-mid-skill tasks and offering potential for long-term economic growth.

In the case of scenario "Talent Hub Tallinn" the potential is realized thanks to having access to global pool of talent who have the right skills to employ technology. Online digital platforms are seen here as an additional way of matching supply and demand on a global level and enabling agility in fast-changing world. This is supported by the blurring line between employment and entrepreneurship—an increasing number of workers are own-account workers providing their services on a global level.

In "Self-Reliant Estonia" the potential of technological development could not be used to its full extent as mobility is restricted and there is a lack of highly skilled people to innovate and use technology. Online labor platforms can offer some alleviation here for the companies suffering from talent shortage but may become a subject of

restrictions to protect the local labor market. An emergence of government or regional work platforms would be likely in such a world to ensure that the local labor market is not threatened by virtual outsiders. In worlds where automation effects prevail and less new jobs are created, the outcomes of high and low mobility scenarios differ vastly when considering the platform ecosystem effects. In case of high mobility, the automation is initially postponed but eventually inevitable, as happens in the "Global Village of Nomads". The role of digital platforms for finding work is a minor one as does not offer sufficient livelihood opportunities for low-skilled people and they prefer migrating instead. When migration is restricted as in the scenario "New World of Work" the online global work could be a substitute income to people who cannot find work in the local labor market. Given the increasing level of unemployment and protective attitudes, a restrictive approach towards private digital platforms is likely. Governments may create platform cooperatives or other alternative models to tackle this threat.

2.2 Meta-Scenarios

Even though Estonian and Singaporean scenarios were developed for the national context they do offer some universally applicable trade-offs concerning the future of platform ecosystems. On the basis of national scenarios and scenarios developed by international organizations three meta-scenarios for the future of platform ecosystems emerge. Particularly as Estonian and Singaporean scenarios also overlap with the OECD and JRC scenarios to some extent—even though they are less technology-centric and emphasize institutional factors as key drivers. In the WEF and Estonian future of work scenarios digital platform ecosystems play an important role in some but not in others. Since WEF created eight scenarios then only four scenarios with accelerated technological change and significant role for digital platforms are included in the comparison. Four scenarios with marginal role for digital platform ecosystems are excluded. As far as governance scenarios are concerned then the Estonian scenario "Entrepreneurial State" and Singaporean "SingaGives.gov" has common characteristics with OECD's "Platform Governments". The JRC's scenario "Over-Regulatocracy" has many elements in common with the Estonian scenario "Caretaker State". The dominance of private digital platforms as envisioned in the OECD's "Corporate Connectors" and in the JRC's "Private Algocracy" is most likely in the Estonian scenario "Night-watchman State" and Singaporean "SingaStore.com". Estonian scenario "Networked Governance" and Singaporean "WikiCity.gov" have many elements in common with the JRC's "Super Collaborative Government" and the OECD's "Artificial Invisible Hands".

The key difference is that both OECD and JRC scenarios place much greater emphasis on digitalization than the Estonian and Singaporean scenarios. It is completely understandable in the case of OECD scenarios because these are digital transformation scenarios, which also discuss some aspects of governance. However, the JRC's scenarios are government scenarios where technology-centricity may narrow down a range of possible and plausible alternatives for digital platform ecosystems. Particularly so because institutional constraints are not likely to be broken down in 10–15 years. In this sense, the Estonian scenarios are more general and offer a wider range of alternatives. For instance, both JRC's "DIY Democracy" and "Super Collaborative Government" could be subscenarios of "Networked Governance". The latter would work under generous budget

Table 1. Three meta-scenarios for digital platform ecosystems

Meta-scenario	Characteristics and implications	Matching scenarios
Private platforms	Dominance of large private platforms globally, particularly from China and the US. Global private gated communities and minimal role for government. Small states have weak bargaining power in private rule-making	Corporate Connectors (OECD), Private Algocracy (JRC), SingaStore.com (IPS), Night-watchman State and Ad Hoc Governance (FC), Agile Adapters (WEF), Talent Hub Tallinn (FC), Global Village of Nomads (FC)
Government platforms	Dominance of government platforms, spliternet globally. Global openness, economies of scale and network effects are reduced. Larger countries have leverage over smaller in regional ecosystems	Platform Governments (OECD), Over-regulatocracy (JRC), Entrepreneurial State (FC), SingaGives.gov (IPS) Caretaker State (FC), Robot Replacement (WEF), Polarized World (WEF), Self-Reliant Estonia (FC)
Decentralized platforms	Diversity of platform ecosystems, divergence in regulatory approaches. Multiple private, public and community solutions are available. System robustness emerges from decentralization. Different capacities and capabilities lead to inequalities and risks	iChoose and Artificial Invisible Hands (OECD), DIY Democracy, Super Collaborative Government (JRC), Productive Locals (WEF), WikiCity.sg (IPS) Networked Governance (FC), New World of Work (FC)

constraints while the former in the case of tight budget constraint. Most importantly, the combination of these different scenarios allow mapping out three meta-scenarios for digital platform ecosystems. The meta-scenarios are the following: (1) Private platform ecosystems (2) Government platform ecosystems (3) Decentralized platform ecosystems. Table 1 summarizes the key elements of each scenario and indicates which scenarios from OECD, JRC, IPS, WEF, and FC correspond to this classification. The future of work scenarios offers a narrower story on the development of digital platforms but in their essence supports the idea of platform ecosystems. Some of these scenarios assume a linear development of global private platform ecosystems. The ideas of government or decentralized platform ecosystems are not straightforward in the scenarios but can be likely results of interpreting the scenarios. WEF's scenarios are a good example of how the institutional and social factors are accounted for—in addition to technological development when placing digital platforms into a wider scenario context. The Foresight Centre's future of work scenarios sees online work as part of a techno-optimistic globalized world as well as a substitute for real mobility options. In the latter case, however, the shift from private global platforms could take place towards government platforms

or decentralized local platforms as governments try to protect their labor markets and workers.

3 Conclusion

In order to grasp the implications of digital platforms, it is important to deepen our understanding of digital platforms and ways of approaching the future. We explored alternative futures for digital platform ecosystems on the basis of scenario planning instead of extrapolation of current trends on the basis of forecasting or prediction. This allowed us to develop three meta-scenarios for the future of digital platform ecosystems emphasizing the importance of governance. Governance plays a fundamental role in structuring platform ecosystems. These three meta-scenarios imply trade-offs for countries. The purpose of the scenario planning approach is not to offer concrete policy suggestions but rather to indicate potential future developments for policy-makers, which allows formulating a framework for policy responses. A robust policy suggestion is that policy-makers have to be prepared for alternative scenarios and radical changes rather than rely on one vision or strategy for thinking about the future of digital platforms.

Acknowledgements. Authors are thankful to the generous support provided for this project by the Foresight Centre at Estonian Parliament, ASTRA TalTech Development Program, and Doctoral School in Economics and Innovation (project code no. 2016-2020.4.01.16-0032).

References

1. Noe, T.H., Parker, G.G.: Winner take all: competition, strategy and structure in the internet economy. J. Econ. Manag. Strateg. **14**, 1 (2005)
2. Brynjolfsson, E., Hu, Y., Smith, M.D.: From niches to riches: the anatomy of the long tail. MIT Sloan Manage. Rev. **47**, 4 (2006)
3. Zysman, J., Kenney, M.: The rise of platform economy. Issues Sci. Technol. **32**, 3 (2016)
4. Cutolo, D., Kenney, M.: The Emergence of Platform-dependent Entrepreneurs: Power Asymmetries, Risks and Uncertainty. SSRN April 15, 2019
5. Jacobides, M.G., Cennamo, C., Gawer, A.: Towards a theory of ecosystems. Strateg. Manag. J. **39**, 8 (2018)
6. Iversen, T., Soskice, D.: Democracy and Prosperity. Reinventing Capitalism Through A Turbulent Century. Princeton University Press, Princeton (2019)
7. Ramirez, R., Wilkinson, A.: Strategic Reframing: Oxford Scenario Planning Approach. Oxford University Press, Oxford (2016)
8. Schoemaker, P.J.: Forecasting and scenario planning: the challenges of uncertainty and complexity. In: Koehler, D.J., Harvey, N. (eds.) Blackwell Handbook of Judgement and Decision Making. Blackwell Publishing, Oxford (2004)
9. OECD: OECD Scenarios for Digital Transformation. OECD, Paris (2018)
10. Vesnic-Alujevic, J. et al: Government 2030+. European Union, Luxembourg (2019
11. Keskus, Arenguseire: Valitsemine 2030. Arenguseire Keskus, Tallinn (2018)
12. Institute for Policy Studies: IPS Prism Report. NUS, Singapore (2012)
13. World Economic Forum: Eight Futures of Work. World Economic Forum, Geneva (2018)
14. Arenguseire Keskus: Tööturg 2035. Arenguseire Keskus, Tallinn (2018)

Azimuth Tree-Based Self-organizing Protocol for Internet of Things

Veena Anand[3(✉)], Prerana Agrawal[3], Pothuri Surendra Varma[3], Sudhakar Pandey[1], and Siddhant Kumar[2]

[1] National Institute of Technology Raipur, Raipur, CG, India
[2] Delhi Technological University, New Delhi, India
[3] Department of Computer Science and Engineering, National Institute of Technology Raipur, Raipur, CG 492010, India
vanand.cs@nitrr.ac.in

Abstract. An azimuth tree-based network algorithm is proposed on a uniform random distribution of scattered points which is actually graph-based. In the proposed azimuth on a tree-based network, first tree is established through the algorithm for finding the weight function of the tree would constrain the number of points to be considered during azimuth routing, thereby limiting the search space, reducing time complexities, and inducing further optimizations. The idea behind our paper would be applying the azimuth algorithm on a selected number of points, which would be achieved through the efficient tree-based routing protocol. So, firstly, we would apply our tree-based protocol on the uniform random distribution of points. The limited number of points output by this algorithm would be fed as input to azimuth routing. And in the protocol proposed by us, the aggregation of data in this tree-based network helps in reducing the network load and the energy consumption. And this research work mainly revolves around the criteria to be taken into consideration for balancing factors like these and construct a tree-based network which is better in terms of both lifetime of the network and the successful routing of protocols. With the help of simulation implemented using C++ and then constructing its corresponding graph, we have shown how considering all the factors possible has improved the performance of tree-based network. And finally, a comparative analysis is done where our proposed model is compared to the already existing traditional routing protocols namely AODV, DSDV, LEACH, Azimuth-based algorithm.

Keywords: Wireless sensor network · Internet of things · Azimuthal routing protocol · Network lifetime · Tree-based network · Successful routing ratio

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_34

1 Introduction

Many current routing algorithms do not consider the dense architecture of IoTs while they are designed. Thus, it is a great summon in the field of Internet of Things to render such algorithms that are adaptive to continuously changing and varying requirements of IoTs (sensor-based) applications as the routing policies in these cases are usually predetermined and because of this reason, they are not energy efficient and lack fault tolerance.

The main motive behind this paper is to implement a routing protocol for IoT based networks which are energy-efficient and also have prolonged lifetime with the main aim of successfully routing the packets. Also, to implement a routing procedure that ensures successful routing even when the number of nodes is very high. Our aim is to derive a routing protocol that ensures successful routing from node to sink while making sure that it is also energy efficient. The protocol that we have implemented considers the following parameters or metrics for making a particular node as intermediate sink node: (1) Distance between the nodes, (2) Number of hops or average hop of network, (3) Number of child nodes.

While during the self-organizing portion of the algorithm following parameters or metrics have to be considered: (1) Distance between the nodes (2) Average hops of the network (3) Residual energy of the network (4) Number of child nodes.

The detailed discussion upon these criteria and the parameters used are done in the upcoming sections.

2 Related Work

The sensor networks comprise of many sensor nodes that are less in cost and less in power that are randomly distributed throughout the area under observation. These nodes communicate with one another in order to gather information and forward the data. With ever-increasing need of the IoTs and their scale increasing these networks needs to be updated; this, on the other hand, results in a more complex network of systems. But simultaneously there is a limit to the memory, energy associated, and computing ability of these network nodes [1–6]. And thus, to attain the maximum network lifetime lots of research have been done which are aimed at controlling the network topologies and building a better data transmission path and simultaneously making sure that the energy consumption of nodes is balanced. And after a detailed study [1, 7], it was found that a major point of consideration now has become as to how to construct a tree-based network that has the highest life. But the complexity of choosing a real tree with the highest lifetime is an NP Complete problem. So we selected to consider a sub-optimal sensor network, so as to meet the needs of the real-time network. A network dependent on tree is not possible to be built in polynomial time. Thus, a spanning tree was constructed using subset division that required polynomial time, even in the worst-case scenario this protocol performed in exponential time. WSTDO [8], a data transmission distributed technology that is dependent on the concept of spanning tree, its performance is dependent on the density of sensor nodes. And in the case of sparse networks, this achieves comparatively better performance. LBT is helpful in maximizing network life.

One-hop node's energy efficiency and load balancing are taken into consideration for construction of tree-based network. This algorithm can actually reduce the consumption of energy in the network. Data aggregation technology is not actually used in the case of the above two protocols. So, when data aggregation is applied, the methods result in increasing the energy consumption and load on the network which is undesirable.

3 Proposed Algorithm

The assumptions that are made while developing the protocol are: Unit disk graph is used to characterize the connectivity between the sensor nodes [9]. It is also assumed that power of transmission of all the nodes is equal and the nodes have an equal transmission range that is circular by nature which is explained by [10]. Because of such assumptions, the complete topology can be scalable down or up to set the transmission range equals to 1 (which is actually done by the division of all possible distances by the of the node's range of transmission), independent of the path loss and transmission power. The network should be connected. At least one path exists between the sink and each sensor node. If in the case where there is more than one sink, then it is assumed that the sensor nodes know the nearest sink position. This ensures that the packet does not keep on circulating indefinitely about the closest sink near to the origin node for which in-fact no path might actually exist. The three states on which the azimuthal protocol working depends are pure azimuth, greedy/azimuth, and pure greedy. In each state, the best forwarder node is selected based on some criteria which will be explained in the subsections. The extra data which is present in the packet header is used to make the choice of the forwarder. Packet contains the following data:

- Positions of the last two nodes, that is the two nodes prior to the actual transmitting node that needs to hold the packet.
- Least distance value: It is the least distance to reach the sink node that has ever been possible and reached during the process of forwarding that particular data packet.
- Direction bit: This bit represents the forwarding direction, either clockwise or counterclockwise.
- Change bit: Whether there exists any modification in the forwarding direction of the packet or not is represented by this bit.

Azimuth might sometimes fail to deliver packets to the destination and is very rare, i.e., failure rate is only about 6%. In the case of low-density networks, the average of number of neighbours are only 1–3, the topology is rather simple and the shortest path to the sink node is the only path possible, so making it easier to reach the destination while the density of the network is more. Therefore, the mid-density network performance is very crucial.

Although the implementation of the azimuth protocol involves great importance as far as a network are concerned but there were certain problems as well. Following are the limitations of the azimuth routing algorithm:

- This routing algorithm may be applied only to devices having a GPS sensor to get its location, which is costly. Thus, not all devices may employ this algorithm.
- Azimuthal routing may not always succeed. If routing fails, the source should be informed about the failure. So, a mechanism is to be implemented to notify the source of any failure.
- The positions of the one-hop neighbours also need periodic updation, to prevent usage of stale incorrect values. This requires a constant beacon broadcast from each node, which in turn requires the usage of bandwidth for control data.

While going through various research papers we found that tree-based network, in general, have better performance in case probability of successful routing when compared to other traditional protocols like AODV, DSDV or LEACH while in case of network lifetime the protocol performs comparably equal to that of LEACH and far better than AODV and DSDV. So keeping in mind this result, we have implemented azimuth on a tree-based network.

First, we would apply our tree-based protocol on the uniform random distribution of points. The limited number of points output by this algorithm would be fed as input to azimuth routing. And thereafter, the algorithm proceeds as before [11].

As the name suggests, this algorithm basically functions by self-organizing the tree-based network. There are basically three main phenomenon taking place in the implemented protocol:

(1) Self-Organization of the Network (2) Dynamic adjustment of Topology (3) Evaluation of Network Performance.

(1) *Self-Organization of the Network*

During construction of the network, the network nodes search their child nodes by broadcasting the network packets. The non-network nodes which are nothing but the nodes which are not yet connected in the network select their intermediate sink node depending on the received broadcasted packets which contains the routing information. The picking of intermediate sink nodes depends upon the average number of hops, number of child nodes, residual energy, the distance between two nodes.

So in order to decide which node will become the intermediate sink node based on balancing the number of hops, residual energy, distance between the nodes and child nodes which is given by:

$$W_i = \frac{\alpha}{D_i} + \frac{\beta}{N_i + 1} + \lambda E_i + \frac{\delta}{H_i + 1} \tag{1}$$

Here, W_i is the node$_i$ weight, E_i is the node$_i$'s residual energy. Distance between the current node to the node$_i$ is given by D_i. The hop of root node is defined as 0 and the hop of other nodes is given by hop of intermediate sink node plus 1. H_i is defined as the node$_i$ hop. α, β, and δ are the normalized parameters for the above four factors.

The value of α is 15, if the maximum distance of transmission is 15 m. If the highest number of child nodes is taken as 5 then we set β as 6. The starting energy of the network node is set randomly, initial maximum energy is set to 20 J. Thus, gamma is set as 1/20. If the highest allowed hop is 10, then delta is 11.

Nodes of the sensor network are categorized into three main types: Sensor nodes, intermediate sink node, and root node. The node with the unlimited energy is classified as root node and is always active throughout the time. Root is the starting node in the network and sends the broadcasted packets to search the child nodes and convert those non-network nodes into the network nodes in the beginning. Then the broadcasted packets are saved by the non-network node and the weight for the sink node is calculated on the basis of the above equation. Finally, the non-network node with the highest weight is chosen as the best intermediate sink node and joins the network.

In order to balance its energy levels, an intermediate sink node might refuse the non-network node from joining the network or it may even remove its child node after some time period. In such cases, the non-network node selects the next best intermediate sink node to join the network.

(2) *Dynamic Adjustment of Topology*

Reconstruction of the network is done in the following two cases: (1) Energy Consumption (2) Link Failure.

Energy Consumption: The intermediate sink node along with gathering the data of its own sensor is also responsible for aggregating data of all of its child nodes and thus the consumption of energy for these nodes is more than that of normal sensor nodes. The node at the farthest distance from the node when its energy drops down lower than a threshold value of R% is deleted. R% depends upon the residual energy of the previously changed topology, which means the intermediate sink node can add or remove a child node. Now, removing the child node initiates the need of sending the packets to inform all the related child nodes to re-elect their intermediate sink node and simultaneously delete all the information of that child node from the child table. It is beneficial for balancing the energy consumption if the child node which is the farthest joins the network's remaining branches. To compute this parameter namely R, we need to be aware of the number of child nodes N of the network.

$$R = N/(N+1) \tag{2}$$

For example: Let us assume N as the number of child nodes at t_0 and residual energy as E_0. So the residual energy at t_1 is E_1 and it can be calculated using the above formula to obtain $E_1 = 5E_0/6$. Now for balancing the consumption of energy, the farthest child node is deleted so the number of child node decreases to 4, i.e., $N = 4$. So at time instant t_2, the residual energy is given as E_2 and it is calculated as $E_2 = 4E_1/5$ and we need to adjust the topology again.

As can be seen in Fig. 1, at the beginning the Node1 is intermediate sink node of Nodes 2, 5, and 6. After a certain amount of time due to continuous exchange of packets, the residual energy of Node1 decreases and it deletes the farthest child node depending upon the equation mentioned above. Here Node6 is assumed to be the farthest child node thus it is deleted and thus Node1 sends packets to inform Node6 for re-electing the intermediate sink node. Node1 has the largest weight and thus Node6 saves the states and remaining nodes remain as it is.

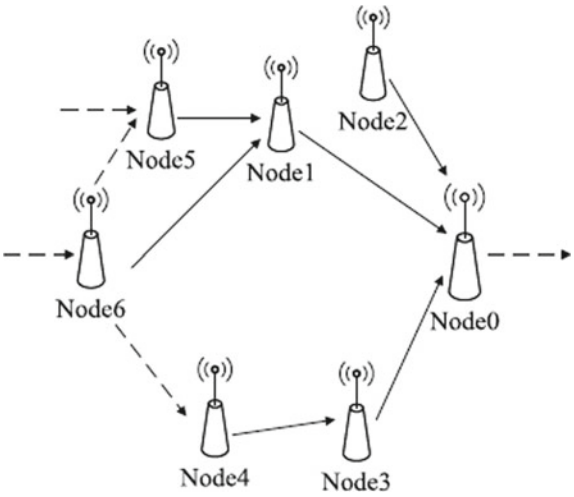


Fig. 1. A wireless sensor network topology

- Node 4 or 5 has the maximum weight. First the node6 transfers the packet to the Node1 requesting the node to leave the sensor network, and then the packets are sent to Node5 or Node4 to call them to join the network.
- The weights of Node5 and Node1 are similar or that of Nodes 1 and 4 are similar thus in this case need arises to compare the residual energy, number of hops, distance between the nodes, and the child nodes of the two nodes. The priority given are in the following manner: less number of hops, greater ratio of the (amount of residual energy)/(child nodes + 1)), less inter-node distance. So after considering the above quantities and their criteria we find the Node1 better. And since the Node1 is already the current intermediate sink node of the Node6 we need not change anything.
- The weights of Nodes 5 and 4 are being same, need arises to compare the number of hops, distance between the two nodes, and the child nodes of the two nodes, residual energy. Assumption is taken as to all the factors for the Node5 and Node4 is similar, so the need arises to check their corresponding sink nodes. Node1 is the intermediate sink node of Nodes 5 and 1 is the intermediate sink node of Node6. Node6 then re-elects intermediate sink node for balancing energy consumed by the Node1. On selecting.Node5 as the intermediate sink node for Node6 the length covered by the packet between Node1 and Node5 will increase.

Thus, leads to increase in energy consumption by Node1. However, we find that Node4 not being part of the branch of Node1 it is beneficial if Node6 connects in the network branch of Node4 as it would help in balancing the consumption of energy by Node1. Assumption is made that Node6 elects Node5 as its intermediate sink node. But after a period, Node1 can no longer work as an intermediate sink node. Node5 should search for its intermediate sink node but does not have any in its one-hop transmission range thus it broadcasts the packets to notify all its child nodes to re-elect intermediate sink node. During this process of reorganization of the network, all the child nodes need to modify their value of hops if the value of hop of that particular intermediate sink node changes.

Case of Link Failure: The child node transfers the data packet to the intermediate sink regularly and the sink transfers responses to corresponding child nodes regularly to make sure that the links are properly connected.

But in a certain period of time, if no packets are received by the node from its child then it thinks that the link is not successful and thus discards the child node from the table that has all the child node mentioned in it. And similarly, if the child node does not receive any response packet from its intermediate sink node for a particular period of time, it will think that the link between them has become unsuccessful and re-elects the intermediate sink node.

4 Evaluation of Network Performance

In a sensor network, since all the nodes are randomly aligned need arises in this case for the formation of a network as soon as possible so that self-organization can also take place as quickly as possible as self-organizing efficiency is of great importance [12–15]. With the increment in the number of hops, forwarding time also increases. Thus this proves our point that the nodes' energy and the average number of hops are quite important aspects for network performance evaluation.

B. *Algorithm1: Formation of tree-based network*

At the beginning before the network is formed, there are a number of inactive nodes and an active root node having infinite energy. To initiate the formation of the tree, this root node broadcasts messages to all nodes under its transmission range. Some of the nodes accept the root node root as their sink and respond correspondingly.

As they become a part of the network, they broadcast messages to their neighbours for becoming an intermediate sink node. All the nodes which are still not a part of the network, receive these messages from many possible intermediate sink nodes in a given period of time. The node chooses the best intermediate sink node. The node with the highest weight is selected as the intermediate sink node.

For a node that joins the network after the tree has already formed, it forwards a join request to all nodes in tree under the nodes transmission range and starts timer. The nodes of network receive the message and if they are willing to act as their sink, they give a positive reply. The node selects its intermediate sink node by calculating the weights of all nodes that gave a positive reply before the timer expired. Pseudocode for selection of best intermediate sink node for a given node:

```
Initialize index to 0.
Initialize maxW to 0.
Initialize sink to 0.
While index is less than number of available sink
  Calculate weight of the sink
  If weight of sink is greater than maxW
          Set sink to the index of current sink
          Set maxW to the weight calculated
          Set index to index + 1
Return sink
```

Explanation for above pseudo code

Steps 1 through 3 initialize the different algorithms as well as loop parameters such as indices, the max weight, and index of the sink. Steps 4–9 comprise of the iterative procedure, which is the main driver of the concerned algorithm. The iteration goes on until we reach the end of the available sinks. Step 5 calculates the weight of the sink in that particular iteration. If the calculated weight comes out to be greater than the maximum weight found till now, all the parameters are updated. Step 6 updates the sink, 7 updates the maximum weight, and 8 increments the loop index.

C. *Algorithm 2: Balancing energy consumption*

The residual energy of every node is checked in a fixed interval of time to balance the amount of energy consumed by the nodes. The balancing of energy will help the network live longer. The node always stores the energy that the node had when the last self-organization process took place. Immediately after the node joins the network, a timer constantly runs in the node. Whenever the timer expires, it is checked if the node is losing energy rapidly. This is done by calculating the ratio of current energy to that of the last self-organization. This is compared with a threshold value R. N is the number of child nodes of the node. If the ratio comes out to be less than the threshold value, then reorganization needs to take place. This is done by deleting the child node of the node which is the longest distance away from the node. This done by sending a delete signal signifies that this node is no longer your intermediate sink node.

(1) *Pseudocode for checking for consumption of energy of a node:*

```
Initialize n to No. of child nodes and threshold r to n/(n+1)
If energy timer has become zero
  If ratio of available energy to the energy after as reorganization is less than r
        Send delete packet to farthest child
```

(2) *Pseudocode for delete procedure:*

```
If delete packet is received
  Remove record of the farthest child
  Decrease the Count of child nodes by one.
 If count of child nodes is 0
  The node is not a network node
Else
  Start timer for energy checking
Return
```

Explanation for above pseudo code: Steps 1–2 initialize the threshold *R*, and the No. of child nodes, *n*. If the energy timer is determined to be zero and the ratio of the available energy and energy after reorganization turns out to be less than the threshold *R*, then the farthest child is signalled to be deleted, using the notation for the delete packet. This has been implemented from steps 3–5. At the node where the delete packet is received, the record of the farthest child is removed, decrement the child node count by 1, done across steps 1–3. If the count of the child nodes becomes zero, the node cannot be a network node, otherwise, the timer is initiated in order to check the energy.

D. *Algorithm 3: Re-electing the intermediate sink node*

(1) *Explanation*

The node which gets deleted during reorganization needs to find a new sink node. A node after receiving the delete message searches for a new intermediate sink node. It does so by broadcasting to all nearby nodes and starting a timer. The network nodes get the message and if they are willing to act as their sink, they give a positive reply. The node selects its intermediate sink node by calculating the weights of all nodes that gave a positive reply before the timer expired. If no reply is received, it chooses one of its child nodes as the new intermediate sink node. If it has no child nodes, it becomes a non-network node.

(2) *Pseudocode for re-electing the intermediate sink node*

```
Initialize sink to -1.
Initialize index to 0
Initialize is_sink to false.
Initialize num_avail_sink to 0.
Broadcast the sink search packet to all neighboring nodes
Start the timer
While the timer has not expired
  If acknowledgement is received from a neighboring node
        Increase num_avail_sink by 1
        Add the node to list optional_sink
If num_avail_sink is 0
  If  node is not a network node
        Go to start and follow the steps again
  Else
        Select one of the neighboring nodes as the sink
Else
  Find the best sink node from the array optional_sink
If the hop count of node changes
  Update the hop count of the child nodes.
Return
```

Steps 1 through 4 initialize the various parameters such as index of sink, loop index, number of available sinks, and the Boolean variable to check if node is sink. Step 5 broadcasts the sink search packet to all the neighbouring nodes. The timer is started in step 6. Steps 7–10 comprise of the iterative procedure. It terminates when the timer expires. For each acknowledgement received from a neighbouring node, the number of available sinks is incremented by 1, and the node is added to the list of optional sinks. After the timer expires, if number of available sinks still remains zero, and if concerned node is not network node, the steps are repeated, otherwise one of the neighbouring nodes is selected as sink. If there are available sinks, the optimal sink node is found from the sequence of optional sinks, done in step 17. The hop count of the child nodes is updated if the hop count of the node changes.

5 Results and Discussions

To evaluate the proposed algorithm, the protocol was implemented in C++ extensive simulations are performed. Evaluation was done on a wide range of parameters such as the size of network, routing cost, number of sink nodes, and transmission error. Unit disk graph model is utilized to characterize the connection between the sensor nodes. This model is prepared based on the assumption that all nodes have equal transmission power and the same circular range. So because of this assumption, the complete topology can be scaled down to obtain the sensor range equal to 1, which is not dependent on the transmission power and path loss.

Generating the topology randomly within a two-D uniform distribution of the sinks and the nodes in a square area ensures the randomness of nodes. While simulating, 100 nodes were considered in the area of 100 * 100 m^2.

As it can be seen from Fig. 2, which represents comparing the updated azimuth routing with that of greedy. In case of mid and low-density networks, i.e., one where average number of neighbours are 4 and 6, the performance of azimuth is extremely well with success of about 99.3% and 98.8, respectively, But with greedy, success is only about half i.e. 45.38 and 50.17 approximately.

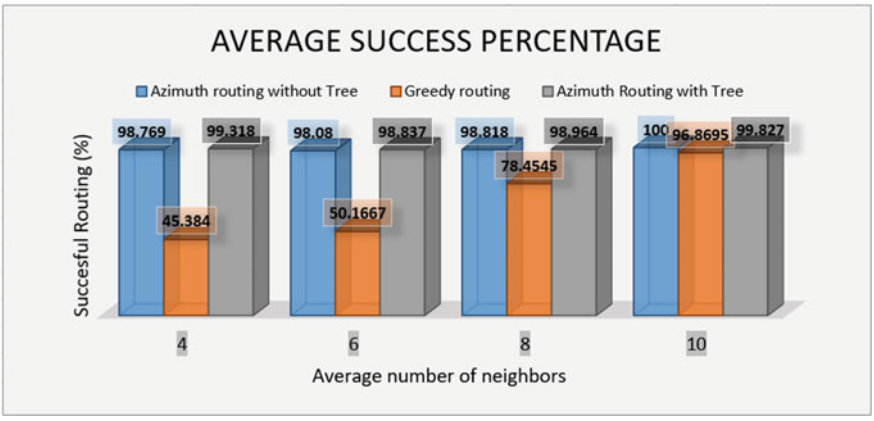


Fig. 2. Comparison of updated azimuth with greedy

As it can be seen from Fig. 3, in the case of azimuthal, the percentage of successful routing with tree is always between 95 and 100% but the performance of greedy routing is dispersed between about 20–100%. But in case of high density networks (number of neighbours is about 12 or more), we find that both greedy and azimuth perform equally well, greedy achieves the minimum path to the sink easily when the density increases.

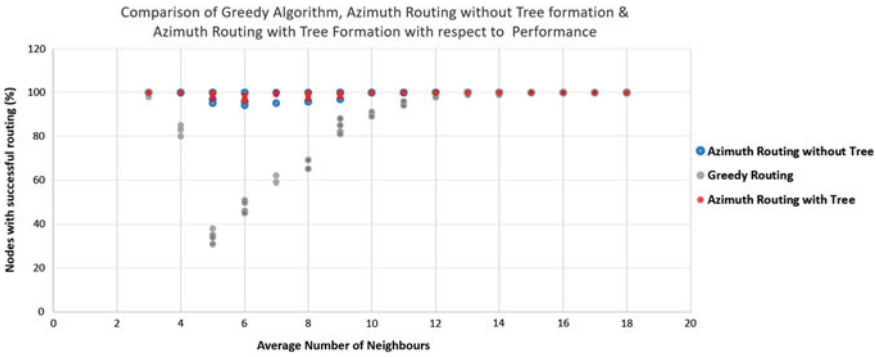


Fig. 3. Comparison between greedy and azimuth routing

In Fig. 4 it can be clearly seen the formation of a tree-based network from the 100 nodes has taken place where all the nodes are connected to the root node. For ease of picturisation we have considered an area of 5 * 5 units. Thus, increasing the dimensions of area while keeping the same number of sensors ensured variable density. To ensure the mobility of nodes in as WSN for IoT, nodes were randomly generated in the implementation of protocol.

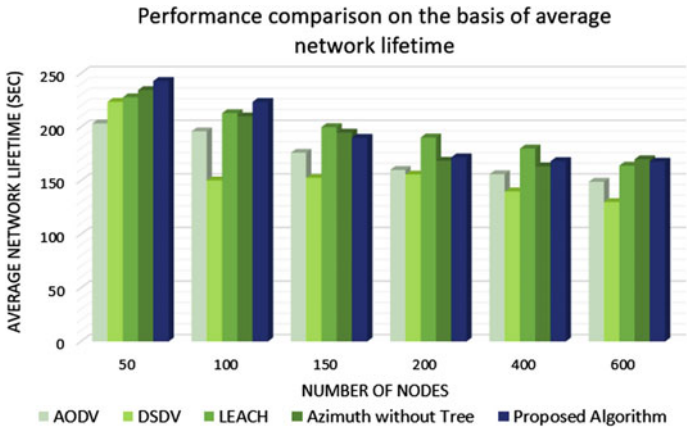


Fig. 4. Comparison average network lifetime between azimuth and greedy routing

From Fig. 4, we see that the network lifetime of the proposed protocol performs extraordinarily well compared to AODV and DSDV in terms of network lifetime because it kind of periodically check the residual energy of the network by sending energy check packets at 0.8 s and thus self-organize and reorganize the hot area to get efficient energy consumption. But we find that the proposed protocol has lesser network lifetime compared to LEACH which is due to the property of LEACH which basically revolves only around energy consumption and finding the cluster head with the maximum of the energy thus ensuring longer lifetime.

Figures 5 and 6 show the comparison of all the four protocols on the basis of probability of the successful routing. The success rate for the packets in the proposed protocol outperforms all the other three protocols, and as can be seen, it remains stable even with growing No. of nodes in the network while AODV and DSDV decrease in their efficiency w.r.t. the successful routing of packets with increment in the No. of nodes.

Figure 7 shows the average probability of successful routing. The routing is successful in both greedy and azimuth when the number of neighbouring nodes is more. But azimuth protocol performs extremely well in low-density networks as demonstrated below.

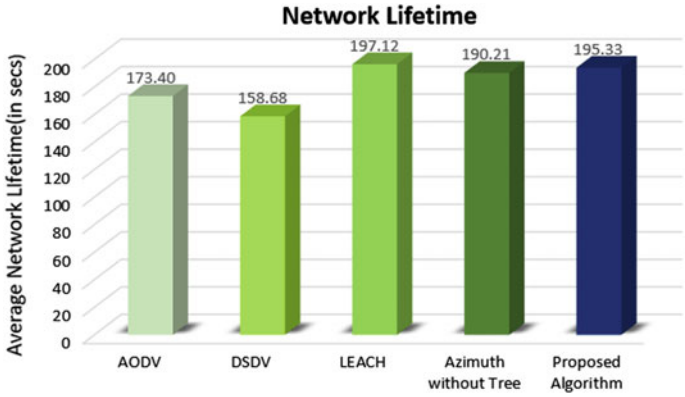


Fig. 5. Comparison between greedy and azimuth routing

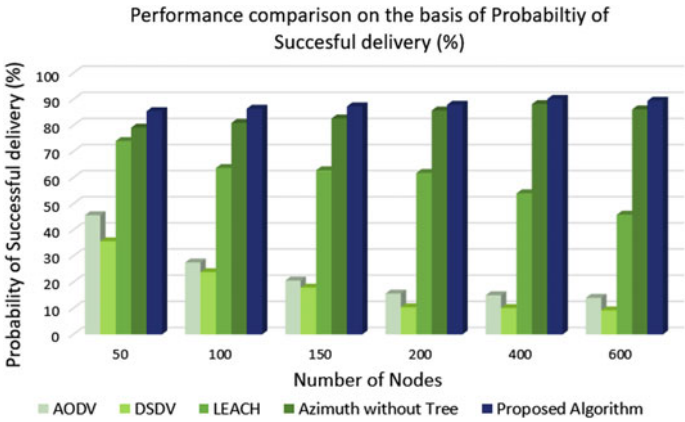


Fig. 6. Comparison for successful delivery

6 Conclusion

In this paper, Now in order to verify that our proposed protocol is efficient, experiments were conducted on different scales. The number of nodes are varied for construction of tree in different possible options and then checking their result. Hop node represents that choosing of intermediate sink node is based on the lowest value of hop in between the available intermediate sink node and the node. “Distance parameter shows that the choosing sink depends on the lowest distance from the available sink node to the node. Choosing sink depends on the highest residual energy of the available sink node. This is indicated by Left Energy. Choosing sink is based on the Number of each child node’s available sink node, which is indicated by Child Number. Greedy performs extremely poor in case of low or mid-density, while azimuth in that case also results in higher percentage of successful routing. So this supports our idea of implementing this protocol

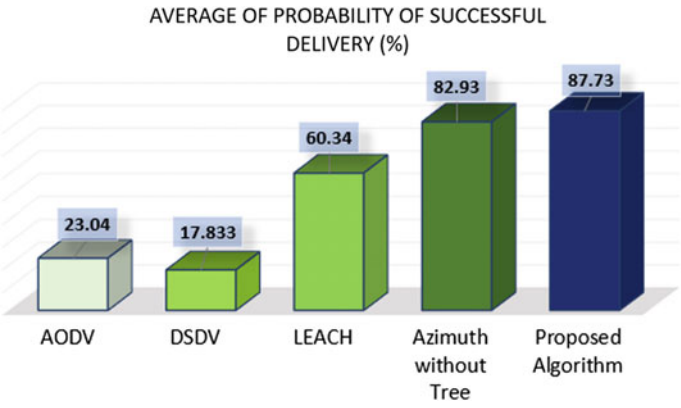


Fig. 7. Comparison for successful delivery

for large-scale mobile ad hoc networks with varying network density where some parts of it are of low or mid-density.

References

1. Agnihotri, S., Ramkumar, K.R.: A survey and comparative analysis of the various routing protocols of Internet of Things. Int. J. Pervasive Comput. Commun. **13**(3), 264–281 (2017)
2. Qiu, T., Liu, X., Feng, L., Zhou, Y., Zheng, K.: An efficient tree-based self-organizing protocol for internet of things. IEEE Access **4**, 3535–3546 (2016)
3. Noorul, T.M., Pramila, R.S., Islam, N.: An analysis of routing protocols in manets and Internet of things. In: IEEE International Conference on IoT and its Applications, ICIOT 2017 (2017)
4. Jeba, N., Kamala, V.: A survey on routing protocols for internet of things. Int. J. Adv. Res. Sci. Eng. Technol. **3**(5), 1993–1996 (2016)
5. Palan, N.G., Barbadekar, B.V., Patil, S.: Low energy adaptive clustering hierarchy (LEACH) protocol: a retrospective analysis. In: Proceedings of the International Conference on Inventive Systems and Control, ICISC 2017 (2017)
6. Rose, K., Eldridge, S., Lyman, C.: The internet of things: an overview. Internet Soc. (October), 53 (2015)
7. Ye, J., et al.: Lifetime optimization by load-balanced and energy efficient tree in wireless sensor networks. Mobile Netw. Appl. **18**(4), 488–499 (2013)
8. Kulakowski, P., Egea-lopez, E., Garcia-haro, J., Orozco-barbosa, L.: ARROW : Azimuth-Range ROuting for large-scale Wireless sensor networks, pp. 1–19 (2013)
9. Kulakowski, P., Egea-Lopez, E., Garcia-Haro, J., Orozco-Barbosa, L.: ARROW: azimuth-range routing for large-scale wireless sensor networks. EURASIP J. Wirel. Commun. Netw. **2013**(1), 93 (2013)
10. Singh, S.K., Singh, M.P., Singh, D.K.: Energy efficient homogenous clustering algorithm for wireless sensor networks. Int. J. Wirel. Mobile Netw. (IJWMN) 2(3) (2010)
11. Agrawal, P., Anand, V., Tripathi, S., Pandey, S., Kumar, S.: A solution for successful routing in low-Mid density network using updated Azimuthal protocol. In: Proceedings of the ICICC 2018, p. 29 (2018)
12. Rault, T., Bouabdallah, A., Challal, Y.: Energy efficiency in wireless sensor networks: a top-down survey. Comput. Netw, **67**, 104–122 (2014)

13. Kalosha, H., Nayak, A., Rührup, S., Stojmenović, I.: Select-and-protest-based beaconless georouting with guaranteed delivery in wireless sensor networks. In: Proceedings—IEEE INFOCOM, pp. 995–1003 (2008)
14. Li, S., Da Xu, L., Zhao, S.: The internet of things: a survey. Inf. Syst. Front. **17**(2), 243–259 (2015)
15. Miorandi, D., Sicari, S., De Pellegrini, F., Chlamtac, I.: Internet of things: vision, applications and research challenges. Ad Hoc Netw. **10**(7), 1497–1516 (2012)

Evaluation of Confidentiality and Data Integrity Based on User Session for IoT Environment

Alwi Bamhdi[1], Omar Abdulgadir[2(✉)], Akram Kargar[3], Ahmed Patel[4], and Bandar Al Ghamdi[2]

[1] Computer Science College, Umm AlQura University, Alqunfodah, Saudi Arabia
[2] Department of ITC, Arab Open University, Riyadh, Saudi Arabia
Luk_amri@hotmail.com
[3] Technical and Vocational Training Organization, Khorasan, Iran
[4] Computer Networks and Security Laboratory, UECE, Fortaleza, Brazil

Abstract. Confidentiality and data integrity are two important security services that are provided by the most popular VPN protocols such IPsec and SSL/TLS. It is a well-known fact that security comes at the cost of performance, and performance is affected by the cryptographic algorithms and their execution. In this study, confidentiality and data integrity processes were implemented using client--server Java applications in a real LAN to assess their performance based on different encryptions methods. The execution time of data encryption, data decryption, and data integrity verification was measured and total session times were computed for the following encryption algorithms AES, Blowfish, 3DES, RC2, MD5, and SHA-1. The observation results have been analyzed and the performance of these encryption algorithms and digest ciphers was compared, interpreted, and outlined in terms of total session time and different security parameters (data integrity, encryption, decryption, and data integrity verification).

Keywords: Confidentiality · Data integrity · Encryption algorithm · Digest cipher · Hash function · Session time performance · Internet of things

1 Introduction

With the advent of the computerization and digital communication, the security of data passing through networks has gained importance and became a major issue, because of criminal exploits and attacks. Douglas E. Comer (2018) listed the techniques used in security attacks as: wiretapping, replay, buffer overflow, address spoofing, name spoofing, denial of service (DOS) and distributed denial of service (DDOS), SYN flood, key breaking, port scanning, and packet interception [1]. These security attacks prove the need for reliable and efficient security mechanisms to protect data and applications. William Stallings (2016) defined the security mechanism as any process (or a device incorporating such a process) that is designed to detect, prevent, or recover from a security attack [2]. Examples of mechanisms are encryption algorithms, digital signatures, and

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_35

authentication protocols. Cryptography is the study and implementation of techniques to hide information from being read [3] to make them secure and immune to attacks. Security services are implemented by using security mechanisms. Security services include authentication, access control, data confidentiality, data integrity, non-repudiation, and availability. Such services are becoming very prevalent and of paramount importance in the field of the Internet of things (IoT) [4], cloud computing, big data [5], and data analytics. This paper focuses on confidentiality and data integrity (message authentication) which are offered by the most popular security protocols such as SSL/TLS and IPsec. We implemented confidentiality and data integrity using Java client--server applications in a real implemented LAN. We also run a group of tests on this network and evaluated, interpreted, and outlined the performance of various encryption algorithms among with digest ciphers based on total session time as performance metric. Encryption algorithms are used to provide confidentiality while digest ciphers (hash functions) are utilized to perform data integrity. Total session time equals to the time consumed from the beginning of the secure session on the sender side until its ending on the receiver side.

The rest of this paper includes sections as follows: Sects. 2 and 3 discuss about data confidentiality and data integrity, respectively. Section 4 describes the implementation of data integrity and confidentiality using Java and performing the tests. Section 5 provides the computations and outlines the experimental results. Finally, Sect. 6 concludes the paper.

2 Data Confidentiality

The concept of how to achieve message confidentiality or privacy has not changed for thousands of years [6]. Confidentiality is a security technique that uses encryption algorithms to make data unintelligible to unauthorized parties. Using this technique, data is encrypted at the sender side and decrypted at the receiver side. Many encryption algorithms are widely available and used in information security [7]. They are classified into symmetric key and asymmetric key. In symmetric-key cryptography, only one key is used for encryption at sender side and decryption at receiver side. This key has to be shared before transmission between parties. The key has an important role in strength of the cryptographic algorithms, since the algorithms are public and standard. Encryption algorithms are used with different key sizes. DES uses one 64-bit key. Triple DES (3DES) uses three 64-bit keys while AES uses various (128,192,256) bits keys. Blowfish uses various (32-448); default 128 bits while RC6 is used with various (128,192,256) bits keys [8]. Symmetric cryptography is the choice for applications that cannot afford computational complexity [9]. However, it is faster than asymmetric cryptography and is used for large amount of data. Asymmetric encryption uses a pair of keys which are mathematically linked. Public key is used for encryption, and private key is used for decryption. It is computationally infeasible to determine the decryption key given only knowledge of the cryptographic algorithm and the encryption key [10]. Public key is known to the public, but private key only is known to the receiver. Public keys do not need a secure channel for their distribution, but they do need an authentic channel. Security protocols utilize asymmetric-key cryptography for authentication and key exchange between parties to facilitate confidentiality.

3 Data Integrity

Data integrity refers to the property that data is not altered or destroyed in some unauthorized way, and hence data integrity services are to protect data from unauthorized modification [11]. Data integrity is also called message authentication. While source authentication ensures a receiver that the received data originates from the claimed source, message authentication guarantees that the data from the source is fresh and unchanged [12]. In order to perform data integrity, data message is run through a hash function, before it is transmitted between the parties. A cryptographic hash function takes a message of arbitrary length and creates a message digest of fixed length [13], known as a digest or hash. For data integrity purpose, the sender sends data message with the encrypted digest to the receiver. Then, receiver side decrypts the received digest and utilizes the same hash algorithm to produce the digest from received data message and verifies if the two digests are the same. Hash algorithms are effective because of the extremely low probability that two different plaintext messages will yield the same hash value [14]. The hash functions are on-way algorithms and also called digest cipher as we refer to it in this paper. Two popular digest ciphers are Message Digest 5 (MD5) and Standard Hash Algorithm 1 (SHA-1).

4 Confidentiality and Data Integrity Implementation

This paper implemented confidentiality and data integrity in a real client--server network to test and analyze the performance of these security services based on the session time. The real experimental network is a local area network (LAN) that consists of two laptop devices as the server and the client which are connected together through a switch supplying Ethernet connection. The client and the server perform confidentiality and data integrity using client and server Java programs running on each device. Firstly, an initial Java program was written to generate and distribute secret key, public key, and private key. Secret key will be used for encryption at the sender Java program and for decryption at the receiver Java program. Public key and private key belong to the sender side. Private key will be used for asymmetric encryption of the digest in the data integrity stage in server Java program while public key will be used for its decryption at the client side. The sender Java program performs as:

- Gets the data message and encrypts the message with an encryption algorithm and using the shared secret key
- Stores the time of the encryption process
- Produces a digest using a digest cipher and encrypts the digest using its private key
- Stores the time of the digest generation and encryption as data integrity time
- Sends the encrypted message and digest to the receiver using programming sockets.

The receiver Java program performs as:

- Receives the encrypted message and digest.
- Decrypts the message using the shared key and extracts the data message

- Stores the process time of decryption
- Produces a new digest using the same digest cipher used on the server side
- Compare the new cipher with the received one to verify if they are the same
- Stores the process time for new digest production and the comparison as the data integrity verification time.

Figure 1 illustrates the confidentiality and data integrity processes executing on the server and the client and interaction between these two. The server Java program runs on the sender device, and the client Java program runs on the receiver device. Java Development Kit (JDK) and Eclipse software were installed on the client and the server devices to help run the Java programs. The Ethernet LAN (192.168.0.x) was set up to connect the two devices. The specifications of the devices deployed in experimental network are described in Table 1. The tests are taken between the server and the client for the same amount of input data and are performed by changing encryption algorithms and digest ciphers at the sender and receiver Java programs in each test.

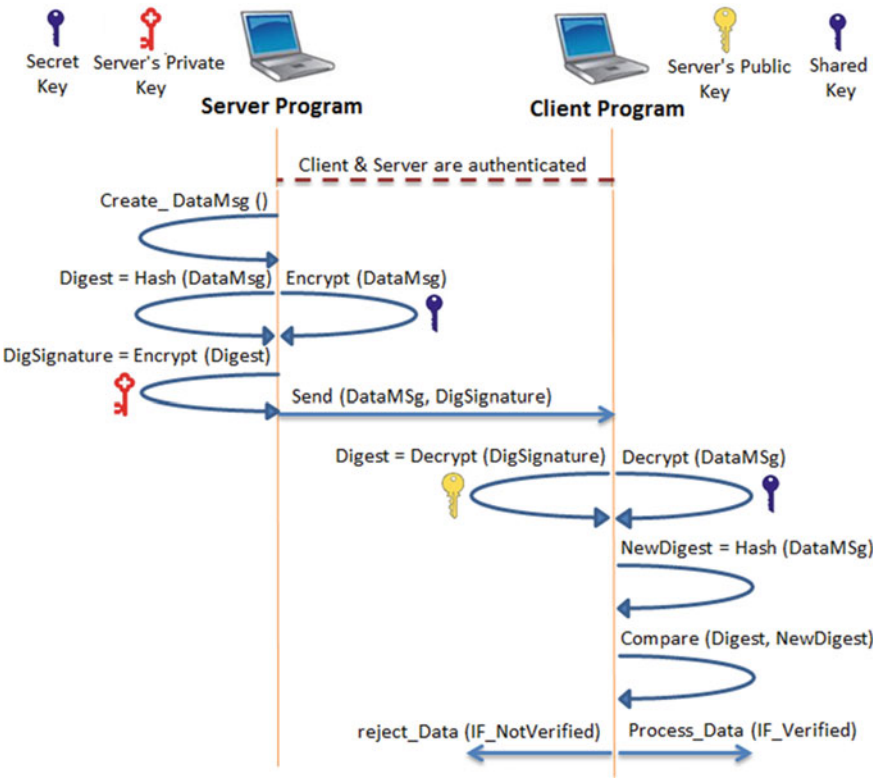


Fig. 1. Processes running for confidentiality and data integrity using Java program

The sender side stores and outputs the time consumed for data encryption and data integrity. Data integrity in Java programs includes digest generation and digest encryption

Table 1. Specifications of the experiment devices

System	Description
Server	Operating system: Windows 7, service pack 1 Processor information: Intel(R) Core(TM)2 Duo T7110 @ 1.80 GHz Real memory: 2.00 GB Link 4965AGN Software: Eclipse 4.3.0, JDK (Java Development Kit)1.7.0_40, JRE (Java Runtime Environment)7, Iperf
Client	Operating system: Windows 7, service pack 1 Processor information: Intel(R) Atom(TM) N2800 @ 1.86 GHz Real memory: 2.00 GB Software: Eclipse 4.3.0, JDK (Java Development Kit)1.7.0_40, Iperf
Switch	Datasheen D316S 16-Port /D324S 24-Port Fast Ethernet Web-Smart Switch

using private key and RSA asymmetric encryption algorithm. The sender Java program also outputs the size of the final data that is ready to send through the network. The receiver Java program outputs the time consumed for data decryption and data integrity verification. Data integrity verification includes a new digest production, decryption of the received digest, and comparison of the two digest. All outputted data in both Java programs is gathered to calculate the total session time for each test. The session time is the total time consumed for: (1) data encryption and (2) data integrity on the sender side, (3) transferring data through the network plus with (4) the data decryption and (5) data integrity verification on the recipient side. The data transfer time through the network can be obtained by dividing the transferred data size to network throughput. Before running the tests, Iperf (Internet Performance Working Group) was installed on the server and the client to measure network throughput. The throughput of the implemented network is 87.5 Mbps. Data transferred size is actually the size of encrypted data plus with the size of encrypted digest which are ready to send through the network and is equal to 342 Bytes. Therefore, the calculated data transfer time is 0.03 ms which is too smaller than other parameters that can be ignored. Various encryption algorithms were chosen to be varied during the tests. These encryption algorithms include AES, Blowfish, 3DES, and RC2. MD5 and SHA-1 were selected as digest ciphers which were set and changed during the tests. In order to make the results more accurate, the tests were replicated three times for each set of network parameters.

5 Experiment Results

The initial observation results of the tests are listed in Table 2. The stored times for different security parameters during the tests are specified in this table. The term DIT refers to data integrity time which itself is the sum of the digest generation time and digest encryption time and was computed in the sender program, the term DET refers to data encryption time, the term DDT refers to data decryption time, and the term DIVT refers to data integrity verification time which itself is the sum of the times for new

digest generation, decryption of the received digest, and the comparison between the two digest which was also computed in the receiver program. All the parameters are listed in milliseconds (ms). The term DTT refers to data transfer time. The input data file for all the test is a text file with the size of 50 kilobytes. The total session time can be obtained using the following equation:

Table 2. Time consumption of different security parameters for various encryption algorithms and digest ciphers used

Digest ciphers	Encryption algorithms				
		AES	Blowfish	3DES	RC2
MD5	DIT	815.32	755.26	780.38	812.25
	DET	134.49	30.97	33.63	26.28
	DDT	291.07	251.87	247.22	253.29
	DIVT	3.99	4.15	4.01	4
	DIT	789.34	763.22	806.14	776.04
	DET	128.18	36.28	33.75	26.03
	DDT	285.13	254.26	267.44	247.5
	DIVT	4.03	4.05	4.07	4.08
	DIT	798.79	782.96	773.25	768.71
	DET	127.84	30.92	34.06	26.16
	DDT	298.18	247.06	260.23	269.83
	DIVT	5.64	4.16	4.06	4.03
SHA-1	DIT	792.80	765.48	814.13	792.85
	DET	123.44	33.36	34.2	26.18
	DDT	300.03	244.08	242.41	241.47
	DIVT	3.97	4.1	4.07	4.04
	DIT	813.17	785.05	814.73	760.03
	DET	126.43	31	35.54	26.06
	DDT	281.59	275.46	273.17	263.16
	DIVT	3.99	4.12	4.7	4.49
	DIT	743.46	789.21	768.15	785.68
	DET	122.95	40.62	34.12	26.43
	DDT	298.41	267.32	273.23	259.73
	DIVT	4	4.19	3.96	5

$$ST = DIT + DET + DDT + DIVT + DTT \quad (1)$$

The values of the first four terms in Eq. (1), for each test, are listed in Table 2. However, DDT can be computed by dividing the data transfer size by network throughput, but compared to other terms, DDT is a too small value that can be ignored from the equation. Therefore, Eq. (2) can be used for computing total session time:

$$ST = DIT + DET + DDT + DIVT + \varepsilon \approx DIT + DET + DDT + DIVT \quad (2)$$

The session time was calculated using (2) for initial observation results of Table 2, and then the mean session time (of the three replications) was computed and gathered in Table 3. Figure 2 illustrates the variation of session time (ST) for different encryption algorithms and digests ciphers based on data in Table 3.

Table 3. Mean session time (ST) for various encryption algorithms and digest ciphers (in ms)

Digest ciphers	Encryption algorithms			
	AES	Blowfish	3DES	RC2
MD5	1227.34	1055.06	1082.75	1072.74
SHA-1	1204.75	1081.33	1100.81	1065.04

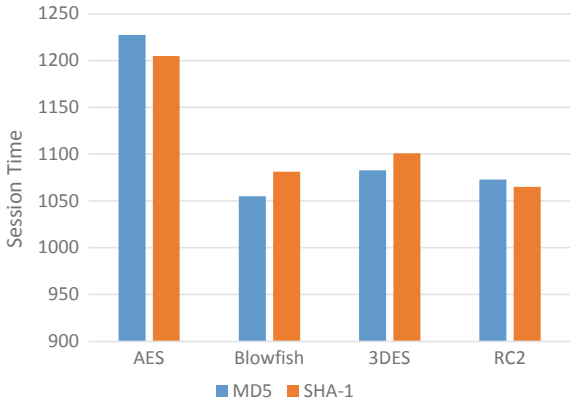


Fig. 2. Variation of session time for encryption algorithms and digest ciphers

The total session time is the sum of to the time taken for confidentiality and data integrity in sender side, the transfer time through the network and the time taken for confidentiality and data integrity in receiver side. By comparing the two graphs in Fig. 2, it is apparent that using MD5 for data integrity and AES for encryption produces the highest session time while using MD5 with Blowfish shows the lowest session time. Generally, using Blowfish and 3DES encryption algorithms among with MD5 produces lower session time than when they used with SHA-1, while using AES and RC2 among with MD5 shows higher session time than when they used with SHA-1. Shorter total session time distinguishes higher performance and higher total session time indicates lower performance.

Confidentiality and data integrity are two security services that are provided by the most popular security protocols such as IPsec and SSL/TLS. These protocols apply an encryption algorithm and a digest cipher to provide confidentiality and data integrity, respectively. Therefore, a set of encryption algorithm and digest cipher is an important factor which can represent the security performance. According to Fig. 2, we can arrange the sets of encryption algorithm and digest cipher from highest performance to the lowest performance as follows:

MD5-Blowfish, SHA1-RC2, MD5-RC2, SHA1-Blowfish, MD5-3DS, SHA1-3DES, SHA1-AES, and MD5-AES.

In addition, using data in Table 2, we can calculate the mean of each security parameters (DIT, DET, DDT, and DIVT) separately. Then, we can calculate the ratio of each parameter to the session time or compare different security parameters of each encryption algorithm-digest cipher set and draw them out. We can also realize the ratio of the time consumed on each side of the network (sender, receiver) to the total session time. The ratio of each security parameter time to the session time can be obtained using the following formula:

$$R = \frac{\text{Mean(SPT)}}{\text{ST}} \times 100 \tag{3}$$

where R = Ratio of each security parameter such as DIT, DET, DDT, and DIVT to ST. SPT term refers to security parameter time (one of DIT, DET, DDT, or DIVT), and the ratio of time consumed in each side of the network to the session time is obtained using the following formulas:

$$\text{SST} = \frac{\text{DET} + \text{DIT}}{\text{ST}} \times 100 \tag{4}$$

$$\text{RST} = \frac{\text{DDT} + \text{DIVT}}{\text{ST}} \times 100 \tag{5}$$

where SST refers to sender-side time and RST refers to receiver-side time.

Table 4 lists the calculation results of the mean consumption time of three test replications based on security parameter. Figure 3 illustrates and compares the processing time of different security parameters including DIT, DET, DDT, and DIVT for a single encryption algorithm using data of Table 4. This chart drew out for AES encryption algorithm. According to this figure, the highest time-consuming security service is data integrity process and the least time-consuming security service is data integrity verification process. Additionally, data decryption consumes more time than data encryption. These results are noticeable in security parameter of all encryption algorithm-digest cipher sets.

After computing the ratio of different security parameter using formula (3) and data in Table 4, the percentage of each security parameter to the session time were gathered in Table 5. Figure 4 demonstrates the percentage of DET to the ST for different encryption algorithm where the digest cipher is SHA-1. According to the chart in this figure, encryption algorithms from having the most percentage of session time to the least percentage are: AES, Blowfish, 3DES, and RC2. Similarly, the ratio of sender-side time consumption to session time and the ratio of receiver-side time consumption to session time are also extractable from data in Table 4 and formulas (4) and (5).

Table 4. Mean time consumed for different security parameter

Digest cipher	Mean parameters (ms)	Encryption algorithm/key size			
		AES	Blowfish	3DES	RC2
MD5	DIT	801.15	767.15	786.6	785.67
	DET	130.17	32.72	33.81	26.16
	DDT	291.46	251.06	258.3	256.87
	DIVT	4.55	4.12	4.05	4.04
SHA-1	DIT	783.14	779.91	799	779.52
	DET	124.27	34.99	34.62	26.22
	DDT	293.34	262.29	262.94	254.79
	DIVT	3.99	4.14	4.24	4.51

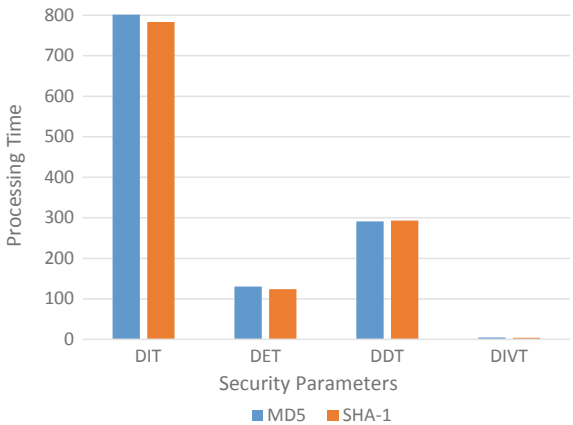


Fig. 3. Processing time of different security parameters

6 Conclusion

This paper has tested and analyzed the performance of confidentiality and data integrity which are security service that are provided by the most popular security protocols such a IPsec and SSL/TLS. To do so, client--server model Java programs were implemented and set up in a real experimental LAN to compute the times consumed for data encryption and data integrity process on the sender side plus with the data decryption and data integrity verification process on the receiver side. The experiment performed in this paper analyzed the performance of four encryption algorithm including AES, Blowfish, 3DES, and RC2 along with two digest ciphers of MD5 and SHA-1. Gathered data has been successfully assessed for correctness, analyzed and evaluated, which has produced the results to estimate the total session time for sets of encryption algorithm-digest ciphers, to compare time consumption of different security parameters (data integrity,

Table 5. Ratio of different security parameters to session time

Digest cipher	Ratio to ST	Encryption algorithm/key size			
		AES (%)	Blowfish (%)	3DES (%)	RC2 (%)
MD5	DIT	65.28	72.72	72.65	73.24
	DET	10.61	3.11	3.13	2.44
	DDT	23.75	23.8	23.86	23.95
	DIVT	0.38	0.4	0.38	0.38
SHA-1	DIT	65.01	72.13	72.59	73.2
	DET	10.32	3.24	3.15	2.47
	DDT	24.35	24.26	23.89	23.93
	DIVT	0.34	0.39	0.39	0.43

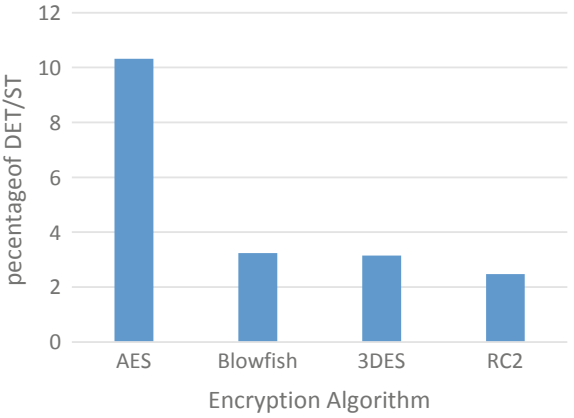


Fig. 4. Percentage of data encryption time to the session time

encryption, decryption, and data integrity verification) and to compute the ratio of each security parameter to the session time to advance a better understanding of them and the underlying pros and cons of using them. In our future research work in this subject area, we want to investigate the use of the encryption algorithms and security parameters in the specifics of intrusion detection and prevention of smart grid system [15], IoT systems, cloud computing, big data, and data analytics.

References

1. Comer, D.E.: Computer networks and internets, 6th edn. Person Education, Boston (2018)
2. Stallings, William: Cryptography and Network Security Principles and Practice, 7th edn. Prentice Hall, New Jersey (2016)
3. Mittal, Mohit: Performance evaluation of cryptographic algorithms. Int. J. Comput. Appl. **41**, 1–6 (2012)

4. Abdulkader, O., Bamhdi, A.M., Thayananthan, V., Al Ghamdi, K.J.B, Patel, A.: A secure lightweight mutual authentication and message exchange protocol for IoT environments based on the existence of active server. In: 4th International Congress on Information and Communication Technology in concurrent with ICT Excellence Awards (ICICT 2019) held at London, United Kingdom | 25–26 February 2019
5. Martins, F.L.J., Júnior, J.C., Gomes, R.L., Patel, A., Agoulmine, N.: BE-SYS: BigData E-health system for analysis and detection of risk of septic shock in adult patients. In: 53rd IEEE International Conference on Communications (ICC): SAC E-Health Track - E-Health, 20–24 May 2019, Shanghai, China, Empowering Intelligent Communications (2019). https://doi.org/10.1109/icc.2019.8761500
6. Forouzan, B.A.: Data Communications and Networking, 5th edn. McGraw-Hill, New York (2013)
7. Ismail, I.Amr, Amin, M., Diab, H.: A digital image encryption algorithm based a composition of two chaotic logistic maps. Int. J. Netw. Secur. **11**, 1–10 (2010)
8. El-Sisi, A.B., Mousa, H.M.: Evaluation of encryption algorithms for privacy preserving association rules mining. Int. J. Netw. Secur. **14**, 284–291 (2012)
9. Razvi Doomun, M., Soyjaudah, K.M.S.: Analytical comparison of cryptographic techniques for resource-constrained wireless security. Int. J. Netw. Secur. **9**, 82–94 (2009)
10. Furnell, S., Katsikas, S., Lopez, J., Patel, A.: Securing Information and Communications Systems Principles, Technologies and Applications. Artech House, Boston (2008)
11. Oppliger, R.: SSL and TLS Theory and Practice. Artech House, Boston (2009)
12. AI-Mashhadi, H.M., Abdul-Wahab, H.B., Hassan, R.F.: Secure and time efficient hash-based message authentication algorithm for wireless sensor networks. In: Global Summit on Computer & Information Technology (GSCIT), pp. 1–7. IEEE (2014)
13. Garg, P., Shanna, V.: An efficient and secure data storage in mobile cloud computing through RSA and hash function. In: International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), pp. 334–339. IEEE (2014)
14. Abutaha, M., Hamamreh, R.:, New one way hash algorithm using non-invertible matrix. In: International Conference on Computer Medical Applications (ICCMA), pp. 1–5. IEEE (2013)
15. Patel, A., Alhussian, H., Pedersen, J.P., Bounabat, B., Júnior, J.C., Katsikas, S.: A nifty collaborative intrusion detection and prevention architecture for Smart Grid ecosystems. Comput. Secur. **64**, 92–109 (2017). https://doi.org/10.1016/j.cose.2016.07.002

Quantales for Uncertainty in Design Engineering

Patrik Eklund(✉)

Department of Computing Science, Umeå University, 90187 Umeå, Sweden
peklund@cs.umu.se

Abstract. In this paper we provide recommendations on how to use quantales as algebraic structures to represent uncertainty and many-valuedness in design engineering using relational views for connecting and combining information. Information is further detailed as based on underlying signatures of types and operators, providing expressions and terms that also become subjected to many-valued qualifications. Machine and engineering design, and related design structures usually adopt rather trivial relational models, and shallow expressions to describe various conditions. In particular, uncertainty, e.g., in prediction and risk estimation, is often based on quite rudimentary and ad-hoc probabilities of events that are mostly just named rather than described in detail. The objects in question being just named items, without elaborating on the internal structure of these objects, makes these descriptions to be simple constants, and truth valuation remains as only binary. We will show how objects can be structured, and how structured objects can be related using various algebraic structures. This enables to provide a richer model also on many-valuedness from a logical point of view. Specifically we will look at the algebraization of the Design Structure Matrix (DSM).

Keywords: Condition · Expression · Fuzzy term · Many-valuedness · Quantale · Relational algebra

1 Introduction

Information involves observation and assessment typically using either numerical or symbolic values. Numerical values can be discrete or continuous whereas truth values can be seen as special symbolic values. Ordinal values can appear both using numbers as well as symbols.

Further, we 'operate' on values using operations having various algebraic properties, and we 'relate' values similarly with relations having certain properties. Order is a typical relation. Operation, result of operation, as well as relations between operations and results, can all be annotated logically in a many-valued domain of truth values represented in a complete lattice. Even for smaller num-

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_36

bers of elements in base sets, the variety of existing complete lattices enriches application development, and the number of quantale operations over such lattices is surprisingly large.

In this paper we look deeper into some and indeed some typical design structures approaches. These are, respectively, the Design Structure Matrix (DSM) [9], Axiomatic Design [15] and TRIZ [1,2]. Algebraically they are, on the one hand, organized as relational structures, and, on the other hand, suitable for enabling algebraic modeling of their information content and scope, in particular as far as uncertainties are concerned, and whenever uncertainty modeling complements efforts to provide stochastic modeling.

2 Uncertainty in Design Engineering

Faults and failures are typical examples where binary classifications are insufficient, and many-valuedness indeed desirable, as faults and failures reduce system and system-of-systems functioning capacity either partly or completely, and the progression of function loss often happens gradually rather than abruptly. In traditional engineering, functioning and its related classification is not connected with corresponding classifications of faults. Further, products or systems that are connected or interact as subsystems within a supersystem also interact with the environment involving activities and people. A common language for representing faults and functioning, and how they interact, is basically missing. Including people and their skills and competencies, and how that is affecting quality of products and production, adds complexity to this modelling problem. Furthermore, humans and products, the human and physical worlds progress and proceed through and within various activities and tasks, as part of complex processes. Organizing structure enables to control the overall system complexity a bit similar like preprocessing of data enables to reduce network complexity in machine learning [5]. Clustering and the use of rough sets [6] represent topological methods.

This triple, e.g. in DSM [9], involves modularity of component, people and activity, and in the Axiomatic Design model [15] involving customer, function, product and production, involving the interaction and combination of information and process, is not sufficiently structured and enriched by mathematical models. A unified conceptual framework for information and process for products and production is obviously too much to ask for, but deeper algebraic models of information combined with enriched many-valued logic supporting process modeling will contribute to such conceptual unification. Or the other way around, shallow information and process models only lead to shallow conceptual frameworks with poor management of interdisciplinarity.

3 Design Structures

3.1 Design Structure Matrix (DSM)

The design structure matrix is simply an unstructured two-valued relation ρ: $X \times X \to \epsilon$, where X may consists of components, people or activities. For example,

as components, actuators and temperature controls may be denoted as simply as actuator and EATC, or more structured as expressions like

$$\text{actuator}(f(x_1), \ldots, f(x_n)) \quad \text{and} \quad \text{EATC}(g(y, z)).$$

This requires component expressions as terms over a signature to be supported by some underlying component signature Σ_{Co}, so that e.g. the subsystem

$$\{\text{actuator}(f(x_1, \ldots, f(x_n), EATC(g(y, z))\} \subseteq \mathsf{T}_{\Sigma_{\text{Co}}} X,$$

where $\{\text{actuator}, f, EATC, g\} \subseteq \Omega_{\text{Co}}$.

Faults and failures are identified and assessed differently, but in a rule base they appear jointly and connected. Engineers typically use fault trees based on two-valued propositional logic, so typically using variables like x_{actuator} and $x_E ATC$ as boolean variables. A faulted actuator would, in a fault tree appear as $x_{\text{actuator}} = \texttt{false}$. Clearly, this information is quite shallow, whereas a many-valued qualification of actuator$(f(x_1), \ldots, f(x_n)$ is based on internal information residing within the expression. Furthermore, if the term functor works over, say, a Goguen category $\texttt{Set}(\mathfrak{Q})$, where $\mathfrak{Q}$ is a quantale, even actuator is assigned a truth value in the quantale.

DSM's "documentation of interaction between elements" [13] requires a many-valued logical description of interaction. Based on the 5-chain

$$L = \{\text{Detrimental, Undesired, Indifferent, Desired, Required}\},$$

and specifically detailed, respectively, for Spatial Scale, Energy Scale, Information Scale and Materials Scale, gives a relation of the form ρ: $X \times X \to L^4$.

DSM further has an intuitive notion of clustering based on ordering of elements into blocks, and ordering blocks. The idea goes back to precedence matrices [3,10,12]. In precedence matrices, relations are two-valued, and investigated graph-theoretically rather than logically.

3.2 Axiomatic Design (AD)

The Axiomatic Design model [15] involves a relation A, as a matrix between (a list of) functional requirements (**FR**) and (a list of) design parameters (**DP**). The $\mathbf{FR} - \mathbf{FR}$ pair structured through $\mathbf{FR} = A \times \mathbf{DP}$ basically connects potentially many-valued subsets of **FR** and **DP**, respectively. The sparser the matrix, the smaller the subsets. This is the main ideology of AD. One of the 'axioms' in AD states that information content should be minimized. It mainly means that the scope of different data types should be less rather than related information structures being sparse.

The intertwining of information and process models addresses the challenge to combine components and product taxonomies with activity and process hierarchies, in order to support decision-makers, engineers and customers. Many-valued logic enriches the anticipated combination of DSM with the Axiomatic

Design model [15], which in effect goes far beyond approaches dealing only with the unstructured matrix view. Typical process models are SysML and BPMN.

Whereas the DSM view involves modularity of component, people and activity, the Axiomatic Design model involves respective domains for customer, function, (physical) product and (observable) process.

Decoupled design using triangular matrices enables sequentiality. Axiomatic Design theory starts from customer attributes supporting functional requirements, then relates to design parameters, and finally arrives at process variables.

Attributes, requirements, parameters and variables are all just elements in sets (or lists), i.e., they can be seen as 0-ary operators in a signature consisting of sorts and constants only. Independence and Information Axioms are formulated using these elements, and decomposition within and between (zigzagging) information is then basically unstructured, and simply building up hierarchies of subsets.

Axiomatic Design relates information between domains, but not within domains as in DSM. Integrating DSM into Axiomatic Design using the unstructured and traditional models is suggested in [16]. Since this integration comes basically without structure on domains, structure-preservation makes no sense.

3.3 TRIZ

In comparison to Axiomatic Design, TRIZ has a focus on process variables. TRIZ is the acronym (in Russian) for the "Theory of Solving Inventive Problems". The focus is on inventions and formulation of patents, where a contradiction matrix for "The 40 Principles" is a crucial construction. Technical contradictions between the principles arise at various points in the design process.

TRIZ is about an inventor's thought process, with 'creativity' meaning the skill to state the problem correctly, and how to choose a road, the best road if possible, towards the correct solution. Contradiction, or physical contradiction, in the sense of TRIZ means that variables and values measurable and observable outside a chosen objective function affect the values of that objective function even when variables and values appearing within the model represented by the objective function may remain unchanged. This means e.g. that stochastic based optimization cannot be performed with sufficient reliability.

TRIZ' notion of "problem" is beyond just a problem to be solved like equations have solutions. Kolmogorov's "Aufgabe" [11] as task or assignment was intented as more than just a mathematical equation to be solved. However, Kolmogorov never defined "Aufgabe" but rather explained it using examples (*Wir definieren nicht, was eine Außgabe ist, sondern erklären dies durch einige Beispiele.*).

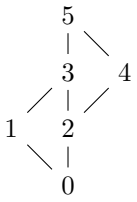
4 Semigroups, Lattices and Quantales

We already mentioned the many-valued relation $\rho\colon X \times X \to L^4$, with L typically a distributive lattice, i.e., with the meet operation distributing over the join

operation. This is a special kind of a *quantale*, which consists of a semigroup and a complete lattice, and with the semigroup operation distributing over the lattice join operation. With ρ: $X \times X \to Q^4$, and Q being a quantale, we have more operations at hand, which is useful in many practical situations. In the case of DSM, and the use of L^4, we may also consider smaller structures, such as a 2-pointed quantale Q_2 with a 3-pointed quantale Q_3, both, respectively, over a lattice as a chain. In this case the cartesian product $Q_2 \times Q_3$ is a 6-point quantales over a lattice structure that is not a chain.

$\wedge$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	0	1	0	1
2	0	0	2	2	2	2
3	0	1	2	3	2	3
4	0	0	2	2	4	4
5	0	1	2	3	4	5

$\vee$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	1	3	3	5	5
2	2	3	2	3	4	5
3	3	3	3	3	5	5
4	4	5	4	5	4	5
5	5	5	5	5	5	5



There are 1268 quantales for that particular lattice [14].

A *semigroup* $(X, *)$ consists of a (base) set X and an operation $* : X \times X \to X$ that satisfies

$$x * (y * z) = (x * y) * z$$

for all $x, y, z \in X$, i.e., a semigroup is an associative magma. A semigroup is *commutative* if

$$x * y = y * x$$

for all $x, y \in X$.

A *lattice* $(X, \wedge, \vee)$ consists of binary operations $\wedge$ and $\vee$ that are, respectively idempotent, commutative and associative, and further satisfying the absorption laws

$$x_1 \vee (x_1 \wedge x_2) = x_1$$
$$x_1 \wedge (x_1 \vee x_2) = x_1$$

for all $x_1, x_2, x_3 \in X$. For lattices, idempotency of $\vee$ and $\wedge$ follows from the absorption laws.

A finite lattice, i.e., a lattice with a finite base set, is a *complete lattice* $(X, \wedge, \vee, \bot, \top)$ with $\bot$ as the smallest element, given the partial order generated by the lattice, and $\top$ as the largest element.

A lattice is *distributive* if

$$x_1 \vee (x_2 \wedge x_3) = (x_1 \vee x_2) \wedge (x_1 \vee x_3)$$
$$x_1 \wedge (x_2 \vee x_3) = (x_1 \wedge x_2) \vee (x_1 \wedge x_3)$$

for all $x_1, x_2, x_3 \in X$.

A *quantale* $(X, *, \wedge, \vee, \bot, \top)$ is a semigroup $(X, *)$ operating over a complete lattice $(X, \wedge, \vee, \bot, \top)$ where the semigroup operation is join-preserving in both variables, i.e., satisfying

$$x * (\bigvee_{a \in A \subseteq X} a) = \bigvee_{a \in A \subseteq X} (x * a)$$
$$(\bigvee_{a \in A \subseteq X} a) * x = \bigvee_{a \in A \subseteq X} (a * x)$$

for all $x \in X$ and all $A \subseteq X$. Given the convention $\bigvee \emptyset = \bot$, we have

$$x * \bot = \bot * x = \bot$$

for all $x \in X$.

A *prequantale* is a magma and a complete lattice, where the magma operation is join-preserving in both variables.

There are 2 quantales on a 2-pointed set, 12 quantales on a 3-pointed sets, 129 quantales on a 4-pointed set, 1852 quantales on a 5-pointed set, and 33391 quantales on a 6-pointed set.

A *balanced quantale* is a quantale satisfying

$$\top * \top = \top.$$

A *unital quantale* $(X, *, e, \wedge, \vee, \bot, \top)$ is a quantale where $(X, *, e)$ is a monoid, and where the unit e satisfies

$$x * e = e * x = x$$

for all $x \in X$. An *integral quantale* is a unital quantale satisfying

$$e = \top.$$

In applications, we will often be interested in quantales that are simultaneously unital (and therefore also balanced), and non-integral.

Uncertainty and many-valuedness typically represented by finite algebraic structures complement numerical modelling of probability and risk. Whereas valuation in probability uses the unit interval $[0, 1]$, finite algebraic structures combine operation ($*\colon X \times X \to X$) with order ($\leq : X \times X \to \{\text{False, True}\}$).

A typical and general operation is provided by a semigroup $(X, *)$, with $*$ being an associative operation. In the finite situation, a complete partial order $(X, \leq)$ corresponds to a (complete) lattice $(X, \wedge, \vee)$, with $\wedge, \vee\colon X \times X \to X$ fulfilling lattice properties. In the finite case a lattice is always complete, i.e., has smallest and elements $\bot, \top \in X$ corresponding, respectively, to $False$ and $True$. We usually write $\mathbf{2} = \{\text{False, True}\}$.

Finite many-valuedness is more than just an extension or generalization of two-valuedness. Many-valuedness involves relations between values as well as

operations on values. Binary relations suffice and mostly requires transitivity, and binary operations suffice as mostly requiring associativity. Partial orders giving (finite) lattices are then useful since order as relation reshapes into lattice operations.

A binary operation in a semigroup, distributing over an order structure as its counterpart is the starting point for quantales as Semigroups in Complete Lattices [8].

Even for smaller numbers of elements in base sets, the number of quantales is very large, where the Catalogue of Quantales [14] supports the view on *Practicability of Quantales*.

Switching between design structures would need to lean on mappings and homomorphisms between structures. Whereas morphisms between algebraic structures connected with logic involves transformations of many-valued terms [7], morphisms between algebraic structures connected with topologies may involve pairs of morphisms [4] acting in opposite directions.

5 Many-Valued Expressions

Common to many design structure approaches is that attributes or parameters are unstructured elements in sets, i.e., they are just named variables holding values. Types are not recognized, nor operators producing terms as expressions including variables and values. Categorically, the underlying category is always the category of sets and functions $\mathtt{Set}$, i.e., the underlying category does not provide a canvas for any structure, over which expressions (terms) are effected with respective to their valuations. To be more precise, if we have an underlying signature $\Sigma = (S, \Omega)$ with sorts (types) in S and operators in Ω, a many-sorted $\mathbf{T}_\Sigma$ over $\mathtt{Set}_S$, the many-sorted category of sets and functions, where $\Sigma = (S, \Omega)$ is a signature, will have, for a sort (i.e. type) $\mathtt{s} \in S$, specific functors $\mathsf{T}_{\Sigma,\mathtt{s}} \colon \mathtt{Set}_S \to \mathtt{Set}$.

The term functor construction ([7]) can be extended so that $\mathsf{T}_\Sigma \colon \mathtt{C} \to \mathtt{C}$ operates more generally over monoidal biclosed categories $\mathtt{C}$. If $\mathtt{C}$ is $\mathtt{Set}$, we have the construction above, and with the Goguen category $\mathtt{Set}(Q)$, where Q is e.g. a quantale, we have a multivalent and typed situation enabled by the signature acting over the selected underlying category.

Whenever design structures involve relations they start off from being unstructured, i.e., they are relations as two-valued $\rho \colon X \times X \to \mathtt{2}$, where $\mathtt{2} = \{\bot, \top\}$ the two-valued lattice, or as $\rho \colon X \times Y \to \mathtt{2}$. More elaborate relations will be many-valued $\rho \colon X \times Y \to \mathfrak{A}$, with $\mathfrak{A} = \{\bot, \top\}$ being algebraic structures like lattices, boolean algebras, or quantales involving semigroup like operators over lattices. Instead of having just sets X and Y of unstructured elements, we might have expressions over signatures, so that relations might look like $\rho \colon \mathsf{T}_{\Sigma_1} X_1 \times \mathsf{T}_{\Sigma_2} X_2 \to \mathfrak{A}$.

Expressions like $actuator(f(x_1), \ldots, f(x_n))$ and $EATC(g(y, z))$ can now be annotated as uncertain or many-valued, aggregating the underlying uncertainties residing, respectively, in $actuator$, f, $x_1, \ldots, x_n$, $EATC$, g, y and z. The aggregation algorithm is derived from the term functor construction [7].

6 Conclusion

Applicability of algebraic structures and relational algebra making use of such structures is obviously not generic or appear in generality with respect to our understanding of it in respective application scenarios. Much of it, we should say, is 'in casu'. For instance, in the case of the unit in a unital quantale representing a "not yet known" is sometimes more like "(still) missing", sometimes it is known but not all that comparable to the other elements. Sometimes such a "not yet known" qualification of something related with and within a scale, say in the style of small-normal-large, is qualifiable but but more like "this does have an effect, but I cannot say how". This we often see in identifying customer needs in product design and development [17].

References

1. Altshuller, G.S.: Creativity as an Exact Science. Gordon & Breach (1984)
2. Altshuller, G.S.: The Innovation Algorithm: TRIZ, systematic innovation and technical creativity, Technical Innovation Center, Worcester, MA (1999)
3. Barankin, E.W.: Precedence matrices, University of California Management Sciences Research Project, Research Report No. 26 (1953)
4. Eklund, P.: A comparison of lattice-theoretic approaches to fuzzy topology. Fuzzy Sets Syst. **19**, 81–87 (1986)
5. Eklund, P.: Network size versus preprocessing. In: Yager, R., Zadeh, L. (eds.) Fuzzy Sets, Neural Networks and Soft Computing, Van Nostrand Reinhold, New York, pp. 250–264 (1994)
6. Eklund, P., Galán, M.A.: Rough objects in monoidal closed categories. In: Mani, A., Cattaneo, G., Düntsch, I. (eds.) Algebraic Methods in General Rough Sets, Trends in Mathematics Series, Birkhäuser Mathematics, pp. 641–655. Springer, Berlin (2018)
7. Eklund, P., Galán, M.A., Helgesson, R., Kortelainen, J.: Fuzzy terms. Fuzzy Sets Syst. **256**, 211–235 (2014)
8. Eklund, P., Gutiérrez García, J., Höhle, U., Kortelainen, J.: Semigroups in Complete Lattices: Quantales, Modules and Related Topics, Developments in Mathematics, vol. 54. Springer, Berlin (2018)
9. Eppinger, S.D., Browning, T.R.: Engineering Systems: Design Matrix Methods and Applications. MIT Press, Cambridge (2012)
10. Harary, F.: On the consistency of precedence matrices. J. ACM **7**, 255–259 (1960)
11. Kolmogorov, A.N.: Zur Deutung der intuitionistischen Logik. Mathematische Zeitschrift **35**, 58–65 (1932)
12. Marimont, R.B.: A new method of checking consistency of precedence matrices. J. Assoc. Comp. Mach. **6**, 164–171 (1959)
13. Pimmler, T.U., Eppinger, S.D.: Integration analysis of product decompositions. In: ASME Design Theory and Methodology Conference, Minneapolis, September 1994
14. Shamsgovara, A., Eklund, P., Winter, M.: A Catalogue of Finite Quantales. In: GLIOC Notes, December 2019
15. Suh, N.P.: Axiomatic Design: Advances and Applications. Oxford University Press, Oxford (2001)

16. Tang, D., Zhu, R., Dai, S., Zhang, G.: Enhancing axiomatic design with design structure matrix. Concurrent Eng. Res. Appl. **17**, 129–137 (2009)
17. Ulrich, K.T., Eppinger, S.D.: Product Design and Development, 6th ed. McGraw-Hill Education, New York (2016)

Deciphering of the gpsOne File Format for Assisted GPS Service

Vladimir Vinnikov[1(✉)] and Ekaterina Pshehotskaya[2]

[1] Higher School of Economics, Department of Computer Sciences, Pokrovsky Blvd, 11, 109028 Moscow, Russian Federation
vvinnikov@hse.ru|https://www.researchgate.net/profile/Vladimir_Vinnikov
[2] Department of Information Security, Moscow Polytechnic University, Avtozavodskaya st., 16, building 4, 115280 Moscow, Russian Federation

Abstract. The paper is concerned with deciphering the data format of a gpsOneXTRA binary file for A-GPS web service. We consider mandatory data content of the file and reveal the changes of this content at different moments of time. The frequency of the changes hints on the location of records for current GPS date and satellite orbits information. Comparing the repeating data patterns against reference orbits information, we obtain meaning of data fields of the orbit record for each operational satellite. The deciphered file header and GPS almanac data layout are provided as tables within the paper.

Keywords: Assisted gps · Almanac · gpsOneXTRA service · Deciphering · Data structure · Binary format

1 Introduction

Recent advances in microelectronics and space technologies enabled the wide spreading of geo-positioning functionality implemented in various devices like smartphones, digital cameras and fitness trackers. The geo-positioning relies on global navigation satellite systems (GNSS) used to determine range (as time delay of traveling signal) from spaceborne emitting antenna to receiving antenna. The performance of GNSS-positioning depends on the number and mutual disposition of satellites on the sky hemisphere. However, the coordinates of the satellites are a-priori unknown. By default, the receiver has to lock-on to satellite signal and download navigational messages, containing data on satellite orbits. The orbits are represented as rough almanac and more accurate ephemeris parts. However, the locking-on and downloading takes significant time (up to 12.5 min for GPS worst-case scenario) due to low throughput of the of the satellite signal broadcast with data rate at only 50 bit/s. This circumstance impacts usability

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_37

of the freshly started device. To overcome this inconvenience, it was proposed to deliver the almanac and the ephemerides into receiver software via faster technical means. Since most of the consumer devices are within the reach of the Internet, the downloading of respective data files from web services was considered the best solution.

The above-described approach to improving performance of GNSS positioning is known as "Assisted GPS" or AGPS (e.g., see [1]). The A-GPS technology is implemented as a web service, the frontend of which consists of a binary datafile and respective URL to retrieve it from a server. Almost all smartphone manufacturers provide their own A-GPS services. The widely used services with A-GPS content are provided by Google (for GlobalLocate chipset), Qualcomm (for gpsOne chipset), Mediatek (for SiRFStarIII chipset). The respective URLs are given in Table 1. Considering smartphones with Android OS, these URLs are substituted into ini-file as values for the keys "XTRA_SERVER_1", "XTRA_SERVER_2", "XTRA_SERVER_3", "XTRA_SERVER_S".

However, to this day, there is no standard format for A-GPS files. Moreover, binary files have proprietary formats with a priori unknown mapping to data structures. The exposure of A-GPS file data layout is a significant factor for improving information security and reducing risks of various exploits designed to compromise end-point user devices. For example, the gpsOne service was considered vulnerable at least in two issues (see [2,3]). One issue was concerned with MitM-attack through unsecured HTTP-protocol able to substitute correct binary file with the fake. The other issue was on ingestion of the fake binary file of large size leading to system crash of Android OS. These vulnerabilities allowed cumulative exploits undetectable by any antiviral scans due to unknown structure of the binary files. Although the issues were fixed, the knowledge of the file structure permits one to safely parse the data fields and check for any inconsistencies, thus facilitating protection against potential exploits. Our aim is to outline a deciphering technique for A-GPS files in a tutorial manner using classical cryptography attacks based on data redundancy and repetition.

The paper is structured as follows. Section 2 outlines the properties of GPS orbits and the respective data formats they are packed in. Section 3 introduces common cryptanalysis approaches to decipher the A-GPS binary file. Section 4 describes the deciphered layout of the A-GPS binary file in gpsOneXTRA format. Section 5 speculates on the future work of deciphering GPS ephemeris in the said format. Finally, Sect. 6 presents our conclusions.

2 General Considerations on the Data File Content

The mandatory contents of A-GPS file include almanac and ephemerides of the considered GNSS for the actual timeframe. To this day, there are four global satellite systems, namely GPS, GLONASS (GLN), BEIDOU (BDS), and GALILEO (GAL). The almanac represents long-term valid orbit approximations for respective GNSS. The ephemeris, as term suggests, is a short-lived data. For the GPS, it expires about every two hours and is updated about every hour.

Table 1. Providers of free A-GPS services

Provider	Ver	URL
Google	–	http://gllto.glpals.com/4day/glo/v2/latest/lto2.dat
Qualcomm	1	http://xtra1.gpsonextra.net/xtra.bin
		http://xtra2.gpsonextra.net/xtra.bin
		http://xtra3.gpsonextra.net/xtra.bin
		https://xtrapath1.izatcloud.net/xtra.bin
		https://xtrapath2.izatcloud.net/xtra.bin
		https://xtrapath3.izatcloud.net/xtra.bin
	1 SSL	https://ssl.gpsonextra.net/xtra.bin
	2	http://xtrapath1.izatcloud.net/xtra2.bin
		http://xtrapath2.izatcloud.net/xtra2.bin
		http://xtrapath3.izatcloud.net/xtra2.bin
	2 SSL	https://ssl.gpsonextra.net/xtra2.bin
	3	https://xtrapath1.izatcloud.net/xtra3grc.bin
		https://xtrapath2.izatcloud.net/xtra3grc.bin
		https://xtrapath3.izatcloud.net/xtra3grc.bin
	3 SSL	https://ssl.gpsonextra.net/xtra3grc.bin
Mediatek	–	http://nsdu.atwebpages.com/packedephemeris.ee
	–	http://epodownload.mediatek.com/EPO.DAT
Sony	–	http://control.d-imaging.sony.co.jp/GPS/assistme.dat
Nikon	–	https://downloadcenter.nikonimglib.com/en/download/fw/111.html
Olympus	–	http://sdl.olympus-imaging.com/agps/index.en.html

Other systems have similar expiration times for corresponding ephemerides. Due to short expiration times, the A-GPS file also contains as well as various timestamps denoting data validity as well as ephemerides prediction for 7–28 days in order tuce o reddevice dependence on the web connectivity.

The initial GNSS-broadcasted navigation messages have rigid data format, so it is safe to assume that A-GPS file is also similarly structured into fixed data fields. Such kind of data structures manifest themselves in binary files as periodic patterns with simply deducible size. We also assume that data is stored in efficient way, so only numeric types are used, and minimum required number of bytes is allocated for them. We should emphasize that one should keep in mind the possibility of two different binary bitwise representations, namely "Little Endian" and "Big Endian." The former is usually implemented in CPUs of x86 architecture, while the latter is used in ARM processors of mobile devices like smartphones. Thus, it is more likely to encounter "Big Endian" encoding in A-GPS files. Since our study is predominantly educational, for the sake of clarity, we consider Qualcomm-provided service for "gpsOne" chipset with first version of the binary file, containing almanac and ephemerides only for GPS (file "xtra.bin" at URL "xtra1"). An example of the binary file header is presented in table 2.

Table 2. Binary content of the file

Offset	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E	1F
0x0000	01	1B	08	01	02	15	01	24	05	BB	13	00	00	96	DE	08	24	16	01	DF	FD	08	24	15	E3	07	00	06	1C	01	00	25
0x0020	14	07	10	0F	0E	0D	0C	0B	0A	0C	37	08	10	10	0F	0E	0D	0C	0A	09	0E	53	08	11	0F	0E	0E	0C	0C	09	08	0D
0x0040	96	02	0B	05	02	03	03	C2	1F	01	00	4B	6A	90	17	81	FD	62	00	A1	0C	CA	FF	F2	84	DF	00	1E	F2	F4	00	5B
0x0060	E1	04	FF	07	FF	FD	08	24	02	00	A0	8E	90	09	D6	FD	55	00	A1	0C	6D	FF	EF	8B	58	FF	BB	3F	64	00	67	D6
0x0080	9B	FE	7B	FF	FE	08	24	03	00	15	65	90	0E	51	FD	43	00	A1	0C	B8	00	1C	EE	72	00	1F	D7	39	00	2C	17	D1
0x00A0	FF	C4	FF	FE	08	24	04	FF	03	5D	7B	0B	15	FD	55	00	A1	0D	7A	00	48	E5	9A	FF	93	67	AF	FF	8F	0A	EC	FF
0x00C0	E8	FF	FF	08	24	05	00	2F	87	90	05	C9	FD	36	00	A1	0C	0A	00	1B	B8	75	00	20	CB	86	FF	D0	8D	52	FF	FB
0x00E0	00	00	08	24	06	00	0D	F9	90	17	40	FD	65	00	A1	0D	6B	FF	F2	2E	BB	FF	CF	03	85	00	6A	75	E5	FF	58	FF
0x0100	FD	08	24	07	00	6C	2A	90	07	B9	FD	50	00	A1	0D	52	00	72	C3	B3	FF	9D	60	35	00	3F	3A	B1	FF	49	FF	FE
0x0120	08	24	08	00	28	08	90	11	E2	FD	45	00	A1	0C	70	FF	C6	FB	81	FF	F6	DF	3E	FF	BD	BB	1A	FF	EE	00	00	08
0x0140	24	09	00	0E	03	90	06	4A	FD	4A	00	A1	0B	DF	00	46	F7	7C	00	44	A6	6A	FF	C9	94	0C	FF	89	FF	FD	08	24

3 General Considerations on the Cryptography Attacks Exploiting Data Redundancy

It is established that the binary A-GPS files have proprietary format, but are not truly encrypted in narrow sense. However, the lack of information on the data structure of the files can be treated the same way as the encryption. This kind of encryption originates from the well-known paradigm "security through obscurity." The paradigm itself is widely recognized as unreliable, and its implementations are considered as bad practice. However, this circumstance makes the recovery of the data structure to be easier than in the case of proper classical encryption.

The data structure of the binary file is completely defined by the numeric values that fill it. By establishing matches between numeric values and their reference counterparts, one can determine the underlying data structure. Since the sought-for numeric values vary with time, we implement quasi-differential cryptanalysis to reveal change patterns within the data on different timescales. Contrary to the true differential cryptanalysis, this one is based on a partial quazi-known-plaintext attack instead of a chosen-plaintext attack. The quazi-known-plaintext attack assumes that attacker still lacks the original plaintext, but has some approximation at least with numerical values of the same magnitude and sign.

4 Analysis of A-GPS Binary File

The GNSS-positioning technology by design relies on a timing, so the primary parameter is the timestamp of data origin. This timestamp is expressed in terms of GPS-week and GPS-day numbers (e.g., [4]) as well as seconds elapsed from some reference instance. Usually, the GNSS-positioning operates on the timescale of milliseconds, so it is reasonable to expect also a data field holding the number of milliseconds.

At initial stage, we obtain seven triplets of binary files. The binary files of the first triplet are to be downloaded at a temporal distance of about 45 minutes, and the files of every next triplet are to be downloaded about 24–25 hours after last file of the previous triplet.

Next, we perform byte-to-byte comparison of the downloaded files via one of the hexadecimal viewers. The results reveal that WORD-variable occupying offsets 0x0F and 0x10 undergo changes between triplets, but not within a triplet (with rare exceptions). Moreover, these changes fit in a pattern of consequential daily increments. Assuming "Big Endian" format and referencing, e.g., [4], one can deduce that the offsets 0x0F and 0x10 essentially hold FULL GPS week since first epoch. The same goes to the offsets 0x15 and 0x16.

Analyzing the DWORD gap (at offsets 0x11–0x14) between GPS week timestamps, one can see that the value changes within each triplet. The computed difference between DWORD-values in consecutive files of a triplet corresponds to about 30 min expressed in milliseconds. The difference between said values in

consecutive triplets confirms the guess by approximately corresponding to the number of milliseconds within a day. Thus, the DWORD-values of seven triplets form the sequence of milliseconds. Extrapolating this sequence backwards to zero, we deduce that the DWORD-value holds actual milliseconds of GPS week for the orbits data in the respective file. The partially deciphered header for the A-GPS binary file is presented in Table 3. To proceed with more refined analysis, one should use larger volume of data, say monthly worth set of files with temporary spacing 30 minutes. Such level of detail is useful for deciphering current and predicted ephemerides.

At the second stage of analysis, we consider blocks of data that are almost constant for all files, downloaded within a day or a week. Our aim is to locate offset and length of a byte sequence corresponding to GPS almanac. Relying on the common sense, one can assume that the almanac is represented as a set of uniform-sized blocks arranged in ascending order of respective satellite numbers (GPS PRN designators, PRN $\in$ [1;32]). Moreover, each block should contain nine parameters, describing orbital elements (see, e.g., [5]) as the set: reference time, eccentricity, orbital inclination, rate of right ascension, a semi-major axis, argument of perigee, mean anomaly, clock correction, and rate of clock correction. As one can see, orbital elements can be divided into two categories by a property of having a constant sign across all satellites. Thus, reference time, eccentricity, orbital inclination, rate of right ascension, and semi-major axis constitute one category, while the remaining parameters form another category with varying sign across all satellites.

Using these categories, we can match orbital elements and data fields within the blocks via sign-to-sign comparison. However, direct matching by value comparison is problematic due to issues of floating-point precision and format of element record. Consulting [6] on the topic of "SEM Almanac Description," we estimated the most likely byte lengths for the respective orbital elements.

After performing sign matching, we deduced that the first block of satellite almanac starts at global offset 0x0049. This block corresponds to GPS PRN designator equal to one. The second block corresponds to PRN = 2 and so forth. Thus, the size of the sought-for GPS PRN SLOT is 31 bytes. The revealed layout of data fields is provided in Table 4. For the purposes of sign matching, we used fields of the longitude of ascending node $L\Omega$ and the argument of perigee ω. The first byte at each of these respective two fields has distinctive values of either 0x00 or 0xFF representing the sign of 32-bit integer.

While analyzing second version of A-GPS binary file, we discovered that the header was extended, and the offsets of the GPS PRN slots were described as an expression (Address 0x0073+0x001E*(PRN-1)). In contrast to the first version, the second one contains GPS week number with respect to rollover procedure (i.e., modulo 1024).

Table 3. File header (address 0x0000, big Endian)

Offset	Type	Content	Value	Comment
0x00	U1	Header	0x01	–
0x01	U1	Format ver	0x1B	version 1
			0x34	versions 2, 3
0x02	...	Undeciphered	...	...
0x0B	U4	Length of file	0x00	0x000096DE – v.1
0x0C			0x00	0x00007E6C – v.2
0x0D				0x00006347 – v.3
0x0E				
0x0F	U2	GPS week since 1st epoch without rollover		
0x10				
0x11	U4	Actual seconds of GPS week for provided data, [ms]		Value changes every 30 min and some seconds
0x12				
0x13				
0x14				
0x15	U2	GPS week since 1st epoch without rollover		
0x16				
0x17	U4	Reference seconds of GPS week for provided data, [ms]		Value changes every hour and some seconds
0x18				
0x19				
0x1A				
0x1B	...	Undeciphered	...	...
0x46	U2	GPS almanac length	0x03C2	$32 \times 31 = 962$
0x47				
0x48	U1	GPS PRN slot length	0x1F	Begin of almanac

Table 4. Block structure for GPS almanac (address 0x0049+0x001E*(PRN-1), big Endian)

Offset	Type	Content	Value	Comment
0x00	U1	PRN	0x01–0x20	
0x01	U1	Undeciphered	0x00	Health ???
0x02	U2	e—Eccentricity	$\tilde{e}$	$e = \tilde{e} \cdot 4.77\text{E}{-7}$
0x03				
0x04	U1	Undeciphered		
0x05	I2	i—Orbital inclination, (deg)	$\tilde{i}$	$i = 180{\cdot}(0.3 + \tilde{i} \cdot 1.91{\cdot}\text{E}{-6})$
0x06				
0x07	I2	dΩ/dt—Rate of right ascension W (deg/s)	$\dot{\tilde{\Omega}}$	$d\Omega/dt = 180 \cdot \dot{\tilde{\Omega}} \cdot 3.64\text{E}{-12}$
0x08				
0x09	U4	A—Semi-major axis (km),	$\tilde{A}$	$A = (\tilde{A} \cdot 4.88\text{E}{-04})^2$
0x0A				
0x0B				
0x0C				
0x0D	I4	$L\Omega$—Longitude of ascending node on 00h.00min.00sec. base date, (deg)	$L\tilde{\Omega}$	$L\Omega = 180 \cdot L\tilde{\Omega} \cdot 1.19\text{E}{-7}$
0x0E				
0x0F				
0x10				
0x11	I4	ω—Argument of perigee, (deg)	$\tilde{\omega}$	$\omega = 180 \cdot \tilde{\omega} \cdot 1.19\text{E}{-7}$
0x12				
0x13				
0x14				
0x15	I4	m—Mean anomaly (deg),	$\tilde{m}$	$m = 180 \cdot \tilde{m} \cdot 1.19\text{E}{-7}$
0x16				
0x17				
0x18				
0x19	I2	$af0$—Clock correction (s),	$a\tilde{f}0$	$af0 = a\tilde{f}0 \cdot 9.54\text{E}{-7}$
0x1A				
0x1B	I2	$af1$—Rate of clock correction $af0$, (sec/sec)	$a\tilde{f}1$	$af1 = a\tilde{f}1 \cdot 3.64\text{E}{-12}$
0x1C				
1x1D	U2	Reference time without rollover		Full GPS week 1st epoch for 2 days ahead
0x1E				

5 Results and Discussion

Using the layout of the binary data file outlined above, we obtained the GPS almanac. Comparison revealed good qualitative agreement with reference values of [5]. The algorithm for reading the GPS almanac from an A-GPS binary file as well as its program implementation are outside of the scope of this paper and straightforward, once all offsets and record formats for orbital elements are established. To our knowledge, the provided results in the form of GPS almanac are the first fruitful attempt to decipher A-GPS binary file of the proprietary gpsOneXTRA format. We discovered no similar results in the publications.

Considering [5], one can see that information on the reference GPS almanac helps quite a lot in revealing underlying parameters represented by the record fields of the binary A-GPS file. However, this problem is of moderate complexity, since the orbit elements are updated in slow pace in comparison to the ephemeris updates of A-GPS files. Deciphering the ephemeris, especially the predicted ephemeris, is a more complex challenge. The prediction of orbit depends significantly on the employed numerical method and precision of initial data. Moreover, the long-term (7–24 days) predictions of the ephemeris always suffer from precision degradation and can become incomparable at all.

Nevertheless, we plan to reveal record format for GPS ephemeris in A-GPS file of gpsOne web service as well as to reveal complementary information on ionosphere. Moreover, we consider to recover almanac and ephemeris formats for GNSS other than GPS that are contained in second version of the A-GPS file. The scope of our future work would also be concerned with file formats of other A-GPS services, mentioned in this paper. On all data formats are revealed, it is possible and plausible to create a program utility for data conversion between the formats. Among practical benefits we expect the increase of availability, versatility and robustness of the A-GPS services, providing shorter satellite acquisition times.

6 Conclusion

In the presented study, we considered the proprietary layout of a binary A-GPS file for Qualcomm gpsOne web service. Employing differential cryptanalysis in the form of quazi-known-plaintext attack, we deduced the structures of the header and GPS almanac entries for each operational satellite. The comparison of the deciphered almanac with reference data showed good agreement on the values of respective orbital elements. We also outlined the future study of more sophisticated GPS ephemeris entries along with ionosphere data in the A-GPS files of the said binary format. However, we should highlight possible limitations of the present approach to decipher A-GPS binary files like the use of block ciphers, especially with time dependent keys, which would demand substantially larger corpus of ciphertexts for successful differential cryptanalysis.

References

1. Qualcomm Press Release, Qualcomm Introduces gpsOneXTRA Assistance to Expand Capabilities of Standalone GPS, San Diego, 12 Feb 2007. https://www.qualcomm.com/news/releases/2007/02/12/qualcomm-introduces-gpsonextra-assistance-expand-capabilities-standalone
2. Nightwatch Cybersecurity, Advisory: Insecure Transmission of Qualcomm Assisted-GPS Data [CVE-2016-5341]. https://wwws.nightwatchcybersecurity.com/2016/12/05/cve-2016-5341/
3. Nightwatch Cybersecurity, Advisory: Crashing Android devices with large Assisted-GPS Data Files [CVE-2016-5348]. https://wwws.nightwatchcybersecurity.com/2016/10/04/advisory-cve-2016-5348-2/
4. GPS Date Calendar. http://navigationservices.agi.com/GNSSWeb/
5. Information and Analysis Center for Positioning, Navigation and Timing, Korolyov, Russia. https://www.glonass-iac.ru/en/GPS/ephemeris.php
6. Official U.S. government information about the Global Positioning System (GPS) and related topics, Interface Control Document ICD-GPS-240. https://www.gps.gov/technical/icwg/ICD-GPS-870A.pdf

Learning Effects of Different Learning Materials About Automated Driving Level 3: Evidence from a Propensity Score Matching Estimator

Maki Arame[1,2(✉)], Junko Handa[1,2(✉)], Yoshiko Goda[2], Masashi Toda[2], Ryuichi Matsuba[2], Huiping Zhou[3], Makoto Itoh[3], and Satoshi Kitazaki[4]

[1] Polytechnic University of Japan, Tokyo, Japan
{arame,j-handa}@uitec.ac.jp
[2] Kumamoto University, Kumamoto, Japan
{ygoda,matsuba}@kumamoto-u.ac.jp, toda@cc.kumamoto-u.ac.jp
[3] University of Tsukuba, Tsukuba, Japan
zhouhp@css.risk.tsukuba.ac.jp, itoh.makoto.ge@u.tsukuba.ac.jp
[4] National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan
satoshi-kitazaki@aist.go.jp

Abstract. There are always big issues to use new technology. The users need to know how to use it and what will happen if they do not use it properly because of lack of knowledge. Automated vehicles equivalent to driving automation levels 3 and 4 and advanced driving support systems in the levels are the same. Necessary knowledge and information to be acquired by drivers and pedestrians as well as effective educational methods should be identified. Furthermore, there are individual differences what one needs to know about automated driving as a driver and how to relate to an automated vehicle as a driver. In this research, safe driving education prototype contents were developed that can absorb differences in personal attributes such as learning styles, ages, and personal traits. This study examined how the traffic safety education, such as safety training for driving schools, should be provided as effective educational methods.

Keywords: Conditional driving automation · Career resilience · Propensity score matching

1 Introduction

Drivers and pedestrians should acquire necessary knowledge about driving automation before introducing levels 3 and 4 of driving automation, and advanced driving assistance system is on the load [1]. This study examined the effective educational methods using different learning materials. The driver is not only necessarily the "user" of automated vehicle, but also he/she should have a responsibility as a driver in a society. There are individual differences when one learns about automated driving and how to relate to an automated vehicle as a driver [2].

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_38

There has been no research on the trial production of safe driving education contents that can absorb differences in personal attributes. In this study, pilot training courses were developed for traffic safety education about driving automation and automated vehicle, for driving schools when they get renewal of driver's licenses. Thus, the contents of teaching materials should be taught about five minutes.

The teaching materials include driving automation levels and the roles of the driver when they drive an automated vehicle. This study focused on the differences of understanding based on individual attributes, such as age, learning style, and personality characteristics.

2 Literature Review

In recent years, attention has been focused on improving preparedness and resilience as risk reduction techniques for complex systems. Preparedness means preparedness in disaster risk management [3]. Based on the premise that completely protecting the area from disasters is impractical in terms of both technology and social cost, raising preparedness awareness and improving preparedness through disaster drills will improve preparedness [4, 5]. It has been suggested to lead to a similar concept as the word resilience, which means "elasticity" and "restoration." There are three types of research on resilience which emphasizes individual factors, environmental factors, and interaction. Research on resilience includes research that emphasizes individual factors, research that emphasizes environmental factors, and research that emphasizes interaction. However, it is important to increase resilience to reduce the negative consequences of risk factors. If the learning materials developed for the general public are able to raise traffic safety consciousness, it should be considered that differences in understanding due to individual attributes can be absorbed.

3 Purpose of This Study

This study investigated whether the three types of learning materials developed to raise traffic safety awareness had the effect of absorbing the difference in the degree of understanding due to personal attributes.

4 Method

4.1 Participants and Procedure

The survey data used in this study were a secondary analysis of the results of 3240 surveys conducted on the Internet in February 2019. In the survey, in addition to age and gender, learning styles and career resilience characteristics were surveyed as basic attributes of respondents. Tests were conducted before and after learning the basic items related to automated driving level 3 of driving automation. Based on the difference between the pre-test and post-test scores, when the group that scored higher was determined

the ascending group, and the group that did not rise or fell was determined the non-elevating group. Then, the age, gender, and learning style attributes of the ascending and non-elevating groups were identified.

If there is no difference, it is considered that the learning material has absorbed the difference in learning style. In addition, this study investigated the factors that reduce the difference in proficiency due to career resilience characteristics by examining the personal attributes of the ascending groups for each type of learning material.

4.2 Measurement

The same contents were created in pamphlet format, quiz format, and video format, and examined which learning materials were effective irrespective of the personal attribute. In the WEB survey (Fig. 1), learning style [6] and career resilience were surveyed as personal types. This study focused on career resilience [7].

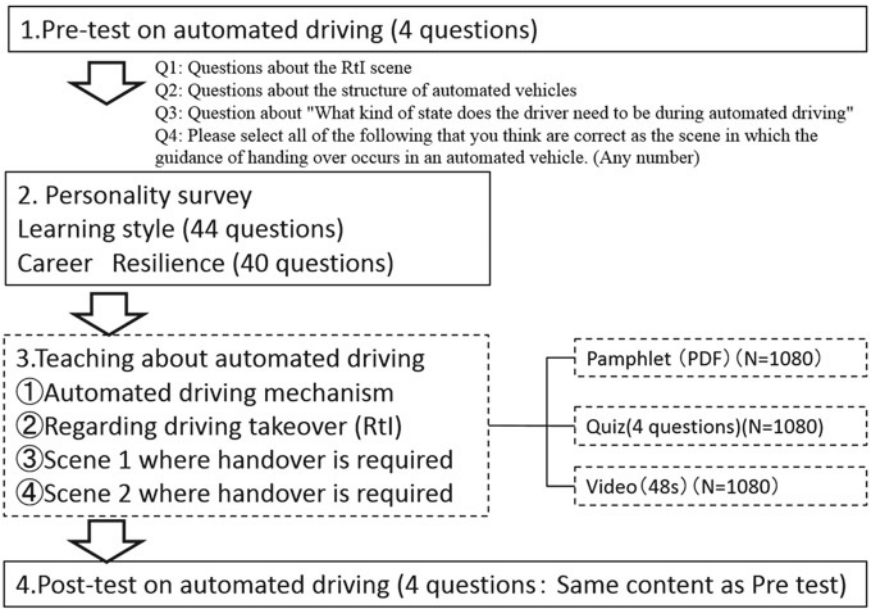


Fig. 1. WEB survey structure

Five measurements of career resilience were used based on Kodama [7] ("Factor 1: Challenge," "Factor 2: Social Skills," "Factor 3: Novel/Diversity," "Factor 4: Future-oriented," and "Factor 5: Assistance-oriented"). A total of 34 questions were asked, and answers were calculated for each item from a five-point scale, from "very much (5 points)" to "not at all (1 point)."

4.3 Data Analyses

In the analyses, in order to match the group whose score increased through learning (up group) with the group that did not increase (non-up group), the propensity score was calculated using variables that affect the probability assigned to the up group and the non-up group. Then, data were calculated and matched those who had a small difference in propensity score, and assigned and compared them to the ascending group and the non-ascending group.

5 Results

5.1 Results of Pre-post-tests by Learning Material Type

Table 1 shows the basic attributes of respondents. Table 2 compared the results of the pre-post-test for each teaching material type. Test results after learning were significantly higher, regardless of the type of learning material. The score increase rate of the pre-post-test for each teaching material type was 1.61 times for pamphlets, 1.67 times for quizzes, and 2.04 times for videos.

Table 1. Respondent attributes

			Pamphlet		Quiz		Video	
Age	Male	Female	Male	Female	Male	Female	Male	Female
20–35	153	667	52	218	51	219	50	220
36–45	285	525	88	182	95	175	102	168
46–60	477	333	164	106	152	118	161	109
61~	615	195	207	63	212	58	196	74
Total	3240		1080		1080		1080	

Table 2. Pre-post-test score comparison

		Pamphlet $N = 1080$	Quiz $N = 1080$	Video $N = 1080$
Pre-test	Mean	0.75	0.78	0.78
	SD	0.79	0.79	0.78
Post-test	Mean	1.21	1.30	1.59
	SD	1.10	1.02	1.10

$N = 3000$

5.2 Results of Pre-post-tests by Teaching Material Type

Factor 1: α (challenge $= 0.87$), Factor 2: social skills ($\alpha = 0.85$), Factor 3: novelty/diversity ($\alpha = 0.86$), Factor 4: future-oriented ($\alpha = 0.81$), and Factor 5: help-seeking preferences ($\alpha = 0.84$), and there was no problem with internal consistency. Table 3 compares the average scores of the learner's five factors of career resilience. There was no significant difference between the learning material types.

Table 3. Pre-post-test score comparison

		Pamphlet $N = 1080$	Quiz $N = 1080$	Video $N = 1080$
Factor 1: challenge	Mean	2.65	2.64	2.63
	SD	0.40	0.43	0.41
Factor 2: social skills	Mean	2.38	2.37	2.36
	SD	0.49	0.53	0.49
Factor 3: novelty/diversity	Mean	2.69	2.68	2.70
	SD	0.54	0.56	0.55
Factor 4: future-oriented	Mean	2.46	2.43	2.43
	SD	0.60	0.62	0.62
Factor 5: help-seeking preferences	Mean	2.81	2.79	2.81
	SD	0.60	0.60	0.60

5.3 Logistic Regression Analysis Results

In order to confirm how much career resilience is affecting learning, logistic regression analysis was performed for each teaching material type, and the results were compared. Table 4 shows the results of logistic regression analysis of pamphlet learning materials. R1 was statistically significantly affected at the 5% level, and a $+0.50$ change in R1 resulted in a 1.644-fold increase in the group. Similarly, R2 had a statistically significant effect at the 15% level, and a rise of -0.79 in R2 increases the increase by 0.45 times. In other words, when estimating the ascending group from the student's R1 score, it was increased by 1.644 times, and estimating the ascending group from the R2 score is 0.45 times. In the quiz materials shown in Table 5, R1, R2, and R4 have a statistically significant effect. The change in R1 $+0.70$ doubles the increase group, and the change in R2 -0.79 increases the increase group 0.45 times. In the video learning materials in Table 6, the R2 rises by -0.54, and the rise group becomes 0.45 times. It can be seen that the video learning materials were least affected by the career resilience factors.

5.4 Propensity Score Matching Estimator

After matching with propensity scores, selecting and assigning similar resilience characteristics between the non-up group and the up group eliminates or alleviates these tendencies, and reduces the number of statistically significant places.

Table 4. Pamphlet analysis results

	B	SE	Wald	Significant	Exp (*B*)
Factor 1	0.50	0.23	4.68	0.03*	1.64
Factor 2	−0.79	0.16	23.55	0.00**	0.45
Factor 3	0.25	0.16	2.38	0.12	1.28
Factor 4	−0.04	0.13	0.09	0.77	0.96
Factor 5	0.19	0.13	2.11	0.15	1.21

$N = 1080$
$^{*}p < 0.05$, $^{**}p < 0.01$ Cox–Snell R^2 0.03

Table 5. Quiz analysis results

	B	SE	Wald	Significant	Exp (*B*)
Factor 1	0.70	0.22	10.28	0.00**	2.00
Factor 2	−0.52	0.15	12.54	0.00**	0.59
Factor 3	−0.08	0.15	0.31	0.58	0.92
Factor 4	−0.24	0.13	3.52	$0.06^{\dagger}$	0.79
Factor 5	0.17	0.12	1.87	0.17	1.18

$N = 1080$
$^{*}p < 0.05$, $^{**}p < 0.01$ Cox–Snell R^2 0.03

Table 6. Video analysis results

	B	SE	Wald	Significant	Exp (*B*)
Factor 1	0.31	0.22	0.22	0.16	1.36
Factor 2	−0.54	0.16	0.16	0.00**	0.58
Factor 3	0.01	0.15	0.15	0.94	1.01
Factor 4	0.01	0.13	0.13	0.93	1.01
Factor 5	0.19	0.12	0.12	0.11	1.21

$N = 1080$
$^{*}p < 0.05$, $^{**}p < 0.01$ Cox–Snell R^2 0.01

Table 7 shows the results of pre-post-tests by learning material type after propensity score matching. After matching by the propensity score, the score difference is smaller than before. It is thought that video learning materials have the effect of absorbing personal characteristics other than career resilience.

Table 7. Propensity score matching results

		Pamphlet $N = 834$	Quiz $N = 1000$	Video $N = 984$
Pre-test	Mean	0.78	0.79	0.79
	SD	0.03	0.03	0.03
Post-test	Mean	1.41	1.32	1.53
	SD	0.04	0.03	0.04

$N = 2818$

6 Discussion

This study indicated that career resilience affected learning. It was shown that only Factor 2 had a significant effect on the video learning materials, and that other factors could be absorbed. Since social skills had a negative coefficient (B), it was thought that low social skills would not significantly affect the type of learning materials implemented this time. The following two points were found in this study.

1. When they study with pamphlets and quiz materials, it may be effective for them to enhance Factor 1: challenge such as problem solving and adaptability in advance by boosting their motivation.
2. Video materials were likely to be able to absorb individual attributes. Thus, using video materials may be more effective for elderly people.

Using video learning formats may also absorb characteristics of personal attributes other than career resilience. It will be necessary to continue to implement further studies and improve educational methods in the future.

7 Conclusion

In this study, the learning materials of traffic safety education used in the classroom for general citizens were provisionally developed in order to examine effective learning methods for different personal characteristics. Then, the study verified whether the learning materials could absorb personal characteristics. The video learning materials were least affected by the career resilience factors. As a result of using propensity score matching, it was confirmed that Factor 1: challenge, Factor 2: social skills, and Factor 4: future-oriented career resilience might affect the learning effect. Research on necessary knowledge about driving automation for drivers and pedestrians and the effective educational methods are urgent issues. Further research should be implemented continuously.

8 Limitations and Recommendations

There are limitations to our research design. In this survey, gender, marriage status, place of residence, etc., were investigated, but such demographic data have not been sufficiently examined.

In addition, only three types of learning materials—pamphlets, quizzes, and videos—were investigated in this study without considering interactive learning that is frequently used in these days. As further studies, using interactive learning materials, other new technologies such as AR and VR should be investigated if they are more effective for all types of learners.

Acknowledgements. This work was supported by Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), entitled "Human Factors and HMI Research for Automated Driving" (funding agency: NEDO).

The authors would like to thank Mr. Miki Akamatsu in Tokyo Business Service Co., Ltd. to support this project.

References

1. Zhou, H., Itoh, M., Kitazaki S.: How does knowledge about system limitations contribute to interventions into partial automation among elderly drivers? In: Proceedings of the 2018 IEEE international Conference on Systems, Man, and Cybernetics (SMC2018), pp. 851–820. Miyazaki, Japan (2018)
2. Zhou, H., Itoh, M., Kitazaki, S.: Long-term effect of experiencing system malfunction on driver take-over control in conditional driving automation. In: Proceedings of the 2019 IEEE international Conference on Systems, Man, and Cybernetics (SMC2019), pp. 1950–1955. Bari, Italy (2019)
3. Nemeth, P.C., Hollnagel, E. (eds.), Kitamura, M. (trans.): Guidelines for Application of Resilience Engineering: To Become a Resilient Organization. Nikka Giren Publisher, (2017)
4. Arame, M., Hideshima, E., Kanda, K.: Study on scaffolding effect of video learning materials in participatory learning for improving preparedness. JSCE D3 **68**(5), 35–43 (2012)
5. Arame, M., Tamaki, K.: Consideration of resilience improvement effect in educational organization management applying P2M. Proc. Int. Assoc. P2M **14**(1), 264–280 (2019)
6. Felder, R.M., Silverman, L.K.: Learning styles and teaching styles in engineering education. Eng. Educ. **78**(7), 674–681 (1988)
7. Kodama, M.: Examination of construct of career resilience and development of measurement scale. Psychol. Res. **86**(2), 150–159 (2015)

A Health Decision Support Framework for the Prediction of Cognitive Ability in Older People

Hui Yang(✉) and Peter A. Bath

The University of Sheffield, Sheffield S1 4DP, UK
hyang@sheffield.ac.uk

Abstract. Deterioration in cognitive functioning has become a serious public health burden due to aging of the population. Strategies for maintaining cognitive abilities with age are needed critically. The objectives of this study were: first to build a system that could accurately predict reduced cognitive function in older adults and; second, through this, to identify features that predict reduced cognitive function. Three tests of cognitive ability were investigated using data from the English Longitudinal Study of Aging (ELSA). Six machine learning algorithms were separately implemented in the system and their performance was compared in terms of the three cognitive tests. For each cognitive test, potentially important risk factors were identified as protective factors against cognitive aging. The findings from this study enhance our understanding of the underlying mechanisms that affect cognitive aging.

Keywords: Mental disorders · Cognitive science · Decision support systems · Gerontechnology

1 Introduction

The proportions and number of older people with cognitive impairment are increasing, due to the rapidly growing population of older people around the world. Mild cognitive impairment (MCI), which involves a slight but noticeable and measurable decline in cognitive abilities, has a dramatic impact on the personal lives of patients and caregivers, such as the loss of independence in daily activities in patients and increased caregiver burden. Persons with MCI are at increased risk of it progressing to Alzheimer's disease or other dementias. It is estimated that approximately 15–20% of people aged 65 or older have MCI, and the prevalence of MCI will continue to rise in the coming decades [1]. MCI in older age leads to substantial costs in social and economic resources. It has thus emerged as a major public health burden in later life, and strategies for maintaining cognitive functions with age are needed.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_39

Given the current absence of effective disease-modifying treatment and increasing awareness that cognitive decline develops over many years prior to clinical symptoms becoming evident, reducing the risk of MCI or delaying the onset of cognitive aging now has added importance. Identifying the risk and potentially protective factors associated with cognitive aging is therefore imperative.

Substantial evidence [2–5] suggests that lifestyle (e.g., alcohol intake and physical activity), social engagement, psychological factors (e.g., depression), and cognitive function in older adulthood are closely related. Moreover, observational studies [6, 7] have suggested that the deleterious role of cardiovascular (e.g., hypertension) and non-cardiovascular (e.g., arthritis) risk factors on the incidence and progression of cognitive disorders in older adults.

The contribution of our study to cognition aging research in older adults is demonstrated in two main perspectives: (1) it investigates cognitive functioning in older people using advanced data analytic approaches on the basis of a large national dataset, and (2) it identifies possible important psychosocial factors associated with cognitive aging in elderly people. We firstly investigated a wide diverse of variables from multiple domains. Secondly, two cognitive aging tests (i.e., *word recall* and *verbal fluency*) and one cognitive impairment test (i.e., *time orientation*) were studied to compare the difference of important factors in separate tests. Different machine learning algorithms were employed to build the predictive models.

2 Materials

2.1 Studying Data

This study was conducted based on a sample from the English Longitudinal Study of Aging (ELSA), a nationally representative cohort survey of people living in England aged 50 and above [8]. The cohort was drawn by postcode sector, stratified by health authority, and proportion of households in non-manual socioeconomic groups. The data used for our study comes from two recent wave surveys. The details about the data are described in Table. 1.

2.2 Measurement of Cognitive Function (Outcomes)

In this study, three cognitive tests were used as measures of cognition ability, which examine cognitive processes relating to daily functioning and are sensitive to age-related cognitive disfunction.

Word recall. In this test, participants were presented with a list of 10 common words aurally by the computer with every 2 s for one word. They were asked to name as many words as possible immediately and again after a short delay during which other tests were performed. Four 10-word lists were randomly selected and the count of correctly recalled words was calculated, higher scores indicating better memory (range: 0–20 words). The test has good construct validity and consistency [9].

Verbal (semantic) fluency. Respondent were asked to name as many animals as they could in 1 min [10]. The task score sums up the produced words that exclude non-animal words and repeated words. The observed range at baseline was from 0 to 50 named animals.

Table 1. Statistical information about three cognitive functions

	Wave-7 (2014–15) (Training)	Wave-8 (2016–17) (Test)
Instances	9666	8445
Age (years)	67.2 ± 10.1	68.8 ± 9.5
Sex		
Female	5368 (55.5%)	4695 (55.6%)
Male	4298 (44.5%)	3750 (44.4%)
Cognitive test scores		
Word recall	9.79 ± 4.62	9.79 ± 4.62
Time orientation	3.44 ± 1.31	3.46 ± 1.26
Verbal fluency	19.73 ± 8.89	20.23 ± 8.85

Time orientation. Time orientation with scores ranging from 0 to 4 was obtained from the Mini Mental Status Examination [11] and assessed using questions relating to the date, i.e., year, month, day of month and day of week, marking one score for each correct answer. In general, higher scores mean better cognitive performance.

Correlation between different cognition tests. Figure. 1 shows the correlation matrix between three cognitive tests obtained by Pearson's test. There was some correlation between the three cognitive tests. The correlation scores ranged between 0.65 and 0.70: the correlation between *word recall* and *verbal fluency* was highest.

	time_orientation	verbal_fluency	word_recall
time_orientation	1.000000	0.645164	0.663591
verbal_fluency	0.645164	1.000000	0.699580
word_recall	0.663591	0.699580	1.000000

Fig. 1. Correlation matrix between different cognitive tests

2.3 Cognition-Related Factors

A total of about 400 variables relevant to cognition aging were either extracted from the ELSA surveys or derived from the ELSA items related to one specific measure (e.g., `5 SWLS scale items -> Satisfaction with Life (SWLS)*`). The selected variables were drawn from several domains, such as demography and economy, social engagement and network, physical and mental health, and lifestyle. The details about these cognition-related factors are described in the previous work [12].

3 Methods

3.1 ML-Based Predictive Models

Different machine learning algorithms were applied to build predictive models for the purpose of performance comparison. They include two gradient boosting machine algorithms, light gradient boosting machine (LGB) and eXtreme gradient boosting (XGB), and one deep learning algorithm, keras-based convolutional neural network (K-CNN). In addition, some commonly-used ML algorithms for cognition prediction such as Lasso regression (Lasso), support vector regression (SVR), and linear regression (Linear) were also implemented in this study.

3.2 Experiment Metric

The outputs of the three cognition tests were treated as continuous variables, so that the prediction for each cognition test could be regarded as a regression task. For the experiment metrics, the root mean square error (RMSE) was used to measure the distance between observation points and predicted points, and is formulated below:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{0}^{n-1}\left(y_i - \hat{y}_i\right)^2} \tag{1}$$

where y_i is the observed value for the ith observation, and $\hat{y}_i$ is the predicted value. RMSE is frequently used in regression tasks to evaluate predictive results. Lower scores indicate better system performance.

At the training phase, ten-fold cross-validation was used for the training and validation of the predictive models.

4 Results

4.1 Algorithm Performance in Three Cognitive Functions

Table. 2 shows prediction performance with respect to different ML-based models in the three cognitive tests. The LGB worked the best on the test data in terms of the three cognitive tests, followed by the Lasso. The SVR performed relatively poorly on both *time orientation* and *verbal fluency* while the linear was the worst on *word recall* with the highest RMSE errors.

4.2 Feature Selection

As discussed above, a wide range of features from multiple domain areas were originally selected to investigate their utility in predicting cognitive function. Because not all factors have an influence on cognitive ability, we wanted to examine how the system performs when using the top-ranking features to build predictive models.

Table 2. System performance (RMSE scores) of different ML algorithms (full feature set)

	Word recall (validate/test)	Time orientation (validate/test)	Verbal fluency (validate/test)
LGB	3.103/3.141	0.612/0.625	6.249/6.319
XGB	3.070/3.217	0.623/0.638	6.204/6.358
K-CNN	3.159/3.258	0.636/0.628	6.396/6.379
Lasso	3.163/3.203	0.615/0.626	6.324/6.350
SVR	3.142/3.240	0.654/0.679	6.356/6.464
Linear	3.137/3.320	0.616/0.627	6.350/6.462

Several selected ML algorithms (e.g., LGB, XGB, and Lasso) have a built-in function that provides the importance score for each feature based on its effect on the prediction. Table. 3 lists the top-25 ranking factors in the three ML-based models with respect to the three cognition tests. It is shown that, in one particular cognitive test, the LGB and XGB have a high overlapping rate in the top-ranking features, but both of them have a feature list different from that of the Lasso, especially in the *time orientation* test. A possible explanation is because the computation of both LGB and XGB uses a GBM-based architecture that is distinct from that used by the Lasso. This implies that each ML algorithm acquires different cognition characteristics.

To further improve model performance, feature selection was conducted on the basis of the top-ranking factors from three ML algorithms. Different top-ranking feature subsets were generated, each of which merged the top-N factors in different algorithms (recall Table. 3).

Table. 4 presents the performance comparison between the different sets of top-ranking features in the individual cognitive tests. It is clear that feature selection based on the top-ranking features generally worked better with reduced RMSE errors than the full features in terms of three cognition tests (recall Table. 2). This implies that feature selection based on the top-ranking features is a useful step for refining the predictive models, as well as for reducing commonalities among features and minimizing the risk of bias. It is also observed that the optimal feature number for the best performance depends on the selected algorithms and the chosen cognition test. The optimal feature numbers for *word recall*, *time orientation*, and *verbal fluency* ranged from 44–79, 50–112, and 63–111, respectively.

5 Discussion

The evidence from our study indicates that the impact of one factor on cognitive performance might differ in individual cognitive tests. When a large number of influencing factors act together, some of the factors (e.g., `age`, `social engagement`) were consistently important, although others (e.g., `cardiovascular diseases`, `smoking`) became less important or even lost their importance. Moreover, some new factors such

Table 3. Top-25 ranking factors in the three ML-based models in terms of different cognitive tests

LGB	XGB	LASSO
Word recall		
Age	**Age**	**CES-D based depression***
Gait walk	**Gait walk**	**Gender**
Gender	**Living region**	Ethnicity
Social close: family	**Social support: all***	Transport: self-drive
Living region	**CASP-19 quality of life***	Volunteer: organize
Social: Volunteer work*	Social contact: all*	**Organization and Club: education**
CASP-19 quality of life*	**Transport: transport tools***	Volunteer: lead group
Social: Activity during last week*	**Social close: family**	**Gait walk**
CES-D based depression*	**Gender**	Unpaid work: look after
Social: Local amenity access*	Cohabitant	Physical activity: 4-tag*
Self-report hearing	**Social close: friends**	**Self-reported general health**
Food: fruit and veg*	**Satisfaction with life (SWLS)***	Depression: make effort
Social support: all*	**Social: Local amenity access***	**Social: Activity during last week***
Self-reported general health	Transport: public tool use	Social: Civic/culture activities*
Social: Unpaid work*	**Self-reported hearing**	**Social: Unpaid work***
Social: organization and club*	Social support: family*	Loneliness: in turn with people
Transport: transport tools*	**Food: fruit and veg***	**Age**
Social close: friends	Social contact: friends*	Physical activity: mild
Disability: ADL and mobility*	Marital status	Unpaid work: write
Food: vegetable portions	Social contact: family*	Physical activity: moderate
Transport: public tool use	Social support: friends*	Social: Volunteer work*
Local amenity: gallery	**Social: Volunteer work***	Eyesight: read newspaper
Social contact: child by msg.	**Self-reported general health**	Volunteer work: raise money
Alcohol: day no. in last 7 days	**Local amenity: gallery**	**Self-reported hearing**
Satisfaction with life (SWLS)*	Social contact: children*	**Social: organization and club***
Verbal Fluency		

(*continued*)

Table 3. (*continued*)

LGB	XGB	LASSO
Age	**Age**	**Ethnicity**
Gait walk	**Gait walk**	CES-D based depression*
Ethnicity	**Living region**	Transport: self-drive
Living region	CASP-19 quality of life*	**Organization and Club: education**
Social close: family	Social support: all*	Civic activities: holiday in UK
Local amenity: gallery	Social contact: all*	Volunteer work: other
Social: Volunteer work*	Cohabitant	Volunteer work: organize
Social: Activity during last week*	**Social close: family**	Volunteer work: lead group
Social close: friends	Transport: transport tools*	Unpaid work: look after
Social: Local amenity access*	**Social close: friends**	Transport: driving in the past
Social: organization and club*	**Social: Local amenity access***	Physical activity: 4-tag exercise*
Social: Civic/culture activities*	Satisfaction with life (SWLS)*	Physical activity: mild
Social: Unpaid work*	**Country of birth**	Volunteer work: campaign
Social close: children	Transport: public tool use	Mobility aid: wheelchair(manu)
Self-report eyesight	Marital status	Joint replace: right knee
Eyesight: read newspaper	**Ethnicity**	**Social: Activity during last week***
Alcohol: drink in last 12 mon.	**Local amenity: gallery**	Depression: feel depressed
Organization and Club: education	Gender	Mobility impairment: lift
Food: vegetable portions	Food: fruit and veg*	**Eyesight: read newspaper**
Disability: ADL and mobility*	**Social: Volunteer work***	Disease: Parkinson
Social contact: family by email	**Social support: family***	**Physical activity: moderate**
Country of birth	Food: vegetable	Volunteer work: education
Physical activity: moderate	**Self-reported general health**	Unpaid work: cooking
Local amenity: concert	Employment status	**Self-reported general health**
Social support: family*	Social contact: family*	**Gait walk**
Time Orientation		

(*continued*)

Table 3. (*continued*)

LGB	XGB	LASSO
Age	**Age**	Disease: non-CVD*
Gait walk	**Gait walk**	Disability: ADL*
CES-D based depression*	Living region	Disease: cancer
Social: Activity during last week*	Cohabitant	Disease: arthritis
Transport: transport tools*	**Transport: transport tools***	Disease: asthma
Food: fruit and veg*	Social contact: all*	**CES-D based depression***
Transport: self-drive	Social support: all*	Disease: osteoporosis
Social contact: friends*	**Social: Activity during last week***	Disease: psychiatric
Social: Civic/culture activities*	Transport: lift	Disease: lung
Transport: public tool use	**Gender**	Mobility aid: wheelchair(manu.)
CASP-19 quality of life*	**Social: Civic/culture activities***	Disease: Parkinson
Mobility impairment*	Self-reported general health	**Social: Unpaid help***
Limited work due to health	Marital status	ADL: bath
Hearing: problem in bkgrd. noise	Religious	Mobility aid: crutches
Mobility aid: wheelchair(manu)	Employment status	**Social: Activity during last week***
Gender	Country of birth	Disease: heart failure
Social: Volunteer work*	**CASP-19 quality of life***	Activities (last week): paid work
Food: vegetable portions	Satisfaction with life (SWLS)*	ADL: eat
Disease: CVD and non-CVD*	Pension	Pain: month
Social support: friends*	**CES-D based depression***	Civic and culture activities: other
SWLS: Change almost nothing	Age group	Joint replace: left knee
Physical activity: 4-tag*	**Transport: public tool use**	Mobility aid: wheelchair(electric)
Self-report hearing	**Food: fruit and veg***	Depression: sleep trouble
Eyesight: recognize friends	Mobility aid*	Mobility aid: walker
Social: Unpaid help*	Social close: family	Mobility aid: buggy

Note Variables appearing in two or more of the top-25 ranked factors are highlighted in bold; The asterisk (*) mark indicates the derived variables

Table 4. System performance (RMSE scores) of different ML algorithms at different top-ranking feature sets

Word recall							
	Top-10 (23 F)	Top-20 (44 F)	Top-30 (59 F)	Top-40 (79 F)	Top-50 (98 F)	Top-60 (119 F)	Top-70 (135 F)
LGB	3.167	3.122	3.050	**3.018**	3.038	3.043	3.042
XGB	3.122	3.052	**2.993**	3.018	3.022	3.037	3.043
K-CNN	3.104	**2.987**	3.054	3.062	3.066	3.070	3.079
Lasso	3.136	3.109	**3.106**	3.148	3.163	3.156	3.159
SVR	3.134	3.121	**3.113**	3.130	3.135	3.154	3.176
Linear	3.156	3.129	**3.121**	3.164	3.172	3.175	3.180
Time orientation							
	Top-10 (26 F)	Top-20 (50 F)	Top-30 (72 F)	Top-40 (92 F)	Top-50 (112 F)	Top-60 (131 F)	Top-70 (152 F)
LGB	0.603	0.586	**0.546**	0.553	0.553	0.563	0.564
XGB	0.627	0.606	0.573	**0.545**	0.562	0.573	0.572
K-CNN	0.627	0.608	0.564	0.542	**0.537**	0.553	0.568
Lasso	0.631	0.626	0.606	0.575	**0.558**	0.563	0.582
SVR	0.602	**0.585**	0.596	0.609	0.609	0.613	0.613
Linear	0.630	0.606	0.586	**0.556**	0.562	0.564	0.583
Verbal fluency							
	Top-10 (24 F)	Top-20 (42 F)	Top-30 (63 F)	Top-40 (81 F)	Top-50 (98 F)	Top-60 (111 F)	Top-70 (125 F)
LGB	6.320	6.308	6.207	6.175	6.155	**6.145**	6.169
XGB	6.299	6.204	6.177	**6.158**	6.163	6.188	6.183
K-CNN	6.366	6.312	6.206	6.161	**6.102**	6.135	6.154
Lasso	6.442	6.305	**6.234**	6.274	6.283	6.308	6.312
SVR	6.472	6.407	6.353	6.301	6.247	**6.212**	6.227
Linear	6.442	6.366	**6.237**	6.241	6.258	6.285	6.310

Note for individual algorithms, the best performance with the optimal features is marked in bold

as social well-being measures (e.g., `CASP-19 quality of life` and `Satisfaction with life (SWLS)*`) were firstly recognized as important by some of the selected ML algorithms.

Compared with previous studies that applied advanced ML techniques in the detection of cognitive impairment, our study differs in two main aspects: first, most previous studies focused on genetic, clinical, or neuroimaging data related to cognitive functioning, which are generally numerical. However, the data used for this study consider the

sociodemographic, health, interpersonal, quality of life, and subjective well-being variables, which contain a large number of categorical or binary variables. These differences in data characteristics requires some specific attention when suitable ML techniques are selected to build predictive models.

Second, several data mining techniques such as random forest, decision tree J48, SVM, and deep learning had been implemented on the cognitive-related applications. In this study, several ML algorithms were used to predict cognitive ability. Besides, the commonly-used algorithms (e.g., SVM, CNN-based deep learning, and Lasso) in the previous study of cognitive functioning, two new algorithms, the XGB and LGB, were firstly used in our predictive system for the detection of cognitive functioning. They worked competitively in cognition prediction.

6 Conclusion

In this paper, we used three cognitive tests to measure cognitive functioning in older age using a large nationally representative dataset, ELSA. Several ML algorithms were used to build predictive models for performance comparison. Feature selection based on the top-ranking factors was conducted to refine the predictive models. A subset of important risk factors was identified from a wide variety of features in multiple domains. This may help develop appropriate strategies to prevent or delay cognitive decline in old age.

There are several limitations in our current study on cognitive functioning in elderly people: first, this study emphasizes the identification of import influencing psychosocial factors in relation to cognitive functioning. However, the pathogenesis of cognitive impairment in late life is inherently complex, so should be explored in multiple domains. Second, in this study, the measure of cognitive ability in elderly people is restricted to the selected three cognitive tests, *word recall*, *time orientation*, and *verbal fluency*. However, there are other cognitive tests, each of which could reflect one specific perspective of cognitive function. Further research could investigate what significant factors might be closely associated with these cognitive tests using the advanced ML algorithms studied here. Third, the data used for the analysis of this study were cross-sectional and restricted to each of two single waves of the survey. Future work could develop a model to utilize information from consecutive waves to predict the probability of an individual's risk of cognitive decline over time.

References

1. Petersen, R.C.: Mild cognitive impairment. Continuum (Minneap Minn) **22**, 404–418 (2016)
2. Kornanong, Y., Khajonsak, P., Somrudee, H.: Physical activity and depression in older adults with and without cognitive impairment. Dement. Neuropsychol. **12**(1), 12–18 (2018)
3. Peters, R., Peters, J., Warner, J., Beckett, N., Bulpitt, C.: Alcohol, dementia and cognitive decline in the elderly: a systematic review. Age Ageing **37**, 505–512 (2008)
4. Hikichi, H., Kondo, K., Takeda, T., Kawachia, I.: Social interaction and cognitive decline: Results of a 7-year community intervention. Alzheimers Dement. **3**(1), 23–32 (2017)
5. Brailean, A., et al.: Late-life depression symptom dimensions and cognitive functioning in the Longitudinal Aging Study Amsterdam (LASA). J. Affect. Disord. **201**, 171–178 (2016)

6. Ligthart, S.A., van Charante, E., van Gool, W.A., Richard, E.: Treatment of cardiovascular risk factors to prevent cognitive decline and dementia: a systematic review. Vasc. Health Risk Manag. **6**(1), 775–785 (2010)
7. Veeranki, S.P., Downer, B., Jupiter, D., Wong, R.: Arthritis and risk of cognitive and functional impairment in older Mexican adults. J. Aging Health **29**(3), 454–473 (2017)
8. Steptoe, A., Breeze, E., Banks, J., Nazroo, J.: Cohort profile: the English longitudinal study of ageing. Int. J. Epidemiol. **42**, 1640–1648 (2013)
9. Baars, M.A., et al.: Predictive value of mild cognitive impairment for dementia. The influence of case definition and age. Dement. Geriatr. Cogn. Disord. **27**, 173–181 (2009)
10. Wu, Y.T., Prina, A.M., Jones, A., Matthews, F.E., Brayne, C.: The built environment and cognitive disorders: results from the cognitive function and ageing study II. Am. J. Prev. Med. **53**(1), 25–32 (2017)
11. Norris, D., Clark, M.S., Shipley, S.: The mental status examination. Am. Fam. Physician **94**(8), 635–641 (2016)
12. Yang, H., Bath, P.A.: The use of data mining methods for the prediction of dementia: evidence from the English longitudinal study of ageing. IEEE J. Biomed. Health (2019). (preprint)

Incorporating Stability Estimation into Quality of Service Routing in an SDN-Based Data Centre Network

Ayotuyi T. Akinola(✉), Matthew O. Adigun, and Pragasen Mudali

University of Zululand, Kwadlangezwa, RSA
ruthertosin@gmail.com

Abstract. Software defined networking (SDN) has emerged as effective paradigm in which the working principle was based on the separation of data plane from the control plane. This approach has proved to be quite advantageous in terms of its flexibility and programmability, thus enhancing better service provisioning and innovative network management. Despite the high results and improved expectations from the approach, one of the major challenges that the system faces is the level of network stability that the approach offers. We then tend to opined that based on our findings in literature, only few articles have addressed the level of network stability that SDN paradigm is offering as well as its improvements. Several approaches include basically the deployment of efficient routing protocols to ensure that the network attains some level of resilience. Even though the routing plays important role in the network environment, the control of the flow setup connection between the switches and controllers helps to instill better packet delivery with avoidance of network failure. We introduced a dynamic mapping orchestrator engine into the network setup which helps to unburden the controllers, as well determine the need for switching to either single, double or multiple mapping of switches to controllers. We evaluated the proposed approach by determining the average flow setup time, the QoS produced in terms of throughput and the fairness of the switching between the mappings of switches to controllers. The stability of the proposed approach was estimated to have been improved by 22% over the existing approach.

Keywords: SDN · Routing protocol · Controller · Switch mapping · Stability · QoS

1 Introduction

The past few decades have witnessed a fast growth in the networking world due to the amount of mobile devices that are connected together and the need to send information from one node to another. Several of these communications are wireless which are mostly an infrastructure-less network connections usually access one or two data storage somewhere for information update. These data storage reservoirs are referred to as

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_40

the data centres which are vascularized interconnections of several computer systems wherein several data of network users are kept for reference anytime [1]. They generally include pool of resources (network, storage, computational) that are interconnected via the communication network with redundant or backup components having capabilities for constant power supply, data communications, several environmental control measures and security procedures [1, 2].

Due to the vascularised environment that the data centre network created and as well as the need to ensure proper management of the environment, network administrators are vastly shifting to the adoption of the software defined networking (SDN) approach [3]. This novel approach advocates the separation of the data plane from the control plane. This separation leads to the realization of a centralized control system for the network management. The centralized system (controller) is described to be logical in such a manner that the management of the switch flow table wherein resides the rules for forwarding various kind of arriving network packets [4, 5]. With the innovation that SDN is expected to bring to the networking system which includes holistic enterprise management, high granular security, lower operating cost, hardware savings and reduced capital expenditures among others; several challenges are experienced in the course of implementing SDN especially in the case where there are new flow entries with no corresponding matching rule and the switch has to initiate setup request to the respective controller to take necessary action [6]. Some sort of packet-out messages are returned from the controller in response to the sorting out of the traffics. Within some short time frame, the reoccurrence of these events when not properly put to check results in instability via packet loss and path breaking. Most of the times, this delay or network instability is usually caused by the propagation delay between the responding controller and the targeted switch or the complexity of the routing algorithm which influences the accumulation of the load on the controllers [7].

With the literature having established the adoption of distributed controller architecture primarily to enhance that issues like latency, single point of failure and scalability are addressed, however, it introduces other challenge like how the flow traffics emanating from the switches are properly channelled to prevent flow setup delay and subsequent setup failures/breakages. Among the adopted approach for proper channelling of the traffics includes the one-to-one mapping of the switches to the controller. This is what is referred to as the single mapping scheme for switch to controller connections. The other approach is the mapping of the switch to at most two other controllers for failure purposes which is referred to as the double mapping scheme. The need for a more resilient network environment has warranted that a kind of multiple approach should be deployed between switches and controllers. This approach is regarded as the multiple mapping scheme which enables the switches to have several routes to access the controllers thereby avoiding failure or packet drops [8].

Several scholars have deployed these approaches to address some challenges in SDN-based networking environment; however, our view is to improve the stability of a network system by deploying a dynamic mapping orchestrator engine **(DMOE)** which monitors the proper channelling of the flow setup rules to ensure that stability is maintained in the system [9]. This is carried out by controlling the fluctuating traffic flow request when there is a possibility of failure and as well as permission of utilizing the appropriate mapping mechanism to be deployed in specific situation.

Hence, the major contributions of this paper are:

(1) Proposition of a dynamic switching of controller–switch mapping for an efficient flow setup within an SDN-based data centre network
(2) Implementation of a dynamic mapping orchestrator engine for monitoring and improved stability of control flows
(3) Deployment of an efficient way of resource utilization within the network environment.

With all these in view, the remainder of this paper is organized as follows. Section 2 discusses the problem formulation and its analysis. Section 3 discusses brief explanation of the flow analysis while Section 4 analysis discusses simulation setup experiments which explains the environment coupled with the implementation details for DMOE. The simulation results were presented and explained in Sect. 5 while Section 6 explains the related works. Section 7 concludes the paper.

2 Problem Formulation

We model the problem of network stability by considering the controllers connected in a distributed hierarchical manners with switches connected to them. All the switches are either connected via another switch to the controllers or connected directly. The master controller in SDN usually has the whole network-wide view that can be used to manage the network; however, our model description does not entirely depend on the master controller for management. We advance the network design through modelling an OpenFlow network that enables the switches to dispatch flow setup requests. We assume that by observing the network over a period of time frame, specific number of packet-in messages was considered. Also, we considered the fact that over a period of time frame, specific number of switches and controllers were considered. For example, when the flow request of 1000 fps (flows per seconds) is generated, we referred to it as an efficiently single mapping flows. The average of 20,000 fps will work efficiently with double mapping while any flow over and above 20,000 fps is recommended as multiple mapping.

The diagram in Fig. 1 shows a network design with network switches, controllers and the incorporated **DMOE** (dynamic mapping orchestrator engine) for the task of stabilizing the network system. We model the controllers as a queuing model using a batch-arrival process of $M^k/M/1$ [10]. Thus, the model flow request from switches

toontrollers h the arrival rate of λ_1 and λ_2 based on the specific controllers as revealed in Fig. 1. Furthermore, by focussing on the theoretical analysis of the controller flow request execution, we try to elaborate the probability of flow arrival process as:

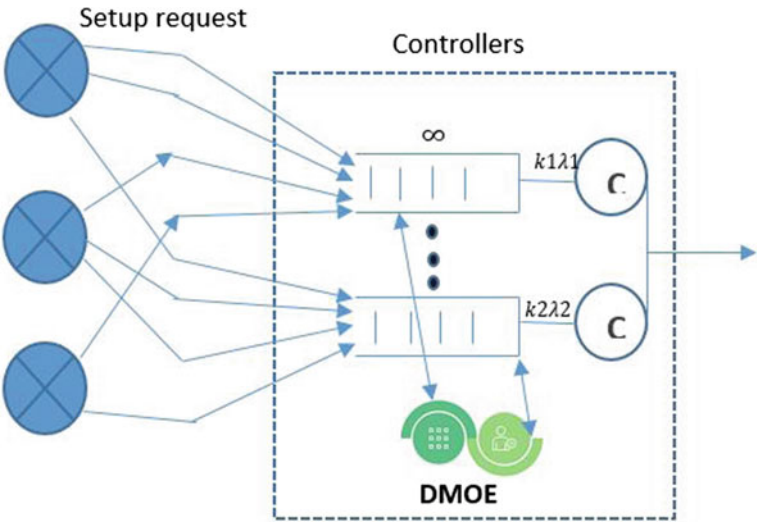


Fig. 1. Control flow setup modelling

$$P_{\text{nc}} = P_{\text{e}}\{\text{using an 'nc' number of controllers}\}$$
$$g_i = P_{\text{e}}\ \{\text{random batch of request size}\, i : 1 \leq i \leq k\}$$

We assumed that the batch sizes follow the pattern of Poisson distribution [11]; hence, it is represented in Eq. 1 as:

$$g_i = P_{\text{e}}\{X = i\} = \frac{\lambda^i}{i!}\text{e}^{-\lambda} \tag{1}$$

In Fig. 1, the mapping orchestrator engine is in charge of ensuring that the switching approach is correctly administered for the batch request size that is currently running in the network environment. The management is strictly controlled by an algorithm which is explained in Algorithm 1 as flow request processing.

Algorithm 1: Flow request processing.
Input: the set of switches managed by the hierarchical controllers **C**,
S: flow request sent from switch *s* and with matching field *f.match*.
rid ← *H*(*f.match*)
C ← LocateByRule(*rid*)
R ← *{orchestrator}*
for all *c* ∈ *C* do
if *c* is available then
R ← SearchRules(*c, rid*)
break
end if
end for
if *R* == *{orchestrator: Single or Double or Multiple}*
Then; *(Handle the flow request)*
Determine the appropriate mapping
end if
for all *r* ∈ *R* do
if *r.switch* ∈ *S* then
Install rule *r* at switch *r.switch*
else
c ← LocateBySwitch(*r.switch*)
Call controller *c* to install rule *r*
end if
end for

The above algorithm 1 enables the flow requests in the network to be properly managed such that the possibility of network congestion is avoided, having taken into consideration the mentioned assumptions in this work. The orchestrator is designed to check the average flow setup time of the environment as well as ensure that the performance of the network is stabilized using appropriate mapping in the course of packet transfers.

3 Flow Analysis

Acknowledging that there are *n* flows in pipeline to be processed via the based flow arrivals, each flow is not expected to be waiting indefinitely, but to allow the orchestrator to quickly determine the rate of flow as allotted to 10,000, 20,000 and above 20,000, respectively, for appropriate network mappings. Therefore, the flow request **S** is given by:

$$S = \sum_{i=1}^{k} \frac{i}{\mu} P_{\mathrm{e}}\{\text{for the } i\text{th flow in } k - \text{batch arrival}\}$$

$$= \sum_{i=1}^{k} \frac{i}{\mu} \cdot \frac{1}{k}$$

$$= \frac{k+1}{2\mu} \tag{2}$$

Considering the mean service time of the traffic request flow, we express this in Eq. 3 as:

$$\begin{aligned} T &= \sum_{n=0}^{\infty} E\{\text{service time}|\text{arrival queue is } n\}.P_{\text{e}} \\ &= \sum_{n=0}^{\infty} \left(\frac{n}{\mu} + \frac{k+1}{2\mu}\right).p_{\text{e}} \\ &= \frac{\lambda k.T}{\mu} + \frac{k+1}{2\mu} \end{aligned} \tag{3}$$

Combining the above expressions, we have Eq. 4 as a representation thus:

$$T = \frac{k+1}{2(\mu - \lambda k)} (\mu > \lambda k) \tag{4}$$

With the introduction of Little's law, the mean queue size of the OpenFlow network controller is otherwise expressed in Eq. 5 thus:

$$N = \lambda k.T = \frac{\lambda k(k+1)}{2(\mu - \lambda k)} (\mu > \lambda k) \tag{5}$$

The uniqueness of the mappings is accessed to encourage proper utilization of the network resources to their optimal performance. This enables the network to be more stable through appropriate mapping selection for the data flow request. The notations used in the above analysis were depicted in Table 1.

Table 1. The analysis representations

Symbols	Representations
N	Mean queue size
T	Mean flow service time
S	Mean flow processing time of the network controller
λ	Mean flow arrival rate from each switch
μ	Mean flow processing rate of the controllers
k	The number of accessible switches
σ	Flexibility constraint

4 The Simulation Setup

4.1 Environment

For proper analytic study of the performance of the proposed approach, we considered using network with 2200 switches and 500 controllers in OMNET++ simulator. We selected a random topology for the switch networks while using a hierarchical design for the controller setup. We chose the propagation delay between controllers and switches to range between [0.1, 1] ms. The controllers were set to have a processing capacity of 1500 packets-in messages per second. The packet-in message rate at every switch is randomly chosen to exist only within the range of [150, 550] packet-in messages per second. Now, we try to describe the numerical analysis that we conduct to show the performance in terms of optimal results from the deployment of dynamic mapping orchestrator engine **(DMOE).**

4.2 Performance Criteria

We generated random series of traffics and tried to deploy DMOE control. At each random series of traffics, we found the appropriate mapping solution for the traffics generated. The performance was tested under different traffic flow arrival rate, along with the fairness of traffic switching. We also evaluated the performance with respect to the throughput to determine how efficient the routing system has performed. Demonstrating the mean flow arrival rate among the three approaches to controller–switch mappings, considering the approximation of Eq. 4 when μ is varying between 10,000 and 50,000, in approximate to λ, then we have:

$$T = \frac{k+1}{2(\mu - \lambda k)} = \frac{\sigma}{2} \cdot \frac{1}{\frac{\mu+\lambda}{k+1} - \lambda} \approx \frac{\sigma}{2} \cdot \frac{1}{\frac{\mu}{k+1} - \lambda} \tag{6}$$

Thus, in this manner, we obtained series of numerical results for the single, double and multiple mapping solutions in which the graphical diagram is shown in Fig. 2. The fairness of the network is measured as the duration of traffic stability when there is no drastic variation or fluctuation in the deployed solutions to address the flow traffic. This is measured as a function of timestamps for non-fluctuating network flow despite varying traffic flow sizes from the switches.

5 Simulation Results

The first simulation results depicted the performances of the described single, double and multiple mapping approaches for traffic flow control. All the approaches achieved similar flow setup time for lower number of switches as depicted in the diagram but however began to show varying time durations with increasing number of traffic flows in the network. Multiple mapping enhances several alternative routes for packet routing though it comes with flow statistic messages such as reports, samplings and triggers for confirming the access and availability of alternative routes. This experiment did not consider the high influx of the statistic messages but strictly evaluated the quick rate of

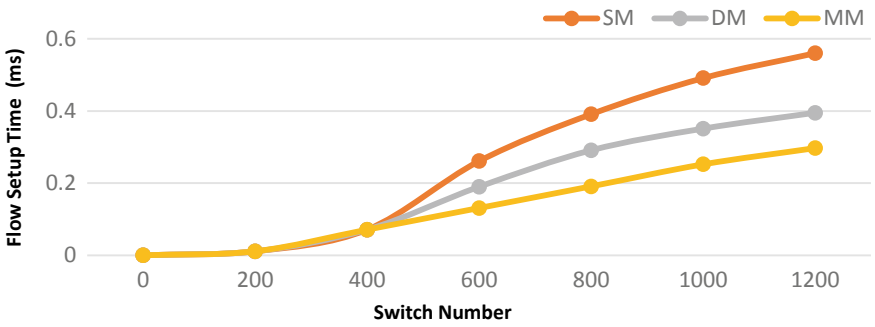


Fig. 2. The performance of mapping switching approaches

salvaging the congested situation through alternative routes. It was found that the single approach takes much longer time for flow setup to be achieved when compared with others. Moreover, the double also takes more time for setup than the multiple, as the multiple approach can provide several alternatives while still choosing the fastest route among the available ones.

The performance of the system is tested to access the level of improvement in the quality of service that is achieved in the orchestrator deployment. The OMNeT++ simulator enhances such quality to be measured directly without having issues. The throughput level speaks of the number of successful packet delivery in the kind of environment we are looking at here. Our expectation in this study is to have a comparative improvement on the existing system without the impact of the orchestrator engine. Using statistical explanation, the DMOE had a comparative average improvement of 20.75% in the number of successful or completed packet delivery than the system without it. This is not farfetched from the impact of the reduced number of control flow that is dynamically monitored by the orchestrator engine.

This proposed approach also ensures that the network resources are properly utilized for efficiency through determining the fluctuation in the network requirement by sudden increase in the network flow. For example, fewer resources are supposed to be used when there are fewer requests issued. In situation where only multiple mappings are implemented, the whole network resources are to be used even though less resources are needed. The DMOE dynamically attends to increase service requests accordingly for better quality of service enhancement and service delivery. Figure 3 depicts the improvement that accrued due to the deployment of the DMOE proposed in this work.

The last test was to verify the effectiveness of the DMOE being deployed in traffic switching to avoid the unnecessary flow statistic messages (control messages) that could result into network congestion. We also ensured that network resources were properly managed in the course of deploying the DMOE through effective network management. The diagram on Fig. 4 typically shows the impact of the DMOE towards regulated traffic flow management. We conducted the performance of the three mapping approaches and determined their average or mean rate of switch–controller mappings (S-C). This average was compared with the proposed DMOE performance as depicted in Fig. 4. We performed this average estimation because we envisaged that using each of the

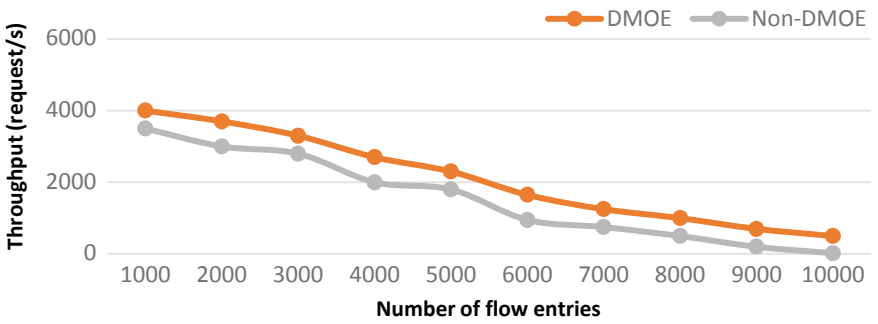


Fig. 3. The throughput performance of DMOE and non-DMOE

mappings are not supposed to give us the best result due to the fact that there are definitely fluctuations in the network requirements based on the level of network users. The DMOE-enabled network produces a reduced rate of switching performance which ranges from 0.06 to 0.11. The percentage rate of stability as depicted from this result is relatively equal to 54% compared to 76% from Non-DMOE approach. This implies that the use of dynamic mobile orchestrator engine was able to reduce network instability by 22% (which is close to 20.75% under throughput test) which we could by statistical estimation refer to as an improvement on the network stability.

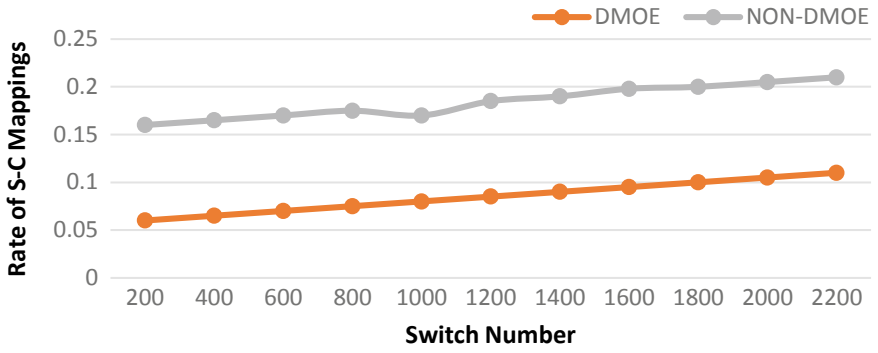


Fig. 4. The impact of DMOE deployment

6 Related Works

Several scholars have written diverse forms of articles on the subject of optimal network performance purposely to ensure that better uses experience is realized. These articles have cut across several fields in networking which ranging from MANET to WAN, ad hoc networks to sensor networks. However, the major challenge is that these proposals are network tailored to address mostly the nature of the intended network. The first work to be reviewed was the work of Weiying Zhu and others in [12]. The work proposed a

model for estimation of relative stability of paths and links of the network. In this work, a distributed QoS routing scheme called ticket-based probing (TBP) was enhanced. Stability factors were introduced to select the path with the longest lifetime based on the updated procedure for path discovery and section process. Four performance models were selected to test the better version of TBP called TBP-SE that was proposed. It was found that the efficiency of the proposed stability estimation model was higher in detecting paths with high stability for the packet routing process. However, this work was tailored for a MANET kind of network in which the network nodes are constantly changing positions. This is not adaptable to a data centre kind of network which is not migratory around the network environment.

We also reviewed the work in [13] which proposed a high-speed routing engine for improving the scalability of the OpenFlow controller. A hardware routing engine that uses an on-chip diorama network (a virtual emulated network in a chip) was used. A prototype of the routing engine was carried out on an actual dynamically reconfigurable processor. The results show that the rate of execution of the shortest path calculation was carried out at 19 times faster than the existing systems. The challenge with this work is that proper mechanism for ensuring the stability of the extended routes is needed to ascertain that this solution for extending the network is efficient. The work in [14] proposed a hybrid fault tolerance operation approach to improve the performance of SDN/OSPF networks. The approach here was to use a proper SDN portioning system to cater for sudden traffic bursts of inherent network failures to assure uninterrupted networks. Despite the portioning scheme proposed, the authors did not provide or speak to the level of stability that the proposed mechanism has to offer the considered network environment.

Brust and others in [15] proposed a heuristic for link stability in ad hoc networks. These networks are majorly dynamic communication networks in which the position keeps changing. A local topological information technique was solely deployed in the heuristic approach to identify links with higher or longer lifetime than other within the network environment. The approach deployed here is purely unsuitable for a data centre network that is not changing locations as mobile like ad hoc networks does.

The work in [16] discussed a mathematical analysis on stability of network control systems using industrial Ethernet ring network. The work presented a kind of sufficiency conditions for asymptotically stability for stochastic discrete linear time-variant system. The approach was simulated in MATLAB and the result according to the report agreed with the established theorem 1, therefore confirming the proposed approach. Lastly, the work of [8] used the option of multiple mapping alone for the network traffic control management, thereby controlling the flow setup in a network environment.

7 Conclusions and Future Works

The idea of centralized network control in SDN architecture seems to be the future direction that is considered and recommended for modern day networking environment. However, the increasing level of packet traffics in form of network setup flow which often increases at a rapid rate necessitates that appropriate approach or mechanism is deployed. Majorly, single, double and multiple mapping approaches were the techniques used in

switch–controller mappings. Several approaches have been deployed to enhance flow setup management but we have designed a dynamic mobile orchestrator engine **(DMOE)** to assist the network controller in the process. The DMOE regulated the management of network resources, reduces flow setup time and increases the performance of the network in terms of the throughput. We therefore hope to design a DMOE-tailored routing protocol for this proposed work as the generally used OpenFlow limits the performance of our proposed idea in this study.

References

1. Hammadi, A., Mhamdi, L.: A survey on architectures and energy efficiency in data center networks. J. Comput. Commun. **40**(1), 1–21 (2014)
2. Akinola, A.T., Adigun, M.O., Mudali, P.: Effects of scalability measures on control plane architectural designs in different data centre architectures. In: Proceeding of Southern African Telecommunication and Network Applications Conference (SATNAC 2018), pp. 198–203 (2018)
3. Tan, W., Zhang, J., Peng, C., Xia, B., Kou, Y.: SDN-enabled converged networks. IEEE Wirel. Commun. **21**(6), 79–85 (2014)
4. Jovicic, A., Wang, H., Viswanath, P.: On network interference management. IEEE Trans. Inf. Theory **56**(10), 4941–4955 (2010)
5. Zhang, B., Hao, J., Mouftah, H.T.: Bidirectional multi-constrained routing algorithms. IEEE Trans. Comput. **63**(9), 2174–2186 (2014)
6. He, M., Basta, A., Blenk, A., Kellerer, W.: Modeling flow setup time for controller placement in SDN: evaluation for dynamic flows. In: IEEE International Conference on Communications (ICC), pp. 1–7 (2017)
7. Song, P., Liu, Y., Liu, T., Qian, D.: Controller-proxy: scaling network management for large-scale SDN networks. J. Comput. Commun. **108**(1), 52–63 (2017)
8. Sridharan, V., Gurusamy, M., Truong-Huu, T.: On multiple controller mapping in software defined networks with resilience constraints. IEEE Commun. Lett. **21**(8), 1763–1766 (2017)
9. Naghmash, M.S., Alobidi, S.: Design and simulation of traffic controller in communication networks using MATLAB. J. Theor. Appl. Inf. Technol. **96**(17) (2018)
10. Jarschel, M., Oechsner, S., Schlosser, D., Pries, R., Goll, S., Tran-Gia, P.: Modeling and performance evaluation of an OpenFlow architecture. In: Proceedings of the 23rd International Teletraffic Congress, pp. 1–7 (2011)
11. Chen, Y., Mucchi, L., Wang, R., Huang, K.: Modeling network interference in the angular domain: interference Azimuth spectrum. IEEE Trans. Communications. **62**(6), 2107–2120 (2014)
12. Zhu, W., Song, M., Olariu, S.: Integrating stability estimation into quality of service routing in mobile ad-hoc networks. In: Proceeding of 4th IEEE International Workshop on Quality of Service, pp. 122–129 (2006)
13. Gao, S., Shimizu, S., Okamoto, S., Yamanaka, N.: A high-speed routing engine for software defined network. J. Sel. Areas Telecommun. (JSAT) 1–7 (2012)
14. Caria, M., Jukan, A.: Link capacity planning for fault tolerant operation in hybrid SDN/OSPF networks. In: Proceeding of Global Communications Conference (IEEE GLOBECOM 2016), pp. 1–6 (2016)
15. Brust, M., Ribeiro, C., Rothkugel, S.: Heuristics on link stability in ad hoc networks. In: IEEE/IFIP Network Operations and Management Symposium, pp. 738–741 (2008)
16. Zhang, Y.H., Li, D., Chu, Q.X., Guo, H.L., Zhang, Y., Jiang, L.: Stability analysis of networked control systems based on industry Ethernet ring networks. In: Seventh International Symposium on Computational Intelligence and Design, pp. 93–96 (2014)

Convergence Speed up Using Convolutional Neural Network Combining with Long Short-Term Memory for American Sign Language Alphabet Recognition

Busarakorn Supitchaya and Varin Chouvatut(✉)

Department of Computer Science and Graduate School, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand
busarakorn_sup@cmu.ac.th, varinchouv@gmail.com

Abstract. In sign language alphabet recognition problem, the scope of study limits only static hand gestures which not cover all gestures of sign language. This paper aims to find an approach for recognizing the static and dynamic gestures of American Sign Language (ASL) alphabet and apply GANs to generates synthetic images to increase dataset size. The proposed method combines convolutional neural networks (CNN) with long short-term memory (LSTM) networks to extract the features and classify images of the American Sign Language alphabet along various dimensions. With two consecutive images, this proposed method has an accuracy of over 97% and on 1D vector images, accuracy reaches 90% in large batch size when were tested on various batch sizes and epochs. Thus, this method is more appropriate for two consecutive images than on 1D vector images. For dynamic features, the performance of the proposed CNN-LSTM on two consecutive images is lower than the simple CNN at the beginning epoch, but the accuracy converged quickly, and finally, it reaches to the accuracy of simple CNN in a few epochs. Our proposed approach offers good results and better than simple CNN for dynamic ASL alphabet gestures, especially on 1D vector images.

Keywords: ASL recognition · Convolutional neural network · Data augmentation

1 Introduction

Communication is important for human because everyone needs to interact with other people. The communication may be either speaking, writing, or performing gestures. For people who impaired hearing and speech, sign language is a tool for communication in deaf community. People use sign language gestures to express their thought and emotions. Sign language is used by the symbolic hand, hand movements, body, or lips in expressions to communicate instead of speech. There are a few normal people who learned sign language, so it is difficult to understand the hearing impaired.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_41

Sign Language are widely language such as English, American, Indian, Chinese, sign language alphabet. Each sign language alphabet is different, and some character has movement but some character is static gesture. The problem of sign language recognition is not a new problem in computer vision which was previously developed by variety of sign language recognition methods. Including the research by [1] used hidden Markov model (HMM) to recognize facial expressions from video combining with Bayesian network classifiers. Hand detection using skin color is still having trouble of detecting hand area [2]. From past, studies were limited in static hand gesture which not cover all gesture of sign language.

This research is studying in the American Sign Language (ASL) recognition. The proposed technique combines CNN and LSTM to recognize gestures. The CNN-LSTM is trained for solving the recognition problem. Therefore, the CNN can extract static features, and LSTM is able to extract temporal features from sequences of image. For the dataset, we apply original images and augmented images of ASL dataset. A set of 24 ASL gestures is used to train and test the performance of the proposed method. The augmented image is synthesized from generative adversarial networks (GANs) which applied to increase the variation of images for improving the performance of classifier in learning. The comparison of simple CNN and our method on 1D vector images and the two consecutive images presented in our approach is better results than simple CNN, and our model is able to recognize the dynamic hand gestures.

Part of this study is divided into five parts as follows: Sect. 1 is an introduction, Sect. 2 explains an overview of the literature studies, and Sect. 3 performs the dataset and proposed method. The experimental results and discussion are shown in Sect. 4. Finally, the conclusion is summarized in Sect. 5.

2 Related Work

Convolutional neural network (CNN) is a kind of artificial neural network that has more complexity and layers for increasing performance to extract importance feature from images. The CNN model is applied to hand, gesture, and sign recognition. For example, gesture recognition using image processing and CNN method was proposed by [3]. The paper uses various image processing techniques to get the image of hand and uses CNN to extract feature from image and classify class. Work by [4] using NN combining with GA, EA, and PSO to recognize the Indian Sign Language.

Chinese Sign Language recognition is proposed by [5] performs how to extract feature of hand from video. After that, feed data to model for training. This method has been experimented on a set of Chinese vocabulary in daily life, the result was giving high accuracy due to the hand segmentation which is better, and it has a good avoidance of data loss in feature extraction. The paper of [6] uses deep CNN to classify images of ASL, both the letters and digits. In CNN, many parameters have affected to accuracy such as number of layers, number of filters, and batch size. Research by [7] presents the development and implement of an ASL fingerspelling translator based on CNN (excluding j and z). From experiment, found that changing the learning rate or structure of training model will affect little on the top-1 and top-5 accuracies, but the lack of variation in data will affect the accuracy and the process of image processing before recognition which will affect the classifier.

The paper by [8] presented Chinese Sign Language recognition using a long short-term memory (LSTM) network which a recurrent neural network (RNN). The dataset is recorded by Kinect 2.0 and focused on the trajectory of four skeleton joints. The proposed model can learn temporal information of the sign language by LSTM automatically. The experiments demonstrate the effectiveness of approach compared with HMM method. In ASL character recognition, it has both static and dynamic gestures. Therefore, only static gesture recognition is not covering all characters. So, [9] proposed the approach to recognizing ASL character using deep learning. The proposed method consists of two models, one is CNN model for recognizing static gestures and LSTM model for recognizing dynamic gestures. The work mentioned uses CNN for recognizing spatial features and RNN for recognizing temporal features on ASL dataset. They train and evaluate the CNN and RNN model independently, but using the same training and test samples. The paper by [10] is improving and developing the LSTM method for recognizing sequential gesture by dividing data into units and creating model with artificial neural network for considering each sub-unit. The experiment is tested on 942 Indian Sign Language sentences that use 35 different words. The average of accuracy on sentences and words were 72.3% and 89.5%.

3 Dataset and Proposed Method

The section explains a description of the dataset and preprocessing method for preparing the images data before fed into CNN-LSTM model. The architecture of the proposed method is explained in this section.

3.1 Generative Adversarial Networks

Generative adversarial networks (GANs), which is one of neural networks that comprised of two networks, provided by [11] have been a one of the interested topics in machine learning research in recent years. GANs are generative models that use deep learning methods. In GAN framework, the learning process consist of two networks, a generator model which generates synthetic data given a random noise vector and a discriminator model which discriminates between real data and the synthetic data that generated from generators model.

3.2 American Sign Language Dataset

In this paper, The American Sign Language dataset by [12] from Kaggle (excluding J and Z) is applied (see Fig. 1) and using GANs which image generator was applied to generate synthetic images and added the generated data to the dataset. The example of the synthetic images (see Fig. 2). Then, the data augmentation technique was applied to increase diversity of images data with zoom range, shear range, rotation, height shift, and width shift. Each image is in JPEG format and a size of 28×28 pixels. All images were converted from RGB to grayscale level because the tone of skin color, saturation, or lighting variation will affect to classification. To reduces the complexity of processing on our model, grayscale image is used in the training and test model.

Fig. 1. Example of ASL alphabet images dataset

Fig. 2. Example of synthetic images by GANs

Two consecutive images

In this work, the length of sequence frames was defined as two frames consecutively. For all characters (excluding J and Z), the first frame is the ordinary image, and the next

frame is the augmented image that was using the augmentation technique. The shape of a dataset in this paper should be (N, 28, 56, 1).

4 1D Vector Image

As two consecutive images are transformed to 1D vector images. Each pixel was mapped to an element in the vector, and each element was represented as a column. In this paper, the shape of dataset should be (N, 1, 1568) where N is number of samples, 1 is a row, 1568 is columns created by transforming two frames consecutively (see Fig. 3).

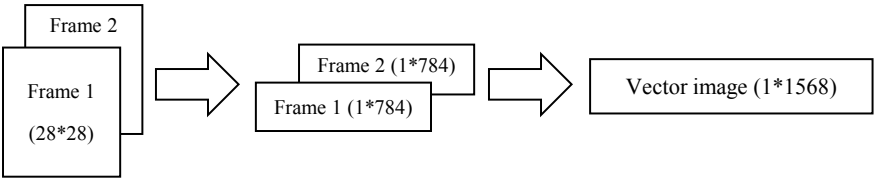


Fig. 3. Transforming two consecutive images to 1D vector image

4.1 Proposed Method

The proposed CNN-LSTM architecture consists of multiple convolutional, max pooling, batch normalization, dropout, and dense layers. The first part was a general CNN architecture, including three groups of convolutional layers with a ReLU function followed by batch normalization and max pooling. Next, a layer called flatten to convert data to a vector before connected to a group of the fully connected layer which followed by batch normalization and dropout layers. Then, connected to LSTM layers and followed by a group of the fully connected layer. The time-distributed layer is used to share the same weight between CNN and LSTM. The last dense layer is the output layer with a softmax function to predict gesture of ASL alphabet.

The CNN and LSTM are combined together. The CNN part extracted features from the frames and pass them through LSTM to predict temporal features. The overview of the proposed method ASL recognition. First, the input images are fed to train the CNN-LSTM model after that the model is build. Then, test the model by fed the test data to model. Finally, classify the data and give an output (see Fig. 4).

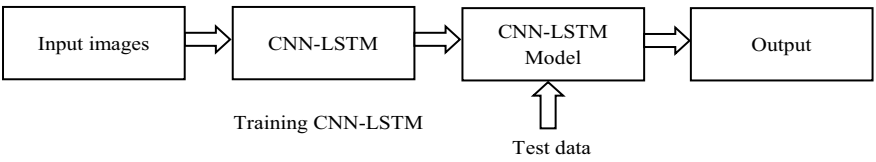


Fig. 4. Overview of proposed ASL recognition model

4.2 Evaluation Measures

To evaluate the performances, the proposed method was assessed by accuracy, precision, and recall. Accuracy is the percentage of correct classification using Eq. (1) to calculate the accuracy. Precision is the percentage of relevant results, and recall is the percentage relevant results that correctly classified by model as shown in Eqs. (2) and (3) where TP, FP, TN, and FN are true positive, false positive, true negative, and false negative, respectively.

$$\text{accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \times 100 \tag{1}$$

$$\text{precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{2}$$

$$\text{recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{3}$$

5 Experimental Results and Discussion

For the proposed CNN-LSTM model, the dataset includes 24 classes (excluding J and Z). In the experiments, the dataset was split in 70% for training, 30% for testing. Experiments were performed using Python, tensorflow, and Keras libraries on GPU GEFORCE GTX. Our model is evaluated on the ASL dataset and compared with different the number of batch sizes, epochs, and dimensions. Model is trained to minimize loss by cross-entropy cost function ADAM. We compare the performance of proposed method on different dimension of input either 1D vector images and two consecutive images. The experimental results were shown in Table 1. The accuracies of proposed model compared with simple CNN on 1D vector images, and two consecutive images were given in Table 2.

Table 1 presents the results of ASL alphabet recognition base on our method using 1D vector images and two consecutive images on dynamic gestures. In the results, we observed over 97% accuracy of two consecutive images, and the accuracy of 1D vector images was not high in small batch sizes, but in large batch sizes, the accuracy was better. Table 2 shows the accuracy of simple CNN compared with CNN-LSTM on both 1D vector images and two consecutive images. When tested on dynamic gestures, the performance of our method is better than simple CNN on large epoch and batch size, and we also observe that the various dimensions affected to accuracy rates.

From the overall of the experimental results showed that the training and test accuracies with different the number of epoch and batch size on two consecutive images are good performance, but the accuracy of 1D vector image are over 70% on large batch size. The training and test data will be shuffled by random, so training and test data are different in each of learning. We thought that the input dimensions, batch sizes, epochs, and network structure affected classification, and thus, we needed to find an appropriate model that addressed these parameters. To increase the dataset and remove overfitting problem, the augmentation technique is used to increase the diversity of images, and

Table 1. Result of our CNN-LSTM model for 1D vector images and two consecutive images recognition

Epoch	Batch size	1D vector images			Two consecutive images		
		Accuracy (%)	Precision	Recall	Accuracy (%)	Precision	Recall
5	32	36.9	0.41	0.37	98.9	0.99	0.99
	128	61.6	0.71	0.62	99.6	1.00	1.00
	512	69.1	0.73	0.69	99.8	1.00	1.00
20	32	63.0	0.71	0.63	99.8	1.00	1.00
	128	81.3	0.72	0.57	99.8	1.00	1.00
	512	81.9	0.85	0.82	100	1.00	1.00
50	32	75.5	0.78	0.76	100	1.00	1.00
	128	82.1	0.86	0.82	100	1.00	1.00
	512	93.0	0.93	0.92	100	1.00	1.00

Table 2. Result of simple CNN and CNN-LSTM for 1D vector images and two consecutive images recognition

Epoch	Batch size	1D vector images		Two consecutive images	
		CNN (%)	CNN-LSTM (%)	CNN (%)	CNN-LSTM (%)
5	32	38.8	36.9	98.5	99.3
	128	63.8	61.6	100	99.9
	512	70.2	69.1	99.9	99.7
20	32	48.3	63.0	99.8	99.9
	128	73.9	81.3	100	100
	512	70.3	81.9	100	100
50	32	73.4	75.5	99.9	99.9
	128	83.1	82.1	99.9	100
	512	89.6	93.0	100	100

GANs is applied to generate the synthetic images from base images. The new images are added into dataset to improve the model's accuracy and performance. We would classify a sequence of images, so the length of sequence is considered. If we picked all frames, it would have taken a long time to train, and some frame had no action, and thus, we needed to avoid them. In this research, we defined that the number of frames is two because the gesture has two movements as start and end gestures, thus only two frames sufficed to recognition.

The comparison of results with CNN-LSTM and simple CNN which the batch size and epoch are 512 and 50 as shown in Fig. 5. The performance of CNN-LSTM as Fig. 5a

converges to a peak fast and better than simple CNN on 1D vector images although the accuracy on first epoch is low, but on more epochs, the accuracy reaches 90%. As Fig. 5b, the comparison results of two methods on two consecutive images, the accuracy of CNN-LSTM is lower than simple CNN at the beginning epoch, but in a few epochs, it is able to converge quickly, and the performance is reach to simple. The dynamic gestures have various positions and orientations on each frame thus the CNN-LSTM is more flexible than simple CNN to recognize. From this study, the performance of two methods on two consecutive images is not different because the dataset was defined in a short sequence of frames, but on 1D vector image, our approach is achieving better results than simple CNN.

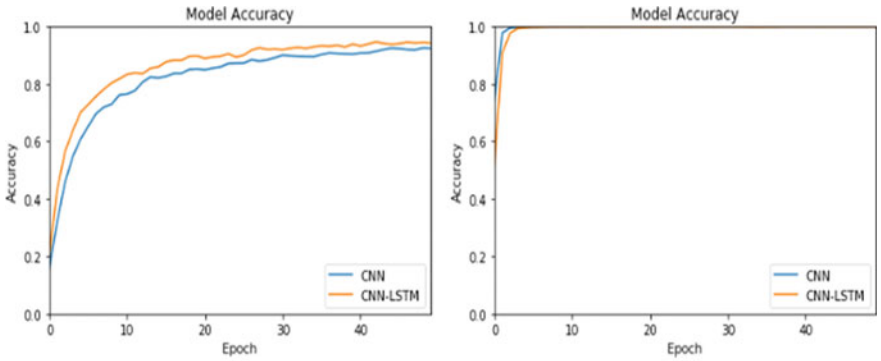


Fig. 5. Recognition accuracy with simple CNN and proposed CNN-LSTM

6 Conclusion and Future Work

This paper proposes CNN-LSTM model for American Sign Language alphabet recognition that have both of static and dynamic hand gestures. GANs were used to generate synthetic images to increase the size of datasets. The dataset consists of original data and synthetic data. For proposed method, CNN extract features from image input and pass them through the LSTM to predict a gesture. Time-distributed layers were defined in our model to enable movement prediction and recognition. The accuracy was over 97% with two consecutive images, while with 1D vector images, accuracy was lower with small batch sizes and higher with large batch sizes and batch size. Thus, this proposed approach is an efficient technique to learn both of static and dynamic ASL gestures. The comparison of simple CNN and our method on 1D vector images, and the two consecutive images demonstrated our approach achieving better results than simple CNN. In the future, we will study dynamic gestures recognition which long sequence of frames, as well as adding various noises and backgrounds to generate synthetic images from GANs or other image generation techniques.

References

1. Cohen, I., et al.: Facial expression recognition from video sequences: temporal and static modeling. Comput. Vis. Image Underst. **91**, 160–187 (2003)
2. Shahriai, S., et al.: Real-time American Sign Language Recognition using skin segmentation and image category classification with convolutional neural network and deep learning. In: 2018 IEEE Region 10 Conference, 2018, pp. 1168–1171
3. Salian, S., et al.: Proposed System for sign language recognition. In: Proceedings 2017 International Conference on Computation of Power, Energy Information and Communication (ICCPEIC), 2017, pp. 58–62
4. Hore, S., et al.: Indian Sign Language Recognition Using Optimized Neural Networks. Information Technology and Intelligent Transportation Systems. Springer, Cham, 553–563 (2017)
5. Yang, S., Zhu. Q.: Video-based Chinese Sign Language Recognition using convolutional neural network. Proceedings—2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN), 2017, pp. 929–934
6. Bheda, V., Radpour, D.: Using deep convolutional networks for gesture recognition in American Sign Language. ArXiv (2017)
7. Garcia, B., Viesca, S.A.: Real-time American Sign Language Recognition with convolutional neural networks. In: Convolutional Neural Networks for Visual Recognition, Stanford University, 2016
8. Liu, T., et al.: Sign language recognition with long short-term memory. In: Proceedings—International Conference on Image Processing, ICIP, 2016, pp. 2871–2875
9. Bantupalli, K., Xie, Y.: American Sign Language Recognition using deep learning and computer vision. In: Proceedings—2018 IEEE International Conference on Big Data, 2019, pp 4896–4899
10. Mittal, A., et al.: A modified LSTM model for continuous sign language recognition using leap motion. IEEE Sens. J. **19**(16), 7056–7706 (2019)
11. Goodfellow, I., et al.: Generative adversarial nets. Adv. Neural Inf. Process. Syst. (NIPS), 2014, pp. 2672–2680
12. Sign Language MNIST: http://www.kaggle.com/datamunge/sign-language-mnist, last accessed 2019/10/9

Pain Detection Using Deep Learning with Evaluation System

Kornprom Pikulkaew[1], Ekkarat Boonchieng[1,3], Waraporn Boonchieng[2], and Varin Chouvatut[1(✉)]

[1] Department of Computer Science and Graduate School, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand
varinchouv@gmail.com, champ_mekung@hotmail.com
[2] Faculty of Public Health, Chiang Mai University, Chiang Mai, Thailand
[3] Center of Excellence in Community Health Informatics, Chiang Mai, Thailand

Abstract.. Recent evidence has appeared that major enhancements in patient results can be increased by the periodical observing patient pain levels by medical staff in hospitals. Nevertheless, owing to the responsibility and pressure that the staffs have, this kind of observation has been complicated to withstand; thus, a system that works automatically could be the solution. Using an automatic facial expression system to detect pain which pain can be defined via several facial action units (AUs). To simplify pain detection using deep learning, data were collected from the UNBC database, which contains sequences of images that show participants' faces while they were doing an arrangement of movement-of-motion tests. To improve pain detection using facial expressions, this research proposes a pain detection technique that uses deep learning. Finally, the resulting of the experiment will be compared with the self-reports, and doctors will be asked to evaluate the system.

Keywords: Classification · Deep learning · Evaluation system · Facial expression · Human faces · Pain detection

1 Introduction

Usually, when someone gets sick, they go to the hospital, and the medical staff analyzes the patient based on their symptoms, which in some cases requires 24-hour monitoring and thus contributes to stress for staff. Following this method, the automatic pain detection method will be useful to doctors or nurses because it can detect pain although in a difficult situation and send an alarm to staff. The common technique for facial expression description and coding, facial action coding system (FACS), which is a measurement of facial movement [1].

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_42

FACS defined facial expressions through coding facial muscle activities into 44 action units (AUs). A few great works which can make a characteristic between distinctive human emotive conditions like impartial, unhappiness, worry, astonishment, rage, pleasure, and hatred have been achieved based on this procedure. Various related algorithms are found in Brahnam et al. [2–4]. The FACS provides a reliable measure of pain. The information received from this system can be used for training a real-time automatic system that offers a significant benefit in patient care and cost reduction. The most common database used to calculate pain detection algorithms is the UNBC. At first glance, it seems like some researchers can detect pain or recognize its intensity. This area of research is open for study, and several studies have been done to achieve a more reliable and robust system.

A sequential image of human faces has been used to recognize and classify pain in this method that identifies the moving faces and postures of humans [5]. Today, deep learning achieves outstanding performance on many important problems in image processing, machine learning, and language processing [6].

This study proposes use of the deep learning techniques for facial pain recognition and classification. It focuses on the facial expressions of patients with shoulder pain. The results are compared with the ground truth that got from medical staff.

2 Methodology

2.1 Pain Estimation and Intensity Metric

We used the Prkachin and Solomon Pain Intensity Metric (PSPI) [7, 8] guide estimation pain that requires FACS. The ache can evaluate an ordinal magnitude. Numerous endeavors have been made in human behavior that consider recognizing dependable and substantial facial markers of pain [9–12]. In this research, pain is generally categorized using the facial muscles and coded that matching AUs: brow lower (AU4), cheek raiser and lid tightener (AU6, AU7), nose wrinkle and upper lip raiser (AU9, AU10), and eyes closed (AU43) that kept much data about pain. Prkachin and Solomon presented the PSPI metric as Eq. (1) and Eq. (2):

$$\text{PSPI} = \text{AU4} + \max(\text{AU6 or AU7}) + \max(\text{AU9 or AU10}) + \text{AU43} \tag{1}$$

$$\begin{aligned}\text{Pain} = {} & \text{intensity(AU4)} + (\max \text{intensity AU6, AU7}) \\ & + (\max \text{intensity AU9, AU10}) + \text{intensity(AU43)}\end{aligned} \tag{2}$$

Except of AU43 which is two value (0,1), each activity is rated at 6-point magnitude (0 = unnoticeable, 5 = highest) using FACS. The PSPI calculations result in a 16 level. Figure 1 displays examples of pain intensity. Moreover, Fig. 2 illustrates a few of the FACS action units (AUs). Using (1, 2) and the 6 scales of the AU level, the PSPI could differ from 0 to 16.

Fig. 1. Examples the resultant pain level utilizing the PSPI metric

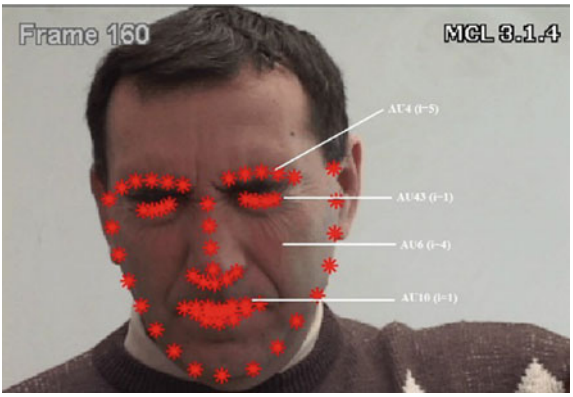


Fig. 2. Plot AAM landmarks to find AUs and calculate pain

2.2 Data Arrangement

The consecutive picture of the facial expression is carried from the UNBC dataset [13]. The UNBC includes about 50,000 images consist of 25 topics and 225 folders. Members consist of sixty-three men and sixty-six women ($\bar{x}$ age $= 42$ years, $SD = 14.5$) that they got a dilemma with shoulder pain. After each test, partakers evaluated the greatest soreness it had created utilizing the tactile and emotion descriptors and the visual analog scale (VAS).

Every human face activity is defined in terms of 1 of forty-four personal AU. Prkachin and Solomon [7] used AU4, AU6, AU7, AU9, AU10, AU12, AU20, AU25, AU26, AU27, and AU43 in this process. Apart from AU 25, 26, 27 and 43, for each activity was coded on five-stage intensity dimensions (1 = trace 5 = maximum).

2.3 Facial Detection and Deep Learning Method

For calculating the human face, the Dlib library by Davis King is used [14], which is up-to-date facial identification developed with deep learning. This technique has an accuracy of more than 99% upon the Labeled Faces in the Wild benchmark [15].

Training deep learning patterns can be hard and take time. The reason is how to find optimization value with learning rates. However, there is a straightforward method from statistics we can use to choose if the loss is still increasing, and when to decrease it. By

this method, it is trivial to fully automate the above technique. Moreover, the ResNet pre-model by Microsoft Research is utilized [16] that demonstrates a quick forecast time and high precision.

2.4 Evaluation System

The strategies that called self-report are commonly utilized but still undergo from disadvantages as reactivity to recommendations, endeavor at impression administration, duplicity, the distinction between patient's and clinician's perception of torment. To reply to questions associated with self-report methods, an observer rating scale (ORS) is implemented. The key problem of the observer rating scale is its impossibility and inability in situations where checking requires a long cycle of time like supervising an individual in hospitals. Table 1 summarizes how to classify pain.

Table 1. Summary of pain classification

Category/Type	No pain	Initial pain	Pain
ORS (0–5)	0–1	2–3	4–5
VAS (0–10)	0–3	4–7	8–10
AFF (0–16)	0–4	5–9	10–16
SEN (0–16)	0–4	5–9	10–16

Accomplishing successful pain administration needs cautious evaluation and standard audit of a patient's encounter of sickness. As illness could be a subjective side effect, and pain measurement instruments are typically depended on a patient's comprehension of their ache and its harshness. Several unique pain measurements are accessible, involving those for newborn children, kids, grown person and patients with communication troubles [17].

- Visual Analog Scale (VAS) is a unidimensional degree of pain magnitude in a grown person. For this scale, patients show their current intensity of discomfort on a 0–10 scale that begins with no pain at 0 and the most terrible pain at 10.
- Observer Rating Scale (ORS) is both unidimensional scales and multidimensional scales, and it connects to previous or present behavior and experiences. Medical staff can rate patient's pain using a straightforward scale started at 0–5, while zero is no pain, as well as five is the most pain [18].
- Affective Motivational Scale (AFF) is a specific difference variable that describes inclinations for regulating emotions. The scale ranges from 0 to 16 which illustrates increasing discomfort from the initial value with no pain at 0 to the final value indicating the most serious pain at 16 [19].
- Sensory Scale (SEN) is the aptitude to control and rating reactions to the sensory environment hence that responses to physical feedback are suitable to the requirements

of everyday life. In this magnitude, patients reveal their existing level of pain between 0 and 16 scale that begins with no pain at 0 and pain at 16 [20].

3 Result and Discussion

3.1 Categorization and Identification Using Deep Neural Network

Deep neural networks are exceedingly advantageous to discovering pain [21], and it is more extraordinary than the previous advanced techniques [22]. Some earlier experiments have used it to sort emotion [23] because of its satisfactory precision. For example, Zeng et al. [23] and Cohn et al. [24] research is focused on emotion detection that they can detect many states on feeling such as neutral, happiness, or unhappiness. After that, Lucey et al. who are the first to study pain. They used the active appearance model (AAM) to track face by manually labeled frames on people's faces and used support vector machines (SVM) to classify pain. This research, pain is categorized from human facial expressions. Moreover, it used to recognize the intensity of pain. We used 80% of sequential images for training and 20% for testing. Another factor to consider is the time complexity that we process on CPU intel i7 with ram DDR4 16 GB, and it takes more than 3 months to classify all pain class. Table 2 shows the results obtained.

3.2 Deep Learning Versus Evaluation System

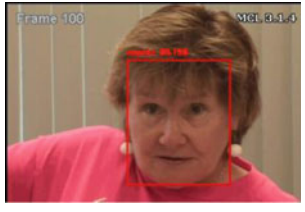
A warning system with sound effect needs to be added because we try to create a system that supports doctors and nurses in hospitals and to show that the system can be used in real life and meet the goal of the research. Furthermore, we used self-report and observation from a doctor/expert invited by the Center of Excellence in Community Health Informatics to evaluate the system. Table 3 summarizes the classification phase and ground truth.

4 Conclusions

According to research, the new method aimed at pain detection using deep learning was defined to resolve the hospital's and patient's problems. Human faces that sick from shoulder pain were used as a model to verify that the suggested technique efficiently performs in situations that cannot be answered by monitoring patients for more than 24 h. We have suggested a technique that improved permanence and correctness for pain detection using neural networks not only to situations where monitor cases but also to other complicated environments. Figure 3 summarizes the process of our system. The primary box is a pre-processing phase that we received from the UNBC. Next, the secondary box tried to train sequential images and classify pain. Finally, the tertiary box is the result that we got pain intensity.

As Table 3, the result shows that no-pain class, we can reach up to 98.75% accuracy together with pain class at 95.05%. However, the initial pain group still has some problems because the UNBC dataset has more data on the no-pain category than other classes, and this is a reason that why accuracy can drop down to 75.49%.

Table 2. Example of pain detection in human faces

Examples	Results	PSPI	ORS	VAS	AFF	SEN
	No pain	0	5	10	14	14
	Pain	10	5	10	14	14
	No pain	0	3	3	3	4

(*continued*)

Table 2. (*continued*)

Examples	Results	PSPI	ORS	VAS	AFF	SEN
	Initial pain	6	3	7	7	11
	No pain	0	3	6	7	10
	Initial pain	5	4	9	14	13

(*continued*)

Table 2. (*continued*)

Examples	Results	PSPI	ORS	VAS	AFF	SEN
	Initial pain	8	5	4	4	8

Table 3. Summary of classification compare with doctor diagnosis

Experiments	Ground truth	Accuracy (%)
(1) No pain	No pain	92.32
(2) Pain	Pain	95.05
(3) No pain	No pain	98.75
(4) Initial pain	Initial pain	79.87
(5) No pain	No pain	98.02
(6) Initial pain	Initial pain	92.98
(7) Initial pain	Initial pain	75.49
(8) No pain	No pain	96.49
(9) Pain	Pain	76.22

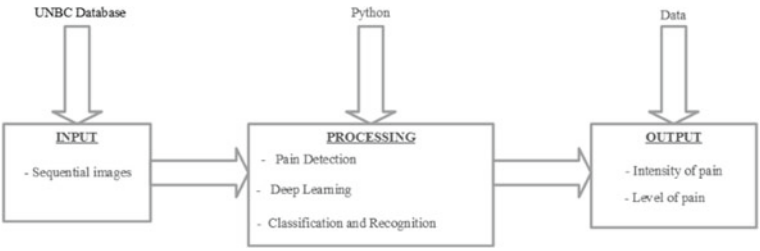


Fig. 3. Flowchart of the proposed method

As part of our future study, we intend to add sequential images because the level of pain between initial pain and pain is not balancing. The generative adversarial network (GAN) and imbalance techniques will be applied to use that can increase the precision of the method. Furthermore, our research can be applied to the detection of activities of humans and feelings.

Acknowledgements. This research was funded under the Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0226/2561) by the Thailand Research Fund.

References

1. Lucey, P., Cohn, J.F., et al.: Painful data: the UNBC-McMaster shoulder pain expression archive database. Image Vis. Comput. J. **30**, 197–205 (2012)
2. Brahnam, S., Nanni, L., Sexton, R.: Introduction to neonatal facial pain detection using advanced face classification techniques. Stud. Comput. Intel. **48**, 225–253 (2007)

3. Gholami, B., Haddad, W.M., Tannenbaum, A.R.: Relevance vector machine learning for neonate pain intensity assessment using digital imaging. IEEE Trans. Biomed. Eng. **57**, 1457–1466 (2010)
4. Huang, J., Xu, X. et al.: Emotion classification using deep neural networks and emotional patches. In: Bioinformatics and Biomedicine (2017)
5. Ashraf, A., Lucey, P., et al.: The painful face: pain expression recognition using active appearance models. Image Vis. Comput. **27**, 1788–1796 (2009)
6. Cornell University Homepage. https://arxiv.org/abs/1406.2661. Accessed 1 Jan 2020
7. Hammal, Z., Cohn, J.: Automatic detection of pain intensity. In: ACM International Conference on Multimodal Interaction, pp. 47–52. (2012)
8. Kaltwang, S., Rudovic, O., Pantic, M.: Continuous pain intensity estimation from facial expressions. In: International Symposium on Visual Computing, pp. 368–377 (2012)
9. Rathee, N., Ganotra, D.: A novel approach for pain intensity detection based on facial feature deformations. J. Vis. Commun. Image R. **33**, 247–254 (2015)
10. Hong, X., Zhao, G., Zafeiriou, S., Pantic, M., Pietikainen, M.: Capturing correlations local features for image representation. Neurocomputing **184**, 99–106 (2016)
11. Florea, C., Florea, L., Vertan, C.: Learning pain from emotion: transferred HoT data representation for pain intensity estimation. In: European Conference on Computer Vision: Workshop on Assistive Computer Vision and Robotics, 8927-LNCS, pp. 778–790 (2014)
12. Faisal, A. Eko, Y. Mauridhi, P.: A novel approach on classification of infant activity post-surgery based on motion vector. In: IEEE International Conference on Signal and Image Processing, pp. 213–218 (2017)
13. Lucey, P., Cohn, J., Prkachin, K., Solomon, P., Chew, S., Matthews, I.: Painful monitoring: automatic pain monitoring using the UNBC-McMaster shoulder pain expression archive database. Image Vis. Comp. **30**, 197–205 (2012)
14. Dlib Homepage. http://dlib.net. Accessed 1 Jan 2020
15. MIT Homepage. http://vis-www.cs.umass.edu/lfw. Accessed 1 Jan 2020
16. Cornell University Homepage. https://arxiv.org/abs/1811.05233. Accessed 1 Jan 2020
17. Prkachin, K.M., Solomon, P.E.: The structure, reliability and validity of pain expression: evidence from patients with shoulder pain. Pain **139**, 267–274 (2008)
18. LNCS Homepage. https://link.springer.com/chapter/10.1007/978–3-319-04627-3_2. Accessed 25 Dec 2019
19. NCBI Homepage. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2873215. Accessed 25 Dec 2019
20. NCBI Homepage, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4153553. Accessed 25 Dec 2019
21. Lauzon, F.Q.: An introduction to deep learning. In: Information Science, Signal Processing and Their Applications (2012)
22. X. Zhou, W. Gong et al., Application of deep learning in object detection, Computer and Information Science, 2017
23. Zeng, Z., Hu, Y., et al.: Audio-visual spontaneous emotion recognition, artificial intelligence for human computing (2007)
24. Cohn, J.F., De la Torre, F.: Automated face analysis for affective computing. In: IEEE International Conference on Automatic Face and Gesture Recognition (2015)

Cooperative Location Acquisition of Mobile Nodes Without Direct Distance Measurement

Masaaki Namekata and Hiroaki Higaki(✉)

Department of Robotics and Mechatronics, Tokyo Denki University, Tokyo, Japan
{namekata,hig}@higlab.net

Abstract. In wireless networks composed of numbers of mobile wireless nodes, their location information is required to be achieved for fundamental network services such as routing of data messages and determination of server nodes providing various services such as name services and for supporting network applications based on locations of the mobile wireless nodes, for instance, sensor network applications and ITS applications. Until now, various methods for achieving distances between two mobile wireless nodes have been proposed; however, some of them requires too expensive devices and others achieves too low-resolution results. Most of the methods using RSSI of transmitted wireless signals between two mobile wireless nodes cannot provide enough high resolution due to large deviation of transmission delay and the multi-pass problem. This paper proposes a novel method for achieving distances between mobile wireless nodes by cooperation of 3 mobile wireless nodes with cameras and a device for measuring its own migration length. The cameras are only used for measuring angles with high accuracy between 2 mobile wireless nodes which are independent of their shapes, sizes and distances.

Keywords: Relative location estimation · Distance between nodes · Mobile Wireless Networks · Measurement accuracy

1 Introduction

Recently, researches and developments of mobile ad-hoc networks (MANETs) where node-to-node wireless communication is available without help of wireless base stations have become active. Here, a wireless network consists of multiple mobile wireless nodes with wireless communication devices. For support of wireless network applications, fundamental network functions such as addressing, naming and directory services and routing of data messages are required to be provided. Location information of mobile wireless nodes is critical not only for realize such fundamental functions but also for various wireless network applications. Usually, mobile wireless nodes are small-size, light-weight and inexpen-

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_43

sively implemented especially in wireless sensor networks. Thus, highly precise location information of mobile wireless nodes are required to be achieved by using widely available devices mounted on them. Though various methods for remote measurements of distances between mobile wireless nodes such as distance estimation by using RSSI (Receive Signal Strength Indicator) have been proposed, it is said to be difficult to achieve enough precision of distance measurement due to vulnerability to environmental effects such as noise signals and multi-path propagation. This paper proposes a novel method for mobile wireless nodes to achieve highly precise relative locations of their neighbor mobile wireless nodes with robust and widely available measurements and cooperation among multiple neighbor mobile wireless nodes. One measurement is an angle between two observable neighbor mobile wireless nodes achieved by cameras mounted on them and the other one is migration distance of a mobile wireless node locally measured by using a simple and widely available device such as a rotary encoder. Results of measurements are exchanged among neighbor mobile wireless nodes and relative location information is achieved by calculation in each node.

1.1 Related Works

Some network applications for mobile ad-hoc networks require absolute location information such as latitudes and longitudes of mobile wireless nodes and the others require relative location information among them such as distances between two mobile wireless nodes and angles among three mobile wireless nodes. GPS (Global Positioning System) and QZSS (Quasi-Zenith Satellite System) provides absolute location information of mobile wireless nodes where distances between satellites whose absolute location information is given and each each mobile wireless node is measured for achieving its location information relative to the satellites and then absolute location information is induced from the relative location information. In most of currently available methods for achieving absolute location information, distances between two mobile wireless nodes and/or angles among three mobile wireless nodes are measured for achieving relative location information among neighbor mobile wireless nodes and absolute location information is induced by using the achieved relative location information and absolute location information of some anchor nodes whose absolute location information is given in advance or calculated in some trivial ways.

There are three well known methods for measurement of distances between two mobile wireless nodes; TOA (Time of Arrival), TDOA (Time Difference of Arrival) and RSSI [1]. In TOA, a distance between two mobile wireless nodes is measured based on transmission delay of wireless signals between them. Here, the transmission delay is achieved as the difference between the clock values in a sender node when it transmits a wireless signal and in a receiver node when it receives the wireless signal. For precise measurement, these two local clocks are required to be precisely synchronized; however, such clock synchronization is difficult to be implemented in mobile wireless networks due to unpredictable transmission delay of wireless signals caused by collisions and contentions in wireless LAN protocols such as IEEE802.11. Thus, some methods measure not one-way

transmission delay but round-trip transmission delay for avoidance of the clock synchronization problem. In TDOA, two signals whose propagation velocities are different are simultaneously transmitted between two neighbor mobile wireless nodes for measure of distances between them. In [2], a method using different wireless signals, i.e. electro-magnetic wave and supersonic wave, has been proposed. Here, no clock synchronization between the mobile wireless nodes is required; however, transceivers for different wireless signals are required to be equipped in each mobile wireless node. In RSSI, transmission power and receipt power of wireless signal in a sender and a receiver mobile wireless nodes are measured, respectively. Then, the distance between these nodes is estimated in accordance with a propagation model of wireless signals. However, it is said to be difficult for RSSI to achieve highly precise measurements of distances due to various environmental effects such as existence of obstacles, multipath propagation of wireless signals and characteristics of antennas mounted on mobile wireless nodes [3]. In all these methods, remote measurements of distances between two mobile wireless nodes with simple and widely available devices cannot provide highly precise results and it is also difficult to achieve highly precise relative location information among multiple mobile wireless nodes.

Fig. 1. Picture achieved by camera mouted on mobile wireless node.

Kato et al. [4] has proposed a method for achieving relative location information among mobile wireless nodes based on measurements of angles among three neighbor nodes by using pictures taken by cameras mounted on the nodes as shown in Figs. 1. The angle is expected to be measured highly precisely since the directions of neighbor mobile wireless nodes are determined precisely independently of the distances to them from the node taking pictures by a camera. However, it is impossible to estimate the distances to the nodes in the pictures from the node taking them only by the pictures. In [4], the angles measured by

mobile wireless nodes are gathered, combinations of three angles whose summation equals 2π are detected, which results in detection of sets of three mobile wireless nodes configuring a triangle. Thus, relative location information without its scale is achieved without help of identification of mobile wireless nodes in the pictures taken by them. In order to get the scale, at least one distance between two neighbor mobile wireless nodes is required to be measured. Especially in a mobile wireless network with sparse distribution of mobile wireless nodes, many distances between two neighbor mobile wireless nodes are required to be measured, which makes difficult to assure the precision of achieved location information.

2 Proposal

This paper proposes a novel method to achieve relative location information among multiple mobile wireless nodes without remote measurements of distances between two neighbor mobile wireless nodes in which unacceptable measurement errors are inevitable. Here, it is assumed that each mobile wireless node equips a camera for achieving directions to neighbor mobile wireless nodes in its visual field in order to get an angle made by two neighbor mobile wireless nodes and a device to measure its own migration distance such as a rotary encoder. In addition, in order to exchange the achieved angles and distances among neighbor mobile wireless nodes, a wireless communication interface with a certain wireless LAN protocol is also equipped in each mobile wireless node. Here, same as [4], a camera in each mobile wireless node is used only for measurements of angles made by two neighbor mobile wireless nodes. On the other hand, each mobile wireless node is required to get only its own migration distance between given two time instances and is not required to get its migration direction. Here, it is enough for a mobile wireless node with wheels for its mobility to equip rotary encoders for measurement of a migration distance where its path is almost straight.

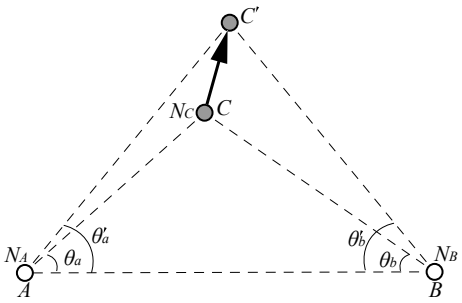


Fig. 2. Calculation of Distance between Nodes by Proposed Method.

Figure 2 shows assumptions for our proposal. Three mobile wireless nodes N_A, N_B and N_C are neighbor ones which are mutually included in their visual

fields each other and are mutually included in their wireless signal transmission ranges, i.e. these three mobile wireless nodes can communication one another in accordance with a wireless LAN protocol such as IEEE802.11. These three nodes exchange control messages for synchronization among them and N_A and N_B suspends their mobility at locations A and B, respectively, at time instance t. Now, N_A and N_B take pictures by their mounted cameras and measure angles $\theta_a := \angle CAB$ and $\theta_b := \angle CBA$ where C represents the current location of N_C at t. After that, only N_C continues to move, i.e. N_A and N_B keep their locations with their suspension of mobility and reaches another location C' at a certain time instance $t' > t$. Now, N_A and N_B take pictures and measure angles $\theta'_a := \angle C'AB$ and $\theta'_b := \angle C'BA$ where C' represents the current location of N_C at t'. During this mobility of N_C, N_C measures its own migration distance $|CC'|$ in the time interval $[t, t']$. These measured angles and distances are exchanged among N_A, N_B and N_C by using wireless ad-hoc communication and three distances $|AB|$, $|BC'|$ and $|C'A|$ are calculated by the following method. After the measurements at t', N_A and N_B restarts their own mobility.

Regarding triangles $\triangle ABC$ and $\triangle ABC'$, the following equations are satisfied due to the law of sines:

$$\frac{|AB|}{\sin(\theta_a + \theta_b)} = \frac{|CB|}{\sin\theta_a} = \frac{|CA|}{\sin\theta_b} \tag{1}$$

$$\frac{|AB|}{\sin(\theta'_a + \theta'_b)} = \frac{|C'B|}{\sin\theta'_a} = \frac{|C'A|}{\sin\theta'_b} \tag{2}$$

Then, regarding triangles $\triangle ACC'$ and $\triangle BCC'$, the following equations are satisfied due to Eqs. (1) and (2) and the law of cosines:

$$\begin{aligned} |CC'|^2 &= |CA|^2 + |C'A|^2 - 2\,|CA|\,|C'A|\cos(\theta'_a - \theta_a) \\ &= \left(\frac{\sin^2\theta_b}{\sin^2(\theta_a + \theta_b)} + \frac{\sin^2\theta'_b}{\sin^2(\theta'_a + \theta'_b)} \right. \\ &\quad \left. - \frac{2\sin\theta_b \sin\theta'_b \cos(\theta'_a - \theta_a)}{\sin(\theta_a + \theta_b)\sin(\theta'_a + \theta'_b)} \right) |AB|^2 \end{aligned} \tag{3}$$

$$\begin{aligned} |CC'|^2 &= |CB|^2 + |C'B|^2 - 2\,|CB|\,|C'B|\cos(\theta'_b - \theta_b) \\ &= \left(\frac{\sin^2\theta_a}{\sin^2(\theta_a + \theta_b)} + \frac{\sin^2\theta'_a}{\sin^2(\theta'_a + \theta'_b)} \right. \\ &\quad \left. - \frac{2\sin\theta_a \sin\theta'_a \cos(\theta'_b - \theta_b)}{\sin(\theta_a + \theta_b)\sin(\theta'_a + \theta'_b)} \right) |AB|^2 \end{aligned} \tag{4}$$

Therefore, distance $|AB|$ between N_A and N_B is calculated as follows:

$$\begin{aligned} |AB| &= \left(\frac{\sin^2 \theta_b}{\sin^2(\theta_a + \theta'_b)} + \frac{\sin^2 \theta'_b}{\sin^2(\theta'_a + \theta_b)} \right. \\ &\quad \left. - \frac{2 \sin \theta_b \sin \theta'_b \cos(\theta'_a - \theta_a)}{\sin(\theta_a + \theta_b) \sin(\theta'_a + \theta'_b)} \right)^{-\frac{1}{2}} |CC'| \\ &= \left(\frac{\sin^2 \theta_a}{\sin^2(\theta_a + \theta_b)} + \frac{\sin^2 \theta'_a}{\sin^2(\theta'_a + \theta'_b)} \right. \\ &\quad \left. - \frac{2 \sin \theta_a \sin \theta'_a \cos(\theta'_b - \theta_b)}{\sin(\theta_a + \theta_b) \sin(\theta'_a + \theta'_b)} \right)^{-\frac{1}{2}} |CC'| \end{aligned} \tag{5}$$

For this calculation, $\theta_a + \theta_b \neq 0, \pi$ and $\theta'_a + \theta'_b \neq 0, \pi$ should be satisfied. That is, at both time instances t and t', all N_A, N_B and N_C are not on one line. In addition, $|BC'|$ and $|C'A|$ are also calculated by using the result $|AB|$ in (5) and the equations in (2).

In our proposed method, distances between two neighbor mobile wireless nodes are never remotely measured as in most conventional methods for avoidance of unacceptable measurement errors and much more precise distances are expected to be achieved since angles made by two neighbor mobile wireless nodes in a visual field of another mobile wireless node measured by using pictures taken by a camera are expected to be more precise due to their independence of the distances. In addition, though our proposed method requires mobile wireless nodes to measure its migration distance during a specified time period, it is measured locally, i.e. directly, not remotely by using a simple and widely available devise. Furthermore, the direction of the mobility is not required to be measured, which is an important property of our proposed method.

The proposed method is based on cooperation among three neighbor wireless nodes and the assignment of roles in the method is allowed to be arbitrary. Thus, only one of the three nodes is required to possess a device for measurement of its own migration distance and the others are required to possess a camera and to suspend their mobility temporarily. Hence, each mobile wireless node is not required to equip both a camera and a rotary encoder and the assignment of the roles can be determined in an on-demand manner based on their possessed devices.

Control messages carrying measured angles and distances and ones carrying calculated distances can be transmitted in one-hop wireless ad-hoc communication or in wireless mulithop transmissions with help of multiple intermediate nodes forwarding these control messages. In addition, the calculation of distances can be carried out in a centralized manner by one of the mobile wireless nodes or in a distributed manner by all the three ones.

For example, a synchronization protocol among two mobile wireless nodes which suspend their mobility and another one which measures its migration distance during their mobility suspension under assumptions that these three nodes are mutually included in their wireless signal transmission ranges, i.e., these

nodes can communicate directly, and that these nodes are mutually observed by using their cameras, i.e., these nodes are included in their visual fields one another, is as follows:

[**Cooperative Distance Measurement Protocol**]

(1) A mobile wireless node S_1 suspends its mobility temporarily and broadcasts a distance measurement request control message $Dreq(S_1, S_2, M)$ to all its neighbor mobile wireless nodes in its wireless signal transmission range.
(2) A mobile wireless node S_2 which receives $Dreq(S_1, S_2, M)$ from S_1 also suspends its mobility temporarily, measures an angle $\angle S_1S_2M$ and transmits a distance measurement reply control message $Drep(S_1, S_2, M)$ to S_1.
(3) A mobile wireless node M which receives $Dreq(S_1, S_2, M)$ from S_2 begins measurement of its migration distance.
(4) On receipt of $Drep(S_1, S_2, M)$ from S_2, S_1 measures an angle $\angle S_2S_1M$.
(5) When a predetermined time interval T passes after the receipt of $Drep(S_1, S_2, M)$, S_1 broadcasts a measurement result request control message $Mreq(S_1, S_2, M)$ to all its neighbor mobile wireless nodes in its wireless signal transmission range.
(6) On receipt of $Mreq(S_1, S_2, M)$ from S_1, S_2 measures $\angle S_1S_2M$ and transmits a measurement result reply control message $Mrep(S_1, S_2, M)$ carrying the two results of measurements of $\angle S_1S_2M$ to S_1.
(7) On receipt of $Mreq(S_1, S_2, M)$ from S_1, M terminates measurement of its migration distance measures and transmits $Mrep(S_1, S_2, M)$ carrying the result of measurement of its migration distance.
(8) On receipt of $Mrep(S_1, S_2, M)$ from S_2, S_1 measures an angle $\angle S_2S_1M$.
(9) After receipt of both $Mrep(S_1, S_2, M)$s from S_2 and M, S_1 calculates three distances $|S_1S_2|$, $|S_1M|$ and $|S_2M|$ and broadcasts a result notification control message $Rntf(S_1, S_2, M)$ to all its neighbor mobile wireless nodes in its wireless signal transmission range. □

3 Experiments

The proposed method for an indirect measurement of distance between two mobile nodes is implemented for evaluation. An evaluation field is a $10\,\mathrm{m} \times 10\,m^2$ and two nodes are fixed at middle points of opposite sides. Additional 14 mobile nodes are randomly located and our proposed method is applied to measure the distance between the two stationary nodes. The angle of visibility of the web cameras is $60°(=\pi/3\mathrm{rad})$ and the cameras have 3280 horizontal pixels. Hence, the resolution of angle measurements is $3.19 \times 10^{-4}\,\mathrm{rad}$. On the other hand, the diameter of wheels is $5.00 \times 10^{-2}\,\mathrm{m}$ and the 10 pulse waves are created in each rotation. Hence, the resolution of migration distance measurements is $1.57 \times 10^{-2}\,\mathrm{m}$.

In our experiment, 6 mobile nodes are kept in the common fields of view of the two stationary nodes. Thus, the same result distance between the two stationary nodes is induced by 6 combinations of measured angles and a migration distance theoretically. However, it is not always true due to the resolutions of

measurements of angles and migration distances and the calculation errors. Two graphs in Fig. 3 show the distribution of the results distances. Here, 10 and 8 mobile nodes are in a field of view of a web camera of one stationary node at the beginning and the end of observation and 11 and 11 mobile nodes are in a field of view of a web camera of the other one. Since the images of mobile nodes are not tracked in our experiments, that is, the angles at the beginning and the end of observation are independently measured, number of possible combinations for calculation of distance between the two stationary nodes is 135,520. The threshold values 1.0×10^{-6} and 1.0×10^{-5} are introduced to assume the same value in Fig. 3, respectively. In the left graph, the expected result 10.00000 m is acquired; however, in the right graph, we have a wrong result 0.0185 m. As shown in our experimental results, it is expected for our proposed method to measure distances precisely between two nodes without direct distance measurement; however, an adequate threshold for the same result distance should be detected for acquisition of the precise distance.

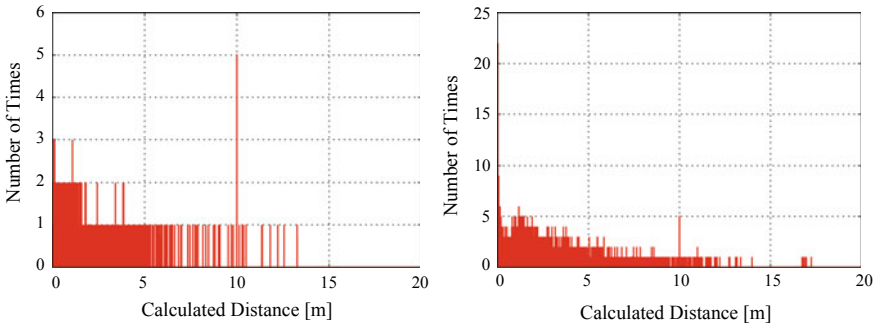


Fig. 3. Distribution of calculated distance.

4 Concluding Remarks

This paper has proposed a novel method to achieve relative locations of mobile wireless nodes. Different from the conventional methods, the proposed method avoids remote measurements of distances between nodes. In the proposed method, it is assumed that mobile nodes possesses cameras and rotary encoders to measure angles made by two neighbor nodes in their visual fields and to measure their own migration distances. The former is expected to suffer less errors since the angle measurement is independent of distances and the latter is also expected to be more precise since the measurement is local, i.e. direct. Since it requires to high calculation overhead to identify multiple mobile wireless nodes in pictures taken by the cameras, we have proposed a heuristical method to achieve the correct distance even without identification of the wireless nodes in

the pictures. In addition, the proposed method is implemented in an experimental mobile network. Here, it is expected for our proposed method to provide precise distance measurement without direct distance measurement. However, a method to overcome the errors of measurement and calculation is required. In future work, we evaluate the precision achieved by the proposed method and the communication and calculation overhead required in our proposed method in comparison with the conventional methods.

References

1. Shchekotov, M.: Indoor localization method based on Wi-Fi trilateration technique. In: Proceeding of the 16th Conference of Fruct Association, pp. 177–179 (2014)
2. Boukerche, A., Oliveira, H.A.B.F., Nakamura, E.F.: Localization systems for wireless sensor networks. IEEE Wirel. Commun. **14**(6), 6–12 (2007)
3. Hamada, K., Numao, M.: Adaptation techniques for environmental change for indoor location estimation by machine learning with Wi-Fi RSSI. IPSJ Technical Report, vol. 2017-DPS-169, no. 3, pp. 1–8 (2017)
4. Kato, K., Ishiguro, H., Barth, M.: Identification and localization of multiple robots using omnidirectional vision sensors. Tech. Rep. IEICE D-II **7**(6), 1270–1278 (2001)
5. Arao, A., Higaki, H.: Estimation of relative offset and drift for synchronization of local clocks in wireless sensor networks. In: Proceedings of the 11th International Conference on Mobile Ubiquitous Computing, pp. 108–114. Systems, Services and Technologies (2017)

E-Learning and Industry 4.0: A Chatbot for Training Employees

Fabio Clarizia, Massimo De Santo, Marco Lombardi, and Domenico Santaniello(✉)

DIIN, University of Salerno, Fisciano, Italy
{fclarizia,desanto,malombardi,dsantaniello}@unisa.it

Abstract. Within what is called the Fourth Industrial Revolution, one of the problems that companies encounter frequently, in order to ensure themselves and their products and services over time, is the need for continuous training of the people who work in synergy for them. If continuing education is a problem, e-learning is its natural solution. The research concerns the realization of a system capable of providing a constant, reliable, and friendly help through a practical and nice chatbot based on context processing. In particular, the proposed chatbot acts as a reminder and follows the user during his personal corporate training, ready to provide, when needed, the necessary and useful teaching material to complete the educational course.

Keywords: Industry 4.0 · E-learning · Chatbot · Context-aware computing · Recommender systems

1 Introduction

The term Industry 4.0 refers to the so-called Fourth Industrial Revolution: a set of enabling technologies that facilitate the automation of processes and the exchange of huge amounts of data along the production chain [1]. This technological trend includes the Internet of things (IoT) [2–4] cyber-physical systems, cognitive computing, and cloud computing: technologies that can transform a traditional factory into a smart factory [5].

In this regard, the increasing use of technology and its entry into the most varied sectors and production cycles has ensured that large companies, as well as those of small and medium size, could experiment new ways of running the business and also new methods of imparting knowledge and communicating between employees [6].

In particular, e-learning solutions can satisfy business training needs, creating projects of which it is possible to properly evaluate and measure the effects. This commitment allows to build agile, engaging, and meaningful educational experiences for the companies that use them and, at the same time, for the people who use them directly. In this regard, there is a new set of emerging technologies that introduces new man–machine communication paradigms: chatbots [7].

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_44

2 Related Works

2.1 Chatbot in Industry 4.0

Use of natural language in human–computer interaction is attractive because it is one of the most flexible, efficient, and natural means to communicate. Recently, Natural Language Processing applications have become very popular in the Industry 4.0 where a cooperation between human and machine takes place in mutual interest [8].

For this reason, chatbots are being integrated in numerous industries: Smart Factory makes it possible to monitor, in real time, all the physical processes and make effective decentralized decisions and chatbots are able to assist the human operators in accomplishing their tasks.

For example, FinderBuddyis is a chatbot intended to simplify procurement processes in industries. This application enables the main features like procuring best deals of materials from suppliers, updating quantities of a item in the inventory, changing or cancelling orders still in processing, review of past orders, checking availability of various items in the inventory, etc. [9] In this way, the chatbot makes the procurement process user-friendly and smooth with minimal human intervention.

A further example is [10]: a chatbot that provides an interface as a platform for communication between the bot and the client.

2.2 Chatbot for E-Learning

In recent times, chatbots have become more important in the field of education, particularly in e-learning [11, 12].

They are practical and natural interfaces for users because they are able to offer everyone a tailored support, exploiting their interests and capabilities [13].

One of the most studied fields in the use of chatbot for e-learning is tutoring. An example is AutoTutor, an intelligent system for tutoring based on natural language conversations between chatbot and human users [14]. The evaluation of this system has shown that the learning acquired through the help of human tutors is equivalent to that acquired through AutoTutor: this works well when the knowledge shared between the tutor and the user is low or moderate. If the shared knowledge is high, both the chatbot and the user will expect a higher level of accuracy of mutual understanding and, therefore, the risk that they will not understand each other is higher. Nevertheless, the tool continues to have the practical advantage of helping children and adults learn.

A further example is in the field of algebra learning, in which a system called Ms Lindquist was implemented, which does not offer explicit instructions but allows students to learn through practice: it was found that students using this system solve fewer problems but they learn just as well, and sometimes even better, than those who do not use the bot [15]. Furthermore, the use of the system has shown that a conversational method helps students stay motivated and remain vigilant.

The effectiveness of chatbots in the field of learning is the fact that the dialogic method is a natural method of communication with which students are very familiar. In addition, some systems allow to forward chatbot conversations to teachers who can give additional support when the chatbot should be lacking. The support of natural language

in e-learning encourages users to make a self-assessment of their ability to answer questions correctly on a given topic, helping them to understand their own strengths and the possible gaps that should instead be filled [16–18].

3 System Architecture

The core of the proposed architecture (Fig. 1) is the Inference Engine, composed of four sub-modules: Text pre-processing, LDA analyzer, query classifier, and interaction manager. These modules allow the analysis of the application, the attribution of meaning, the ontology mapping, and the formulation of a response (final or intermediate) to the user. They lean on the Knowledge Base, that is the databases related to courses and teaching materials or offered services.

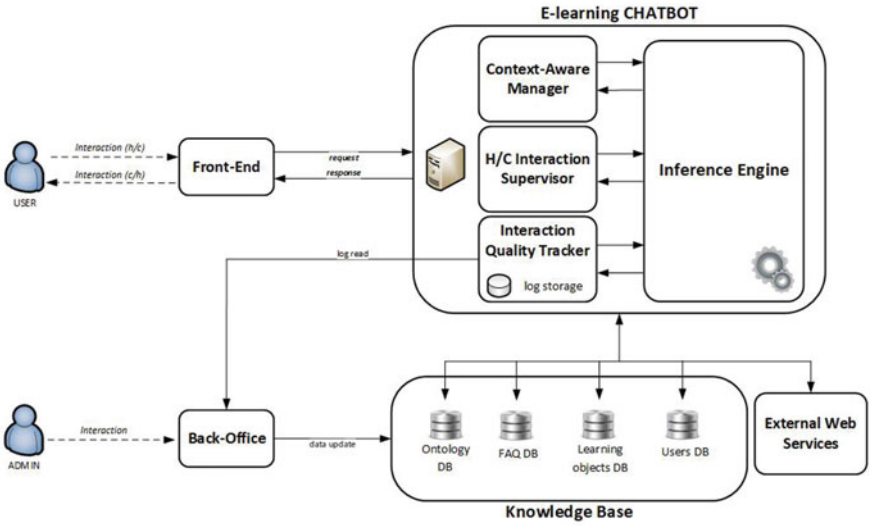


Fig. 1. System architecture

In particular, the Inference Engine includes the analysis of the text and the elaboration of the user's real intentions and needs obtained through collaboration with other minor modules and, in particular, through the use of the module that allows context extraction based on the Context Dimension Tree (CDT) [19]. This last module indeed has a crucial role: the text generated by a chat can be seen as a concoction of contextual information and the use of some words contributes to defining the different context elements useful in the research for the context itself.

In this respect, Latent Dirichlet Allocation (LDA) represents the appropriate model for making a series of observations and for explaining the correlation between keywords and topics [20]. In particular, according to the proposed approach, each topic must refer to a specific context element [21]. In doing so, the chabot—first through the analysis of the chat and then through the elaboration of the current context—defines the real

intention of the user and it is able to better satisfy his needs, recommending properly filtered training content.

Through the textual analysis, it is possible to know the employee (factory worker, office worker, manager, etc.), where he is (production department, administrative office, meeting room, etc.), the purpose of his work (assembly, contracts, project management, etc.), and what it needs (technical specification of a machine, legal advice, telephone contact of a customer, etc.). For example, a possible context could describe a user who must process personal data of employees and wants to do so in accordance with the directives contained in the GDPR.

3.1 Proposed Approach

The proposed approach is based on the probability that a word *W* in the chat is related to a concept node *Nc* of the CDT. This probability is proportional both to the number of times the word has been used for that concept and to the number of times the same concept has been identified. The application of this model provides a fully automatic characterization of the conversations between user and chatbot without the need to indicate the semantic value of the words in the text [22, 23].

In this regard, Latent Dirichlet Allocation is a model that allows to make a series of observations to explain the correlation between keywords and some topics. In this specific case, for example, it is assumed that each message in a chat is a mixture of a small amount of contextual information and that the use of some words can therefore be associated with a context topic, or a context element.

In Natural Language Processing (NLP), LDA is often used as a topic model: a type of statistical model that attempts to discover the abstract topics that are occurred in a collection of documents. Intuitively, since each chat is related to a specific context, specific words are expected to be part of the conversation more or less frequently. A topic model captures this intuition in a mathematical statistical model, allowing to examine a set of documents and discover, based on purely statistical reasoning, what the topic of each specific document might be.

In practice, through appropriate Bayesian keywords-based filters, a message sent to the chatbot is divided into shorter and simpler sentences (clusters), in order to identify the appropriate contextual elements. So, once the system has managed to identify all the context elements, it will be able to define a specific context and therefore recommend content to users [24, 25].

According to the LDA approach on a set of conversations related to the same domain (in this case, e-learning and industry), a Mixed Graph of Terms (mGT) can be automatically extracted and used both for the design of the context tree and the associated constraints. This will be used to detect the context in real time from the user's chat with the bot. In particular, LDA was mainly used to generate topics (context elements) within text documents (chats). These topics were processed by the system as contextual elements useful during the use of the Context Dimension Tree [26].

In other words, each context element, identified through the analysis of conversations with the chatbot, corresponds to a relevant section of information stored in the database: each context element corresponds to a partial view on the database while the contextual query is performed by automatically combining the different partial views. The

same mechanism is used not only to select useful data but also to recommend services according to the identified context.

4 Experimental Results

According to the proposed approach, a prototype chatbot was designed and implemented (Fig. 2), together a server-side component. In this first experimental phase, the application is used to support the employees of a southern Italian company in the training related to the General Data Protection Regulation (GDPR).

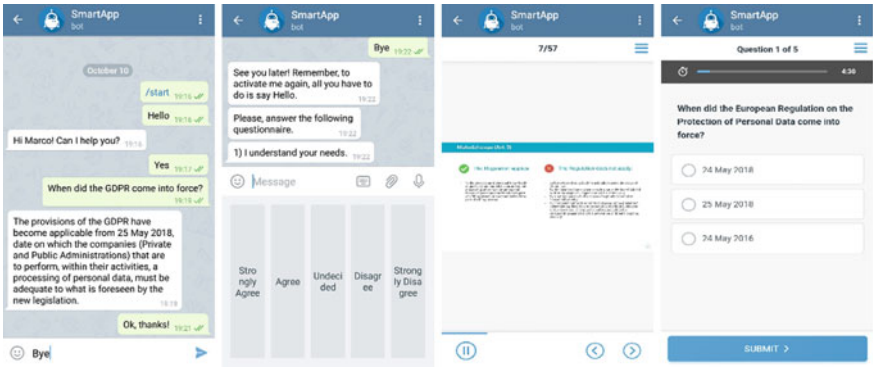


Fig. 2. Some screenshots of the chatbot

After interacting with the chatbot and according to the Likert scale, users responded to a questionnaire consisting of five sections. In particular, users responded to each assertion of a specific section, using five possible answers: TD—"Totally Disagree," D—"Disagree," U—"Undecided," A—"Agree," and TA—"Totally Agree." In total, 2349 interactions were recorded.

Section A: contextual recommendation

1. User needs have been correctly identified
2. The context and personal preferences of the user have been used correctly.

Section B: chatbot conversation

1. User intentions have been correctly identified.
2. The dialogue with the chatbot was smooth and uninterrupted.

Section C: presentation of information

1. The content provided by the chatbot was exhaustive.
2. The content provided by the chatbot was adequately presented.

Section D: usability of the system

1. The interface is user-friendly.
2. Response times are short.

Section E: future development

1. Other users (work colleagues) may be included in the chat to improve the learning experience.
2. The same approach could be applied in other scenarios (not only professional training).

The results for each section are shown in Table 1. As can be seen from the answers, users are largely satisfied.

Table 1. Analysis of results

Section	Percentage		
	Negative (%)	Neutral (%)	Positive (%)
A	5.64	2.85	91.51
B	10.54	5.45	84.01
C	10.09	4.92	84.99
D	9.54	4.45	86.01
E	9.17	4.32	86.51

Looking at the number of negative feedbacks, the main problems are related to the conversation with the chatbot which does not always take place smoothly: according to some users, the conversation is not as natural as the conversation with a human. Indeed, the main goal of this work is not intended to have long and fluid conversations with the chatbot, instead to create a personal assistant that, in a few steps, can understand the context and respond to user needs through the recommendation of Web services and available information. The problem related to the dialogue can be solved also considering

dialect words or grammatical errors not yet learned by the system in this initial state: the results suggest paying attention to natural language processing techniques in order to improve human--machine interaction.

Furthermore, Sect. A, relative to the recommendation module based on the context, shows very interesting results: the proposed training contents met the needs of the user, dynamically adapting to the current context.

5 Conclusions

This work presents an innovative solution which, as demonstrated by the experimental results, is very useful for both companies and employees. The proposed system in fact simplifies learning and significantly reduces the timing. In particular, the chatbot can be adopted both for the education of newly hired employees and for continuous training, which is necessary after the installation of new industrial plants or the introduction of new processes. In this way, it adds a further step to the transformation toward the Industry 4.0.

References

1. Lu, Y.: Industry 4.0: a survey on technologies, applications and open research issues. J. Ind. Inf. Integr. **6**, 1–10 (2017). https://doi.org/10.1016/j.jii.2017.04.005
2. Abate, F., Carratù, M., Liguori, C., Ferro, M., Paciello, V.: Smart meter for the IoT. In: I2MTC 2018 IEEE International Instrumentation and Measurement Technology Conference: Discovering New Horizons in Instrumentation and Measurement, Proceedings, 2018, pp. 1–6 https://doi.org/10.1109/i2mtc.2018.8409838
3. Colace, F., De Santo, M., Lombardi, M., Pascale, F., Santaniello, D., Tucker, A.: A multilevel graph approach for predicting bicycle usage in London area. In: Yang, X.-S., Sherratt, S., Dey, N., Joshi, A. (eds.) Fourth International Congress on Information and Communication Technology. Advances in Intelligent Systems and Computing, vol 1027, pp. 353–362. Springer, Singapore (2020). https://doi.org/10.1007/978-981-32-9343-4_28
4. Colace, F., Santaniello, D., Casillo, M., Clarizia, F.: BeCAMS: a behaviour context aware monitoring system. In: 2017 IEEE International Workshop on Measurement and Networking, M and N 2017—Proceedings, 2017 https://doi.org/10.1109/iwmn.2017.8078374
5. Da Xu, L., Duan, L.: Big data for cyber physical systems in industry 4.0: a survey. Enterp. Inf. Syst. **13**(2), 148–169 (2019). https://doi.org/10.1080/17517575.2018.1442934
6. Gorecky, D., Schmitt, M., Loskyll, M., Zuhlke, D.: Human-machine-interaction in the industry 4.0 era. In: 2014 12th IEEE International Conference on Industrial Informatics (INDIN), 2014, pp. 289–294 https://doi.org/10.1109/indin.2014.6945523
7. Colace, F., De Santo, M., Lombardi, M., Santaniello, D.: CHARS: a cultural heritage adaptive recommender system. In: Proceedings of the 1st ACM International Workshop on Technology Enablers and Innovative Applications for Smart Cities and Communities—TESCA'19, 2019, pp. 58–61 https://doi.org/10.1145/3364544.3364830
8. Quarteroni, S.: Natural Language Processing for Industry. Informatik-Spektrum **41**(2), 105–112 (2018). https://doi.org/10.1007/s00287-018-1094-1
9. Chawla, A., Varshney, A., Umar, M.S., Javed, H.: ProBot: an online aid to procurement. In: 2018 International Conference on System Modeling & Advancement in Research Trends (SMART), 2018, pp. 268–273 https://doi.org/10.1109/SYSMART.2018.8746954

10. Rahane, W., Patil, S., Dhondkar, K., Mate, T.: Artificial intelligence based solarbot. In: 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018, pp. 601–605 https://doi.org/10.1109/ICICCT.2018.8473172
11. Colace, F., De Santo, M., Lombardi, M., Pascale, F., Pietrosanto, A., Lemma, S.: Chatbot for E-learning: a case of study. Int. J. Mech. Eng. Robot. Res. 528–533 (2018). https://doi.org/10.18178/ijmerr.7.5.528-533
12. Lombardi, M., Pascale, F., Santaniello, D.: An application for cultural heritage using a Chatbot. In: 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), 2019, pp. 1–5. https://doi.org/10.1109/cais.2019.8769525
13. Clarizia, F., Colace, F., Lombardi, M., Pascale, F., Santaniello, D.: Chatbot: an education support system for student. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11161 LNCS, 2018, pp. 291–302. https://doi.org/10.1007/978-3-030-01689-0_23
14. Graesser, A.C., Chipman, P., Haynes, B.C., Olney, A.: AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. IEEE Trans. Educ. **48**(4), 612–618 (2005). https://doi.org/10.1109/TE.2005.856149
15. Heffernan N.T., Croteau, E.A.: Web-based evaluations showing differential learning for tutorial strategies employed by the Ms. Lindquist Tutor. 2004, pp. 491–500. https://doi.org/10.1007/978-3-540-30139-4_46
16. Kerry, A., Ellis, R., Bull, S.: Conversational agents in E-learning. In: Applications and Innovations in Intelligent Systems XVI, London: Springer London, 2009, pp. 169–182. https://doi.org/10.1007/978-1-84882-215-3_13
17. Amato, F., Moscato, V., Picariello, A., Sperli, G.: Multimedia social network modeling: a proposal. In: 2016 IEEE Tenth International Conference on Semantic Computing (ICSC), 2016, pp. 448–453. https://doi.org/10.1109/ICSC.2016.20
18. Casillo, M., Colace, F., De Santo, M., Lemma, S., Lombardi, M., Pietrosanto, A.: An ontological approach to digital storytelling. In: Proceedings of the The 3rd Multidisciplinary International Social Networks Conference on SocialInformatics 2016, Data Science 2016—MISNC, SI, DS 2016, 2016, pp. 1–8 https://doi.org/10.1145/2955129.2955147
19. Daniel, F., Matera, M., Quintarelli, E., Tanca, L., Zaccaria, V.: Context-aware access to heterogeneous resources through on-the-fly mashups. 2018, pp. 119–134. https://doi.org/10.1007/978-3-319-91563-0_8
20. Krestel, R., Fankhauser, P., Nejdl, W.: Latent dirichlet allocation for tag recommendation. In: Proceedings of the third ACM conference on Recommender systems - RecSys'09, 2009, p. 61. https://doi.org/10.1145/1639714.1639726
21. Casillo, M., Clarizia, F., D'Aniello, G., De Santo, M., Lombardi, M., Santaniello, D.: CHAT-Bot: a cultural heritage aware Teller-Bot for supporting touristic experiences. Pattern Recognit. Lett. (2020). https://doi.org/10.1016/j.patrec.2020.01.003
22. Colace, F., De Santo, M., Greco, L., Amato, F., Moscato, V., Picariello, A.: Terminological ontology learning and population using latent dirichlet allocation. J. Vis. Lang. Comput. (2014). https://doi.org/10.1016/j.jvlc.2014.11.001
23. D'Aniello, G., Gaeta, M., Loia, V., Orciuoli, F., Tomasiello, S.: A dialogue-based approach enhanced with situation awareness and reinforcement learning for ubiquitous access to linked data. In: 2014 International Conference on Intelligent Networking and Collaborative Systems, 2014, pp. 249–256. https://doi.org/10.1109/INCoS.2014.73
24. Amato, F., Moscato, V., Picariello, A., Piccialli, F.: SOS: a multimedia recommender system for online social networks. Futur. Gener. Comput. Syst. **93**, 914–923 (2019). https://doi.org/10.1016/j.future.2017.04.028

25. Amato, F., Castiglione, A., De Santo, A., Moscato, V., Picariello, A., Persia, F., Sperlí, G.: Recognizing human behaviours in online social networks. Comput. Secur. **74**, 355–370 (2018). https://doi.org/10.1016/J.COSE.2017.06.002
26. Amato, F., Moscato, V., Picariello, A., Colace, F., De Santo, M., Schreiber, F.A., Tanca, L.: Big data meets digital cultural heritage. J. Comput. Cult. Herit. **10**(1), 1–23 (2017). https://doi.org/10.1145/3012286

A Secure Smart City Infrastructure Framework for E-Service Delivery Within a Developing Country: A Case of Windhoek in Namibia

Licky Richard Erastus[1], Nobert Rangarirai Jere[2], and Fungai Bhunu Shava[1](✉)

[1] Faculty of Computing and Informatics, Namibia University of Science and Technology, Windhoek, Namibia
LErastus@mtc.com.na, fbshava@nust.na

[2] Information Technology Department, Walter Sisulu University, Buffalo City Campus, East London 5200, South Africa
njere@wsu.ac.za

Abstract. As cities embrace technologies, they become more intelligent ubiquitous and unlimited connection. This is enabled by fast broadband and better supporting infrastructure. African cities are making effort to improve service delivery using technologies. As these developments are witnessed, information sharing, infrastructure requirements and security issues become very common. There are a number or study papers on smart cities; however, less has been studied on the smart city security and Infrastructure. Since it is clear that within a smart city, high Internet speed, intelligent systems efficient services are crucial, there is need for a secure supporting information infrastructure. Qualitative data was gathered from participants within a case study approach. Results should the current ICT initiatives to support smart cities. At the same time, all participants engaged agree that a secure infrastructure framework is key for African cities to become smart. The paper presents a comprehensive Secure Smart City Infrastructure Framework (SSCIF) for the City of Windhoek in Namibia. The framework aims to enable e-service delivery within a smart city.

Keywords: Smart city · Cybersecurity · Infrastructure · E-services · Smart services · Emerging technologies · ICT security · Developing country

1 Introduction

The growth of population in African cities is evident. The projections by the United Nations show that by 2050, almost 70% of the population in the world will be living in urban areas with only 30% or less in rural areas [1]. This view is also shared by Escher Study Group who mentioned that cities will continue to grow as rural population migrates to urban centres for better service [2]. Urban growth presents new challenges to city governance and service provisioning as a result of pressure from the ever-increasing

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_45

population. Some of the notable challenges as population increase in urban areas as mentioned by [3] include air pollution, traffic congestion and competition for resources. There is need for proper planning for many cities in the world to cope with noted population increases. [4] noted that "central and local governments from around the world have devised plans for existing and emerging cities to become both smarter and greener".

With the ICT developments personal information, business transactions and the city services are likely to be exposed. This creates security challenges and may affect the success of service delivery within the smart city. Implementing smart city projects will not only bring the city to meet the qualities of today's world but also improve the quality of people's living, infrastructure and economy of the country. There is clear evidence that as the population increases in cities, Namibian cities will also be affected. This study considers that the creation of smart cities could redress the social and economic gaps that are currently being experienced in Namibia in response to population growth in its cities. There is no doubt that ICT services are becoming visible in Namibia; however, the adoption of the services is still low. The common challenge as observed during the data collection in this paper is lack of trust on the e-services. Having noticed that smart cities will require improved ICT infrastructure leading to Internet of things, the paper focuses on designing an enabling Secure Smart City Infrastructure Framework (SSCIF), which aims to improve adoption of smart cities.

This paper assesses the current e-services and emerging technologies within the African city. The paper combines security, infrastructure and e-services for sustainable development. The paper is based on the African city that is not yet a smart city and prioritizes the security before services are deployed. We consider current ICT initiatives within the City of Windhoek (COW) and present the views of ICT experts engaged. The purpose of the paper is to provide a secure infrastructure framework that could enable e-service delivery within African cities. The paper contributes and adds some key guidelines to smart city planners and citizens' needs. The paper explains the SSCIF that is meant to improve service delivery in Namibia. The paper is structured as follows: literature on technologies in teaching and learning, common teaching approaches, the methodology which was applied, key findings and the proposed recommendations.

2 Related Literature

It should be noted that ICTs play an important role in enabling service provision within a smart city. The reliance and use of pervasive ICTs to derive smart cities call for laid down security structures. Reference [5] supports this view by suggesting that the preservation of privacy is of prime importance if smart cities are to be a reality. They went on to propose conceptual framework of privacy. Reference [6] expressed concern over the smart city systems' security vulnerabilities at the expense of unknowing users. This is attributed to the lack of clear theory on law and rights that defines the capabilities and scopes of smart cities systems. This calls for security architectures for smart city systems that participants and the shared data is protected. Security is the largely ignored element in current smart city models as authors put more focus on defining components that define a smart city environment.

The security issues are explained based on possible threats that may happen within the city [7]. The threat categories taken into account may include availability, integrity and accountability threats. The basic elements of the security information infrastructure of smart cities may have the following layers to provide the city with ICT security Levy [8]: data transmission network, data processing, data aggregation connectivity and smart processing.

The phrase "smart city" has been used to describe the integrated urban infrastructure and services, i.e. public safety, water distribution, electrical, transportation and buildings [9]. Some of the important solutions within the smart city are as in Fig. 1.

Fig. 1. Some of the important solutions within the smart city. Retrieved from [10] https://www.accenture.com/us-en/insight-smart-cities

Reference [9] reviewed the meaning of the phrase "smart" in the "smart city context". Smartness is placed on user perspective in marketing terms. Because of the need for approval by community members; "smart" is user-friendly and serves better than the term "intelligent", which is narrowed to having a speedy mind and being responsive to feedback. This explanation suggests that "smart" is more than "intelligent", and the smartness is recognised when the system adjusts itself to the user necessities. The smart cities require good supporting infrastructure.

There are many types of infrastructure/s that exist. Some of the common ones are presented in the next section.

2.1 Institutional Infrastructure

Institutional infrastructure is representative of activities that relate to governance, planning and management of a city for its improvement. Cities need high quality governance, with a solid local say in decision making, which is critical for smart cities [11]. Therefore, the principle to be followed is "Governance by Incentives rather than Governance by Enforcement". Making all information effortlessly obtainable in the public domain and governance must be made possible through the social media will encourage public participation [12]. Then there is the physical infrastructure. According to [13, 11], physical infrastructure refers to its stock of cost-effective and intelligent physical infrastructure such as the urban mobility system, the housing stock, the energy system, the water supply system, sewerage system, sanitation facilities, solid waste management system, drainage system just to mention a few which are all combined through the use of technology.

2.2 Security in a Smart City

The previous section deliberated on pillars of a smart city according to the reviewed literature. The safety and confidentiality of information in a "smart city" are of interest to authors. The motive being in order to secure the existence of vital services such medical care, governance and amenity issues in a "smart city", the information security must be air-tight. The aspects taken into consideration in order to recognise the concerns in information security in a "smart city" are described in the next sections.

Reference [14] identified, explained and proposed clarifications to the information security concerns by bearing in mind the stated influences. Quite a number of studies centre on the security and confidentiality concerns, the components and architecture of a smart city are discussed in current literature.

The Internet of things (IoT) is the main concentration of authors as it is the central technology on which the "smart cities" are being established and [15]. According to [15, 16], the significant obstacles and problems faced regarding security and confidentiality are deliberated, with the context of technological standards in mind. From such data, the "smart cities" can extract very vital information serving real-time analysis and pervasive computing. Reference [17] expounds that nevertheless the enormous data offers numerous prospects for smarter life, still it brings challenges of security and confidentiality.

Reference [6] addressed the cybersecurity challenges and their focus was on two key challenges—security and confidentiality. They present a mathematical model depicting the alliance between people, IoT and servers which are susceptible to information security threats. Moreover, [18] proposed a disseminated framework for IoT applications, which assures security, trust and confidentiality in information delivery. As IoT applications play a fundamental part on structuring the "smarter city", so some information security matters in a "smart city" can be addressed through the distributive framework.

2.3 Security Challenges in Smart Cities

It is not easy to ensure that all the services and the data accessed within a smart city are secure. The main security challenges as explained by a couple of authors such as [2, 7, 16] include.

Data Privacy and protection concerns: Personal data is very sensitive and should always be protected. Smart city technologies allow data storage of many uses posing a data privacy risk [2]. The other security challenges include: lack of surveillance, too much information and complicated systems and data leakage and insecure hardware are the other challenges.

The importance of understanding the common security challenges for this study is to be able to propose a framework that considers such challenges and provides a solution to these. It is also important to consider a number of challenges that affect smart cities. There are critical challenges facing smart cities. The key challenges are as indicated in Fig. 2.

Current literature shows that smart city challenges come in form of the infrastructure, service providers competing for resources, citizens competing for resources, the different perceptions and the citizens who are city dwellers. The next section provides some

Fig. 2. Smart city challenges. Adopted from http://smartmycity.ro/smart-city-challenge

possible security practices that have been tried and tested in other areas. Some of these practices could be critical in ensuring sustainable security solution.

2.4 Good Practices Against Intentional Security Attacks in Smart Cities

There is no doubt that security issues are one of the biggest challenges that many institutions face. These challenges become worse in situations where everyone is connected. With the increase in hackers and internet attackers, it makes it difficult to ensure that a smart city is secure [7]. Therefore, it is not a one man's duty to ensure security is maintained by making sure that the citizens are aware of cybersecurity threats and also well-trained to act properly. However, [7, 8] propose some of the good practices including setting up virtual private networks; encryption of data; deployment of physical protection and maintenance of backups.

3 Study Approach

The study uses the qualitative approach within a case study. This methodology is suitable for this study as it gives the author numerous options of sourcing data that enhances the understanding of the phenomenon understudy. The case study strategy is used. In this case, the City of Windhoek that intends to be a smart city by 2022 is considered. [19] define a case study as "an exploration of a "bounded system" or a case (or multiple cases) over time through detailed, in-depth data collection, involving multiple sources of information rich in context" [20]. The study uses semi-structured interviews to propose the SSCIF based on smart cities experiences. This was supported by an online questionnaire. The qualitative approach to design a security framework is selected as this has been used in developing other related cybersecurity frameworks by professional organisations and other established authors such as [21–24] Actual findings from the approaches selected are presented in Sect. 5.

3.1 Study Population and Case Selection

When using qualitative approach, where interviews are used as one of the methods, selection of the participants is very critical [25]. The selection of the participants was

based on their roles and participations in smart city involvement. The ITU Report [21] provides the main stakeholders that are supposed to be engaged for smart city implementation. Based on the possible stakeholders that could be involved for the city planning, the study classifies the stakeholders indicated below:

- **Active**: all stakeholders that support smart city initiatives.
- **Beneficiaries**: individual and entities that will benefit from smart city developments.
- **Affected**: these are the individuals or entities that will be somehow affected by the deployment of smart city solutions. These could be either potential supporters or potential opponents [21].

Referring to the agreed common stakeholders by a study done by ITU as indicated in the [21], the population was derived from COW employees, government officials, ICT private departments, NGOs, academic institutions and research institutions and Namibian citizens.

Simple random sample was used to pick from the above-mentioned stakeholders. Participants for the interview should have the required information to assist the author to have deep understanding on the study. From each of the stakeholders, one would be randomly picked to give a total of seven (7) participants for the interviews.

3.2 Description of the Case

Namibia's projected population for 2018 is approximately 2,413,600 with Windhoek as the main city that has a total population estimate to be roughly 447,600 which is 15% of the total population when using the census done in 2011 [26]. This coupled with the current ICT developments in Namibia has motivated this study. ICT developments have been supported by initiatives from the current ICT service providers such as Telecom, MTC, NGOs and the Namibian government. According to the City of Windhoek Strategic Plan Report [27], the City of Windhoek (COW) is the major driver of social and economic transformation within Namibia. COW is the capital city of Namibia and is the centralised city with all the major services required for the country. It plays a big role in nation—building and also leading other local authorities. COW is currently creating a viable and accountable governance structure. In coming up with the current guiding document, the: Transformational Strategic Plan (2017–2022), collaborative meetings with the citizens and key stakeholders was done [27].

The city is guided by the national and international strategies such as the Harambe Prosperity Plan, the Vision 2030 and National Development Plan (NDP) 5 and the African Union Agenda 2063.

The common problems as pointed out by the COW Chief Executive Officer in the strategic plan include shortage of water and pressure of land.

Water shortages come as a result of less rainfall and continuous droughts (COW Strategic Plan, 2017). The pressure on the land is as a result of population growth and an increase in application for land ownership. Just like other African countries, the need for land is critical in Windhoek as well. Due to these two factors, land allocation and housing have been identified as the highest city priorities.

The other main projects that have been identified as pointed in COW strategic plan [27] include concentrating on: improved business systems, business continuity and the implementation of a paperless business environment.

The services are meant to improve customer services enabling simpler, faster, friendlier, efficient and meaningful service to all citizens.

As indicated in the COW strategic plan [27], a number of themes have been considered. These are: financial sustainability, governance, social progression, economic advancement and infrastructure development.

The themes are formulated based on the national guiding documents and existing strategies. These initiatives and themes are aimed at achieving the COW goal of "becoming a Smart and Caring City by 2022 through absolute dedication and commitment of staff, a sound financial base, strong political will and responsive structures, processes and systems" [27]. All these are meant to enhance service delivery to the Namibian citizens. Figure 3 is a snapshot of COW, the case where the highlighted strategies need to materialise.

Fig. 3. Snapshot of city of Windhoek as extracted online

3.3 ICT Initiatives Towards a Smart City in Namibia

The growing population within COW calls for proper ICT infrastructure planning [3]. Although Windhoek has not been declared a smart city, the current ICT infrastructure and developments encourage a smart Windhoek [3]. Currently, there is a 20-year plan that was adopted to improve the public transportation in Windhoek and the surrounding areas. Though this focuses only on transport services, it motivates this study, as the smart transport is one of the pillars of a smart city.

The rapid technological advancement within COW contributes to the potential of a smart city. The emphasis by the Namibian government through the HPP Plan [27]

on the need to have developed ICT infrastructure adds on to the potential of COW to become smarter. According to the HPP [27], the country has modern ICT infrastructure that supports smart city initiatives. The telecommunication services have improved from the mid-1990s to-date [28]. In addition to this, the growth and access to mobile network have also contributed to better communication among the urban dwellers in Namibia [28].

Current electronic and digital services also enable a smart Windhoek in the future [28]. Namibia has been improving in terms of connectedness. The availability of internet and different efforts by the government and the private sector to make Namibia connected have been witnessed [28]. The government has enabled the creation of supporting entities such as Communication Regulation Authority (CRAN) and the significance of ICT is highlighted in the HPP [27]. The major plan of action as outline in the HPP [27], which supports a smart Windhoek are: Better broadband infrastructure; supporting electronic and cybercrime bill; expansion of connectedness; improved internet access; implementation of electronic services.

The current initiatives are also supported by the private sector through software developments and emerging technology applications within COW. There are some examples of smart cities that this study refers to and learn from.

4 Findings

This section presents the key findings that were considered for the paper.

4.1 Current Smart City Initiatives in Windhoek

The City of Windhoek (CoW) Council adopted a strategy to further develop the smart city through coordination of the City's work on digitalisation as it embraces a "smart cities" concept and redirects its efforts towards rendering efficient and effective services through a vision to be "a Smart and Caring City by the year 2022".

COW is embarking on a number of digitalisation projects. These include:

1. A smart and connected city through monetisation of City of Windhoek's existing Fibre Infrastructure;
2. Improve broadband services and provision of seamless connection in all areas of the city;
3. working with service providers to properly management and control of these infrastructure assets;
4. providing an Open Access for ICT services to create a platform where anyone can connect to anyone in a technology-neutral framework that encourages innovative, low-cost delivery to residents; and
5. CCTV surveillance network deployed in all major high safety risk areas of the City.

The way forward to make Windhoek a smart city is innovative and sustainable in economic, and socially sustainable markets.

4.2 Factors of Smart City Adoption in Namibia

The study has considered related studies that were conducted in Namibia. One of those is the study by [29, 30] and their findings informed a proposal on factors of smart city adoption in Namibia. These factors are arrived at following a consolidation of propositions in the previous sections and an understanding of Namibia. In line with findings by [29, 30], factors of smart city adoption in Namibia include perceived usefulness, perceived ease of use, social influence, and perceived security and privacy risk. These factors are discussed in the following section:

Perceived ease of use: Perceived ease of use is defined as the extent to which one expects using smart city facilities to be effortless. Thus, users are more likely to adopt a technology that does not require a lot of effort to understand and use.

Social influence: Social influence could be defined as the influence of important others on one's decision to use smart city facilities. Traditionally, within the African context, one of the principles in the Ubuntu philosophy is the role of other societal members in influencing decision making. These could include friends, relatives and workmates. Hence, it is argued that the opinion of close relatives, friends and or workmates is likely to influence Namibians into a decision to adopt smart city technologies.

The authors are aware of many other ICT-related initiatives and organisations that are working hard in ensuring that the dream of Windhoek to be smart by 2022 is achieved. These organisations and related policies are summed up in Table 1. The next section presents findings from questionnaires.

4.3 Finding from Online Questionnaire

4.3.1 Challenges Facing Smart Cities

Eight challenges facing smart cities were identified as presented in Fig. 4. A lack of Secure ICTs and No Supporting ICT Infrastructure were the most common challenges. Below is a list of all the challenges identified: population growth, high crime rates, lack of parking, lack of secure ICTs, no clear urban planning, no supporting ICT infrastructure, complicated ICT and lack of required technical skills

Smart City projects
Respondents had to identify smart city projects that are visible within their cities. The top three smart city projects that were identified by respondents include education and skills, competitiveness and energy. However, respondents suggested that all projects in Fig. 5 were visible in their city.

4.4 Findings from Interviews

The participants in Windhoek were engaged and interviewed for 45–60 min, with all data being recorded. Data was later transcribed and translated. For the purpose of this paper, only data related to security infrastructure and security within the smart cities is presented. To protect the privacy of the participants, different names and codes were

Table 1. Namibian supporting organisations and related policies

Name of policy/document/entity	Smart city initiative
National Development Plan (NDP5)	A Namibian Fifth National Development Plan of Namibia is part of a series of seven consecutive five-year strategic plans that the country has embarked upon to reach its objectives contained in the Vision 2030 strategy and plan
Harambe Prosperity Plan	A plan to explain key priority areas for Namibia within a specific time frame
Electronic transaction and cybercrime bill	Provides general framework for the promotion of the use of electronic transactions
Universal Access and Service policy for Information and Communications Technologies	Provision of how and what ICTs should be available for the Namibian citizens
Communications Regulatory Authority of Namibia (CRAN)	This was established to ensure adequate regulation of the ICT sector, which in turn promotes the availability, and accessibility of Telecommunications, Postal and ICT services at affordable prices
National Commission on Study Science and Technology (NCRST)	An entity responsible for driving study in Namibia and make recommendations to the nation
Academic Institutions	Key stakeholders in transforming and studying on Namibia and future developments

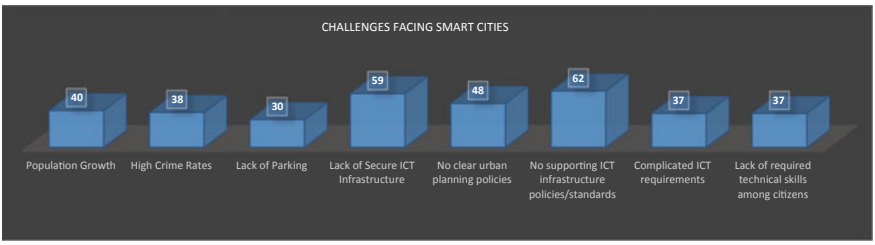


Fig. 4. Challenges facing ICT smart cities

used as follows: R1 for Steve, R2—Wesly, R3—Ester, R4—Belinda, R5—Collin, R6—Angeline, R7—Abbie and R8—Kendrick.

Infrastructure management

Steve… we have under the communications aspect of ICT you have digitisation, you have infrastructure man

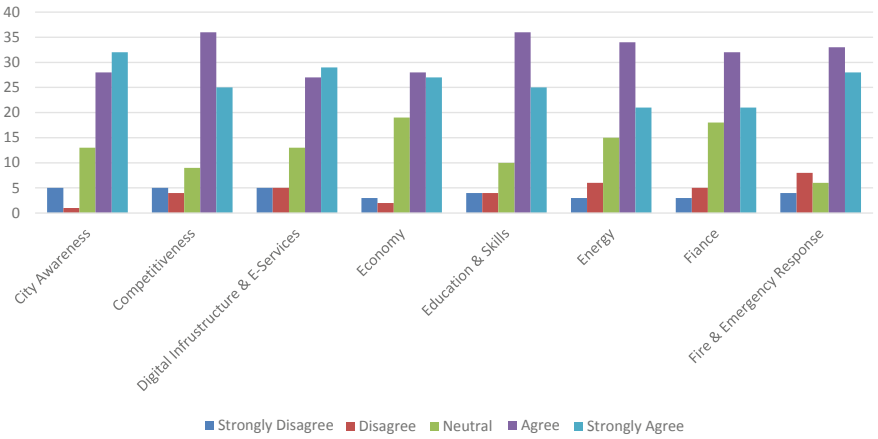


Fig. 5. Smart city projects

Kendrick… I mean, there must be some core infrastructure very high powerful service of machines that I believe and then be stationed for one at the city of Windhoek premises and obviously and some other forms of backup or parallel installation to ensure that in the event the information get lost you will always have a full backup.

Abbie … Some of these key terms are like first of all we are starting with infrastructure, basic infrastructures it's like water, electricity reduction networks, sanitation and solid waste management and this is also to do with mobility within the town you must also consider public transportation.

Collin… information sharing and traffic management, crime reduction and efficiency. This is the areas that are the pillars of smart cities and it depends on where you focus whether you focus on the transport system, the electricity system, the energy water or sewage and communication between different entities in the smart city or traffic so there are all the pillars of smart city.

Wesley… The first one is Bergan access combine Bergan access of fibre like 2G, 3G or 5G Wi-Fi an IT network which can provide connection of the relevance facilities and services and crowd competing crowd competing services is to low the overall cost of service delivery and enable mechanize for quickly responding to the needs of the public and big data mining and analytics collect and analysis in the most the amounts of data from various sources supporting the decision or the …

Wesley… And smart transportations analyse traffic data created by census and adjusted traffic signals in rule time to metadata the traffic congestion and smart breath and smart water. Improve sewage efficiency and enable the network and water components to find and resolve problems quickly.

Wesley. …infrastructure which supports the smart city.

Belinda… So smart cities aims to enable smart grids in terms of traffic, power and communication all which leads to improved efficiencies, mobility and infrastructure integration of indexes within the city. It also means that it, a city monitors and integrates conditions of all of its critical infrastructure including roads, bridges, tunnels, railways, subways, airports, communications, water, power and even major buildings and can optimise better resources even including preventing maintenance activities.

Infrastructure

Wesley… we need **right infrastructure** we need big tools in **connectivity powerful** competing capacity, census and **maintenance of the devices and the supporting system** and the platforms and the reliant and advanced energy and so on.

Improved security

Wesley. … if we can improve the security level and provide national rule on demanding networks and so on.

Ester… Is of movement in town service, security services and all that. So these are the benefits that will come with. And you know having a city that's has all the services does contribute to making it a preferred destination of business establishment of investments and that intend will contribute to the economy of the town and the country as a whole.

5 SIIF Design Process

In order to have a proposed SIIF, a series of steps were considered. These steps were derived from some tried and tested theories, strategies and techniques for designing security framework. The main phases that were borrowed were as outlined by [31]. Of course, not all of the phases were considered, but the main ones that were borrowed were the following: Detailed reading and classification of related data; key concepts naming and identification; integration of the concepts; classification and categorisation and making sense of the concepts; validation of the framework.

Secure Information Infrastructure Components

The proposed framework was designed following the outlined phases, guided by the findings, application of related theories and the author's own interpretation of some findings. The framework could be termed a 3-in-1 generic framework. This is so, as it accommodates the views of three categories of study fields and hence is based on ideas from those different experts. For example, the framework accommodates the views of experts from ICT, security and information infrastructure fields.

The main components of the framework, could also be explained as key steps that one could follow in designing such a framework. The authors understand the significance of keeping the basics and mantaining simplicity, even when the study itself had a number of complex topics. In this case, the framework components are explained as follows:

Setting the basis, prioritie, stakeholders, governance and citizen engagement—this means that for any author in the field or anyone to consider this framework, there is need to engage all key stakeholders.

Developing smart infrastructure and integrated platform—this aspect was important, as the second step. After getting the input from different stakeholders, the study had to consider other possible smart ideas that would work in the context. This was done through engaging key smart city experts in Namibia.

Identifying and developing smart service—once the infrastructure is available, the next step is to engage all the key stakeholders to draw up the intended smart services. As indicated from the findings and also summarised in the framework, the services needed were outlined.

Defining key perfomance indicators, standards and monitoring—when the stakeholders have being engaged, integrated platform is set up, smart services are defined, the next step is to outline the KIPs. This defines the supporting policies on how to manage, maintain and measure the success.

Ensuring accountability and responsibility including data security—For this framework to be adopted, someone somewhere has to be in charge. There is need to draw up reporting and related policies on who does what and who is responsible for what? This is necessary to avoid the blame game. For the citizens to trust and feel secure to use services, there should be reporting mechanisms and backup strategies in place. Those responsible for service provision and maintenance should always be alert and aware. This enables those in charge to respond and react accordingly in case of an emergency.

Having considered all the results, initial experts reviews and the current literature, the authors applied all these and drew up the initial SSCIF as shown in Fig. 6. This was later revised after the experts reviews. The main points that were picked in considering the framework were summarised based on the following questions that are asked to participants: What services are needed in smart city? What are the main components of a smart city? What are the security challenges of smart cities? Who are the major stakeholders? How do we ensure citizen centric in building a security framework?

We have applied interpretivism paradigm in drawing up the key components of the framework. Based of the findings, interviews feedback, online questionnaire feeedback, document review mainly related to the Namibian environment were consired. The framework exposes the different elements that could were conisdered and could be critical in enabling the adoption of secure information infrastructure. The main components that the security should address are outlined in the SSCIF.

To satisfy the engaged theories and the design science approach, this framework has to be evaluated. The author understands that this evaluation could take many years and may not be possible to do during the study, as it may require this framework to be implemented and monitored. For the purpose of the academic requirements, the

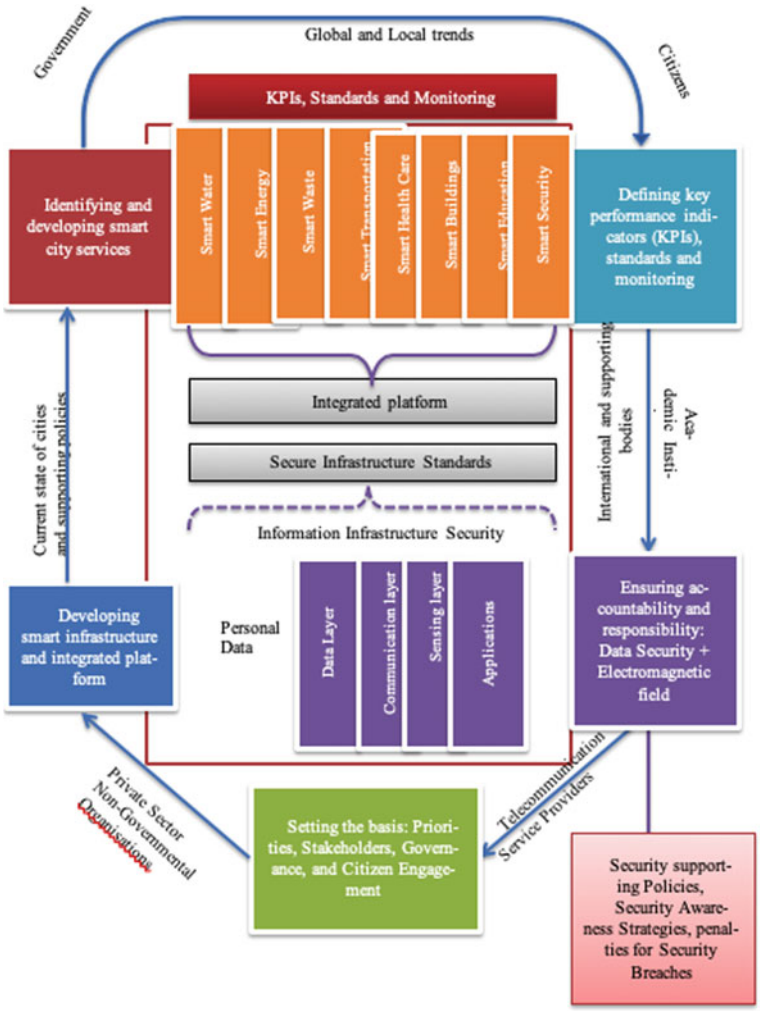


Fig. 6. SSCIF for the city of Windhoek

evaluation of the framework was done using the necessary related literature and other security models. In addition to this, evaluation was also done by engaging the Namibian experts.

6 Conclusion

The study proposes a guiding SSCIF to assist smart city stakeholders in planning and introducing smart and secure services in Namibia. This was developed by engaging Namibian citizens and related literature. The paper outlines the importance of security within smart cities. As the City of Windhoek intends to be a smart city by 2022, there is no doubt that this framework will be referred to and be of use in the Namibian

environment. The framework is developed based on lessons learnt from current smart cities, making it more realistic and suitable. The most important contribution is the fact that this study closes the current identified gap on smart cities which have not been studied on, i.e. ICT security. The City of Windhoek is expected to benefit from this framework as it is benchmarked with other current smart cities and all smart city stakeholders could have a point of reference. Most importantly, the framework combines three main aspects related to smart city services namely: smart cities, infrastructure and security. The study combines the views of the Namibian participants with current security models and benchmarks the solution to the smart city success stories in other countries.

References

1. United Nations: *World Urbanization Prospects* (2014). Retrieved 06 15, 2017 from http://esa.un.org/unpd/wup/highlights/wup2014-highlights.pdf
2. Escher Group: Five ICT essentials for smart city. In: *A White Paper for Business* (2012)
3. Amugongo, L.: *Smart Cities: Namibia Should Not Be Left Out* (2016). From http://thepatriot.com.na/index.php/2016/04/08/smart-cities-namibia-shouldnt-be-left-out/
4. Lee, J., Hancock, M.G., Hu, M.: Towards an effective framework for building smart cities: lessons from Seoul and San Fransisco. Technol. Forecast. Soc. Change **89**, 80–99 (2014)
5. Martinez-Balleste, A., Perez-Martinez, P., Solanas, A.: The pursuit of citizen's privacy: a privacy-aware smart city is possbile. IEEE Commun. Mag. **51**(6), 136–141 (2013)
6. Elmarghraby, A., Losavio, M.M.: Cyber security challenges in smart cities: safety, security and privacy. J. Adv. Res. **5**(4), 491–497 (2014)
7. Symantec Report: *Internet Security Threat Report-Symantec* (2016). Retrieved July 21, 2017 from www.symantec.com/content/dam/symantec/docs/reports/istr-21-2016-en.pdf
8. Levy-Bencheton, H., Darra, E.: *Cyber Security for Smart Cities* (2015). Retrieved July 21, 2019 from https://www.enisa.europa.eu/publications/smart-cities-architecture-model/…/fullReport
9. Harrison, C., Donnelly, I.A.: A theory of smart cities. In: *Proceeding of the 55th Annual jMeeting of ISSS* (2011)
10. Accenture: (2018) Retrieved 8 January 2020 from https://www.accenture.com/us-en/insight-smart-cities
11. Sriindia, E.: *Smart Cities as Envisioned by MOUD* (2014). Retrieved July 19, 2019 from http://www.esriindia.com/~/media/esri-india/files/pdfs/news/arcindianews/Vol9/smart-cities-envisioned-by-MoUD.pdf
12. Priyanka, N., Aditya, V.: Smart cities. Int. J. Eng. Sci. Invent. **4**(9), 43–49 (2015)
13. Visionas: *Smart Cities: An Answer to Problems of Uranisation in India* (2016). Retrieved July 17, 2019 from http://visionias.in/home/resources/open_doc/current_affairs/imp_issues/Smart
14. Ijaz, S., Munam, A.S., Khan, A., Ahmed, M.: Smart cities: a survey on security concerns. (IJACSA) Int. J. Adv. Comput. Sci. Appl. **7**(2) (2016)
15. Dohler, M., Vilajosana, I., Vilajosana, X., Llosa, J.: Smart cities: an action plan. In: Barcelona Smart Cities Congress (2011)
16. Bartoli, M., Hernandez-Serrano, J., Soriano, M., Dohler, M., Kountouris, A., Barthel, D.: Security and privacy in your smart city. In: Proceedings of the Barcelona Smart Cities Congress, pp 1–6 (2011)
17. Schmitt, C.: Security and privacy in the era of big data (2014). A RENCI/National Consortium for Data Science White Paper https://www.renci.org/wp-content/uploads/2014/02/0213WhitePaper-SMW.pdf\

18. Bohli, J., Langend-orfer, P., Skarmeta, A.: Security and privacy challenge in data aggregation for the IoT in smart cities. River Publisher Series in Communications p 225 (2013)
19. Beverland, M., Lindgreen, A.: What makes a good case study? A positivist review of qualitative case research published in industrial marketing management, 1971–2006. Ind. Mark. Manage. **39**(1), 56–63 (2010)
20. Yin, R.: Case Study Research: Design and Methods. Sage, Los Angeles, CA (2009)
21. ITU Report.: (2012) Setting the Stage for Stakeholders Engagement in Smart Sustainable Cities
22. Kulikova, O., Heil, R.V.: Cyber crisis management. A decision-support framework for disclosing security incident information. In: Inrernation Conference on Cyber Security (Cyber Security) (2012)
23. OCED.: Cyber security policy making at a turning point: analysing a new generation of national cyber security strategies for the internet economy and non-governmental perspective on a new generation of national cyber security strategies: contributions from BIAC, CSISAC and ITAC (2012)
24. European Union Agency for Network and Information Security (ENISA): *Security Framework Guidelines for Trust Services Providers* (2012). From https://www.enisa.europa.eu/
25. Cresswell, J.: Research Design: Qualitative, Quantitative and Mixed Methods Approaches. Sage Publications, London (2013)
26. Namibia Statistics Agency: Statistic Report (2018). Retrieved from https://nsa.org.na/
27. City of Windhoek: Sustainable Urban Transport Master Plan (2015). From http://www.movewindhoek.com.na/sites/default/files/downloads/sutmp_project_overview_and_summary_report.pdf
28. Harambee Prosperity Plan: (2016). From http://www.gov.na/documents/10181/264466/HPP+page+70-71.pdf/bc958f46-8f06-4c48-9307-773f242c9338
29. MTC Report: *MTC Annual REport 2015* (2015). Retrieved July 21, 2017 from www.mtc.com.na/sites/annual-reports/2015/index.html
30. Giovanis, A.N., Binioris, S., Polychronopoulos, G.: An extensionof TAM model with IDT and security/privacy risk in the adoption of internet banking services in Greece. Euromed J Bus **7**, 24–53 (2012)
31. Martins, et al.: Understanding the Internet banking adoption: a unified theory of acceptance and use of technology and perceived risk application. Int. J. Inf. Manage. **34**(2014), 1–13 (2013)
32. Jabareen, Y.: Building a conceptual framework: philosophy, definitions, and procedure. Int. J. Qual. Methods. Technion—Israel Institute of Technology Technion City, Haifa, Israel (2009)

A Multigraph Approach for Supporting Computer Network Monitoring Systems

Francesco Colace[1], Muhammad Khan[2], Marco Lombardi[1], and Domenico Santaniello[1](✉)

[1] DIIN, University of Salerno, Fisciano, Italy
{fcolace,malombardi,dsantaniello}@unisa.it
[2] New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
mk7406@nyu.edu

Abstract. The pervasiveness of information technologies has reached very high levels: most human activities involve the use of sensor-based systems connected to the network. The increasingly widespread use of the Internet of things has significantly improved our quality of life but has introduced a series of new problems, especially from the security point of view. Protecting these systems from cyber-attacks has become a priority as possible malfunctions can lead to issues with a significant social impact. Imagine, for example, computer attacks on smart cars connected to the network or remotely controlled electrical or water systems. Protecting this type of system is a complex task as there are many elements to consider and the data to be monitored. An analysis able to foresee eventual attacks through the study of the data and their variations could be a useful tool to prevent malfunctions. This paper proposes a methodology based on the integrated use of three graphic models to address the problem of preventing attacks on pervasive systems from three different perspectives: probabilistic, contextual, and ontological. The paper proposes the use of Bayesian networks built through an ontological definition of the problem dropped on a particular context represented by a Context Dimension Tree—the proposed approach experiments in a real scenario providing satisfactory results.

Keywords: Network security · Knowledge management · Probabilistic graphical models · CDT · Bayesian network · Ontology

1 Introduction

Today's society is heavily oriented toward digitization, which increasingly affects the management of cities and services addressed to the citizens. This relentless process takes advantage of the innovation boost and the growing needs of users related to the smart solutions of their daily activities. The paradigm that best describes this propulsive thrust

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_46

is represented by the Internet of things (IoT), where "things" are interconnected and able to exchange information, accessing local network services, and the Internet [1–5]. In this scenario, based on the continuous exchange of information on the network, an increasingly significant role is played by systems able to guarantee data security [6–9]. This problem is usually dealt with through methodologies and technologies oriented to network control, an example is represented by the Intrusion Detection System (IDS). These tools analyze packets containing information exchanged within the network to identify possible threats. Unfortunately, these systems, in many cases, work simply by comparing the packets content with a database containing the threats identified, remaining utterly blind to any new attempts to attack, and not ensuring the opportunity to build a picture of what will happen in the future. Furthermore, it remains necessary to deepen the study of techniques capable of supporting these systems, and focusing on methodologies not only able to recognize and classify a Denial-of-Service (DoS) attack based on traffic but also to concern data behavior.

This work aims to introduce an approach to recognize and manage problems related to network security through the adoption of different graphs to represent several aspects of the problem. This multilevel approach involves the use of various graphical formalisms capable of inferring the automatic identification and resolution of the issues related to network security. The three graphic structures identified (CDT, Ontology, and Bayesian network) allow, based on the occurrence of certain events, to identify, find the best solution, or predict any problems. After presenting the related works and the main issues related to network security, the proposed methodology will be introduced, its application in a real scenario, and the relevant experimental results will close the article.

2 Related Works

In literature, several articles deal with the problem of network security in the Internet of things or scenarios typical of the Smart City [10–13]. The general approach provides the introduction of a framework to collect information coming from networks and to manage any attacks or malfunctions. An exciting first approach is provided by Elsaeidy [14], which introduces a system based on deep learning and user data behavior. This approach guarantees proactively identifying any attacks or recognition of patterns, which may highlight potential attacks. A specialization of the previously introduced method is found in [15], where an approach based on artificial intelligence techniques is applied in the industrial sector to identify potential problems or risks. This procedure involves the definition of trees of possible attacks to develop defence strategies. Besides, the graph-based approach seems to provide promising results, as witnessed by Hawrylak's article [16], where an attack modeling system called Hybrid Attack Graph (HAG) is introduced. This approach provides graphic modeling that allows the immediate combination of the physical and software component of the attacks to form a complete picture of the potential risk. Moreover, from this analysis, it is easier to identify possible solutions approaches. A further graph-based method is introduced in [17] where employing probabilistic graphic structures, and data mining techniques are used to determine the type of attack. In particular, techniques based on multilayer perceptron, Naive Bayes, and random forest have been introduced. The method showed in this article shows that the advantage of

these experiences, adds a multilevel graph approach to allow a more in-depth analysis of the status of a network. Through the combined use of Context Dimension Tree, Bayesian network, and Ontology, it is possible to integrate semantic, probabilistic, and context-aware approaches in a single methodology.

3 The Proposed Methodology

The proposed approach would be a methodology able to predict and capable of adapting itself to a context useful in several fields [18–20]. In this paper, we will evaluate the proposed approach to forecast cybersecurity attacks. The proposed methodology is based on the combination of three graph approaches as Ontology, Context Dimension Tree (CDT), and Bayesian network. Bayesian networks are probabilistic graphical models to predict the probability of an event [21], adequately supporting scenarios already managed through context representation approaches. The CDT is a tree spread used to represent all possible contexts [22]. Many of the CDT applications provide for the management and personalization of information to be distributed to specific users [23]. One of the most used tools for reality representation are Ontologies. An Ontology can provide a formal, shared, and explicit description of a domain of interest, through which it is possible to support context-aware systems [24]. Bayesian networks and Ontologies are strongly interconnected; in particular, it is possible to build Ontologies through the use of Bayesian networks [25], and vice versa, it is possible to construct Bayesian networks through domain Ontologies [26].

Figure 1 shows the system architecture, which, by collecting raw data, uses them to return the appropriate usage application. In the first phase, there is the collection of data from the IDS and other sensors allocated in the computer network, which are stored raw. These data are harmonized and sorted in the pre-processing phase and stored in a database that powers the inference engine. Inside the inferential engine are the three graph views previously described (CDTs, Ontologies and Bayesian networks) which provide an interpretation of the knowledge acquired.

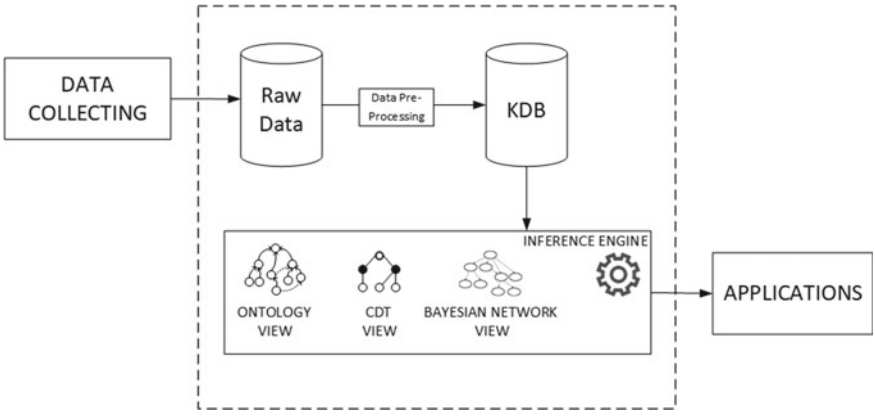


Fig. 1. System architecture

In practice, the detection of malicious attacks, given a particular context, could be done through the described architecture that exploits the right information characterized by innovative elements based on formal context representation, knowledge management organization, and inferential engines [27].

4 Experimental Results

To perform the experimentation phase, an ad hoc dataset was used, which contains data from the Intrusion Detection System to monitor a university computer network. This dataset, which includes over eleven thousand instances, is a reduced sample of the global monitoring database that analyzes this network since the year 2008. In particular, the dataset used contains almost one thousand attempts to attack, of which only about 5% succeeded in penetrating the computer network.

To predict cyber-attacks occurring in a Computer network, analyzing the performance of the presented approach, it was necessary to divide the dataset into the testing set (10%) and training set (90%). Two cases were considered to perform the analysis: Case 1 involves the only use of the structural learning algorithm and Case 2 in which the proposed methodology is used. The obtained results are shown in terms of confusion matrix overall accuracy (1), Recall (2), and Precision (3).

$$A = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3}$$

Starting from the selected dataset, five categories of attacks have been selected:

1. Distributed Denial of Service—DDoS.
2. Denial of Service—DoS.
3. Web Deface—WD.
4. Spear Phishing—SPh
5. Password Harvesting—PH.

The confusion matrix of Case 1 (Table 1) refers to the Bayesian network structure learned using only the structural learning algorithm. As shown in Table 1, the Bayesian network can correctly classify some events, in particular, many DoS attacks. Due to the high number of incorrectly organizing events, this Bayesian network does not guarantee excellent results in terms of overall accuracy, nor, as shown in Table 3, in terms of Precision and Recall.

The confusion matrix of Case 2 (Table 2) represents the case in which the proposed methodology was used to build the Bayesian network. In this case, the confusion matrix improves in terms of correct classification of events. The number of correctly classified

Table 1. Confusion matrix Case 1

		Reference				
		DDoS	DoS	WD	SPh	PH
Prediction	DDoS	104	48	33	21	51
	DoS	52	142	17	35	43
	WD	11	51	143	25	24
	SPh	14	15	13	95	32
	PH	7	33	24	17	97
	Overall accuracy: 50.65%					

Table 2. Confusion matrix Case 2

		Reference				
		DDoS	DoS	WD	SPh	PH
Prediction	DDoS	192	34	7	5	19
	DoS	31	221	10	9	18
	WD	9	22	189	21	13
	SPh	20	9	15	117	8
	PH	3	12	2	12	149
	Overall accuracy: 75.68%					

events is increased, and the incorrectly classified activities are decreased, compared to the previous case.

This improvement can be seen in the increase of overall accuracy, which exceeds 75% and highlighted in Table 3, in terms of Precision and Recall, where can be seen a significant increase.

Table 3. Precision and Recall parameters Case 1 and Case 2

		DDoS	DoS	WD	SPh	PH
Case 1	Precision (%)	40.47	49.13	56.30	56.21	54.49
	Recall (%)	55.32	49.14	62.17	49.22	39.27
Case 2	Precision (%)	74.71	76.47	74.41	69.23	83.71
	Recall (%)	75.29	74.16	84.75	71.34	71.98

Even if the value over 75% of accuracy may seem a reasonable result, compared to that obtained by the network learned through the machine learning algorithm, it could not

be a great result in absolute terms of forecasting. However, it can be said that our system (different from common IDS systems) does not only just do a package analysis based on a comparison with a malicious packages archive but also attempts to recognize cyber-attacks based on several other factors that affect the network. In particular, it can monitor and acknowledge related patterns based on other data, such as the bandwidth demand in a given time, the types of demand, the number of particular requests compared to time, and so on. This approach is, therefore, able to predict possible unknown cyber-attacks, exploiting other data, such as contextual information provided by graph approaches: Context Dimension Tree and Ontology. Such a system is, therefore, not intended to replace modern methods for the detection of cyber-attacks but is particularly suitable for the support of such systems. Overall, the proposed approach can increase its performance according to the increase in available data.

5 Conclusions

This article aimed to experiment with a multilevel graph methodology for the prediction of attacks on the computer network at the service of a university institution. The dataset used contains almost one thousand attempts at the attack, of which only about 5% managed to penetrate the computer network. The analysis was carried out considering two cases, the first where the Bayes network is learned through an automatic learning algorithm, and the second where the proposed methodology is used. Subsequently, another aspect is evaluated, which is to understand how the network learned in the second case behaves in prediction of the only events that have been successful in the cyber-attack.

As can be seen from the confusion matrices (Tables 1 and 2) and Table 3, compared to forecasts obtained through a Bayesian network traditionally built with an automatic learning algorithm, our system has been able to maintain good performance in terms of Accuracy, Prediction, and Recall. However, even if this result is not convincing in absolute classification terms, the real strength of the system lies in the use of graph approaches that allow contextualizing the problem by predicting and classifying events based on the behavior of the data.

As analyzed, we can say that the proposed system could be used in combination with traditional intrusion detection systems acting in cases where the events that occur are not known or are not detected, by comparison, further increasing the security of computer networks. According to the nature of the approach used, we can say that there are two factors for which this methodology could improve its performance. The first is to increase the data available; in fact, an increase in available data would help to build a more reliable Bayesian network. The second is related to the nature of the methodologies used for the designed system. In fact, according to graphical approaches such as CDT and Ontologies, it is possible to allow the system to exchange information with other systems that face similar problems on different domains, exploiting the competences of the other systems, to enhance their performances.

References

1. Ashton, K.: That 'internet of things' thing. RFID J. **22**(7), 97–114 (2009)
2. Carratu, M., Ferro, M., Pietrosanto, A., Sommella, P., Paciello, V.: A smart wireless sensor network for PM10 measurement. In: 2019 IEEE International Symposium on Measurements & Networking (M&N), pp. 1–6 (2019). https://doi.org/10.1109/IWMN.2019.8805015
3. Ugwiri, M.A., Carratu, M., Liguori, C., Paciello, V.: Bridge dynamic analysis for accurate control and monitoring: Feasibility study. In: IECON 2019—45th Annual Conference of the IEEE Industrial Electronics Society, pp. 5307–5311 (2019). https://doi.org/10.1109/iecon.2019.8927692
4. Casillo, M., Clarizia, F., Colace, F., Lombardi, M., Pascale, F., Santaniello, D.: An Approach for Recommending Contextualized Services in e-Tourism. Information **10**(5), 180 (2019). https://doi.org/10.3390/info10050180
5. Amato, F., Mazzeo, A., Moscato, V., Picariello, A.: Semantic management of multimedia documents for e-government activity. In: 2009 International Conference on Complex, Intelligent and Software Intensive Systems, pp. 1193–1198 (2009). https://doi.org/10.1109/cisis.2009.195
6. Castiglione, A., Dambrosio, C., De Santis, C., Castiglione, A., Palmieri, F.: On secure data management in health-care environment. In: Proceedings—7th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IMIS 2013, (2013). https://doi.org/10.1109/imis.2013.120
7. Castiglione, A., Palmieri, F., Colace, F., Lombardi, M., Santaniello, D.: Lightweight ciphers in automotive networks : a preliminary approach. In: 2019 4th International Conference on System Reliability and Safety, pp. 142–147 (2019)
8. Khan, M., Al-Raweshidy, H.S.: Exploiting the capacity-routing ability of a cloud radio access network. In: 9th International Conference on Information and Communication Technology Convergence: ICT Convergence Powered by Smart Intelligence, ICTC 2018, (2018). https://doi.org/10.1109/ictc.2018.8539427
9. Amato, F., Moscato, F., Moscato, V., Colace, F.: Improving security in cloud by formal modeling of IaaS resources. Futur. Gener. Comput. Syst. **87**, 754–764 (2018). https://doi.org/10.1016/j.future.2017.08.016
10. Essa, A., Al-Shoura, T., Al Nabulsi, A., Al-Ali, A.R., Aloul, F.: Cyber physical sensors system security: threats, vulnerabilities, and solutions. In: 2018 2nd International Conference on Smart Grid and Smart Cities (ICSGSC), pp. 62–67 (2018). https://doi.org/10.1109/icsgsc.2018.8541316
11. Chakrabarty, S., Engels, D.W. (2016) A secure IoT architecture for smart cities. In: 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), pp. 812–813 (2016). https://doi.org/10.1109/ccnc.2016.7444889
12. Pacheco, J., Hariri, S.: IoT security framework for smart cyber infrastructures. In: 2016 IEEE 1st International Workshops on Foundations and Applications of Self Systems (FAS*W), pp. 242–247 (2016). https://doi.org/10.1109/fas-w.2016.58
13. Amato, F., Mazzeo, A., Penta, A., Picariello, A.: Knowledge representation and management for e-government documents. In: E-Government ICT Professionalism and Competences Service Science, Boston, MA: Springer US, pp. 31–40. https://doi.org/10.1007/978-0-387-09712-1_4
14. Elsaeidy, A., Elgendi, I., Munasinghe, K.S., Sharma, D., Jamalipour, A.: A smart city cyber security platform for narrowband networks. In 2017 27th International Telecommunication Networks and Applications Conference, ITNAC 2017, vol. 2017-Janua, pp. 1–6 (2017) https://doi.org/10.1109/atnac.2017.8215388

15. Falco, G., Viswanathan, A., Caldera, C., Shrobe, H.: A master attack methodology for an ai-based automated attack planner for smart cities. IEEE Access **6**, 48360–48373 (2018). https://doi.org/10.1109/ACCESS.2018.2867556
16. Hawrylak, P.J., Haney, M., Papa, M., Hale, J.: Using hybrid attack graphs to model cyber-physical attacks in the Smart Grid. In: Proceedings—2012 5th International Symposium on Resilient Control Systems, ISRCS 2012, pp. 161–164 (2012) https://doi.org/10.1109/isrcs.2012.6309311
17. Alkasassbeh, M., Al-Naymat, G., Hassanat, A.B., Almseidin, M.: Detecting distributed denial of service attacks using data mining techniques. Int. J. Adv. Comput. Sci. Appl., **7**(1), (2016)
18. Colace, F., Lombardi, M., Pascale, F., Santaniello, D., Tucker, A., Villani, P.: MuG : a multilevel graph representation for big data interpretation. In: IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems, pp. 1410–1415 (2018). https://doi.org/10.1109/hpcc/smartcity/dss.2018.00233
19. Clarizia, F., Colace, F., Lombardi, M., Pascale, F., Santaniello, D.: A multilevel graph approach for road accidents data interpretation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11161 LNCS, pp. 303–316 (2018) https://doi.org/10.1007/978-3-030-01689-0_24
20. Colace, F., De Santo, M.. Lombardi, M., Pascale, F., Santaniello, D., Tucker, A.: A Multilevel graph approach for predicting bicycle usage in London Area. In: Yang, X.-S., Sherratt, S., Dey, N., Joshi, A. (eds.) Fourth International Congress on Information and Communication Technology. Advances in Intelligent Systems and Computing, vol. 1027, pp. 353–362. Springer, Singapore (2020). https://doi.org/10.1007/978-981-32-9343-4_28
21. Weber, P., Medina-Oliva, G., Simon, C., Iung, B.: Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. Eng. Appl. Artif. Intell. (2012). https://doi.org/10.1016/j.engappai.2010.06.002
22. Colace, F., De Santo, M., Lombardi, M., Santaniello, D.: CHARS: a cultural heritage adaptive recommender system. In:Proceedings of the 1st ACM International Workshop on Technology Enablers and Innovative Applications for Smart Cities and Communities—TESCA'19, pp. 58–61 (2019). https://doi.org/10.1145/3364544.3364830
23. Casillo, M., Clarizia, F., D'Aniello, G., De Santo, M., Lombardi, M., Santaniello, D.: CHATBot: a cultural heritage aware teller-bot for supporting touristic experiences. Pattern Recognit. Lett. (2020). https://doi.org/10.1016/j.patrec.2020.01.003
24. Chen, H., Finin, T., Joshi, A.: An ontology for context-aware pervasive computing environments. Knowl. Eng. Rev. **18**(3), 197–207 (2003)
25. Colace, F., De Santo, M.: Ontology for E-learning: a bayesian approach. IEEE Trans. Educ. **53**(2), 223–233 (2010)
26. Helsper, E.M., Van Der Gaag, L.C.: Building bayesian networks through ontologies. In: ECAI2002, Proceedings of the 15th European Conference on Artificial Intelligence, pp. 680–684 (2002)
27. Clarizia, F., Colace, F., De Santo, M., Lombardi, M., Pascale, F., Santaniello, D., Tuker, A.: A multilevel graph approach for rainfall forecasting: a preliminary study case on London area. Concurr. Comput. Pract. Exp. e5289 (2019). https://doi.org/10.1002/cpe.5289

Comparing Phonetic Reduction in Russian as First and Second Language: From Psycholinguistics to Natural Language Processing

Elena I. Riekhakaynen(✉)

Saint-Petersburg State University, Universitetskaya emb. 7/9, 199034 St. Petersburg, Russia
e.riehakajnen@spbu.ru

Abstract. In the paper, we argue that it is necessary to collect and analyze casual speech of the speakers from different age groups, both native speakers and those who study a language as a second one, in order to understand the mechanisms of spoken word production and recognition and to improve current automatic systems of natural language processing. We provide a short overview of the corpora of adult, children and adolescent Russian speech we develop and then focus on the methodology and results of a study of phonetic reduction in the speech of 16 Chinese students learning Russian as a foreign language. We found out similar tendencies of phonetic reduction in the speech of the Chinese students and in the speech of native speakers of Russian. At the same time, the speech of Chinese students, unlike the speech of Russian-speaking children aged four to six years, is characterized by a large number of examples with sound changes. The second language learners of Russian usually have different realizations of one and the same word in their speech. The results we obtained can be used for spoken word recognition modeling, as well as for various educational purposes including teaching Russian as a foreign language.

Keywords: Speech signal processing · Russian · Phonetic reduction · Corpora of spontaneous speech

1 Introduction

Voice assistants such as Siri, Google Assistant, Alexa, etc., are becoming more and more popular nowadays. Normally people want to talk to them in their natural manner, i.e., not changing the speed of their speech and their pronunciation habits. However, casual speech still causes problems to automatic speech recognition models, even those based on neural networks [1–3]. We assume that there are at least two reasons for the shortcomings of voice assistants and automatic speech recognition models in general.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_47

The first one is the omission of sounds (so-called phonetic reduction) that occurs both in prepared and casual speech but is much frequent in the latter. Although there are many linguistic studies on reduction in different languages (see, for example, [4–8]), we still do not know the rules of reduction in casual speech and how a listener copes with it while processing spontaneous speech. The second reason for the imperfection of the automatic speech recognition systems is probably the fact that they are usually trained on the speech of adult native speakers who do not have any speech disorders. However, such factors as the age of a speaker and his/her level of proficiency in a language can influence the speech a lot. For these reasons, it is necessary to collect and analyze casual speech of the speakers from different age groups, both native speakers and those who study a language as a second one.

From a psycholinguistic point of view, it is worth comparing the production of spontaneous speech in native and foreign languages, because the analysis of mistakes made while speaking first and second languages allows us to put forward hypotheses on the differences between first and second language acquisition, which is necessary for a more detailed understanding of the speech processing in general. The results of such studies can be used for evaluating second language learners' communication skills.

As it was mentioned before, the unprepared oral speech of a native speaker is characterized by a number of features, the most striking of which include phonetic reduction of word forms and various types of hesitations. The hesitations in the speech of second language learners of Russian have already become the object of several studies (e.g., [9]), whereas the analysis of reduced word forms in Russian as a second language, as far as we know, has not been previously conducted. In studies performed on the material of other languages (see, for example, [10–13]), the focus is primarily on the perceptual aspect, i.e., the ability of foreigners to choose the correct interpretation of a phonetically incomplete signal.

However, if we assume that at least part of the reduced realizations can be stored in the mental lexicon of a speaker and a listener (see the discussion of this issue, for example, in [14]), the data on the production of reduced units by foreigners in comparison with native speakers should definitely help to understand the ways in which reduced units enter the mental lexicon. The results of such a study can be used for modeling the structure of the mental lexicon of a listener as a part of an automatic algorithm of spoken word recognition.

We consider a word form to be reduced if it has at least one omitted sound if compared to the canonical realization prescribed by the norms of standard Russian. About 20–25% of all word forms are usually reduced in casual speech [4–8]. The frequency of word forms is believed to be one of the main reasons for phonetic reduction [8]. There are numerous studies of the phonetic realizations of frequent word forms in casual speech (see, for example, the following papers for the Russian language: [14–17]). Most high-frequency word forms have several different variants of phonetic realization, and different variants can be found even in the speech of one and the same speaker within the same message.

The rest of the paper is organized as follows. In Sect. 2, we provide an overview of the spoken data we have for different age groups of native Russian speakers: adults,

children and adolescents. In Sect. 3, we focus on the results we got while analyzing the speech of second language learners of Russian. The conclusions are provided in Sect. 4.

2 Our Data from Native Speakers of Russian

2.1 Adult Speech

We have been developing a corpus of spontaneous Russian speech since 2009. It currently contains around three hours of recordings (talk shows, radio interviews, announcements by professional newsreaders) provided with an orthographic and acoustic-phonetic annotation (more than 22 thousand words). A frequency wordlist that contains all pronunciation variants of each word was created based on the corpus. Both the corpus and the wordlist are available online (http://russpeech.spbu.ru/). The transcription we use for the annotation of the corpus is close to X-SAMPA symbols. We use one and the same principles of annotation for all our corpora. For more information on this corpus, as well as on the corpus of children's speech, see [17].

2.2 Children Speech

In order to get the information on how children pronounce frequent words, we conducted an experiment. The study involved 71 children from middle and senior groups of two kindergartens in St. Petersburg, Russia. The experiment was conducted with each child separately and lasted an average of 5–10 min. The main part of the experiment was a role-playing game "shop keeping", in which each child acted as both a shop-assistant and a customer. In addition, the experimenter asked every child questions about what the child likes to do during the day in kindergarten and about his favorite toys. The experimenter's questions were composed so that they provoked the child to use a number of high-frequency words (for example, the words *only, when, because, if, now,* etc.). While choosing such high-frequency units, we focused on a list of high-frequency words, for which it is known that they often undergo reduction in the speech of adult native speakers of the Russian language (see, for example, [14–17]). The parents of the participants signed a written consent to record their children's speech in both parts of the experiment. The experiment was approved by the administration of the kindergartens.

2.3 Adolescent Speech

The corpus of adolescent speech includes the recording that we also got in an experiment. 28 school children aged from 13 to 17 years performed several tasks: took part in an interview, played a role-playing game, described several pictures and read two dialogues. The experiment was conducted in accordance with the Declaration of Helsinki and the existing Russian and international regulations concerning ethics in research. We received written consents from the parents of all participants to record their children's speech. As we used very similar tasks and the principles of data analysis for the experiment with the adolescents and the experiment with second language learners (see Sect. 3), further details on the experimental procedure will be provided in the next section of the paper.

The main reason for using similar tasks is that we are going to compare the tendencies of reduction in the speech of these two groups. We created a database with all realizations of every frequent word in adolescent speech. Thus, we could compare the realizations in the speech of the adolescent native speakers to the realizations of the same words in the speech of second language learners.

3 Realizations of Frequent Russian Words in the Speech of Chinese Students

3.1 Material

To get the data for second language learners of Russian, we conducted an experiment that was very similar to the experiment with the Russian adolescents. The experiment consisted of four blocks. The first block was an interview with the experimenter. It included several questions regarding everyday topics (studying at university, living in Russia, free time, day plans, etc.). After that, the participants were invited to talk about their family. In the second block of the experiment, informants were asked to play a role-playing game with the experimenter and act first as a customer, and then as a seller in a university cafeteria. The third task was to describe several pictures and answer to questions about them. The first three pictures depicted some of the actions of people (jumping from a cliff; meeting a homeless puppy; watching a melodrama). The informants were asked to look at the pictures and answer the question of whether they are capable of committing such acts. The response model was as follows: "I am capable of this only if…" or "Never at all." The following three pictures depicted funny dogs with different emotions. The informants were asked to look at the pictures and answer the question as quickly as possible, starting the answer with the words "Oh, it's me when…". We provided the participants with the options for starting the answers, because we wanted to get the realizations of the frequent word forms *kogda* "when", *nikogda* "never", *tol'ko* "only", *esli* "if", *voobsche* "at all". In the last part of the experiment, we asked the informants to read aloud two dialogues that contained many word forms often subjected to reduction in spontaneous speech. All recordings were made in WAV format, sampling frequency was 44,100 Hz. 64 files were recorded with a total duration of 3 h 53 min 24 s.

16 Chinese students studying Russian aged 21–24 years (eight female) took part in the experiment. All of them are native speakers of the Chinese language studying Russian for at least four years and having already passed B1 or B2 test of proficiency in the Russian language. 11 informants are native speakers of the northern Chinese dialect (seven fem.) and five informants are native speakers of the southern Chinese dialect (one fem.). The experiment was conducted in accordance with the Declaration of Helsinki and the existing Russian and international regulations concerning ethics in research. Before starting the recording, the experimenter received the consent of each informant to conduct an interview and to record the informant's speech on the recorder.

The orthographic annotation of all records was performed in the Praat program (http://www.fon.hum.uva.nl/praat/). An acoustic-phonetic transcription was carried out for those frequent word forms which (taking into account previous studies with native

Russian speakers) we expected to be reduced in the speech of our participants. A word form with a complete quantitative loss of at least one element was considered reduced. Qualitative changes in word forms were also recorded and analyzed separately. In this case, the quality of an unstressed vowel after a hard consonant was not taken into account, since the vowels are often subjected to qualitative and quantitative reduction in this position in Russian, even in prepared speech, and in some cases, it is difficult to determine the quality of the vowel.

3.2 Results

We revealed 1380 realizations of the mentioned above words of high frequency, of which 505 (37%) underwent reduction. In 553 examples (40%), we found sound substitutions. Reduced variants and variants with sound substitutions are intersecting sets, because there are realizations with both loss and substitutions of sounds. The data on the number of reduced realizations and realizations with sound changes in the "reading" part and in all talking blocks together are presented in Table 1.

Table 1. Comparison of the realizations in "reading" and "talking" parts of the experiment

	Reading	Talking
Number of realizations analyzed	815	565
Number and percentage of reduced realizations	251 (31%)	254 (45%)
Number and percentage of realizations with substitutions of sounds	316 (39%)	237 (42%)

Most of the examples of the words selected for the analysis (815 out of 1380) were found in the "reading" block, since the tasks of this block contained many high-frequency words. In both blocks, the number of unreduced implementations of the words analyzed is significantly higher than the number of reduced ones (according to the binomial test, $p < 0.001$ for the "reading" block; $p = 0.009$ for the "talking" block); in the "speaking" block, reduced realizations are found much more often than in the "reading" block ($\chi 2 = 28.8$, df $= 1$, $p < 0.01$). This trend is generally expected and corresponds to what is observed in the speech of native speakers: when reading, which is a more prepared form of speech, the phonetic reduction of word forms is less common.

The difference between the blocks in the number of realizations with sound changes is not statistically significant ($\chi 2 = 1.4$, df $= 1$, $p = 0.237$). It is worth mentioning that in the "reading" block, implementations with sound substitutions were even more common than reduced realizations.

The largest number of different realizations was found for words *ochen'* "very" (37 different variants), *potomu chto* "because" (34) and *zdravstvujte* "hello" (33). Most probably, for the words *potomu chto* "because" and *zdravstvujte* "hello", this is due to the length of the words and, therefore, the fact that these words potentially have more positions in which qualitative or quantitative changes can occur than in one- and two-syllable words. In addition, the word *zdravstvujte* "hello" has several consonant

clusters which can be simplified (cf. numerous variants of this word form in the speech of Russian-speaking children).

For the word form *ochen'* "very" one can offer several explanations for such a significant variety of options. Firstly, since there are diphthongs in the Chinese language, informants often pronounced two vowels at the beginning of the word: [uo], [əo], [ao], [au]. Secondly, instead of the affricate [tɕ], the Chinese students pronounced voiceless soft consonant [tʲ], and sometimes voiced soft [dʲ]. It is worth noting that despite the fact that in the texts of the "reading" block the word form *ochen'* "very" was mentioned only once and, therefore, we expected to get only 16 realizations of this word (one from each informant), it turned out to be the most frequent for the whole study: we obtained 137 realizations of this word, i.e., 121 examples were obtained from the "talking" block.

The smallest number of different realizations (three and four, respectively) was found for word forms *sebja* "myself" and *govorju* "I say". This is due to the fact that these words were not included in the "reading" block, and informants almost did not use these words in conversational blocks.

For 26 words, canonical realizations prevailed over the reduced ones. However, there are words whose reduced realizations were more frequent in our material than canonical ones: *segodnja* "today", *pozhalujsta* "please", *zdravstvujte* "hello", *budet* "will be", *kak skazat'* "how to say it". In the speech of adult native speakers of the Russian language, some of the high-frequency word forms we have analyzed are also more often used in the canonical form than in a reduced one, but words like *menja* "me (Gen./Acc.)", *esli* "if", *tebja* "you (Gen./Acc.)", *vsjo-taki* "nevertheless" usually appear in the speech of native speakers of the Russian language in a reduced form [14], whereas in the speech of the Chinese students, the canonical variants turned out to be predominant.

4 Discussion

The main tendencies of the phonetic reduction demonstrated by the Chinese students are similar to the trends observed both in adult native speakers of Russian and in children aged four to six years. This is the loss of elements of consonant clusters (in particular, initial consonants (**s***pasibo* "thank you") and the omissions of the first consonants in the intervocalic consonant clusters (*sko**l'**ko* "how many/much", *ko**g**da* "when", etc.), loss of the consonant [j] and the reduction of unstressed vowels. In addition, some typical reduced realizations in the speech of the Chinese students were the same as the typical reduced realizations in the speech of native speakers of the Russian language ([kə'da], [mnʲæ], [sʲi'vonʲæ], etc.).

The substitutions of sounds we revealed in the material are primarily due to the influence of the phonological system of the native (Chinese) language. This finding indicates that even with the level of Russian language proficiency not lower than B1, not all oppositions forming the phonological system of the Russian language are acquired by the Chinese students. 30% of all substitutions of sounds encountered in the material were the substitutions [d] <-> [t] or [dʲ] <-> [tʲ]. In 18% of the examples, the substitutions [g] <-> [k] and [gʲ] <-> [kʲ] were observed.

We also analyzed the tendencies of high-frequency words usage in the speech of each of the informants. It turned out that unlike Russian-speaking children aged four to

six years the foreign students, as a rule, do not give preference to any of pronunciation options. The variability of realizations was observed in the speech of all informants (for example, in the speech of one student the following realizations of the word *kogda* "when" were encountered: [ə'gda], [k'əta], [kə'kta], [kə'gda]).

We were not able to identify any distinctive features of the realizations of high-frequency words in Russian speech of native speakers of the northern or southern dialects of the Chinese language. However, it can be noted that in the speech of the girls, the reduced realizations were more common than in the speech of the boys. For example, only girls did often use the variants with the omission of the initial consonants ([ətɕtʲu:], [ətɕtʲi] for *pochti* "almost"; [nʲe:] for *mne* "me (Dat.)"). In addition, for the word *ochen'* "very" the realization [otɕi] turned out to be the most frequent in the speech of the girls, and the [otɕin] option was the most frequent in the speech of the young men, regardless of the informant's native dialect.

5 Conclusions

The study showed that in the speech of the Chinese students who have B1 or higher level of the Russian language, the same reduction trends are observed that are characteristic of native speakers of the Russian language if we regard individual sounds. Apparently, the formation of these strategies is one of the basic mechanisms for mastering natural spoken language, as the analysis of the speech of the Russian-speaking children aged four to six years also revealed the same strategies [17].

At the same time, both in the speech of the Chinese students and in the speech of the Russian-speaking children, there are a significant number of substitutions of sounds in high-frequency words (compared to the speech of adult native speakers of Russian). However, the causes of this phenomenon for children and for foreigners are most probably different. Whereas native speakers replace sounds because of the imperfection of articulatory skills, the Chinese students who have been studying Russian for at least four years still encounter a number of phonetic difficulties in Russian spoken language caused by the influence of the phonological system of their native language. The differences in mastering the phenomenon of reduction of words when acquiring the Russian language as a mother tongue and while learning it as a foreign language are also evident from the fact that most children prefer to use one pronunciation option for each high-frequency word form, while in the speech of every foreign student there are normally at least several different options (primarily due to the sounds substitutions, since the number of positions in which sound omissions may occur is more limited).

The results of the research are relevant for automatic speech recognition and synthesis. On the one hand, the tendencies of sound omissions that are common for native speakers and second language learners should be regarded as the first to be modeled while synthesizing spontaneous speech. On the other hand, the differences in pronunciation of frequent words described in the study can be incorporated in the models of spoken word recognition as indicators of native versus non-native speech signal. Further analysis and comparison of the data from native and non-native casual speech should definitely contribute to the elaboration of stricter rules.

Acknowledgements. The study is supported by the grant No MK-6776.2018.6 from the President of the Russian Federation. I would like to thank my student Iuliia Telova who helped me to gather and annotate the data.

References

1. Mizera, P., Pollak, P.: Improving of LVCSR for Causal Czech Using Publicly Available Language Resources. In: Karpov, A., Potapova, R., Mporas, I. (eds.) Speech and Computer. SPECOM 2017. LNCS, vol. 10458, pp. 427–437. Springer, Cham (2017)
2. Ravanelli, M., Serdyuk, D. Bengio, Y.: Twin Regularization for online speech recognition. In: Proceedings of Interspeech (2018). https://arxiv.org/pdf/1804.05374.pdf. Last accessed 30 Dec 2019
3. Kipyatkova, I.: Improving Russian LVCSR using deep neural networks for acoustic and language modeling. In: Karpov, A., Jokisch, O., Potapova, R. (eds.) Speech and Computer. 20th International Conference SPECOM 2018. LNAI, vol. 11096, pp. 291–300. Springer, Cham (2018)
4. Nigmatulina, Ju, Rajeva, O., Riechakajnen, E., Slepokurova, N., Vencov, A.: How to study spoken word recognition: evidence from Russian. In: Anstatt, T., Gattnar, A. (eds.) Slavic Languages in Psycholinguistics: Chances and Challenges for Empirical and Experimental Research, Tübinger Beiträge zur Linguistik, vol. 554, pp. 175–190. Narr Francke Attempto Verlag, Tübingen (2016)
5. Johnson, K.: Massive reduction in conversational American English. In: Yoneyama, K., Maekawa, K. (eds.) Spontaneous speech: Data and analysis. Proceedings of the 1st Session of the 10th International Symposium, pp. 29–54. The National International Institute for Japanese Language, Tokyo (2004)
6. Ernestus, M.: Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface, PhD Thesis, Utrecht (2000)
7. Shockey, L.: Sound patterns of spoken English. Blackwell Publishing, Oxford (2003)
8. Jurafsky, D., Bell, A., Gregory, M., Raymond, W.: Probabilistic relations between words: Evidence from reduction in lexical production. In: Bybee, J., Hopper, P. (eds.) Frequency and the emergence of linguistic structure, pp. 229–254. John Benjamins, Philadelphia PA (2000)
9. Prokaeva, V.: Hesitations in the first (Japanese) and the second (Russian as learned) languages. In: 26th International Scientific Conference of Jonas Jablonskis Linguistic Diversity in the Modern World: Language Power and Prestige, Lithuania, Vilnius, 3–4 October 2019, e-book of Abstracts, pp. 78–79. Institute of the Lithuanian Language, Vilnius (2019)
10. Ernestus, M., Dikmans, M., Giezenaar, G.: Advanced second language learners experience difficulties processing reduced word pronunciation variants. Dutch J. Appl. Linguistics **6**(1), 1–20 (2017)
11. Nouveau, D.: Limites perceptives de l'e caduc chez des apprenants néerlandophones. Revue Canadienne de Linguistique Appliquée **15**, 60–78 (2012)
12. Wong, S.W., Mok, P.P., Chung, K.K.H., Leung, V.W., Bishop, D.V., Chow, B.W.Y.: Perception of native English reduced forms in Chinese learners: its role in listening comprehension and Its phonological correlates. TESOL Q. **51**(1), 7–31 (2017)
13. Brand, S., Ernestus, M.: Listeners' processing of a given reduced word pronunciation variant directly reflects their exposure to this variant: evidence from native listeners and learners of French. Q. J. Exp. Psychol. **71**, 1240–1259 (2018)
14. Raeva, O., Riekhakaynen, E.: Frequent word forms in spontaneous Russian: realization and recognition. Linguistica Lettica **24**, 122–139 (2016)

15. Klivadenko, E.: Phonetic realizations of politeness formulas in spontaneous Russian [in Russian], BA Thesis. St. Petersburg State University, St. Petersburg (2017)
16. Stojka, D.: Reduced word forms in Russian speech: Linguistic and extralinguistic aspects [in Russian], PhD Thesis. St. Petersburg State University, St. Petersburg (2017)
17. Riekhakaynen, E.: Corpora of Russian spontaneous speech as a tool for modelling natural speech production and recognition. In: Proceedings of the 10th IEEE Annual Computing and Communication Workshop and Conference, Las Vegas, United States, 6–8th January 2020, pp. 406–411. University of Nevada, Las Vegas (2020)

Ensuring Continued Use of Information Systems in an Organization

Mohd Zuhan Mohd Zain[1](✉) and Ab Razak Che Hussin[2]

[1] School of Computing, Universiti Teknologi Malaysia (UTM), Johor Bahru, Johor, Malaysia
mohdzuhan@gmail.com
[2] Azman Hashim International Business School (AHIBS), Universiti Teknologi Malaysia (UTM), Johor Bahru, Johor, Malaysia
abrazak@utm.my

Abstract. The information systems (IS) play a significant role in an organization's achievement, especially in digital era. IS can accelerate organization thru information handling and give substantial affects of IS success and later to its continuous use. The growth in today's business in terms of technology and commercial enterprise models requires comprehensive elements to support usage and continuance use of IS. Therefore, this research has a look to develop an IS continuance version through the continuation of expectation confirmation mode (ECM) by means of integrating new factors from other related theories. The proposed model is evaluated using questionnaire among contributors of the user of endowment firms that use e-endowment system. The feedback from the reliable respondents had been analysed using the partial least squares (PLS) primarily based on structural equation modelling (SEM) technique. The results show that support, usage, and technology suitability complement behavioural intention elements to IS continuance intention. The results of this study contributed to the new understanding in IS continuance environment and gives a possibility for developing an influential plan of IS continuance inside the organizations.

Keywords: Information system · Continuance model · Organization model

1 Introduction

Information systems (IS) are defined as strategies used of information to improve decision making, communication, and managerial purpose in corporations [18]. They provide valuable statistics associated with the overall performance of an organization and might indicate levels of client satisfaction. Despite the growth of IS, a scarcity of long term use is an obstacle in assessing the advantages of IS implementation [1]. Many previous researches investigated firm's preliminary adoption of ISs [2], but focused less on evaluating their long term [3]. The key in IS success is related more to the continuance and

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_48

effective control of the IS [4]. IS continuance is crucial for organizations because long term success is dependent on endured usage in place of first time usage [1, 5]. Therefore, this research investigates the elements that have an effect on the continuance of ISs by the use of ECM as the basis for the comprehensive model.

2 Motivation of Study

Many ISs are at risk of failure. The failure of an undertaking may be caused by many elements such as a trade-in requirements, change in process or requirement. Furthermore, ISs initiatives are situation to different problems that make certain even after venture completion; there may be still no guarantee of success. ISs have to break thru many barriers related to their continuance or groups may most effective use the machine for a short quantity of time.

Most IS continuance (ISC) studies are primarily based at the ECM that has been produced in 2001 [3, 6–8]. ECM states that confirmation, perceived usefulness, and pride are the fundamental factors or constructs contained in literature. The development of latest era, consumer centred and open access systems have caused new constructs gaining importance in relation to ISC. Considering the modern day literature available, the constructs seem like scattered and contributes in exceptional way. There is not any empirical proof that constructs have been incorporated into one single model to reveal the impact towards ISC. Therefore, it is critical that constructs like prior experience, self-efficacy, utilization, task-technology fit, and perceived support are investigated to be incorporated to the ECM in measuring the effects on ISC.

Integrating the relevant constructs into one comprehensive model offers a much wider view of ISC. A complete ISC version is useful for organizations, researchers, IT developers, and online organizations to assist in improving the planning, strategy, and implementation of ISs. This can result in enhancements in efficiency in areas like manufacturing and operations. Thus, this research taking this opportunity to investigate and proposed an ISC model that could explain these kinds of constructs as one inclusive ISC model.

3 The Concept of ISs Continuance and Its Research

A variety of organizations use ISs to directly connect with customers, suppliers, and clients. Examples include e-service, e-government, e-revenue, e-tendering, e-procurement, e-donation, and e-booking. These are often used on a global scale by businesses to improve work quality, production, and efficiency. This study used data collected on an e-service system called e-endowment which is a type of IS that acts as a collection agent for donations made online. The fund is used for poverty alleviation, endowments, and for the development of facilities like schools and housing. It is an IS that was built to serve this specific purpose and is comprised of several components including a payment system, a database system, and a security system. E-endowment is designed to perform endowment for distributing funds to provide welfare [2]. It was designed in the hope that it would deliver an improved performance compared to a traditional endowment system. Through e-endowment, contributing funds is made easier

as it is convenient to access from any location at any time. Nonetheless, it appears that crowd funding, which functions in a similar way, is more favourable for both individuals and businesses [5].

ISs continuance (ISC) refers to a sort of post-adoption behaviour, and it is similar to the time period "post-adoption phase". There's a lack of a theoretical models referring to ISC the identification of factors that influenced the continuous use of ISs remains crucial with regards to long time period success. The discussion from the literature suggests that ISC should not be based totally on confirmation, perceived usefulness, and user pleasure or satisfaction as in ECM. Constructs in ECM have to be supported with other different constructs including the individual ability, experience, generation utilization, and suitability. This is because of the evolution of recent technology connecting people and new commercial enterprise models concentrate on people and an open access system. Therefore, this study explores on integrating some constructs from different models, consisting of task-technology fit (TTF), social cognitive theory (SCT), social support theory (SST), and unified theory of acceptance and use of technology (UTAUT) with ECM to expand and develop an IS continuance model.

ECM predominantly predicts and aware the continuance aim on the consumer behaviour research, along with post-buy behaviours and satisfaction [1]. It pursuits to investigate the post-adoption behaviour by way of utilizing delight. TTF has been widely employed to analyse the statistics generation can lead to performance, and look at the concord when coupling the assignment and current era characteristics. The technology ought to match the venture, where it is able to lead an overall performance effect for IS to positively impact the use of generation. The suit of technology to the task is the volume to the era functions consistent with the assignment requirements [9].

SCT particularly focused on idea of self-efficacy (SE) that is practically one of the useful standards enhanced from current psychology. SCT presume an aggregate of internal self-affect constructs and outside social systems able to encourage and alter individuals' behaviour [10]. SE is the main element relating to individuals' perception of abilities to arrange, control, and explore a sequence of movements to obtain favoured performance. UTAUT offers four essential constructs, such as overall performance expectancy, effort expectancy, social influence, and facilitating conditions to immediately and not directly investigate and predict aim to use and the used behaviour. Four moderators, which include gender, age, experience, and voluntariness of use, are used to examine the impact on aim to apply and use behaviour [4, 11]. The theory drastically inherited a few exceptional from the theory of planned behaviour, particularly, the model's structure.

SST is an idea with multidimensional elements together with emotional, instrumental, appraisal, and informational. Emotional assist has some reference to observe and expressing concern or belief. The social aid acts as an essential mechanism in keeping the excellent of life. Without this, users may also easily surrender and abandon the usage of the technology because it poses a few degree of problem for them. Users might appreciate help once they are confronted with barriers, for this reason, bringing a positive final result on persisted usage of a technology.

4 Proposed Research Model

The ECM suggested that information from the literature that persisted intention closer to IS should not only be based on confirmation, perceived usefulness, and person satisfaction. In fact, constructs associated with technology, support, and people background history should not be ignored in understanding the sector of IS continuance intention. Therefore, to become aware of the other elements or constructs worth thinking about for IS retain intention, 21 previous studies on IS continuance specializing in different case study domains have been analysed. After that, the conceptual model is proposed as stated in Fig. 1.

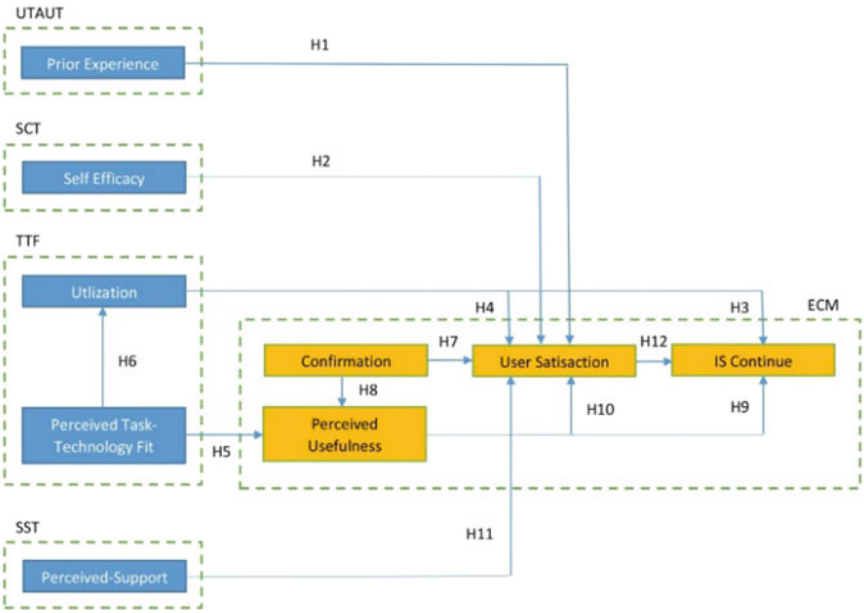


Fig. 1. Proposed IS continuance model

5 Model Evaluation and Data Analysis

The records analysis process used Smart PLS v3 to assess the conceptual model. Measurement model evaluation is conducted to take a look at the general validity and reliability of the objects used within the survey. The structural version is assessed to check the correlation between the constructs. The effects are used to verify the very last version and all suggested hypotheses.

5.1 Respondent Profile

Among the 115 valid respondents who answered the main survey, 64 respondents are male (56%) and 51 respondents are female (44%). On average, 36% of the respondents

are between 25 and 30 years old and 39 respondents are between 30 and 35 years old (34%). Six respondents are more than 35 years old and 29 respondents under the age of 25. Education stage is high with the majority of respondents having graduated from university (54%), 34 are college graduates (30%), and the last respondents have completed excessive college education. With regard to working revel in the organization, 34 respondents (30%) have less than 5 years' enjoy and 7 respondents have extra than 15 years' working enjoy within the organization. The majority of the respondents have working revel in of among five and 14 years. 72 (62%) of respondents are currently holding the position for more than five years, which means that a few of the respondents have good information of their job. Sixty-seven (58%) have more than 3 years' experience in the use of e-endowment systems. Only 48 respondents (42%) have less than 1 year experience on the use of e-endowment systems.

5.2 Data Analysis

The statistics analysis system contain of two most important series of techniques in instruction to test the model. Firstly, the measurement model is used to test every construct and its items. This evaluation employs the PLS set of rules to look at the construct's reliability and validity. It is used to ascertain to what dimensions the indicators measure the construct properly. Secondly, the structural version is classified once the construct's measures are confirmed. Structural model evaluation is applied to take a look at the predictive functionality of the model and the connection between the construct. To acquire this, several tests need to be applied and are discussed within the following subsections.

5.3 Internal Consistency Reliability

The dimension model evaluation specifies how the constructs are measured. There are two different sorts of dimension models: reflective and formative measurement. Each has a different sort of test. In this research, all of the constructs are reflective. In order to check the reflective size models, a set of tests are used to evaluate the internal consistency and examine the convergent validity and discriminant validity. This is end by the use of Cronbach's alpha, composite reliability, factor loading, average variance execrated (AVE), cross-aspect loading and the Fornell–Larcker criterion as the statistical tests.

The first specification to be analysed in the survey records is internal consistency reliability. One of the broadly used exams for internal consistency of the signs is Cronbach's alpha. Cronbach's alpha contributed an approximate of an indicators' reliability derive from its inter-correlations. The Cronbach's alpha has to be extra than or equal to 0.6. The second take a look at used for appraising internal consistency is composite reliability, which to achieve value over than 0.7. The composite reliability cost between 0.6 and 0.7 is acceptable, whilst values between 0.7 and 0.9 are considered to be satisfactory. The results of these two tests can provide proof of the internal consistency reliability. In this research, the outcomes of both Cronbach's alpha and composite reliability suggest that the minimum value for Cronbach's alpha is 0.7. The final constructs achieved greater than that. On the opposite hand, the composite reliability fee is more. These results show that each one of the constructs has acceptable internal consistency reliability, which suggests all the constructs' indicators are dependable.

5.4 Convergent Validity

In convergent validity, the volume which a trademark is compatible with positivity to opportunity signs of the similar construct. In assessing the convergent validity of a construct and its indicators, the outer loading of the indicators and the AVE needs to be considered. The generally accepted cost for the outer loading has to be 0.7 or higher, and the AVE price must be greater than 0.5. However, indicators with values between 0.4 and 0.7 have to be removed from the scale when this deletion leads to an increment in the value of AVE of the precise construct.

5.5 Discriminant Validity

The very last criterion for assessing the measurement version is discriminant validity. This refers to the volume to a particular construct is wonderful from other constructs by way of experiential levels. The established order of discriminant validity implies that the construct is distinctive and occupy the phenomenon that is not always found by means of different constructs. Two measures are being relevant to assess the discriminant validity. Firstly, the cross-loading of the indicators. Each indicator's outer loading of the specific construct should be extra than any of its cross-loading. The second approach is the use of the Fornell–Larcker criterion, which compares the square root of the AVE values with the feasible variable correlation. The outcomes of the Fornell–Larcker check included the AVE rectangular produced every construct are greater than the off-diagonal pass all any other construct. It is contributed that the measurement model of this examine has distinct discriminant validity. Fornell–Larcker criterion compared the rectangular root of AVE and issue correlation coefficients in a diagonal manner. The cross-loading matrix reveals that the measurement version exceeds the threshold.

5.6 Collinearity Assessment

Once the measurement version is installed and the assemble measures are confirmed to be dependable and authentic, the second one phase within the manner can progress with the assessment of the model. This involves inspecting the predictive abilities of the version, in addition to the relationship between the constructs. To gain this, this takes a look at follows the systematic technique by previous researches. The first step in this method is to study the collinearity of the version, which is completed by way of engaging in tolerance and variance inflation thing (VIF) tests. The second step is to evaluate the important and relevance of the model, which is done by using the analysing of p-value and t-value. The first step involves the collinearity assessment being carried out the usage of IBM SPSS v22. In this assessment, every set of predictor constructs are tested one by one for each dependent construct. All the values of consequences tolerance and VIF exceed the edge values that are extra than 0.20 for tolerance and much less than 0.5 for VIF. As a result, no construct had to be eliminated from the model.

5.7 Path Coefficient

The path coefficient is tested to represents the hypothesized relationships in the midst of the model's constructs. The path coefficient has standard values between -1 and $+1$.

The path coefficient values estimated close to +1 constituted strong positive relationship that is statistically significant. Besides, the closer approximated values approach to 0 is considered weak relationships. On the case of *t*-value is larger than the critical value, the path coefficient is significant. In this study, one-tailed tests are used due to the hypothesized relationships being directional or expected to be positive. The critical values for one-tailed tests are 1.28 (significant level = 10%), 1.65 (significant level = 5%), and 2.33 significant level = 1%). Table 1 provides an outline of the statistical outcomes of t-value, p-value, and significant levels associated with each hypothesis. The outcomes from the bootstrapping test indicate that 11 out of the 12 hypotheses are significant. One construct, namely prior experience, has an insignificant impact on satisfaction.

Table 1. Summary of the statistical results

Path	Path coeff (ß).	*T*-values	*P*-values	Significant level
PE → SAT	0.058	1.037	0.150	NS
SE → SAT	0.129	1.764	0.039	*
UT → ISCI	0.110	1.725	0.042	*
UT → SAT	0.155	1.642	0.050	*
PTT → PU	0.461	4.723	0.000	***
PTT → UT	0.606	10.309	0.000	***
CO → SAT	0.284	3.070	0.001	***
CO → PU	0.309	3.617	0.000	***
PU → ISCI	0.404	5.693	0.000	***
PU → SAT	0.167	1.912	0.028	*
PS → SAT	0.221	2.195	0.014	*
SAT → ISCI	0.628	7.493	0.000	***

Note ***$p \leq 0.001$, **$p < 0.01$, *$p \leq 0.0.05$, and *NS* not significant

6 Conclusion

The studies study is guided by the aims formed at the start of the vicinity of continuous use of IS. The main studies query is "How to expand an answer for understanding the long term use of IS in an organization?" To discover a significant answer of the question, a well-deliberate SLR is performed to understand the prior studies that had been completed via seeking to overcome the issue concerning continuous use of IS in an organization. The final research model which consist of ECM, TTF, SCT, and SST theory able to provide significant understanding on the long term use of IS in an organization.

References

1. Bhattacherjee, A.: Understanding information systems continuance: an expectation-confirmation model. MIS Q. **25**(3), 351–370 (2001)
2. Sifa, E.N.: E-Waqf as an alternative solution for infrastructure development based on crowdfunding. Int. Adv. Board, **178** (2016)
3. Terzis, V., Moridis, C.N., Economides, A.A.: Continuance acceptance of computer based assessment through the integration of user's expectations and perceptions. Comput. Educ. Elsevier Ltd. **62**, 50–61 (2013)
4. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. MIS Q. **27**(3), 425–478 (2003)
5. Yusof, M.M., Papazafeiropoulou, A., Paul, R.J., Stergioulas, L.K.: Investigating evaluation frameworks for health information systems. Int. J. Med. Inform. **77**, 377–385 (2008). https://doi.org/10.1016/j.ijmedinf.2007.08.004
6. Chang, Y.P., Zhu, D.H.: The role of perceived social capital and flow experience in building users' continuance intention to social networking sites in China. Comput. Hum. Behav. **28**(3), 995–1001 (2012)
7. Chen, C.P., Lai, H.M., Ho, C.Y.: Why do teachers continue to use teaching blogs? the roles of perceived voluntariness and habit. Comput. Educ. **82**(1), 236–249 (2015). https://doi.org/10.1016/j.compedu.2014.11.017
8. Lu, H.P., Yang, Y.W.: Toward an understanding of the behavioral intention to use a social networking site: an extension of task-technology fit to social-technology fit. Comput. Hum. Behav. **34**, 323–332 (2014)
9. Mouakket, S.: Factors influencing continuance intention to use social network sites: the Facebook case. Comput. Hum. Behav. **53**, 102–110 (2015). https://doi.org/10.1016/j.chb.2015.06.045
10. Bandura, A.: Social cognitive theory: an agentic perspective. Annu. Rev. Psychol. **52**(1), 1–26 (1989)
11. Luo, Y., Ling, H.: Exploration and exploitation of information systems usage and individual performance. Procedia Comput. Sci. **22**, 863–872 (2013). https://doi.org/10.1016/j.procs.2013.09.169

A Cost-Effective 3D Acquisition and Visualization Framework for Cultural Heritage

Hosameldin Osman Ahmed[1], Abdelhak Belhi[2,3], Taha Alfaqheri[1(✉)], Abdelaziz Bouras[2], Abdul H. Sadka[1], and Sebti Foufou[4]

[1] Brunel University, London, UK
{hosameldin.ahmed2,taha.alfaqheri2,abdul.sadka}@brunel.ac.uk
[2] CSE, Qatar University, Doha, Qatar
abdelhak.belhi@qu.edu.qa, abdelaziz.bouras@qsu.edu.qa
[3] DISP Laboratory, University Lumière Lyon 2, Lyon, France
[4] Le2i Lab, University of Burgundy, Dijon, France
sfoufou@u-bourgogne.fr

Abstract. Museums and cultural institutions, in general, are in a constant challenge of adding more value to their collections. The attractiveness of assets is practically tightly related to their value obeying the offer and demand law. New digital visualization technologies are found to give more excitements, especially to the younger generation as it is proven by multiple studies. Nowadays, museums around the world are currently trying to promote their collections through new multimedia and digital technologies such as 3D modeling, virtual reality (VR), augmented reality (AR), and serious games. However, the difficulty and the resources required to implement such technologies present a real challenge. Through this paper, we propose a 3D acquisition and visualization framework aiming mostly at increasing the value of cultural collections. This framework preserves cost-effectiveness and time constraints while still introducing new ways of visualization and interaction with high-quality 3D models of cultural objects.

Keywords: Cultural heritage · Deep learning · Artificial intelligence · 3D modelling · CEPROQHA project · 3D interaction · Motion controller

1 Introduction

Art and culture are the essence of humanity as they pack a lot of historical information that cannot be found elsewhere. Cultural artifacts are distinguished by their variety, shape, type, and value. Hence, the preservation of cultural heritage is an important process to curate and maintain these assets along with their provenance information for current and future generations. However, the physical preservation of these assets is a tedious and delicate process that requires a lot of time, resources and needs to be

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_49

undertaken by highly skilled professional curators. As a cost-effective and reliable way for art and culture preservation, digital technologies offer additional ways to preserve and further give more value and excitement around cultural heritage [1–3]. A lot of effort was undertaken to provide information technology (IT) solutions in the cultural domain. Some of the existing applications are geared toward digital preservation and the documentation of collections to ease the management and retrieval of assets. Other applications focus mainly on the end-user experience with innovative ways to increase the value and attractiveness of assets using latest data acquisition and visualization technologies such as 3D, VR, AR, and other immersive technologies to enable new ways of content consumption in the domain of cultural heritage.

In this paper, we mostly focus on technologies related to 3D imaging acquisition, visualization, and content consumption focusing on a specific use case related to interaction with museum objects. A high-value cultural asset is often put in a glass box, under a certain type of lighting conditions and with limited interaction. This traditional exhibition scheme seems not to attract younger people and is often reported to be "boring." Several 3D visualizations and interaction frameworks exist in this context, but their generalization is costly. The 3D modeling of assets takes a considerable amount of effort and requires the use of expensive hardware. Through this paper, we aim at tackling this challenge of 3D content acquisition, adaptation, and visualization using consumer-level hardware offering a more cost-effective yet attractive framework to allow 3D visualization and interaction with high-value museum assets. In this framework, first, a Digital Single Lens Reflex (DSLR) camera, lighting setup, and a basic turntable are used for the 3D content acquisition. Then, our framework uses photogrammetry as the main technology to model 3D cultural objects. Some optimizations are introduced to this process where artificial intelligence techniques such as super-resolution and motion interpolation are used to generate high-resolution input images for the photogrammetry process. Finally, for the interaction with the created 3D cultural objects, a leap motion controller is used to capture end-user hand motion, and this motion is then translated into controls that are used to interact in real-time with the generated 3D model allowing 360° rotation, movements, and zoom actions with the asset [4].

The rest of this paper is organized as follows. In section two, we present the works related to the 3D acquisition, visualization, and interaction with cultural heritage assets. In section three, we present the methodology and the implementation steps of our framework focusing mostly on data acquisition, data preprocessing, photogrammetry, and motion interaction. Section four concludes the paper and outlines some perspectives of future work.

2 Related Work

The digitization of cultural heritage plays an important role in long-term preservation as it is more reliable and less difficult to implement and maintain assets in a digital form. As an added challenge, museums and heritage institutions want to promote their collections using new digital content consumption techniques especially to attract visitors and to give more value to their collections. In this regard, a lot of work has been undertaken for the design and implementation of 3D acquisition and visualization technologies for

cultural heritage serving a wide range of use cases. In this work, our primary focus is for application dedicated to exhibitions in addition to end-users and not for professional applications as the system requirements for later are usually strict about quality and do not necessarily focus on cost-effectiveness and interaction which are the main drivers behind our study.

Most of the 3D acquisition techniques fall into two main categories: laser scanning and photogrammetry. For cultural heritage, several works [5–9] have addressed this challenge. Of these techniques, laser scanning seems to attract more wealthy institutions that aim at achieving the best possible quality. Thus, the applications of laser scanning for cultural heritage are mostly for extremely valuable assets. The most iconic scanner used in this context is the CultLab3D developed in Germany [6] where costs per scan are around 1000 USD. Other solutions such as shape from a stereo, shape from motion, shape from shading, and shape from silhouette are used as cost-effective alternatives [10]. Our focus in this paper goes to two of these methods: Shape from motion and shape from a silhouette. These two methods are reported to give average results in terms of quality when used in an uncontrolled environment while still achieving the best cost-effectiveness.

For end-user interaction, many approaches that implement the concept of virtual museums have been proposed. These solutions try to reconstruct an immersive museum visit experience by modeling museum architecture as well as assets [11]. Other approaches use virtual reality headsets and motions controller to provide a more immersive experience for end-users [12].

3 Methodology

In this section, we present the design and implementation details of our cost-effective cultural 3D acquisition, pre-processing, visualization, and interaction framework. As shown in Fig. 1, the proposed framework consists of five main stages, these include data acquisition, preprocessing, photogrammetry, 3D model adaptation, and motion-based 3D visualization. The following subsections discuss these stages in more detail.

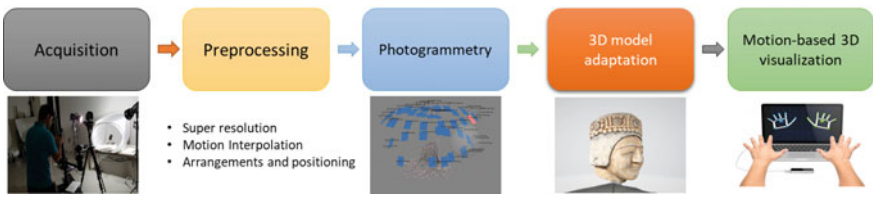


Fig. 1. From data acquisition to display and interaction

3.1 Data Acquisition

The data acquisition stage is crucial for our framework as high-quality images are required for the process of photogrammetry in order to generate good 3D models of the assets. The registration accuracy of the object geometry depends on the object texture and size.

The capturing equipment includes lighting photographer stands, lighting box, measuring distance device, and semi-professional 2D camera. Figure 2 shows the data acquisition setup for data registration. To get good quality capturing results, all the capturing equipments are installed and configured in the best possible way.

Fig. 2. Capturing setup in Museum of Islamic arts/media digital laboratory in Qatar

In the camera setup, the camera is positioned into three different angles. The angle degree of the camera position is set to (45°) as depicted in Fig. 3. More details on the pre-processing step are presented in Sect. (3.2). The captured data set includes 13 assets with different surface materials and sizes using the same capturing conditions.

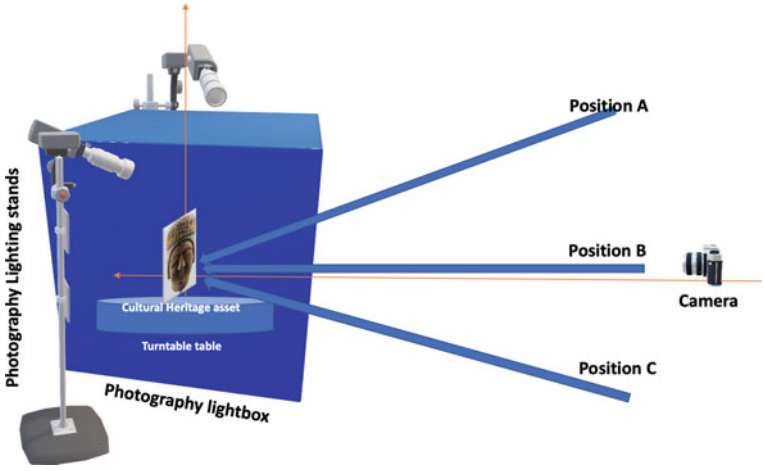


Fig. 3. Camera position setup

3.2 Data Preprocessing

After performing the data acquisition using the setup presented in the previous section, a preprocessing step is required in order to filter, enhance the quality, and organize the captured visual content to be used for the 3D model generation. Photogrammetry has the main advantage of not requiring expensive 3D scanning hardware as it only works on single 2D images captured using a 360° shooting setup. However, if one or some of the captured pictures do not meet the strict quality, positioning, and environmental parameters required, the whole 3D modeling process will most likely fail.

As a result, to preserve cost-effectiveness while still having good quality results, we use super-resolution and motion interpolation as software-based solutions to enhance the quality of the acquired image and generate more frames for the photogrammetry process as shown in Fig. 4.

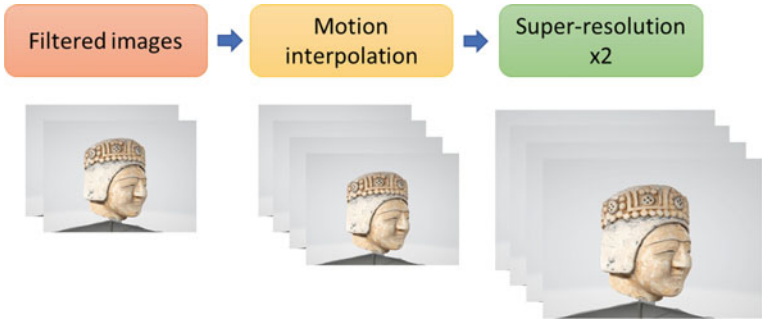


Fig. 4. Data preprocessing stages

3.3 Photogrammetry and 3D Model Adaptation

Photogrammetry is a technique for generating three-dimensional shapes through the analysis, measurements, and interpretations from a group of images acquired using a set of strict guidelines.

For our framework, we compared two of the most used tools for photogrammetry in the context of cultural heritage: Autodesk Recap Photo (Fig. 5) and AgiSoft PhotoScan (Fig. 6). Both performed almost similarly but using our setup, Autodesk Recap Photo yielded consistent results.

Once preprocessed, the images are put into the Recap Photo software; the 3D model is then generated using the Autodesk Cloud service. After that, the 3D models are tweaked using the provided tools and exported to either OBJ or FBX formats.

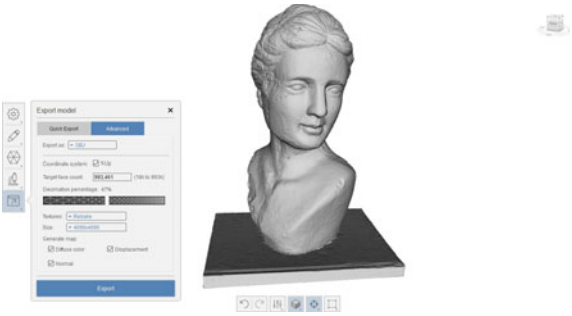


Fig. 5. Photogrammetry using Autodesk Recap photo

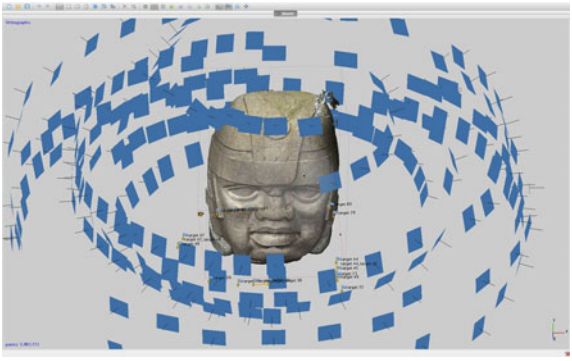


Fig. 6. Photogrammetry using AgiSoft PhotoScan

3.4 3D Visualization and Interaction

The 3D models created using our proposed framework can be utilized as virtually reinstated forms of the original Cultural Heritage (CH) assets to be presented in virtual museums and for further advancement of historical research and education. For the presentation of these 3D models, we utilized Unity 5.6 software in our framework for visualization and interaction of the produced models. Recently, there has been an increased interest in virtual reality (VR) techniques for various applications due to the developments in the field of VR. An important example of these applications is the virtual museum installation. Hence, to facilitate human interaction to the virtual reality of the acquired 3D models, we have used the LEAP motion controller (see Fig. 7) that can provide an informative representation of hands and fingers motion. The device has two infrared cameras that work to track the user's hands. The device was designed to allow integration with a wide range of platforms without too much hassle for developers as it is shipped with a high-level API that simplifies application development. Using this setup, one can interact with the acquired 3D model of the CH assets.

It is now well established from a variety of studies, that recent human--computer interaction (HCI) techniques are based on computational vision owing to its capability to track and recognize human gestures typically. Commonly, these techniques employ

Fig. 7. Leap motion controller (https://developer.leapmotion.com/)

images taken from a camera or a stereo pair of cameras to determine the hand pattern at each time instant [13, 14].

In our proposed framework, we have developed a visualization and interaction system that can help researchers and museum visitors to interact with the created 3D objects more simply and practically. Our system consists of a laptop, leap motion controller, unity 3D software, and C# programming language. The 3D object visualization comprises preprocessing object data by assigning different attributes through inspector and C# to control the location, rotation, scale of the 3D object in Euclidean X/Y/Z space, adding collider, i.e., boundary, to the simulated 3D, control how the model is displayed via mesh render, creating and applying material to the model that contains the texture information. The leap motion controller is used to track the motion of bare hands to allow natural interactions with a 3D object in a controlled area. The leap motion controller includes three infrared emitters and two infrared cameras that can be used to track the image of the hand. Then, based on this tracking, the controller extracts information and transmits it to the laptop to be used by the unity 3D software for the interaction with the targeted object. In our system, the extracted information is used to present virtual hands that can be used to interact with the object including, turning, moving, grasping, zooming, pushing forward and backward.

To test our system, ten users used the leap motion controller to interact with a 3D head model using virtual hands presented in three-dimensional virtual space as can be seen in Fig. 8. The test results indicate that the users found the practice positive and that it was easy for them to interact with the targeted 3D object. Figure 8 also displays screenshots from various interaction types with a 3D head model.

Fig. 8. A user using the proposed visualization and interaction system

4 Conclusion

In this paper, we presented the design and implementation details of a framework dedicated to the acquisition, processing, 3D model generation, 3D model visualization, and human interaction of digitized 3D models of cultural heritage. The framework was mainly designed to address cost-effectiveness concerns as well as to minimize the time required to digitize and generate 3D models of cultural heritage objects. Preliminary results show that the framework achieves good quality results due to the use of artificial intelligence techniques to perform super-resolution and motion interpolation on the acquired data before the generation of 3D models using photogrammetry. In the future work, we aim at improving its quality and usability and conduct further testing and result validation for more types of assets.

Acknowledgements. This publication was made possible by NPRP grant 9-181-1-036 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors (www.ceproqha.qa).

The authors would also like to thank Mr. Marc Pelletreau, the MIA Multimedia team, the Art Curators and the management staff of the Museum of Islamic art, Doha Qatar for their help and contribution in the data acquisition.

References

1. Belhi, A., Bouras, A., Alfaqheri, T., Aondoakaa, A.S., Sadka, A.H.: Investigating 3D holoscopic visual content upsampling using super-resolution for cultural heritage digitization. Sig. Process. Image Commun. **75**, 188–198 (2019)
2. Belhi, A., Gasmi, H., Bouras, A., Alfaqheri, T., Aondoakaa, A.S., Sadka, A.H., Foufou, S.: Machine learning and digital heritage: the CEPROQHA project perspective. In: Fourth International Congress on Information and Communication Technology, pp. 363–374. Springer (2018)
3. Belhi, A., Bouras, A., Foufou, S.: Towards a hierarchical multitask classification framework for cultural heritage. In: 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA), pp. 1–7. IEEE (2018)

4. Belhi, A., Bouras, A., Foufou, S.: Digitization and preservation of cultural heritage: the CEPROQHA approach. In: 2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), pp. 1–7. IEEE (2017)
5. Scopigno, R., Callieri, M., Cignoni, P., Corsini, M., Dellepiane, M., Ponchio, F., Ranzuglia, G.: 3D models for cultural heritage: beyond plain visualization. Computer 48–55 (2011)
6. Singh, G.: CultLab3D: digitizing cultural heritage. IEEE Comput. Graphics Appl. **34**, 4–5 (2014)
7. Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V., Chamzas, C.: Methods for 3D digitization of cultural heritage. J. Cult. Herit **8**, 93–98 (2007)
8. Stanco, F., Battiato, S., Gallo, G.: Digital Imaging for Cultural Heritage Preservation: Analysis, Restoration, and Reconstruction of Ancient Artworks. CRC Press (2011)
9. Belhi, A., Foufou, S., Bouras, A., Sadka, A.H.: Digitization and preservation of cultural heritage products. In: IFIP International Conference on Product Lifecycle Management, pp. 241–253. Springer (2017)
10. Gomes, L., Bellon, O.R.P., Silva, L.: 3D reconstruction methods for digital preservation of cultural heritage: A survey. Pattern Recogn. Lett. **50**, 3–14 (2014)
11. Bastanlar, Y., Grammalidis, N., Zabulis, X., Yilmaz, E., Yardimci, Y., Triantafyllidis, G.: 3D reconstruction for a cultural heritage virtual tour system. Int. Arch. Photogramm., Remote Sens. Spat. Inf. Sci., Beijing **37**, 1023–1028 (2008)
12. Gonizzi Barsanti, S., Caruso, G., Micoli, L., Covarrubias Rodriguez, M., Guidi, G.: 3D visualization of cultural heritage artefacts with virtual reality devices. In: 25th International CIPA Symposium 2015, pp. 165–172. Copernicus Gesellschaft mbH (2015)
13. Manresa, C., Varona, J., Mas, R., Perales, F.J.: Hand tracking and gesture recognition for human-computer interaction. ELCVIA Electron. Lett. Comput. Vis. Image Anal. **5**, 96–104 (2005)
14. Weichert, F., Bachmann, D., Rudak, B., Fisseler, D.: Analysis of the accuracy and robustness of the leap motion controller. Sensors **13**, 6380–6393 (2013)

Practical Vulnerability Analysis of Mouse Data According to Offensive Security Based on Machine Learning

Kyungroul Lee[1], Hoon Ko[2], Hyoungju Kim[3], Sun-Young Lee[4], and Junho Choi[5(✉)]

[1] R&BD Center for Security and Safety Industries, Soonchunhyang University, 31538 Asan, South Korea
carpedm@sch.ac.kr
[2] IT Research Institute, Chosun University, 61452 Gwangju, South Korea
skoh21@chosun.ac.kr
[3] Department of Computer Engineering, Chosun University, 61452 Gwangju, South Korea
snowlisakim@gmail.com
[4] Department of Information Security Engineering, Soonchunhyang University, 31538 Asan, South Korea
sunlee@sch.ac.kr
[5] Division of Undeclared Majors, Chosun University, 61452 Gwangju, South Korea
xdman@chosun.ac.kr

Abstract. We demonstrate a security threat of mouse data by differentiating the real mouse data from the dummy mouse data by deriving features to have high accuracy based on data science. Features appearing between the mouse coordinates input by the user are analyzed, and the feature is defined as a feature for machine learning models to derive a method of improving the accuracy. As a result, we found a feature where the distance between coordinates is concentrated in a specific range. When the distance is used as a feature, we verified that the mouse data is stolen more accurately.

Keywords: Practical security · Offensive security · User authentication · Machine learning · Vulnerability analysis

1 Introduction

Because of the emergence of online society, technologies for user authentication online have been required. Representative online user authentication technology is a password-based authentication technology, the information, ID and password, necessary for authentication is input from the keyboard [1]. However, because of keyboard data attack techniques, keyboard data input from a user is exposed to attackers [2–4]. To counteract these security threats of keyboard data, image-based authentication technology has

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_50

been presented [5–7]. This technology uses a specific location chosen by the user as a password in the image displayed on the screen. Nevertheless, as with the keyboard data, a security threat has been found that does not ensure the security of image-based authentication, because of the mouse data exposure likewise keyboard data [8–11].

Among attack techniques, an attack technique based on machine learning that effectively classifies the actual mouse position input by the user among all collected the mouse positions has emerged [12]. This technique not only requires the attacker to have a high level of attack techniques, but also has a high accuracy rate. Although this technique has high accuracy, attackers want and require a technique to more completely steal mouse data. Therefore, in this study, we study the feasibility of an attack technique that improves the accuracy by constructing features to the existing datasets [12].

2 Related Works

We describe the mouse data logging and protection technologies. To do this interaction with user as well as a mouse device, the operating system places the mouse cursor on the screen and provides APIs related to the mouse, GetCursorPos() function, for supporting requests from application programs. To extract the mouse positions, the attacker collects the mouse coordinate by periodically calling this GetCursorPos() function at short intervals and traces then the mouse movement by connecting them for visualization.

In order to protect the mouse coordinate from being exposed by the attack technique, a defense technique has been proposed to protect the mouse data [10]. This technique prevents exposure of the real mouse position by generating random mouse positions in a short period of random times by the defender.

An attacker wants and requires a technology to more completely steal the mouse movement, and a security threat is related to authentication information exposure under the mouse data protection technique [12]. This study collects both the dummy mouse data and the user-input data, and we derive mouse data vulnerability based on data science. This technique verified that the mouse data was stolen with a high attack success rate by defining the distance between coordinates as features even if the attack technique that makes it impossible to attack mouse data exposure is used. Therefore, we propose an attack method to have high accuracy by analyzing the feature appearing between mouse coordinates input by the user.

3 Feature Extraction

The previous attack technique based on machine learning used the elapsed time and the collected coordinates as a feature. In order to analyze the characteristics of this feature, the distribution of coordinates according to elapsed time was derived, and Figs. 1 and 2 show the analyzed results.

As a result, it is possible to derive the distance between the mouse coordinates input by the user, which is a yellow triangle, and random coordinates generated by the defender, which is a blue circle. This is considered to be possible to classify the coordinates.

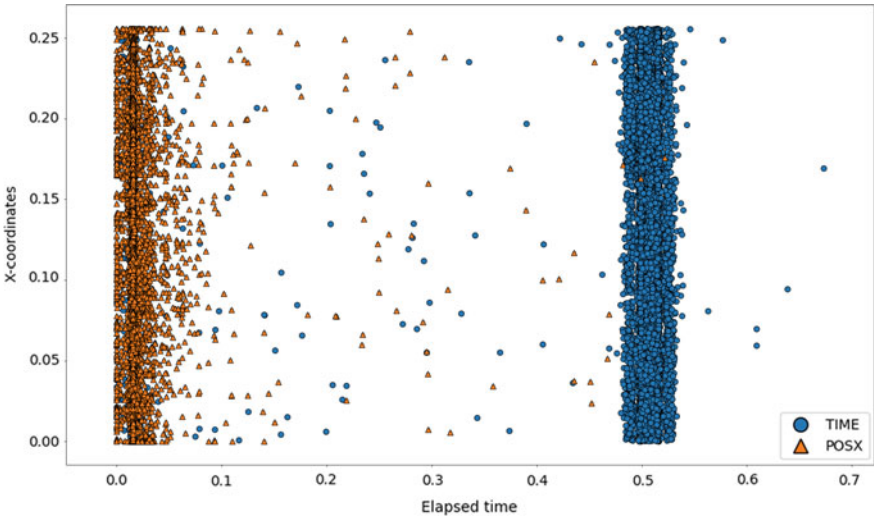


Fig. 1. *X* coordinates distribution according to the elapsed time (*X*-axis: elapsed time, *Y*-axis: *X* coordinate)

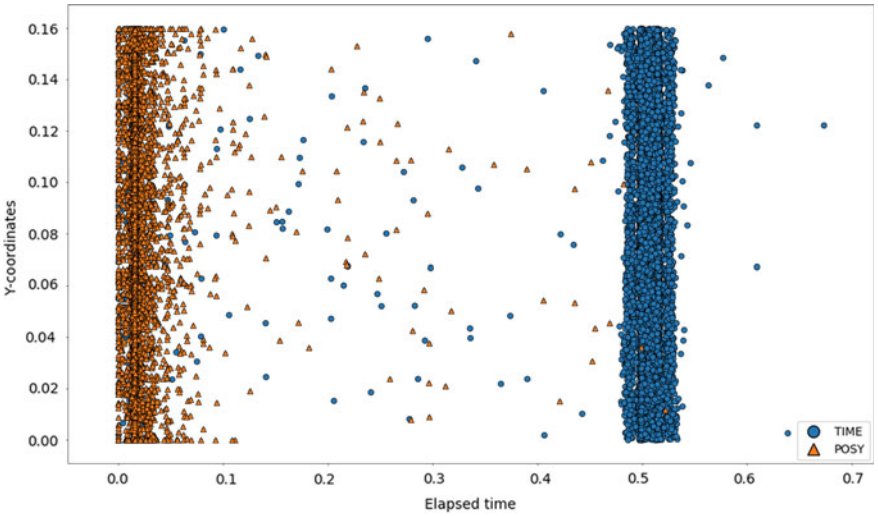


Fig. 2. *Y* coordinates distribution according to the elapsed time (*X*-axis: elapsed time, *Y*-axis: *Y* coordinate)

4 Experiment Results

We use various machine learning models [13–18] such as K-nearest neighbors (KNN), linear model, decision tree, kernel trick, and deep learning and describe the performance evaluation results of model validation, accuracy, precision, recall, F1-score, and AUC in this section. For the experiments, datasets were classified into any number to overcome

overfitting and underfitting. Table. 1 shows the training and testing results based on dataset used in [12].

Table 1. Training and testing results of Dataset 1-1 with distance

Model	Parameters	Training set score	Validation set score	Test set score
KNN	n_neighbors = 4	0.98	0.98	0.98
Logistic regression	$C = 10$, L2 regularization	0.75	0.76	0.75
Linear SVC	–	0.75	0.76	0.75
Decision tree	max_depth = 6	0.98	0.98	0.98
Random forest	n_estimators = 12	1.00	0.98	0.98
Gradient boosting	learning_rate = 0.01, max_depth = 9	0.99	0.98	0.98
SVM	$C = 100000$	0.98	0.98	0.98
MLP	max_iter = 1000, alpha = 0.01	0.97	0.98	0.98

As shown in the table, the training set has the highest score with the random forest of 1.0, and the validation set and the test set have a score of 0.98 with most of the models without the logistic regression. To derive the significant meaning of experiment results, Fig. 3 shows the performance evaluation according to Datasets 1–1 to 1–4 used in [12].

As shown in the figure, most of the datasets work well except for the logistic regression. The lowest performances are 0.75, 0.88, 0.96, and 0.97 for the logistic regression in Datasets 1-1, 1-2, 1-3, and 1-4, respectively. The highest performances are 0.981 for the random forest in Dataset 1-1, 0.995 for the decision tree, random forest, and gradient boosting regression tree in Dataset 1-2, 0.998 for the KNN, decision tree, random forest, gradient boosting regression tree, and SVM in Dataset 1-3, and 0.998 for KNN, decision tree, gradient boosting regression tree, and SVM in Dataset 1-4. When compared to the dataset that does not include the distance with the dataset does include the distance, the performances are improved in most of the models except for logistic regression and decision tree in Dataset 1-1, and performances are improved in most of the models except for logistic regression in Dataset 1-2. In Dataset 1-3, random forest and SVM improved performance, and only random forest improved performance in Dataset 1-4. Experiment results do not show improved performance in many models. However, in most recall and AUC results, there were performance degradation of only 0.0018 on average (17 items, a total decrease of 0.031), and all other items increased by an average of 0.073 (141 items, a total increase of 10.351); hence, the increase rate is 4.055% higher than the decrease rate.

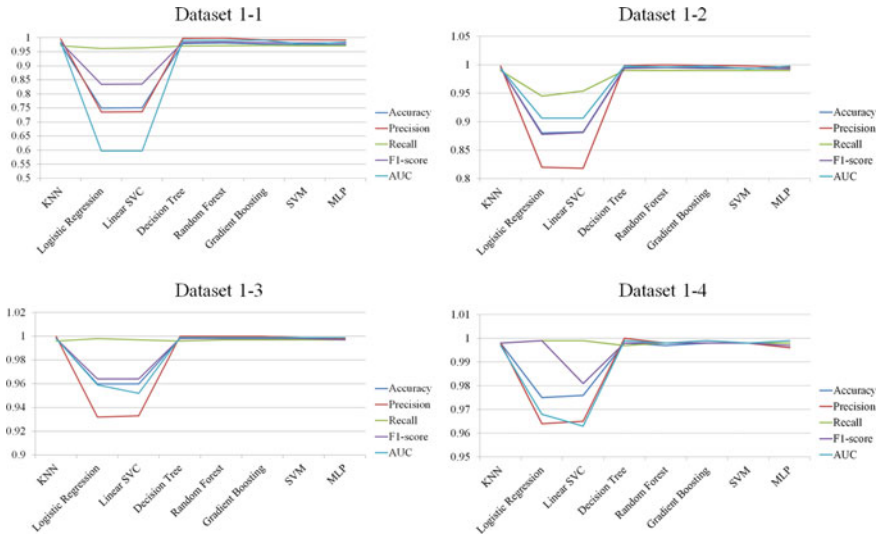


Fig. 3. Performance evaluation results

In this study, we evaluated the performance of datasets with the elapsed time and the X and Y coordinates, which is the existing attack technique, and the datasets including the distance between the previous and current coordinates. Consequently, we verified through experiments that the performance of datasets including distance is higher than that of datasets including elapsed time only. Therefore, the results of the performance evaluation are changed by defining features that can be included in the datasets, and we proved the contribution of this paper by comparing the performance evaluations. Figure. 4 shows the comparison results according to datasets 1-1 to 1-4 used in [12].

As shown in the figure, on the left is the result of the Dataset without distance, and on the right is the result of the Dataset with distance. The performance tends to be higher toward the right, and the performance of the datasets with distance is significantly higher than that of the datasets without distance. In conclusion, using the dataset with distance, we can effectively classify random mouse data with up to 99% accuracy. This means that an attacker steals the user-input mouse data in the real world.

5 Conclusion

This study demonstrated a security threat of the mouse data by deriving features to further improve the accuracy using machine learning based on mouse data in image-based authentication. In this study, we analyzed the distribution of coordinates based on the distance between coordinates according to elapsed time. We found that the distance between the mouse coordinates input by the user is concentrated in a specific range. Therefore, we configured datasets that defined the distance between the previous coordinates and the current coordinates as a feature. As experiment results, the datasets with

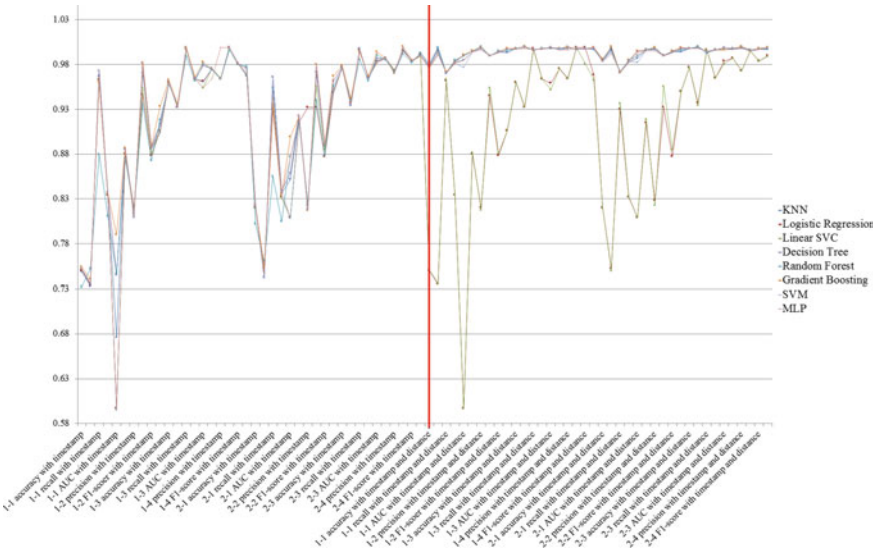


Fig. 4. Comparison of performance evaluation according to datasets 1-1 to 1-4 by features

distance classify the real user-input mouse data which is significantly higher than conventional attack techniques using the datasets without distance. The accuracy was over 99%, which means that the user-input data is almost completely classified.

Acknowledgements. This work was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2018R1A4A1025632) and by Korea Electric Power Corporation (Grant number: R18XA06-12).

References

1. Conklin, A., Dietrich, G., Walz, D.: Password-based authentication: a system perspective. In: 37th Annual Hawaii International Conference on System Sciences, pp. 1–10. IEEE, Big Island, HI, USA (2004)
2. Lee, K., Yim, K.: Password sniff by forcing the keyboard to replay scan codes. In: 5th Joint Workshop Information Security, pp. 9. Guangzhou, China (2010)
3. Lee, K., Yim, K.: Keyboard security: a technological review. In: 5th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, pp. 9–15. IEEEE, Seoul, South Korea (2011)
4. Oh, I., Lee, K., Lee, S., Do, K., Ahn, H., Yim, K.: Vulnerability analysis on the image-based authentication through the PS/2 interface. In: 12th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, pp. 212–219. Springer, Matsue, Japan (2018)
5. Takada, T., Koike, H.: Awase-E: Image-based authentication for mobile phones using user's favorite images. In: 5th International Conference on Mobile Human-Computer Interaction, pp. 347–351. Springer, Berlin, Heidelberg (2003)

6. Parekh, A., Pawar, A., Munot, P., Mantri, P.: Secure authentication using anti-screenshot virtual keyboard. Int. J. Comput. Sci. Issues **8**(5), 534–537 (2011)
7. Newman, R.E., Harsh, P., Jayaraman, P.: Security analysis of and proposal for image-based authentication. In: 39th Annual 2005 International Carnahan Conference on Security Technology, pp. 141–144, IEEE, Palmas, Spain (2005)
8. Lee, H., Lee, Y., Lee, K., Yim, K.: Security assessment on the mouse data using mouse loggers. In: 11th International Conference on Broad-Band Wireless Computing. Communication and Applications, pp. 387–393. Springer, Asan, South Korea (2016)
9. Oh, I., Lee, Y., Lee, H., Lee, K., Yim, K.: Security assessment of the image-based authentication using screen-capture tools. In: 11th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, pp. 156–161, Springer, Torino, Italy (2017)
10. Lee, K., Oh, I., Yim, K.: A Protection technique for screen image-based authentication protocols utilizing the SetCursorPos function. In: 18th World Conference on Information Security Applications, pp. 236–245, Springer, Jeju Island, South Korea (2017)
11. Lee, K., Yim, K., Lee, S.: Vulnerability analysis on the image-based authentication: through the WM_INPUT message. Concurrency Comput Pract Experience **2019**, e5596 (2019)
12. Lee, K., Esposito, C., Lee, S.: Vulnerability analysis challenges of the mouse data based on machine learning for image-based user authentication. IEEE Access **7**, 177241–177253 (2019)
13. Zhang, M., Zhou, Z.: ML-KNN: a lazy learning approach to multi-label learning. Pattern Recogn. **40**(7), 2038–2048 (2007)
14. Cheng, W., Hüllermeier, E.: Combining instance-based learning and logistic regression for multilabel classification. Mach. Learn. **76**(2–3), 211–225 (2009)
15. Sinclair, C., Pierce, L., Matzner, S.: An application of machine learning to network intrusion detection. In: 15th Annual Computer Security Applications Conference, pp. 371–377, IEEE, Phoenix, AZ, USA (1999)
16. Rovert, E., Lawrence, O., Kevin, W., Kegelmeyer, W.: A comparison of decision tree ensemble creation techniques. IEEE Trans. Pattern Anal. Mach. Intell. **29**(1), 173–180 (2006)
17. Jan, S.U., Lee, Y., Shin, J., Koo, I.: Sensor fault classification based on support vector machine and statistical time-domain features. IEEE Access **5**, 8682–8690 (2017)
18. Yin, C., Zhu, Y., Fei, J., He, X.: A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access **5**, 21954–21961 (2017)

Preliminary Study and Implementation of Chiang Mai Tourism Platform Based on DOSA

Vorada Panpeng[1], Miao Fang[2], Aniwat Phaphuangwittayakul[1], and Tanarat Rattanadamrongaksorn[1(✉)]

[1] International College of Digital Innovation, Chiang Mai University, Chiang Mai 50200, Thailand
tanarat.r@cmu.ac.th

[2] The Big Data Research Institute, Chengdu University, 610106 Chengdu, China

Abstract. To cope with data complication systematically, a tourism platform is proposed through a newly-defined concept named Data Oriented Security Architecture or DOSA. The framework of DOSA shifts an attention from "Application" in a conventional method to "Data". The main idea is to neutralize the numerously different sources of data into single entity called Data Register Center and strengthen the security of data by utilizing the public- and private-key encryption. Thus it could be simply adaptable to the varieties of applications and easily extensible to sustainable development. The advantages of DOSA include (1) maintainability of data ownership, (2) improvement in security and privacy, and (3) reduction in storage that, as a result, are suitable to modern application development. The application prototype is implemented in web application and piloted by a set of tourist destination in Chiang Mai. This work contains the implementation from Data Register Center in DOSA through the Tourism Service Platform. Although the sample does not include all features of the DOSA, the results are comprehensive and ready for the future extension and improvement.

Keywords: Big data · Variety · Smart city · Object-based · Data-oriented · Chiang Mai · Tourism · Digital economy · Data-oriented architecture · Database management

1 Introduction

Tourism is an important economic indicator to Thailand's economy. It is crucial not only to the Gross Domestic Product (GDP) as major revenue but also academically interesting because of new emerging trends from the changing behaviors [1]. Chiang Mai, on the other hand, is known as a provincial capital and has been ranked one of the top ten of the most tourist's favorite destination in

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_51

Thailand by Travel and Leisure magazine since 2016 [2–4]. The city has many cultural places and facilitated by all kinds of major transportation hubs convenient and attractive to a wide range of tourism, as well as the rural migrants and neighboring-country work forces [5].

As an information intensive business, a tourism industry depends heavily on varieties of data. The modern tourism applications, such as digital tourism, intelligent tourism, and virtual tourism alter the nature of production, distribution and consumption by optimizing operations and, as a consequence, achieve rapid growth in the tourism economy [6]. Smart tourism is an advanced stage of tourism informatization—a process that extends its capability and scope of information technology. Both tourists and organizations can easily be fed with information of activity, consumption, and resources instantly integrated by communications technology through a variety of end-user devices [7].

Tourism Service Platform (TSP) is an application involving in smart tourism offering information on catering, transportation, accommodation, traveling, shopping, etc. [8]. In other words, TSP is considered as One-Stop Service for tourists depending on their different traveling requirements. For example, TSP will provide information respecting the shortest route from origin to the destination, famous places nearby the destination, the information of each place along the way during traveling, special notice or advice from the government in that period and so on when tourists need to travel from one place to another place. Another sample is that tourists can access and collect the public supported information, surrounding accommodation details and reviews based on their preferences and updated current promotions when they require to take a rest or sleep. With the volume and variety, the data grow and are accumulated as Big Data that challenges of security and privacy are concerned [9]. In addition, the functions of current TSP, are expected to support different and multiple requirements, processes, models, places, functions and stakeholders. For example, it is no longer simple for the support of a primitive website on a traditional platform. To support these requirements, the modern TSP is needed to be extensible for sustainable development. However, to our best knowledge, no method using traditional software development engineering can align with this goal hence Data-Oriented Security Architecture (DOSA) [10,11] is proposed.

The Data-Oriented Security Architecture (DOSA) is data-concentrated demonstrating differently to the traditional *Transaction-Oriented* architecture which is implemented by the Data Base Management Systems (DBMSs). The DOSA focuses on data itself, for example, data security application, data ownership authorization, etc. [10] that become the basic requirements for any Information Technology (IT) platform. On the tourism industry, big data technology and Internet of Things (IoTs) are utilized in offering tourists on the right management and personalized services [12] and to suit tourists' preferences at the right time [13]. There are two main purposes of applying the DOSA in smart tourism i.e. (1) to overcome the data sharing problems according to the varieties of data sources, data structures, business departments, data owners and systems

and (2) to solve the security and privacy problem in open data and big data in order to protect the privacy and the guarantees the integrity of data received.

In this paper, we propose the methodology by applying DOSA to the TSP and demonstrate the example of the TSP for restaurants in Chiang Mai that is supported by DOSA as discussed in the following sections.

2 Methodology

Tourism Service Platform (TSP) is an application, service and analytic system which is composed tourist-related information, traveling information such as transportation and accommodation, and information from the government. It provides support and convenience to the tourists and make their lives more comfortable during the trip. As a result, TSP needs to be evolved to accommodate the rapid growth, security and privacy.

2.1 Data-Oriented Architecture (DOA) and Data-Oriented Security Architecture (DOSA)

DOA was first proposed by Fang Miao [14,15]. The concept bases on data-oriented or data-centric thinking. Three key facts distinguish the DOA from other approaches. First, DOA does not bind data to a specific location but can access them by address link. Data is defined by the application which produced it and not an address of a location where it is currently stored. Second, DOA is a holistic approach that comprises of suitable data structure, data access methods and message exchange protocol. Thus, DOA can be easily implemented and utilized without lead time. Third, even not implemented in place, the user can make a decision on which data they want to keep private and which data they want to be public. The shared data becomes"public goods" that are not owned by a specific entity but belongs to the community. Data used in DOA is structured as a key-value-pair completed by header that holds information on how to process the key-value-pair [16]. Data-Oriented Security Architecture (DOSA) is known as the trusted data and platform. The DOSA is built based on DOA that prioritizes on the data and security architecture rather than the application. DOSA is composed with two main components that are data ownership and CA/PKI (Certificate Authorization / Public Key Infrastructure) [17]. Public Key Infrastructure (PKI) is one of the most widely implementations for security solution in all application domains. The PKI system contains a trusted third party known as a Certificate Authority (CA) that identifies the owner of a public key [18]. The objective of using DOSA is to perform a trusted data and its application in the open data and big data era. It is not only address the security problem but also preserve data ownership by encryption. The concept of DOSA can be generally described as the one body with two wings with generalized key-value-pair structure which is similar to DOA structure. The first wing of DOSA is the Key of registered data and the second wing of DOSA is the Value of data which is encrypted by the owner's public key. This second wing of DOSA makes DOSA

differs from DOA. The general basic principle of DOSA is shown as in Fig. 1. The DOSA is comprised of the Data Register Center (DRC), Data Authority Center (DAC), Data Exception Center (DEC) and Data Application Units (DAUs) [19].

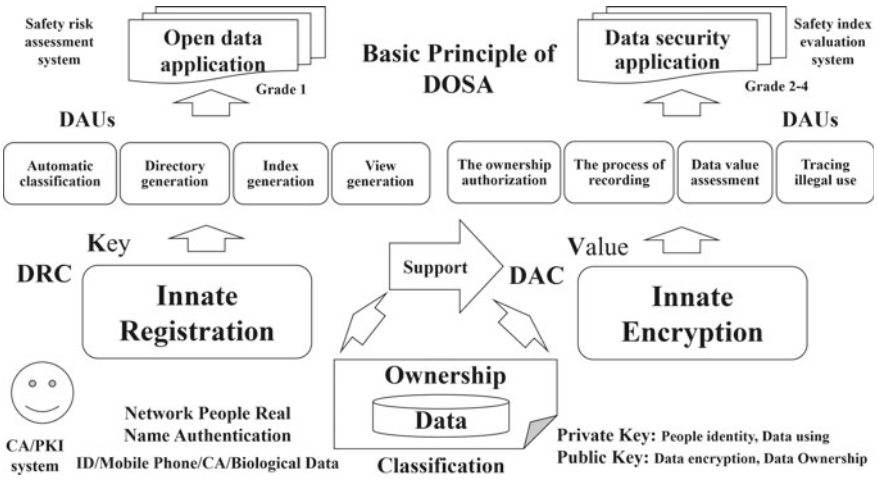


Fig. 1. Tourism applications supported by DOSA platform and data sources

2.2 DRC Data Registration

The DRC is a core component of DOSA. It stores all kinds of data attribute information registration and private data. DRC is used to construct the pool of logic data resources through the establishment of index and search engine, implementation, application service and management. The DRC can be set up as a lone server or associated with other DRCs so as to realize the data sharing [17].

Data registered in DRC comes from diverse sources therefore the techniques to register the data into DRC is identified. In this work, the automatic and semi-automatic web scraping is developed for gathering the information from a website [20]. In case of unstructured data, web scraping considered as automatic program is applied to retrieve metadata of image, audio and video files. Since DOSA only keep the metadata instead of original unstructured data, it can reduce the storage consumption of DRC. On the other hand, structured data are collected from selected websites by semi-automatic program and cleaning data process is involved if needed.

3 DOSA for Smart Tourism

In DOSA, the first step is to define the users that involve in smart tourism application. The structure of smart tourism are included three levels that are the

level of tourist, that of manager (government and tourism enterprises) and that from a technical perspective [8]. Hence, users are classified into three groups in this paper that are; (1) Tourist, (2) Enterprise and (3) Government. These groups require and provide different information that can be concluded in Tables 1, 2 and 3.

Table 1. Provided and required data for tourist

Provided data	Required data
Name	Hotel information
Gender	Restaurant and attraction
ID (Passport/Citizen id)	Transportation
Nationality	Recommended place
Departure country	Trusted guidance
Question and answer from Chatbot/Forum/Fanpage	Advisable sign
Comments/reviews in form of photo, video, note, rating	Call centers, Email and supported services
Booking history	Loyalty programs
Latitude-Longitude/Location (GPS) survey	Culture, nature and history introduction
Social media information	

Table 2. Provided and required data for enterprises

Provided data	Required data
Contact	Data from tourists
Address/Location	Data from government
Room profile	
Number of total employees	
Availability of rooms/services	
Tour packages/promotion	
Advertisement	
Travel enquiries	

According to the provided and required data from the tables, we can generate the groups of users as data sources. The DOSA platform supports the tourism application as shown in Fig. 2.

The relations between the groups of users define our framework into three tiers i.e. (1) Data sources, (2) DOSA Platform, and (3) Tourism Applications.

Table 3. Provided and required data for government

Provided data	Required data
Arrival and departure country of tourists	Data from tourists
Number of tourists in both domestic and international arrivals	Data from enterprises
Number of hotels and availability of hotel rooms	
Tourist attraction's information	
Support information such as nearest police station, embassy contact, public emergencies, transportation route and contact	
Advertisement information such as main or special events	

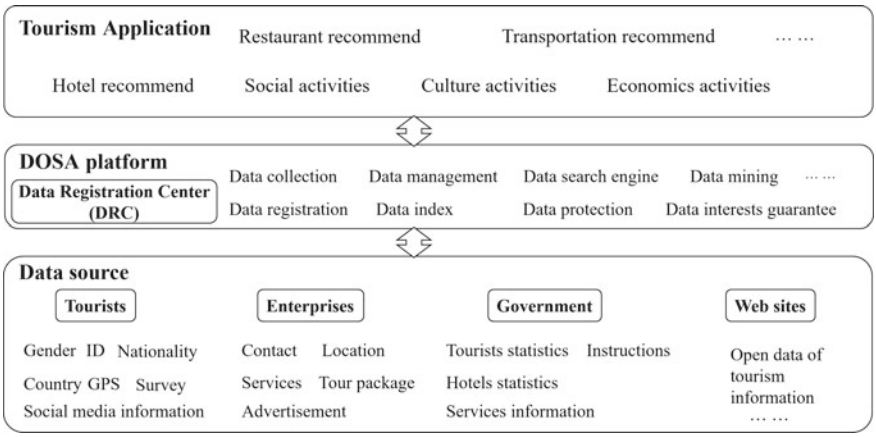


Fig. 2. Tourism applications supported by DOSA platform and data sources

The description of each entity and layer is common and self-contained. The tier of DOSA platform is inserted between other end-user interface and database in order to standardize data for both sides. Without the middle layer, it is too restricted or, even, impossible to utilize the data and extend the application.

The next step of DOSA is to determine type of provided and required data from each user, for example, text, photo, audio, video, location and document. The data are divided in Tables 4, 5 and 6.

Table 4. Data type of tourists

Data type	Data
Text	• Personal information that are Name, Gender, ID, Nationality, Departure Country
	• Question and Answer from Chatbot/Forum/Fanpage
	• Comments/Reviews
	• Booking history
	• Survey
	• Social media information
Photo	• Photos from social media platform such as Instagram, Facebook, Flickr, Twitter, Pinterest, etc
	• Photos in reviews or comments
Video	• Videos in social media platform such as youtube, Instagram, Twitter, Facebook, etc
	• Videos in reviews or comments
Location	• Check-in address
	• Tagged address
	• Latitude, Longitude
Document	• Word file
	• ppt file
	• pdf file
	• excel file
Audio	• Voices
	• Songs

Table 5. Data type of enterprises

Data type	Data
Text	• Basic information that are Contact, Room profile, Number of total employees, Availability of rooms/services
	• Tour packages/Promotion
	• Advertisement, Marketing information
	• Travel enquiries
Photo	• Promoted photos
Video	• Promoted videos
Location	• Address
	• Latitude, Longitude
Document	• Accounting data
	• Presentation documents

Table 6. Data type of government

Data type	Data
Text	• Arrival and Departure country of tourists
	• Number of tourists in both Domestic and International arrivals
	• Number of hotels and availability of hotel rooms
	• Tourist attraction's information
	• Support information such as nearest police station, embassy contact, public emergencies, transportation route and contact
	• Advertisement information such as main or special events
Photo	• Promoted photos about places and events
Video	• Promoted videos about places and events
Location	• Address
	• Latitude, Longitude
Document	• Private documents

Based on the provided, required data and classified data type, we can create various applications in smart tourism. There are many applications related to tourism industry can be proposed. The sample applications are, for example, Energy and Utility Management,Improve customer supports, Understand behavior of tourist to create tour promotion and etc.

The third step is to design DRC standard. The DRC standard is basically a database that stores data attributes including structured data and metadata of unstructured data are registered in the DRC.

4 Results and Discussion

In this implementation, we illustrate a tourism service platform for restaurants in Chiang Mai that is supported by the DOSA. The DRC standard will be designed as Fig. 3.

There are two types of DRC proposed in our TSP. The example of the application for food recommendation contains the "metadata_entities" and "restaurant_detail" DRCs. The former stores metadata of unstructured data that are video files, audio files, documents, images and so on which come from different sources including websites and internal storages. The metadata of unstructured data stored in the DRC are in the form of text and uniform resource identifier (URI) including the name and address that link to original sources. Storing metadata instead of original data provides two main advantages. Firstly, the TSP is not required to extend the storages which keep the data frequently as the sizes of metadata are less than the sizes of original data. Secondly, the TSP can prevent and solve the data privacy issue since the stored data are belonging

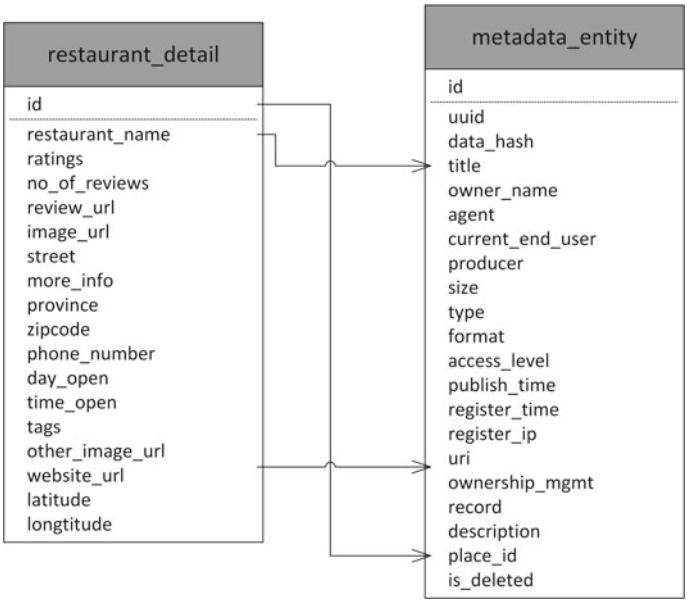


Fig. 3. DRC standard example of tourism service platform for restaurant in Chiang Mai

to the original sources. Value in uuid attribute is created by hashing algorithm with the values in other attributes hence the uuid value will be unique. The latter DRC called "restaurant_detail". This DRC stores data in form of structured data which is provided from public website. Some attributes of restaurant_detail DRC will be stored in metadata_entity DRC as well. In the diagram shows that the id, restaurant_name, website_url attributes in restaurant_detail DRC are stored as place_id, title, uri attributes in metadata_entity DRC.

Once the DRC standard is implemented in MySQL. The tourism service platform was built with PHP connecting MySQL database. The TSP for the restaurant in Chiang Mai has two main web pages. The DRC for Data Management is used by an administrator in managing the DRC's database. The metadata_registry DRC is displayed in Fig. 3. The rights of database management are set to add, edit and delete a record in the DRC. As a matter of fact, the concept of "Edit" in DRC is not included in DOSA. Instead, the DRC saves every change into a new record once we edit any existing record. The value of is_delete attribute value is set to 1 in case of deleting and the deleted record will be filtered out from the Data Management web page. The DRC, in addition, can be sorted by selecting any attribute name in ascending or descending order as well as query by typing the keyword in a search function on the screen.

On the front end, the user would not have the different user experience comparing to the other common types of TSP. However, the complexity of application is managed and hidden from the user in the back end of the DOSA. In

this pilot application, all of information shown in detail of restaurant page are extracted from the metadata of unstructured data and private data stored in the DRCs. The implementation is designed according to varieties in the source of data. Unlike other conventional method, once is the application finished, the modification or extension can be made easily. For example, there might be a need to revise a list of sources due to a popularity of a new platform. Simply, a new website and its related information could be registered properly and ready to go.

This work presents a newly-defined trusted data and platform named Data-Oriented Security Architecture (DOSA) as well as demonstrates the DOSA application through a tourism service platform (TSP) for restaurants in Chiang Mai. Since DOSA is based on data-centric thinking with security and privacy of data, designing Data Register Center (DRC) process is crucial. Once a DOSA is developed and deployed, multiple applications can be built upon the DOSA. Our work also shows that DOSA is applicable in different domains besides tourism depending on stored data in DRC. Additionally, designing and managing DRC standard is less complicated and more accessible compared to the traditional data management method.

5 Conclusion

Using Data Oriented Security Architecture (DOSA), the application can be managed in different types and sources as well as support various types of computer, web and mobile application because they can be unified by the DRC type of data sharing. Since DOSA is appropriate for the varieties of application requirements and extensible for sustainable development, applying DOSA method in a tourism service platform is introduced and demonstrated through a simple application prototype in this paper. Data Register Center or DRC which is the key component of DOSA was built to register all possible data by automatic and semi-automatic programming. The DRC is used to access all data by the application after the data registration. In our experiment, we represent the whole processes to create the application under DOSA covering designing DOSA, defining DRC, accessing data and creating a prototype.

Acknowledgment. Thanks to International College of Digital Innovation (ICDI) of Chiang Mai University (CMU), Thailand and Big Data Research Institute (BDRI) of Chengdu University (CDU), China for our long collaboration in the topics of Digital Innovation and Tourism.

References

1. Uansa-ard, S., Binprathan, A.: Creating the awareness of halal mice tourism business in Chiang Mai, Thailand. Int. J. Tourism Policy **8**(3), 203–213 (2018)
2. Terzian, P.: The world's top 15 cities (2018). https://www.travelandleisure.com/worlds-best/cities

3. Lieberman, M.: The world's best cities in 2016 (2017). https://www.travelandleisure.com/worlds-best/2016/cities
4. Kelso, S.: The world's top 15 cities in 2017 (2017). https://www.travelandleisure.com/worlds-best/cities-2017
5. Çakmak, E., Lie, R., McCabe, S.: Reframing informal tourism entrepreneurial practices: capital and field relations structuring the informal tourism economy of Chiang Mai. Ann. Tourism Res. **72**, 37–47 (2018)
6. Tian, G., Junping, D., Zengqi, S., Yingmin, J.: Digital tourism integrated service system realization. In: 2010 International Conference on Networking, Sensing and Control (ICNSC), pp. 232–237. IEEE (2010)
7. Ling-Yun, Z., Nao, L., Min, L.: On the basic concept of smarter tourism and its theoretical system. Tourism Tribune/Lvyou Xuekan **27**(5) (2012)
8. Li, Y., Hu, C., Huang, C., Duan, L.: The concept of smart tourism in the context of tourism information services. Tourism Manage. **58**, 293–300 (2017)
9. Zhang, D.: Big data security and privacy protection. In: 8th International Conference on Management and Computer Science (ICMCS 2018). Atlantis Press (2018)
10. Miao, F., Fan, W., Yang, W., Xie, Y.: The study of data-oriented and ownership-based security architecture in open internet environment. In: Proceedings of the 3rd International Conference on Cryptography, Security and Privacy, pp. 121–129. ACM (2019)
11. Fang, M.: Discussion of data-oriented software architecture. J. Front. Comput. Sci. Technol. **10**(10), 1351–1364 (2016)
12. Buhalis, D., Amaranggana, A.: Smart tourism destinations enhancing tourism experience through personalisation of services. In: Information and Communication Technologies in Tourism 2015, pp. 377–389. Springer, Berlin (2015)
13. Guo, Y., Liu, H., Chai, Y.: The embedding convergence of smart cities and tourism internet of things in China: an advance perspective. Adv. Hospital. Tourism Res. (AHTR) **2**(1), 54–69 (2014)
14. Miao, F., Yang, W., Ye, A., Yu, X., You, Y., Xiong, Y.: A data-oriented security architecture method and system. CN Patent application 201511025657.4, Public No. CN105450669A (2015)
15. Miao, F., Yang, W., Li, Y., Yu, B., Zhou, X., Song, C., Wang, G.: A data-oriented information technology system. CN Patent 201410341092.X, PCT/CN2015/080013 (2014)
16. Vorhemus, C., Schikuta, E.: A data-oriented architecture for loosely coupled real-time information systems. In: Proceedings of the 19th International Conference on Information Integration and Web-based Applications & Services, pp. 472–481. ACM (2017)
17. Miao, F., Yang, W., Ye, A., Chen, H.: The architecture for data, security and application in digital earth platform. In: IOP Conference Series: Earth and Environmental Science, vol. 46, p. 012050. IOP Publishing (2016)
18. Jain, A., Gupta, S., Vyas, M., Pathy, D., Khare, G., Rajan, A., Rawat, A.: Open source EJBCA public key infrastructure for e-governance enabled software systems in RRCAT. In: ICT Based Innovations, pp. 127–139. Springer, Berlin (2018)
19. Wang, H.: A research on the security of wisdom campus based on geospatial big data. In: AIP Conference Proceedings, vol. 1967, p. 020025. AIP Publishing (2018)
20. Vargiu, E., Urru, M.: Exploiting web scraping in a collaborative filtering-based approach to web advertising. Artif. Intell. Res. **2**(1), 44–54 (2013)

Factors of Implementing Citizen-Centric e-Government in Developing Countries: Namibia

Karin Frohlich[1(✉)], Marko Nieminen[1], and Antti Pinomaa[2]

[1] Alto University, Espoo, Finland
karin.frohlich@aalto.fi
[2] Lapperanta-Lahti University of Technology, Lappeenranta, Finland

Abstract. The low adoption and use of e-Government have motivated interest in researching citizen-centric e-Government. This is important to developing countries where e-Government is seen as promoting social exclusion due to differences in income, access to ICTs and literacy rates across the populace. This study used a prototype m-Government application to investigate factors of citizen-centric e-Government. The study targeted one of the Ministries in Namibia and investigated challenges faced by citizens when accessing government service. While m-Government and e-Government could play a key role in addressing some of the challenges faced, it was noted that the government need to strategically implement e-Government given a number of contextual factors. For instance, while an urban-based citizen could afford a smartphone and operate one, the case is different with a citizen living in rural communities; rural populace, characterised by the aged cannot afford smartphones nor operate ones without assistance. This is mainly down to the fact that part of the rural populace cannot read and write in English or their native language. Hence, this study suggests understanding ICTs accessible to the populace, ICT skills, attitude towards technology, ICT infrastructure, costs, security and using a community-based partnership approach as key factors of e-Government in Namibia.

Keywords: M-Government · e-Government · Rural electrification · ICTs · Citizen-centric · Namibia

1 Introduction

The debate on factors of citizen-centric electronic government (e-Government) has become a topic of interest [1, 2]. A low electronic participation (e-participation) and the adoption of less interactive e-Government [3] platforms are some of the reasons why understanding citizen-centric e-Government in Africa has become important. Often, African governments struggle to align their e-Government implementations to existing technologies and governmental structures [4, 5]. This study uses a prototype mobile

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_52

government (m-Government) application to establish factors of e-Government implementation. Namibia is experiencing a fast-mobile phone adoption rate as reflected by an increase in mobile phone users from 4.32 per 100 inhabitants in the year 2000 to 105.79 per 100 inhabitants at the end of 2017 [6]. However, m-Government provision and ownership of a mobile phone do not guarantee e-participation [7]. In addition, research on citizen-centric e-Government has been focused on economically developed countries or developing countries outside Africa that may present unique adoption characteristics that are not comparable to those in Africa [8]. This is critical given that a number of factors play a key role in e-Government adoption including culture [9] something that limits the transferability of findings from one context to another without further confirmation. All these arguments motivated the need to investigate factors of e-Government in Namibia. This study focuses on factors of e-Government design and implementation in a selected government Ministry in Namibia. Findings from this study can be used as a point of reference when formulating policies for e-Government and m-Government adoption in Namibia and other countries that may exhibit similar characteristics to the referenced case.

2 Literature Review

2.1 Factors Influencing e-Government Design and Implementation

There is a growing interest in citizen-centric e-Government [10–12]. It is believed that citizen-centric e-Government has the potential to create public value as expected by the citizens [11, 13]. This study is guided by factors of e-Government design and implementation in Ochara and Mawela [5] and Jaeger and Bertot [14]. Accordingly, citizens' means of access to ICTs, ICT skills, community-based partnerships, information needs, ICT infrastructure and attitude towards ICTs are the key factors that were considered for citizen-centric e-Government in this study [5, 14].

Access. Ochara and Mawela [5] proposed that better access to ICTs could reduce the digital divide. In support of this proposition, Jaeger and Bertot [14] commented that, "a system that resides on technologies to which the citizen does not have access will also guaranty failure". As such, e-Government platforms should correspond with the technologies used by citizens [14]. This may involve making sure that Namibia's e-Government platforms are compatible with mobile phones that are widely used by the populace [3, 15, 16]. Approximately 80% of the Namibians use mobile phones to access the Internet [3]. In addition, language is one of the issues that may influence the access of e-Government services [4]. Namibia has at least eleven ethnic groups with at least eleven different officially recognised languages even though English is Namibia's official language [17].

ICT Skills. Computer illiteracy can promote digital divide across the population [5]. Jaeger and Bertot [14] noted that "developing a Web-based e-Government service that requires a broadband connection, high-end computer, and advanced technology competencies will immediately exclude a segment, or multiple segments, of the intended service population". As such, ICT skills are important to Namibia who is gearing towards a

"knowledge-based economy by 2030". Accordingly, the Namibian government is working on equipping its employees with ICT skills that could enable the use of e-Government services" [16].

Community-based partnership. Jaeger and Bertot [14] suggested that integrating e-Government with the immediate community's public resources such as libraries and schools has the potential to increase e-Government acceptance. Reference [4] echoed a similar view by suggesting that e-Government must be implemented in such a way that it adapts to the context not the other way around. As such, Namibia may implement its e-Government through schools and public libraries. Such an initiative is key as it avoid the idea of "off-loading" e-Government services to a community without overseeing other critical activities of training and user support [14].

Citizen information needs. Jaeger and Bertot [14] suggested that understanding how citizens search, acquire, and use information could provide basis for developing citizen-centric e-Government. This is important given that a significant number of users in the USA appear to trust information sourced through social media instead of government websites [18].

ICT infrastructure. The provision of ICT infrastructure in economically developing countries has been a concern for some time. The majority of developing countries, Namibia included, are dominated by rural areas where there is still sparse connectivity and have no electricity [19, 20]. Indications are that, rural areas are often excluded from e-Government access due to their geographical location and economic standing [6, 21–23]. This is so despite the growing use of mobile phones in the rural areas [6, 21–23].

Attitude. Attitude towards ICTs use is one of the factors considered important in determining the likelihood of e-Government use [5]. Thus, the willingness to use ICTs by the citizens can work in favour of e-Government use [16]. Attitude, arguably, reflects the extent of shift in the traditional cultural belief that ICTs can be used to facilitate the provision and access to e-Services without the need for human physical contact. Therefore, understanding the attitude towards ICTs can help ascertain the amount of effort the government should put in promoting e-Government and the likelihood of success.

3 Methodology

This study uses a qualitative case study strategy. The authors had no control over the subject study something that makes it suitable to conduct a case study as purported in Maoneke and Isabrye [24]. A selected Namibian government Ministry was considered as the case study for this research. The selected Ministry offers a wide range of government services that are critical to all Namibian citizens. The study makes use of a prototype m-Government application to gather data for evaluating factors of e-Government design and implementation. The prototype assumed one of the services (application for an identification —ID-document) offered by the selected Namibian government Ministry. The aim was to establish the perception of participants if a mobile application was used to apply for an ID instead of the current manual system. Below is an overview of the prototype, design factors, data collection, and analysis methods.

3.1 ID Application Prototype Design

This study made reference to the Finnish and Estonian electronic identity services in order to come up with a prototype of e-Government that is compatible with Namibia. Estonia is widely regarded as one of the leading countries in terms of e-Government development with close to 100% digital ID adoption [25]. The process of applying for an ID in Namibia is comparable to that of the Finland. The steps involved include completing and signing an application form that is followed by capturing the applicant's biometrics (fingerprints and photo). In addition, ID application in Namibia requires that applicants submit a copy of their birth certificate. This functionality was included in the prototype even though the Finish ID application process does not require a copy of the birth certificate. The wide adoption and use of mobile phones to access the Internet motivated the thought of basing the prototype on a mobile application. A mobile authentication option based on SIM cards were adopted for the prototype such that participants could be differentiated according to their mobile number. This was done for the sake of the experiment in the study and a more appropriate form of authentication need to be researched. Some screenshots of the prototype that was used in this study are presented in Fig. 1.

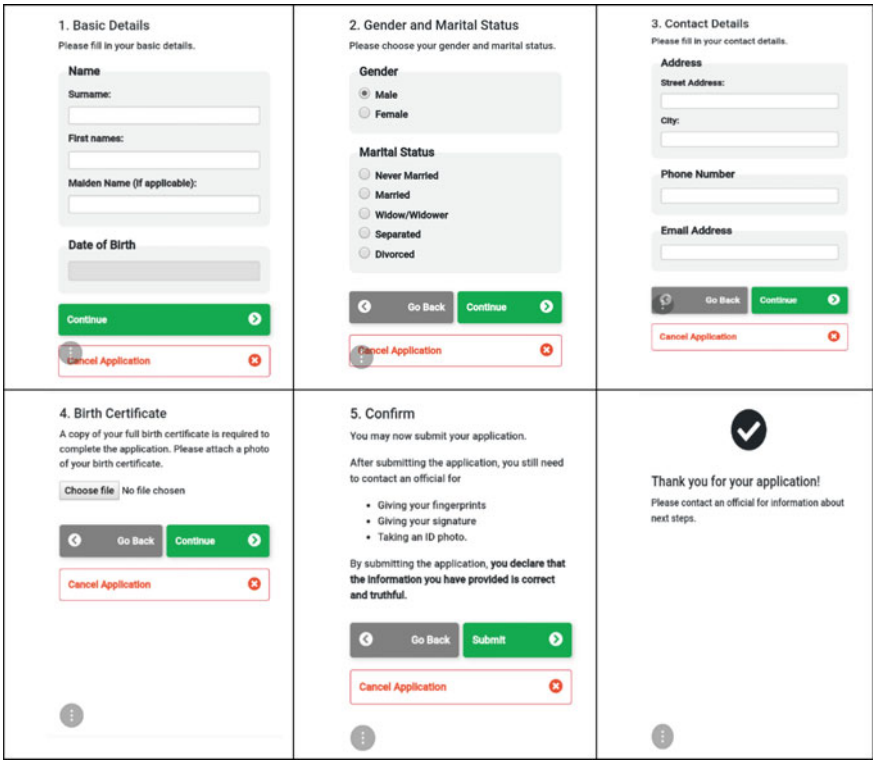


Fig. 1. Activities of the m-Government prototype

3.2 Data Collection and Analysis

In addition to participating in a fictitious ID application using a mobile prototype, participants were interviewed. Face-to-face interviews were done prior to exposure to the mobile application in order to gather data on background information about the participant and their perception on government service. Participants were then exposed to the prototype and asked further questions on their perception about using m-Government. All interviews were tape-recorded. Additional data was gathered through observations during face-to-face interviews and participation in the prototype experiment. All participants were randomly selected. Data analysis started with the transcription of tape-recorded interviews. This was followed by an analysis of response from each participant that saw findings on factors of m-Government implementation being coded according to those found in the literature. In addition, emerging themes were coded on their own. Lastly, themes on factors of m-Government implementation were put together, combining themes from all the participants and making comparisons taking note of similarities and differences.

4 Data and Sample Profile

Fifteen participants took part in the prototype experiment and face-to-face interviews. Participants were given labels (e.g., participant 1: P1) to avoid the use of their actual names for privacy reasons. Only Namibians were engaged.

4.1 Factors of e-Government Design and Implementation

This section presents study findings. Focus was on evaluating the importance of factors that influence the design and implementation of e-Government discussed in Sect. 2.

Access. It was observed that the majority of the participants owned a smartphone. Participant P4 lost her phone so she relied on her young son's (P5) phone. In addition, participant P6 indicated that he owns a feature phone. Mobile phone adoption is also extending to marginalised rural areas though at a slower pace compared to urban settings. Participant P15 from Omatako, a marginalised rural area, gave a brief overview of ICTs owned by people from his village "*let me say 60% of people have-have cell phones that they use. But not all can afford it, so only a few. Yeah, especially the young—the young people they do have smartphones but the elder people they just use their normal phones to communicate…*" Furthermore, there are suggestions that there is poor access to other ICTs. Participant P15 states that "*for computers I cannot say much, computer it's, it's really rare with our communities, I think not even… Not even more than ten people have computers, unless it's just teachers. Mostly, yeah. So, teachers I think depend on how many teachers we have at our school, I think most of them they have. But I think the community maybe just few, not more than ten people, maybe highest it's not more than 15. But the teachers I think it's even more than 15 now, so, that has…*"

ICT Skills. The majority of participants except, P4, demonstrated basic ICT skills of at least completing an online form, uploading a file and submitting a form. Participant P4 asked her son to participate in the prototype experiment on her behalf. This suggests older people may not be confident about their ICT skills; hence, they may need to engage their children on next of kin. In addition, participant P3 appeared shaky when completing the forms and skipped choosing a file (see step number 4 in Fig. 1) something that could be due to lack of knowledge about the feature. Nevertheless, the general feeling was that the prototype application was easy to use as participant P4 summarised his perception by saying that the process was "*very easy, [if the government introduce the app] they will bring that thing it will be very easy to apply for ID*". Participants indicated that they have different skills that include using the social media, email, electronic banking, browsing the Internet and text messages. However, the same cannot be said about rural dwellers, especially the elders, due to low literacy rates. Participant P15 indicated that "*most of the people [at my village] stay at the farms and they cannot even speak English. They just speak their home language or they use Afrikaans but that is difficult for them…..even for the written language, we the San people, we don't know how to read our own language*".

Community-Based Partnerships. Community-based partnerships are important for they could provide the necessary technical skills to support m-Government use at local level and provide access to others. Participant P2 indicated that, such an initiative "*it's gonna [going to] take a lot of convincing people*" for them to adopt e-Government. Furthermore, P4, a 47-year-old lady was not comfortable trying out on the system. This demonstrated a need for technical support if e-Government is to be used by the elderly. When P4 was asked to complete forms on a prototype, she stated that "*oh no I don't know what exactly [is] about a form*". She continued to state that "*can I use my baby, my son? [Son's name]. My son he will come and help you….He knows about mobiles*". Furthermore, participant P3 skipped a step to upload a fictitious birth certificate something that could be due to a lack of knowledge of the feature. On the other hand, participant P15 from a rural village appear to recommend the need of community-based partnerships that could help locals to adopt m-Government. P15 suggest of a "*hub*" or "*campsite*" where the community centre could go and gain access to ICTs for m-Government. Due to the language problem, P15 went on to recommend the use of translators as he stated that "*I think there should be even more than three people [at the hub], one should be of San speaking language and the other two they can speak English, then this one can translate for the people*". These findings suggest that community-based partnerships that could help citizens through these processes and steps involved are a critical requirement for e-Government in Namibia.

Attitude Towards Technology. Participants expressed mixed reaction on using IT in improving government service. There was a group of participants who showed a positive attitude towards using e-Government and or m-Government in light of the hardships faced with travelling to the government offices and waiting in long queues. Participant P4 states that "*maybe people will uh will take, go for that [using Internet to access government services], because now is the time for you know this– how do you call that? The Internet and that?*" Participant P8 stressed that the IT "*application would definitely work*". Similarly, P9 thought it was a good idea "*as long as there could be an option*

to verify the information submitted by the user. The people the ministry should be able to know the information is true without trusting the user to apply correct information". Participant P15 weighed in as he suggested that such an idea would reduce transport costs. He explained that "*it's less costly because people don't need to, to just to go move from one place to another just to fill in the details, I think it's something very good for the community as well. Cause they will just be at home and them everything will be completed without moving up and down*". In addition, P15 indicated that such a system could "*it saves time, and I think maybe for one-one-one house or maybe for people come together in a place where you can do this, it's faster than [travelling] to towns*" to apply for an ID.

However, a thought of people in the rural areas made participants to have some degree of reservation over the feasibility of using IT-driven system. Participant P15 explains "*it will need somebody trained that can help them, cause my community members, especially the elders and even the younger ones, they do not go to school, it's a problem. Cause our community members, most of them they, they have a communication problem when it comes to communicating in this official language-English*". Similarly, P10 raised concern over how such systems would be used by those based in rural areas. It was stated that, "*always coming up with these fancy ideas is always an issue [given] that the majority of the people, unfortunately, are staying in rural areas. So, it's always an issue as to how you would get this out and make it easy for them to use*". These views were shared by P8.

ICT Infrastructure. Interviews with participants revealed a number of ICT infrastructural concerns. Participant P2 indicated that "*they have a lack of infrastructure*" in rural areas. She went on to explain that rural areas "*don't have the MTC [Namibia's leading mobile phone service provider] towers close. They don't have access to Internet you understand*". In addition, P10 suggest there is poor network access in rural areas as she stated that "*the closer the people are to the rural areas they wouldn't have, you know, access to for example this e-services that government uh actually advocating for. So, I think that's a big issue currently*". Participant P15 went on to suggest that, while there are sections with base stations in the rural areas, signal reception remains poor. Participant P15 stated that "*at least, at our village we have Internet tower. But it's not that high, but it can cater for the surrounding. The other village we were yesterday, it's a problem, there's no Internet connection*". Similarly, one would need to position themselves strategically to get mobile phone reception in the villages as P15 went on to state that "*Only maybe when you try to call you can at least access but you have to stand in a way that the phone can be…..stable in order for you to communicate….network is very weak*". The same applies to supporting infrastructure for other ICTs. For instance, there are indications that there are poor radio signal reception rural areas. P15 states that "*we don't have radio connection. So we need an aerial in order for us to connect to radio*". Similarly, electricity access is poor in the village. Most households are not connected to the national electricity grid. "*We don't have electricity at all, and it's only the school and the clinic that has. Even with the phones, its very difficult cause we have to go to the school and ask if they can help to charge our phones*" says P15.

Costs. ICTs access costs are an important factor that imaged from the data. This appeared common to all participants. For instance, there was an impression that data may be expensive as most of the participants indicated that they did not have data in their mobile phones. Even though participants were not required to use their phones in testing the prototype, every time the interviewer asked her co-researcher about data availability, participants would jump into indicate they do not have data. For example, during conversations with participant P1, the interviewer asked: "*Ana, do you have uh data, could you share it?*" and participant P1 quickly responded "*no I don't have*". "*No, I'm asking her*", the interviewer had to clarify. Similarly, participant P2 stated that "*but I don't have data at the moment*". In addition, there are suggestions that ICT access costs might be higher in the rural areas. For instance, P15 indicated that "*the base that can help with printing or making copies is the school, and since it's only the school that can help, they tend not to*" out of fear that they may run of the stationery or experience breakdown in a location where there is limited sources of ICT support. In addition, printing and photocopying is "*very expensive even if the communities want to use it they won't be able cause it's expensive. They have to pay more than fifteen [Namibian] dollars just for one copy*". That is approximately US$1 per page or copy. These findings suggest data remain expensive for the majority of citizens and ICT infrastructure remains a critical requirement for a successful citizen-centric e-Government.

Security. Another theme that emerged from data analysis is the subject of security. It should be noted that the prototype recommended two authentication measures to participants. The first one was the authentication used at the bank or a mobile phone number. Nearly all participants were not comfortable to use the authentication they use at the bank. In particular, participants were not comfortable integrating bank authentication with that of a government applications system. Participant P7 states that "*when it comes to bank, they're like okay what if tomorrow I wake up and my account is empty*". Similarly, P8 explains that "*You know when it comes to bank, money, you know, you're just like ahhh, let me take mobile, it's a bit safer. I felt more comfortable with the mobile, yes*". P11 weighed in by stating that "*Mm, yeah, I'm not very comfortable with my bank details being shared at the ministry*".

5 Discussion, Recommendation and Conclusion

Government service delivery remains poor in developing countries. Information technology is expected to play a key role in enhancing service delivery. As such, citizen-centric e-Government is seen as one of the solutions that could improve government service delivery [1, 2]. This study investigated factors that should be looked at when implementing citizen-centric e-Government. Focus was on a selected government Ministry in order to establish how such an initiative could improve service delivery. The study found access to ICTs, ICT skills, community-based partnerships, attitude towards technology, citizen information needs, ICT infrastructure, costs and security as important factors for citizen-centric e-Government.

In particular to access to ICTs, the study found that the majority of participants own a mobile phone. However, this statistic may change in rural settings were few people own a phone and the most commonly owned phones in rural areas are the feature phones.

This could be explained by the fact that rural areas are twice as much likely to be poor compared to their urban counterparts [22]. As indicated by P15, only a few are able to own and use a smartphone. So, it is mostly "*teachers and the other elder people that have been working*" who can own and use a smartphone. When asked if people own computers in rural communities, participant P15 responded that "*for computers I cannot say much, computer, it's really rare with our communities, I think not even… not even more than ten people have computers*". As such, a citizen-centric e-Government suitable for Namibia would be one that is compatible with technologies owned by citizens. In this case, m-Government that is accessible on both smartphones and feature phones would be ideal. In addition, ICT skills were found popular among the youth and the working class. Elders based in rural areas are likely to struggle when it comes to operating a mobile phone. This is down to the fact that some of the rural dwellers cannot read. Hence, they mainly use phones for making calls and playing music. Besides, some of the elders need assistance to make a call. As such, the government need to look at how e-Government and m-Government can be delivered to such a population. The use of community-based partnerships can be one of the options given the low literacy rates and limited access to ICTs. Clinics and schools in rural areas that have access to electricity can be instrumental in providing centres for community-based partnerships in e-Government and m-Government. These study findings have already shown that rural populace makes use of local schools to charge their phones. Hence, providing e-Government and m-Government access through schools could be an ideal alternative. Again, such schools can also provide the necessary support where citizens find it difficult to operate e-Government systems. Furthermore, the attitude towards technology appears positive in light of challenges faced. However, more need to be done in promoting awareness, imparting the necessary skills and supporting ICT infrastructure. The study noted that mobile phone reception remains a challenge in rural areas. So does access to electricity. Namibia Statistics Agency [22] shows that the majority of rural households (85%) have limited or no access to the main electricity grid as they use firewood for cooking. Another key factor is costs where e-Government and m-Government should be accessible at relatively low costs. These study findings suggest that even the urban populace find the cost for data high. In the case of the rural population, the majority (31.1%) of the population's source of livelihood is farming a seasonal activity that may not be adequate to finance high costs of Internet access [22]. Lastly, the Namibian government should look into e-Government security design and mechanisms that is acceptable to the populace. While countries like Estonia have succeeded in its e-Government partly due to its digital IDs with digital signatures that interlink the private and public sector [25], there appears to be a lack of trust when it comes to interlinking information between government platforms and banks. As such, security and user authentication is one of the key issues, especially for the investigated Ministry considering the fact that, the Ministry offer services to a populace that does not have any form of formal identification document.

References

1. Lee, H., Irani, Z., Osman, I.H., Balci, A., Ozkan, S., Medeni, T.D.: Research note toward a reference process model for citizen-oriented evaluation of e-government services. Transform. Gov.: People, Process Policy **2**(4), 297–310 (2008)
2. Tsohou, A., Lee, H., Irani, Z., Weerakkody, V., Osman, I.H., Anouze, A.L., Medeni, T.: Proposing a reference process model for the citizen-centric evaluation of e-government services. Transform. Gov.: People Process Policy **7**(2), 240–255 (2013)
3. United Nations: United Nations e-Government Survey 2018. Gearing e-Government to Support Transformation Towards Sustainable and Resilient Societies. United Nations (2018)
4. Ochara, N.M.: Emergence of the E-Government artifact in an environment of social exclusion in Kenya. Afr. J. Inf. Syst. **1**(1), 18–43 (2008)
5. Ochara, N.M., Mawela, T.: Enabling social sustainability of e-participation through mobile technology. Inf. Technol. Dev. **21**(2), 205–228 (2013)
6. ITU: ITU releases 2018 global and regional ICT estimates. Available at: https://www.itu.int/en/mediacentre/Pages/2018-PR40.aspx last accessed: 6/11/2019
7. Rana, N., Janssen, M., Sahu, G.P., Baabdullah, A., Dwivedi, Y.: Citizens' perception about M-Government Services: results from an exploratory survey. In: Proceedings of the 52nd Hawaii International Conference on System Sciences (2019)
8. Nam, T., Sayogo, D.S.: Who uses e-Government? Examining the digital divide in e-government use. In: Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance, pp. 27–36. ACM (2011)
9. Shareef, M.A., Kumar, V., Dwivedi, Y.K., Kumar, U.: Service delivery through mobile-government (mGov): driving factors and cultural impacts. Inf. Syst. Front. **18**(2), 315–332 (2016)
10. Sorn-in, K., Taumsuk, K., Chaopanon, W.: Factors affecting the development of e-government using a citizen-centric approach. J. Sci. Technol. Policy Manage. **6**(3), 206–222 (2015)
11. Sigwejo, A., Pather, S.: A citizen-centric framework for assessing e-government effectiveness. EJISDC **74**(8), 1–27 (2016)
12. Osman, I.H., Anouze, A.L., Irani, Z., Al-ayoubi, B., Lee, H., Balc, A., Weerakkody, V.: COBRA framework to evaluate e-government services: a citizen-centric perspective. Gov. Inf. Q. **31**, 243–256 (2014)
13. Karunasena, K., Deng, H., Singh, M.: Measuring the public value of e-government: a case study from Sri Lanka. Transform. Gov.: People, Process Policy **6**(3), 258–282 (2012)
14. Jaeger, P.T., Bertot, J.C.: Designing, implementing, and evaluating user-centered and citizen-centered e-government. Int. J. Electron. Gov. Res. **6**(2), 1–17 (2010)
15. United Nations: United Nations e-Government Survey 2014. E-Government for the Future We Want. United Nations (2014)
16. Republic of Namibia: e-Government Strategic Action Plan for the Public Service of Namibia (2014–2018) (2014)
17. Peters, A., Winschiers-Theophillus, H., Mennecke, B.: Cultural Influences on Facebook Practices: A Comparative Study of College Students in Namibia and the United States. Digital Repository, Iowa State University (2015)
18. Mergel, I.: Social Media in the Public Sector: A Guide to Participation, Collaboration and Transparency in the Networked World. Wiley, New York, NY (2016)
19. Gumbo, S., Jere, N. Terzoli, A.: A qualitative analysis to determine the readiness of rural communities to adopt ICTs: a Siyakhula living lab case study. In: Proceedings of the IST-Africa 2012 Dar es Salaam, Tanzania (2012)
20. Pade-Khene, C., Plamer, R., Kavhai, M.: A baseline study of a Dwesa rural community for the Siyakhula information and communication technology for development project: understanding the reality on the ground. Inf. Dev. **26**(4), 265–288 (2010)

21. Namibia Statistics Agency: Poverty dynamics in Namibia: a comparative study using the 1993/94, 2003/04 and the 2009/10 NHIES surveys. Retrieved from: https://cms.my.na/assets/documents/p19dnar71kanl1vfo14gu5rpbkq1.pdf. Last accessed: 20/11/2019
22. Namibia Statistics Agency.: Namibia Inter-censal Demographic Survey 2016 Report. Retrieved from: https://cms.my.na/assets/documents/NIDS_2016.pdf. Last accessed: 1/11/2019
23. Republic of Namibia: Namibia's 5th National Development Plan (NDP5). Retrieved from: https://www.npc.gov.na/?wpfb_dl=294. Last accessed 25/1/2019
24. Maoneke, P.B., Isabirye, N.: An e-Commerce model for hospitality SMMEs, a case of the Eastern Cape. In: Proceedings of the Southern African Institute for Computer Scientist and Information Technologists Annual Conference 2014 on SAICSIT 2014 Empowered by Technology, p. 50. ACM (2014)
25. Kattel, R., Mergel, I.: Estonia's digital transformation. Mission Mystique and the Hiding Had. A digital success story. In: M.E. Compton P.'T. Hart (eds.) Great Policy Success. New York: Oxford University Press (2019)

Integration of Blockchain and Remote Database Access Protocol-Based Database

Muhammad Jafar Sadeq[1(✉)], S. Rayhan Kabir[1,2], Marjan Akter[2], Rokeya Forhat[2], Rafita Haque[1], and Md. Akhtaruzzaman[1]

[1] Asian University of Bangladesh, Dhaka, Bangladesh
{jafar,rayhan923,rafitahaque,azaman01}@aub.edu.bd, rayhan561@diu.edu.bd
[2] Daffodil International University, Dhaka, Bangladesh
{marjan563,rokeya35-1028}@diu.edu.bd

Abstract. Many companies are relying on software to manage their businesses. Usually, the software, especially those used by smaller companies, is not secure against unauthorized or unethical data manipulation on the database level. This paper recommends and demonstrates the use of blockchain for securing small enterprises against hacking by alerting the management whenever a change is made to the data without using the authorized channels. This is done through blockchain technology's inherent hash replication and mining algorithm. The paper shows an application where this idea has successfully been implemented for a desktop application built upon a remote database access (RDA) protocol-based relational database management system (RDBMS) such as remote MySQL and remote Oracle database.

Keywords: Blockchain · Remote database access (RDA) · Relational database management system (RDBMS) · Enterprise application integration (EAI) · Remote MySQL · Remote Oracle database · Information security

1 Introduction

Many small businesses are using desktop business software such as inventory management systems, accounting systems and/or purchase–sales systems. Depending on the size of the business, the database supporting the application might be stored in a single computer at the business premises or in a cloud to which the application connects. When the database is separate from the application, remote database access (RDA) protocols may be used for connecting the application to the database server. RDA-protocol-based remote relational database management systems (RDBMSs) such as remote MySQL and remote Oracle database are popular among software developers, but they suffer from threats including unauthorized access and data manipulation at the database level bypassing the business application [1].

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_53

To protect against the threat of unauthorized data manipulation, blockchain's resistance to data modification deserves consideration. Blockchain is an innovation for data and information security that is generally used for digital cryptocurrency such as Bitcoin, but it is being proposed for future uses in enterprise or business applications [2, 3]. Blockchain resides on a decentralized peer-to-peer (P2) network where there are no central server and database and the P2P network ensures the security of the data by using decentralized consensus [4].

There are already trends towards integration of blockchain with RDBMS. OurSQL [5] is a new technology that applies blockchain to a MySQL database using blockchain-MySQL proxies called OurSQL as a protection layer between the application and the actual MySQL database, and its use has been proposed in an IoT system for agriculture quality control [6]. However, OurSQL does not support RDA protocols for connecting to remote MySQL servers. In 2017, Oracle Corporation launched Oracle Blockchain [7] but the architecture of Oracle Blockchain is too complex [8]. Rather than integration of the blockchain with the RDBMS, Oracle Blockchain integrates several third-party Software as a Service (SaaS) and Platform as a Service (PaaS) with blockchain by using REST API [9].

In addition to these tools, it would be useful to have a tool that integrated blockchain with RDA protocol-based remote RDBMS. There are already some efforts to use blockchain and local RDBMS (local MySQL, SQLite), but work still needs to be done for remote RDBMS such as remote MySQL and remote Oracle database where RDA protocol is required. It was also found through anecdotal evidence by the present authors that industry software developers found the previous blockchain-RDBMS models hard to use for enterprise desktop applications.

Thus, the challenge is to develop a technique that is simple for the software developer and that can provide data security at the database level. To maintain simplicity, the technique should employ the fewest number of third-party libraries. Therefore, Blockchain-Integrated Remote Database (BIRD) has been developed for integrating blockchain with a remote RDBMS server. Specifically, this paper shows that an application has been implemented using remote MySQL and a Java-based desktop application, where the configuration file on the local computer stores the blockchain's blocks and encrypted hash values of data from the remote database server. Therefore, this paper provides the following contributions:

1. Integration of blockchain with remote RDBMS that is using RDA protocols.
2. Development of a blockchain platform using Java's client-side configuration files.
3. Development of a successful Java-based desktop ERP without any third-party libraries.
4. Demonstrating the ability to secure against database-level manipulation attacks.

2 Related Works

Integration of databases with blockchain is a topic of interest among various researchers. For example, a local SQL database has been used for storing and querying Bitcoin information [10]. SQLite, which is a local database and not a network database, has also been

used in an Ethereum-based blockchain system for securing SQL queries [11]. A PostgreSQL relational database-based blockchain implementation demonstrates a concept without providing mining process and node verification methods [12]. ETDB-Caltech is a blockchain-based method for distributing electron tomography data sets that takes advantage of the P2P nature of blockchain, where a local MySQL server has been used [13]. Specifically for IoT devices, blockchain was proposed for the use of storing device configuration files for distribution by the central system and for download by the devices [14].

BIRD intends to contribute to the increasing attempts of using blockchain with databases by showing how remote RDBMS using RDA protocols may be secured using blockchain.

3 Blockchain-Integrated Remote Database (BIRD)

BIRD has been developed to work on currently running systems by using existing files to support the blockchain mechanism. Any file that is hosted on the client machines in the business ERP can be used at a blockchain node by tagging the blockchain information onto that file. For example, configuration files such as the config.properties file in a Java-based desktop ERP are suited for this task. In this particular implementation of BIRD, the config.properties file was made to store the blocks and hash values of the remote MySQL database. Blockchain mining then compares the remote MySQL data with the configuration files' encrypted hash values to verify the local nodes. When a discrepancy is detected, an alert is raised.

Figure 1 shows how BIRD works. Two computer machines, Node 1 and Node 2, are connected to a remote MySQL database through separate desktop application installations. The applications are connected to the remote MySQL database through the Internet. Two different configuration files are located in the two computers as part of the respective desktop applications, which are able to read and modify the configuration files, which are usually located in the software installation folder.

A mining process continuously verifies the nodes by checking 'consensus', which means that the nodes all agree on the hash values stored in the configuration files. When any mismatch is detected by the mining process, an alert is given to the system administrators.

A test case may be as follows:

1. An invoice number is inserted into InvoiceList table of the remote MySQL server from a source node's application. Upon successful transaction, the source node modifies its blockchain.
2. After successful transaction, applications of all nodes except the source node are notified through the enterprise application to read the InvoiceList table and generate a new hash value for the new invoice.
3. The application reads the last CurrentHash of the configuration file (config.properties). It then creates a new block value of BlockID, PreviousHash and CurrentHash, where the BlockID is the new block's number, CurrentHash of the previous block is assigned as PreviousHash value of new block, and hash output of a getHash (newInvoice) method is assigned as a CurrentHash for the new block.

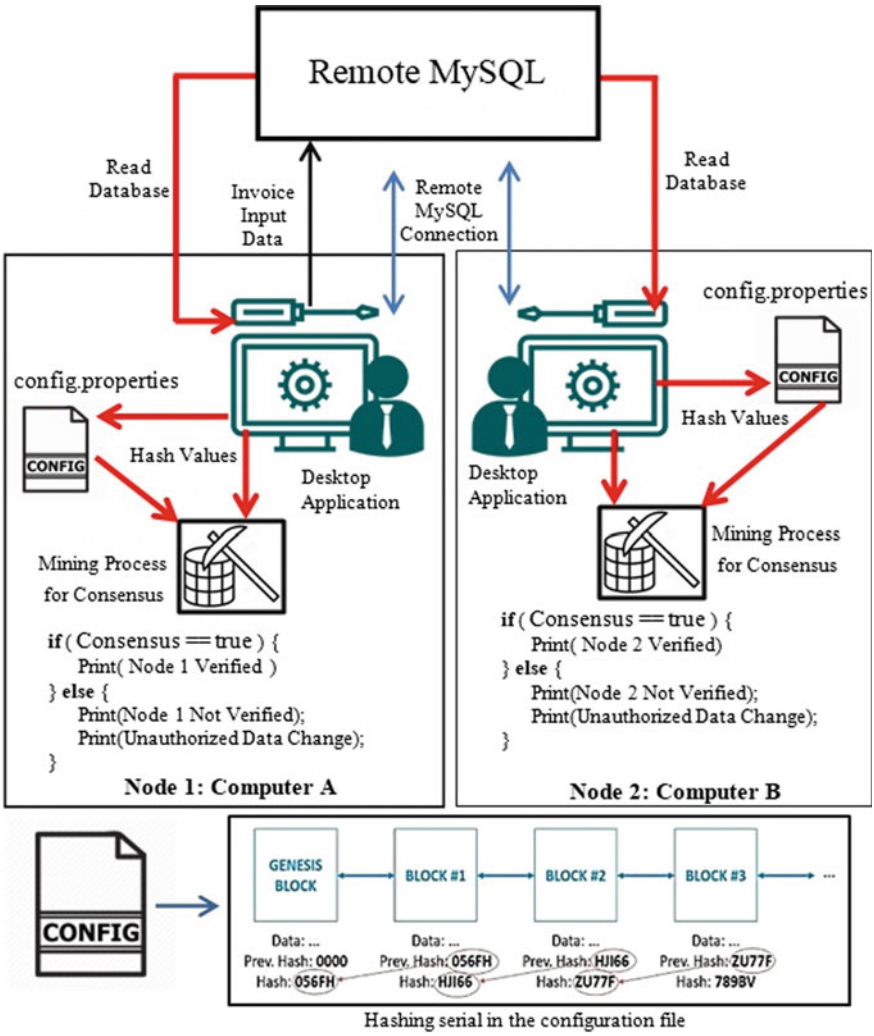


Fig. 1. BIRD model showing two nodes

4. The mining process runs continuously. If there is any mismatch between the hash values calculated from the remote database and the hash values stored in the nodes, an alert is given for unauthorized data change.

In summary, BIRD ensures that unauthorized data manipulation is detected by making changes to the blockchain at the application level. Thus, any changes that avoid the application will not have a resultant effect on the blockchain and hence cause a mismatch between the database and the blockchain.

4 Results

BIRD represents an approach where a developer can use blockchain for remote RDBMS servers for securing enterprise data by storing blockchain hash values in configuration files on client devices. While Java and MySQL have been used, the concepts presented are not limited to these tools.

Figure 2 displays a sample configuration file which includes a few blocks with BlockID, PreviousHash and CurrentHash. Any unauthorized alteration to the data associated with any block causes mismatch with two blocks: the data's actual block and the next block where the current block's hash value is kept as the PreviousHash.

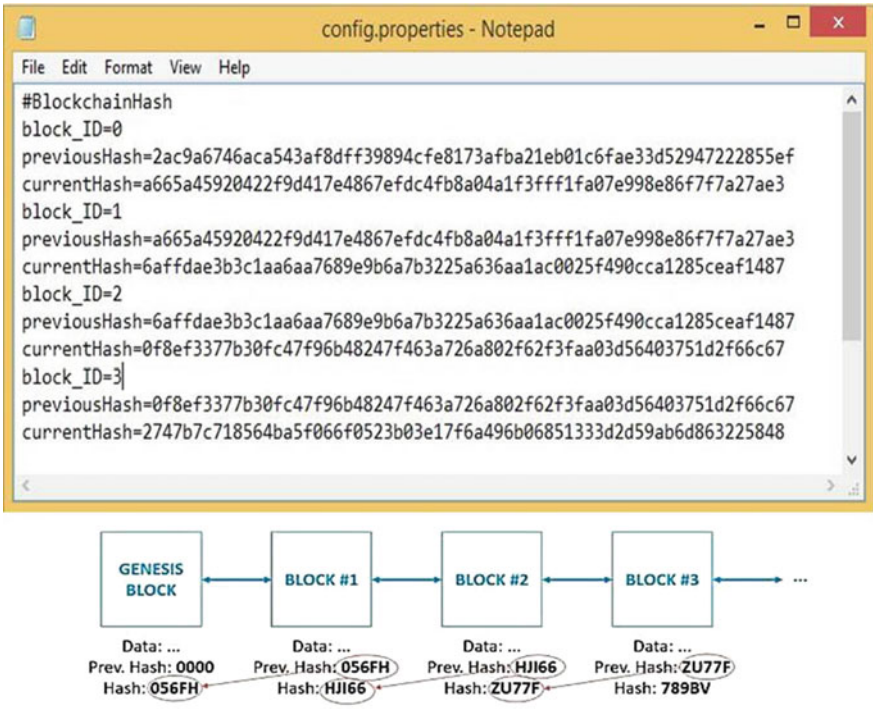


Fig. 2. Blockchain hashing sequence in the configuration file

Thus, BIRD can detect four kinds of unauthorized manipulation because they will all lead to hash mismatches, as follows:

1. Delete data from or add data to the database—this will cause the block sequence in the nodes' configuration tables to not match the data sequence in the table.
2. Change data in the database—this will cause mismatch in the CurrentHash value of a block and the PreviousHash value of the next block.
3. Change hash value in the configuration file—this will cause a mismatch in CurrentHash value with the data in the database, and also with the PreviousHash value of

the next block, and also a mismatch with the P2P network of blocks inherent in the blockchain.
4. Add or delete a hash block—this will cause the block sequence in the nodes' configuration tables to not match the data sequence in the table and also a mismatch with the P2P network of blocks inherent in the blockchain.

5 Conclusion

For small businesses, the inhibiting cost of maintaining a database server often results in the use of cloud-based remote databases upon which their entire applications are built. RDA protocols are used to access these remote databases. Also, enterprise applications owned by small businesses are less likely to be rigorously secure from unauthorized data manipulation. Thus, Blockchain-Integrated Remote Database has been presented in this paper to show a technique where unauthorized data changes can be detected using blockchain's inherent hash replication and mining process. This has been done through the storage of the hash values in configuration files of the enterprise application's client nodes, and modification of these files is only done through the enterprise application. This ensures the detection of all changes made to the data at the database level.

The current work has been implemented, but its effectiveness and robustness from hacking have not yet been identified. Also, since there are regular updates to the blockchain, applications with large amount of transactions and large data sets may have performance issues. These are all matters of future study.

References

1. Sanfilippo, J., Abegaz, T., Payne, B.R., Salimi, A.: STRIDE based threat modeling for MySQL databases. In: Arai, K., Bhatia, R., Kapoor, S. (eds) Proceedings of the Future Technologies Conference (FTC), Volume 1070 of Advances in Intelligent Systems and Computing, pp. 368–378. Springer, Cham (2019)
2. Al-Jaroodi, J., Mohamed, N.: Blockchain in industries: a survey. IEEE Access **7**, 36500–36515 (2019)
3. Konstantinidis, I., Siaminos, G., Timplalexis, C., Zervas, P., Peristeras, V., Decker, S.: Blockchain for business applications: a systematic literature review. In: Abramowicz, W., Paschke, A. (eds) Business Information Systems, Volume 320 of Lecture Notes in Business Information Processing, pp. 384–399. Springer, Cham (2018)
4. Acharjamayum, I., Patgiri, R., Devi, D.: Blockchain: a tale of peer to peer security. In: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 606–617, November 2018
5. Roman Gelembjuk. Our SQL: MySQL + blockchain
6. Surasak, T., Wattanavichean, N., Preuksakarn, C., Scott, C.-H.: Thai agriculture products traceability system using blockchain and internet of things. Int. J. Adv. Comput. Sci. Appl. **10**(9) (2019)
7. Zsigri, C., Lehmann, C.: Oracle launches its blockchain cloud service at openworld 2017. resreport, 451 Research, October 2017
8. Egberts, A.: The oracle problem an analysis of how blockchain oracles undermine the advantages of decentralized ledger systems. SSRN Electron. J. (2017)

9. Blockchain technology for the enterprise
10. Yue, K.-B., Chandrasekar, K., Gullapalli, H.: Storing and querying blockchain using SQL databases. Inf. Syst. Educ. J. **17**(4), 24–41 (2019)
11. Han, J., Kim, H., Eom, H., Coignard, J., Wu, K., Son, Y.: Enabling SQL-query processing for ethereum-based blockchain systems. In: Proceedings of the 9th International Conference on Web Intelligence, Mining and Semantics WIMS2019. ACM Press (2019)
12. Nathan, S., Govindarajan, C., Saraf, A., Sethi, M., Jayachandran, P.: Blockchain meets database: design and implementation of a blockchain relational database. Proc. VLDB Endowment **12**(11), 1539–1552 (2019)
13. Ortega. D.R., Oikonomou, C.M., Jane Ding, H., Rees-Lee, P., Jensen, G.J.: ETDB-caltech: a blockchain-based distributed public database for electron tomography. PLOS ONE **14**(4), e0215531
14. Košťál, K., Helebrandt, P., Belluš, M., Ries, M., Kotuliak, I.: Management and monitoring of IoT devices using blockchain. Sensors **19**(4), 856 (2019)

Crossing the Artificial Intelligence (AI) Chasm, Albeit Using Constrained IoT Edges and Tiny ML, for Creating a Sustainable Food Future

Chandrasekar Vuppalapati[1](✉), Anitha Ilapakurti[2], Sharat Kedari[2], Raja Vuppalapati[2], Jaya Vuppalapati[3], and Santosh Kedari[4]

[1] Computer Engineering, San Jose State University, San Jose, USA
chandrasekar.vuppalapati@sjsu.edu
[2] AI and IoT, Hanumayamma Innovations and Technologies, Inc, Fremont, USA
{ailapakurti,sharath,raja}@hanuinnotech.com
[3] AI and IoT, Hanumayamma Innovations and Technologies PVT Limited, Hyderabad, India
jaya.vuppalapati@hanuinnotech.com
[4] Customer Success and Product, Hanumayamma Innovations and Technologies PVT Limited, Hyderabad, India
skedari@hanuinnotech.com

Abstract. Big data surrounds us. Every minute, our smartphone collects huge amount of data from geolocations to next clickable item on the e-commerce site. Data has become one of the most important commodities for the individuals and companies. Nevertheless, this data revolution has not touched every economic sector, especially rural economies; e.g., small farmers have largely passed over the data revolution, in the developing countries due to infrastructure and constrained compute environments. Not only this is a huge missed opportunity for the big data companies, but it is one of the significant obstacles in the path toward sustainable food and a huge inhibitor closing economic disparities. The purpose of the paper is to develop a framework to deploy artificial intelligence models in constrained compute environments that enable remote rural areas and small farmers to join the data revolution and start contribution to the digital economy and empower the world through the data to create a sustainable food for our collective future. In a nutshell, close the digital gap by crossing the AI chasm to democratize AI for poor and helpless farmers and help ourselves by creating sustainable food future.

Keywords: Edge · IoT device · Artificial intelligence · Kalman filter · Dairy cloud · Small-scale farmers · Hardware constrained model · Tiny ML · Hanumayamma · Cow Necklace

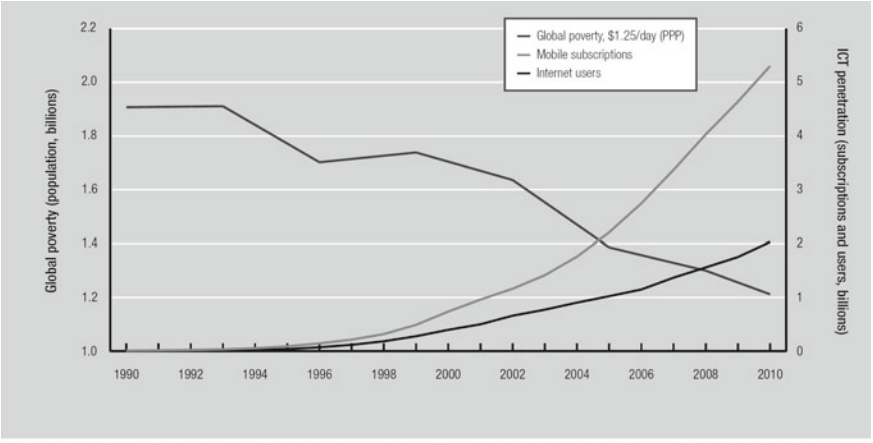
X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_54

1 Introduction

Artificial intelligence (AI) stands out as a transformational technology of our digital age—and its practical application throughout the economy is growing apace [1]. One of the chief reasons why AI applications are getting prominence and industry acceptance is AI software ability to learn, albeit continuously, from real-world use and experience, and its capability to improve its performance [1]. It is no wonder that the applications of AI span from complex high-technology equipment manufacturing to personalized exclusive recommendations.

Nevertheless, this data revolution has not touched every economic sector, especially rural economies; e.g., small farmers have been largely passed over the revolution, in the developing countries due to infrastructure and constrained compute environments even when AI is critical for ensuring food sustainability[1] [2].

Although deeper ICT penetrations lead to falling absolute global poverty rates (see Fig. 1), ironically, there is striking imbalance of ICT spread in developing economies. As a result, though developing economies' agriculture sector is the largest employer of the world (see Fig. 2), there is almost no dissemination of AI to help small farmers and to create a sustainable future for all.



Sources: World Bank *PovCal* database (1990, 1993, 1996, 1999, 2002, 2005, 2008, 2010); authors' calculations and interpolation, ITU *World Telecommunication/ICT Indicators* database June 2013.

Fig. 1 Falling global absolute poverty and rising ICT penetration [3]

In noting the promise and challenge of AI, the McKinsey Global Consulting Firm noted numerous use cases across many domains where AI could be applied and for these AI-enabled interventions to be effectively applied, several barriers must be overcome [5]. These include the challenges of data, information and communication technology (ICT),

[1] How to Sustainably Feed 10 Billion People by 2050, in 21 Charts. https://www.wri.org/blog/2018/12/how-sustainably-feed-10-billion-people-2050-21-charts.

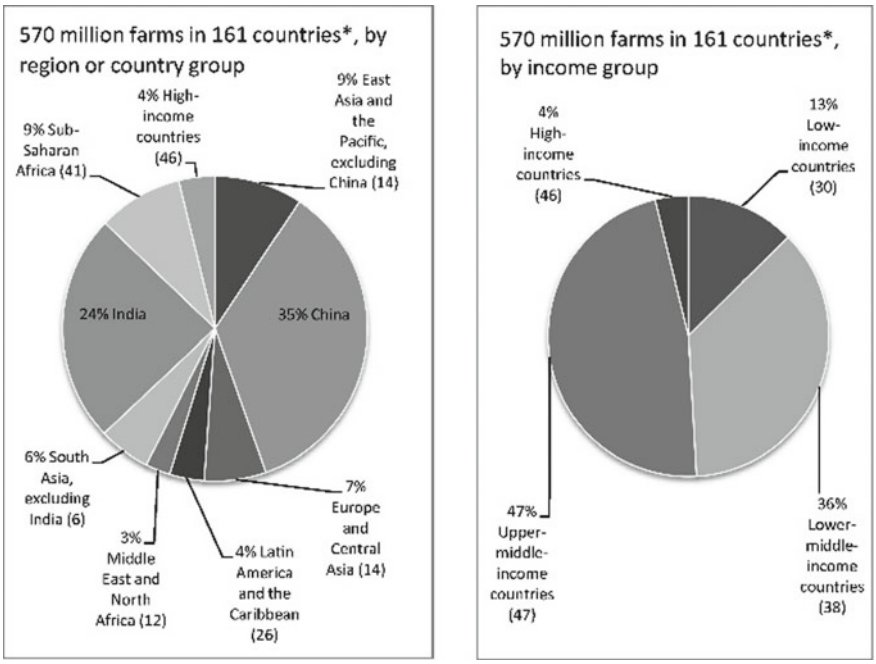


Fig. 2 Small farms [4]

computing, and talent availability, as well as more basic challenges of access, infrastructure, and financial resources that are particularly acute in remote or economically challenged places and communities.

One chief reason, importantly, for AI applications not touched every economic sector is the current AI algorithms are only made to run on very powerful research workstations without considering the real-world implications of constrained hardware. Machine learning in embedded systems specifically targets embedded system to gather data, process data, and apply mathematical rules to produce insights [6]. The embedded systems typically consist of low memory, low RAM, limited power compared to regular computers. Increase in factors such as processing power [7] leads to higher accuracies—the cost to bear is battery life.

To successfully disseminate AI to masses and enable successful democratization, we need to bring rural communities to digital revolution (people, technology, and data together) and the purpose of the paper is to achieve such digital revolution. The paper proposes AI deployment to small-footprint (tiny) IoT edge device that operates in constrained device and presents the data collected in real production environment. Our goal is to achieve AI for all, a true Fourth Industrial Revolution [8].

2 Democratization of AI

The democratization of AI is need of the day. The current AI is more applicable for businesses and end-consumers who are mostly city dwellers. Lack of data that could potentially help local businesses and societies is one of the most significant challenges in AI adoption. One limiting challenge is the availability of data for social good use cases, especially in rural communities. Lack of technologies, importantly, in the hands of the users in rural settings exacerbating the proliferation and democratization of AI issue. For instance, global penetration of Internet in rural areas is very low compared to suburban and urban area and this is persisting wider digital gap between rural and urban areas [8]. Lack of Internet connectivity is causing inhibition of digital data services dissemination, resulting into sparse vital dataset capture for better governance purposes and preventing rural population to participate in the digital economy.

2.1 The Calling

As the global population grows from 7 billion in 2010 to a projected 9.8 billion in 2050,[2] and incomes grow across the developing world, overall food demand is on course to increase by more than 50%, and demand for animal-based foods by nearly 70% [3]. Yet today, hundreds of millions of people remain hungry, agriculture already uses almost half of the world's vegetated land, and agriculture and related land-use change generate one-quarter of annual greenhouse gas (GHG) emissions [3].

The menu for a sustainable food future includes [3]:

1. Reduce growth in demand for food and other agriculture products.
2. Increase food production without expanding agricultural land.[3]
3. Protect and restore natural ecosystems and limit agricultural land-shifting.
4. Increase fish supply.
5. Reduce greenhouse gas emissions from agricultural production.

The World Resources Institute explicitly calls for technology enablement and diffusion to produce sustainable and green agriculture to keep the planet in a good form for our future generations.

2.2 Data

The world is data [9]. Small farmers make big contribution to agriculture and dairy production in developing countries [10]. Unlike the dairy farms of the west, milk originates in highly decentralized villages with the help of small farmers who own three to five cattle and they bring milk twice a day to milk collection centers to get paid [11]. Simply put, the livelihood of roughly 2 billion people (26.7% of the world population) of small

[2] Creating a sustainable food future, https://wrr-food.wri.org/.

[3] Course 2 Increase Food Production without Expanding Agricultural Land-https://wrr-food.wri.org/course/increase-food-production-without-expanding-agricultural-land-synthesis.

farmers in developing world depends on agriculture and the climate change adversely impacting their survival (see Fig. 2) [12].

Additionally, lack of data for serving these farmers puts food sustainability and food security in a huge risk mode [13].

2.3 Paradigm Shift—Compute Shifting to the Edge

Data generation and processing are shifting from corporate data centers or central servers to edge servers or machines [14]. Four big technology drivers underlie the massive shift of data to the edge[4][5]:

1. Artificial intelligence (AI) has become cost-effective and practical.
2. Billions of IoT devices are being deployed.
3. Wireless operators are upgrading their networks to the fifth generation of cellular mobile communications (5G).
4. Innovations in edge data centers are solving for the complexities of distributed facilities and unit cost economics.

Nevertheless, the processing at the edge comes with its own processing and constraints [15]. Here is the list of conditions that the edge devices should endure to provide resilient technologies (see Fig. 3):

- Harsh conditions
- Remote locations
- Quality of service (QoS)/low latency
- No-touch/self-healing
- Global data experience.

2.4 The Last Mile—Constrained Compute Environments and "AI Chasm"

Fourth wave of compute has spurred the development of edge devices. Edge devices come in various forms and shapes with varying compute capacities (Class 0, Class 1, and Class 2) [16]. Class 0 and Class 1 devices collect vast amount of environmental and geolocation data on a periodic basis. Due to constrained environments, the Class 0 devices require external devices such as gateways and mobile phones to relay to the Internet. However, these device deployments for small farmer are sporadic.

(1) *Class 0 Devices*

Class 0 devices are very constrained sensor-like motes. These devices are so severely constrained in memory and processing capabilities that most likely they will not have the resources required to communicate directly with the Internet in a secure manner.

[4] Data at the Edge https://www.seagate.com/enterprise-storage/what-is-it-4-0/.

[5] State of the Edge https://www.seagate.com/www-content/enterprise-storage/it-4-0/images/Data-At-The-Edge-UP1.pdf.

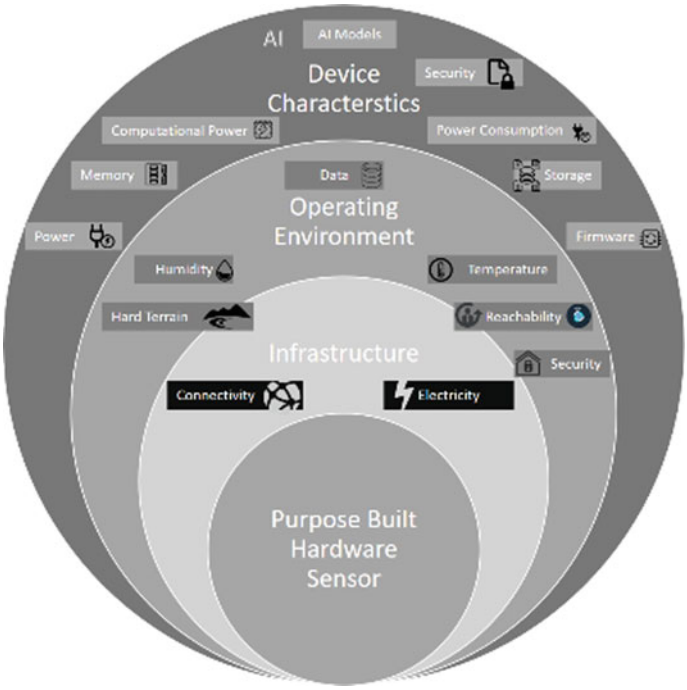


Fig. 3 Edge device operating environment

(2) *Class 1 Devices*

Class 1 devices are ***quite constrained*** in code execution space (stack and register level) and processing capabilities, such that they cannot easily talk to other Internet nodes employing a full protocol stack such as using Hypertext Transfer Protocol (HTTP), Transport Layer Security (TLS), and related security protocols and Extensible Markup Language (XML)-based data representations [16].

2.5 Constrained Device (Tiny IoT Edge) Architecture

As shown in Fig. 4, constrained devices are limited by the compute power, memory, storage space, stack space, and work in limited infrastructure. In general, the devices have a central microcontroller as a processing unit with sensors tied to the device unit. The sensors collect data on time–frequency—frequencies of collection would affect the battery useful time.

On a broad level, constrained devices are grouped on three characteristics:

- Model accuracy
- Connectivity
- Hardware economy.

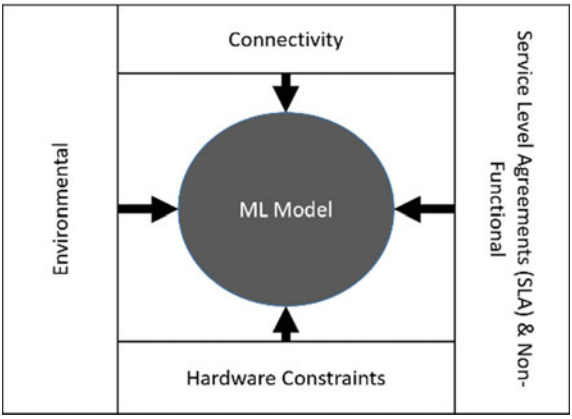


Fig. 4 Tiny ML or constrained edge

2.6 The Trade-off Model

The trade-off model is used in finance and manufacturing. The conventional trade-off model states that unless there is some slack in the system, improving any one of the four basic manufacturing capabilities—quality, dependability, speed, and cost—must necessarily be at the expense of one or more of the other three.

The trade-off model is used in finance and manufacturing. The conventional trade-off model[6] states that unless there is some slack in the system, improving any one of the four basic manufacturing capabilities—quality, dependability, speed, and cost—must necessarily be at the expense of one or more of the other three capabilities.

The trade-off model (see Fig. 5) provides seesaw view of a system by dissecting the influence of one of the key lever modeling/variations on the entire workings/penalties/compromises of the two key characteristics of the system. Generally, the variation of one key lever with respect to other important design parameters is necessary to develop a resilient, robust, and fault-tolerant system. For example in the below system, trade-off model depicts the performance implications of a ML model, namely: in-memory large dataset plus computational complexity (O (n log (n))) impact on the battery longevity. Of course, in the below system only two scalar components of a system are depicted. In real-world production system, multiple interplay of different key design levers is compared and performance studied on other parameters. Put it plainly, a cross-influence matrix of system performance needs to be considered.

3 ML Mode Framework

We have deployed model as part of Cow Necklace. The following hardware consists of accelerometers, gyroscope, temperature, humidity and onboard Bluetooth connectivity.

The sensor module is built (see Fig. 6) on working in constrained environments [17].

[6] Trade-off Models https://www.ifm.eng.cam.ac.uk/research/dstools/trade-off-models/.

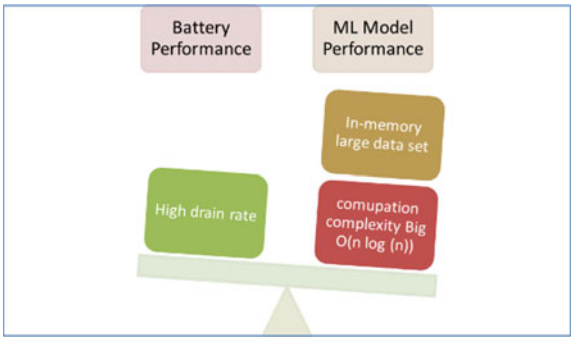


Fig. 5 Trade-off

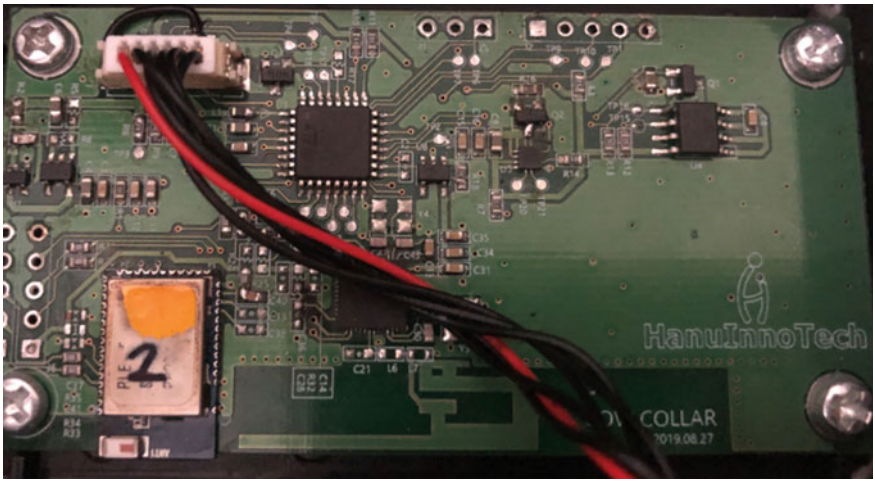


Fig. 6 Cow Necklace (tiny IoT edge)–PCB board

The Cow Necklace sensor connects to mobile (see Fig. 7) using Bluetooth Low Energy (BLE) and uploads data to the Dairy Analytics Cloud. The sensor collected data (see Fig. 8).

The machine learning model in constrained environment is subjected to various constraints and trade-offs [18].

- Hardware to model accuracy
- Model accuracy to connectivity
- Connectivity to hardware.

The balance has to be drawn with respect to applicability versus model accuracy (see Fig. 9). For instance, if ML deployed model is active learner (e.g., *K*-means cluster), the power consumption is taxed very high as the algorithm dynamically allocates *K* values. On the other hand, if deployed model is lazy learners, the model evaluation is based on

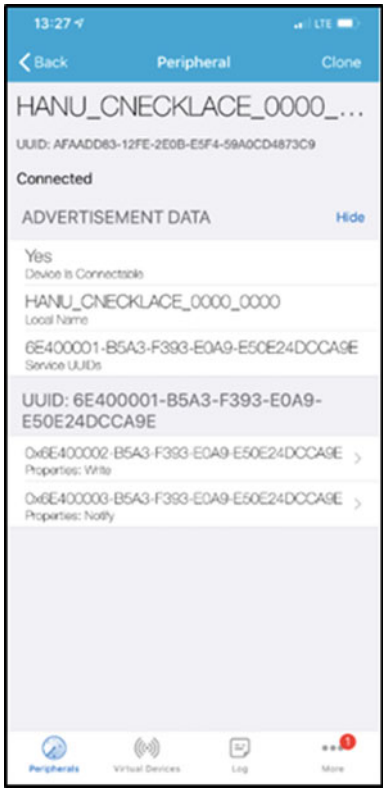


Fig. 7 Cow Necklace mobile app

	0 Date	Time	X	Y	Z	SensorTestAmbient_1	SensorTestAmbient_1	Heat_indx	Heat_indx	Warnings_	Warnings_	Kalman_e	vatflag_ar	Kalman_e	vatflag_humidity
[101]	11/30/2019	14:00:01	-991	-188	-16185	11.4	11.7	69.5	66.7	51.2	50.74	Normal	Normal	19.9	FALSE	NaN	FALSE
[102]	11/30/2019	15:00:01	-991	188	16185	11.4	11.7	69.5	66.7	51.2	50.74	Normal	Normal	17.17	FALSE	NaN	FALSE
[103]	11/30/2019	15:00:01	-991	-188	-16185	11.4	11.7	69.5	66.7	51.2	50.74	Normal	Normal	15.34	FALSE	NaN	FALSE
[104]	11/30/2019	16:00:01	803	230	16183	11.4	11.7	68.6	67	51.22	50.7	Normal	Normal	14.13	FALSE	NaN	FALSE
[105]	11/30/2019	17:00:01	-914	-158	-16153	11.4	11.7	68.4	67	51.22	50.69	Normal	Normal	13.32	FALSE	NaN	FALSE
[106]	11/30/2019	18:00:01	-901	210	16184	11.3	11.6	68.4	67	51.02	50.49	Normal	Normal	12.75	FALSE	NaN	FALSE
[107]	11/30/2019	19:00:01	-884	-100	-16164	11.3	11.6	68.6	67	51.03	50.5	Normal	Normal	12.36	FALSE	NaN	FALSE
[108]	11/30/2019	20:00:01	-908	-210	-16149	11.2	11.5	69	67.2	50.89	50.32	Normal	Normal	12.08	FALSE	NaN	FALSE
[109]	11/30/2019	21:00:01	-933	210	16166	11.3	11.5	69	67.3	50.83	50.52	Normal	Normal	11.88	FALSE	NaN	FALSE
[110]	11/30/2019	22:00:01	-935	-238	-16155	11.3	11.6	72.7	67.4	51.04	50.69	Normal	Normal	11.79	FALSE	NaN	FALSE
[110]	11/30/2019	22:00:01	-935	238	16155	11.3	11.6	72.7	67.4	51.04	50.69	Normal	Normal	11.73	FALSE	NaN	FALSE
[110]	11/30/2019	22:00:01	-935	-238	-16155	11.3	11.6	72.7	67.4	51.04	50.69	Normal	Normal	11.68	FALSE	NaN	FALSE
[110]	11/30/2019	22:00:01	-935	238	16155	11.3	11.6	72.7	67.4	51.04	50.69	Normal	Normal	11.65	FALSE	NaN	FALSE
[110]	11/30/2019	22:00:01	-935	-238	-16155	11.3	11.6	72.7	67.4	51.04	50.69	Normal	Normal	11.64	FALSE	NaN	FALSE
[110]	11/30/2019	22:00:01	-935	238	16155	11.3	11.6	72.7	67.4	51.04	50.69	Normal	Normal	11.62	FALSE	NaN	FALSE
[111]	12/1/2019	0:00:01	-931	-240	-16155	11.8	12	70.3	68.1	51.86	51.57	Normal	Normal	11.75	FALSE	NaN	FALSE
[112]	12/1/2019	0:00:01	-921	-240	-16155	11.8	12	70.3	68.1	51.86	51.57	Normal	Normal	11.93	FALSE	NaN	FALSE
[113]	12/1/2019	0:00:01	921	240	16155	11.8	12	70.3	68.1	51.86	51.57	Normal	Normal	11.89	FALSE	NaN	FALSE

Fig. 8 Sensor data

the memory-resident stack space algorithm evaluation. Here, the rules are in high-drive mode to execute the model.

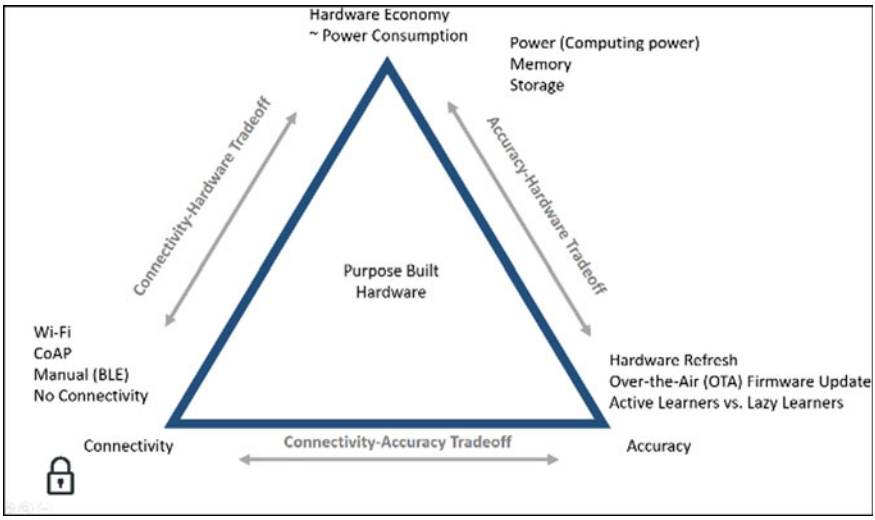


Fig. 9 Hardware-ML model-connectivity framework

3.1 Hardware-Model Accuracy Table (Constraint–Connectivity)

For evaluating various conditions that are subjected to hardware to model accuracy, holding connectivity, an infrastructure aspect, as a constraint the following are to be considered:

Model: Hardware versus model accuracy

Constraint: Connectivity

The connectivity, which is infrastructure service, could vary based on the geographical location:

Connectivity options:

- Wi-Fi
- Manual (Bluetooth Low Energy)
- No connectivity.

Constraint–Wi-Fi Connectivity

With Wi-Fi availability, the model could be updated during the hardware refresh or via over the air (OTA) (see Fig. 10).

Over the Air Model Update

This option provides more flexibility as it has influence on the model in-memory and storage options (see Fig. 11):

Hardware—Memory: Low

- The most optimized and updated models could be deployed on the sensor.

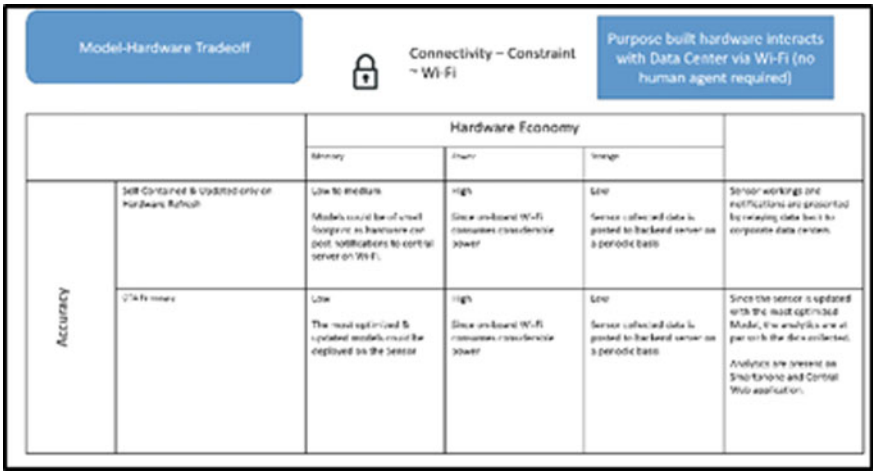


Fig. 10 Connectivity: Wi-Fi

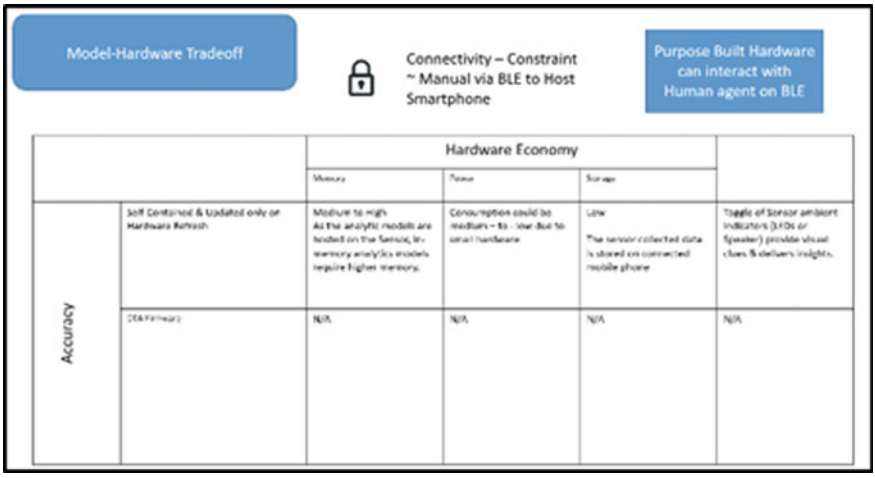


Fig. 11 Connectivity: manual

Hardware—Power: High

- Since onboard Wi-Fi consumes considerable power.

Hardware—Storage: Low

- Sensor collected data is posted to backend server on a periodic basis.

Constraint–Manual Connectivity

With manual connectivity, either Bluetooth Low Energy, the model execution and hardware have huge performance or tax penalties. Let us look the following cases:

Model update during hardware refresh, the following are considerations:

Hardware—Memory: High

High due to self-contained model with high memory—host models (for historical and outlier detection)

Hardware—Battery: High

- High—to support in-memory and compute operations

Hardware—Battery: Storage

- Since no connectivity, the data collected to be saved on

Hardware design consideration: Toggle of sensor ambient indicators (LEDs or speaker) provides visual clues and delivers insights.

- **Kalman Model Code**

The following code predicts Kalman temperature [19]:

```
    # Formulas
    def
TempPrediction(PreviousEstimate,currentMeasurement,Pre
viousErrorInEstimate):
      ErrorInEstimate = 2
      ErrorInMeasurement= 4
      KalmanGain = ErrorInEstimate /(ErrorInEstimate +
ErrorInMeasurement)
      CurrentEstimate = PreviousEstimate +
KalmanGain*(currentMeasurement - PreviousEstimate)
      # step2
      ErrorInEstimate = (1-
KalmanGain)*(PreviousErrorInEstimate)
 return CurrentEstimate,ErrorInEstimate
```

4 Conclusion

Democratization of artificial intelligence is the need of the day. It is our responsibility to develop models and hardware equipment that enable the collection of the data from the constrained environments so as to model the AI for food sustainability and threats that we face as humans—climate change. Finally, it is our ardent belief that the data is our best defense and the savior against the negative effects of climate change. The sooner we embark on democratization of AI to small farmers, the better we leave our progeny a wonderful life on the earth, i.e., better than what we have inherited.

Acknowledgements. We are very thankful to the management of Hanumayamma Innovations and Technologies Inc., for providing sensor and sensor data to publish as part of the paper.

References

1. Chui, C., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., Malhotra, S.: Notes from the AI frontier: applications and value of deep learning (2018). https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning
2. Amaral, L.: The data revolution hasn't yet hit agriculture. (05 March, 2019). https://www.wri.org/blog/2019/03/data-revolution-hasnt-yet-hit-agriculture
3. Searchinger, T., Waite, R., Hanson, C., Ranganathan, J.: In: Creating A Sustainable Food Future. (2018). https://wrr-food.wri.org/
4. Lowder, S.K., Skoet, J., Raney, T.: The number, size, and distribution of farms, smallholder farms, and family farms worldwide. pp. 16–29. (Nov 2016). https://www.sciencedirect.com/science/article/pii/S0305750X15002703
5. Manyika, J., Bughin, J.: The promise and challenge of the age of artificial intelligence. (Oct, 2018). https://www.mckinsey.com/featured-insights/artificial-intelligence/the-promise-and-challenge-of-the-age-of-artificial-intelligence
6. Van Ranst, W.: Real world applications of artificial intelligence on constrained hardware. Web (2019)
7. Coldewey, D.: Overclocked smartwatch sensor uses vibrations to sense gestures, objects and locations. (21 November, 2016). https://techcrunch.com/2016/11/21/overclocked-smartwatch-sensor-uses-vibrations-to-sense-gestures-objects-and-locations/
8. Perrin, A.: Digital gap between rural and nonrural America persists. (May 2019). https://www.pewresearch.org/fact-tank/2019/05/31/digital-gap-between-rural-and-nonrural-america-persists/
9. Thomas, R., McSharry, P.: In: Big Data Revolution: What Farmers, Doctors And Insurance Agents Teach Us About Discovering Big Data Patterns, 1st edn, Wiley. (2 March 2015). ISBN-10: 9781118943717
10. Rao, D.R.: Small farmers make big contributions: dairy production in India. (7 February 2017). https://ucanr.edu/blogs/blogcore/postdetail.cfm?postnum=23184
11. Prahalad, C.K., Hart, S.L.: Strategies for the bottom of the pyramid: creating sustainable development (1999).http://pdf.wri.org/2001summit_hartarticle.pdf
12. Watson, R.T.: In: Industrial Agriculture and Small-scale Farming. (2009). https://www.globalagriculture.org/report-topics/industrial-agriculture-and-small-scale-farming.html
13. Schumacher, E.F.: In: Small is Beautiful: Economics as if People Mattered Paperback. Perenial Library/Harper and Row, (August 1975). ISBN-10: 9780060803520
14. Morris, J., Trifiro, M., Smith, J.:In: State of the Edge. https://www.seagate.com/www-content/enterprise-storage/it-4-0/images/Data-At-The-Edge-UP1.pdf. Accessed (25 December 2019)
15. Lui, K., Karmiol, J.: AI infrastructure reference architecture-IBM systems. (June 2018). https://www.ibm.com/downloads/cas/W1JQBNJV. Accessed (25 December 2019)
16. Bormann, C., Ersue, M., Keranen, A.: Terminology for constrained-node networks. (May 2014). https://tools.ietf.org/html/rfc7228
17. Ilapakurti, A., Vuppalapati, J.S., Kedari, S., Kedari, S., Chauhan, C., Vuppalapati, C.: iDispenser—big data enabled intelligent dispenser. In: 2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService) (2017)

18. Han, J., Kamber, M., Pei, J.: Data mining: concepts and techniques. 3 edn, Morgan Kaufmann, (6 July 2011)
19. Rajaraman, A., Ullman, J.D.: Mining of massive datasets. Cambridge University Press (30 December 2011)

Design and Development of System for Post-infection Attack Behavioral Analysis

Toivo Herman Kamati[1,2(✉)], Dharm Singh Jat[2], and Saurabh Chamotra[3]

[1] Office of the Prime Minister, Windhoek, Namibia
Toivo.kamati@opm.gov.na
[2] Namibia University of Science and Technology, Windhoek, Namibia
dsingh@nust.na
[3] Center for Development of Advanced Computing, Mohali, India
saurabhc@cdac.in

Abstract. Attacks targeting government and public networks for getting political, and strategic advantage is a common phenomenon. Most of the time these targeted attacks are state-sponsored and banks heavily on the social engineering techniques to get a successful foothold in the targeted network. As these attacks use a combination of social engineering tactics along with zero-day exploits, their detection using conventional signature-based detection systems is a challenge. The first step towards the mitigation of these targeted attacks is to develop a mechanism to analyze and characterize these attacks. This study proposed a conceptual framework for capturing, collection, and analyses of targeted attacks in an enterprise network scenario. This framework provides a mechanism for capturing and performing post-infection behavioral analysis on the captured adversary.

Keywords: Deception · Targeted attack · Post-infection attack

1 Introduction

As per the threat reports published by security companies, there is a considerable increase in the number of users targeted by the malware [1]. According to the 2019 Internet Security Report, there is an unexpected increase in malware and an unknown attacker steal millions in cryptocurrency [2]. These days' cyber-attacks have used advanced attack techniques to breach security with unknown malicious code which is difficult to detect using traditional firewall and intrusion detection systems [3]. Further, as the services offered through the internet become an integral part of the corporate ecosystem the relevance of the stolen credentials for attackers starts gaining more importance. In earlier incidents most of the time these incidents were to show the superiority or strategic gains but later these hacks were done with a purpose for gaining persistent and escalated access to the victim network. Attackers are using the stolen credentials as a key to

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_55

the enterprise network and it has become a technique to perform the lateral movement in organizational network. Also, this has been observed in the past that cyber-attacks targeting critical organizational networks employ social engineering along with multiple propagation vectors to infiltrate inside the organizational networks. Factors such as use of social engineering techniques as propagation vector, persistent nature, targeting specific functional domains, tendency to remain dormant for a long time, and use of zero-day vulnerability exploits makes capturing and analysis of these targeted attacks a challenge. These attacks use unique attack vectors and custom-built tools tuned for the particular target, making detection very challenging whether either signature or anomaly detection techniques are used. As per recent the threat report of popular technology security company Attivo [4] networks, 9 out of 10 companies have shared that they have been breached in the last three years with 59% stating that at least one of these attacks resulted in physical damage. In addition, the most recent Verizon Data Breach Investigation Report [5] pointed out that currently deployed protection mechanisms are not adequate to address APT malware. The report stated that 56% of the breaches took months or years to which goes undiscovered. Furthermore, 29% of these attacks used stolen credentials to infiltrate the systems. All these instances clearly bring out the fact that conventional preventive security solutions alone are no longer seen as a reliable line of defense against advanced cyber-attacks targeting enterprise networks.

In the work presented in this paper, have proposed a framework that uses deception as a tool for capturing, analysis, and modeling an adversary in an enterprise setup. The current study is focused on the detection of targeted attacks and post-infection attack behavior profiling. During this study systems, software, and models are developed to capture, analyze, visualize and model targeted attacks in an enterprise network. The study has used deception as a tool to deceive and capture the adversary tools and tactics inside the enterprise network. The data hence collected could be used for similarity and connectivity-based grouping with the purpose of developing better attack and malware binary signatures, and hence contributes to the mitigation of such attacks. The aim of this paper is to investigate various methods used by cyber-aggressor to conceal their activities in an enterprise network without being noticed and develop mechanisms to capture and analyze such techniques. Techniques to profile the behavior of a compromised system is required to build correlation models in order to cluster and classify observed network traffic and system events. These observations are derived from the automatic analysis of the behavior of malware captured in the honeynet and security server deployed on the public cloud of the Office of the Prime Minister in Namibia.

1.1 Related Work

Over the last three decades, the concepts of deception experienced rising popularity in information technology. Researchers have come across varied forms of deception implemented at various layers of interaction with the attackers. In their work [6] has presented a survey of such diverse forms of deception envisaged by the researchers. In their work [7] established the need to show how cyber-attacker behave to exploit an information ecosystem. They have shown that a comprehensive outline of observable aggression patterns is a good source of intelligence for setting baselines to discriminate between the cyber-attacker and normal user so as to their respective actions. Reference

[8] shared the experiences from an empirical analysis of cyber-attacker behaviors based on high-interaction honeypot technology, the study focused on observing activities to establish the extent of attack paths after she had gained access through SSH service.

As per the automated attack graph generator models are seldom applied to illustrate the atomic-level vulnerability of a system [9]. In one familiar problem, completion of tedious and erroneous manually drawn vulnerability-based attack graphs, given system vulnerability and connectivity, a detailed attack graph model has yielded a network attack graph structure composed of hosts, network connectivity, trust, services, intrusion detection events, actions, intruder and omitted complication is computed using open source tools [10]. To formulate a model-based attack graph for automatic generation, an attack graph is defined as

$$G = (S, \tau, S_0, S_8)$$

where S is the states, τ is a transition relation, S_0 is a state of initial states and S_8 is a set of success states [11].

The problem of Payload encoding methods (base64), malware polymorphism and self-packers when analyzing malware is solved by dynamic analysis of malware in a sandbox environment, by clustering and classification of observations using Machine Learning techniques, the earlier method is for grouping unknown patterns whereas the later assigns observed patterns into known malware classes [12]. Study had mapped attack events to the cyber kill chain phases in order to analyze malware infections, cyber-attack cycle is grouped into seven phases (i) Reconnaissance (ii) Weaponize (iii) Delivery (iv) Exploitation (v) Installation (vi) Command and Control (vii) Act on Objectives, malware implantation, and communication activities between infected nodes and cyber-attacker with an objective of Data exfiltration and or lateral movement after the payload has been turned-on happens in the last three phases [13].

Designed an instruction detection system based on Bro-Zeek to remedy the security situation at a hospital. Patient records with personal, financial, and diagnostic information were stolen from a health management system [14]. Malicious IP detection, malicious SSL certificate detection, domain flux detection, and TOR connection detection are the composite modules for detecting C & C communication. Trained Random Forest, Support Vector Machine, and Extreme Machine learning to classify communication network intrusion [15]. Random Forest classifiers based on the Recall parameter have outperformed Random Forest based on the F-Measure parameter by 3.5% [16]. Studies proposed using a population-based searching algorithm which reduces computational and storage strain of training data [17, 18].

Propose k-Fold cross-validation resampling procedure to assess the versatility of the Botnet classifier [16]. The randomly selected data sets are clustered into 10 folds, $k = 10$ assuming 10-fold cross-validation, iteratively, the first fold is used for validation and the remaining nine folds are for training the classifier, evaluation completes with quantifying Recall, Precision, F-MeasureBot, and F-Measurenormal related to the aggressive user.

2 Post-infection Behavior Analysis and Problem

Post-infection behavioral analysis system models the behavioral property of a vulnerable target after implanted malware has been activated. According to some existing work, implantation happens once the attacker has installed malware on the target [19]. This study analyzes sequences of events sequence of events in the last phases namely Command & Control and actions of the Lockheed model. Collected malware binaries are executed in the sandbox environment to discover network trace, instruction trace, and system call trace comparable to a technique used in an experiment [20]. The system can be applied to investigate malware-induced network connections and internal system processes execution. This research designs and develops a framework for the systematic study of automatic analysis of malware in reference to the Lockheed's Cyber kill chain model and proposes machine learning model to reveal prolonged and undetected activities of cyber-aggression in targeted network enterprise using statistical algorithms for clustering and classification of anomalous events empirically observed in a honeynet and features extracted from malware simulated in sandbox environment.

Perused the improvement of the detection and prediction of the unknown "zero-day" online fast-flux botnet that occasionally hosts malicious content [21]. Proposed a model for the automatic analysis of malware behavior using machine learning [12]. Research had presented a technical review of Stuxnet, Duqu, Flame, Red October, Mini Duke and, highlighting the common characteristics and techniques of advanced malware and clearly bringing out why the existing security defenses have failed to protect against APT attacks [22].

Due to the number of effective and diverse attack propagation vectors available (i.e., spear phishing, drive-by download), the breaches in enterprise networks perimeter security has become inevitable. Further, the use of unique attack vectors and custom-built tools tuned for the particular target makes the detection of such attacks a challenge for the conventional detection mechanisms [23]. All these instances clearly bring out the fact that conventional preventive security solutions alone are no longer seen as a reliable line of defense against advanced cyber-attacks targeting enterprise networks [24]. As a result, researchers have started exploring approaches such as honeypot technologies for providing real-time alerts of threats, improved incident response, and to mitigate the risks associated with data and employee credential exfiltration [25].

In the current study, investigators envisage addressing the problem of detection of such targeted attacks, by developing a honeypot-based attack capturing and sandbox environment-based analysis framework. The proposed framework as shown in Fig. 1 will perform post-infection attack data analysis on the data collected by the honeypots. Researchers trust that the data captured by such a framework at various layers of abstraction of the organizational network provides sufficient inputs for the detection of the presence of adversary inside the organizational network.

3 Data Gathering

Nebula, a high-interaction honeypot is deployed with HonSSH honeypot software [26]. HonSSH intercepts all the interaction established between a cyber-attacker and the honeypot through SSH secure tunnel, it captures connection attempts, copies files transferred

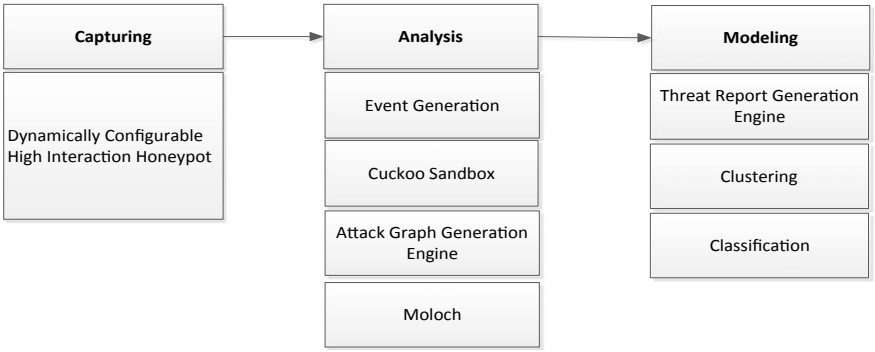


Fig. 1. Proposed conceptual model

with wget or scp for replay during analysis. Nebula server sits between the cyber-attacker and the targeted honeypot shown in Fig. 2.

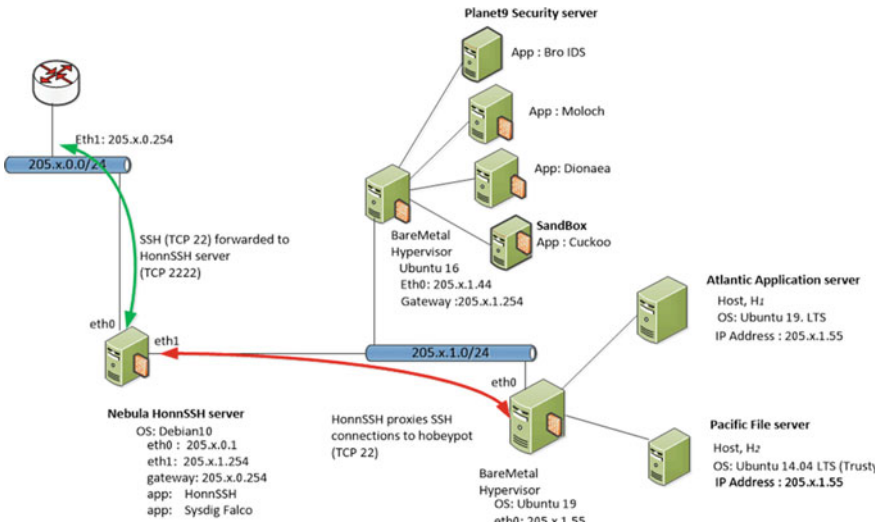


Fig. 2. Experimental network topology

3.1 Planet 9 Security Server

The security server is hypervisor installation with Dionaea honeypot, Moloch, Sandbox, and Bro Intrusion Detection system as guest hosts.

Dionaea

Dionaea is a medium interaction honeypot coded in Python that implements vulnerabilities in Sever Message Block, Domain Name Services, File Transfer Protocol, and SQL database [25].

It uses LibEmu to capture copies of binaries and malware sent by a cyber-attacker onto vulnerable nodes in the enterprise network. In addition, Dionaea honeypot hosts a database that records all accepted and rejected connection flows between nodes under attack and the cyber-attackers system. In this work researchers have deployed Dionaea honeypot to capture copies of malware and binaries implanted on vulnerable nodes, an early test run of Planet 9 security server (beta version) has captured malware shown in Table 1.

Table 1. Binaries and bistreams extracted from planet 9 security server

Dioea	Binary snapshot
1	smbd-445-172.24.48.74-BlvKAZ 2019-09-20 11:06 ... 74-BlVKAZ File 11 600 KB smbd-445-172.24.48.74-CLetEE 2019-09-20 11:06 ... 74-CLETEE File 2 KB smbd-445-172.24.48.74-sckucO 2019-09-20 11:06 ... 74-SCKUCO File 3 KB
2	smbd-445-172.24.48.74-BlvKAZ 2019-09-20 11:06 ... 74-BlVKAZ File 11 600 KB smbd-445-172.24.48.74-CLetEE 2019-09-20 11:06 ... 74-CLETEE File 2 KB smbd-445-172.24.48.74-sckucO 2019-09-20 11:06 ... 74-SCKUCO File 3 KB
3	c0d149a7828c3ad6046da2d897bcff0c 2019-10-28 02:25 ... File 5 145 KB

Randomly selected samples of binaries and bistreams captured by Planet 9 security server-alpha version are as follows:

Moloch is a full packet capture open source tool written by Andy Wick and Eolin Miller for recoding network traffic in PCAP format. The downloaded or imported files are stored in an indexed manner, searches can be done on IP address, GeoIP, and hostname or URL [26].

Cuckoo sandbox

Cuckoo provides a secure sandbox environment for running and executing malware samples collected from the experimental network. Execution steps and Operating System patterns are recoded to re-engineer the purpose of the malware, for example, checking registry keys altered, files accessed, tasks executed by abnormal processes [27].

Bro Intrusion Detection System

Bro intrusion detection is behavior and signature-based IDS capable of analyzing anomaly HTTP traffic, DNS, SSH, and anomaly behaviors [28].

3.2 High-Interaction Honeypot

Primarily, the deployed targets discovering zero-day attacks, denial of service attack, exfiltration methods, covert channel, and Web application exploits. The deployed honeynet is depicted in Fig. 2.

Pacific File server

Pacific handles Secure Shell Socket (SSH) and Server Message Block (SMB) traffic,

it offers a secured remote connection and file sharing services. Reference [8] Remote connection services listen for SSH connection requests and set up an encrypted tunnel for providing a secure remote session to the honeypot from the internet. Whereas, [29] file sharing service provides storage for documents and electronic media (audio, presentation slides, installation images, and movie files) to be shared amongst anonymous and authenticated users connecting via the internet. For experimental setup, the SSH suite is OpenSSH 6.1.1p1, and file sharing suite is SAMBA 4.3.11-ubuntu, which is deployed on an unpatched Ubuntu 14.04 LTS, Trusty Tar.

Atlantic Application server
Dynamic applications honeypots based on Glastopf templates are easily detected by a well-seasoned cyber-attacker because of the manner in which the content is structured [30].

Hence, the Atlantic Application server shall not be based on Glastopf but it will implement a dummy e-Government ticketing management service for managing the request of the citizens. For experimental purposes, it is based on the installation of Apache HTTP server 2.0-ubuntu, PHP 5.0-Ubunut, and Mysql-5.7 are installed. All attack sessions targeting the Atlantic Application and Pacific file server high-interaction honeypot are recorded by Nebula HonSSH server for replay during incremental attack analysis.

Data processing
The gathered network and system will be filtered, compressed, encrypted and logically fused with the Metadata regarding the honeypot and subsequently converted into a relational database format, the structure of the proposed database schema is shown in Fig. 3. This logical correlation of the captured attack data will enable us to trackback the attacks to the vulnerability exploited correlated network traces to traffic sessions indexed in the Moloch database in order to derive details about URLs, ports, source IP geolocation and access to other network resources [28].

In this study, researchers have processed events and will be labeled and stored in a relational database for future analysis by adopting the database structure similar to their schema [29]. The events collected are stored in a centralized database that feeds the analysis module interlinked to a web interface.

4 Analysis

Vulnerability analysis is useful in planning the generation of attack graphs for visualization and analysis of activities of a cyber-attacker in order to draw statistical descriptive of the concerned network and system events, for input into the Machine learning model.

Modeled attack graph structure to simplify the representation of state changes in attacked nodes, the proposed host-based attack graph, and connectivity-based attack graphs are shown in Figs. 4 and 6, respectively, [31].

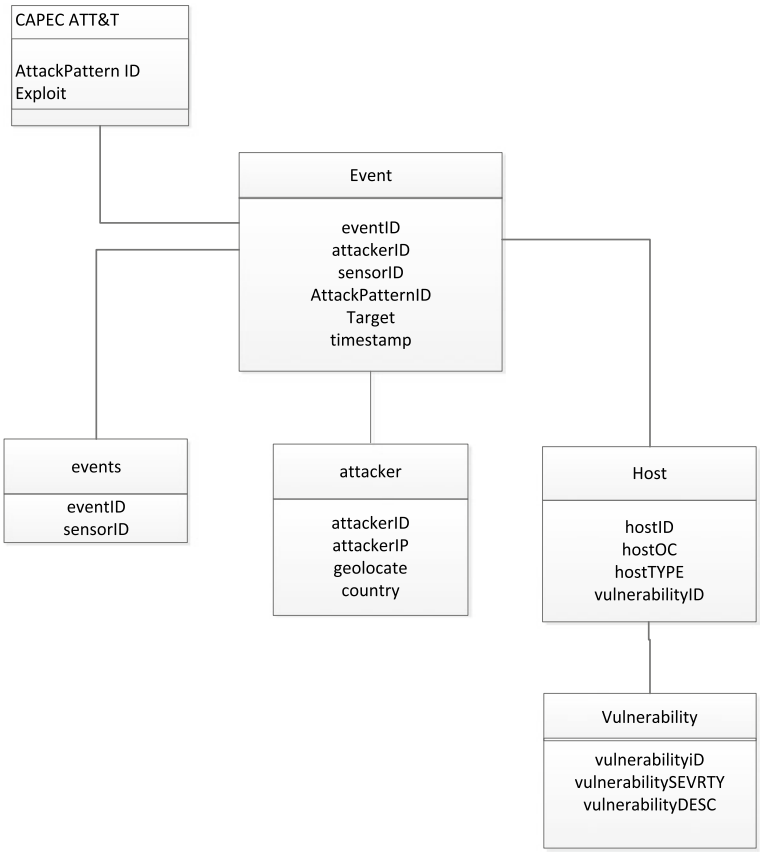


Fig. 3. Generated events archiving respository

5 The Post-infection Learning Model

The conceptual proposed model is based on events generated during the execution of malware samples in Cuckoo sandbox environment, see the proposed system workflow in Fig. 5. Further, this work is to identify cyber-attacks using high-interaction honeypots which intercepts all cyber-attacker interactions and binaries transferred to nodes in the honeynet. Nebula and Planet9 security server stores the intercepted copies of malware, binaries, and data files exfiltrated from the Atlantic Application and Pacific file server.

Cuckoo sandbox execution
Collected malware samples are exported to Cuckoo sandbox for execution, examples of traces produced are TCP, UDP network connection details, and full read, write, delete instruction of system memory, registry, and file. The network and systems events generated are sequenced and indexed using timestamps for export into the centralized database,

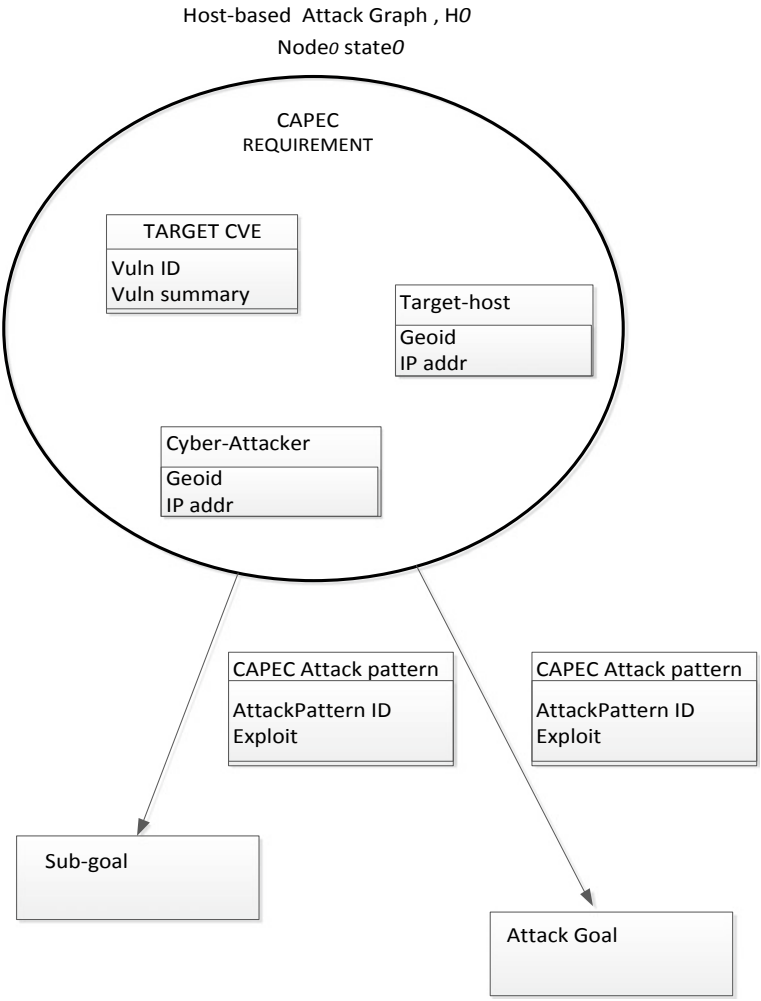


Fig. 4. State-based host attack graph structure

which serves as the repository of checking-in and checking-out events to append new details to the attack graphs corresponding to indexed nodes.

Attack graphs

The network and system events generated are applied to the attack graph generation engine to output host-based attack graph and connectivity-based attack graphs, the structures of the respective graphs are shown in Figs. 4 and 6.

The ultimate goal of the proposed post-infection behavior analysis system is to generate threat reports, documenting the details of the detected targeted attacks in the enterprise network. The statistical algorithms used for training and testing the learning model are support vector machine (SVM), decision tree (DT) and k-nearest neighbors

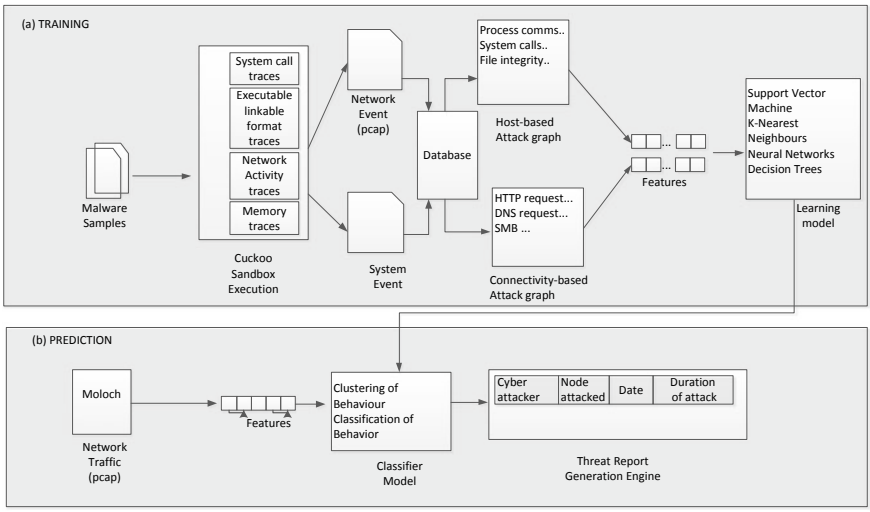


Fig. 5. Post-infection behavior analysis system workflow

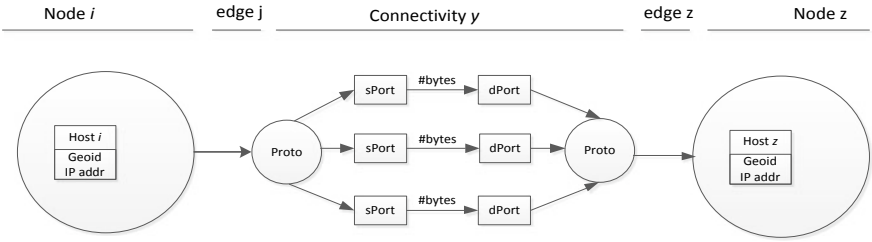


Fig. 6. Connectivity-based attack graph structure

(KNN). The algorithms are optimized to identify anomalous behavior given a pool of features extracted from network activity and internal system events observed in Cuckoo sandbox runs.

6 Conclusion and Discussion

The study elaborated on the design of the multi-layered malware-behavioral analysis system: Capturing attack using high-interaction honeypots, Analyzing with Cuckoo sandbox and Moloch and Modeling using machine learning technique, in order to create threat reports outlining attacks. The reports generated can be used for the detection of targeted attacks and post-infection attack behavior profiling in the Office of the Prime Minister, Namibia. The conceptual proposed system will be infective if this study overlooks Cuckoo sandbox environment vulnerabilities and other related anti-analysis techniques inherent in some malware codes, patching the sandbox environment vulnerabilities hardens the analysis process against evasion during malware analysis runtime.

Regardless of the stealthy nature of some malware, traces of malware behavior resulting from Linux kernel-based approach, Process Control Block based approach, System-call-based approach, Executable Linkable Format based approach and Hybrid-based approach produce sufficient malware-features to optimize clustering and classification algorithms thereby classifying target attack behavior to precision.

Acknowledgements. Researchers would like to thank Department Public Service Information Management, Office of the Prime Minister, Windhoek, Namibia especially to Deputy Executive Director Mr. Erastus Amutenya for providing the infrastructure of the honeynet, the data capturing of this research.

References

1. 2019-crowdstrike-global-threat-report. https://www.crowdstrike.com/resources/reports/2019-crowdstrike-global-threat-report/, last accessed 2020/01/29
2. Internet Security Report—Q1 2019: http://watchguard.com/wgrd-resource-center/security-report-q1-2019, last accessed 2020/01/29
3. Seo, J., Lee, S.: Abnormal detection to identify infected systems using the APChain algorithm and behavioural profiling. Secur. Commun. Netw. (2018). https://doi.org/10.1155/2018/9706706
4. Attivo Report-2019: https://attivonetworks.com/fortinet-fabric-ready-partner-spotlight-attivo-networks/, last accessed 2020/01/29
5. DIBR 2019: https://www.cs.columbia.edu/~smb/classes/f19/Files/(3)_2019_Verizon_DBIR.pdf, last accessed 2020/01/29
6. Fraunholz, D., Anton, S.D., Lipps, C., Reti, D., Krohmer, D., Pohl, F., Tammen, M., Schotten, H.D.: Demystifying Deception Technology: A Survey (2019)
7. Mallikarjunam, M.N., Prabavathy, S., Sundarakantham, K., Shalinie, S.: Model for cyber attacker behavioural analysis. In: 2015 IEEE Workshop on Computational Intelligence, 2016
8. Nicomette, V., Kaaniche, M., Alata, E., Herrb, M.: Set-up and deployment of a high-interaction honeypot: experiment and lessons learned. J. Comput. Virol. **7**(2), 143–157 (2011)
9. Ghazo, A.T., Mariam, I., Hao, R., Kumar, R.: A2G2V: automated attack graph generator and visualizer. In: The 1st ACM MobiHoc Workshop, vol. 1, pp. 1–4, 2018
10. Sheyner, O., Wing, J.: Tools for generating and analyzing attack graphs. In: Proceedings of Formal Methods for Components and Objects, Lecture Notes in Computer Science, pp. 344–3713 (2004)
11. Sheyner, O., Haines, J., Jha, S., Lippmann, R., Wing, J.M.: Automated generation and analysis of attack graphs. In: Conference: Security and Privacy, 2002. Proceedings. (2002) IEEE Symposium, 2002
12. Rieck, K., Willems, C., Holz, T.: Automatic analysis of malware behavior using machine learning. J. Comput. Secur. **19**(4), 639–668 (2011)
13. Yadav, T., Mallari, R.A.: Technical Aspect of Cyber Kill Chain, pp. 1–7 (2016)
14. Ghafir, I., Prenosil, V., Hammoudeh, M., Baker, T., Jabbar, S., Khalid, S., Jaf, S.: Botdet: A System for Real Time Botnet Command and Control Traffic Detection. pp. 38947–38953 (2018)
15. Ahmad, I., Basheri, M., Iqbal, M.J., Rahim, A.: Performance comparison of support vector machine, random forest and extreme learning machine for intrusion detection. IEEE Access **6**, 33791–33794 (2018)

16. Beiknejad, H., Vahdat-Nejad, H., Moodi, H.: P2P Botnet detection based on traffic behaviour analysis and classification **6**(1), 4–10 (2018)
17. Aslan, M., Beskirli, M., Kodaz, H., Kıran, M.S.: An improved tree seed algorithm for optimization problems. Int. J. Mach. Learn. Comput. **8**(1), 20–22 (2018)
18. Ahmad, I., Basheri, M., Iqbal, M.J., Rahim, A.: Performance comparison of support vector machine, random forest and extreme learning machine for intrusion detection. IEEE Access **6**, 33791–33794 (2018)
19. Hoffman, R.: In: 2019 International Conference on Military Communications and Information Sytems (ICMCIS). Markov Models of Cyber Kill Chains with Iterations, pp. 1–5, 2019
20. Cali, L., Torres, R., Modelo-Howard, G., Tongaonkar, A., Jha, S.: In: 2017 IEEE Conference on Computer Communications (IEEE INFOCOM 2017). Botnet Protocol Inference in the Presence of Encrypted Traffic, 2017
21. Almomani, A.: Fast-flux hunter: a system for filtering online fast -flux botnet Neural. Neural Comput. Appl. **29**(7), 483–493 (2018)
22. Virvilis, N., Tsalis, N., Mylonas A., Gritzalis, D.: Mobile devices: a phisher's paradise. In: Proceedings of the 11th International Conference on Security and Cryptography (SECRYPT-2014), pp. 79–87, ScitePress, Austria, 2014
23. Yadav, T., Mallari, R.A.: Technical Aspect of Cyber Kill Chain, pp. 1–7. https://doi.org/10.1007/978-3-319-22915-7_40 (2016)
24. Lin, H., Yan, Z., Chen, Y., Zhang, L.: A survey on network security-related data collection technologies. In: IEEE (February 2017), pp. 1, 9–10, pp. 17–19, 2017
25. Ali, P.D., Kumar, T.G., Malware capturing and detection in dionaea. In: Innovations in Power and Advanced Computing Technologies (i-PACT-2017), pp. 2–3, 2017
26. Uramova, J., Segec, P., Moravcik, M., Papan, J., Mokos, T., Brodec, M.: Packet capture infrastructure based on Moloch. In: Proceedings of the of the 15th International Conference on Emerging eLearning Technologies And Applications (ICETA-2017), pp. 1–3, 2017
27. Jamalpur, S., Navya, Y., Raja, P., Tagore, G., Rao, G.R.K.: Dynamic Malware analysis using Cuckoo Sandbox. In: Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT-2018), pp. 1057–1059, 2018
28. Demertzis, K.: The Bro Intrusion Detection System, pp. 3–5 (2018)
29. Guoan, X.: Install and Configure Samba Server on Ubuntu 16.04 for File Sharing. Retrieved from https://linuxbabe.com/ubuntu/install-samba-server-ubunut-16-04
30. Sokol, P., Pekarcik, P., Bajtos, T.: Data collection and Data Analysis in Honeypots and Honeynets, pp. 5–8, 2015
31. Mphango, B.: Deception in web application honeypots: case of Glastopf. Int. J. Cyber-Secur. Digit. Forensics (IJCSDF) **6**(4), 179–180, 2017
32. Nicholson, T.: HonSSH. https://github.com/tnich/honssh, last accessed 2020/01/29
33. Kaynar, K.: A Taxonomy for attack graph generation and usage in network security. J. Inf. Secur. Appl. (2016), pp. 3–5, 9–12

Multi-threaded Multilayer Neural Network for Character Recognition

Arun Vaishnav[1(✉)], Manju Mandot[2], and Priyanka Soni[2]

[1] Department of Computer Science, Mohanlal Sukhadia University, Udaipur, Rajasthan 313001, India
arun.vaishnav@gmail.com

[2] Department of Computer Science and IT, Janardan Rai Nagar Rajasthan Vidhyapeeth (Deemed-to-Be University), Udaipur, Rajasthan 313001, India
manju.mandot@gmail.com, priyanka12kimmi@gmail.com

Abstract. New techniques are evolving day by day to recognize characters in any image, text, cards, etc. Optical character recognition is a well-accepted technique to recognize characters in scene or optically captured image. In this present work, we have evolved new method which employs feed forward back propagation (FFPB) technique interlaced with multi-threaded multilayer perceptron neural network (MTMLNN) to recognize the characters with speedup in operations and better accuracy in character recognition. In present methodology, we have employed structured similarity index measure (SSIM) in MTMLNN to compare image and sigmoid function for neuron activation with stochastic gradient descent method for optimization, to train the neural network to recognize the input character with better accuracy.

Keywords: Feed forward back propagation (FFBP) · Multi-threaded multilayer perceptron neural network (MTMLNN) · Optical character recognition (OCR) · Structured similarity index measure (SSIM)

1 Introduction

In computer vision, character and object recognition plays vital role. Importance of character extraction and recognition from the images has been receiving more attention in recent years [1–4]. Various techniques and methods have been working in this area to recognize characters or objects efficiently with précised accuracy. Character recognition is generally done by optical character recognition (OCR), which means it uses optical medium to read the characters. In general, image is captured through the camera device and applied image processing methods to recognize correct pattern in an image. Several applications are working on OCR. Several techniques have evolved to recognize the character with better accuracy such as template matching (TM), support vector machine (SVM), and neural network (NN). One of the novel approaches to recognize the characters is NN.

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_56

Artificial intelligence (AI) is playing an important role to recognize pattern using neural networks in which neuron is trained from the available dataset of images, and then, it will identify the input pattern. As the number of layers of neurons increases, the network will move toward the more accuracy. Neural network working is similar to our brains neurons using the concepts of multi-threading in neural network which will improve the time limit and utilization of cores in available system. In the present work, multi-threading is used in multilayer neural network and trained the neurons to identify the given pattern in minimum time with better accuracy. We have discussed important literature to better understanding as per the available and invented methodologies.

2 Work Done

Template matching (TM) is the simplest technique that is used to recognize optical character. For finding the similarity between the images, correlation coefficient method is employed. Vaishnav and Mandot [5] have proposed modified method using TM and applied on vehicle number plate to recognize the vehicle license number. They have also analyzed necessary environment for their methodology. All the above-published researches take lot of time to process the image and also require suitable environmental conditions and proper structural information to recognize correct characters, otherwise lead to wrong outputs. To improve the speed and accuracy, we move toward parallelization in image processing, and currently neural networks are in extensive use. A plenty of work has been done on neural networks (NN), although many great deeds have already been done to implement NN in character recognition. Rajavelu et al. [6] had proposed a layered feed forward neural network with one hidden layer, which enormously reduces image processing time while maintaining efficiency and versatility. Zhu and Wu [7] have been proposed handwritten digital recognition system basically based on multilevel quantum neural network (QNN). In this, QNN consists of two sub-classifier layers, first is coarse classifier and subsequent layer is QNN numeral pair classifier.

Masood et al. [8] have been presented work for detection and recognition characters of the license number plate through a sequence of deep convolution neural network (CNN) in an end-to-end system capable of detecting and recognizing license plates with low error rates and robust toward different variations in conditions and license plate templates. Authors have pipelined the complete task in three stages by training three separate deep CNN's: license plate detection, character detection, and character recognition. They compared their results against open automatic license plate recognition (ALPR) open source [9] and open ALPR cloud API [10] and proved that their work on open ALPR sighthound's method is a single unified model and good for license plates for USA and EUROPE regions.

Sahare and Dhok [11] have been presented a robust algorithm for segmenting and recognizing the multilingual Indian document images of Latin and Devnagari scripts. In their proposed work, primary segmentation is done by structure formation or graph distance theory. And segmentation results are validated by support vector machine (SVM) classifiers. In their proposed algorithm, three new geometrical structure-based features are calculated. First two features are formed for the center pixel of given character, and for the third feature, neighborhood information of text pixel is employed. They have used

K-nearest classifier to recognize the input character. The authors have proved that their proposed algorithm has better performance compared to other algorithms with highest segmentation and recognition rate of 98.86% and 99.84%, respectively. Vaishnav et al. [12] have been recently published work through SVM for identifying the characters of license number plate and reported better accuracy of their methodology.

Wu et al. [13] have worked for improvement of Chinese handwriting recognition. For this, they have compared feed forward neural network language model (FNNLM) and recurrent neural network language model (RNNLM) combined with back-off language model to construct hybrid language model. They have proved that NN language model (LM) improves the recognition rate, and hybrid RNNLM outperforms than the other language model.

Abobakr et al. [14] have been proposed a new nonlinear structured similarity index measures (SSIM) layer to the deep CNN, which mimics the function of human visual system (HVS) by focusing on the learning structural properties of its input. In its core computation, the structural similarity metric is evaluated and allows the kernel to learn to compare structural information. They conducted their experiment on CIFA-10 benchmark and showed the better convergence of SSIM layer than the traditional convolutional layer.

Yin et al. [15] used deep CNN for traffic sign recognition. In their proposed neural network, they have used ten layers, where first seven layers are alternate convolutional layers and block layers and the other three are fully connected layers. They have implemented scaled exponential linear units (SELU) as an activation function, having self normalizing properties. They have validated their work on German traffic sign recognition benchmark (GTSRB) and achieved 99.67% accuracy.

Above available literature shows the importance of the character recognition in different computer vision applications using various techniques like TM, SVM, neural network, and CNN. The purpose of present paper is manifold to develop a technique for character recognition using FFBP method to train our multi-threaded, multilayer perceptron neural network (MTMLNN) to correctly recognize the characters.

Present technique explores the possibility of character recognition technique to identify alphanumeric characters. To minimize the timing for character recognition, we use two threads as per the number of cores in the system which will work in parallel. The newness of the present methodology is that we have considered structural similarity of characters by making their pairs and created two threads, each will process horizontally splitted half part of an input image to work in parallel, with neural network.

3 Present Methodology

In the current work, we have used FFBP neural network interlaced with multi-threaded, MTMLNN to speed up the character recognition process with better accuracy in recognition rate.

MTMLNN consists of basically three layers namely (a) input layer, (b) hidden layer, and (c) output layer as shown in Fig. 1.

The present work is divided into majorly four stages to recognize character as shown below:

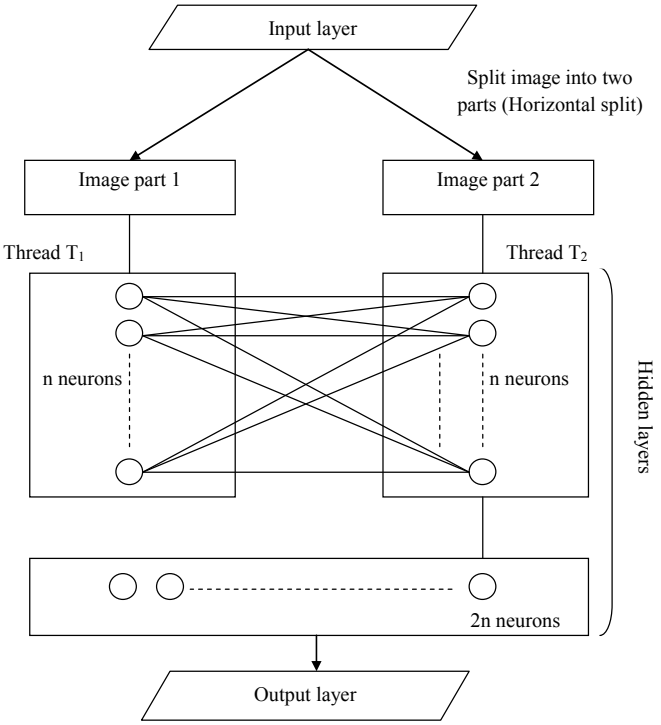


Fig. 1. MTMLNN architecture

Stage 1: Horizontal splitting (Input Layer)

We horizontally split our input image into two parts as per the number of cores in our computer system. It is to get their best utilization of processing power of each core (processor). We can increase the number of subparts of an image depending upon the availability of cores on working system.

Stage 2 and 3: Parallelization process (Hidden SSIM Layer)

In present architecture, we have used structural data parallelization technique. For this, we have created two threads, each one of is dedicated to a subpart of an input image. In proposed architecture, hidden SSIM layer has basically three layers. Two of among all the three are considered as threads, and third layer compiles the outputs of these two SSIM layers. Each thread is constructed of identical neural network having '*n*' number of neurons as we have constructed '*n*' pairs from the character set considering the similarity between their formation patterns by statistics. Pairs of the characters are given below:

Pairs: (B, D), (B, P), (B, R), (C,G), (C, 6), (E, F), (G, 6), (I, J), (I, L), (I, 1) (J,1) (K, R), (L, 1), (M, N), (O, Q), (P, R), (U, V), (U, W), (V, W), (X, Y), (Z, 2), etc.

Due to the structural similarity of O and 0, they are considered as same. If pair is not possible for any character, then we use itself as a second element of pair, and we compare it from itself only like as (A, A).

Stage 4: Output layer

Output layer is to show recognized character with accuracy as an output.

4 Detailed Description of Architecture

In the present architecture, image is divided into two equal parts horizontally using matrix formulation and assigns each part to an individual core to perform processing in parallel fashion to reduce processing time.

To compare input image subpart within the given pairs, two approaches are available, i.e., mean squared error (MSE) and structured similarity index (SSIM) (2018).

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left[o(i,j) - t(i,j) \right]^2 \tag{1}$$

where m and n are the number of pixels, o means observing image and t means trained image in (1).

SSIM is calculated between two images o and t of common size 14 * 28 as shown in (2).

$$\text{SSIM}(o,t) = \frac{(2\mu_o\mu_t + c_1)(2\sigma_{ot} + c_2)}{(\mu_o^2\mu_t^2 + c_1)(\sigma_o^2 + \sigma_t^2 + c_2)} \tag{2}$$

where

μ_o is the average of observing image, and μ_t is the average of trained image,
σ_o^2 is the variance of observing image, and σ_t^2 is variance of trained image,
σ_{ot} is the covariance of observing and trained image, and
c_1 and c_2 are variables used to stabilize the division when there is weak denominator.

MSE is worked by calculating mean square error between each pixels of comparing two images. But the issue with MSE is that it tends to have randomly high numbers, due to which it becomes hard to standardize and if the MSE differs randomly for given image, it will be harder to tell anything. Apart from this, SSIM is worked on similarities between pixels in comparing images. SSIM puts everything on scale of -1 to 1. 1 means that the images are very similar, and -1 means that they are very different. The SSIM metric is nonlinear, and its operations are differentiable, fulfilling the requirements for back propagation and descent gradient-based optimization.

$$u_i^m = \sum_{j=1}^{n} x_j^{m-1} w_{ij}^m + t_i^m \tag{3}$$

where x_j^{m-1} is weight of jth neuron of $(m-1)th$ layer, calculated by sigmoid function [16],

w_{ij}^{m} is weight of connecting edge of *ith* neuron to *jth* neuron of *mth* layer, and t_i^m is SSIM value calculated for lower half.

And it emits output

$$x_i^m = f(u_i^m) \tag{4}$$

where f is sigmoid function used for neuron activation, $f = \frac{1}{1+e^{-x}}$.

Finally, for the optimization in back propagation, stochastic gradient method is used.

$$Q(w) = \frac{1}{n}\sum_{i=1}^{n} Q_i(w) \tag{5}$$

In Eq. 5, $Q(w)$ is weight of connecting edge.

$$w = w - \eta\nabla Q = w - \eta\sum_{i=1}^{n}\frac{\nabla Q_i(w)}{n} \tag{6}$$

In (6), η is learning rate and used for weight adjustment.

In this, (5) and (6) are used for calculating the weights of nodes and adjusting them by using stochastic gradient descent (SGD) method [17, 18].

We can easily prove that multi-threading used in multi-core systems can increase the speed of character recognition, by using speedup, Amdhal's law [19] as follows:

$$S(n) = \frac{1}{(1-p) + \frac{p}{n}} \tag{7}$$

In (7), $S(n)$ is theoretical speedup, p the fraction of the algorithm which is parallel (in our case it is ½) and n the number of C.P.U. threads (in our case it is 2)

$$S(n) = 1.33 \tag{8}$$

In the above equations, we have proved that when multi-threading is used to implement parallelization using multiple cores (here only 2 cores), speedup in operations will be achieved, and in our case, speed is up by 1.33 as shown in (8).

In Fig. 1, thread T_1 (HL1), thread T_2 (HL2), and hidden layer 3 (HL_3) use SSIM metric as a computational layer in this MTMLNN.

5 Working Example of Proposed Methodology

Input images are of 28*28 dimension and are horizontally split into two parts of 14*28 and 14*28 dimensions. Among two threads, each thread having n neurons is used one of half part of the image to compare using SSIM. In this way, complete image is to be processed in just half amount of time using threads. Calculated weight of nodes from one layer is passed to another layer by sigmoid activation function and back propagation stochastic gradient descent (SGD) method with number of iterations in our training, and it minimizes the error and recognizes correct image with better accuracy, with learning rate 0.01.

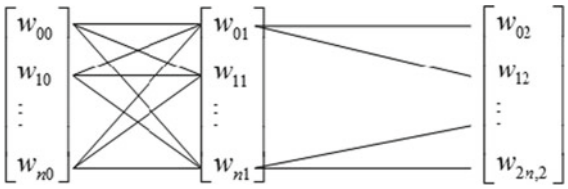
Stage 1:

$$\begin{pmatrix} a_{0,0} & \cdots & a_{0,28} \\ \vdots & \ddots & \vdots \\ a_{28,0} & \cdots & a_{28,28} \end{pmatrix} \tag{9}$$

Stage 2:

$$\begin{pmatrix} a_{0,0} & \cdots & a_{0,28} \\ \vdots & \ddots & \vdots \\ a_{14,0} & \cdots & a_{14,28} \end{pmatrix} \begin{pmatrix} a_{15,0} & \cdots & a_{15,28} \\ \vdots & \ddots & \vdots \\ a_{28,0} & \cdots & a_{28,28} \end{pmatrix} \tag{10}$$

Stage 3: Weight matrix



(11)

Weight will be adjusted by stochastic gradient descent method.

Recognition process of 'B':

As shown in Fig. 2, the input image of character 'B' in 28*28 dimension as (9) will split the image 'B' in two parts 14*28 and 14*28 dimensions as (10). Upper half is straight line and vertical arc, and similarly, lower half also contains straight line and vertical arc. Assign each half part to an individual core which will process that part of an input image using thread. Each thread consists of n number of neurons, where each neuron is separately dedicated to do comparison in a pair, formed by considering the structural similarity between two characters. While considering above example, upper half of character 'B' is similar to pairs (B, P), (P, R), and lower half is only within part of (B, P). Weight of each neuron will be calculated using SSIM and propagate it to next layer using sigmoid function [16] as an activation function for neuron and weights will be adjusted by SGD method of back propagation [17, 18] as (11). Finally, after adjusting the weights from HL_1, HL_2, and HL_3 in output layer, only 'B' will be shown as a final output.

6 Conclusion and Future Scope

In the present investigation, we have proposed an efficient technique of character recognition using FFBP with MTMLNN, leading to higher degree of accuracy in character recognition within better time limit. We have divided MTMLNN into four stages. In the first stage, we horizontally split the image to use structural data parallelism by using

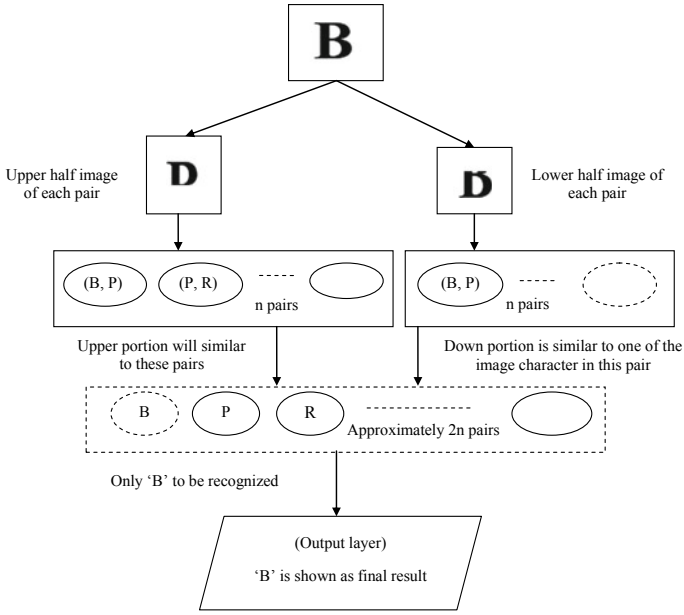


Fig. 2. Working example of MTMLNN in case of 'B'

threads. In the next two stages, SSIM layers are used to calculate weights of nodes and adjusted them using SGD method to train the neural network. And the last stage is our final output layer which shows the final image as a correctly recognized character, using MTMLNN. It is seen that MTMLNN is a multilayer perceptron NN interlaced with multithreading, which is a supervised machine learning technique that leads to better time limit, and for two cores we have achieved 1.33 speed up by original speed, and MTMLNN leads to remarkable results. We are working on the implementation of this proposed algorithm and will discuss in our future research work with proper proofs.

References

1. Lucas, S., Panaretos, A., Sosa, L., Tang, A., Wong, S., Young, R.: ICDAR 2003 robust reading competitions. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR'03), IEEE, pp. 682–687 (2003)
2. Lucas, S.M.: ICDAR 2005 text locating competition results. In: Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05), IEEE, pp. 80–84 (2005)
3. Lucas, S.M., Panaretos, A., Sosa, L., Tang, A., Wong, S., Young, R., Ashida, K., Nagai, H., Okamoto, M., Yamamoto, H., Miyao, H., Zhu, J., Ou, W., Wolf, C., Jolion, J., Todoran, L., Worring, M., Lin, X.: ICDAR 2003 robust reading competitions: entries, results, and future directions. IJDAR **7**, 105–122 (2005)
4. Shahab, A., Shafait, F., Dengel, A.: ICDAR 2011 robust reading competition challenge 2: reading text in scene images. In: International Conference on Document Analysis and Recognition, IEEE, pp. 1491–1496 (2011)

5. Vaishnav, A., Mandot, M.: Template matching for automatic number plate recognition system with optical character recognition. In: Information and Communication Technology for sustainable Development, Advances in Intelligent Systems and Computing, pp. 683–694. Springer (2018)
6. Rajavelu, A., Musavi, M.T., Shirvaikar, M.V.: A neural network approach to character recognition. Neural Networks **2**, 387–393 (1989)
7. Zhu, D., Wu, R.: A multi-layer quantum neural networks recognition system for handwritten digital recognition. In: Third International Conference on Natural Computation, IEEE (2007)
8. Masood, S. Z., Shu, G., Dehghan, A. and Ortiz, E. G., License Plate Detection and Recognition using Deeply Learned Convolutional Neural Networks, *arXiv* (2017)
9. www.github.com/openalpr/openalpr
10. www.openalpr.com/cloud-api.html
11. Sahare, P., Dhok, S.B.: Multilingual character segmentation and recognition schemes for indian document images. IEEE Access **6**, 106003–106017 (2018)
12. Vaishnav, A., Ahuja, B. L., Mandot, M.: Feasibility of efficient number plate recognition using morphological dilation and support vector machine. Inter. J. Recent Technol. Eng. **8**(3), (2019)
13. Wu, Y.C., Yin, F., Liu, C.L: Improving handwrittern chinese text recognition using neural network language models and convolutional neural network shape models. In: Pattern Recognition, pp. 1–48 (2016)
14. Abobakr, A., Hossny, M., Nahavandi, S.: SSIMLayer: towards robust deep representation learning via nonlinear structural similarity, pp. 1–11. arXiv (2018)
15. Yin, S.H., Zhang, J.C., Du, J.Y.: Traffic sign recognition based on deep convolutional neural network. Optoelectron. Lett. **13**(6), 0476–480 (2017)
16. Jamel, T.M., Khammas, B.M.: Implementation of a sigmoid activation function for neural network using FPGA. In: 13th Scientific Conference of AI-Ma'moon University college, vol. 13 (2012)
17. Bottou, L.: Stochastic gradient descent tricks. In: Neural Networks: Tricks of the Trade, pp. 421–436 (2012)
18. https://en.wikipedia.org/wiki/Stochastic_gradient_descent
19. Hill M.D., Marty, M.R.: Amdahl's law in the multicore era. IEEE Computer Society, pp. 33–38 (2008)

Threat Detection in Social Media Images Using the Inception-v3 Model

Shatha AbdulAziz AlAjlan(✉) and Abdul Khader Jilani Saudagar

Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Kingdom of Saudi Arabia
{samalajlan,aksaudagar}@imamu.edu.sa

Abstract. Threat detection in images within social media content has become an important aspect of content monitoring. This task can be achieved using various image object detection and classification techniques. Recently, object detection has become an important task of machine learning, with significant studies dedicated to using deep learning techniques, especially the convolutional neural network (CNN), in the computer vision field. The current study involves an experiment using transfer learning technology to retrain the Inception-v3 model from TensorFlow in terms of the collected dataset related to known threat content. The novelty of this research work lies in the threat detection of images shared on social media which was not addressed before. The model achieves a high accuracy of around 96% in threat detection. The results of this research will be helpful in monitoring and tracking social media image content in terms of the detection of threats in the images shared among users, while the system can be used as a standalone system or as part of larger systems.

Keywords: Detection · CNN · Images · Threat · Transfer learning · Inception-v3

1 Introduction

Recently, the world has witnessed a significant evolution in social interaction on the Internet through what is generally known as "social media." Social media has become a means of communicating using media content such as images, text and videos. People upload and share billions of images on social media every day and these images have the power to reveal a great deal of information about what people are sharing. Computers and mobile phones deal with images as a set of pixels that must be classified into objects and then transformed into real meanings. Nowadays, image classification plays a key role in multidisciplinary applications [6] like image retrieval, biomedical imaging, brand marketing, video surveillance, etc. "Threat" is one of the negative content categories on social media posts. Threats can be very harmful and inappropriate, especially in terms of the safety of teenagers and children, and of society as whole. No country, including the Arab countries, is willing to accept any kind of activity that threatens society or national

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7_57

security. The threat related to social media images may become a further source of threats that could endanger the national security of any country or the safety of teenagers or children in their homes. In this research, the term "threat" refers to different kinds of weapons or drugs. The aim of the research is to detect these threats within social media images using deep learning techniques.

In the late twentieth century, the deep learning neural network [8] has received a great deal of attention. Deep learning refers to machine learning models where many levels of representation are used to accurately understand various data (e.g., texts, images and sound) [5]. One of the most commonly used deep learning neural networks is the convolutional neural network (CNN). In fact, CNN is the most well-known deep learning model that is used in the computer vision field, especially in terms of image classification and object detection [10]. There are many CNN image classification architectures that won in image classification competition (ILSVRC12) [2], including AlexNet, VGGNet and GoogLeNet (Inception model). The components of CNNs include input layer, output layer and one or more hidden layers. The difference between a convolutional network and a regular neural network is that the layers are arranged in three dimensions (width, depth and height), which allows CNNs to train various layers with nonlinear mapping to classify multiple dimensional inputs into sets of classes at the output layer. Compared to other standard feedforward neural networks with the same sized layers, CNNs have less parameters and connections, which makes them easy to train and learn [7]. CNNs are very effective when applied to large datasets. However, the lack of available threat image datasets is the main problem in training a CNN. The use of a transfer learning technique [9] through a pre-trained model can help to avoid this problem.

In this research, the Inception-v3 model is trained using the transfer learning technique on the TensorFlow platform. This is an effective model that can be implemented to detect threats in images based on collected datasets, which involves a short training time and results in high accuracy. It could be used as a threat detection system for social media images as a standalone system or as part of a larger, hybrid solution.

The research paper is organized as follows. Section 2 discusses the work related to image object detection before Sect. 3 presents and explains the methodology and the experiment design. Then, the experiment results are shown in Sect. 4. Finally, the conclusion and the recommendations for future work are presented in Sect. 5.

2 Related Work

Objects in realistic settings have considerable variability, which means in order to learn how to classify them, it is necessary to use a large training dataset. One of the most recent larger datasets is the ImageNet dataset [3]. This is an image dataset organized based on the WordNet hierarchy (only the nouns) in such a way that each node of the hierarchy is represented by hundreds or thousands of images. There are many CNN image classification architectures, including AlexNet, VGGNet and GoogLeNet. These architectures were first trained in relation to ImageNet.

2.1 AlexNet

The study in [7] was the first to present a CNN for image classification. Here, the authors trained a large CNN in terms of a subset of ImageNet, achieving good results compared to other studies of that time. Their final network contained five convolutional layers followed by three fully connected layers. The final error rate was 15.3% and the number of parameters was around 60 million. However, their network is too big and is limited due to (a) the amount of memory in the available GPUs of that time (even if they split the training across two GPUs) and (b) the length of the training time, which amounts to five to six days with their GPUs.

2.2 VGGNet

This model is a deep convolutional network that uses up to 19 weighted layers and which can be used for large-scale image classification tasks. The VGG stack of convolutional layers (up to 19 layers) is followed by three fully connected layers. Much like AlexNet, VGGNet was trained in terms of the ImageNet dataset. The number of parameters is 138 million and the model achieves an error rate of 7.3% [11].

2.3 GoogleNet (Inception Model)

The Inception model is one of the CNN models that is used for image classification and one that has a low error rate and a very deep network compared to AlexNet and VGGNet. It was designed to perform efficiently, even when under strict constraints in terms of memory and computational budget. GoogLeNet (using the Inception model) was developed by Google and has a 6.67% error rate and employs only five million parameters, which is 12 times less than AlexNet and nearly 30 times less than VGGNet. This lower number of parameters allows for high quality training in terms of various different training sets [12].

Table 1 shows the key figures related to the training of these networks in terms of the ImageNet dataset, while the results may vary based on the dataset.

Table 1. Image classification CNNs based on ImageNet dataset

CNN	Year	Developed by	Top-5 error rate (%)	No. of parameters
AlexNet	2012	Krizhevsky, Sutskever and Hinton	15.3	61 million
VGGNet	2014	Simonyan and Zisserman	7.3	138 million
GoogLeNet (Inception-v3)	2015	Szegedy, Vanhoucke, Loffe, Shlens and Wojna	6.67	5 million

2.4 Transfer Learning

Image classification and recognition models have millions of parameters. In order to train them from scratch, a huge of computing power and labeled data are required. Transfer learning is a technique that makes use of the advantage of the pre-trained models for a related task and attains the aspect of reusability. Many studies have used this technique to train the model in terms of their own datasets related to specific categories. The need for transfer learning [13] relates to the lack of target training datasets, where the data is either rare, expensive to collect or inaccessible. In recent times, transfer learning is being widely used in numerous deep learning applications. TensorFlow provides support in terms of many tutorials on retraining the final layer for new classes using transfer learning [14].

3 Construction of Experiment

The experiment (see Fig. 1) involves the detection of threats using the Inception-v3 model from the TensorFlow platform. TensorFlow is an open-source machine learning platform released by Google that was developed to support deep learning algorithms through experimenting with new models, training them in terms of specific datasets and moving them into production. TensorFlow assists in ensuring rapid training and runs trained models on different platforms. TensorFlow [1] is flexible enough to support the research and experimentation involving new models.

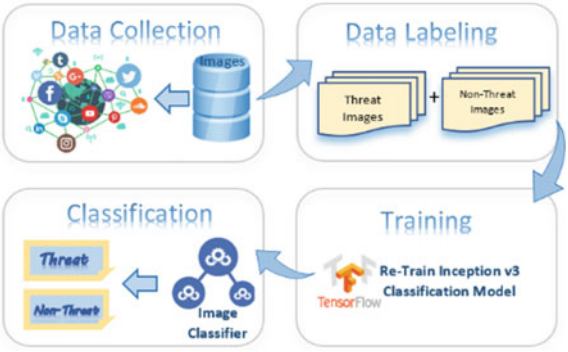


Fig. 1. Workflow of research experiment

3.1 Data Collection

Threat and non-threat dataset image examples are shown in Fig. 2. The dataset collected and labeled. It includes more than 1000 images, with 600 being threat images and the remaining non-threat images.

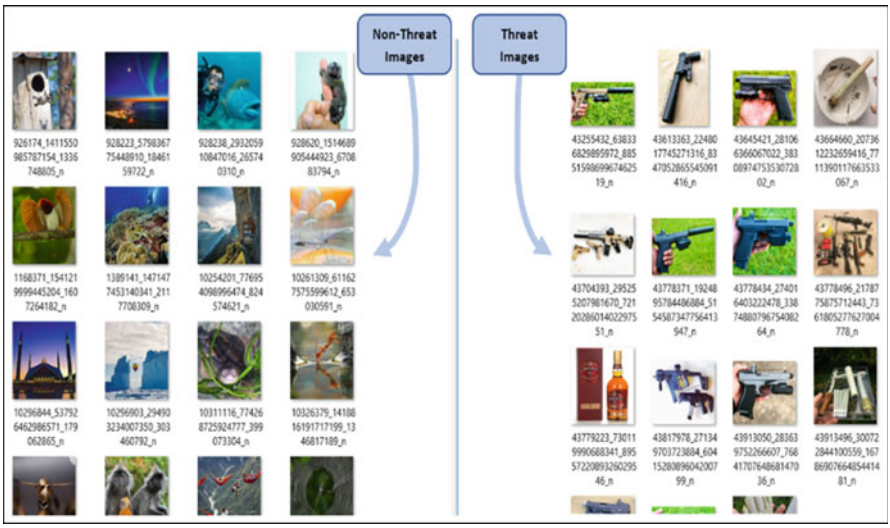


Fig. 2. Threat/non-threat image dataset

3.2 Training

The Inception-v3 model [12] is a deep CNN that was pre-trained on an ImageNet dataset that contains more than 1000 different classes and more than 1.2 million images. It is a generic model that can classify images in terms of the pre-trained 1000 classes.

As noted above, many studies have used the technique of transfer learning to train a model in terms of their own datasets related to specific categories [14]. The transfer learning method based on CNN model is shown in Fig. 3.

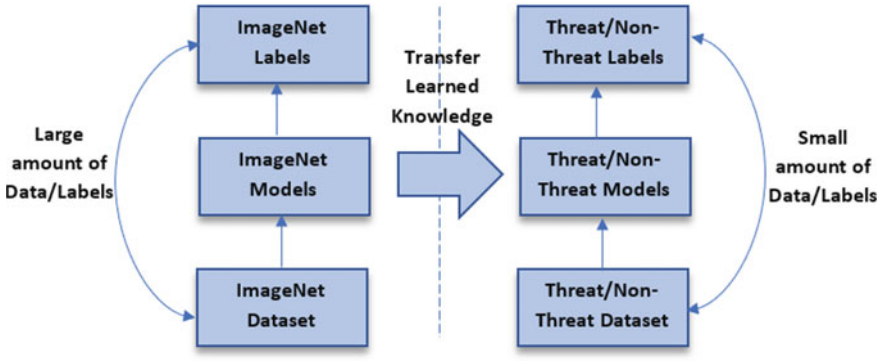


Fig. 3. Transfer learning based on CNN model [5]

With limited computation power, it is difficult for researchers to train the model from scratch [15]. TensorFlow provides support in terms of many tutorials on retraining the final layer for new classes using transfer learning. Here, the threat dataset is used as

input to retrain the new final layer, where the number of nodes will be changed to two (two classes): "threat" and "non-threat."

4 Experiment Results

The dataset used during the retraining was separated in terms of training and validation by cross-validation. Figure 4 shows the accuracy achieved during retraining based on the dataset, while Fig. 5 shows the cross-entropy loss during retraining based on the collected dataset. Validation accuracy is the percentage of the correctly detected images from a random selection that was not in the original training dataset. Cross-entropy is a loss function that provides a measure of how the learning is progressing. Here, lower numbers are better as discussed in [3, 4]. The orange line represents the training set, while the blue line represents the validation set.

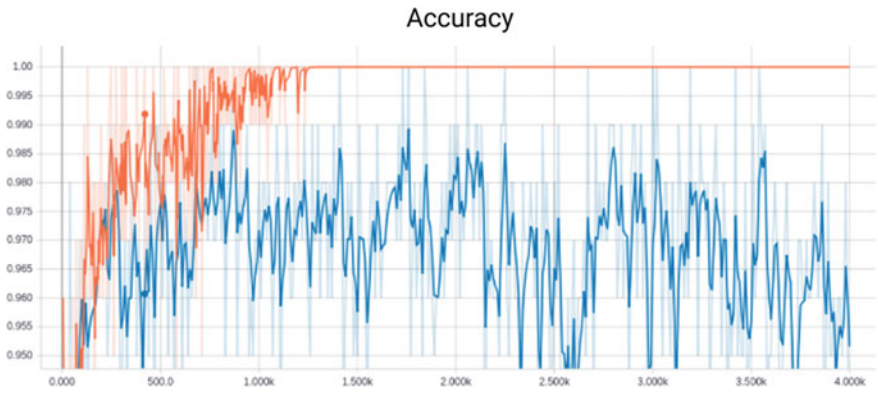


Fig. 4. Accuracy in terms of the threat/non-threat dataset

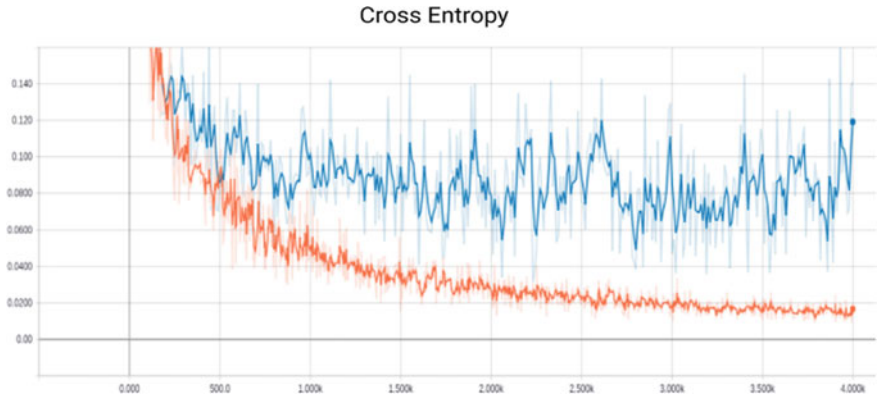


Fig. 5. Cross-entropy loss in terms of the threat/non-threat image dataset

In order to verify whether the model achieves its objectives and aims, the results evaluated. Consequently, a confusion matrix technique applied to calculate the number

of predicted images that were counted as true positive, true negative, false negative and false positive.

The confusion matrix of the classification of threat and non-threat images for test dataset contains 50 images is illustrated in Table 2.

Table 2. Confusion matrix for image detection

50 images	Predicted: non-threat	Predicted: threat
Actual: non-threat	36	1
Actual: threat	1	12

The confusion matrix presented in Table 2 and the calculations presented in Table 3 show that the proposed method for the classification of images into threat and non-threat images—that is, the retrained Inception-v3 model—achieved an overall accuracy rate of 96% with 4% error rate. In additional to accuracy and error rate measures, other measures are also calculated and shown in Table 3. Here, the number of images classified correctly for each class is clarified, while the number of images that were not classified correctly for each class is also highlighted.

Table 3. Calculations of measures for image detection

Measure	Calculated value (%)
Accuracy	96
Misclassification rate (error rate)	4
True positive rate (recall)	92
False positive rate	3
True negative rate (specificity)	97
Prevalence	26
F-Score	92

5 Conclusion

This research involved an experiment related to image threat detection using Google's model, the Inception-v3, which is a CNN architecture-based model on the TensorFlow platform. The transfer learning technique was used to retrain the Inception-v3 model in terms of a threat dataset. The experimental results of the proposed method show that the transfer learning method ensured the training was rapid and that the results were satisfactory. The detection accuracy of the model is 96%.

In terms of future work, this study can be used in relation to a threat detection system for social media images that could be a standalone system or part of a larger, hybrid solution.

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: a system for large-scale machine learning. In: USENIX 12th International Proceedings on Operating Systems Design and Implementation, pp. 265–283. Savannah, GA, USA (2016)
2. Han, D., Liu, Q., Fan, W.: A new image classification method using CNN transfer learning and web data augmentation. Expert Syst. Appl. **95**, 43–56 (2018)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. IEEE, Las Vegas, NV, USA (2016)
4. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861 (2017)
5. Islam, S.M.S., Rahman, S., Rahman, M.M., Dey, E.K., Shoyaib, M.: Application of deep learning to computer vision: a comprehensive study. In: 5th International Conference on Informatics, Electronics and Vision (ICIEV), pp. 592–597. IEEE, Dhaka, Bangladesh (2016)
6. Kamavisdar, P., Saluja, S., Agrawal, S.: A survey on image classification approaches and techniques. Int. J. Adv. Res. Comput. Commun. Eng. **2**(1), 1005–1009 (2013)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (2017)
8. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E.: A survey of deep neural network architectures and their applications. Neurocomputing **234**, 11–26 (2017)
9. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
10. Ponti, M.A., Ribeiro, L.S.F., Nazare, T.S., Bui, T., Collomosse, J.: Everything you wanted to know about deep learning for computer vision but were afraid to ask. In: 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), pp. 17–41. IEEE, Niterói, Brazil (2017)
11. Simonyan, K., Zisserman, A.: Very deep donvolutional networks for large-scale image recognition. arXiv:1409.1556 (2014)
12. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. arXiv:1512.00567 (2015)
13. Weiss, K., Khoshgoftaar, T.M., Wang, D.: A survey of transfer learning. J. Big Data **3**(9), 1–40 (2016)
14. Wu, Y., Qin, X., Pan, Y., Yuan, C.: Convolution neural network based transfer learning for classification of flowers. In: IEEE 3rd International Conference on Signal and Image Processing (ICSIP), pp. 562–566. IEEE, Shenzhen, China (2018)
15. Xia, X., Xu, C., Nan, B.: Inception-v3 for flower classification. In: 2nd International Conference on Image, Vision and Computing (ICIVC), pp. 783–787. IEEE, Chengdu, China (2017)

Author Index

X.-S. Yang et al. (eds.), *Proceedings of Fifth International Congress on Information and Communication Technology*, Advances in Intelligent Systems and Computing 1184, https://doi.org/10.1007/978-981-15-5859-7