4th Conference on Production Systems and Logistics

Beyond Pareto Analysis: A Decision Support Model for the Prioritization of Deviations with Natural Language Processing

Yuxi Wang[1], Lukas Longard[1], Christian Hertle[2], Joachim Metternich[1]
[1]Institute of Production Management, Technology and Machine Tools, Darmstadt, Germany
[2]SFM Systems GmbH, Darmstadt, Germany

Abstract

In the manufacturing domain, the systematic problem-solving (SPS) process is essential to eliminate the root causes of deviations from expected performance. The major goal of SPS is to prevent the recurrence of known deviations. However, due to time and resource limitations, the deviations that occur on the shop floor should be prioritized before applying SPS. Therefore, a method to support the decision-making process for prioritization of deviations is required. Traditional methods, such as the Pareto analysis, are widely accepted and applied for easy use. But their performance is no more sufficient for the production environment with large fluctuations nowadays. Therefore, this paper proposes a decision support model – the error score – to prioritize deviations on the shop floor. The error score is calculated based on the process data as well as textual data found in the deviation documentation. As the quality of textual data in the deviation documentation has great effects on the performance of the model, Natural Language Processing (NLP) methods are developed to pre-process the unstructured text. To validate the model, it is applied to a real-world use case in the automotive industry to demonstrate and evaluate the performance. The study shows that the proposed model can effectively support the decision-making process on the shop floor and is superior to traditional methods.

Keywords

Systematic problem-solving (SPS); Deviation management; Natural language processing (NLP); Shop floor management; Production and manufacturing

1. Introduction

In the manufacturing domain, the systematic problem-solving (SPS) process is an essential process to ensure the stability of production. The goal of SPS is to find the root causes of occurring problems and to develop solutions as well as new standards to prevent recurrences. However, due to time and resource limitations, the deviations that occur on the shop floor should be prioritized before applying SPS [1]. In some cases, shop floor managers decide on priorities based on their personal experience. Moreover, it is possible to support the decision-making process based on historical data by means of data analysis [2]. For example, Pareto analysis has been widely accepted and applied in production. However, their performance is no longer adequate for today's highly volatile production environment, since they can only consider one aspect (e.g. number of occurrences) at a time [3].

The development of digitalization provides the possibility to get access to more information from the production process and related activities [4]. More process- and problem-related data is documented through terminals on the shop floor, such as the digital shop floor management (dSFM) system, which enable the use of data analytics and machine learning algorithms to automatically analyze the data in real-time [5].

DOI: https://doi.org/10.15488/13483
ISSN: 2701-6277

However, in practical applications, there is often unstructured data, such as textual data, which can cause difficulties in the automatic analysis process. To overcome this issue, the methods of natural language processing (NLP) can be utilized in data pre-processing to remove the ambiguities present in the unstructured text [6]. Based on this, the major research contributions of this work can be summarized as follows:

- A decision support model to prioritize deviations on the shop floor is developed
- NLP is used in the preprocessing steps of the model to raise the data quality of the textual data
- An application of the model on a real-world use case in the automotive industry where the proposed solution is demonstrated and evaluated.

The remainder of this paper is structured as follows. Section 2 introduces the background and related work of this study. Section 3 establishes the decision support model aiming at prioritizing the deviations for the SPS on the shop floor. Section 4 conducts a case study on an automotive company using data from the rework process. The results are listed and discussed in Section 5. Finally, Section 6 concludes this study and suggests several directions for future research.

2. Background and related work

2.1 Shop floor management

The methods of shop floor management (SFM) are widely used in industry to control and improve production processes on a daily basis [7]. The shop floor control loop is based on identifying deviations through performance indicators (KPIs), discussion about the deviations in the regular shop floor meeting, initiating SPS and the stabilization as well as standardization of the processes [8]. With the increasing popularity of digitalization, the use of dSFM systems is now gradually increasing in production companies [9]. During the use of the dSFM system, data about deviations and problems is collected and stored in the system. This data includes both structured and unstructured data. The structured data includes information like the time of the deviation as well as the product number, while the unstructured data is mainly textual and includes for example, the description of the deviation, defined measures.

When deviations occur, there are usually two ways to counteract: On the one hand, if the cause is known, quick measures should be taken. On the other hand, if the cause is unknown, a SPS should be triggered [10]. Besides, quick measures have to be taken to prevent the problem from escaping to the customer. Compared to SPS, taking quick measures skips a thorough analysis of the deviation and can quickly restore production from the deviated state back to standard. However, in order to prevent recurrences of the issue, a SPS must be applied to analyze and solve the root cause of the problem [11]. Considering the time and labor costs, it is necessary to prioritize the deviations before applying SPS. There are not many existing methods in the literature that specifically mention how to prioritize deviations; the existing methods mainly include empirical-based and data analysis-based approaches. Shop floor managers determine the prioritization by referring to their own experience and anticipating the likely outcome of the problem. The severity of the problem can also be evaluated by analyzing the historical data of its occurrence. As a commonly used method for decision support processes, Pareto analysis is widely used on the shop floor, which can helps managers to identify top errors with Pareto chart [12]. Advanced statistical tools are also being applied to assist in the decision-making process, such as presenting important problems as graphs by means of text clustering [13].

2.2 Natural language processing

NLP is an interdisciplinary research area that combines computational linguistics, computational science, cognitive science and artificial intelligence to inquire how computers process human language to perform useful tasks [14]. Textual data is an important form of human language in the production, most of which are unstructured or semi-structured. Because few restrictions can be applied to those data, the likelihood of errors and inconsistencies is higher than in structured data [15]. Figure 1 shows typical data errors and inconsistencies in textual data on the shop floor. These errors can be separated into orthography and the content. Misspellings and synonyms are the significant properties that affect the quality of textual data [16]. As an essential part of NLP, the preprocessing step of textual data normalizes the unistructural data into a computer readable form. It consists of three steps: tokenization, normalization and vectorization [17]. In a recent study, Mueller et al. have identified best practices for the handling of shop floor textual data [6].

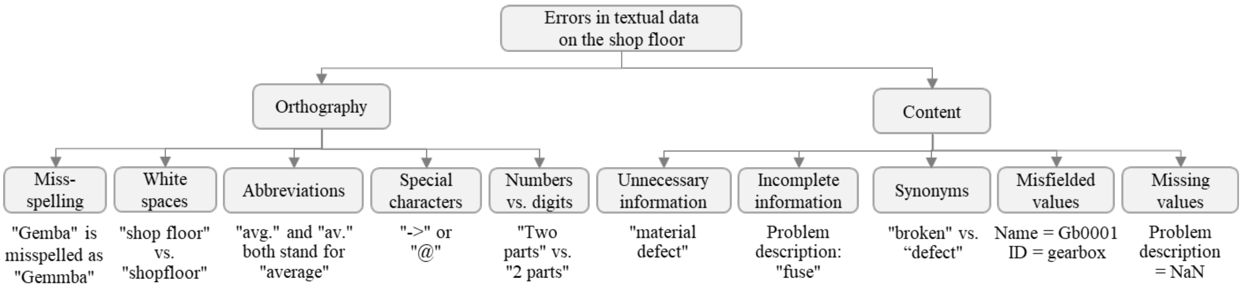

Figure 1: Typical errors in textual data on the shop floor

3. Structure of the decision support model

To improve resource allocation in form of SPS capacity on production deviations, this paper introduces a new approach to support deviation prioritization. The model in this study primarily consists of three parts: the data preprocessing with NLP, the calculation of the error score and the weight optimization. The data preprocessing analyses the deviation descriptions and utilizes NLP to raise the textual data quality. The calculation of the error score delivers the deviation ranking list based on the preprocessed textual data and the certain weights. The weight optimization adjusts the weights in a certain rhythm to enable a specific fit to the company's production environment in order to provide a better decision-making support.

For the sake of clarity, Figure 2 depicts the structure of the proposed decision support model.

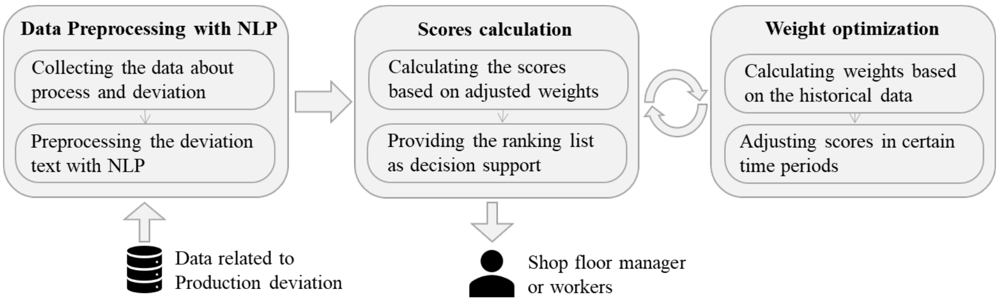

Figure 2: Structure of the decision support model

3.1 Data preprocessing with NLP

The calculation of the error score is based on the type of deviations. This process is highly related to the quality of the textual from the workers. If the quality is not sufficient, e.g., containing semantic similar description or spelling errors, this will affect the result of the designed decision-support system. Based on the analysis in Section 2.2, misspellings and abbreviations are selected to be addressed through NLP.

3.1.1 Preprocessing of misspellings

The open-source python module "Pyspellchecker" provides a function to correct misspelling. It lists a set of candidates for the input word by identifying whether the input word is included in the word frequency list, and gives the best correction for the input word [18]. However, this word frequency list is trained from existing corpus, which do not contain the words related to activities on the shop floor, and therefore cannot be directly applied to handle typos in textual data. Some adjustments are required. As shown in Figure 3, there are two operations in the preparation phase, i.e., enriching the existing dictionary and preprocessing the input textual data. At first, the shop floor-specific words including standardized abbreviations should be embedded into the word frequency list. The list employed to enrich the existing dictionary must be without spelling errors. After identifying the typos in the preprocessed data, the words that are not in the enriched dictionary are returned. Depending on the volume of the textual data, two different approaches can be used to correct the typos. If the volume is manageable, the system can list the candidates for each detected typo to be corrected manually. For large volumes, the system can automatically suggest the best correction for each identified typing error, inducing the risk of false corrections.

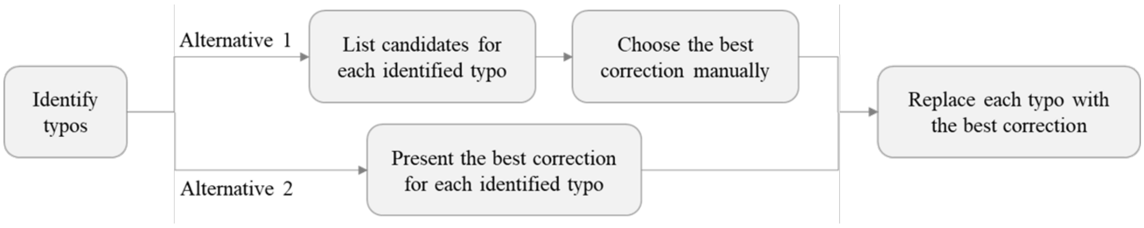

Figure 3: Pre-processing of misspellings

3.1.2 Preprocessing of abbreviations

The purpose of dealing with abbreviations is to normalize the abbreviations and thus reduce ambiguity. Based on the analysis of possible patterns in abbreviations, a rule-based algorithm is proposed. As shown in Figure 4, there are mainly three steps to check whether a token is an abbreviation. The process starts with the tokenization and filter of words. Only the abnormal words, which have potential to be abbreviations, are contained in this process to be checked. The check of regular form contains the following conditions:

- The token contains at least one and less than three letter
- The token only contains digits and letter belonging to the language in which the text is written
- The token does not begin with digits

If a token satisfies all these three conditions, it can then be saved to the list of candidates of abbreviations. If not, it is possible that the token contains special characters that affects the first check. The special characters refer to all characters, excluding digits, letters, and dots. Therefore, the special character should be transformed into a whitespace character. Those transformed spaces that locate at the beginning and the end of the original token shall be then removed. Subsequently, the process returns to the first step again. If no more special characters can be found in the token, the third step is to be performed. In the check of special form of abbreviation, the following conditions should be fulfilled:

- A special pattern can be observed in the token, which is: a letter is followed directly by a dot
- This pattern appears at least once but less than four times in the token

If a token satisfies all these three conditions, it can then be saved to the list of candidates of abbreviations. After the three steps, a list with abbreviations is generated. With the help of expert knowledge, the meaning of the abbreviations is explained and stored to normalize the abbreviation in the whole dataset.

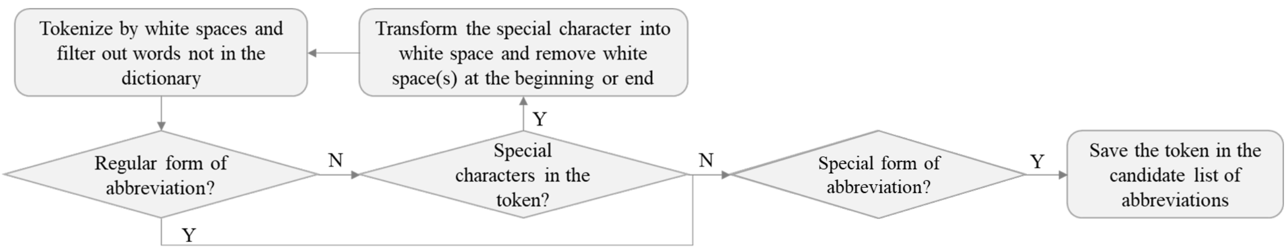

Figure 4: Pre-processing of abbreviations

3.2 Calculation of error score

A previous study by Longard et al. has shown the effectiveness of error scores in the field of rework processes in production [19]. To extend this model to more fields related to problem-solving, this work applies several adjustments to the original model, which are explained in the remainder of this section.

3.2.1 Error Score

Data analytics is applied to calculate an error score based on documented errors from the past. The following parameters are included: the total cost for the deviation and the trend of the occurrence. The goal is to faster identify errors and to prioritize which errors to focus on [20]. The error score is defined as

$$S_i^{error} = \frac{S_i^{uptrend} \cdot w_{trend} + S_i^{cost} \cdot w_{\text{cost}}}{w_{\text{trend}} + w_{\text{cost}}} \tag{3.1}$$

, where S_i^{error} is the error score for deviation i, S_i^{trend} and S_i^{cost} are the scores of the trend of occurrence and the cost of deviation. w_{trend} and w_{cost} are the weights for sub scores, with the constraint

$$w_{\text{trend}} + w_{\text{cost}} = 1 \qquad (3.2).$$

This constraint is set to ensure that the optimal solution is always found during the optimization process and that the two weights do not move in the same direction at the same time. The calculation of the S_{trend} and S_{cost} should consider both process data and deviation documentation in the historical data. The formular is

$$S_i^{trend'} = \begin{cases} \frac{\sum_{d=0}^{m} C_i^{x-d}}{\sum_{d=0}^{n} C_i^{x-d}}, if \sum_{d=0}^{m} C_i^{x-d} > \sum_{d=0}^{n} C_i^{x-d} \; and \; \sum_{d=0}^{n} C_i^{x-d} \neq 0 \\ 0, otherwise \end{cases} \tag{3.3}$$

, where $S_i^{trend'}$ is the score of time tendency without standardization. C_i^{x-d} is the cost of deviation i on day $x - d$. $\sum_{d=0}^{m} C_i^{x-d}$ is the sum of the cost for the deviation i in the last m days, and $\sum_{d=0}^{n} C_i^{x-d}$is the sum of the cost for the deviation in the last n days. With this function, the score considers whether the deviation cost in the recent m days is greater than the period of n days (considering $n > m$). If $S_i^{trend'}$ is zero, which means the error shows no trend in the recent. This way, the deviations with a rising frequency of occurrence are identified. After the normalization, the score of time tendency is defined as

$$S_i^{trend} = \frac{S_i^{trend'} - \min\limits_{1 \le i \le N} S_i^{trend'}}{\max\limits_{1 \le i \le N} S_i^{trend'} - \min\limits_{1 \le i \le N} S_i^{trend'}} \tag{3.4}$$

, where N is the total number of deviations. And

$$S_i^{cost'} = \sum_{d=0}^{s} C_i^{x-d} \tag{3.5}$$

, where $\sum_{d=0}^{s} C_i^{x-d}$ means the sum of the cost in the last s days. s is determined empirically, and usually in reality, when a deviation arises, the longer it causes to the production line, the larger the value of s. This gives a more complete image of the impact of the error. After the normalization, the score of cost is

$$S_i^{cost} = \frac{S_i^{cost'} - \min\limits_{1 \le i \le N} S_i^{cost'}}{\max\limits_{1 \le i \le N} S_i^{cost'} - \min\limits_{1 \le i \le N} S_i^{cost'}} \tag{3.6}$$

, where N is the total number of deviations. After the calculation of S_i^{trend} and S_i^{cost}, the error score can be calculated with the pre-defined weights.

3.3 Automatic weight determination and optimization

Since the severity of a deviation is affected by trends and costs differently at different times, the two weights controlling trends and costs, w_{trend} and w_{cost}, need to be updated at intervals so that the algorithm is adapted to the current state of production in each period. This is done by a novel optimization algorithm. The basic concept of the algorithm is described in Figure 5.

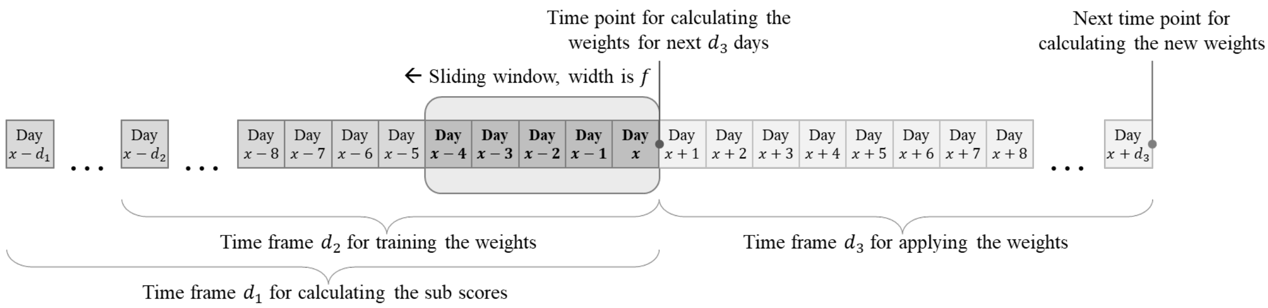

Figure 5: Weight optimization

The weight optimization process runs as follows:

- The weights w_{trend} and w_{cost} are updated in every d_3 days. The default length of d_3 is one month.
- On the day of weight optimization, the time frame for calculating the sub scores is d_1, which means $s = d_1$ and $m, n < d_1$. The default length of d_1 is three months.
- The weight optimization process considers the performance of error score in the past period, the length of this time frame is d_2. This is to make sure that the weight is not only adjusted to one day but a whole period. The default length of d_2 is one month.
- The performance of the error score is how well the error score can predict the top errors in the next f days. Therefore, the evaluation is always based on f days. This is considered as one sliding window. The default length of f is five days.
- In each siding window, one shop floor meeting is simulated, error scores are calculated, and the top deviations are ranked. With the deviation cost of the next f days, the performance of the error score is also evaluated. The performance of each sliding window is marked as P_j. In the time frame d_2, the sliding window moves to the left and simulates all the shop floor meetings in the period as shown in Figure 6. The total number of sliding windows is $d_2 - f + 2$.

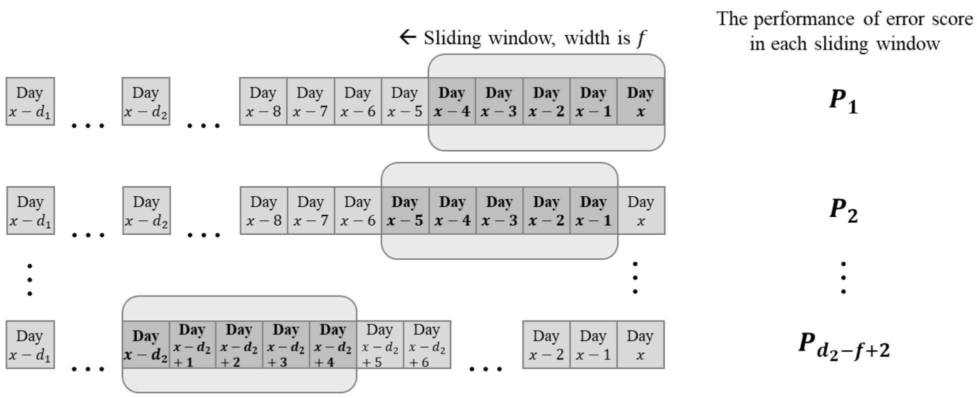

Figure 6 Weight optimization with sliding window

- The average performance P is then calculated based on the performance of all sliding windows. And the target of the algorithm is to obtain the right weights by iteration, which can achieve a better overall performance.

The average performance P is the optimization target of the algorithm. In order to achieve a higher P, the algorithm uses the method of gradient descent to find out the best weights with historical data. And the best weights are used for the calculation of error score in the next period of d_3 days.

4. Case study

In the automotive production, an effective problem-solving process is essential, as the right and immediate recognition of the deviation can prevent the severe deviation that may cause the huge consequence in the future [19]. In this work, the proposed algorithm is tested with the data from a production line of an automotive company. At the end of the production line, the workers check the quality of the products with both test equipment and eye check. The deviations are recorded in the dSFM system through the terminals at the quality gate. Products that don't fit the defined requirements are sent to the rework station. After finishing the rework, the workers record the measurements and time needed in the system. In the every-day shop floor meeting, the managers go through the rework records and identify the most serious deviations that may occur in the next days, so that the measurements can be taken as soon as possible, and the personnel plan can also be adjusted.

4.1 Dataset

The dataset contains three different sub datasets, namely a process dataset, a quality check dataset and the rework dataset. The process dataset is collected from the manufacturing execution system and contains the basic information about the product and the production process. The quality check dataset is collected from the terminal of the SFM system at the quality gate at the end of the production line, contains the results of the quality check. The rework dataset is collected from the terminals at the rework station. As shown in the section 3.1, the datasets are merged according to the product number. The failure description is in natural language. With the methods described in section 3.2, the abbreviations and misspellings are corrected and standardized, so that every failure type is identical and can be further used in the error score calculation. An excerpt of the final data, used for the error score calculation, is shown in the following table.

Table 1: Excerpt of final data for error score calculation

Product No.	Deviation Description	Rework Duration	Rework Time	Production Time
1000001	Steckplatz Drehmoment n.i.O.	60.0	09.06.2021 18:54	09.06.2021 10:12
1000003	Fehlendes Getriebe	60.0	09.06.2021 18:54	09.06.2021 13:31
1200057	DGM Einstellung	90.0	09.06.2021 18:54	09.06.2021 10:16
2257009	Beschädigte Schrauben	30.0	09.06.2021 18:54	09.06.2021 08:07

4.2 Methodology

To verify the model, the proposed method is compared with the Pareto analysis, which is often used in the production line to identify and prioritize the most serious deviations. The goal of both error score and the Pareto analysis is to generate a list of deviations, displaying them from highest to lowest severity. Therefore, the evaluation is done by comparing the overlap between the deviations recommended by the list and the deviations that occurred. For the five largest deviations that occur, if the percentage of costs resulting from deviations predicted by the error score is greater than the results of the pareto analysis, then the performance of error score is superior. Based on this understanding, the metrics are defined to evaluate the performance of the decision support system. They can be defined as follows:

$$PE(cost) = \sum_{i=1}^{5} r_i * C_i \tag{4.1}$$

$$PE(percentage) = \frac{\sum_{i=1}^{5} r_i * C_i}{\sum_{i=1}^{5} C_i} \tag{4.2}$$

$$C_i = \sum_{d=0}^{4} C_i^{x+d} \tag{4.3}$$

, where C_i^{x+d} is the cost of the deviation i on day $x + d$, $C_1, \ldots, C_5$ is the cost of the top deviations. r_i means whether the deviation is suggested by the decision support system.

As the test of the algorithm lasts for a period, which means that there are more shop floor meetings and decision-making processes involved. Therefore, to get the performance of the algorithm in a period, the average performance of the algorithm is calculated as follows:

$$PE_{AVG} = \frac{\sum_{j=1}^{M} PE_j}{M} \tag{4.4}$$

, where PE_j is the performance in one day, s is the number of days in the period when the same weight is used. The larger the PE_{AVG}, the better the performance of the decision support system over the period.

To better observe the changes in the rework process after the error score is applied, the Cox-Stuart test [21] is used to assess whether there is a downward trend in rework duration.

5. Results and discussion

The proposed method is tested with the data from 06.2020 to 08.2021. After merging the process data, deviation data and the rework documentation, this results in a total of 7771 records. In the data preprocessing process, all the text related to deviation are analyzed with the algorithms to correct the misspellings and abbreviations. The 62 abbreviations automatically found by the algorithm. With the help of the expert knowledge from the shop floor, the abbreviations are replaced in the data. A new dictionary with 49 new words is trained and imported to the system, so that the algorithm for misspelling can also recognize and correct the specific words in the rework documentation. After the preprocessing with NLP, the original 1316 failure categories are lowered to 1063. Through this process, 19.33% of the original records were fused with records that had the same meaning, enhancing the accuracy of the data.

For the evaluation of the error score, the shop floor meetings from January to July are simulated. As comparison, the performance of the Pareto analysis is also calculated. As shown in Table 2, the performance of error score is overall better than the Pareto analysis.

Table 2: Results of the evaluation

Simulated Date	Performance of error score		Performance of Pareto analysis	
	$PE(cost)_{AVG}$	$PE(percentage)_{AVG}$	$PE(cost)_{AVG}$	$PE(percentage)_{AVG}$
01.2021	893, 34 min	31.84%	830.75 min	29.28%
02.2021	1215.75 min	58.27%	950.75 min	45.52%
03.2021	1007.5 min	52.56%	651.0 min	34.17%
04.2021	942.0 min	47.71%	819.0 min	40.44%
05.2021	957.75 min	47.25%	826.75 min	41.47%
06.2021	643.25 min	28.44%	522.75 min	25.25%
07.2021	678.75 min	38.33%	117.5 min	9.60%
average	**905.48 min**	**43.49%**	**647.07 min**	**32.17%**

Besides the results of simulation, the rework duration after applying the error score is also observed. Figure 6 shows the rework duration since the official Go-Live in September 2021 to September 2022. The Cox-Stuart test identifies a significant downward trend with $p - valure = 0.0063$ (see Figure 7).

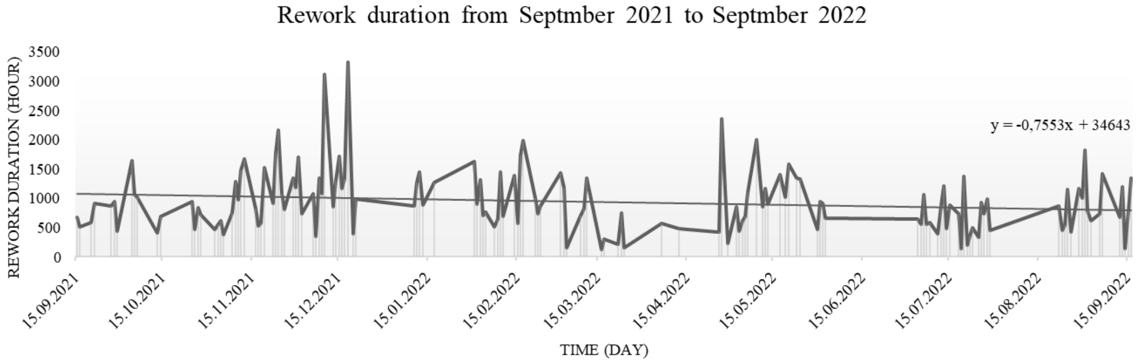

Figure 7: Rework duration from September 2021 to September 2022

This case study reveals that the proposed decision support model can be effectively and practically applied for the prioritization of the deviation in the problem-solving process. A comparative analysis with the pareto method also shows that the proposed method has a good predictive performance for the problems that may arise in the future. In addition, a trend of gradual reduction in the cost of rework time caused by errors in practice was also observed during the year when the method was applied to the actual production process. This also reflects the effectiveness of the method in practice.

6. Conclusion and future work

This paper establishes a comprehensive decision support model that utilizes process and deviation information to assist the deviation prioritization for the SPS on the shop floor. The proposed method is able to achieve better performance than the traditional Pareto analysis. It is also demonstrated by deploying the method in a real production environment. The results show that the deviation prioritization process can be improved to a certain degree with the help of this decision support system. This paper makes some attempts in dealing with heterogeneous data in production, combining structured and unstructured data for analysis. In the future, there are many more directions for the analysis of heterogeneous data, such as the graph databases that can be used to model the data involved in SFM, which can better provide the data basis for other decision support systems. In the direction of NLP, there is also a need to further analyze and study the textual data in SFM, so that more information for problem solving can be extracted.

Acknowledgements

This research is funded by the European Regional Development Fund, the Hessian State Chancellery – Hessian Minister of Digital Strategy and Development, and the European Institute of Innovation & Technology (EIT) Manufacturing.

References

[1] Künzel, H. (Ed.), 2016. Erfolgsfaktor Lean Management 2.0: Wettbewerbsfähige Verschlankung auf nachhaltige und kundenorientierte Weise. Springer Gabler, Berlin, Heidelberg, 447 pp.

[2] Xu, Z., Dang, Y., Munro, P., 2018. Knowledge-driven intelligent quality problem-solving system in the automotive industry. Advanced Engineering Informatics 38 (2–3), 441–457.

[3] Zaidat, A., Boucher, X., Vincent, L., 2005. A framework for organization network engineering and integration. Robotics and Computer-Integrated Manufacturing 21 (3), 259–271.

[4] Khan, M., Wu, X., Xu, X., Dou, W., 2017. Big data challenges and opportunities in the hype of Industry 4.0. 2017 IEEE International Conference on Communications (ICC), 1–6.

[5] Longard, L., Meißner, A., Müller, M., Metternich, J., 2020. Digitales Shopfloor Management – Wohin geht die Reise? Zeitschrift für Wirtschaftlichen Fabrikbetrieb: ZWF 115 (9), 645–649.

[6] Müller, M., Longard, L., Metternich, J., 2022. Comparison of preprocessing approaches for text data in digital shop floor management systems. Procedia CIRP 107 (7-8), 179–184.

[7] Pötters, P., Schindler, P., Leyendecker, B., 2018. Status quo Shopfloor Management. Zeitschrift für wirtschaftlichen Fabrikbetrieb 113 (7-8), 522–525.

[8] Hertle, C., Tisch, M., Metternich, J., Abele, E., 2017. Das Darmstädter Shopfloor Management-Modell. Zeitschrift für Wirtschaftlichen Fabrikbetrieb: ZWF, Carl Hanser Verlag 112 (3), 118–120.

[9] Meißner, A., Hertle, C., Metternich, J., 2018. Digitales Shopfloor Management – Ihr Weg zur vernetzten Fabrik. Zeitschrift für Wirtschaftlichen Fabrikbetrieb : ZWF, Carl Hanser Verlag 113, 281–284.

[10] Furlan, A., Galeazzo, A., Paggiaro, A., 2019. Organizational and Perceived Learning in the Workplace: A Multilevel Perspective on Employees' Problem Solving. Organization Science 30 (2), 280–297.

[11] Choo, A.S., Nag, R., Xia, Y., 2015. The role of executive problem solving in knowledge accumulation and manufacturing improvements. Journal of Operations Management 36 (1), 63–74.

[12] Kumar, S., Sosnoski, M., 2009. Using DMAIC Six Sigma to systematically improve shopfloor production quality and costs. International Journal of Productivity and Performance Management 58 (3), 254–273.

[13] Müller, M., Alexandi, E., Metternich, J., 2021. Digital Shop Floor Management Enhanced by Natural Language Processing. Procedia CIRP 96, 21–26.

[14] Hapke, H., Howard, C., Lane, H., 2019. Natural Language Processing in Action: Understanding, analyzing, and generating text with Python MEAP V10 - Chapter 1, 40 pp.

[15] Rahm, E., Hong, H.D., 2000. Data cleaning: Problems and current approaches. IEEE Data Eng. Bull. (23), 3–13.

[16] Müller, M., Metternich, J., 2021. Production Specific Language Characteristics to Improve NLP Applications on the Shop Floor. Procedia CIRP 104, 1890–1895.

[17] Jurafsky, D., Martin, J.H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition 2020.

[18] Bird, S., Klein, E., Loper, E., 2009. Natural language processing with Python: analyzing text with the natural language toolkit, 504 pp.

[19] Longard, L., Brungs, F., Hertle, C., Roeth, J., Metternich, J., 2021. Reduced Rework Through Data Analytics and Machine Learning – A Three Level Development Approach. SSRN Journal.

[20] Longard, L., Bardy, S., Metternich, J., 2022. Towards a Data-driven Performance Management in Digital Shop Floor Management. Hannover: publish-Ing, 11 pp.

[21] Cox, D.R., Stuart, A., 1955. Some Quick Sign Tests for Trend in Location and Dispersion. Biometrika (42(1/2)), 80–95.

Biography

Yuxi Wang, M. Sc. (*1995) is research associate and PhD student at the Institute of Production Management, Technology and Machine Tools (PTW) at the Technical University of Darmstadt since 2021. Her research interests include the further development and introduction of natural language processing to digital shop floor management and problem-solving systems.

Lukas Longard, M. Sc. (*1992) is chief engineer and PhD student at the Institute of Production Management, Technology and Machine Tools (PTW) at the Technical University of Darmstadt since 2019. His research interests include the further development and introduction of machine learning to digital shop floor management and performance management systems.

Dr.-Ing. Christian Hertle (*1987) is CEO Business Development of SFM Systems, a software provider for digital shop floor management software. He did his PhD at Technical University Darmstadt on the topic of shop floor management and how it can be used for continuous competency development in manufacturing.

Prof. Dr.-Ing. Joachim Metternich (*1968) is head of the Production Management division at the Institute of Production Management, Technology and Machine Tools (PTW) at the Technical University of Darmstadt. He is the spokesman of the medium-sized businesses digital competence center Darmstadt; he has been president of the International Association of Learning Factories (IALF) from 2015 to 2021 and he is scientific partner of the Conference on Production Systems and Logistics (CPSL).