SciVerse ScienceDirect
Procedia CIRP 9 (2013) 44 – 48
www.elsevier.com/locate/procedia

2nd CIRP Global Web Conference

Improving data integrity in production control

Guenther Schuh[a], Till Potente[a], Christina Thomas[a], Felix Brambring[a]*

[a]Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University,
Steinbachstraße 19, 52074 Aachen, Germany

* Corresponding author. Tel.: +49-241-80-28241; fax: +49-241-80-22293.*E-mail address*: F.Brambring@wzl.rwth-aachen.de.

Abstract

Studies show that companies which focus on high adherence to promised delivery dates as their main logistic goal, regularly outperform their competitors. Only with a highly accurate production planning and control (PPC) companies can accomplish this goal. However, usually there is a gap between the planned forecast of the Advanced Planning and Scheduling System and the actual output of the production system. One of the reasons for this discrepancy is inconsistent data which is collected on the shop floor and builds the foundation for the planning process. In this paper, a methodology is presented to assure higher integrity in production control data.

Selection and peer-review under responsibility of International Scientific Committee of the 2nd CIRP Global Web Conference in the person of the Conference Chair Dr. Sotiris Makris

Keywords: Production Planning and control; Data Integrity; Data Inconsistency; Data Mining; Association Rule Induction

1. Introduction and state of the art

Driven by an increasing market dynamic, manufacturing companies are facing the challenge of increased customers' demands regarding highly individualized products at low costs and short delivery times [1]. A recent study by the Laboratory for Machine Tools and Production Engineering (WZL) shows that 67% of manufacturing companies claim the adherence to promised delivery dates to be their main logistic target [2]. For these companies, being able to perform an accurate production planning and control (PPC) is a core element in meeting this challenge [3].

Especially in job-shop manufacturing, production planning and control is a complex task characterized by a high variation of possible production sequences and last-minute changes by the customer concerning the delivery date [4]. Therefore manufacturing companies rely heavily on the utilization of specialized IT-systems (e.g. Advanced Planning and Scheduling Systems, APS-Systems) to support PPC-processes [5]. However, the sole application of such an APS-System does not automatically lead to a higher adherence to promised delivery dates [6]. On the contrary, deficits in the reliability of the forecast of planned production orders are commonly found in job-shop manufacturing. Therefore, increasing the achieved adherence to promised delivery dates is a question of minimizing the gap between the original plan and the actual shop-floor activities [7]. For this end, it must be ensured that the gathered data are accurately representing current state of all production orders on the shop-floor. Companies use Production Data Acquisition Systems (PDA-Systems) to gather data directly from the shop-floor. These data can be uploaded automatically by the deployed machines if they are linked to the company's IT-infrastructure. However, to this day most manufacturing companies still use a number of machines which do not offer this functionality, so that machine operators need to manually register finished process steps via terminals exclusively provided for this purpose. Obviously, this method is more prone to errors since human actions are involved. Where information about the same process is gathered via different sources (e.g. automatically and manually), inconsistencies can occur which also question the reliability of the data.

Selection and peer-review under responsibility of International Scientific Committee of the 2nd CIRP Global Web Conference in the person of the Conference Chair Dr. Sotiris Makris
doi:10.1016/j.procir.2013.06.166

For all these various reasons, APS-Systems are forced to work with inaccurate data while simultaneously ignoring the fact that the planning builds on faulty assumptions. As an example, in Figure 1 the distribution of planned and actual throughput times of individual process steps in a company with job-shop manufacturing is shown. Only roughly 2% of these process steps adhered to the planned schedule while for 50% of all process steps the reported throughput time deviated more than 20% from the initially planned throughput time. This evidence illustrates the high potential of improving planning in production.

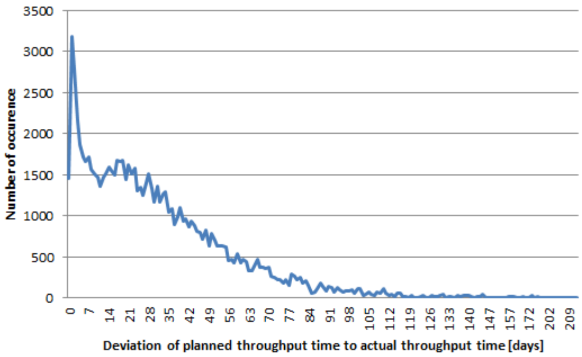

Fig. 1. Distribution of deviation between planned and actual throughput time in days

Substantial research has been carried out concerning the utilization of data mining methods in production-related tasks. Harding et. al. did an extensive review of data mining in manufacturing: They found extensive literature discussing data mining applications for managing production processes, operations, maintenance and product quality improvement while manufacturing planning and shop-floor control have been less considered [8]. Chen used association rule induction in order to compose manageable subsystems in cellular manufacturing systems [9]. Kwak and Yih developed a production control approach based on long- and short-term information which is mined from historical production data [10]. So far, researchers have not explicitly addressed the capability of data mining when it comes to improving data integrity in a production control context.

This paper presents an overview of the common data inconsistencies in production control. A data mining methodology is proposed which helps to overcome the information gap between the planned production activities and the actual shop-floor situation.

2. Data inconsistencies in production control

In principle, the operational data gathered by PDA-Systems contains information about produced quantities of intermediate- and end-products and the corresponding machines or work stations that were used to produce these quantities with specific set-up and processing times [11]. For a manufacturing company, this data easily exceeds hundreds of unique data sets created every work day.

As a representative of the class of German mid-sized manufacturing companies, the data of a manufacturer of clutches and brakes serves as a proper example. This company applies job-shop manufacturing due to a high variety of customized products. The available data set contains information about roughly 16.000 production orders processed over the course of one and a half years. The gathered production data includes planned and actual set-up and processing times of each process step as well as the respective work station. Additionally, the planned work station for the following process step and date-stamps of planned and actual start- and end-date of each process step are reported.

The given data set contains a variety of data inconsistencies, shown in Table 1 with their respective rate of occurrence. Over 10% of the total data are inconsistent or missing which justifies accelerated efforts concerning data integrity.

Table 1. Examples of data inconsistencies with occurrence rate

Data Inconsistency	Occurrence rate
Mismatch between planned work stations in consecutive process steps	6.1%
No feedback about work station	3.2%
Previous process step reported as 'finished' after following process step was already reported as 'started'	2.1%
Process steps with missing start- or end-date	1.4%
Following process step reported as 'finished' earlier than previous process step	1%

In 6.1% of roughly 86.000 process steps reported in the data set, there was a mismatch between the planned work station for the following process step and the actually reported work station. This equals roughly 5000 incidents in which information about the deployed work station is ambiguous leading to faulty assumptions concerning the material flow and utilization of capacities. Without correct information about the whereabouts of production orders and the actual utilization of available capacities, future planning cannot be accurate. Additionally, in further 3.2% of recorded

process steps no work station was reported at all, which further exacerbates the described problems.

In 2.1% of the recorded data sets, a process step was reported as ‘finished’ after its following process step was already reported as ‘started’. Since handling of a production order on one work station can only begin after processing was completed on the previous work station, the data is obviously inconsistent. The following process step even being reported as ‘finished’ earlier than the previous process step, was observed in 1% of all data sets. In 1.4% of all cases, the start- or end-date of the respective process step was missing completely. All these inconsistencies and errors in data, prevent the calculation of actual throughput times and therefore the adherence to delivery dates cannot be estimated correctly. It can be suspected that these problems would be even more apparent with the recording of time-stamps for the respective events instead of only date-stamps as in this example.

These common data inconsistencies prevent APS-Systems from accurately planning the completion dates of production orders and therefore constitute a important reason for low adherence to promised delivery dates. Hence, for an APS-System to be an useful tool, data inconsistencies need to be eliminated as far as possible.

3. Assuring data integrity in production control

To ensure data integrity in production control, data inconsistencies as described in the previous chapter, need to be cleared efficiently and without introducing new bias into the data. This chapter presents a combination of association rule induction and logical reasoning to increase data integrity in production control.

The previously described flaws in production-related feedback data can be clustered in two separate classes. The first class is comprised of the first two entries mentioned in Table 1 and describes problems related to the utilized work stations. The provided information concerning the work station for the respective process step is either inconsistent or missing. The second class is comprised of the last three entries in Table 1 and therefore deals with inconsistent or missing data concerning start- and end-dates of the respective process steps. The proposed method has two steps: First, the data inconsistencies are cleared via an association rule induction. With these results, the missing start- and end-dates in the data can be estimated in a second step. This paper focuses on finding an assumption for the missing work stations.

Association rule induction is an extensively used tool in market research, especially in the so-called market basket analysis. Within this domain, it is used to find regularities in the shopping habits of supermarket an online-shop customers. Its main idea is to conclude from one set of products which other products the customer will also purchase with a high probability [12]. While in market basket analysis, this knowledge is used to advertise these products together or place them next to each other in the store, in production control we propose this method can be used to make assumptions about the utilized work station and therefore increase data integrity.

The greatest impediment of using association rule induction is the high number of theoretically imaginable association rules [12]. The previously mentioned data set from a German mid-sized company contains 167 unique work stations, which theoretically allows to formulate trillions of possible association rules and will grow exponentially with each added work station. Understandably, testing each of these rules individually would be very time consuming. However, a variety of algorithms exist that can be used to limit the number of possible association rules. To make sure that the best-fitting rules are chosen, two indicators are usually calculated: the *support* and the *confidence* of an association rule [12]. Transferred to our production control application, the *support* measures the percentage of production orders in which the specific rule is correct and the *confidence* stands for the conditional probability of a sequence, given an antecedent work station sequence.

One of the best-performing algorithms for association rule induction is the so-called apriori algorithm [13] [14]. The apriori algorithm works in two basic steps [14]: first, all work station sequences with a pre-defined minimal support have to be determined which ensures that the sequences occur at least in a fixed percentage of all production orders. We call this subset the *frequent sequences*. Afterwards, association rules can be generated by testing sequences from this subset for their *confidence*.

Compared to the market basket example, the analysis is even more complex in production control since the sequence of products in the shopping cart is usually not interpreted. However, this circumstance can be used as an advantage: in order to double-check the formulated rules, the respective production technologies of the work stations in sequence can be examined, e.g. fine finish grinding is expected to follow rougher machining steps like milling or turning and not vice versa.

The proposed modification of the apriori algorithm has been applied on the previously described data set of the clutches and brakes manufacturer. A production order with a missing work station in the feedback data was chosen at random. The predecessors of the unknown work station where an oven and a following manual control of the material. In order to get relevant results, the minimal support had to be set to 0.3% which is considerably lower than in a typical market basket analysis. This effect is caused by the exponentially higher number of possible sequences since the chronological order of the work stations are of paramount importance in production control. Setting the minimal support to 0.3% implies that every association rule which we wanted to consider had to be true for at least this percentage of production orders. The generated subset of frequent sequences had a cardinality of 173. One of the applicable sequences in the subset of all frequent sequences had a support of 0.34%. These rule suggested the missing work station to be a manual bundling and counting process step which represents a reasonable activity given the antecedent process steps. This rules confidence was calculated with 27.44%, meaning that this rule is correct in over a quarter of all cases in which it is applicable. Following this reasoning, this manual bundling and counting process step would be the best guess in order to fill in the blanks in the feedback data. Applying this algorithm to the whole data set will eliminate all missing work stations.

The second step of the proposed method is to eliminate the missing start- and end-dates of the affected process steps. With all work stations already known through application of the modified apriori algorithm, the spadework has been done. By switching to a resource-centered view of production circumstances, it is possible to estimate the likely time of arrival of any production order in the queue in front of any work station by taking the average waiting time and process time into account.

4. Conclusion and further research

This paper presents the idea of improving data integrity in production control through a combination of association rule induction and a resource-centered estimation of start- and end-dates. The focus lies on a modification of the apriori algorithm known from market basket analysis. The modified apriori algorithm can be further improved by taking longer antecedent sequences in consideration for estimating the missing work stations, although the minimal support will have to be lowered in order to get a satisfying subset of frequent sequences. Further research is necessary to determine a acceptable minimal support in this field of application.

As with any estimation, the described method may cause a bias within the completed data which can lead to wrong conclusions. One way to validate the method will be to investigate whether it is possible to estimate data which are not really missing in the feedback data set by comparing the algorithm results with the actual data. For this end, the calculation of the algorithm needs to be automated. In order to further validate the results of the algorithm it is necessary to have feedback data which can be presumed correct without any doubt. A cyber-physical demo factory is currently built on campus of RWTH Aachen University as part of the Cluster of Excellence "Integrative Production Technology for High-Wage Countries" which will be able to provide feedback data in the necessary quality.

Ensuring data integrity in production control will become even more important in the future when data is gathered via multiple sensors simultaneously in cyber-physical systems. As more data is gathered, more types of data inconsistencies will become relevant. In order to deal with these new conditions, an extension of the described methods will be necessary.

Acknowledgements

The authors would like to thank the German Research Foundation DFG for the kind support within the Cluster of Excellence "Integrative Production Technology for High-Wage Countries".

References

[1] Wiendahl, H.-P., 2008. Betriebsorganisation für Ingenieure (Enterprise organisation for engineers). Munich: Hanser
[2] Schuh, G., 2011. Produktion am Standort Deutschland (Production in Germany). Survey. Retrieved from http://www.win-d.de
[3] Lödding, H., 2008. Verfahren der Fertigungssteuerung (Methods of Production Control). Berlin: Springer
[4] Stevenson, M.; Hendry, L. C.; Kingsman, B. G., 2005. A review of production planning and control: The applicability of key concepts to the make-to-order industry. In: International Journal of Production Research 43 (5), pp. 869–898.
[5] Schuh, G.; Gottschalk, S.; Höhne, T., 2007. High Resolution Production Management. In: CIRP Annals - Manufacturing Technology 56 (1), pp. 439–442
[6] Schuh, G., 2006. Produktionsplanung und –steuerung: Grundlagen, Gestaltung und Konzepte (Production Planning and Control: Principles, Design and Concepts). Berlin: Springer
[7] Georgiadis, P., Michaloudis, C., 2012. Real-time production planning and control system for job-shop manufacturing: A system dynamics analysis. In: European Journal of Operational Research 216 (1), pp. 94-104
[8] Harding, J. A., Shahbaz, M., Srinivas, Kusiak, A., 2006. Data Mining in Manufacturing: A Review. In: Journal of Manufacturing Science and Engineering 128 (4), pp. 969-976

[9] Chen, M., 2003. Configuration of cellular manufacturing systems using association rule induction. In: International Journal of Production Research 41 (2), pp. 381-395
[10] Kwak, C.; Yih, Y., 2004. Data-mining approach to production control in the computer-integrated testing cell. In: IEEE Transactions on Robotics and Automation 20 (1), pp. 107-116
[11] Buzacott, J. A.; Corsten, H.; Gössinger, R.; Schneider, H. M., 2013. Production Planning and Control: Basics and Concepts. Munich: Oldenbourg
[12] Borgelt, C.; Kruse, R., 2002. Induction of Association Rules: Apriori Implementation. In: Proceedings of the 15th Conference on Computational Statistics, pp. 395-400
[13] Agarwal, R.; Imilienski, T.; Swami, A., 1993. Mining association rules between sets of items in large databases. In: ACM SIGMOD Record 22 (2), pp. 207-216
[14] Agarwal, R.; Sikrant, R., 1994. Fast Algorithms for Mining Association Rules. In: Proceedings of the 20th VLDB Conference 1215, pp. 487-499