

Peter Bubenik - Filip Horak - Viktor Hancinsky *

ACQUIRING KNOWLEDGE NEEDED FOR PULL PRODUCTION SYSTEM DESIGN THROUGH DATA MINING METHODS

Article deals with defining relationships for setting the optimal number of kanban cards in individual circuits of pull production systems, in order to minimize work in progress, while maximizing the number of completed orders in the observed time interval. To achieve this objective, data mining methods were used.

Keywords: Data mining, production planning, knowledge acquisition, genetic algorithm.

1. Introduction

In a pull production system, processes are based on customer demand. The main difference between classic push production systems and pull production systems is that the first one schedules work releases based on demand, and the latter one authorizes work releases based on system status. Among other advantages, pull production systems can provide low unit cost, good customer service, high external quality and flexibility. Unfortunately, pull systems cannot be applied to all business types, because they require initial conditions such as stable demand, low number of product types, etc.

However, by integrating pull systems in some of your production processes, you will be able to reduce your lead times, and perhaps associated costs. One of the most significant parameters of pull production system is a number of circulating kanban cards. Central topic of this article is a new approach to finding out an optimal setting of this crucial parameter by using methods and techniques such as data mining, simulation and genetic algorithms.

2. Pull Production System

Kanban became an effective tool in support of running a production system as a whole, and it proved to be an excellent way for promoting improvement. One of the main benefits of kanban is to establish an upper limit to the work in progress inventory, avoiding overloading of the manufacturing system.

Kanban cards are a key component of kanban and signal the need to move materials within a manufacturing or production facility or move materials from an outside supplier in to the production facility. The kanban card is, in effect, a message that signals that there is a depletion of product, parts, or inventory that, when received, the kanban will trigger the replenishment of that product, part, or inventory. Consumption therefore drives demand for more production, and demand for more products is signalled by the kanban card. Kanban cards therefore help create a demand-driven system [1].

3. Data mining

The manual extraction of patterns from data has occurred for centuries. Early methods of identifying patterns in data include Bayes' theorem (1700s) and regression analysis (1800s). The proliferation, ubiquity and increasing power of computer technology has dramatically increased data collection, storage, and manipulation ability. As data sets have grown in size and complexity, direct "hands-on" data analysis has increasingly been augmented with indirect, automated data processing, aided by other discoveries in computer science, such as neural networks, cluster analysis, genetic algorithms (1950s), decision trees (1960s), and support vector machines (1990s). Data mining is the process of applying these methods with the intention of uncovering hidden patterns in large data sets. It bridges the gap from applied statistics and artificial intelligence (which usually provide the mathematical background) to database management

* Peter Bubenik, Filip Horak, Viktor Hancinsky

Department of Industrial Engineering, Faculty of Mechanical Engineering University of Zilina, Slovakia
E-mail: peter.bubenik@fstroj.uniza.sk

by exploiting the way data is stored and indexed in databases to execute the actual learning and discovery algorithms more efficiently, allowing such methods to be applied to ever larger data sets. Knowledge exists in all business functions, including purchasing, marketing, design, production, maintenance and distribution, but knowledge can be notoriously difficult to identify, capture and manage [2]. Simply stated, data mining refers to extracting or “mining” knowledge from large amounts of data [3]. Due to large amounts of data generated and collected during manufacturing execution, manufacturing is a promising area of application for data mining to extract knowledge for optimization purposes [4].

4. Simulation in terms of data mining

Simulations lend meaning to data and can be updated and adapted as further data come in [5]. It often happens that provided historical data are not sufficient to derive relevant knowledge. This occurs for a number of reasons, the most common ones are:

- The unpredictability of external factors affecting the operation of the production system.
- Inadequate recording of the data needed for analysis.
- Incorrect data caused by human factor.
- Incompatible corresponding resolution of analyzed data.
- Lack of data due to newly implemented system.

In such cases, it is possible to use simulation, in which the listed historical data are used, if possible, in the following tasks:

- Elimination of attributes that do not affect the target parameter.
- Setting the probability distribution of the process attributes.
- Simulation model validation.

If the historical data are not available at all, as often happens in case of designing new production systems, simulation is the only way to obtain data necessary to derive applicable knowledge. The heart of data mining is knowledge discovery, as it enables to discover relevant objects and the relationships that exist between

these objects, while simulation provides a vehicle to represent those objects and their relationships [6].

5. Problem formulation

Figure 1 shows the pull production system, whose parameters are needed to be optimized. Simulation model was created in Simulink environment with use of SimEvents Toolbox [7]. In considered system, there are three circuits, in which kanban cards in the number of kb_1 , kb_2 and kb_3 circulate. The first circuit is shown first from the left in Fig. 1, to which belongs number of cards kb_1 , the next one, transport circuit with block *Trans_kanban* contains kb_2 kanban cards and for following block, kb_3 cards are allocated. The very last block on the right is used to generate production orders based on normal distribution. Also cumulative operations for defining the amount of work in progress run in this block.

The aim of the solution was to formalize relations that describe the impact of the selected combination of numbers of kanban cards on the amount of work in progress and the numbers of finished orders, as these factors are crucial in terms of the efficiency of the proposed system.

Since the chosen maximum number of kanban cards was $max[kb_i]=30$, in the case of simulating every possible case it would leave $n^k = 30^3 = 27000$ simulation runs. To reduce the number of simulation runs, optimization through genetic algorithms in Matlab environment was used in this case.

6. Acquiring data using genetic algorithms

If we label $f_{sim}(kb_1, kb_2, kb_3)$ as a function that describes the behavior of the simulation model, the optimization problem can be written as:

$$WIP_{opt} = \min_{kb_i \in Z^+} \{f_{sim}(kb_1, kb_2, kb_3)\} \quad (1)$$

where WIP_{opt} is the minimal amount of work in progress achieved with properly chosen combination of numbers of kanban cards in each circuit. Utilization of genetic algorithms in this case

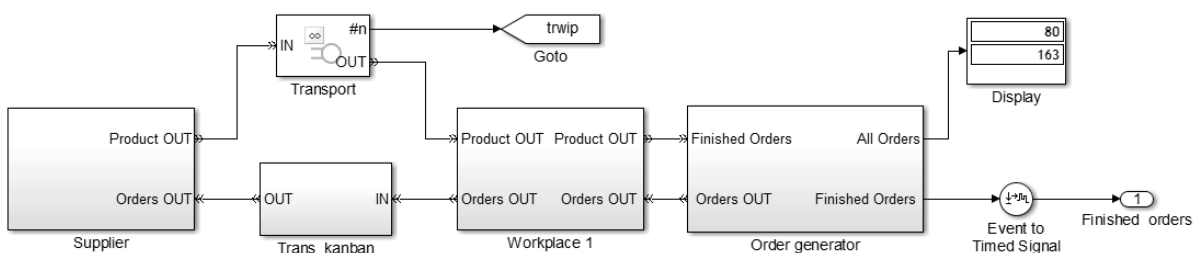


Fig. 1 Analyzed production system created in Simulink environment

reduced the number of simulation runs from 27000 to 1040 for optimization in terms of the amount of work in progress. For optimization in terms of the number of finished orders, a similar formula was used, with the difference that in this case, the aim was to maximize the number of finished orders.

$$FO_{opt} = \max_{kb_1 \in Z^+} \{f_{sim}(kb_1, kb_2, kb_3)\} \quad (2)$$

For optimization in terms of the number of finished orders, 880 simulation runs were executed. Values calculated in order to minimize WIP and maximize FO together make the dataset, which should also contain those combinations of numbers of kanban cards in particular circuits, which can be determined from which knowledge applicable to the production system can be determined. Figure 2 shows the resulting dataset of performed simulation run, where the color attribute describes the amount of work in progress.

From Fig. 2 it is clear that the amount of work in progress is increasing with the increasing number of kanban cards in each circuit. Figure 3 shows the same set of combinations of different

amounts of kanban cards, but in this case the number of finished orders is the evaluation criterion.

Search and formalization of knowledge through data mining algorithms

Following data mining algorithms were used to seek knowledge in the dataset:

- Neural networks.
- Random forests.
- Linear regression.

The solution was carried out in KNIME [8] environment and for testing the models in terms of suitability of use, equation (3) was used to calculate the mean square error.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (3)$$

where \hat{Y} is prediction vector and Y is the vector of actual values obtained from the simulation model. To use a neural network predictor, it was necessary to temporarily convert the data into $\langle 0,1 \rangle$ interval, using combinations of *Normalizer/Denormalizer* nodes. In this case, feedforward neural networks algorithm RProp

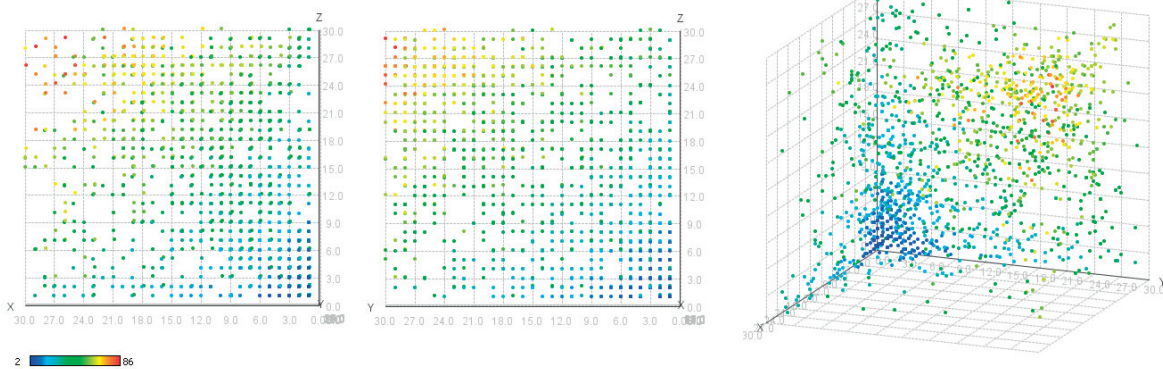


Fig. 2 Effect of combinations of numbers of kanban cards in particular circuits on the amount of work in progress ($x=kb_1, y=kb_2, z=kb_3$)

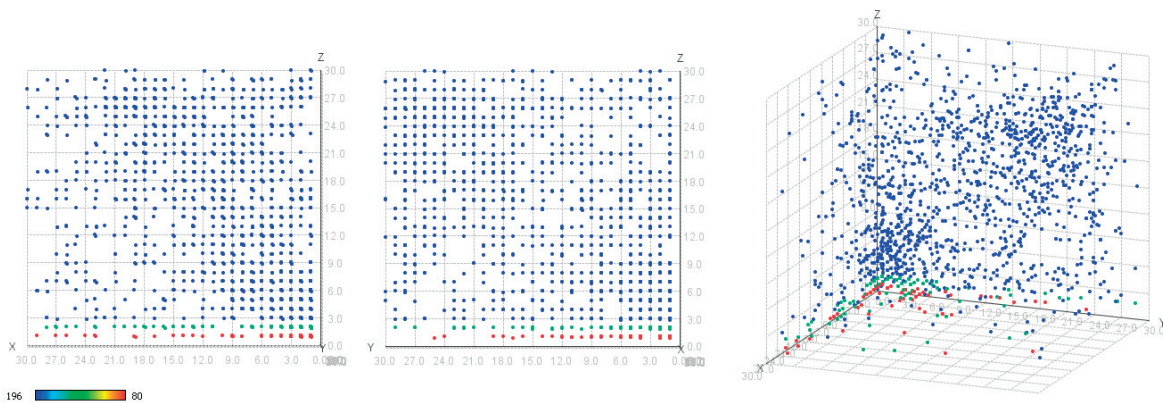


Fig. 3 Effect of combinations of numbers of kanban cards in particular on the number of finished orders ($x=kb_1, y=kb_2, z=kb_3$)

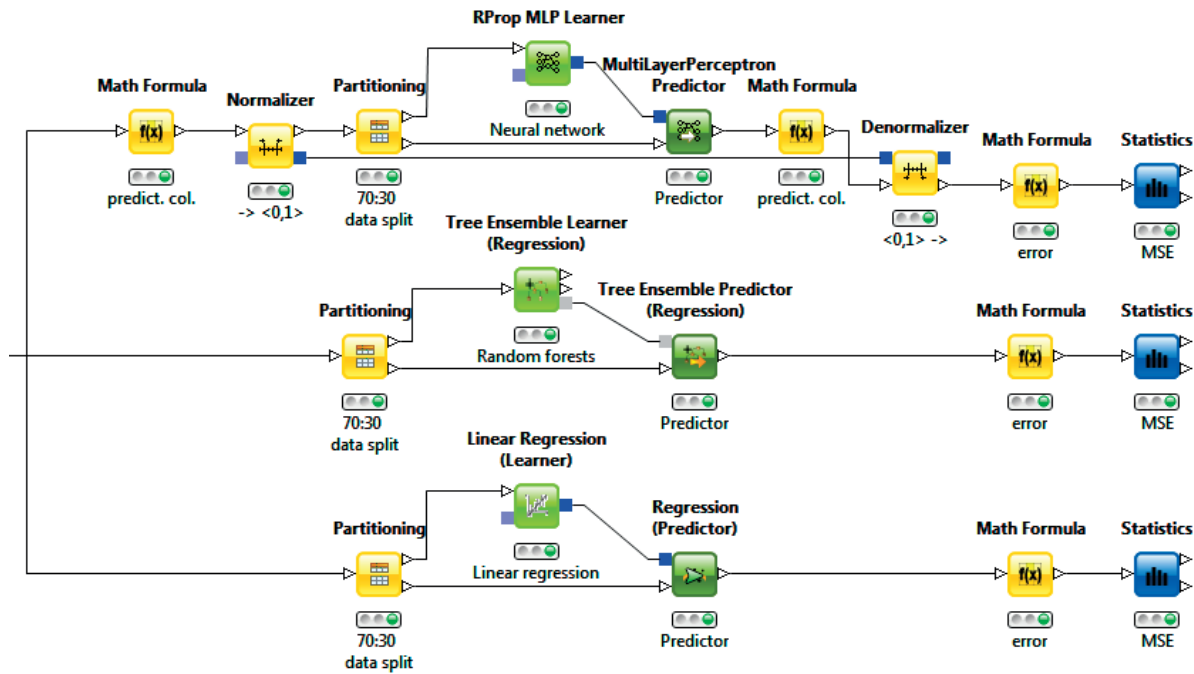


Fig. 4 Data processing stream and knowledge acquisition

MLP was used with the number of 100 iterations. As with other used algorithms, random selection was used to select training set of 30% size of the total analyzed data, based on which the mean square error was calculated according to equation (3) within model testing. Random forests algorithm was implemented by creating 50 random regression decision trees. It was not necessary to normalize the data. Further configuration of the algorithm for linear model prediction based on linear regression method was not necessary. Data processing stream and knowledge acquisition in KNIME environment can be seen in Fig. 4.

In this case, the aim was to seek knowledge about relation of combination of numbers of kanban cards kb_i in individual circuits i to amount of work in progress WIP . In the case of gaining knowledge about target parameter FO - number of finished orders, process topology remained unchanged (as can be seen in Fig. 4), only input dataset and its predicted columns were changed.

After executing transformation and mining operations, the results showed that for defining relations between number of kanban cards and amount of work in progress, the linear regression generated model performs the best. For determining the effect of kanban cards on number of finished orders, the best performing algorithm is random forests. In these cases, the mean square error was the lowest, as is shown in Table 1.

Calculated mean square error values for individual models Table 1

Predicted attribute	MSE		
	Rprop MLP	Random forests	Linear regression
WIP	6.467	37.42	0.106
FO	0.0587	0.011	1121.579

Tested models with the lowest error can be considered established and formalized knowledge. For the amount of work in progress, the model has the form:

$$WIP = 1.005 \cdot kb_1 + 0.9985 \cdot kb_2 + 1.0274 \cdot kb_3 \quad (4)$$

For the number of finished orders, the random forests algorithm derived model:

$$FO = \begin{cases} 80: kb3 < 1.5 \\ 154: kb3 \leq 2.5 \\ 196: kb3 > 2.5 \end{cases} \quad (5)$$

7. Conclusion

Acquired knowledge (4), (5) can be written in PMML format and added to the knowledge base, which is together with inference mechanism and user interface capable to form functioning knowledge system. Therefore, when designing pull production system it is necessary in terms of capacity to consider this knowledge, if we want to achieve the optimal amount of work

in progress together with the highest possible number of finished orders. In the case of appropriate choice of combination of kanban cards, the cost perspective would be also important, but if we consider that in the case of pull systems the highest costs come from delays, which results in a smaller number of completed

orders in due time, the most feasible combination of numbers of kanban cards should be [1 1 3]. Specific solution for the current situation, however, should be offered by the knowledge system, which would retrieve listed findings from knowledge base and optimize the resulting combination also in the terms of cost.

References

- [1] MICIETA, B., GREGOR, M., QUIRENC, P., BOTKA, M.: *KANBAN: Slovak Centre for Productivity (in Slovak)*, Zilina, 2001, ISBN 8096832425
- [2] HARDING, J. et al.: Data Mining in Manufacturing: A Review, *J. of Manufacturing Science and Engineering*, 2006, vol. 128, p. 969-976, ISSN: 1087-1357
- [3] HAN, J., KAMBER, M.: *Data Mining: Concepts and Techniques*, 2nd ed., edited by J. Gray, San Francisco: Morgan Kaufmann Publishers, 2006, p. 5, ISBN: 1-55860-901-6
- [4] GROGER, CH., NIEDERMANN, F., MITSCHANG, B.: *Data Mining-driven Manufacturing Process Optimization*. Proc. of the World Congress on Engineering, 2012, London. ISBN: 978-988-19252-2-0
- [5] LYTTON, W., STEWART, M.: Data Mining Through Simulation. *Neuroinformatics*. ISBN 978-1-58829-720-4, 2007, vol. 401, pp. 155-166
- [6] ALNOUKARI, M.: *An Integrated Data Mining and Simulation Solution*, 2010, p. 929, Available online: <http://www.academia.edu/190456/An_Integrated_Data_Mining_and_Simulation_Solution>
- [7] *SimEvents*, 2012. Available online: <<https://www.mathworks.com/products/datasheets/pdf/simevents.pdf>>
- [8] *KNIME Quickstart Guide*. 2011 Available online: < http://tech.knime.org/files/KNIME_quickstart.pdf> .