SELECTED TOOLS FOR SUPPORTING DECISION MAKING IN THE ERA OF BIG DATA

[1]Radosław KAPLAN, [2]Katarzyna GDOWSKA, [2]Roger KSIAZEK

[1]AGH University of Science and Technology, Faculty of Management, Department of Management in Energetics, Cracow, Poland, EU, rkaplan@zarz.agh.edu.pl

[2]AGH University of Science and Technology, Faculty of Management, Department of Operations Research, Cracow, Poland, EU, kgdowska@zarz.agh.edu.pl, rksiazek@zarz.agh.edu.pl

Abstract

Decision making processes in modern industrial management benefit from big data processing. This paper presents selected tools for processing huge amounts of data, which can support decision makers in following areas: process control, industrial logistics, and urban public transportation planning. We begin with the Statistical Process Control (SPC) tools which can be used to control the entire production or service process. Dealing with the variability of large systems requires using statistical approach to prevent errors and to evaluate the quality of system performance. In the second part we concentrate on logistic flow planning, especially on tools for lot sizing and scheduling, which enable transforming aggregated data on the total demand into feasible and operable production schedules. It is an important aspect of manufacturing control and production management, as it can result in minimizing the total cost of set-up and processing and the total inventory cost. Finally, we focus on transport planning and management which plays a critical role in the development of modern cities and industry. We review the most popular approaches and tools for the transfer synchronization and the interval synchronization, and we discuss their capability to process big data. This paper contributes with the review of analytical methods conducted from the perspective of their usefulness for big data analysis. We show that tools commonly used in managerial practice in process control, industrial logistics, and urban public transportation planning can support decision making process.

Keywords: Big Data, decision making, scheduling, industrial logistics

INTRODUCTION

Nowadays, decision making processes in modern industrial management benefits strongly from processing huge amounts of data generated by devices used in manufacturing and logistics, for example smart manufacturing is a data-driven approach to real-time decision making support [1]. Recent advances in the information and measurement technologies result in higher and higher precision of sensors, which can communicate with each other and create the Internet of Things (IoT) in the industry. Because of the exponential growth of the amount of data generated within the IoT decision makers in manufacturing and logistics need to use so-called Big Data as a source of information on which decision making processes are based. This paper presents selected traditional tools for processing huge amounts of data in order to formulate recommendations for decision makers in three areas: process control, lot sizing and scheduling, and urban public transportation planning and management.

The term ‘Big Data’ is used for large sets of heterogeneous data: structured data, unstructured data, and semi-structured data [2]. Definitions of Big Data usually include 3 main features (3Vs): volume, velocity, and variety [3]. Volume refers to huge amount of data continuously generated by devices and applications. The amount of generated, replicated and used data was estimated to be to ca. 8 ZB in 2015 [4] and in 2020 it is supposed to increase to ca. 40 ZB [5]. Velocity is the speed of data generating, processing, and analyzing, so that they can provide useful information [2]. Variety of data results from the variety of their sources (public or private, complete or partial, centralized or distributed, and of different forms - e.g. movies, documents, notes, signals, etc.) and the ways how they are generated (structured and unstructured).

The remainder of this paper is as follows. In Section 2 we focus on the Big Data in manufacturing, where they are especially useful in process control and quality management. We show how we can use Statistical Process Control for dealing with Big Data processing. Section 3 is devoted to the lot-sizing and scheduling problems. The next section reviews combinatorial optimization oriented methods for transportation planning and management. The main contribution of this paper is the review of selected analytical methods, which are commonly used in practice and discussion of their usability for the Big Data analysis from the perspective of support decision making.

1. BIG DATA IN QUALITY MANAGEMENT

In manufacturing industry all the decision making processes have to be based on hard facts, hence control and monitoring of the selected features of products and processes have always been the key issue of the quality management. Recent technological advances made data collection less expensive and time consuming and in the era of the Internet of Things this process demands minimal effort [1]. In every type of organization we can observe growing tendency of applying approaches which originate from the Total Quality Management (TQM). As a consequence Statistical Process Control (SPC) - well-known approach to quality management [6-8] - becomes more and more popular as a method supporting processes management in the high series manufacturing. The usefulness of the SPC and its applicability in practice are described in [9]: „…there is variation in the characteristics of manufactured articles.... If this variability is considerable it is impossible to predict the value of the characteristic of any single item. Using statistical methods, however, it is possible to take meager knowledge of individual items and turn it into meaningful statements which may then be used to make decisions about the process or batch of products."

Each real process is variable, so it is necessary to use statistical tools to simulate and predict how specific features of a product may behave in time. The SPC tools enable us to monitor particular characteristics of both the process and the product (or the service) and obtained data are used to control the course of the process [9]. The objective is to monitor the variability of the process, so that errors in manufacturing can be avoided and the number of eliminated faults and defects can be reduced. The reason for avoiding defects in high series manufacturing is an exponential dependence between the moment of fault detecting and the costs of its rectification. The costs of elimination of faults and defects can vary depending on the moment when total inspection or skip-lot control was conducted. The costs can differ by 10-100 times [10].

From the Big Data perspective SPC - and in consequence smart manufacturing - benefits from data gathering, since this methodology provide tools for detecting discrepancies from accepted values of variability factors quickly and easily. We should remember that these discrepancies are not always caused by worsening of monitored factors; sometimes they may result from improvements or other changes to which an enterprise needs to react. Once we use SPC to analyze bigger and bigger amounts of data, we can understand better variability of the main processes and improve control. Moreover, we can figure out what factors influence mostly the main processes and the entire enterprise. The challenges faced in the enterprise during adapting the SPC to operate on the Big Data are the same as in case of any other tool. A great advantage of the SPC are both the existing infrastructure and popularity and well-understanding of the SPC rules among employees of manufacturing companies.

2. BIG DATA IN INDUSTRIAL LOGISTICS

The keys to success in industrial logistics and management are planning and scheduling of supply deliveries and shipping of final products. Nowadays, manufacturers constantly receive a great number of highly customized orders; hence it is necessary to gather a huge amount of data and to process information quickly, so that enterprises can quickly react to new orders and breakdowns by adjusting schedules to new conditions. Setting delivery and shipment dates depends mostly on the production schedule, which defines dates and order volumes of each material and semi-manufactured product and also due dates for each production stage. Industrial logistics includes lot-sizing and scheduling, since it provides lot sizes of each product as well as setup and processing times. The objective of the lot-sizing and scheduling problem is to minimize the total production costs, which includes the cost of generating production lots and inventory costs. When lots are relatively big, their number is reduced, so the setup cost is also reduced, but inventory costs increase. In contrary, small-sized lots result in the decrease of the inventory costs, but in the same time the setup costs and the costs of generating production lots increase. When solving the lot-sizing and scheduling problems we try to find the trade-off between these two types of costs.

The lot-sizing and scheduling problems origin from the order quantity planning problems; however similarity between them is gone as we include more and more detailed requirements of planning. Both the order quantity planning problems and for the lot-sizing and scheduling ones are solved using exact methods (based on mixed-integer programming) and heuristics. The most important issue to overcome is the variability of demand. From the very beginning of the research on this issue [11], planners tried to include demand fluctuations. The very first model, Economic Order Quantity (EOQ), assumes that demand can differ between planning periods, but in a given planning period is fixed [11]. The same assumption is adopted in a simple heuristics (Silver-Meal heuristic) [12] and an exact method based on the dynamic programming (Wagner-Within algorithm) [13]. All the aforementioned methods have significant limitation, but their huge advantage is to deal with lot-sizing and scheduling in relatively short time, hence they were used to prepare initial production schedules as well as to re-schedule production in case of an unexpected event.

Technological advances made it possible to solve complex lot-sizing and scheduling problems with heterogonous machines and products. However, we still need to include lot order, setup and processing times, demands at each planning period as well as the availability of machines and staff [14-17]. Due to constantly received orders, planning periods become shorter and shorter, so the shift from deterministic lot-sizing and scheduling methods towards dynamic ones is needed. Lot-sizing and scheduling problems are NP-hard [18], so computing time for exact methods increases exponentially with the size of instances. Newly-developed heuristics, which are capable to use the Big Data and generate feasible solution in real time [16,19,20] are highly desirable assets for smart manufacturing.

3. BIG DATA IN TRANSPORTATION PLANNING AND MANAGEMENT

Robustness of a transportation system is assessed with different criteria, therefore a multi-criteria analysis based on the Big Data provides a powerful supporting tool for decision making in transport planning and management. Strategic and operational planning and timetabling is based on exact methods applied to deterministic problems [21-23]. For instance, in public transportation planning deterministic travel times over a given period of the day are considered - it may be minimal or average travel times between two subsequent nodes in a route in a given period of the day. Such an approach results in reducing computation time; however, fixed travel times may affect synchronization. Detailed reviews of works dedicated to public transport system planning and management are presented in [24-26].

As far as utilization of the Big Data in transportation planning and management is concerned, the main issues of its applicability are the objective of synchronization and solving methods used. The most popular optimization criteria for synchronization problems are: minimizing the average waiting time [e.g. 27-30], minimizing the total travel time [e.g. 31], maximizing the total number of synchronizations [e.g. 25, 32-35], maximizing the number of direct transfers [30], or minimizing the coefficient of variation of intervals between subsequent arrivals [36]. Among frequently used exact methods can be listed: non-linear programming [27], mixed-integer programming [28,29,32], mixed-integer linear programming [25,33-35], integer linear programming [32], mixed non-linear programming [31], and multi-criteria linear programming [36]. In order to deal with the computational complexity of exact models heuristic methods of different types are developed, e.g. genetic algorithms [27, 31], Lagrangian based heuristics [29], constructive heuristics [33], local search algorithms [34], or multi-start variable neighborhood search algorithms [35].

Exact methods proposed for transportation planning and management are based on combinatorial optimization, where computational complexity is a significant issue and it is correlated with the size instances. In other words, computation time increases with the number of trips to be synchronized and it is long even for relatively small instances. It is acceptable, because synchronization is included in timetable generating process and long-term timetables are not prepared every day, so planners can wait for optimal solutions. Nowadays, in public transportation systems and logistics there is a need to adapt schedules dynamically to current conditions, e.g. traffic jams, road accidents, or vehicle breakdowns. It can be possible only if huge amounts of data are collected from the system (which already happens) and are processed in real time, so that they can be used by decision making tools. Therefore, both exact methods and heuristics for generating timetables are developed. There are several approaches to the reduction of computation time. For instance, planning horizons are divided into small planning periods, so that generated instances are solvable with available computing power and memory [36]. Another approach is to develop and implement valid inequalities for timetabling problems [34]. Moreover, metaheuristics for interval synchronization problems are being developed, so that timetables can be re-scheduled on-the-fly [e.g. 33,34].

4. CONCLUSION

In this paper selected tools and methods used commonly in the industry are reviewed from the perspective of possible utilization for the Big Data processing. Nowadays, decision making in the industry more and more frequently faces the problem of processing and analyzing huge amounts of data obtained from process monitoring. We should remember that multithreading and complexity of the problems determine the choice of methods and tools supporting decision making processes, both operational and strategic ones.

Dealing with variability of each large system requires using statistical approach to prevent and monitor errors as well as to evaluate the quality of system performance. Using the Statistical Process Control (SPC) tools we can monitor a single product or a service, but obtained information can be used to control the entire production chain. These tools are commonly used in practice, but in this paper we concentrate on their usefulness for Big Data processing.

Logistic flow planning is an important aspect of manufacturing control and production management, which needs to take into account the dates of material supply and re-supply, the earliest possible dates of parts and final products to be ready, and the due-dates of orders. As logistics flow planning includes re-scheduling and adjusting to the up-to-date conditions it requires processing huge amount of data, so that production schedules can be optimized. Tools used for lot-sizing and scheduling enables us to transform aggregated data on the total demand into a feasible and operable production schedule. The objective is to minimize the total cost of set-up and processing and the total inventory cost. In this paper we review the most popular tools for lot-sizing and scheduling from the perspective of their effectiveness in the Big Data processing.

Public transport plays a critical role in the development of modern cities. Urban public transport planning and management includes route and line planning, generating timetables, vehicle scheduling, and crew rostering. In this paper we focus on analytical methods used for decision support in timetabling synchronization process. We review the most popular approaches and tools for the transfer synchronization and the interval synchronization, and we discuss their capability to process the Big Data and to act as effective tools for support decision making in this area.

The majority of models and methods for transportation planning and management are deterministic, however nowadays a strong tendency to shorten planning periods is observed. In transportation and logistics the Big Data processing can contribute to the robustness of transportation service. Basing on the output of the Big

Data processing traffic dispatchers can adjust bus speed to current traffic conditions, so that the timetable can be executed without delays. In case of rescheduling necessity, new timetable can be generated in real-time and both bus drivers and passengers can be kept updated via the Internet and mobile devices and electronic displays at bus stops.

ACKNOWLEDGEMENTS

This work was supported by AGH University of Science and Technology subject subsidy for maintaining research potential - grants No. 15/11.200.357, 15/11.200.356, and 15/11.200.360.

REFERENCES

[1] HE, Q. Peter and WANG, Jin. Statistical process monitoring as a big data analytics tool for smart manufacturing. *Journal of Process Control*. 2018. vol. 67, pp. 35-43.

[2] OUSSOUS, Ahmed, BENJELLOUN, Fatima-Zahra, LAHCEN, Ayoub Ait and BELFKIH, Samir. Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*. 2018. vol 30, no. 4, pp. 431-448.

[3] FURHT, Borko and VILLANUSTRE, Flavio. *Introduction to Big Data*. In: Big Data Technol. App. Cham: Springer International Publishing, 2016, pp. 3-11.

[4] RAJARAMAN, Vaidyeswaran. Big data analytics. *Resonance*. 2016. vol. 21, pp. 695-716.

[5] KUNE, Raghavendra, KONUGURTHI, Pramodkumar, AGARWAL, Arun, CHILLARIGE, Raghavendra Rao and BUYYA, Rajkumar. The Anatomy of Big Data Computing. *Software: Practice and Experience*. 2016. vol. 46, pp. 79-105.

[6] FEIGENBAUM, Armand V. *Total Quality Control*. 3rd ed. New York: McGraw-Hill, 1983.

[7] JURAN, Joseph M. *Quality Control Handbook*. 3rd ed. New York: McGraw-Hill, 1975:

[8] OAKLAND, John S. and FOLLOWELL, Roy R. *Statistical Process Control*. 2nd ed. Oxford: Butterworth-Heinemann, 1990.

[9] OAKLAND, John S. Statistical Process Control. 6th ed. London: Heinemann, 2008.

[10] KAPŁAN, Radosław, SAŁUGA, Piotr, KAMIŃSKI, Jacek and GRZESIAK, Paweł., Koncepcja wykorzystania statystycznego sterowania procesem w sektorze energetycznym. *Rynek Energii*. 2017. vol. 6.

[11] HARRIS, Ford W. How Many Parts to Make at Once. *Factory, The magazine of management*. 1913. vol. 10, pp. 135-136.

[12] SILVER, Edward A. and MEAL, Harlan C. A heuristic for selecting lot size quantities for the case of a deterministic time-varying demand rate and discrete opportunities for replenishment, *Production and Inventory Management*. 1973. vol. 14, no. 2, pp. 64-74.

[13] WAGNER, Harvey M. and WHITIN, Thomson M. Dynamic version of the economic lot size model. *Management Science*. 1958. vol. 5, pp. 89-96.

[14] BRAHIMI, Nadjib, DAUZÈRE-PÉRÈS, Stéphane, NAJID, Najib M. and NORDLI, Atle. Single item lot sizing problems. *European Journal of Operational Research*. 2006. vol. 168, no. 1, pp. 1-16.

[15] DREXL, Andreas and KIMMS, Alf. Lot sizing and scheduling - Survey and extensions. *European Journal of Operational Research*. 1997. vol. 99, no. 2. pp. 221-235.

[16] GDOWSKA, Katarzyna and KSIĄŻEK, Roger. The structure of the genetic algorithm of lot-sizing and scheduling problem formulated as capacitated lot sizing problem. *Logistyka*. 2012. vol. 2, pp. 559-565.

[17] KARIMI, Behrooz, GHOMI, Fatemi S.M.T. and WILSON, John M. The capacitated lot sizing problem: a review of models and algorithms. *Omega*. 2003. vol. 31, no. 5, pp. 365-378.

[18] JANS, Raf and DEGRAEVE, Zeger. Meta-heuristics for dynamic lot sizing: A review and comparison of solution approaches. *European Journal of Operational Research*. 2007. vol. 177, no. 3, pp. 1855.

[19] HAASE, Knut. *Lotsizing and scheduling for production planning*. In: BECKMANN, Matthew and KUNZI, Hans-Peter, eds. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, 1994.

[20] KSIĄŻEK, Roger and GDOWSKA, Katarzyna. Algorytmy heurystyczne wyznaczania wielkości produkcji dla znanego harmonogramu przezbrojeń dla zadania planowania wielkości i szeregowania partii produkcyjnych. *Logistyka*. 2014. vol. 4, pp. 4597-4608.

[21] GDOWSKA, Katarzyna and KSIAZEK Roger. *Choice of optimisation criteria for the timetabling problem in city public transportation for tram lines sharing intermediate stops*. In: ŁEBKOWSKI, Piotr, ed. Aspects of production engineering and management. Cracow: AGH University of Science and Technology Press, 2011, pp.107-121.

[22] ORTÚZAR, Juan and WILLUMSEN, Luis G. *Modelling Transport*. Chichester-New York-Weinheim-Brisbane-Toronto-Singapore: John Wiley and Sons Ltd., 2001.

[23] SCHOBEL, Anita. A model for the delay management problem based on mixed-integer-programming. *Electronic Notes in Theoretical Computer Science*. 2001. vol. 50, no.1, pp. 1-10.

[24] GUIHAIRE, Valerie and HAO, Jin-Kao. Transit network design and scheduling: a global review. *Transportation Research Part A: Policy and Practice*. 2008. vol. 42, no. 10, pp. 1251-1273.

[25] IBARRA-ROJAS, Omar J. and MUÑOZ, Juan C. Synchronizing different transit lines at common stops considering travel time variability along the day. *Transportmetrica A: Transport Science*. 2015. vol. 12, no. 8, pp. 751-769.

[26] POORJAFARI, Vahid and YUE, Wen Long. Synchronisation of Public Transport Services: Comparison of Methods. In: Proceedings of the Eastern Asia Society for Transportation Studies [on-line]. 2013. [viewed 2014-12-07]. Available from: http://easts.info/on-line/proceedings/vol9/PDF/P237.pdf.

[27] ADAMSKI, Andrzej. Real-time computer aided adaptive control in public transport from the point of view of schedule reliability. In: DADUNA, Joachim R., BRANCO Isabel and PAIXÃO, Jose M.P., eds. Computer-aided transit scheduling: proceedings of the Sixth international workshop on Computer aided scheduling of public transport: Lisbon, Portugal, July 1993. Lecture Notes in Economics and Mathematical Systems 430. Berlin-Heidelberg-New York: Springer-Verlag, 1995, pp. 278-295.

[28] BRUNO, Giuseppe, IMPROTA, Gennaro and SGALAMBRO, Antonino. Models for the schedule optimization problem at a public transit terminal. *OR Spectrum*. 2009. vol. 31, pp. 465-481.

[29] CASTELLI, Lorenzo, PESENTI, Raffaele and UKOVICH, Walter. Scheduling multimodal transportation systems. *European Journal of Operational Research*. 2004. vol. 155 no. 3, pp. 603-615.

[30] CEDER, Avishai, HADAS, Yuval, McIVOR, Michael and ANG, Albert. Transfer Synchronization of Public-Transport Networks. *Transportation Research Record: Journal of the Transportation Research Board*. 2013. vol. 2350, no. 1, pp. 9-16.

[31] CHAKROBORTY, Partha., DEB, Kalyanmoy and SUBRAHMANYAM, Prakya S. Optimal scheduling of urban transit systems using genetic algorithms. *Journal of Transportation Engineering*. 1995. vol. 121, pp. 544-553.

[32] CEDER, Avishai, GOLANY, Boaz and TAL, Ofer. Creating bus timetables with maximal synchronization. *Transportation Research Part A: Policy and Practice*. 2001. vol. 35, no.10, pp. 913-928.

[33] ERANKI, Anitha. A model to create bus timetables to attain maximum synchronization considering waiting times at transfer stops. MA thesis. Florida: University of South Florida, Department of Industrial and Management System Engineering, 2004.

[34] IBARRA-ROJAS, Omar J. and RIOS-SOLIS, Yasmin A. Synchronization of bus timetabling. *Transportation Research Part B*. 2012. vol. 46, no.5, pp. 599-614.

[35] IBARRA-ROJAS, Omar J., LOPEZ-IRARRAGORRI, Fernando and RIOS-SOLIS, Yasmin A. Multiperiod Bus Timetabling. *Transportation Science*. 2015, Articles in Advance, pp. 1-18.

[36] GDOWSKA, Katarzyna. Models and methods for harmonization of urban public transport timetables in a network with radial routes. PhD thesis. Krakow: AGH University of Science and Technology, Faculty of Management, 2016.