Decision making support for the development of new products based on Big Data technology

A I Shinkevich¹, I A Zaraychenko², R P Ivanova³ and E L Vodolazhskaya⁴

¹Department “Logistics and Management”, Kazan National Research Technological University, K.Marx Street, 68, Kazan 420015, Russia
²Department “Logistics and Management”, Kazan National Research Technological University, K.Marx Street, 68, Kazan 420015, Russia
³Department “Logistics and Management”, Kazan National Research Technological University, K.Marx Street, 68, Kazan 420015, Russia
⁴Department of business statistics and Economics, Kazan National Research Technological University, K.Marx Street, 68, Kazan 420015, Russia

E-mail: irina-zar@mail.ru

Abstract. The paper presents an analytical model for constructing a production system based on the application of the concept of big data. There are proposed a model for data collection and the architecture of a dynamic control system based on KPI modeling and construction of predictive models. A model for processing system data is developed, based on the use of a dynamic three-dimensional indexing data in the following areas: equipment, technology and schedule. Based on this approach, it is proposed to use the multiple linear regression model with fixing system constraints to solve the problem of optimizing production resources.

1. Introduction

In the context of building efficient production according to the concept of Industry 4.0, the important role of Big Data technology is determined by the fact that it represents a new computer paradigm that allows to collect, process and analyze of dynamic, diverse and large volumes of structured and unstructured data generated by the physical elements of the production system through Internet of Things (IoT) technologies or Cyber Physics Systems (CPS) [1]. Thus, the modern stage of digitalization of the surrounding world allows us to significantly develop the analytical results of processing various kinds of information to obtain a significant non-digital result, necessary for a deep understanding of the essence of the occurring phenomena, as well as their development trends.

Researches in the field of Big Data in production concentrate on technologies for analyzing significant data arrays with the aim of developing solutions that enhance the efficiency of the production process [2]. The work [3] presents the architecture of the Big Data application consisting of several layers: the system layer, the data collection layer, the processing layer, the modeling / statistics layer, the service layer (request / access), the visualization layer, and the control layer for a large data array. The author of [4] created a structure that includes algorithms for cleaning, compressing, grouping, and storing data from radio frequency beacons (RFID). The authors of the study [5] propose a big data analysis architecture in the context of product life cycle management. Researchers have focused on the application of big data in the manufacturing process and the life support processes of production. J.
Wang et al. [6] proposed a cloud architecture for building a Big Data system in maintenance, algorithms for real-time service and predictive data management by analyzing data collection methods. Gee and Wang [7] proposed a big data analytic approach to fault prediction for workshop planning. They improved the availability of processing resources by identifying potential malfunction / error patterns, including processing errors, processing defects, or problems with service conditions caused by processing conditions or an inappropriate schedule.

2. Materials and methods

Based on the analysis of existing models for embedding Big Data in the production process, a model of a data collection system is proposed to optimize the enterprise’s production resources.

Data on the physical condition of the equipment is recorded by PLC controllers, which include temperature sensors, volumetric weight sensors that record the volume and weight of processed raw materials, pressure sensors, etc. The data generated by the controllers on the equipment goes to the OPC platform (open platform communication), since the data are in different formats. In addition to the processed OPC data, the server client receives data from the MES and PLC databases as a source of historical data. Next, a complete array of data arrives at the processing and analysis server. The query results are transferred to APS, which generates a list of tasks for MES based on the generated model (figure 1).

![Figure 1. The data collection system.](image)

The structure of the processed data is presented in table 1.

The data in the MES database is stored in HBase, which is a NoSQL database. Equipment parameters received from local PLCs are larger and more heterogeneous, so the use of existing relational databases (RDBs) is limited. For this reason, NoSQL was used to collect this type of data, supporting data distribution and parallel processing.

All data is collected on a unit of equipment identifier (ID) to create forecasting models, based on a multiple regression model:

\[ E(x) = f(x_1, x_2, \ldots, x_n), \]  

(1)

E(x) – the efficiency of manufacturing product \(x_1, x_2, x_3, \ldots x_n\) – the parameters of the production system.

**Table 1.** Classification of data collected in the framework of the proposed model.

| Category      | Name                  | Content                                          |
|---------------|-----------------------|--------------------------------------------------|
| Equipment     | Physical performance  | Temperature conditions, working pressure, volume and weight of processed raw materials |
|               | Technical performance | Extreme operating conditions                     |
Based on the presented models, the architecture of the production system was developed, focused on solving the problems of energy and resource conservation in the framework of production processes. This system analyzes the data stream in real time, predicts the quality, performance and resource consumption of individual elements and processes and carries out a dynamic request for rescheduling taking into account deviations through integration with APS and MES systems. The proposed model is presented in figure 2.

**Figure 2.** The architecture of the system of dynamic management of resource-saving production based on Big Data technologies.

The structure of the elements of the presented model:

1. Big Data Warehouse – Data storage from control and monitoring systems (equipment sensors), as well as from the production management system (MES). This includes data such as the physical parameters of the process (temperature, pressure, vibration, metric production parameters, etc.). Process data from MES can include raw materials for each process, production status, work in progress and finished products, quality control results. Considering Since this data type is heterogeneous and also includes a significant array, it is advisable to use NoSQL databases to store events in real time, and to analyze and store historical data when enyat distributed file systems such as hadoop (HDFS);

2. Predictive Model Repository – Building a predictive model based on machine learning is iterative. The model repository allows you to store and search machine learning model;
3. Predictive model generator – Data research and data preprocessing are performed through parallel processing (two-phase analytics models and federated query). Data mining, model generation, and model validation are performed using a machine learning tool. Recommended big data processing platforms include Hadoop MapReduce and Spark. Mahout, Spark MLlib, R, and Python can be used as a machine learning tool;

4. KPI Simulator – Performs resource-intensive analysis using the model created in the previous step. The result of this module may include time-space production parameters that are adequate to current parameters and historical data;

5. Dynamic planning system (APS) – This module generates optimal production parameters, based on the set parameters of KPI resource saving, taking into account the current production situation. The result includes targeted resource costs, equipment performance, etc.;

6. Manufacturing execution system (MES) – It is an element of the system that carries out routine monitoring and control of the production process based on the parameters specified in the previous step.

Data on the physical parameters of production processes and their current state are transmitted via indicator indicators to the Big Data repositories, in addition, these repositories collect information on design and historical data on technical and technological parameters of production processes. Based on this data, through the implementation of various data processing models, adequate forecast data models are determined. In the absence of an adequate model in the repository, the process of generating a new model starts by embedding the current data in historical parameters. Then, the obtained / selected model is transferred to the KPI simulator, where, based on the set values of minimizing resource consumption, the predicted value is determined according to the set KPI values, if they are deviated, corrective measures are developed as part of the integration with the APS dynamic planning system. This system, based on the data on the actual operating parameters of the equipment received from the storage, forms the optimal corrective measures that are transmitted to the MES production system.

3. Results and discussion
In the analysis of the model, the task of determining the optimization criterion $E(x)$, which is the resulting variable of the production process, becomes relevant. The criteria for the efficiency of the production process include: productivity, duration of the production cycle, workload of equipment, etc. In this situation, an unambiguous criterion for the effectiveness of the organization of the production process is the criterion of total costs, and decision-taking in the field of production is reduced to building a decision tree taking into account the limitations of the data in table 1.

In this situation, the data set is a three-dimensional structure [8], oriented to the following directions of the parameters of the production process: equipment, technology, schedule (figure 3). Thus, the optimization of the processing of relevant information is carried out through the use of technology grip file, which allows to optimize the procedures for requesting and processing information about the production process.

![Figure 3. The multidimention data index.](image)
A grid file can be represented as a triple where is the set of dimensions, is the set of stripes, and is the set of chunks. Each stripe corresponds to exactly one dimension and crosses a non-empty subset of the set of chunks [9]. In the intersection cells of these measurements, the data of the parameters of the production system are stored in accordance with the values of various measurements. Accordingly, the optimization problem reduces to finding the extremum of the optimization function.

In addition, the analytical model should take into account restrictions in the indicated areas:

1) equipment: total production capacity:

\[ f(p_{e1}, p_{e2}, \ldots, p_{en}) \leq E, \]  \hspace{1cm} (2)

\( p_{e1}, p_{e2}, \ldots, p_{en} \) – equipment operation parameters;

2) technology: the need to comply with a given sequence of production operations, as well as their technological duration;

3) schedule:

– the need to comply with equipment operation requirements:

\[ \sum_{i=1}^{n} T_{ei} \leq T_{norm}, \]  \hspace{1cm} (3)

\( n \) – number of equipment group names;

\( T_{ei} \) – the duration of the equipment for \( i \)-th operations (taking into account technological breaks);

\( T_{norm} \) – ultimate duration of equipment.

– production plant operating time:

\[ \sum_{i=1}^{n} T_{ei} \leq T_{work}, \]  \hspace{1cm} (4)

\( T_{work} \) – production plant operating time.

Thus, the development of a new product using this toolkit will significantly reduce the costs associated with launching a new production, as well as predicting its effectiveness. In addition, this algorithm will improve the efficiency of existing processes, as well as provide dynamic flexible regulation of existing production systems by updating and information processing.

Acknowledgments

The research was carried out within the framework of the grant of the President of the Russian Federation for state support of leading scientific schools of the Russian Federation, project number NSh-2600.2020.6.

References

[1] Dyrdonova A N, Shinkevich A I, Galimulina F F, Malyshova T V, Zaraychenko I A, Petrov V I and Shinkevich M V 2018 Issues of industrial production environmental safety in modern economy Ekoloji 27(106) 193-201

[2] Chien C F, Chuang S C 2014 A Framework for Root Cause Detection of Sub-Batch Processing System for Semiconductor Manufacturing Big Data Analytics IEEE Transactions on Semiconductor Manufacturing 27 475–88

[3] Tekiner F, Keane J A 2013 Big Data Framework. IEEE International Conference on Systems 1494–9

[4] Zhong R Y, Huang G Q, Dai S L, Chen D X, Zhang T 2015 A big data approach for logistics trajectory discovery from RFID-enabled production data International Journal of Production Economics 165 260–72

[5] Zhang Y, Ren S, Liu Y, Si S 2017 A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products Journal of Cleaner Production 142 626–41

[6] Wan J, Tang S, Li D, Wang S, Liu C, Abbas H, Vasilakos A V 2017 A Manufacturing Big Data Solution for Active Preventive Maintenance. IEEE Transactions on Industrial Informatics.
2039–47

[7] Ji W, Wang L 2017 Big data analytics based fault prediction for shop floor scheduling Journal of Manufacturing Systems 43 187–94

[8] Nievergelt J, Hinterberger H, Sevcik K. 1984 The Grid File: An Adaptable, Symmetric, Multikey File Structure. ACM Transactions on Database Systems 9(1) 38–71

[9] Manukyan M G, Gevorgyan G R 2018 A dynamic indexing scheme for multidimensional data Modern information technology and IT-education 14(1) 111-25