Information Technologies of Processing Big Industrial Data and Decision-Making Methods

Andrey Kupin
Department of Computer Systems and Networks
Kryvyi Rih National University
11, Vitaliya Matusevycha str., Kryvyi Rih, Ukraine
kupin.andrew@gmail.com

Ivan Muzyka
Department of Computer Systems and Networks
Kryvyi Rih National University
11, Vitaliya Matusevycha str., Kryvyi Rih, Ukraine
musicvano@gmail.com

Ivchenko Rodion
Department of Automation Computer Science and Technologies
Kryvyi Rih National University
11, Vitaliya Matusevycha str., Kryvyi Rih, Ukraine
ivchenko.ra@gmail.com

***Abstract*—The scientific methods aimed at introducing of IT for processing large volumes of data with distributed infrastructure based on intelligent agents and parallel algorithms are considered. The emphasis is made on innovative methods based on intellectual agents and principles of Industry 4.0. The implementation and simulation of parallel algorithms for processing big data and decision-making trees are carried out.**

Keywords—Industry 4.0; Big Data; smart-agent; parallel algorithms.

I. Introduction

The issue of implementing intellectual management of repair and maintenance services at large industrial enterprises on the basis of new approaches within the modern concept of Industry 4.0 is considered [1]. The actuality of the problem and potential ways of its solution are presented in the previous papers of authors [2].

Modern trends in development of scientific and technical progress for the world industry are quite often described in such terms as “Smart Factory”, “Smart Manufacture”, “Intelligent Factory” and “Factory of the Future”. Now the development of these research areas is well enough formalized by the concept of the 4th Industrial Revolution (Industry 4.0) [3]. The implementation of the concept is related to the use of some key technological trends such as Big Data Processing, cyber-physical systems, autonomous robots with different intelligent sensors, simulators for 2D- and/or 3D-modelling, 3D-printers, Internet of Things, augmented reality etc. [4]. Thus, according to the estimations of leading world experts, these tendencies will determine the main vector of modern competitive industries [5].

II. Problem Statement

We reviewed the technological complex for monitoring, repair, calibration, search and replacement of unsuitable electronic equipment. Figure 1 shows the receipt and dispatch of equipment requiring maintenance and repair off/on the conveyor, used for transportation of the equipment between workshops of one company to another company.

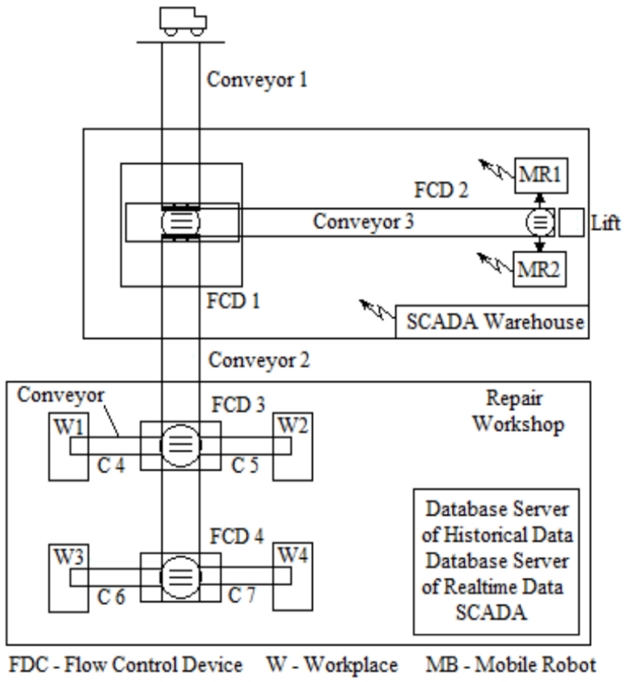

Fig. 1. The layout of the equipment and devices of the storage system of reception and repair shop [6]

The following tasks must be solved:

1) Parallel processing:

- Search for analogues of devices, parts.

978-1-5386-6611-1/18/$31.00 ©2018 IEEE

- Evaluation of the degree of wear on the basis of regression analysis and decision making on necessity for repair or replacement.
- Orders for logistics.

2) Development of monitoring and control algorithms, taking into account actual delays in hardware as well as delays caused by necessity for information processing [7-8].

- Data processing for certain signals:
- Reading of RFID tags – 2 ms.
- Reading of device parameters – 70-100 ms.
- Data transmission via Ethernet 100 Mbps – 0.5-1 μs.
- Reading from server – 2 μs.

3) Determination of the bandwidth of a peer-to-peer network considering a large number of types of devices and their accessories (over 16,000) and simultaneous processing tasks to determine the need for purchase a device or driver, software, delivery time, etc.

4) Logistics of delivery to the warehouse or workshop for calibration.

5) Data acquisition and filling in of the measurement database for apparatus diagnostics using regression analysis.

6) Implementation of the diagnostics without removing a device from the object.

Statement of the problem of the implementation of intelligent management using cyber-physical systems (smart-agents) may be applied to metallurgical enterprises for example, PJSC "ArcelorMittal Kryvyi Rih"). As a result, structural schemes have been presented and generic algorithms (Fig. 2-3) of implementation with similar approaches within the Industry 4.0 concept have been developed [5-6].

Practical realization of the specified tasks is possible only with the condition of effective parallelization of the basic processes.

Taking into account the significant calculation of the complexity of the problem, the implementation of these parallel algorithms and further computer simulation of such technology in the Big Data conditions will be discussed.

III. Methology of Parallelization

The proposed algorithm uses modern approaches for processing data streams using parallelization technology (Fig. 2) [6].

This algorithm provides with decomposition of the task for processing of parts. After the data flow is formed for parallel processing, the mutex mechanism is used to avoid conflicts between parallel processes, which try to access the shared data. After entering the critical section (function Mutex lock ()), the data are processed, and after the shutdown on the shared resource, a critical section is executed, causing mutex lock ().

After processing of parts, they are transmitted as a part (component agent). The information obtained through processing in PLC is transferred to the subsequent processing and storage in the database, which is constructed in accordance with ISA 95, ISA 88, ISO 22400 standards.

For managing tasks and predictions that are difficult to formalize, the effectiveness of system control has to be increased. This approach is based on the fact that for every value of a vector that describes the current situation (3) is a known value of a vector that describes the solution to be taken. Interpolation has to be used if it is impossible to describe all the situations.

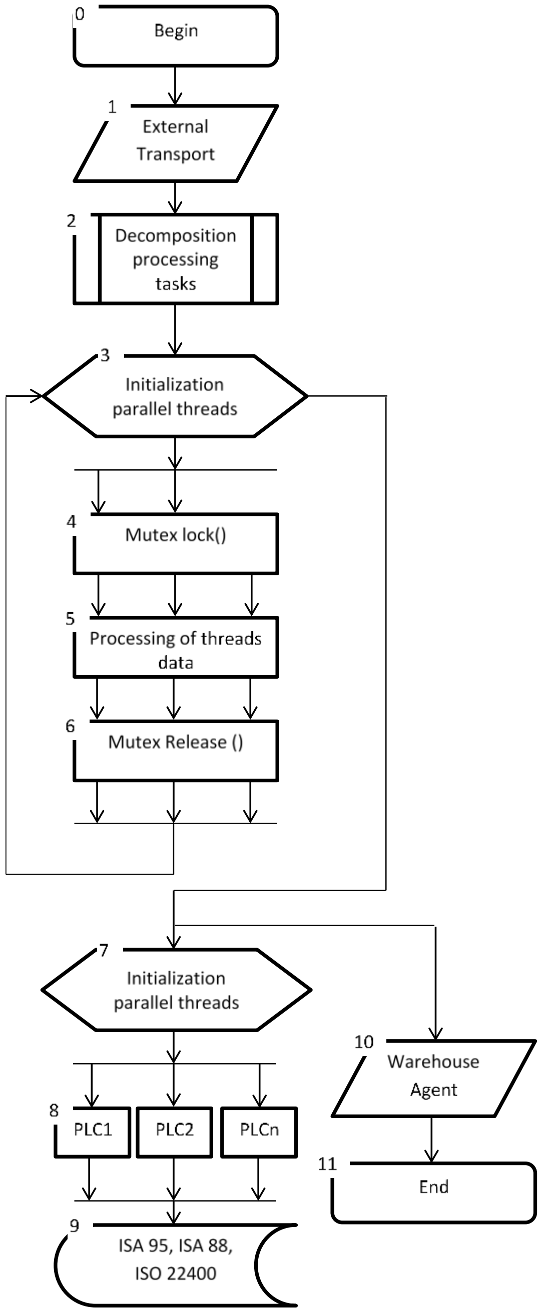

Fig. 2. Parallel algorithm of data processing [6]

The initial situation is characterized by the vector X = {x_1, x_2, ..., x_n}. Y solutions are taken as a component X vector. Y is also a vector Y = {y_1, y_2, ..., y_m}. Associative memory is used by this approach. The basic idea is based on opposition of each situation X with the decision of Y.

It is advisable to provide a system training with an exact model that uses the most accurate calculation of component solutions. This model for study is not only used for a specially developed system of step-by-step training, but also beyond the management of a real cycle with functional control.

The algorithm is shown in Figure 3.

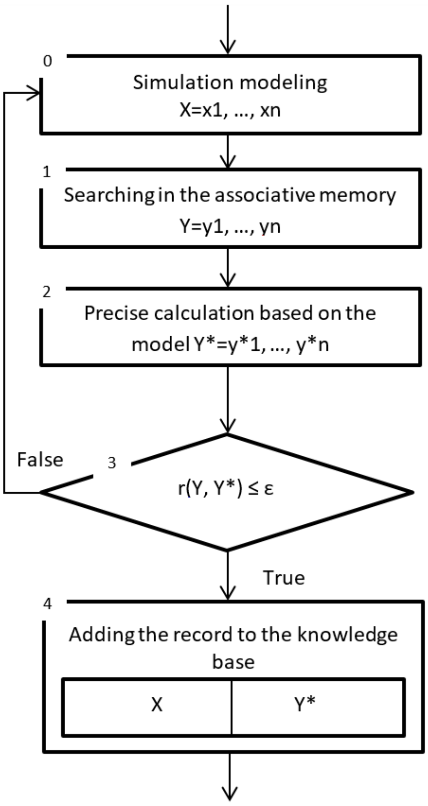

Fig. 3. Fragment of block diagram using for control and predictive tasks [6]

On this block diagram, Y* is vector management actions that are precisely designed for the model, ε – given accuracy, r(Y, Y*) – distance between the vectors Y and Y*.

IV. Methodology of Decision Making

The structure of the tree contains the following elements: “leafs” and “branches”. The attributes are written at the edges (“branches”) of the decision tree, on which the target function depends. The values of the target function are written in the “leaf”, and in other cells there are attributes that distinguish the cases. To classify a new case, we must go down the tree to the leaf and give out a corresponding value. Similar decision trees are widely used in intelligent data analysis. The goal is to create a model that predicts the value of the target variable based on multiple input variables. An example of one of these trees is shown in Figure 4.

Each letter represents the value of the target variable, variable on the way from the root to the leaf. Each internal cell corresponds to one of the input variables. A tree can also be “studied” by the division of output sets of variables into a subset, which are based on the testing of attribute values. This is a process that is repeated on each of the subsets received. Recursion ends when the subset in the cell has the same value of the target variable, thus, it does not add value to the predictions [12].

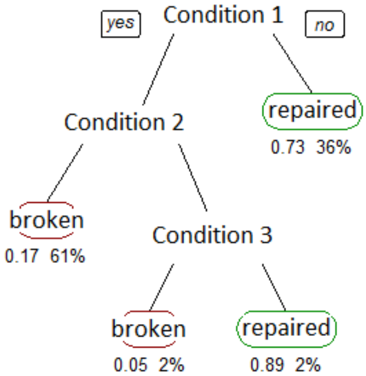

Fig. 4. An example of a decision-making tree

The process of going from top to bottom, the induction of decision trees (TDIDT), is an example of an absorbing “greedy” algorithm, and today it is the most widespread strategy of decision-making trees for data. But this is not the only possible strategy [12]. In the intelligent data analysis, decision trees can be used as mathematical and computational methods to help describe, classify and summarize a set of data that can be stated as follows:

$$(x, y) = \{x_1, x_2, x_3 \ldots, x_k, y\} \qquad (1)$$

The dependent variable Y is a target variable that needs to be analyzed, classified and generalized. Vector X consists of input variables x_1, x_2, x_3, which are used to perform this task.

In the decision analysis, the "decision tree" is used as a visual and analytical decision support tool that calculates the expected values (or expected profit) of competing alternatives.

The decision tree consists of three types of cells in figure 4:

- Decision cells – usually represented by squares.
- Probabilistic cells – represented as a circle.
- Closed cells – are presented in the form of a triangle.

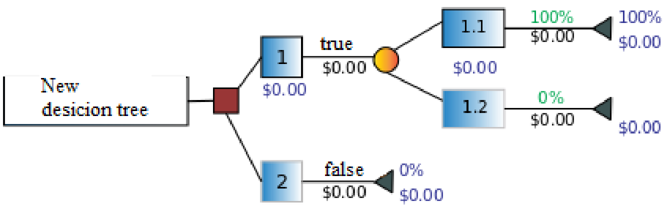

Fig. 5. Decision-making tree and types of cells

In the picture above, the decision-making tree should be read from left to right. The decision-making tree cannot contain cyclic elements, each new leaf can only be split, there are no converging paths. When constructing a tree manually, we may encounter a problem with its dimension, so we can usually obtain a decision-making tree by means of specialized software. Usually the decision-making tree is represented in the form of a symbolic scheme, which makes it easier to perceive and analyze [13].

We are interested in replacing spare parts on the equipment. We know that it depends on a number of parameters; it is hopeless to list all of them, but it is possible to choose the main ones (Table 1).

The methods of planning the existing maintenance and repair measures are classified as follows:

1) By event - for example, the elimination of equipment breakdown is used if the cost of repair is relatively low, and the lack of products resulting from equipment failure is low and will not affect the fulfillment of obligations to customers;

2) Regular maintenance - for equipment with provided modes and maintenance regulations, which initially presupposes the regular application of appropriate maintenance measures, this type of service gives the highest percentage of equipment readiness, but it is also the most expensive, as the actual state of the equipment may not require repair;

3) By state - expert way or with the help of the meters installed on the equipment, an assessment of the state of the equipment is carried out, and on the basis of this assessment, a forecast is made when this equipment is to be repaired.

The state of the method has the highest priority in use. The benefits of this type of service - its cost is less, and the readiness of equipment to run production programs is high enough. This method of "tree adoption" solutions and will be a priority in using it in the repair of equipment [10].

TABLE I. EXAMPLE OF STATISTICS

Run	Type	Manufacturer	Criticality for production	Replacement
R1	T1	Country1	Yes	Yes
R1	T2	Country2	No	Yes
R2	T3	Country1	Yes	No
R2	T4	Country2	No	Yes

V. MANUFACTURING EXECUTION SYSTEM ARCHITECTURE

A manufacturing execution system MES is an information system that connects, monitors and controls complex manufacturing systems and data flows on the factory floor. The main goal of an MES is to ensure effective execution of the manufacturing operations and improve production output.

An MES helps achieve that goal by tracking and gathering accurate, real-time data about the complete production lifecycle, beginning with order release until the product delivery stage for finished goods.

The MES collects data about product genealogy, performance, traceability, material management and work in progress WIP and other plant activities as they occur. This data, in turn, allows decision-makers to understand the current settings of the factory floor and better optimize the production process [14].

The ANSI/ISA-95 standard merged the MESA-11 model with the Purdue Reference Model, creating a functional hierarchy. In this model, MES was established at the intermediate level three, between enterprise resource planning (ERP) at level four and process control at levels zero, one and two [16].

What makes the MES system important is that it serves as a functional layer between the ERP and the process control systems on the factory floor, giving manufacturers real-time workflow visibility, flexibility and insight into how best to improve enterprise-wide manufacturing operations.

MES maintains device history record DHR for each product unit and batch by collecting data, processes and outcomes of the manufacturing process for compliance.

An MES is often integrated with Enterprise Resource Planning ERP, supply chain management, product lifecycle management and other key IT systems. ERP is the integrated management of core business processes, often in real-time and mediated by software and technology. ERP is usually referred to as a category of business-management software — typically a suite of integrated applications—that an organization can use to collect, store, manage, and interpret data from these many business activities.

ERP provides an integrated and continuously updated view of core business processes using common databases maintained by a database management system. ERP systems track business resources—cash, raw materials, production capacity—and the status of business commitments: orders, purchase orders, and payroll. The applications that make up the system share data across various departments (manufacturing, purchasing, sales, accounting, etc.) that provide the data. ERP facilitates information flow between all business functions and manages connections to outside stakeholders.

A key benefit of using an MES includes [15]:

- Increased customer satisfaction.
- Improved regulatory compliance.
- Better agility and time to market.
- Improved supply chain visibility.
- Reduced manufacturing cycle time.
- Elimination of paperwork and manual data-entry processes.
- Reduced order lead time.
- Lower labor costs.
- Reduced WIP inventory.
- Increased machine utilization.

VI. CONCLUSION

The most important results of the project from the point of view of science is to improve the methodology of distributed cyber-physical intelligent management system to solve the problems of optimization of technical maintenance and repair of industrial equipment, in accordance with the concept Industry 4.0.

Preliminary calculations show that if the research and created intellectual technology is implemented by the relevant services at an enterprise, then the following improvements will be made:

- Reduction of downtime to zero due to reduced time, duration of repairs and maintenance.
- Reduction of production costs.
- Optimization of equipment kits at the plant, reducing its diversity and minimizing the required number of parts and devices.
- Reduction of expenses for purchase of spare parts due to more efficient use of functioning parts of the decommissioned equipment.
- Improvement of quality of repair and maintenance;
- Increase of safety, functionality and comfort of personnel operations.
- Creation of new working profiles [9].

Given the novelty and urgency of problems not only in Ukraine, but also in the world, it is considered appropriate to use such solutions at domestic enterprises [10-11].

References

[1] A. Kupin, I. Muzyka, M. Romanov, S. Ruban, O. Savitskyi, and V. Kharlamenko, "Intelligent management of repairs industrial facilities with geographically distributed infrastructure on the basis of CPS",. *Cyber-Physical System: Achievements and Challenges, II scientific seminar,* Lviv Polytechnic National University, pp. 23-29, 2016. (in Ukrainian).

[2] A. Kupin, I. Muzyka, "Analysis of potential opportunities and rationale of cyber-physical systems for mining and metallurgical complex", *Computer science, information technology, automation*, vol. 4, pp. 25-29, 2016.

[3] A. Kupin, "Neural identification of techno-logical process of iron ore beneficiation", *Proceedings of 4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems Technology and Applications* (IDAACS'2007), Dortmund, Germany, pp. 225–227, 2007.

[4] A. Kupin, Y. Kumchenko, "Usage of training methods to parameterization of multilayer neural computing structures for technological processes", *Radio electronic and computer systems*, vol. 5, pp. 100-104, 2014.

[5] M. Hermann, T. Pentek, B. Otto, "Design Principles for Industrie 4.0 Scenarios: A Literature Review", Technische Universitat Dortmund, 2015.

[6] B. Poddubny, A. Kupin, I. Muzyka, O. Savytskyi, S. Ruban, and V. Kharlamenko, "Information Technology for Processing of Industrial Big Data with Distributed Infrastructure on the Basis of Smart Agents and Parallel Algorithms", *International Journal of Engineering and Technical Research (IJETR)*, vol. 7, issue 9, pp. 25-29, 2017.

[7] E. A. Lee and S. A. Seshia, Introduction to Embedded Systems — A Cyber-Physical Systems Approach, http://LeeSeshia.org, 2011.

[8] Mario Hermann, Tobias Pentek, Boris Otto. Design Principles for Industrie 4.0 Scenarios: A Literature Review. Technische Universitat Dortmund. 2015.

[9] Klaus Schwab. The Fourth Industrial Revolution: what it means, how to respond. *World Economic Forum*. 2016.

[10] The official site of PJSC "ArcelorMittal Kryvyi Rih" [Electronic resource] – Available: http://ukraine.arcelormittal.com

[11] The official site of the international program HORIZON 2020 [Electronic resource] – Available: http://ec.europa.eu/programmes/horizon2020/en

[12] Quinlan, J. R. Induction of Decision Trees. Machine Learning 1: 81-106, Kluwer Academic Publishers, 1986.

[13] Breiman, Leo; Friedman, J. H., Olshen, R. A., & Stone, C. J. Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software., 1984.

[14] McClellan, Michael. Applying Manufacturing Execution Systems. Boca Raton, Fl: St. Lucie/APICS., 1997.

[15] Meyer, Heiko; Fuchs, Franz; Thiel, Klaus. Manufacturing Execution Systems: Optimal Design, Planning, and Deployment. New York: McGraw Hill., 2009.

[16] Johann Eder, Schahram Dustdar Business Process Management Workshops., 2006, p. 239