Machine Learning goes Measure Management: Leveraging Anomaly Detection and Parts Search to Improve Product-Cost Optimization

Matthias Walter[1], Volodymyr Vasyutynskyy[2], Duc Anh Trinh[1], and Christian Leyh[1]

[1] Technische Universität Dresden, Chair of Information Systems, esp. IS in Manufacturing and Commerce, Dresden, Germany
{matthias.walter3,duc_anh.trinh,christian.leyh}@tu-dresden.de
[2] SAP SE, Innovation Center Network, Dresden, Germany
{volodymyr.vasyutynskyy}@sap.com

Abstract. In many industries, particularly discrete manufacturing, companies can benefit from conducting product-cost optimization as early as possible. Given the amount of data to be analyzed in the costing process, the lack of dedicated information system support, and the pressure to quickly estimate the cost of new products, the potential for cost optimization is often underexploited. In this paper, we present an approach for leveraging machine learning capabilities, including similarity and anomaly analysis, to improve the identification of product-cost optimization potentials and therefore, improve the quality of early cost estimates. For the approach to succeed, however, ongoing training of a model based on a high-quality dataset is crucial. Thus, we also propose the machine learning approach's integration with our long-term research project toward improving the management of cost optimization during product development.

Keywords: Product costing, product-cost optimization, machine learning, product development, measure management.

1 Problem Context

To ensure long-lasting economic success amid globalization, shortened product life cycles, and growing product diversity, companies seek support for conducting product-cost optimization [1], [2]. Despite the immense potential of early cost optimization [3], [4], our prior research has shown that early optimization lacks dedicated information system (IS) support [2], [5]. In response, our long-term research project grounded on design science research methodology [6], [7] aims to improve IS support for early cost optimization. Based on industry-evaluated requirements and IS implementation hurdles, we developed an approach that aims at improving the management of cost-optimization measures, supporting the phases of identifying, elaborating, evaluating, and implementing optimization measures. This measure management (MM) approach, including optimization examples, is depicted in [5] and [8] in detail.

14th International Conference on Wirtschaftsinformatik,
February 24-27, 2019, Siegen, Germany

In this context, this paper takes up an idea that emerged from our research project: The automated identification of optimization potentials [2], [8]. Such a functionality is especially important in the discrete manufacturing industry, where products such as cars and special-purpose machines comprise up to 100,000 components. Due to this and further practical challenges [2], [8], the complexity of optimization is hardly manageable manually. Therefore, industry demands IS approaches to support the identification of optimization potentials by using available datasets [5], [9].

Thus, on the one hand, this paper showcases machine learning (ML) approaches for identifying optimization potentials. On the other hand, the paper identifies a means of improving the accuracy of ML algorithms by incorporating the MM approach from our long-term research project.

2 Machine Learning goes Measure Management

2.1 Measure Identification

To demonstrate the potential of ML for early product-cost optimization, we first introduce two exemplary optimization use cases from industry [5]. We use these to develop suitable approaches for self-optimizing algorithms. These approaches then form the foundation for demonstrating the interaction between the MM approach based on [8] and self-optimizing algorithms that support the identification of cost optimization potentials. This paper summarizes such self-optimizing algorithms under the term ML according to the definition in [10], which defines ML as a specific system design that allows the system to learn from data and improve with experience [10].

The first approach, a similar parts search, helps users locate similar but preferable (e.g., cheaper) materials [9]. For that, a user selects a certain material in a product costing structure. Based on dedicated material attributes like description, material number, weight, plant, etc., the approach starts the search for materials with similar attributes in the master data and previous costing structures. The single similar attributes are found using fuzzy search [11], with subsequent defining of the overall degree of similarity of the target and similar materials as a weighted sum of single similarity scores. The weight coefficients are defined using an optimization algorithm which analyzes the historical similar products. As result, the ML algorithm recommends a list of similar materials, ranked by their degree of similarity to the one selected [12]. This list of similar but possibly more suitable materials can be used to create an optimization measure as described in [5]. By creating such a measure, the user can accept the respective recommendation to replace the originally selected material.

The second approach involves plausibility checks, based on the work of Vosough and Vasyutynskyy [13], whose main goal was identifying potential user-made errors by assessing the validity and plausibility of items within a costing structure. Since only a small set of examples of such errors is known, which makes usage of the supervised ML approaches difficult, Vosough and Vasyutynskyy concentrated on the unsupervised model-based algorithms for anomaly detection. These algorithms detect potential errors in cost estimates as anomalies, i.e., significant deviations from the values of typical material attributes (e.g., price and weight) or typical costing structures, in which the attribute values and structures are learned from historical cost estimates. The detected anomalies are then converted into human readable messages [13], which can be used to create a dedicated optimization measure [5]. Different anomaly detection algorithms can be used [14], whereas we identified the adjusted approach based on the Z-Score of the data points as most robust for different use cases in our context. In this adjusted approach, the thresholds for anomaly detection are not constant, but are defined dynamically dependent on the number of samples, their statistical distribution as well as user-defined settings (patent filed [15]). The approach was validated and adjusted based on real data and feedback from four large, international companies.

Figure 1 drafts ML recommendations from our research project and how both approaches could potentially be integrated with IS for product costing. The left side shows the similar parts search containing the list of similar materials for a hexagonal screw, which can be replaced by either the same material from another plant (PT1) or another material (M16). The right side shows results of plausibility checks. In addition to the anomalies identified, the potential cost impact of these anomalies on the product-cost estimate are also calculated and displayed.

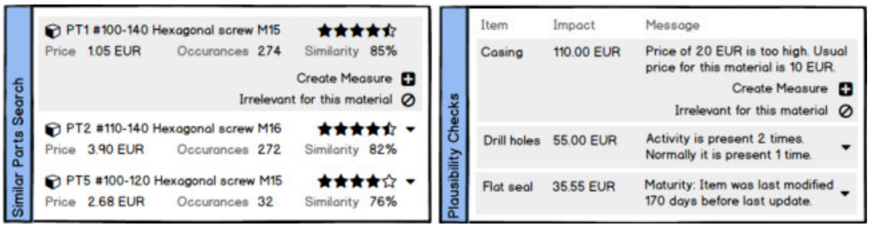

Figure 1. User interface mockups for the similar parts search and plausibility checks

2.2 Problem Case: Labeled Data

Both the similar parts search and plausibility checks can be realized with unsupervised [16] and supervised learning [17]. The latter requires a high quantity and quality of data (e.g., labels, master data, costing information, and bill of materials) to accurately train the ML model. As mentioned above, it is difficult to obtain the required large labeled datasets, given the need for extraordinary manual effort and the variety of potential errors in the costing structures [18]. Moreover, specific domain knowledge in product costing is required to accurately label the data [2], [9], [19].

To address those setbacks, which are quite common in ML [20], the presented approaches start with unsupervised learning using fuzzy search [11] as well as statistical and hierarchical algorithms [14]. Although these approaches are already a good support in improving the product-cost estimate quality, they require company-specific calibration and do not deal well with special use cases, which may result in false positives or missed errors. To combine the ML with the knowledge from the user without overwhelming the latter with additional efforts, we propose integrating the ML algorithms with our approach of MM [8]. We extend the approach with supervised learning to further improve the ML algorithms and thereby raise the quality of optimization recommendations (Figure 2).

With the initial ML model trained by unsupervised learning, we use inference to generate recommendations. The user now has the option to either mark the recommendations as irrelevant or, if it seems reasonable, initiate an optimization measure for further processing (Figure 1). All recommendations marked as irrelevant are flagged as false positives. With the initiation of an optimization measure, the respective recommendation is labeled implicitly as true positive and later labeled again after the measure has successfully been applied to the product [5].

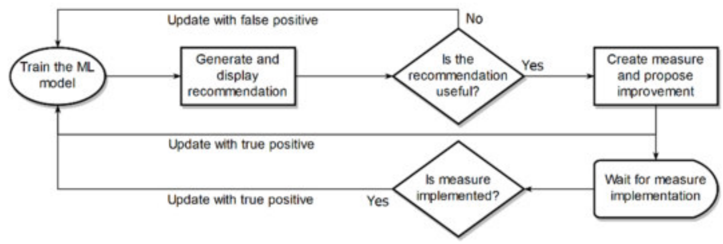

Figure 2. Measure management as an enabler of supervised learning

In sum, positive and negative feedback is used to retrain the models so future recommendations are more accurate (Figure 2). Thus, the user's feedback is reused without additional effort for labeling data. In that way, we can continuously improve learning results for both ML approaches and increase the accuracy of algorithms, even if optimization recommendations are not ultimately applied to the product.

3 Status Quo and Outlook

The initial unsupervised learning of the described optimization approaches has already been implemented in the context of our research partner's system SAP Product Lifecycle Costing (see [21]) to propose optimization potentials [12], [13]. To date, unsupervised learning has been performed with a set of real data representing four large, international companies in the discrete manufacturing industry, each with more than 10,000 employees. For them, plausibility checks detect anomalies and potential errors for up to 5% of all available line items in the provided costing structures. This error rate is surprisingly high, which can be explained on the one hand by undiscovered errors caused by manual editing of costing structures with up to 100,000 line items and on the other hand by the fact that the unsupervised approaches result in false positives in specific use cases. Therefore, we look forward to validating and further training of the ML models with the help of our presented approach.

With continuous improvement of the models and more profound results, further automation of the process is possible. For example, the most certain recommendations can be automatically implemented for entire product development projects with minimal manual effort. Even the creation of new costing structures based on identified typical costing structures and user-provided parameters is conceivable.

References

1. Stark, J.: Product Lifecycle Management: 21st Century Paradigm for Product Realisation, 3rd ed. Springer, London (2015)
2. Walter, M., Leyh, C.: Knocking on Industry's Door: Needs in Product-Cost Optimization in the Early Product Life Cycle Stages. In: Proceedings of the IEEE 19th Conference on Business Informatics (CBI 2017), pp. 330–338. Thessaloniki, Greece (2017)
3. Eigner, M., Stelzer, R.: Product Lifecycle Management: Ein Leitfaden für Product Development und Life Cycle Management. Springer, Heidelberg (2009)
4. Ehrlenspiel, K., Kiewert, A., Lindemann, U.: Cost-Efficient Design. Springer, Berlin (2007)
5. Walter, M., Leyh, C., Strahringer, S.: Knocking on Industry's Door: Needs in Product-Cost Optimization in the Early Product Life Cycle Stages. Complex Systems Informatics and Modeling Quarterly 13, 43–60 (2017)
6. Peffers, K., Tuunanen, T., Rothenberger, M. A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24, 45–77 (2008)
7. Peffers, K., Tuunanen, T., Niehaves, B.: Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research. European Journal of Information Systems 27 (2), 129–139 (2018)
8. Walter, M., Leyh, C., Strahringer, S.: Toward Early Product Cost Optimization: Requirements for an Integrated Measure Management Approach. In: Proceedings of the Multikonferenz Wirtschaftsinformatik 2018 (MKWI 2018), pp. 2057–2068. Leuphana University Lüneburg, Lüneburg (2018)
9. Mörtl, M., Schmied, C.: Design for cost - a review of methods, tools and research directions. Journal of the Indian Institute of Science 95, 379–404 (2015)
10. Bell, J.: Machine Learning: Hands-On for Developers and Technical Professionals. Wiley, Indianapolis (2015)
11. SAP HANA Developer Guide: FUZZY Search, https://help.sap.com/viewer/691cb949c103 4198800afde3e5be6570/2.0.01/en-US/cc602780bb5710148aa2bf6cab3c015b.html (Accessed: 13.11.2018)
12. Heidenreich, M.: Unlock the Potential of Machine Learning to Get Better Cost Estimates in the Early Stage of the Product Lifecycle, https://blogs.sap.com/2017/10/04/unlock-the-potential-of-machine-learning-to-get-better-cost-estimates-in-the-early-stage-of-the-product-lifecycle/ (Accessed: 30.08.2018)
13. Vosough, Z., Vasyutynskyy, V.: Using Parallel Sets for Visualizing Results of Machine Learning Based Plausibility Checks in Product Costing. In: Proceedings of VisBIA 2018: Workshop on Visual Interfaces for Big Data Environments in Industrial Applications, pp. 4–11. Grosseto, Italy (2018)
14. Mehrotra, K. G., Mohan, C., Huang, H.: Anomaly Detection Principles and Algorithms. Springer, Cham (2017)
15. Wasiutinski, V.: Plausibility-Based Validation of Product Costing Structures. US Patent Filed (2018)
16. Hastie, T., Tibshirani, R., Friedman, J.: Unsupervised Learning. In: The Elements of Statistical Learning, pp. 485–585. Springer, New York (2009)
17. Caruana, R., Niculescu–Mizil, A.: An Empirical Comparison of Supervised Learning Algorithms. In: Proceedings of the 23rd International Conference on Machine Learning (ICML '06), pp. 161–168. New York (2006)
18. Settles, B.: Active Learning Literature Survey. Technical report 1648, Computer Sciences. University of Wisconsin-Madison (2009)

19. Rush, C., Roy, R.: Expert Judgement in Cost Estimating: Modelling the Reasoning Process. Concurrent Engineering 9, 271–284 (2001)
20. Zhu, X.: Semi-Supervised Learning Literature Survey. Technical report 1530, Computer Sciences. University of Wisconsin-Madison (2005)
21. Product Costing | SAP Product Lifecycle Costing, https://www.sap.com/products/product-lifecycle-costing.html (Accessed: 3.09.2018)