Toward Real-Time Data Analytics Application for Industry 4.0

Bojana Bajic[1], Nikola Suzic[2], Nenad Simeunovic[1], Slobodan Moraca[1],
and Aleksandar Rikalovic[1(✉)]

[1] Department of Industrial Engineering and Management, Faculty of Technical Sciences,
University of Novi Sad, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia
{bojana.bajic,a.rikalovic}@uns.ac.rs

[2] Department of Management and Engineering, University of Padova, Stradella S. Nicola 3,
36100 Vicenza, Italy

Abstract. The Industry 4.0 is moving the production towards smart production systems, based on new technologies (i.e. Internet of Things, Cyber-Physical Systems, Cloud Computing, Big Data and Artificial Intelligence). Companies rightfully have high expectations of Industry 4.0. However, one of the major obstacles is how to transform reactive, via proactive, to predictive production systems. The predictive production systems are new type of intelligent production systems that enable the implementation of new technologies for discovering hidden information in the raw production data, this way enabling prediction and prevention of errors that could occur in production. The discovery of hidden information in the raw data is enabled by the real-time predictive data analytics on the basis of which the real-time decision making models are applied in the production system. The present study proposes a conceptual framework for real-time data analytics for Industry 4.0. The framework is developed based on the analysis of the available literature and informed with direct experience in implementing Industry 4.0 technologies in practice.

Keywords: Real-time data analytics · Industry 4.0 · Predictive production systems · Edge computing · Cloud computing

1 Introduction

The development of information and communication technologies offers opportunity to transform traditional production into a smart production. This digital transformation concept significantly impacts production processes and drives the fourth industrial revolution - Industry 4.0 [1].

Development of Industry 4.0 technologies (e.g. Internet of Things (IoT), Cyber-Physical Systems (CPS), Cloud Computing and Artificial Intelligence) results with the constant increment of amount of data obtained during production processes. Moreover, generation of large amount of raw data in production systems, called Big Data [2, 3], is becoming a key for optimizing production process and improving competitiveness [4, 5].

Thus, in this way, the data acquired across the product life-cycle can be converted into knowledge with positive impact on all aspects of production [6].

However, Big Data with all the positive expectations also brings challenges not encountered by production companies until now. Specifically, the biggest challenge today's production faces is the processing and analysis of Big Data. Consequently, the Big Data speeds up the development of data analytics technology for detecting hidden information among data collected by different machines and devices in the production processes using advanced analytical techniques (advanced statistical analysis, machine learning and expert systems). In this way, the emphasis is put on eliminating the issues that arise in production systems [7]. Another challenge of using Big Data is that very often it cannot be processed and analyzed using the existing software applications and personal computers due to insufficient processing power [4]. Therefore, new technologies, such as Cloud and Edge Computing, are using advanced data analytic techniques to detect hidden information to overcome the problems of processing and analyzing of the generated data.

The discovery of hidden information in the raw data is enabled by the real-time predictive data analytics. Based on real-time predictive data analytics, the real-time decision making models are applied in the production system [8] transforming a reactive production system into predictive. The predictive production systems enable proactive behavior permitting to anticipate the error before it occurs in the production process and to instantly take appropriate actions to avoid it [9]. Thus, we argue that in the near future of Industry 4.0 the real-time data analytics will play a mayor role in the development of the predictive production systems with ability of processing and analysing the data as it is generated in real-time [10].

The present research contributes to the body of real-time data analytics literature by proposing a conceptual framework for development of real-time predictive model for data analytics. Notably, the proposed framework represents the basis for development of real-time data predictive models based on datasets collected from production process.

The present paper is organized as follows. Section 2 provides a theoretical background of real-time data analytics for smart production systems and details on the research method. Section 3 proposes framework for real-time data analytics for Industry 4.0. Finally, Sect. 4 derives conclusions and provides information on the future research.

2   Theoretical Background and Research Method

This section summarizes the state of the art of the research in the subject fields, namely: Industry 4.0, Predictive production systems, IoT, and real-time data analytics for smart production. Further, the research method is provided.

2.1   Background: Real-Time Data Analytics for Smart Production

Industry 4.0 - Recently, Industry 4.0, has become one of the major research and development topics for industry and academia [7, 8]. The concept of Industry 4.0 can be defined in different ways, depending on the point of view and field of inquiry [1, 7]. In the present research, the used definition of Industry 4.0 observed through lens of production processes in industry and the application of data science based on data analytics,

reads: "Industry 4.0 is a concept that aims to increase production processes efficiency by collecting and interpreting data using data analytic techniques (e.g. data mining, statistical analysis etc.) and developing of predictive models based on which real-time decisions are made and executed" [11].

Predictive Production Systems - Intelligent production systems represent new generation of production systems that include hardware parts of equipment and machines, and software components [2, 6]. These production systems use techniques such as expert systems, fuzzy logic, statistics and machine learning [3] for production processes control [2]. One of the biggest challenges of production processes control is the transition from reactive to proactive production system. On the one hand, reactive production systems apply a posteriori inspections to detect defects (i.e. machine failures or defect products) at the end of the production process and reactively implement changes in the production system [2]. On the other hand, proactive production system use predictive (i.e. a priori) approach that aims to prevent machine failures or defect products from being manufactured in the first place [2]. Predictive production systems are a type of proactive Industry 4.0 intelligent production systems [4]. Thus, according to the previously said, predictive production systems predict an error occurrence inside the system and take adequate actions for avoiding the error to occur at all [4]. The real-time data analytics represents the future of the predictive production systems since it enables the data to be processed and analysed while it is generated in real-time [10].

Internet of Things (IoT) - The increasing availability of production data is changing the way decisions are taken in industry regarding predictive maintenance and quality improvement using data analytical methods [12]. The implementation of advanced Industry 4.0 technology, namely Internet of Things (IoT) allows the higher availability of production data due to it's ability of connecting different devices, communication technologies, sensor networks, Internet protocols, tags for RFID devices and so on [13]. The IoT combined with data analytics, can enable predictive production and networked production environments [4] for real-time data analytics [10].

Real Time Data Analytics - Data analytics, as a part of data science field, represents a practice for revealing hidden information among data collected from various devices further enabling real-time decision making in production systems using real-time data analytics [7]. Real-time data analytics, as a part of data analytics, refers to analytical techniques where data is processed and analysed while it is generated, in real-time [10] or near to real-time [14]. Among the available real-time data analytics research [15, 16] the main focus is put on real-time monitoring and quality controling without predictive abilities. Moreover, the recent systematic literature review [11] showed that data analytics implementation challenges are the most frequently addressed in the Industry 4.0 literature. Specifically, *"the inability to achieve real-time maintenance"* is one of these challenges ([11] based on [17]). The present research adds to discussion present in this literature stream.

2.2   Research Method

According to Phaal et al. [18] the term "conceptual" implies "concerned with the abstraction or understanding of a situation", while "framework" implies "supports understanding and communication of structure and relationship within a system for a defined purpose" (Phaal et al. [18] based on Shehabuddeen et al. [19]). Thus, conceptual frameworks "support understanding of an issue or area of study, provide structure, and support decision making and action" [18]. Conceptual frameworks are "needed to guide thinking about technology management, based on well-founded theoretical principles" [20].

The present research proposes a conceptual framework [18, 20] for real-time data analytics application for Industry 4.0. It does so by reviewing the relevant literature and conceptualizing a framework based on the findings. The proposed framework for real time data analytics is composed of three parts, namely: system characteristics, dataset characteristics, and IoT network infrastructure. Notably, the research presented is supported and additionally informed by researchers' insights obtained during the implementation of Industry 4.0 in industrial practice [21].

3   Proposed Framework for Real-Time Data Analytics for Industry 4.0

In the present research, we propose a framework for application of real-time data analytics for Industry 4.0 in production system (Fig. 1). The proposed framework is composed of: production system characteristics, dataset characteristics, and IoT network infrastructure.
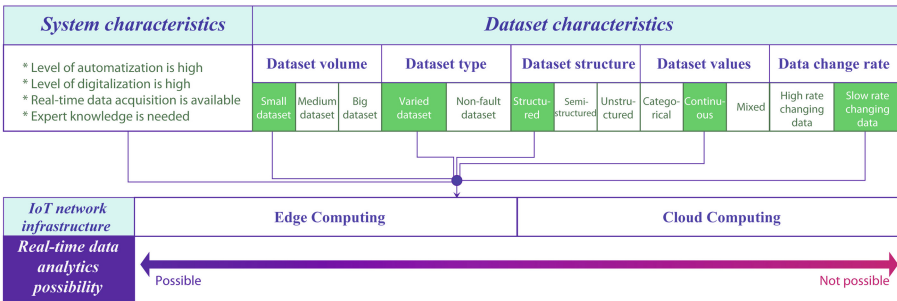

Fig. 1. Proposed framework for real-time data analytics

Production System Characteristics - Notably, we need to know and understand production system characteristics in order to determine if it is possible to develop a real-time data analytical predictive model for the observed production system [22]. Without production system characteristics it is very hard to imagine a real-time predictive model, because in general predictive model development is built on the characteristics of the production system it is meant to serve. Notably, production system that moves towards

implementation of real-time data analytics tends to have a high level of automatization and high level of digitalization which together enable real-time data acquisition [23] (Fig. 1). Also, expert knowledge must not be neglected, because there is a need for complete openness, sincerity, tight collaboration and constant communication between industry and researchers in order to obtain the synergy needed [24] for achieving real-time predictive production system. In this collaboration, industry practitioners provide the know-how on the observed production process, while researchers develop predictive models and test them in production systems [24] to improve the performances of the processes using real-time data analytics. Based on data acquired from production system, the dataset is formed.

Dataset Characteristics - Basic characteristics of acquired dataset are data volume, dataset type, dataset structure, dataset values and data change rate. In the following we elaborate on all five Data characteristics needed for implementation of real-time data analytics.

- Data volume - It is a generally accepted opinion that a large amount of data is necessary for the development of a predictive model for real-time data analytics [25]. However, recent research [26] points out the importance of small dataset in developing predictive models. Therefore, we argue that small carefully selected dataset based on experts' knowledge can provide an accurate prediction that is comparable with predictive models developed based on a medium or large dataset.
- Data type - The development of the predictive model with ability of real-time data analysis largely depends on the type of the data collected over a period of time in production systems [5]. Therefore, it is necessary to select a dataset without incomplete, homogeneous or noise data, which can disrupt the quality of a dataset [27, 28]. Thus, a high-quality dataset is crucial for developing an effective predictive model. A high-quality dataset is considered to consist of a certain number of samples from the periods when production system worked without problems, interruptions and difficulties, as well as a certain amount of data when certain problems occurred in the system [1]. Data samples from the periods when production system worked without any problems, interruptions and difficulties are defined as non-fault data samples, while data samples when certain problems occurred in the system are defined as fault data samples. Notably, the difficulty lies in finding of balance between the number of non-fault data samples and the number of fault data samples in order to obtain a varied dataset.
- Data structure - Production data is usually not in a format that can be used directly for data analysis [29]. Nevertheless, predictive models require a structured dataset. This means that dataset needs to be in the form of a matrix containing rows and columns, that is it is necessary for the dataset to be structured as a two-dimensional matrix [30]. The rows of the matrix represent all identified independent parameters, as well as one dependent parameter determined by the expert knowledge, while the columns correspond to each individual case (sample) recorded during the production process. Even though dataset needs to be in the structured form, in majority of cases the collected dataset is available in semi-structured and unstructured form. If the data is available in a semi-structured or unstructured form, the additional efforts are needed to convert the data into a usable structured form.

- Data values - Different values of dataset can be generated in production process, namely: categorical, continuous or combination of these two. On the one hand, for categorical variables, it is suggested that they are converted into binary value (e.g. value of variable can be either 0 or 1). If there are more than two levels in the categorical variable, a series of dummy variables should be used, where each defining the presence of a level [31]. On the other hand, continuous variables are numeric variables that have an infinite number of values between any two values. In other words, continuous values are numeric values [32]. Therefore, when collecting a dataset for development of a predictive model for real-time data analytics, the best choice is to use continuous value of a production parameter.

- Data change rate - Nowadays, large volumes of data are daily generated at high rate from heterogeneous sources during production process due to use of fast production equipment. This leads to fast change of data rate representing the frequency at which data is generated, captured, and shared [33]. Thus, the data arrives in stream and must be analyzed in real-time. That represents a difficulty for developing predictive model for real-time data analytics based on high-rate changing data. Therefore, in order to develop a reliable predictive model for real-time data analytics, the slow-rate changing data represents better option.

IoT Network Infrastructure - When considering the formed dataset, taking into account the characteristics of the production system, we need to choose IoT network infrastructure. The possibility of developing a predictive model for real-time data analytics depends on the right choice of IoT network infrastructure, i.e. Cloud Computing or Edge Computing. Bajic et al. [34] compare the Cloud and Edge Computing and stress out that these technologies do not rule out one another, but complement each other. However, on the one hand, the disadvantage of Cloud Computing is represented in the longer processing and computing times as well as a slower response time which prevents the implementation of real-time data analytics [34]. On the other hand, Edge Computing enables the data generation at the network edge reducing the distance that data must travel on the network enabling real-time data analytics implementation [34].

4   Conclusions and Future Work

The present research develops a framework for real-time data analytics (Fig. 1). It does so by analyzing the relevant literature and subsequently conceptualizing a corresponding framework. The framework is meant to serve as starting point for companies that would like to implement real-time data analytics as a part of transformation towards Industry 4.0. The proposed framework is composed of three parts, namely: system characteristics, dataset characteristics, and IoT network infrastructure. The research goes in detail with the dataset characteristics analysis.

In the future work, the developed conceptual framework for the real-time data analytics will be further refined and elaborated. Additionally, the research will move towards testing the proposed framework with the companies interested in implementing real-time data analytics.

References

1. Ghobakhloo, M.: The future of manufacturing industry: a strategic roadmap. J. Manuf. Technol. Manag. **29**(6), 910–936 (2018)
2. Holubek, R., Kostal, P.: The intelligent manufacturing systems. Adv. Sci. Lett. **19**(3), 972–975 (2013)
3. Li, B., Hou, B.-C., Yu, W.-T., Lu, X.-B., Yang, C.-W.: Applications of artificial intelligence in intelligent manufacturing: a review. Front. Inf. Technol. Electron. Eng. **18**(1), 86–96 (2017). https://doi.org/10.1631/FITEE.1601885
4. Nikolic, B., Ignjatic, J., Suzic, N., Stevanov, B., Rikalovic, A.: Predictive manufacturing systems in industry 4.0: trends, benefits and challenges. In: 28TH DAAAM International Symposium on Intelligent Manufacturing and Automation, pp. 769–802 (2017)
5. Sivarajah, U., Kamal, M.M., Irani, Z., Weerakkody, V.: Critical analysis of big data challenges and analytical methods. J. Bus. Res. **70**, 263–286 (2017)
6. Pacaux-Lemoine, M.P., Trentesaux, D., Zambrano Rey, G., Millot, P.: Designing intelligent manufacturing systems through human-machine cooperation principles: a human-centered approach. Comput. Ind. Eng. **111**, 581–595 (2017)
7. Kamble, S.S., Gunasekaran, A., Gawankar, S.A.: Sustainable industry 4.0 framework: a systematic literature review identifying the current trends and future perspectives. Process Saf. Environ. Prot. **117**, 408–425 (2018)
8. Wang, B.: The future of manufacturing : a new perspective. Engineering **4**(5), 722–728 (2018)
9. Ming, N.K., Philip, N., Sahlan, S.: Proactive and predictive maintenance strategies and application for instrumentation & control in oil & gas industry. Int. J. Integr. Eng. **11**(4), 119–130 (2019)
10. Milosevic, Z., Chen, W., Berry, A., Rabhi, F.A.: Real-time analytics. In: Big Data: Principles and Paradigms, pp. 39–61 (2016)
11. Bajic, B., Rikalovic, A., Suzic, N., Piuri, V.: Industry 4.0 implementation challenges and opportunities: a managerial perspective. IEEE Syst. J. **15**, 1–14 (2020). In press
12. Susto, G.A., Schirru, A., Pampuri, S., McLoone, S., Beghi, A.: Machine learning for predictive maintenance: a multiple classifier approach. IEEE Trans. Ind. Inform. **11**(3), 812–820 (2015)
13. Bajić, B., Rikalović, A., Morača, S.: Challenges of big data analytics in industry 4.0. In: SPIN 2019 Lin Transformacija I Digitalizacija Privrede Srbije, pp. 20–27 (2019)
14. Cui, Y., Kara, S., Chan, K.C.: Manufacturing big data ecosystem: a systematic literature review. Robot. Comput. Integr. Manuf. **62**, 1–20 (2020). In press
15. Zhang, Y., Wang, W., Wu, N., Qian, C.: IoT-enabled real-time production performance analysis and exception diagnosis model. IEEE Trans. Autom. Sci. Eng. **13**(3), 1318–1332 (2016)
16. Qian, C., Zhang, Y., Jiang, C., Pan, S., Rong, Y.: A real-time data-driven collaborative mechanism in fixed-position assembly systems for smart manufacturing. Robot. Comput. Integr. Manuf. **61**, 1–13 (2020). In press
17. Wan, J., et al.: A Manufacturing big data solution for active preventive maintenance. IEEE Trans. Ind. Informatics **13**(4), 2039–2047 (2017)
18. Phaal, R., Farrukh, C.J.P., Probert, D.R.: A framework for supporting the management of technological knowledge. Future Innov. Stud. **27**, 1–15 (2001)
19. Shehabuddeen, N., Probert, D., Phaal, R., Platts, K.: Management representations and approaches: exploring issues surrounding frameworks. In: BAM Working Paper (2000)
20. Phaal, R., Farrukh, C.J.P., Probert, D.R.: Technology management tools: concept, development and application. Technovation **26**(3), 336–344 (2006)
21. Bajic, B., Suzic, N., Simeunovic, N., Moraca, S., Rikalovic, A.: Real-time data analytics edge computing application for industry 4.0: the Mahalanobis-Taguchi approach. Int. J. Ind. Eng. Manag. **11**(3), 1–16 (2020)

22. Milosevic, Z., Chen, W., Berry, A., Rabhi, F.A.: Real-time analytics. In: Big Data: Principles and Paradigms, no. 2016, pp. 39–61 (2016)
23. Schmidt, R., Möhring, M., Härting, R.-C., Reichstein, C., Neumaier, P., Jozinović, P.: Industry 4.0 - potentials for creating smart products: empirical research results. In: Abramowicz, W. (ed.) BIS 2015. LNBIP, vol. 208, pp. 16–27. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19027-3_2
24. Ankrah, S., Al-Tabbaa, O.: Universities-industry collaboration: a systematic review. Scand. J. Manag. **31**(3), 387–408 (2015)
25. Oussous, A., Benjelloun, F.Z., Ait Lahcen, A., Belfkih, S.: Big Data technologies: a survey. J. King Saud Univ. - Comput. Inf. Sci. **30**(4), 431–448 (2018)
26. Feng, S., Zhou, H., Dong, H.: Using deep neural network with small dataset to predict material defects. Mater. Des. **162**, 300–310 (2019)
27. Zhang, S., Zhang, C., Yang, Q.: Data preparation for data mining. Appl. Artif. Intell. **17**(5–6), 375–381 (2003)
28. Hariri, R.H., Fredericks, E.M., Bowers, K.M.: Uncertainty in big data analytics: survey, opportunities, and challenges. J. Big Data **6**(1), 1–16 (2019)
29. Gupta, D.: Applied Analytics Through Case Studies Using SAS and R, 1st edn. Apress (2018)
30. Abbott, D.: Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst, 1st edn. Wiley, Hoboken (2014)
31. Yenidünya, B.: Robust design with binary response using Mahalanobis Taguchi system (2009)
32. Sarstedt, M., Mooi, E.: Regression analysis. In: A Concise Guide to Market Research: The process, Data, and Methods Using IBM SPSS Statistics, 1st edn., pp. 193–233. Springer, Berlin (2014). https://doi.org/10.1007/978-3-642-53965-7_7
33. Riahi, Y., Riahi, S.: Big data and big data analytics: concepts, types and technologies. Int. J. Res. Eng. **5**(9), 524–528 (2018)
34. Bajic, B., Cosic, I., Katalinic, B., Moraca, S., Lazarevic, M., Rikalovic, A.: Edge computing vs. cloud computing: challenges and opportunities in industry 4.0. In: 30TH DAAAM International Symposium on Intelligent Manufacturing and Automation, no. December, pp. 864–871 (2019)