

MINING DATA STREAMS FOR THE ANALYSIS OF PARAMETER FLUCTUATIONS IN IOT-AIDED FRUIT COLD-CHAIN

Petar Juric, Marija Brkic Bakaric, Xiang Wang, Xiaoshuan Zhang & Maja Matetic



This Publication has to be referred as: Juric, P[etar]; Brkic Bakaric, M[arija]; Wang, X[iang]; Zhang, X[iaoshuan] & Matetic, M[aja] (2016). Mining Data Streams for the Analysis of Parameter Fluctuations in IoT-Aided Fruit Cold-Chain, Proceedings of the 27th DAAAM International Symposium, pp.0756-0761, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-08-2, ISSN 1726-9679, Vienna, Austria
DOI: 10.2507/27th.daaam.proceedings.109

Abstract

The paper gives an overview of the state of the art methods and technologies in the field of stream data mining with applications in the Internet of Things systems for supporting fruit cold chain logistics. As the number of sensors used in on-line monitoring of the process is large, the amount of time series data is increasing rapidly. It is challenging to process such data in order to discover patterns, trends and outliers as a consequence of fluctuations of certain process parameters. In particular, the paper discusses methods for mining stream data collected in fruit cold chain aiming at real time control of fruit quality. A model of the centralized IoT system and the part responsible for monitoring fluctuations of temperature, humidity, and concentration of gases is proposed.

Keywords: Internet of Things, Mining data streams, Outlier detection, Parameter fluctuation, Cold-chain logistics

1. Introduction

Unlike the “classic” Internet which interconnects computer networks, the Internet of Things (IoT) refers to various areas with a common characteristic of connecting networks and devices from everyday life which can be uniquely identified on the Internet. IoT is used for monitoring and measuring parameters of interconnected physical world objects with the aim to improve business processes by data mining models and technics [1].

The first IoT prototype was devised by the scientists at the Auto-ID Centre of the Massachusetts Institute of Technology when they suggested designing a global network which would rely on the Radio Frequency Identification (RFID) technology for connecting devices. This network would enable identification and intelligent control over connected objects [2]. In the same year Kevin Ashton coined the term IoT to refer to an idea of using connected RFID devices within a distribution chain. Since data are generated by RFID and sensors, computers which have access to these data can observe, identify and understand the environment, and communicate (machine-machine) without limitations imposed by manual data input (time, concentration and precision) [3].

RFID technology in combination with wireless sensor networks enables monitoring environmental conditions important for quality preservation of temperature-sensitive food [4]. Beside temperature, it enables real-time humidity monitoring of agricultural products during transportation and storage in order to reduce loss and ensure high quality standards of these products [5]. State of the art devices also monitor other parameters that can affect food quality, such as shock, acceleration, light, temperature, sound, and combine these information with external information delivered over the Internet, such as rainfall, wind intensity, and route conditions [6].

IoT can help solving scalability and communication problems within sensor networks since it enables real time communication between sensor networks and cloud locations where data are analysed [10]. However, classic data mining methods cannot be applied since these systems generate Big Data [11], which is also stream data [15, 16]. This paper focuses on food cold chain data processing and proposes IoT cold chain model and its respective implementation with the support for simultaneous analysis of various environmental parameters and processing multitude of cold chains in both, online and offline modes. An example of processing data in time windows [17, 18] is provided. The purpose is to build an optimized system with quick autonomous reactions for food quality preservation.

2. Mining wireless sensor networks data

Applying wireless sensor networks (WSN) data mining algorithms can be seen as centralized (collection node) at the location of the computer centre and data warehouse or distributed across the network (sensor nodes) [7]. Centralized systems have more computing power at the expense of real-time data processing. Updating and sensor reactions are hence delayed or subsequent. Distributed systems process data real time and react quickly in case of deviation, but they do not enable more complex analysis due to low computing power [8].

Regardless of the system setup, the main goals of data mining entail frequent pattern detection, sequential pattern detection, clustering and classification. Data mining in distributed systems is mostly used for optimizing processing efficiency, energy consumption, and data and memory flow. Centralized systems are mostly used for prediction, raising result precision, and rationalization of operational costs [9].

Data mining within these systems faces scalability and communication problems when other sensor networks need to be incorporated.

IoT can help solving these problems since it enables real time communication between sensor networks and cloud locations where data are analysed. This meets the requirements for smart connecting devices and context aware observation activities [10].

3. IoT and Big Data

Since data generated within these systems has Big Data characteristics, classic data mining methods cannot be applied [11]. Observation, identification and understanding data generated by IoT systems is achievable by applying advanced machine learning algorithms adapted for real time Big Data stream processing.

Big Data is defined by three big Vs: Volume, Velocity at which data are generated and sent to a system, and Variety (structured, half-structured and unstructured dynamic data) [12]. Machine to machine (M2M) communication adds two more Vs to IoT Big Data complexity: Variability, which refers to the oscillating speed of data generation and system input, and Value, which refers to the fact that not all data are equally important for analysis, i.e. those that can improve the process or model are more important. Data mining social networks can additionally raise complexity with other Vs like Veracity, which refers to data accuracy in expressing personal attitudes with regard to subjectivity [13].

4. Mining data streams

Big Data is mainly stream data. Stream data is every ordered pair of sequences (i, t) where i is an n -tuple and t is a positive time interval. Data size cannot be directly limited since stream data sequences are continuously generated in different intensities and spans. Thus it is potentially unlimited [14].

Classic data mining algorithms access file data or relational database data. On the other hand, stream data mining algorithms are applied at the very stream [15]. Stream data mining results are expected in real time. Single pass data processing is common for all stream data mining algorithms. It is usually done by using time windows, micro clustering, limited aggregation and approximation [16].

Windows can be defined physically by time spans or logically by the number of elements [17]. State of the art algorithms seek to determine optimal window size.

Data processing in time windows can handle data streams and important variables in the following manners:

- summative and continuously from the beginning of measurement with equal weights (landmark window)
- summative and continuously from the beginning of measurement with different weights – more obsolete data have lower weights (damped window)
- temporarily within defined time spans, e.g. only last 60 seconds (sliding window) [18].

The process of mining multiple data streams includes alignment, approximation, combination, building the model and adaptation to concept drift [19].

5. Stream data mining model within cold chain logistics

The analysis of wireless sensor network data produced by food distribution chains from the source to end-users can lead to process optimization which, in turn, can affect food quality preservation [20].

IoT solutions in food supply chain make parameters relevant for food preservation visible to all the stakeholders and enable cooperation between manufacturing, storage, transporting, selling and buying in order to perform automated activities in real-time with the aim to achieve maximum food quality, on-time delivery, and food preparation under optimal conditions [21].

Although cold chain and IoT in food transportation imply raising prices for end-users, due to implementation and maintenance costs and due to an increase in energy consumption, the acceptance of such systems facilitates process efficiency. Efficiency is reflected in raising standards for food safety and loss minimization [22].

IoT for supporting cold chain logistics is made possible by integrating the Internet and the existent system of wireless sensor networks. Data mining of these IoT systems should support M2M communication between the system for collecting data, namely sensors, and the system for real time knowledge extraction from multiple data streams. The goal is to build a system which would enable quick autonomous reactions and optimization in logistics of different cold chains for fruit quality preservation.

After careful observation of the IoT cold chain, we propose modelling data with approximate time windows. The window size should be defined by the time needed for detecting deviation of the combination of discretized values of elements that can be influenced upon (temperature, humidity, SO₂ concentration) from their limit values [23]. This model upgrades WSN and the respective model for managed traceability system (MTS) [24] in grape cold chain [25] with autonomous machine (computer) for detecting critical CO₂ values (values which indicate senescence), and for managing temperature, humidity and SO₂ concentration in order to preserve optimal grape quality. An example is shown in Figure 1.

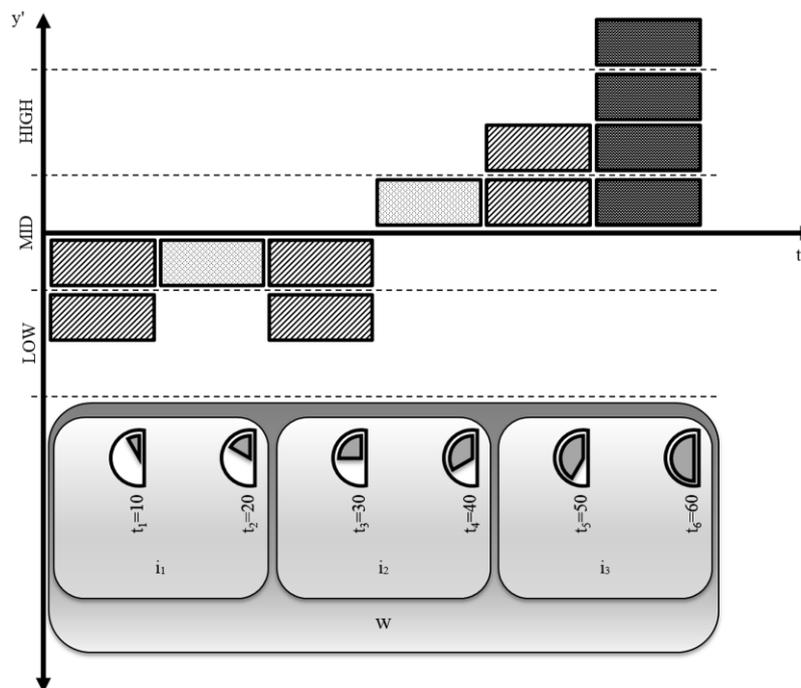


Fig. 1. Example of a discretized stream data mining of WSNMTS within grape cold chain

The window interval is made up of the following: the measured CO₂ concentration (dotted = optimal, diagonal = passing, trellis = critical), time t , time intervals i for data processing, window time span w , and the combination of temperature, humidity and SO₂ concentration y' . The example given in Fig. 1 has a window size of 60 seconds which is updated every 20 seconds. The change in the concentration of CO₂ from optimal to critical is detected after three intervals. The sensor data is sent twice per interval. The last detected CO₂ value in the interval is taken as the reference value for the interval. The window size should be adapted to real conditions in order to achieve timely and precise reactions based on M2M communication.

The problem with centralized WSN systems is reflected in their impossibility for timely reactions, i.e. real time reactions, since data are saved and loaded from the database instead of analysed as they are recorded [26].

The centralized IoT system capable of real time data mining and sending results back to sensors and computer devices within the cold chain in order to adjust to micro climate conditions can be achieved by using the proposed model and the newest technology for processing data streams. The system implementing such a model would need to deal with Big Data scalability, heterogeneity, and timeliness [27] considering that it would simultaneously analyse data of a number of cold chains, and potentially a number of groceries. From the IoT perspective, this system should be built as a cloud service [28].

The system could be based on open-source technologies for machine learning and for Big Data analysis of a third generation which enable real time data processing [29]. Systems suitable for cold chain data processing are Hadoop and Spark [30]. The proposed IoT cold chain model is given in Figure 2.

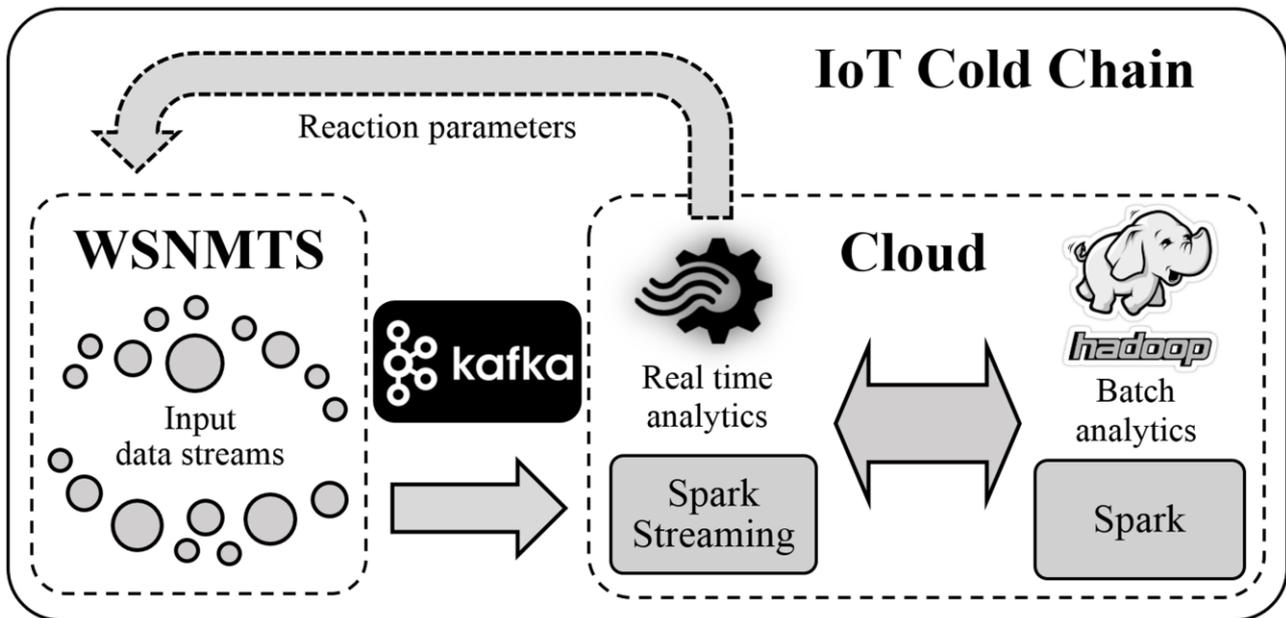


Fig. 2. IoT Cold-chain model

When designing such systems, problems and security threads which can affect data availability and data integrity need to be taken into account [31].

In order to ensure better data availability and data analysis, software technologies supporting clustering (e.g. Apache Kafka) could be used for input and bus architectures for data sent from multiple WSNMTS [32]. The real time stream data processing functionality of the system could be achieved with Spark Streaming [33]. A drawback of this approach is the failure to deliver data and real-time reactions when offline. If it would rely on the Spark architecture [34], the system would also be capable of on-demand batch data processing [35], i.e. Big Data mining over the entire dataset stored in the Hadoop file system [36] and supported data warehouses.

The combination of both data processing modes is a prerequisite for improving classification accuracy and evaluating stream data mining model. The system could be updated in online and offline (batch) modes. Such approach would lessen the problem with Internet connectivity because the system could base input reactions on the last version of the cumulatively trained offline model. The model could even be stored on computer equipment used in warehouses or vehicles, although this might negatively affect energy consumption.

6. Conclusion

The paper deals with the problem of real time communication between sensor networks and cloud locations in food cold chain. We propose a model for the centralized IoT system capable of real time data mining and sending results back to sensors and computer devices within the cold chain with a special focus on the part responsible for monitoring fluctuations of temperature, humidity, and concentration of gases. Lastly, we discuss and evaluate possible technological solutions for the proposed model.

In our future work we will deal with problems which relate to the failure to deliver certain variables important for analysis and prediction due to sensor failure or signal loss. Since in such cases real data are delivered with delay, the problem of resynchronizing multiple stream data occurs [37]. Therefore, system implementation would have to take advantage of techniques memorizing the last known state or compensating the stream with virtual data approximating the trend until then.

7. Acknowledgements

This work has been fully supported by the University of Rijeka under the project number 13.13.1.3.03.

8. References

- [1] Winter, J. S. (2014). Privacy Challenges for the Internet of Things, In: Encyclopedia of Information Science and Technology, Third Edition, Khosrow-Pour, M. (Ed.), pp. 4373-4383, ISBN 978-1-4666-5888-2, IGI Global.
- [2] Xiaoping, X.; Luoxian, L.; Mingyang, L. & Guobin, L. (2012). Perspectives on Internet of Things and Its Applications, Proceedings of the 2012 International Conference on Computer Application and System Modeling, Taiyuan Institute of Science and Technology, Taiyuan, Shanxi, China, ISSN 1951-6851, ISBN 978-94-91216-00-8, pp. 20-24, Atlantis Press, DOI:10.2991/iccasm.2012.6.
- [3] Ashton, K. (2009). That 'Internet of Things' Thing, Available from: <http://www.rfidjournal.com/articles/view?4986>, Accessed: 2016-09-07.
- [4] Ting, P. H. (2013). An Efficient and Guaranteed Cold-Chain Logistics for Temperature-Sensitive Foods: Applications of RFID and Sensor Networks, International Journal of Information Engineering and Electronic Business, Vol. 5, No. 6, pp. 1-5, ISSN 2074-9031.
- [5] Li, Y.; Peng, Y.; Zhang, L.; Wei, J. & Li, D. (2015). Quality Monitoring Traceability Platform of Agriculture products Cold Chain Logistics Based on the Internet of Things, Chemical Engineering Transactions, Vol 46, pp. 517-522, ISSN 2283-9216.
- [6] Capello, F.; Toja, M. & Trapani, N. (2016). A Real-Time Monitoring Service based on Industrial Internet of Things to manage agrifood logistics, Proceedings of the 6th International Conference on Information Systems, Logistics and Supply Chain, Bordeaux, France, Available from: http://ils2016conference.com/wp-content/uploads/2015/03/ILS2016_FB01_1.pdf, Accessed: 2016-10-21.
- [7] Flammini, A. & Sisinni, E. (2014). Wireless Sensor Networking in the Internet of Things and Cloud Computing Era, Procedia Engineering, Vol. 87, pp. 672-679, ISSN 1877-7058.
- [8] Lojka, T. & Zolotova, I. (2014). Distributed sensor network – data stream mining and architecture, Advances in Information Science and Applications, Vol. 1, Proceedings of the 18th International Conference on Computers (part of CSCC '14), Santorini Island, Greece, Recent Advances in Computer Engineering Series, ISSN 1790-5109, ISBN 978-1-61804-236-1, pp. 98-103.
- [9] Mahmood, A.; Shi, K. & Khatoun, S. (2012). Mining Data Generated by Sensor Networks: A Survey, Information Technology Journal, Vol. 11, No. 11, pp. 1534-1543, ISSN 1812-5638.
- [10] Gubbia, J.; Buyyab, R.; Marusic, S. & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions, Future Generation Computer Systems, Vol. 29, No. 7, pp. 1645-1660, ISSN 0167-739X.
- [11] Fouad, M. M.; Oweis, N. E.; Gaber, T.; Ahmed, M. & Snasel, V. (2015). Data Mining and Fusion Techniques for WSNs as a Source of the Big Data. Procedia Computer Science, Vol. 65, pp. 778-786, ISSN 1877-0509.
- [12] Gandomi, A. & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management, Vol. 35, No. 2, pp. 137-144, ISSN 0268-4012.
- [13] Özköse, H.; Ari, E. S. & Gencer, C. (2015). Yesterday, Today and Tomorrow of Big Data, Procedia – Social and Behavioral Sciences, Vol. 195, pp. 1042-1050, ISSN 1877-0428.
- [14] Namiot, D. (2015). On Big Data Stream Processing, International Journal of Open Information Technologies, Vol. 3, No. 8, pp. 48-51, ISSN 2307-8162.
- [15] Hebrail, G. (2008). Data stream management and mining, In: Mining Massive Data Sets for Security, Fogelman-Soulie, F.; et al. (Eds.), pp. 89-102, IOS Press, ISBN 978-1-58603-898-4.
- [16] Han, J. & Gao, J. (2009). Research Challenges for Data Mining in Science and Engineering, In: Next Generation of Data Mining, Kargupta, H.; et al. (Eds.), pp. 3-28, Chapman & Hall, ISBN 978-1-4200-8586-0.

- [17] Joseph, S.; Jasmin, E. A. & Chandran, S. (2015). Stream Computing: Opportunities and Challenges in Smart Grid, *Procedia Technology*, Vol. 21, pp. 49-53, ISSN 2212-0173.
- [18] Shah, H. M. & Kaur, N. (2014). Improve Frequent Pattern Mining in Data Stream, *International Journal of Research in Engineering & Technology*, Vol. 2, No. 5, pp. 143-152, ISSN 2321-8843.
- [19] Spiliopoulou, M. (2012). Mining Multiple Interdependent Streams, Tutorial at European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 24-28 September 2012, Bristol, UK.
- [20] Ruiz-Garcia, L.; Lunadei, L.; Barreiro, P. & Robla, J. I. (2009). A Review of Wireless Sensor Technologies and Applications in Agriculture and Food Industry: State of the Art and Current Trends, *Sensors*, Vol. 9, No. 6, pp. 4728-4750, ISSN 1424-8220.
- [21] Xiaorong, Z.; Honghui, F.; Hongjin, Z.; Zhongjun, F. & Hanyu, F. (2015). The Design of the Internet of Things Solution for Food Supply Chain, *Proceedings of the 5th International Conference on Education, Management, Information and Medicine (EMIM 2015)*, Shenyang, China, ISSN 2352-5428, ISBN 978-94-62520-68-4, pp. 314-318, Atlantis Press, DOI:10.2991/emim-15.2015.61.
- [22] Liu, X.; Xu, M. & Yu, C. (2016). Food Cold Chain Logistics Based on Internet of Things Technology, *Proceedings of the 6th International Conference on Applied Science, Engineering and Technology (ICASET 2016)*, Qingdao, China, ISSN 2352-5401, ISBN 978-94-6252-186-5, pp. 92-96, Atlantis Press, DOI:10.2991/icaset-16.2016.18.
- [23] Kasetty, S.; Stafford, C.; Walker, G. P.; Wang, X. & Keogh, E. (2008). Real-Time Classification of Streaming Sensor Data, *Proceedings of the 20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2008)*, Vol. 1, Dayton, Ohio, USA, ISSN 1082-3409, ISBN 978-0-7695-3440-4, pp. 149-156, IEEE Computer Society, Los Alamitos, CA 90720-1314.
- [24] Zhang, J.; Liu, L.; Mu, W.; Moga, L. M. & Zhang, X. (2009). Development of temperature-managed traceability system for frozen and chilled food during storage and transportation, *Journal of Food, Agriculture & Environment*, Vol. 7, No. 3, pp. 28-31, ISSN 1459-0255.
- [25] Xiao, X.; Wang, X.; Zhang, X.; Chen, E. & Li, J. (2015). Effect of the Quality Property of Table Grapes in Cold Chain Logistics-Integrated WSN and AOW, *Applied Sciences*, Vol. 5, No. 4, pp. 747-760, ISSN 2076-3417.
- [26] Duhaney, J.; Khoshgoftaar, T. M.; Agarwal, A. & Sloan, J. C. (2010). Mining and storing data streams for reliability analysis, *Proceedings of the 16th International Society of Science and Applied Technologies on Reliability and Quality in Design*, Washington D.C, USA, Pham, H. (Ed.), pp. 314-317, ISBN 978-0-9763486-6-5, International Society of Science and Applied Technologies, Piscataway, NJ 08855.
- [27] Cortes, R.; Bonnaire, X.; Marin, O. & Sens, P. (2015). Stream processing of healthcare sensor data: studying user traces to identify challenges from a big data perspective, *Procedia Computer Science*, Vol. 52, pp. 1004-1009, ISSN 1877-0509.
- [28] Gnimpieba, Z. D. R.; Nait-Sidi-Moh, A.; Durand, D. & Fortin, J. (2015). Using Internet of Things technologies for a collaborative supply chain: Application to tracking of pallets and containers, *Procedia Computer Science*, Vol. 56, pp. 550-557, ISSN 1877-0509.
- [29] Agneeswara, V. S. (2014). *Big Data Analytics Beyond Hadoop: Real-Time Applications with Storm, Spark, and More Hadoop Alternatives*, Pearson, ISBN 978-0-13-383794-0, New Jersey 07458, USA.
- [30] Ko, D.; Kwak, Y.; Choi, D. & Song, S. (2015). Design of Smart Cold Chain Application Framework Based on Hadoop and Spark, *International Journal of Software Engineering and Its Applications*, Vol. 9, No. 12, pp. 99-106, ISSN 1738-9984.
- [31] Cvitic, I.; Vujic, M. & Husnjak, S. (2015). Classification of Security Risks in the IoT Environment, *Proceedings of the 26th DAAAM International Symposium*, pp. 0731-0740, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-07-5, ISSN 1726-9679, Vienna, Austria, DOI:10.2507/26th.daaam.proceedings.102.
- [32] Apache Kafka. Available: <http://kafka.apache.org>, Accessed: 2016-09-22.
- [33] Spark Streaming. Available: <http://spark.apache.org/streaming>, Accessed: 2016-09-22.
- [34] Apache Spark. Available: <http://spark.apache.org>, Accessed: 2016-09-22.
- [35] Karau, H.; Konwinski, A.; Wendell, P. & Zaharia, M. (2015). *Learning Spark*, O'Reilly Media, ISBN 978-1-4493-5862-4, Sebastopol, CA 95472, USA.
- [36] Apache Hadoop. Available: <http://hadoop.apache.org>, Accessed: 2016-09-20.
- [37] Kreml, G.; et al. (2014). Open Challenges for Data Stream Mining Research, *ACM SIGKDD Explorations Newsletter - Special issue on big data archive*, Vol. 16, No. 1, pp. 1-10, ISSN 1931-0145.