What Does Data Analytics Offer for Extracting Knowledge from Middle-of-Life Product Data?

Abou Eddahab, Fatima-Zahra; Horvath, Imre

DOI
10.3233/978-1-61499-898-3-1102

Publication date
2018

Document Version
Final published version

Published in
Transdisciplinary Engineering Methods for Social Innovation of Industry 4.0

Transdisciplinary Engineering Methods for Social Innovation of Industry 4.0
M. Peruzzini et al. (Eds.)

doi:10.3233/978-1-61499-898-3-1102

What Does Data Analytics Offer for Extracting Knowledge from Middle-of-Life Product Data?

Fatima-Zahra ABOU EDDAHAB[1] and Imre HORVÁTH
Delft University of Technology, the Netherlands

Abstract. Companies are getting increasingly interested in learning how different customers use their products. Collecting data about the use of products provides useful insights and facilitates design enhancements. Effective data analytics needs dedicated tools. In this paper, we summarize the results of our literature research done with special attention to existing tools. We observed that everything is changing rapidly and getting more complex in terms of data and processing methods and tools. While remarkable attention has been paid to processing big data, much less is being devoted to effective semantic progressing of middle-of-life (MoL) data. One of our findings is that commercialized data analytics tools have not addressed extraction, aggregation, and handling genuine MoL data adequately. Another one is that the currently available tools are in the lack of the capability to adapt themselves to designers needs and to produce results that could be reused in multiple design tasks. Nowadays products are equipped with smart capabilities and this offers new opportunities for exploiting middle-of-life data. The knowledge aggregated in this study will be used in the development of a sophisticated toolbox. This will: (i) integrate various tools under a unified interface, (ii) implement various semantics orientated and smart reasoning-based functions, and (iii) facilitate data transformations by practicing designers in contexts.

Keywords. Data Analytics, Middle-of-Life Product Data, Analytics Tools, Application Practices.

Introduction

The last decade witnessed the merge of product engineering big data analytics and software tool development. They together form a new transdisciplinary field of knowing and making that offers new opportunities for product developers. Consequently, companies can combine (i) static process information with dynamic ones, (ii) product information with information concerning processes and resources, and (iii) human aspect information with business information [1]. However, as reflected by the related literature, most of the efforts were so far dedicated to methodological and computational support of begin-of-life and end-of-life models and activities. Less efforts were made to exploit middle-of-life (MoL) data and to value/knowledge creation based on it. Our background research was stimulated by the observation that there seems to be a lack of tools to support decision-making in product and service design using MoL data. This issue was placed into the position of a

[1] Corresponding Author, Mail: f.aboueddahab@tudelft.nl

concrete research phenomenon to be studied.

The above decision was underpinned by the following argumentation. Effective statistical and semantic processing of MoL data is not only an academic challenge, but also a useful asset for the industry [2]. It is important for product developers and production companies to see how different customers use their products in different application contexts and under different circumstances. This may provide insights in how to transform use patterns into specific product enhancements, and how to avoid deficiencies, which may occur under circumstances not completely known or specified in the development phase of their products. MoL data can be aggregated by (i) field observations, (ii) interrogations of the users, and (iii) studying failure log files and maintenance reports, or (iv) from relevant web resources such as social media and user forums. Alternatively, they can be elicited directly from and by products using sensors or self-registrations. However, due to the dynamic change of sensor data, typically large volumes of data are to be aggregated over time. Due to the unknown nature of data patterns, it is unfortunately not straightforward to perform effective data analysis using the existing traditional techniques [3]. Feeding structured MoL data and use patterns back to product designers is an insufficiently addressed issue [4]. The key challenge is to find ways of using data analytics techniques effectively in purposeful combinations, depending on the application contexts and specific objectives of product designers [5].

1. Organization of the study

1.1. Reasoning model

We completed a comprehensive literature study in two phases. The first phase was a 'shallow exploration'. It was conducted to identify the most relevant domains of knowledge for the study. Based on a wide range of keywords, we tried to develop a topographic landscape of the related publications. It was supposed not only to show the distribution (the clusters of the keyword-related publications), but also the peaks and the plains of the clusters. Based on these considerations, we established four major clusters of papers: (i) changes in the nature of data, (ii) approaches of transformation of data, (iii) tools and packages for data analytics, and (iv) design applications of data analytics. These four domains of knowledge were brought into an implicative interrelationship. The second phase included a 'deep exploration', where various sources such as subscription-based and open access journals, conference proceedings, web-repositories, and professional publications were used for collecting several hundreds of relevant publications. It provided opportunity for a quantitative characterization of the interrelationships among the key terms belonging to the same cluster. The pieces of information from the quantitative part of the literature study were used to develop a reasoning model for the qualitative one. This part of the activities focused on the interpretation of the findings and disclosed semantic relationships. The constructed reasoning model, which was also used for structuring the contents of this paper, is shown in Figure 1. It indicates only the first and second level key terms, while the study was actually completed using more specific key terms of third decomposition level.

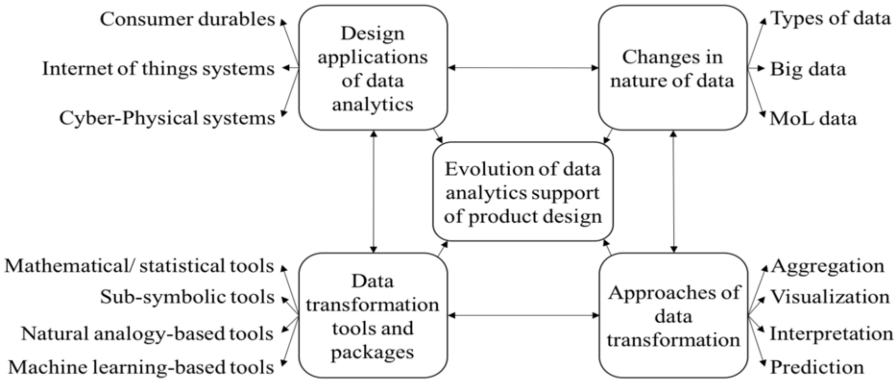

Figure 1. Reasoning model of the literature research.

1.2. Content and structure of the rest of the paper

In the rest of this paper, we provide details about the state of the art in the broad field of data analytics methods and tools, which support extracting product developmental knowledge from middle-of-life product data. In Section 2, we investigate the essence and trend of changes from product-associated data to processing big data. In Section 3, we provide an overview of the various data transformation actions and techniques, and discuss the accompanying challenges. In Section 4, we summarize the findings related to existing software tools for data analytics and discuss how they can be improved according to the literature. In Section 5, the major application domains of various big data analytics approaches are discussed as well as their challenges. The last Section discusses the implications of the findings. In this Section, we present our conclusions regarding what needs to be done to find solutions for the unearthed issues and deficiencies.

2. Overview of the changing nature of data

The overall research objective of our study was to find and analyze scientific and professional publications that discuss the recent changes and trends in the nature of data. However, due to the abundance of the kinds of data, the actually conducted review was restricted to data associated with monitoring products and use of services in real life operation, and to data obtained by user feedback on social media. On the other hand, this scoping of the study made it possible to derive highly relevant conclusions in the narrower context of our research. The practical objective was formulated as to study the kind of data that makes it possible: (i) to extend product and service lifespans, and (ii) to optimize the use of the necessary resources all along their lifecycle. Here, the term ‘lifecycle’ refers to all observable phases of the life of products and services. Thus, it covers: (i) the beginning of life (BoL) (where the product is designed and realized), the middle of life (MoL) (where the product is available on the market and used by the customer), and (iii) the end of life (EoL) (where the product is dismissed or revamped) [6].

The trends of change concern not only the sources and amount (size) of digital data,

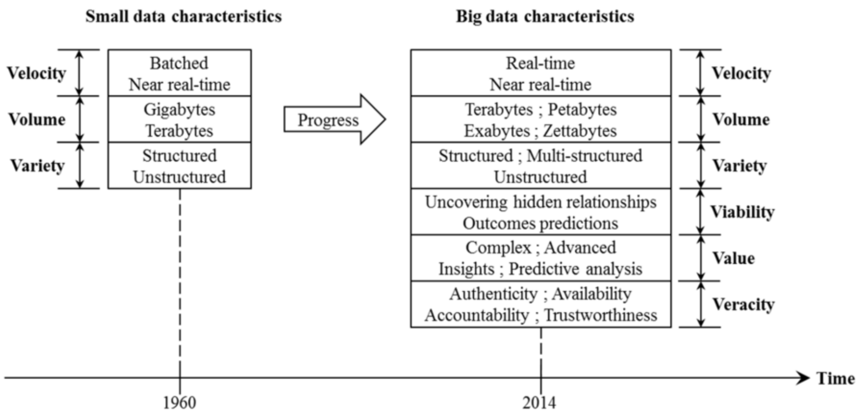

Figure 2. Reasoning model of the literature research Change in the nature of data (designed after [10]).

but also the arrangement (structure) of data. Figure 2 represents this progression. Unstructured data do not support (i) traditional database management [7], (ii) formal content and relationship analyzes, or (iii) applying pattern searching methods [8]. The papers related to the nature of data cluster indicate that individuals, industry, and science face the challenge of dealing with large datasets. This is a result of the proliferation and ubiquity of high-throughput computing technologies and Internet connectivity. The main difficulty is not in the technical handling of large amount of data, but in mining and extracting valuable information and knowledge from them [9]. Decades ago data was characterized by three characteristics (volume, velocity, and variety), recently it has been complemented with three more characteristics (value, veracity and viability) [10]. In order to deal with its scalability and affordability, the literature suggests that big data requires optimized data warehouses and cloud computing [11]. The outcome of the literature review showed that managing and gaining insights from the produced big data is a challenge and a key to competitive advantage [12]. It offers substantial value to organizations that decided to adopt it, but also poses a number of challenges for the realization of such benefit [13].

Based on the outcomes of these studies we are conceptualizing a functional, architectural and information-processing framework for the target toolbox showed that it is important to focus on MoL data that primarily, but not exclusively include use, service and maintenance data [14]. These data allow the observation of the condition and the behavior of products during the usage phase [15]. Furthermore, acquisition of MoL data creates opportunities and encourages a life-cycle orientated approach to product design, which evaluates and enhances all products and services on a continuous basis [16]. In other words, MoL data can be transformed into knowledge that enables a perpetual and long term design improvement, and product innovation and planning.

A recognized difficulty associated with MoL data is that the related elicitation activities should be executed outside the companies, typically with an intense involvement of both the products and the end users. If elicitation of product-related information is interventional, then it may lead to operational inefficiencies. Since conventional information systems used in the definition of products and services cannot handle MoL data and the knowledge, the need for dedicated data analytics approaches and tools have been recognized also by the developers of product life cycle management systems. Nevertheless, in the current industrial practice the potentialities offered by MoL data analytics is seldom utilized sufficiently by product and service developers. The gradually increasing smart behavior of products have been recognized as a key development in collecting and feeding back data and information to designers with regards to modes of use and operation [14].

3. Analysis of data transformation steps and techniques

The term ’data transformation’ (DT) has a broader and a narrower meaning. In the narrower meaning DT is the process of converting data and information from one format to another, usually from the format of a source system into the required format of a destination system [17]. Typical statistical transformations are such as: (i) logarithmic, (ii) square root, (iii) square, (iv) cube root, and (v) reciprocal transformation. In the broader meaning, it refers to all data processing activities that can introduce change in the state, representation, and/or meaning of data [18]. That is, data transformation is the process by which data in a dataset are transformed, or changed, during data cleaning and involves the use of mathematical operations in order to reveal features of the data that are not observable in their original form [19].

The mining of big data primarily focuses on extracting patterns to be evaluated by both manual and automated approaches [20]. Currently, dealing with patterns requires multiple expert interventions (especially after mining) [21]. At the same time, the currently widespread methods of big data transformation do not consider human behavior, which adds uncertainty to the outcome of the process [22]. Extraction of patterns is typically done on historical data, rather than won real-time acquired data. Mining and transforming big data necessitates highly scalable strategies. In order to achieve more effective processing, the literature suggests developing sophisticated data filtering and integration techniques, as well as using advanced parallel computing environments and more effective user involvement [23].

Researchers also observed that, if the process of transforming data to knowledge is time consuming, it may negatively influence the relevance of the extracted knowledge, or its validity in the dynamic context, or it can even make the extracted knowledge invalid [24]. This issue has been addressed by many publications, but the contour of a general solution does not seem to be in formation. However, one issue that does not seem to be sufficiently addressed in the literature is providing meaning to data (automatically or semi-automatically). The importance of this issue originate in that it concerns and may computationally influence all data transformation steps. A hierarchy of concepts interlinked by the assumed relationships, and the axioms able to express the relationships of the concepts and to constrain their interpretation are seen as ingredients of a possible solution [25]. In addition, only limited efforts have been made related to capturing the semantics of transformed data and to interpreting the meaning of transformed data in context [26].

It seems that there are multiple challenges related to the early preparatory activities of data analytics. One of them is data inundation, which may manifest as the major performance bottleneck for processing the output of increasingly complex sensor networks used to monitor product use and life cycle performance [27]. It was argued that data analytics would be challenged significantly by the necessity of combining sensor generated macro data with end-user generated micro data, when the task is to figure out their mutual meaning and influences [28]. Furthermore, associating quantitative data with qualitative data needs specific blending and fusing techniques that are also semantics and context sensitive.

4. Interpretation of data analytics tools and packages

The literature presents, discusses, and compares many tools that have been developed to help understand and process massive and big data. However, the overwhelming majority of these tools are general-purpose statistical tools [29]. Incomparably less number of tools has been developed to assist the improvement of products and services [30]. The general-purpose software means are typically sorted into three categories: (i) single task orientated software tools, (ii) multi tasks orientated integrated software packages (and toolboxes), and (iii) multi-functional development environments. Though the literature discusses many big data mining and analysis tools, the majority of them are still in their infancy [31].

There is a need for computational theories and tools to assist humans in servicing [32], and to extract useful information and knowledge from the rapidly growing volumes of digital data. This is also confirmed by a study done in 2012, which explains that 23% of the digital world is producing data that would be useful for big data analysis, but unfortunately only 3% of these potential data is identified and even less is analyzed [33]. The fact is that there are a very large number of existing commercial and open source tools, and choosing the most appropriate one for a particular data analysis task is already a challenge in itself [34]. The situation is even more complicated when choosing the most convenient tool is also targeted [35]. An apparent technological issue for traditional software tools that amount of data generated and stored at different sources grows rapidly and their handling needs a sufficient level of automation [36]. In the lack of this, it is becoming hard to capture, store, manage, analyze, visualize, and share mass data using typical tools [37]. Required are powerful non-traditional tools that can take care of interpretation of big data in a way that goes beyond human capabilities and offers better comprehension and decision-making [38].

In the case of any comparison of the tools not only the functionalities and the data analytics tasks at hand, but also the user groups, the data structures, the processing methods, the import and export of data, the use of models, as well as the platforms and the licensing have to be taken into consideration. The choice of tools is also influenced by several pragmatic issues, such as the budget and the user experience [39]. As an overall finding, we can argue that no one single tool covers all needs and steps of big data processing [40]. A generally accepted conclusion in the literature is that no tool is better than the others are [38], and that users can select the adequate software package for data analytics only based on a critical analysis of the specific objectives and the application case [41].

5. Data analytics application domains

Due to its fast development, big data is rapidly expending in all science and engineering domains, as well as in physical, biological, and biomedical sciences. The domains of engineering- and product-associated big data processing (BDP) are behind the overall progress because of the shear fact of late recognition. Other issue is the rapid paradigmatic changes in the field due to the converging technologies and embedding software and cyber-ware in practically all products. These imply many changes that are already observable currently. First, engineered products are becoming more-and-more multifunctional, technology intensive, network connected, data dependent, and customized/personalized [42]. Products with these characteristics are often referred to as advanced or sophisticated products. However, the largest paradigmatic change is that they are rapidly becoming knowledge-intensive and smartly operating, or even progressing towards some forms of intelligent operation [43].

There is a debate concerning the relevance of big data processing to design application domains [44]. There are voices that big data exist only on paper and cannot provide immediate benefits for practical applications [45]. In the context of product development, BDP opened the opportunity of not only storing data concerning customers, but also to analyze large volume of data about their behaviors and customs [46]. For the reason that every discipline and application domain has a stake, big data became primordial for multidisciplinary problem solving [47]. In this case, the challenge is how it is possible to use data regardless the application domain [44].

The rapidly proliferating terms 'artifact intelligence' and 'service intelligence' still beg for a comprehensive clarification as well as for a more careful use in the contexts of product functionality and development. Until now, the concept of intelligent products has remained somewhat fuzzy and the use of the term is confusing [48]. It seems that there is a problem with the verbatim interpretation of the term 'intelligent' as well as with the relationships of intelligent products to knowledge acquisition and processing. This entails the need for further work considering the variety of application contexts. In addition, there is a need for a new classification of these products simultaneously considering the achieved level of intelligence and the specific manifestations of these levels [49]. On the other hand, researchers active in various fields of intelligent products tend to agree that there is still a long way to go before different kinds of machines and systems will be able to intelligently communicate with and understand each other, and reason meaningfully in structurally indefinite decision making situations [50]. Some of them believe that much more is needed than that typically provided by ontologies and semantic web-related technologies to be able to produce truly 'intelligent' tools [51].

6. Conclusions and future work

6.1. Conclusions

Based on a statistical and relational study of the literature, we derived a reasoning model, which identified and brought four domains of knowledge into an implicative interrelationship. Current data analytics should deal with data that are largely different from those processed digitally some decades ago. The major difference is not only in the cardinality of data, but also in the complexification of data. On the one hand, this creates new challenges for data analytics, but, on the other hand, it also creates opportunities for new value creation approaches. There seems to be a consensus in the literature concerning the fact that the complexity of big data cannot be properly addressed by the overwhelming majority of the existing (traditional) data processing methodologies and tools, and that exploiting the affordances of big data in various application contexts needs a stronger contextualization of the data transformation processes. The transformation techniques and tools are expected to support real time processing of data as well as the highest possible level of semantic interpretation of data. Time-dependent (and real-time) processing of complex data streams still raises many issues, in addition to the well-known issues of storing, visualizing and fusing heterogeneous data from multiple data sources.

It can be prognosticated that efficiency and reliability of data mining and knowledge discovery will remain the major issues for advanced big data analytics. Processing algorithms and mechanisms should be based on new underpinning theories, which allow dealing with the volume, the distribution, the cognitive complicatedness, and the dynamically changing characteristics of big data. The timed characteristic of big data does not seem to a significant obstacle, but the interplay among all aspects of big data does. In the context of future product and service development, some new sources of data like social media will offer new opportunities for designers to get insights into consumers' purchasing preferences, decisions and behavior, and to uncover information in context that is not possible with traditional product functionalities and life cycle data management approaches.

To put an end to the infinite problems of big data management and processing, some authors proposed specific solutions for a selection of tools and methodological approaches to coping with the complexity of big data in particular application domains. The investigation of these solutions (touched upon earlier in this paper) revealed the fact that the majority of authors are committed (if not attached) to real-time analysis of data and to developing powerful tools and better system architectures so that consumer durable making companies can realize value by understanding their operations, customers, distributors, and the marketplace as a whole.

Design application of advanced (semantic and smart) data analytics seems to be in a premature stage at the time of publishing this paper. As a combined effect of the proliferation of data analytics tools and the Internet of Things connectivity, companies gradually recognize the opportunities and try to convert them into business benefits. Deeply penetrating into real life industrial, social and human processes, products and services enabled by the paradigm of cyber-physical computing also need more sophisticated data processing and inferring capabilities, which are inseparable from their self.* functionalities. However, neither comprehensive methodologies nor dedicated toolboxes seem to be available to facilitate their endeavor.

6.2. Future work

Our on-going and future work concentrates on the development of a novel 'designerly' data analytics toolbox for product designers of a family of data-intensive products. Based on our completed study, we could hypothesize that the currently available data analytics tools miss semantic fusion concerning their output data and suffer from the lack of interpretation of data constructs in various design contexts. Because of these findings, we are conceptualizing a functional, architectural and information processing framework for a next-generation data analytics toolbox for white goods designers.

References

[1] K. Bodenhoefer, A. Schneider, T. Cock, A. Brooks, G. Sands, L. Allman and O. Delannoy, Environmental Life Cycle Information Management and Acquisition – First Experiences and Results from Field Trials, In: *Electronics Goes Green*, Berlin, 2004, pp. 5-8.

[2] M. Franke, P. Klein, L. Schröder and K.-D. Thoben, Ontological Semantics of Standards and PLM Repositories in the Product Development Phase, In A. Bernard et al. (eds.): *Global Product Development*, Springer, Berlin Heidelberg, 2011, pp. 473-482.

[3] A. Katal, M. Wazid and R. Goudar, Big data: Issues, Challenges, Tools and Good Practices. In: *Proceedings of the 6th International Conference on Contemporary Computing*, Noida, 2013, pp. 404-

409.

[4] S. Terzi, A. Bouras, D. Dutta, M. Garetti and D. Kiritsis, Product Lifecycle Management–from its History to its New Role, *International Journal of Product Lifecycle Management*, Vol. 4(4), 2010, pp. 360-389.

[5] W.F. van der Vegte, Taking Advantage of Data Generated by Products: Trends, Opportunities and Challenges, In: *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Charlotte, 2016, pp. V01BT02A025-V01BT02A025.

[6] D. Kiritsis and A. Rolstadås, Ubiquitous Product Lifecycle Management Using Product Embedded Information Devices, In: *International Conference in Intelligent Maintenance Systems, Arles, 2004*, pp. 65-78.

[7] S. Madden, From Databases to Big Data, In B. Blake et al. (eds.): *IEEE Internet Computing*, Vol. 16(3), 2012, pp. 4-6.

[8] D. Dietrich, *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley*, EMC Education Services, USA, 2015.

[9] H.-P. Kriegel, K.M. Borgwardt, P. Kröger, A. Pryakhin, M. Schubert and A. Zimek, Future Trends in Data Mining, *Data Mining and Knowledge Discovery*, Vol. 15(1), 2007, pp. 87-97.

[10] M.M. Fouad, N.E. Oweis, T. Gaber, M. Ahmed and V. Snasel, Data Mining and Fusion Techniques for WSNs as a Source of the Big Data, *Procedia Computer Science*, Vol. 65, 2015, pp. 778-786.

[11] A.J. Jara, D. Genoud and Y. Bocchi, Big Data for Cyber Physical Systems: An Analysis of Challenges, Solutions and Opportunities, In: *Proceedings of the 8th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, Birmingham, 2014, pp. 376-380.

[12] M.D. Assunção, R.N. Calheiros, S. Bianchi, M.A. Netto and R. Buyya, Big Data Computing and Clouds: Trends and Future Directions, *Journal of Parallel and Distributed Computing*, Vol. 79, 2015, pp. 3-15.

[13] P.S. Yu, On mining big data, In J. Wang et al. (eds.): *Web-Age Information Management*, *Lecture Notes in Computer Science, Vol. 7923*, Springer-Verlag, Berlin Heidelberg, 2013, p. XIV.

[14] H.-B. Jun, D. Kiritsis and P. Xirouchakis, Research Issues on Closed-loop PLM, *Computers in Industry*, Vol. 58(8), 2007, pp. 855-868.

[15] A. Bufardi, D. Kiritsis and P. Xirouchakis, Generation of Design Knowledge from Product Life Cycle Data, In A. Bernard and S. Tichkiewitch (eds.): *Methods and Tools for Effective Knowledge Life-Cycle-Management*, Springer, Berlin Heidelberg, 2008, pp. 375-389.

[16] A. Ericson, P. Müller, T. Larsson and R. Stark, Product-Service Systems–from Customer Needs to Requirements in Early Development Phases, In: *Proceedings of the 1st CIRP Industrial Product Service Systems Conference*, Cranfield, 2009, pp. 62-67.

[17] S.-J. Lee and E.-N. Huh, Shear-based Spatial Transformation to Protect Proximity Attack in Outsourced Database, In: *Proceedings of the 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, Melbourne, 2013, pp. 1633-1638.

[18] E. Rahm and H.H. Do, Data Cleaning: Problems and Current Approaches, *IEEE Data Engineering Bulletin*, Vol. 23(4), 2000, pp. 3-13.

[19] L. Berti-Equille, T. Dasu and D. Srivastava, Discovery of Complex Glitch Patterns: A Novel Approach to Quantitative Data Cleaning, In: *Proceedings of the 27th International Conference on the Data Engineering*, Hannover, 2011, pp. 733-744.

[20] L.A. Kurgan and P. Musilek, A Survey of Knowledge Discovery and Data Mining Process Models, *The Knowledge Engineering Review*, Vol. 21(01), 2006, pp. 1-24.

[21] B. Padmanabhan and A. Tuzhilin, A Brief-driven Method for Discovering Unexpected Patterns, In: *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining*, New York, 1998, pp. 94-100.

[22] S. Munir, J.A. Stankovic, C.-J.M. Liang and S. Lin, Reducing Energy Waste for Computers by Human-in-the-loop Control, *IEEE Transactions on Emerging Topics in Computing*, Vol. 2(4), 2014, pp. 448-460.

[23] D. Che, M. Safran and Z. Peng, From Big Data to Big Data Mining: Challenges, Issues, and Opportunities, In: *Proceedings of the International Conference on Database Systems for Advanced Applications*, Berlin Heidelberg 2013, pp. 1-15.

[24] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, The KDD Process for Extracting Useful Knowledge from Volumes of Data, *Communications of the ACM*, Vol. 39(11), 1996, pp. 27-34.

[25] N. Guarino, Formal Ontology in Information Systems, In: *Proceedings of the 1st International Conference on Ontology in Information Systems*, Trento, 1998.

[26] D.T. Tempelaar, B. Rienties and B. Giesbers, In Search for the Most Informative Data for Feedback Generation: Learning Analytics in a Data-rich Context, *Computers in Human Behavior*, Vol. 47, 2015, pp. 157-167.

[27] S. Ramsay, Special Section: Reconceiving Text Analysis: Toward an Algorithmic Criticism, *Literary*

and Linguistic Computing, Vol. 18(2), 2003, pp. 167-174.

[28] S. Berry, J. Levinsohn and A. Pakes, Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market, *Journal of Political Economy*, Vol.112(1), 2004, pp. 68-105.

[29] W.K. Michener and M.B. Jones, Ecoinformatics: Supporting Ecology as a Data-intensive Science, *Trends in Ecology & Evolution*, Vol. 27(2), 2012, pp. 85-93.

[30] D. Gorissen, K. Crombecq, I. Couckuyt and T. Dhaene, Automatic Approximation of Expensive Functions with Active Learning, In: *Foundations of Computational Intelligence, vol. 1*, Springer, Berlin, 2009, pp. 35-62.

[31] M. Chen, S. Mao and Y. Liu, Big Data: A Survey, *Mobile Networks and Applications*, Vol. 19(2), 2014, pp. 171-209.

[32] R. Roy, E. Shehab, A. Tiwari, T. Baines, H. Lightfoot, O. Benedettini and J. Kay, The Servitization of Manufacturing: A Review of Literature and Reflection on Future Challenges, *Journal of Manufacturing Technology Management*, Vol. 20(5), 2009, pp. 547-567.

[33] J. Gantz, and D. Reinsel, The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the Future*, 2012, pp. 1-16.

[34] A. Satyanarayana, Software Tools for Teaching Undergraduate Data Mining Course, In: *American Society of Engineering Education Mid-Atlantic Fall Conference*, Washington, 2013.

[35] R. Mikut and M. Reischl, Data Mining Tools, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 1(5), 2011, pp. 431-443.

[36] T. Pirttioja, *Applying Agent Technology to Constructing Flexible Monitoring Systems in Process Automation*, PhD dissertation, Faculty of Electronics, Communications and Automation, Helsinki University of Technology, 2008.

[37] S. Sagiroglu, and D. Sinanc, Big Data: A Review, In: *Proceedings of the International Conference on Collaboration Technologies and Systems*, San Diego, 2013, pp. 42-47.

[38] A.H. Wahbeh, Q.A. Al-Radaideh, M.N. Al-Kabi and E.M. Al-Shawakfa, A Comparison Study between Data Mining Tools over Some Classification Methods, *International Journal of Advanced Computer Science and Applications*, Special Issue on Artificial Intelligence, 2011, pp. 18-26.

[39] M.A. King, J. Elder, B. Gomolka, E. Schmidt, M. Summers and K. Toop, Evaluation of Fourteen Desktop Data Mining Tools, In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, San Diego,1998, pp. 2927-2932.

[40] K. Collier, B. Carey, D. Sautter and C. Marjaniemi, A Methodology for Evaluating and Selecting Data Mining Software, In: *Proceedings of the 32nd Hawaii International Conference on System Sciences*, Maui,1999, pp. 1-11.

[41] C. Giraud-Carrier and O. Povel, Characterising Data Mining Software, *Intelligent Data Analysis*, Vol. 7(3), 2003, pp. 181-192.

[42] C. Zheng, M. Bricogne, J. Le Duigou and B. Eynard, Survey on Mechatronic Engineering: A Focus on Design Methods and Product Models, *Advanced Engineering Informatics*, Vol. 28(3), 2014, pp. 241-257.

[43] X. Yang, O. Moore and S.K. Chong, Intelligent Products: From Lifecycle Data Acquisition to Enabling Product-related Services, *Computers in Industry*, Vol. 60(3), 2009, pp. 184-194.

[44] C. Bizer, P. Boncz, M.L. Brodie and O. Erling, The Meaningful Use of Big Data: Four Perspectives - Four Challenges, *ACM Sigmod Record*, Vol. 40(4), 2012, pp. 56-60.

[45] J. Li, F. Tao, Y. Cheng and L. Zhao, Big Data in Product Lifecycle Management, *The International Journal of Advanced Manufacturing Technology*, Vol. 81(1-4), 2015, pp. 667-684.

[46] M.J. Shaw, C. Subramaniam, G.W. Tan and M.E. Welge, Knowledge Management and Data Mining for Marketing, *Decision Support Systems*, Vol. 31(1), 2001, pp. 127-137.

[47] M. Brodie, M. Greaves and J. Hendler, Databases and AI: The Twain Just Met, In: *Proceedings of the 2011 STI Semantic Summit*, Riga, 2011, pp. 6-8.

[48] P. Valckenaers, B. Saint Germain, P. Verstraete, J. Van Belle, K. Hadeli and H. Van Brussel, Intelligent Products: Agere versus Essere, *Computers in Industry*, Vol. 60(3), 2009, pp. 217-228.

[49] G.G. Meyer, K. Främling and J. Holmström, Intelligent Products: A Survey, *Computers in Industry,* Vol. 60(3), 2009, pp. 137-148.

[50] K. Främling, J. Holmström, J. Loukkola, J. Nyman and A. Kaustell, Sustainable PLM through Intelligent Products, *Engineering Applications of Artificial Intelligence*, Vol. 26(2), 2013, pp. 789-799.

[51] J. Bézivin, H. Bruneliere, F. Jouault, and I. Kurtev, Model Engineering Support for Tool Interoperability, In: *Proceedings of the Workshop in Software Model Engineering - A MODELS 2005 Satellite Event*, Montego Bay, 2005, pp. 1-16.