Methods of Data Mining for Quality Assurance in Glassworks

Łukasz Paśko, Pawel Litwin

HAL Id: hal-02478736
https://inria.hal.science/hal-02478736

Submitted on 14 Feb 2020

Methods of Data Mining for Quality Assurance in Glassworks

Łukasz Paśko and Paweł Litwin

Faculty of Mechanical Engineering and Aeronautics, Rzeszow University of Technology,
Al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland
{lpasko, plitwin}@prz.edu.pl

Abstract. In manufacturing enterprises implementing the idea of Industry 4.0, devices that generate data are increasingly used. Over time, huge data sets are created. These collections, known as Big Data, are very important to the company because they can contain valuable information. One of the goals of today's enterprises is to discover this information and transform it into knowledge. The aim of the article is to present the methodology of exploration of large data sets from the manufacturing process in glassworks. The result of the research is knowledge about the parameters of the manufacturing process causing defects in the products.

Keywords: big data, data mining, quality assurance, glass industry.

1 Introduction

Today's manufacturing companies increasingly use different types of devices that generate data. This data may contain crucial information. An important area of modern enterprises' activities is discovering this information. Such task requires processing a large amount of data. Then the identified information is transformed into knowledge understood as information that has been confirmed and can be used to support decision-making [1].

The crucial stage of knowledge discovery process is data mining (DM). DM is defined as the automatic search for unknown dependencies and patterns hidden in the data. The detected information should then be presented to the user in an intelligible form, e.g. in the form of logical rules or visualizations [1].

In the research literature on production engineering, an increasing number of DM applications has been observed for several years. Examples include: scheduling and production planning [2], customer needs research [3], machine failure prediction [4], improvement of the assembly process [5], supply chain management [6] or product design [7].

Also in collaborative networks are places where DM methods can be used. A collaborative network has the potential to collect huge amount of data about its collaborative activities. For this reason, there is a need to extract significant patterns out from data and utilize them in a collaborative network. The authors of the paper [8] propose a framework for DM in the design of collaborative virtual environments.

Modeling of collaboration networks is also the topic of the work [9]. The aim of the paper is exploration of the structural and the dynamic features of collaboration to demonstrate that there is a unified modeling approach for reproduce these features in different domains. Proposed model allows understand collaboration patterns in two different domains: research and development alliances between firms, and co-authorship relations between scientists. In the work [10], the authors draw attention to prediction of collaborative relationships in the form of collaboration platform where firms can address works to solve problems together. A similar topic is discussed in the paper [11], which presents an analytic framework for managing extended networks in supply chains.

The authors of the work attempt to demonstrate the suitability of using DM methods to ensure the quality of products and propose the methodology for analyzing industrial data in glassworks.

2   Data Analysis Methods Applications to Quality Assurance

Many applications of data analysis in manufacturing enterprises are related to the product quality. A large part of the research concerns the influence of input parameters of the manufacturing system on the quality of products. Usually the data analyzed includes parameters of the manufacturing process (MP), properties of materials used, employees (e.g. seniority, age), equipment and tools status (e.g. number of days since the last failure, number of hours worked), and the environment of the production system (e.g. atmospheric conditions, day of week).

Many publications using DM in the field of quality assurance come from the metallurgical industry. The aim of the research [12] was to use neural networks and multiple regressions to predict the quality of galvanized steel. Paper [13] presents models of neural networks that predict the influence of process parameters on the properties of products manufactured in a pressure die-casting process. The article [14] presents the use of Kohonen's neural networks to identify key factors affecting the quality of steel after the rolling process. Work [15] shows that data recorded with automatic sensors in the cutting process can be used to monitor the quality of products. In [16], Bayesian neural networks were used to determine the appropriate parameters of the sintering process. In [17], grouping algorithms, similarity measures, and distance measures were used to improve production schedules. In work [18], a method for determining the correlation between a sequence of machines and the quality of products is proposed, based on association rules.

All the above-mentioned works focus on a strictly defined field of industry and describe only selected cases of the use of DM in the context of product quality assurance. There are no solutions in the literature describing the use of DM techniques that comprehensively cover all stages of manufacturing.

In addition, it should be noted that the applications of DM in the area of quality assurance are mainly focused on the processing of metals, and the glass industry is an area in which there is a small number of studies [19].

The aim of this article is to present the concept of methodology for the analysis of large data sets from a production system in glassworks. The result of the research is

expected to be the knowledge about the impact of MP parameters on the occurrence of defects in the products.

3   Problem Statement and Data Source Description

Analysis of industry data is a non-trivial task, because of the attributes that can be assigned to this data. The Big Data concept defines them as 3V: volume, velocity, variety. Volume refers to a large amount of data, velocity concerns the speed of data inflow and analysis, and variety indicates the heterogeneity of data. An additional difficulty is to determine the technique of extracting knowledge that will best suit to a given problem. These difficulties indicate the need to develop some general methods and guidelines that will be useful for researchers. A glassworks factory is an example of a company where large amounts of data are generated. There are 4 glass furnaces and 14 production lines in the analyzed glassworks. The daily production volume is over 5 million. The manufacturing is continuous, and as a result, data regarding process parameters are recorded 24 hours a day, 7 days a week. Data sets created in this way may be used when there are problems with ensuring high quality of products.

In the glassworks, at automatic quality control stations, an anomaly was recorded. The anomaly involves the occurrence of periods in which an increased number of defective products is observed. Periods of the increased number of defective products last from a few to over a dozen days. During this time, the number of defective products is about three times higher than the average. The abovementioned anomaly is that air bubbles occur in the product's head. The bubbles appearing on the surface of the seal are particularly dangerous; product with the defect becomes useless. It should be noted that with the use of the statistical data analysis methods, glassworks' employees were not able to determine the reason for the periods of increased defectiveness of the products. The aim of the research is to identify the parameters that are responsible for periods of increased number of products with the defect. Because the glassworks collects extensive datasets containing the values of hundreds of MP parameters, there was decided to find the reason of increased number of defective products using DM methods.

The quantity and dynamics of the collected data (values of most parameters are registered every second) justifies the use of artificial intelligence and machine learning methods. In addition, the research issue also includes the appropriate presentation of knowledge, so that it is understandable for decision-makers. The knowledge base, which is an integral part of the decision support system (DSS), should contain decision rules acquired by appropriate methods of induction.

Data from the production process monitoring system were recorded within a period of 27 days and can be divided into three groups:
1. The number of products with the defect recorded on three manufacturing lines.
2. Parameters characterizing the MP, concerning: operation of the glass furnace (glass level and temperature, electric and gas heating power); work of three forehearths (glass temperature and heating); cooling of moulds.

3. Meteorological data describing weather conditions outside the production hall, such as air temperature, atmospheric pressure and humidity.

Data regarding the number of defective products are recorded at quality control stations located at the so-called cold end of the production line. Based on them, three variables were defined, which will be explained (output) variables during the test.

Parameters of the MP, together with meteorological data, will be used as explanatory (input) variables. The first device from the hot end is a glass furnace. Based on data obtained from sensors located in the furnace, 11 variables were defined, that describe the operation of the furnace and ventilation, as well as the level and temperature of the liquid glass. The next device in the MP is a forehearth, which receives liquid glass from the glass furnace. Glass moving through forehearth gets the right consistency and temperature. The acquired forehearth parameters refer to the air pressure, the position of the air supply valves and the temperature of the glass. At the end of forehearth, liquid glass is divided into portions (so-called gobs). A total of 139 parameters were registered in the forehearths. Gobs are moving to moulds that give them the shape of the product. Another group of acquired parameters describes the cooling of moulds, where nine parameters were registered. The number of registered parameters is shown in Table 1.

Table 1. The set of explanatory variables

The number of parameters used as explanatory variables				
Glass furnace	Forehearths	Moulds cooling	Meteorological data	TOTAL
11	139	9	3	162

4   Research Methodology and Preliminary Results

In the paper [20], it was shown that DM methods can be used to generate one of four types of information:
– logical rules based on IF (conditions) THEN (conclusion) construction;
– relative significance of explanatory variables expressing their impact on the explained variable;
– predicted values of explained variables;
– results of grouping explanatory variables.

The analysis focuses on obtaining the four types of information, enabling the understanding of the relationship between the parameters of the production process and the number of defective products. Figure 1 presents the assumed research plan, leading to the automatic acquisition of knowledge from production data, and then to develop a decision support system in the quality assurance process. Data selection is aimed at choosing the right data that may potentially contain knowledge about the problem being studied. Preparation for analysis includes handling of missing data and outliers as well as generation of basic statistics describing the data. The principal component analysis (PCA) in the literature is most often used as a method of reducing dimensionality, i.e. reducing the number of explanatory variables.
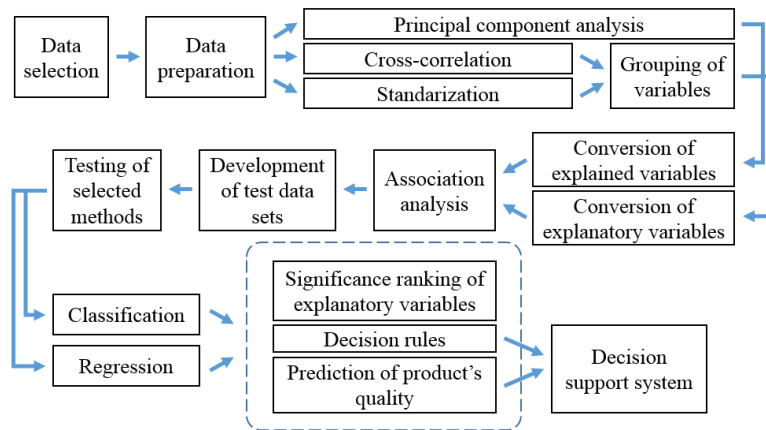
Fig. 1. Proposed research plan

In the proposed approach, PCA plays a slightly different role - it is used for grouping observations. The expected results of PCA are the principal components, which are a linear combination of explanatory variables, and graphs presenting the view of observations (cases) on the plane of the principal components. The graphs will allow to identify groups (clusters) of observations. The clusters will help in the implementation of the next stage - the conversion of explained variables. An example of the application of this approach is the analysis of parameters obtained from one of the manufacturing lines. The eigenvalues of the 20 principal components are presented in the scree plot shown in Figure 2.
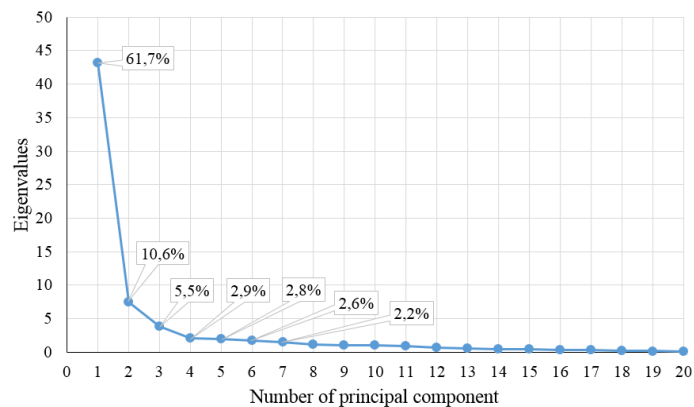

Fig. 2. A scree plot for 20 principal components

The selection of the principal components for further analysis assumes determining the division point at the place where the eigenvalues of subsequent components cease to decrease significantly. For the further analyzes the principal components number 1

to 4 are selected. It is worth noting that the percentage values of the four main components cover in total over 80% of the variability of the explained variables. This example shows the high usefulness of PCA in application to the reduction of the dimensionality of the analyzed dataset.

For each pair of four principal components, observations were projected onto the plane. The points shown in the charts are individual cases from the data set. Charts in which component No. 1 is present contain clear clusters of observations (Figure 3).
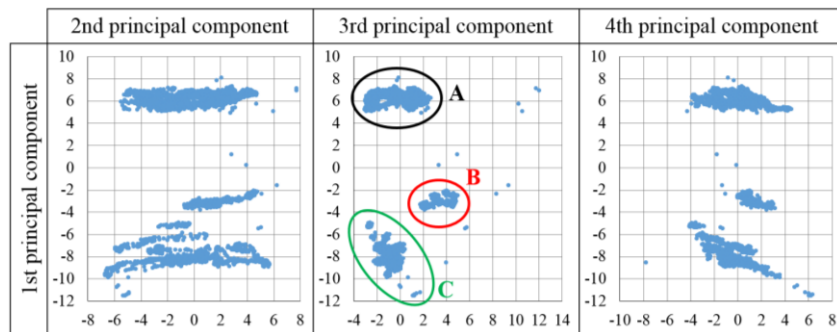

Fig. 3. Observations for selected pairs of four principal components

Particularly compact clusters of observations appear in the graphs of components No. 1 and 3, where they are marked with letters A, B and C.

The next stage of the analysis is the conversion of explained variables. At this stage of the research, data cases belonging to clusters A, B and C were projected onto a chronological graph showing the number of products with the defect (Figure 4).
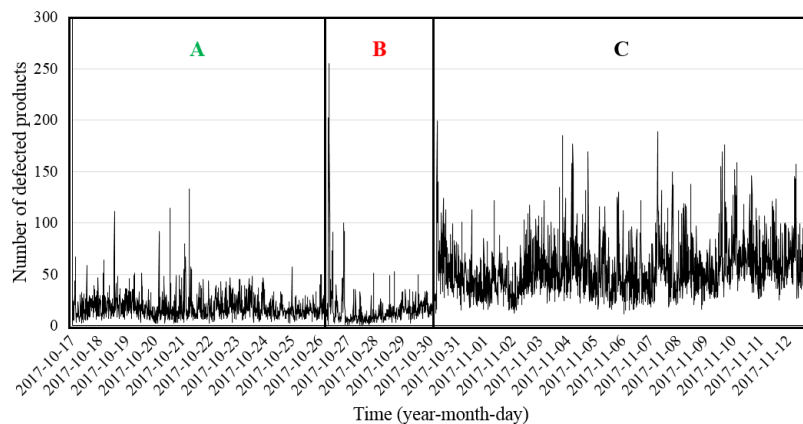

Fig. 4. The diagram of the number of defected products with identified classes

It turns out that data cases belonging to cluster A are at the beginning of the graph, cases from cluster B are in the middle, and the final part of the chart is occupied by

cases from cluster C. Therefore, the clusters found can be used to separate observation classes on the graph of the number of defective products, i.e. to convert a quantitative variable into a qualitative one. In addition to the "increased number of defective products" and "acceptable number of defective products" corresponding to clusters C and A, the PCA suggests also taking into account the third "predicting" class, covering a period of a few days before the sudden increase in the number of defective products.

The research carried out so far allows the first conclusions. The PCA showed the variables, which have the highest impact on the data set's variability. These are temperatures of liquid glass, whose values were measured in three zones of the forehearth. The variables are the main candidates for recognition as a cause of products' defects, because the clusters obtained from PCA are compatible with the variability of the number of defected products, what is depicted in the Figure 4.

5   Summary

The paper presents a DM application for acquiring knowledge from large data sets in the glassworks. The article proposes a proprietary research plan aiming to develop a decision support system in the quality assurance. Part of the plan, which was implemented, allowed for identification of the parameters of the production process responsible for the increased defectiveness of the products. Preliminary results have been positively verified in the glassworks. The implementation of the full range of research (Figure 1) will allow to build and test a DSS for quality assurance. Both the research problem and the proposed solution method can be successfully used in education in the field of Industry 4.0 and Big Data concepts. The plan of further work envisages extending the data analysis to a collaboration network including glassworks' suppliers (materials, industrial automation systems, energy) in order to obtain a more complete picture of the parameters affecting the glass MP.

References

1. Larose, D.T.: Odkrywanie wiedzy z danych: wprowadzenie do eksploracji danych. Wydawnictwo Naukowe PWN, Warsaw (2006)
2. Wang, Y., Zhang, Y., Yu, Y., Zhang, C.: Data Mining Based Approach for Jobshop Scheduling. In: Qi, E., Shen, J., Dou, R. (eds.) Proceedings of 2013 4th International Asia Conference on Industrial Engineering and Management Innovation (IEMI2013). Springer, Berlin, Heidelberg (2014)
3. Kittidecha, C., Yamada, K.: Application of Kansei Engineering and Data Mining in the Thai Ceramic Manufacturing. Journal of Industrial Engineering International, No. 14 (4), pp. 757--766 (2018)
4. Ruschel, E., Santos, E.A.P., Loures, E.F.R.: Establishment of Maintenance Inspection Intervals: An Application of Process Mining Techniques in Manufacturing. Journal of Intelligent Manufacturing, pp. 1--20, Springer, Berlin, Heidelberg (2018)

5. Knosala, R.: Zastosowania metod sztucznej inteligencji w inzynierii produkcji. WNT, Warsaw (2002)
6. Akhtar, M.D., Manupati, V.K., Varela, M.L.R., Putnik, G.D., Madureira, A.M., Abraham, A.: Manufacturing Services Classification in a Decentralized Supply Chain Using Text Mining. In: Abraham, A., Muhuri, P.Kr., Muda, A.K., Gandhi, N. (eds.) Hybrid Intelligent Systems. HIS 2017. Advances in Intelligent Systems and Computing, vol. 734, pp. 186--193, Springer, Cham (2018)
7. Giess, M.D., Culley, S.J.: Investigating Manufacturing Data for Use within Design. DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design, Stockholm (2003)
8. Biuk-Aghai, R.P., Simoff, S.J.: An Integrative Framework for Knowledge Extraction in Collaborative Virtual Environments. In: Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work, pp. 61--70, New York (2001)
9. Tomasello, M.V., Vaccario, G., Schweitzer, F.: Data-driven modeling of collaboration networks: a cross-domain analysis. EPJ Data Science, vol. 6 (2017)
10. Zuo, Y., Kajikawa, Y.: Prediction of collaborative relationships by using network representation learning. 2017 IEEE International Conference on Systems, Man, and Cybernetics, pp. 69--74 (2017)
11. Cox, A., Sanderson, J., Watson, G.: Supply Chains and Power Regimes: Toward an Analytic Framework for Managing Extended Networks of Buyer and Supplier Relationships. Journal of Supply Chain Management, vol. 37, 28--35 (2001)
12. Mere, J.B.O., Marcos, A.G., Gonzalez, J.A., Rubio, V.L.: Estimation of mechanical properties of steel strip in hot dip galvanising lines. Ironmaking & Steelmaking, vol. 31, No. 1, pp. 43--50 (2004)
13. Krimpenis, A., Benardos, P.G., Vosniakos, G.-C., Koukouvitaki, A.: Simulation-based selection of optimum pressure die-casting process parameters using neural nets and genetic algorithms. Int J Adv Manuf Technol, vol. 27, no. 5, pp. 509--517 (2006)
14. Cser, L., Gulyas, J., Szucs, L., Horvath, A., Arvai, L., Baross, B.: Different Kinds of Neural Networks in Control and Monitoring of Hot Rolling Mill. In: Monostori, L., Vancza, J., Ali, M. (eds.) Engineering of Intelligent Systems, pp. 791--796, Springer, Berlin, Heidelberg (2001)
15. Chang, D.S., Jiang, S.-T.: Assessing quality performance based on the on-line sensor measurements using neural networks. Computers & Industrial Engineering, vol. 42, No. 2, pp. 417--424 (2002)
16. Cherian, R.P., Midha, P.S., Pipe, A.G.: Modelling the relationship between process parameters and mechanical properties using Bayesian neural networks for powder metal parts. International Journal of Production Research, vol. 38, No. 10, pp. 2201--2214 (2000)
17. Hu, C.-H., Su, S.-F.: Hierarchical clustering methods for semiconductor manufacturing data. IEEE International Conference on Networking, Sensing and Control, vol. 2, pp. 1063--1068 (2004)
18. Chen, W.-C., Tseng, S.-S., Wang, C.-Y.: A novel manufacturing defect detection method using association rule mining techniques. Expert Systems with Applications, vol. 29, No. 4, pp. 807--815 (2005)
19. Koksal, G., Batmaz, I., Testik, M.C.: A review of data mining applications for quality improvement in manufacturing industry. Expert Systems with Applications, No. 38 (10), pp. 13448--13467 (2011)
20. Perzyk, M., Soroczynski, A.: Comparative Study of Decision Trees and Rough Sets Theory as Knowledge Extraction Tools for Design and Control of Industrial Processes. International Journal of Industrial and Manufacturing Engineering, no. 4 (1), pp. 234--239 (2010)