Impact of Modeling Production Knowledge for a Data Based Prediction of Transition Times

Günther Schuh, Jan-Philipp Prote, Philipp Hünnekes, Frederick Sauermann, Lukas Stratmann

Impact of modeling production knowledge for a data based prediction of transition times

Günther Schuh[1], Jan-Philipp Prote[1], Philipp Hünnekes[1], Frederick Sauermann[1], Lukas Stratmann[1]

[1] Laboratory for Machine Tools and Production Engineering (WZL), RWTH Aachen University, Aachen, Germany
{g.schuh,j.prote,p.huennekes,f.sauermann,l.stratmann}
@wzl.rwth-aachen.de

Abstract. An increasing demand for customer-specific products is a major challenge for manufacturing companies. In many cases, companies attempt to satisfy this demand by increasing the number of product variants. In those companies, cost-oriented production processes have to be transformed into flexible workshop or island production structures in order to be able to produce this variety. This leads to an increasing complexity of production and subsequently planning. In order to reliably meet due dates, it is necessary to improve the quality of planning. This paper presents an approach for predicting transition times, the times between two production steps, by employing machine learning methods. In particular, the influence of the modelling of production knowledge of experienced employees on the prediction quality compared to a pure optimization of the methods' parameters is investigated.

Keywords: Production planning, machine learning, transition times.

1 Introduction

A trend of an increasing demand for customer-specific products can be observed. [1,2] The increasing number of variants forces many companies to transform their former cost-oriented flow production structures into flexible workshop or island production structures. [3] Typically, in the latter ones lead times (LT), logistics efforts and costs tend to increase, while the delivery reliability tends to worsen. [4] Today, most companies' highest logistical target is a high logistical performance and especially a high delivery reliability. [3] In times of an ever increasing globalization, a high logistical performance is crucial for a strong competitiveness, as markets all over the world have changed mostly from sellers' to buyers' markets. [4] In those, a high planning reliability plays a central role in achieving a high logistical performance. [5]

In principle, companies can positively influence the delivery reliability by measures such as additional shifts, employee overtimes, short-term outsourcing or targeted overcapacities. However, those measures come at high costs. Those costs are a challenge for companies, especially for those in high-wage countries. An approach for acting more proactively and achieving a high delivery reliability is a more accurate prediction

of order LT. Those LT consist mainly of transition times (TT), the times between two processing steps, which consist of waiting times before and after processing steps and transport times. TT scatter in general widely, which is why companies struggle to predict TT properly. [6] This paper aims at analyzing the impact of modeling and using production knowledge on the accuracy of TT prediction compared to a pure optimization of a prediction model's parameters using only raw feedback data. The research question is how much the prediction accuracy is influenced by these two tasks.

In the following section the challenges and consequences of a more accurate prediction are discussed briefly. In section three the state of the art of TT prediction is presented. Section four presents the conceptual design on the developed approach of an accurate TT prediction. In section five, the impact of modelling production knowledge of experienced employees is investigated and compared to a pure optimization of parameters being used in a machine learning (ML) method. Section six summarizes the findings and gives an outlook on future research.

2 Challenges and consequences of an accurate transition time prediction

An accurate prediction of TT is a challenging and uncertain task. [7] Production planers often attempt to cope with this uncertainty by allocating time buffers in the manufacturing process. This leads to a phenomenon called vicious cycle of production planning that was already described in the 1970s. Allocating time buffers leads to an earlier release of orders. Assuming orders are not finished substantially earlier, an increasing work load at work stations leads to longer waiting queues. Caused by this, orders need even longer than planned and the LT scatter more widely. [8] Instead of allocating time buffers, the right approach would be to predict LT order-specifically. By this, LT scattering decreases while process and planning reliability increases. However, current planning systems do not support a more accurate prediction of TT. [9] Thus, companies often use static estimations of TT. On the other hand, research shows that a more accurate prediction of TT is possible. [4]

Yet, there are difficult obstacles to overcome when predicting TT. In general, a profound and error-free database is a prerequisite for a robust and accurate prediction. [2] However, companies often do not possess such a database. Often they have a high degree of missing or erroneous data [10] and/or they do not have a large amount of non-redundant feedback data. As TT are related to orders in general and processing steps in particular, there is a limitation in the number of data entries per attribute by the number of orders or processing steps. Each attribute represents a data source, such as sensors, feedback data or master data. Generating more data points per order or processing steps would only lead to a more redundant data set that does not bring substantial benefits. An additional challenge is the data representation. Production data is characterized by a high heterogeneity in type, structure, semantics, organization and granularity, which is difficult to handle for most ML methods. [11]

Besides the mentioned challenges, research states that it is important to model human expertise and especially production knowledge for a transformation of raw data into

useful data. [12,13] Thus, in section four, an approach for modeling this production knowledge into a data based prediction of order-specific TT is presented. In section five, the approach is applied both with and without production knowledge to a data set containing real feedback data of a producing company from the machine equipment industry. By doing so, the impact of modeling production knowledge is investigated and compared to a pure parameter optimization of a ML method.

3 State of the art of the prediction of transition times

In the following, eight papers are presented that have been identified as the most relevant ones in a semi-structured literature review with a total of 75 analyzed papers. In [6] LT estimations based on two different decision tree algorithms are compared for different basic job shop production configurations. It is found that feature selection has a significant impact on the prediction accuracy. At best, a deviation as high as 18 % between predicted and actual LT is achieved. In [4] an approach of LT prediction in a tool shop based on real industry data is presented. In [14] both a feature selection for and prediction of cycle times in a simulation model of a use case from the semiconductor industry is described. In [15] a short term prediction model for LT in a semiconductor factory is proposed. In [16] three deep neural networks are compared for predicting order completion times. In [2] the use of a multivariate linear model, a non-linear random forest and a support vector machine algorithm is analyzed, all using real feedback data from a flow-shop production in the optics industry. In [17] LT in a semiconductor factory are predicted for three process steps by applying eleven different prediction algorithms. In [18] cycle times of wafer lots in a company from the semiconductor industry are predicted when releasing orders into production.

Analyzing the presented papers in detail, certain improvement measures can be derived. First of all, all papers focus on the prediction of LT that do not only contain highly scattering TT but also processing times. As processing times are often dependent on technological influencing factors whereas TT often depend on organizational factors, a separate analysis of both is recommended. In some of the papers significant reductions of the inspection areas have been made. For instance, in [16] only 12 workstations are considered, while in [6] only 6 workstations are considered. With the inspection area being to small, it can be explained, why in some cases selected features do not change their importance for the accuracy of LT prediction with respect to different production configurations. In general, a high usage of simulation data can be observed (in five of the twelve papers). Simulation data has in principle a high quality in terms of fewer instances of missing data or outliers. This characteristic is favorable for data analysis methods but real world applications are different. In four papers, use cases from the semiconductor industry are selected. As per [18], semiconductor factories are among the most digitized ones. With prediction accuracy is highly dependent on the number of training data, it is obvious, that an application in different industries is challenging. Lastly, it can be observed that e.g. in [14] feedback data from production is used for the prediction of LT, which is not known a priori, i.e. at the time when predicting LT. Hence, the accuracy of such a model decreases, when removing such features

that were identified as relevant for prediction. In the following section, an approach is presented that addresses the mentioned challenges and weaknesses.

4 Conceptual design of an approach for an accurate prediction of transition times

Essential for any data analysis is the quality of the underlying data set [19]. Especially the feedback data of highly differentiated and complex processes of medium-sized machine equipment companies pose a challenge for predicting TT. As a result, this approach focuses on generating a high quality and informative data set.

In research, two approaches have shown to be effective methods for increasing data quality (see also section three): *Feature selection* and *feature engineering* [21,20,22]. Features are the attributes which characterize a specific data set [23]. In the process of *feature selection*, an analyst selects the features to decrease the total number of features which improves modeling accuracy [14,22]. In *feature engineering*, a data scientist combines preexisting features or creates new features from scratch. Thus, *feature engineering* enables the opportunity to enrich a data set with knowledge from production experts, increasing the accuracy of TT predictions beyond the level of raw data sets. This task is especially challenging regarding the necessity of understanding the underlying production system as well as the available data set. Neither a data scientist nor a production expert can easily provide both.

Due to a potential extensive effort of generating a significant benefit from *feature engineering*, a systematic approach is fundamental [24]. This challenge has been addressed by several researchers, but mostly with high level and general concepts. In principle, a data scientist should begin by selecting obvious features such as dates and durations, and then continue to generate individual features such as LT or percentiles of a feature [20]. [24] suggest a three-phase method consisting of *exploring* the data, *extracting* relevant features and *evaluating* the engineered features. However, both approaches offer very few concrete actions and instead focus on the ability of a data scientist to understand the underlying system. The presented approach is based on previous work with the extension to include production experts into the process [25].

For *feature selection*, the most common method uses and ranks the set of features according to their relevance. The wrapper method accomplishes that by employing the learning algorithm and returning the features' impact on the prediction accuracy [26].

Having generated the high-quality data set, the learning model's performance can significantly be improved by optimizing its hyperparameters. Those are parameters of modeling algorithms that can be tuned to optimize the fit of a model to a data set [23].

Considering the steps above, we applied and evaluated the impacts of *feature engineering*, *feature selection* and *hyperparameter optimization* on a set of production data of a German medium-sized manufacturing company as follows.

5 Investigating the impact of modeling production knowledge for usage in ML methods

Validating the thesis of creating a significant benefit by enriching production data with additional production knowledge, the used raw set of production data consisted of real production information of approx. one year taken from a company's MES. Features included for analysis were e.g. the number of orders per day, number of operations per order, planned dates for starting and finishing the order, number of parts per order and estimated processing times by the processing production planner. Summing up, the data set contained 25 features and 100.000 data entries. In this, total LT accounts for 458,000 hours, of which TT has a share of 78 %. The used features are all known a priori.

In a first step, we designed a case-based approach to infuse additional knowledge into the data. Then, basic analysis on the data set was conducted, promptly identifying two findings. First, the TT per processing step varied with the time left until the order was completed. Second, the workload of the factory had a significant effect on the TT. By extracting daily data of all processed orders, new features were created accumulating the processing times of orders outstanding (i.e. calculating the workload of that day). Beyond that, the production expert stated that the work schedule (i.e. planned downtimes) causes a great portion of the TT. With this knowledge, the created feature eliminates planned downtimes such as weekends or nights for machines not operating at these times. This led to a decrease of TT to 44 % of LT. After implementing all engineered features, the data set included about 90 features.

In the process of generating an optimal data set, we applied some basic *data preparation* methods to clean the data of missing and incorrect entries. In accordance to [27], instances with missing values were ignored, if their total appearances summed up to less than one percent of total instances and the median was inserted in residual cases. However, for some specific features missing values were interpreted as nulls, if missing values are allowed in the first place, e.g. additional processing times or priority.

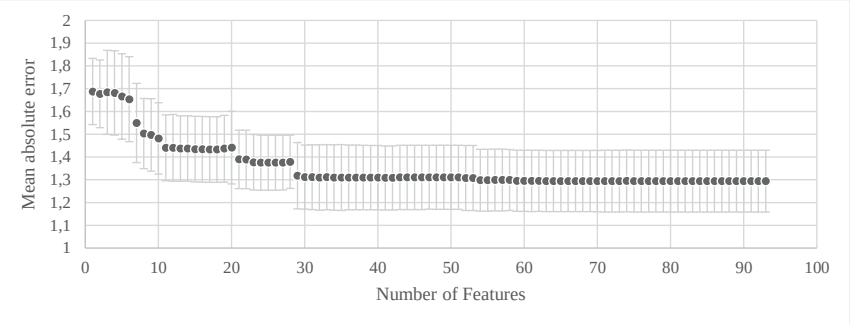

Fig. 1. Results showing the prediction error per number of used features

Following, the wrapper method iteratively computed TT predictions adding a new feature each iteration to find the most influential ones. As seen in **Fig. 1**, the steps indicate a severe improvement when specific features were added to the model. Shown is the mean absolute error (MAE) of the TT predictions in working days that decreases

over the number of features included into the model. Out of the best 25 features, we engineered 14 through expert knowledge. Among the most influential features are e.g. the time prediction by the production planner or the order-specific amount of time left to finish the complete order.

For modeling, we used python as programming language including its ML library. In particular, the scikit Regression Tree was used to perform the TT predictions [28]. Optimizing its performance, the following hyperparameters were adjusted: Decision tree depth, minimum number of samples to be at a leaf node and minimum number of samples required to split a node. To evaluate the engineered features in comparison to the raw data set, four scenarios were analyzed. The scenarios I and II use only the raw data, while III and IV use the enriched data set. While I and III have default hyperparameters, in II and IV those hyperparameters were optimized.

As seen in **Fig. 2**, the TT prediction error is lowest for the data set that includes optimized hyperparameters as well as selected and engineered features. It reaches a 14.6 percent lower MAE than the data set without engineered features and thus shows a significant improvement. The effect of optimizing the hyperparameters is explicitly high for the decision tree due to its tendency of overfitting and therefore shows a great improvement of about 25 percent for the raw data set and 18 percent for the enriched data set. One can assume that engineering features helps the fitting of the model by providing more informative features, which leads to a lower need for optimization. This could prove important in cases of high amount of data, in which an evaluation, such as this, would need too much time to compute.

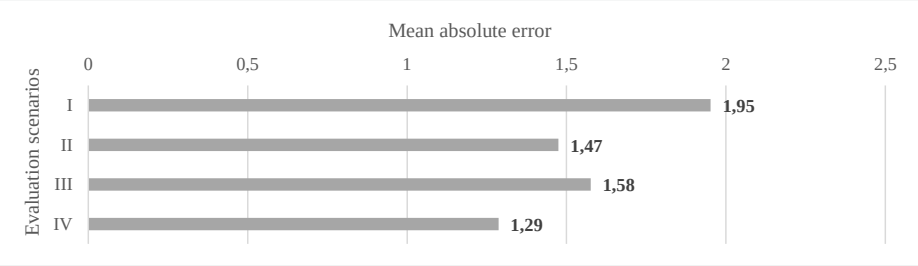

Fig. 2. Mean absolute error of transition time predictions in working days

6 Summary and outlook

In this paper, the importance of a more accurate prediction of TT, that represent the time between production process steps, has been discussed. By using more accurate TT, logistical performance can significantly be approved by simultaneously reducing costs. Based on a CRISP-DM based approach for a order-specific prediction if TT, a validation with real production feedback data took place. Modeling production knowledge in combination with optimizing a ML method's hyperparameters decreased prediction errors by 44 %.

In the future, a profound evaluation is planned on what type of production knowledge can improve the prediction accuracy even further and how that knowledge can be modeled for a use in ML methods. One concrete example will be to evaluate the impact of modeling order sequence corrections at machines. At last, the integration in a company's planning process needs to be conceptualized, including a control loop for a self-learning development of the prediction model in order to maintain a high prediction accuracy. A major challenge for this control loop is to deal with the fact that production managers will actively influence the TT based on the predicted length.

7 Acknowledgement

The authors would like to thank the German Research Foundation DFG for the kind support within the Cluster of Excellence "Internet of Production" (ID: 39062161).

References

[1] ElMaraghy, H., Schuh, G., ElMaraghy, W., Piller, F., Schönsleben, P., Tseng, M., Bernard, A., 2013. Product variety management. CIRP Annals 62 (2), 629–652.

[2] Gyulai, D., Pfeiffer, A., Nick, G., Gallina, V., Sihn, W., Monostori, L., 2018. Lead time prediction in a flow-shop environment with analytical and machine learning approaches. IFAC-PapersOnLine 51 (11), 1029–1034.

[3] Duffie, N., Bendul, J., Knollmann, M., 2017. An analytical approach to improving due-date and lead-time dynamics. Journal of Manufacturing Systems 45, 273–285.

[4] Berlec, T., Starbek, M., 2010. Forecasting of Production Order Lead Time in SME's, in: Füerstner, I. (Ed.), Products and services. Sciyo, Rijeka, Croatia.

[5] Schuh, G., Brettel, M., Reuter, C., Bendig, D., Dölle, C., Friederichsen, N., Hauptvogel, A., Kießling, T., Potente, T., Prote, J.-P., Weber, A., Wolff, B., 2017. Towards a Technology-Oriented Theory of Production, in: Brecher, C., Özdemir, D. (Eds.), Integrative Production Technology. Theory and Applications. Springer, pp. 1047–1079.

[6] Öztürk, A., Kayalıgil, S., Özdemirel, N.E., 2006. Manufacturing lead time estimation using data mining. European Journal of Operational Research 173 (2), 683–700.

[7] Pfeiffer, A., Gyulai, D., Monostori, L., 2017. Improving the Accuracy of Cycle Time Estimation for Simulation in Volatile Manufacturing Execution Environments, in: Wenzel, S., Peter, T. (Eds.), Simulation in Produktion und Logistik 2017. pp. 413–422.

[8] Mather, H., Plossl, G., 1978. Priority Fixation versus Throughput Planning. Journal of Production and Inventory Management (19), 27–51.

[9] Niehues, M. R., 2016. Adaptive Produktionssteuerung für Werkstattfertigungssysteme durch fertigungsbegleitende Reihenfolgebildung. Dissertation, München.

[10] Schuh, G., Reuter, C., Prote, J.-P., Brambring, F., Ays, J., 2017. Increasing data integrity for improving decision making in PPC. CIRP Annals 66 (1), 425–428.

[11] Chen, M., Mao, S., Liu, Y., 2014. Big Data: A Survey. Mobile Netw Appl 19 (2), 171–209.

[12] Guyon, I., Elisseeff, A., 2006. An Introduction to Feature Extraction, in: Kacprzyk, J., Gunn, S., Guyon, I., Nikravesh, M., Zadeh, L.A. (Eds.), Feature extraction. Foundations and applications. Springer, Berlin, Heidelberg, pp. 1–25.

[13] Niggemann, O., Biswas, G., Kinnebrew, J.S., Khorasgani, H., Volgmann, S., Bunte, A., 2017. Datenanalyse in der intelligenten Fabrik, in: Vogel-Heuser, B., Bauernhansl, T., Hompel, M. ten (Eds.), Handbuch Industrie 4.0. Bd. 2 : Automatisierung, 2., erweiterte und bearbeitete Auflage ed. Springer Vieweg, Berlin, pp. 471–490.

[14] Meidan, Y., Lerner, B., Rabinowitz, G., Hassoun, M., 2011. Cycle-Time Key Factor Identification and Prediction in Semiconductor Manufacturing Using Machine Learning and Data Mining. IEEE Trans. Semicond. Manufact. 24 (2), 237–248.

[15] Tirkel, I., 2013. Forecasting flow time in semiconductor manufacturing using knowledge discovery in databases. International Journal of Production Research 51 (18), 5536–5548.

[16] Wang, C., Jiang, P., 2017. Deep neural networks based order completion time prediction by using real-time job shop RFID data. J Intell Manuf 19 (4), 1–16.

[17] Lingitz, L., Gallina, V., Ansari, F., Gyulai, D., Pfeiffer, A., Sihn, W., Monostori, L., 2018. Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer. Procedia CIRP 72, 1051–1056.

[18] Wang, J., Zhang, J., Wang, X., 2018. A Data Driven Cycle Time Prediction With Feature Selection. IEEE Trans. Semicond. Manufact. 31 (1), 173–182.

[19] Rahman, M.G., Islam, M., 2012. A Decision Tree-based Missing Value Imputation Technique. Conferences in Research and Practice in Information Technology Series 121.

[20] Kanter, J.M., Veeramachaneni, K., 2015. Deep feature synthesis: Towards automating data science endeavors, in: Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Campus des Cordeliers, Paris, France. IEEE, Piscataway, NJ, USA, pp. 1–10.

[21] Guyon, I. (Ed.), 2006. Feature extraction: Foundations and applications. Springer, Berlin, 778 pp.

[22] Zhang, C., Kumar, A., Ré, C., 2016. Materialization Optimizations for Feature Selection Workloads. ACM Trans. Database Syst. 41 (1), 1–32.

[23] Witten, I.H., Pal, C.J., Frank, E., Hall, M.A., 2017. Data mining: Practical machine learning tools and techniques, Fourth edition ed. Morgan Kaufmann, Cambridge, MA, 641 pp.

[24] Anderson, M.R., Antenucci, D., Bittorf, V., Burgess, M., Cafarella, M.J., Kumar, A., Niu, F., Park, Y., Ré, C., Zhang, C., 2013. Brainwash: A Data System for Feature Engineering, in: CIDR.

[25] Schuh, G., Prote, J.-P., Luckert, M., Sauermann, F., 2018. Determination of order specific transition times for improving the adherence to delivery dates by using data mining algorithms. Procedia CIRP 72, 169–173.

[26] Kacprzyk, J., Gunn, S., Guyon, I., Nikravesh, M., Zadeh, L.A. (Eds.), 2006. Feature extraction: Foundations and applications. Springer, Berlin, Heidelberg.

[27] Grzymala-Busse, J.W., Hu, M., 2001. A Comparison of Several Approaches to Missing Attribute Values in Data Mining, in: Rough Sets and Current Trends in Computing. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 378–385.

[28] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.