iHPPVis: Interactive Visual Analysis of Industrial data in Heavy Plate Production

Tongkang Zhang, Yufei Ma, Depeng Xu, Changxin Liu, Jinliang Ding

State Key Laboratory of Synthetical Automation for Process Industries,
Northeastern University, Shenyang, 110004, China (e-mail:
tkzhang@stumail.neu.edu.cn, yufeima123@163.com,
neuxudepeng@163.com, {jlding, cxliu}@mail.neu.edu.cn).

Abstract: This paper develops an interactive visualization system, called iHPPVis, to analyze and locate the cause of quality-related faults for the heavy plates production. A time distribution of the products under different operating conditions based on Marey's graph is presented and the process data for corresponding conditions to identify clusters and outliers is visualized, which utilizes the alternative dimension reduction algorithms. The crucial stage that leads to the abnormality of product quality is also diagnosed, and the data distribution of heterogeneous process variables in the crucial stage is exhibited. By integrating alternative algorithms with interactive visual analysis to achieve quality-related fault diagnosis, iHPPVis can facilitate the improvement of heavy plates quality. A case study is conducted to demonstrate its effectiveness and exhibit a pilot application of visual analytics for the heavy plates production.

Keywords: Industrial data; Heavy plates production; Visual analytics; Quality-related fault.

1. INTRODUCTION

As a typical complex industrial process, the hot rolling production of heavy plates characterized by multi-stage, higher-dimensional process variables and huge volumes of data. It generally consists of several main stages (Fig. 1), i.e., heating, rolling, cooling, and hot levering (Tang and Wang, 2011). In the hot rolling process of heavy plates, the production indices of each stage are regarded as a significant foundation to evaluate the operating state. It is important to explore the internal relationship between plant-wide production process variables and product quality. Valuable decision support information can be provided to the field operator. In addition, the visual analytics can explore the valuable information of industrial big data, and realize the intellectualization of hot rolling production.

Data visualization has been continuously improved with the development of information technology in decades. Through the application of visualization technologies, more reliable decisions are provided with diversified fields, such as business (Yue et al., 2020), sports (Wu et al., 2018b), transportation (Chen et al., 2015), etc. Although many commercial companies in the industrial field have begun to engage in related research, including Predix platform from GE Digital, MindSphere platform from Siemens, these researches have not been able to carry out in-depth exploration for visual analysis. In academic research, there are only a few studies analyzing industrial data (Xu et al.,

2017; Wu et al., 2018a). For example, Xu et al. (2017) proposed ViDX, a visual analytics system, to identify patterns of production and diagnose problems in automobile assembly lines. Wu et al. (2018a) developed a visualization system to support on-site operators in monitoring equipment condition.

It should be noted that designing and developing such a visual analytic system with huge volumes of heavy plates production data presents two major challenges. First, given the scale of the industrial big data in Supervisory Control And Data Acquisition System (SCADAS) and Manufacturing Execution System (MES), there is a need to explicitly show the operating conditions of the plant-wide production process and to optimize the data analysis performance. Second, it is difficult to intuitively represent the complicated relationship between production process variables and quality indices of heavy plates.

To address the aforementioned challenges, we propose a highly interactive visualization system called iHPPVis, which aims at identifying patterns of operating conditions and locating the root of quality-related faults. The major contributions of this paper are as follows:

- An interactive visualization system is developed, which assists users in analyzing the big data of heavy plates production.

- A visual analytics method is proposed, which enhances with new features to support users in exploring the root cause of the abnormality on product quality.

- A case study with domain experts is conducted to demonstrates the usefulness and effectiveness of iHPPVis .
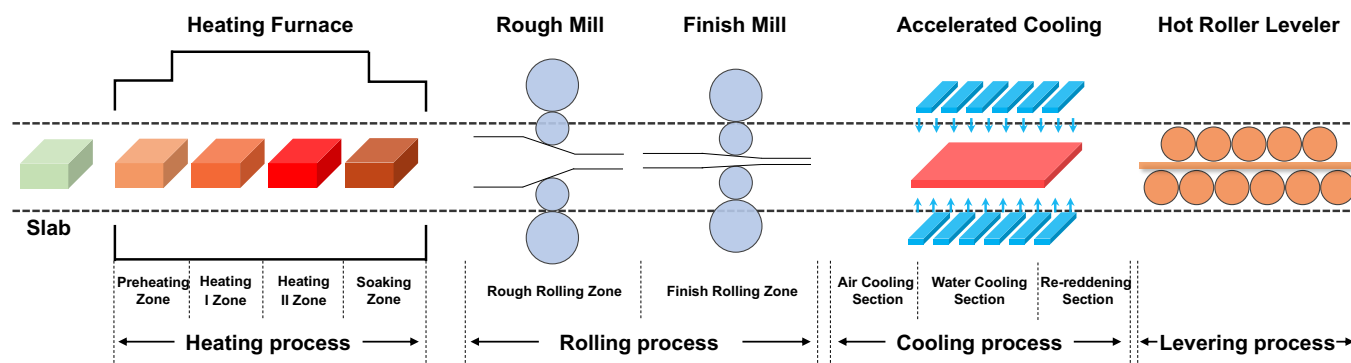
Fig. 1. Heavy plate production process.

2. REQUIREMENT ANALYSIS

In this section, we summarized the most critical requirements, following multiple in-depth discussions with four domain experts, including two experienced on-site operators, one manager of the heavy plate production line, and one scientist who had been working on the heavy plate production for more than ten years. Then we developed an early prototype and then improved the designs iteratively according to the feedback from these domain experts. Finally, we accomplished the current version of the visualization system, and illustrated four requirements which can be grouped into two levels as follow.

This manager proposes two **overall-production-level** requirements to supply an overview of the heavy plate production.

R1: Visualize the root cause of the abnormality on plates quality. The system should not only highlight the products of abnormal quality and show when and why quality-related faults have occurred, but also provide a comprehensive overview for on-site operators.

R2: Display the overview of industrial big data in the plant-wide production process. The visual encoding should present various operating conditions of heavy plate production, furthermore, exhibit the distribution of high dimensional process variables, which provides guidance for further analysis.

Two on-site operators provide three **detailed-diagnosis-level** requirements to focus on a detailed analysis that leads to the abnormality on plates quality.

R3: Present different hierarchies of diagnostic analysis. To support a comprehensive exploration of plant-wide production process, the system should be able to provide different hierarchies of production data.

R4: Integrate feature extraction and dimension reduction algorithms. The system should support interactions to extract the features of industrial data and reduce the dimensions of production indices.

R5: Support real-time and interactive analysis of the large volume of industrial data. The assembly line of heavy plates production sets up hundreds of measuring sensors. It is essential to provide real-time and interactive analysis in the system.

3. SYSTEM ARCHITECTURE

Motivated by the above design requirements, we design a visual analytics system, to allow users to explore the reason of the abnormalities on product quality. iHPPVis is a web-based application with three major parts, a data storage module, a data processing module and a visualization module as shown in Fig. 2. The data storage module (Fig. 2a) is based on MongoDB (Chodorow, 2013) and Hadoop (Shvachko et al., 2010). Furthermore, this module collects massive information from plant-wide production process heterogeneous sensors, such as process data in SCADAS and MES, and product quality data in final quality control system. The data processing module (Fig. 2b) is based entirely on Python3, which cleans and filters the collected data. Then the mathematical model is established for statistical analysis and visual analysis of cleaned data. This Python3-based data analysis module forms the backend and it guarantees the efficiency of our visualization module. Data visualization module (Fig. 2c) provides users with visual analysis and interactive services. It is a front-end interface of the system based on D3.js and Vue.js that connects users. In order to comprehensively analyze the system workflow, we will describe each part of the system architecture in detail.
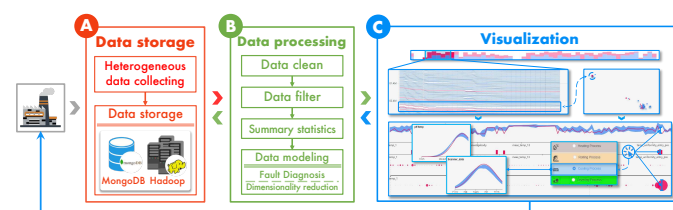

Fig. 2. System architecture. iHPPVis consists of three major parts: a) Data storage module; b) Data processing module; c) Visualization module.

3.1 Data Storage & Data Processing

The raw industrial data is considerably large in size, which is collected from more than 1,200 sensors in the production line. And its overall scale is more than 20tb with about one hundred thousand of products ranging from June 2018 to October 2019. After collecting the raw information from the hot rolling production of heavy plates, further data storage and data processing should be indispensable.

Fig. 3. iHPPVis facilitates the exploration of heavy plates production process.

We first utilize Apache Hadoop2 on two clusters with 35 data nodes and 960 cores to collect and parallel process industrial big data based on Map-Reduce. It transforms data collected from multi-source heterogeneous data sources into uniform format (a set of vectors or matrices). The uniform format is helpful for further analysis and exploration. Furthermore, the processed results are stored in MongoDB3 database, which is deployed in the cloud and supports real-time query, with the size of more than 5TB.

The next step is to process two types of anomalous data which are missing data and infinite values. It's significant to clean and filter anomalous data, which is derived from industrial sensors anomaly.After discussion with domain experts, we determine to replace anomalous data with the appropriate values calculated by interpolation algorithm.

At the last but not least, we reveal that industrial big data still has several issues such as high dimensions and multiple redundancy, which is not conducive to subsequent visual analysis. We adopt Maximal Information Coefficient (MIC) (Reshef et al., 2011) to extract feature variables. MIC algorithm is widely used in various applications for its versatility and efficiency (**R4**, **R5**). After the industrial process data $X_{raw} = \{x_1, x_2, \ldots, x_n\}$ and quality data $Y = \{y_1, y_2, \ldots, y_m\}$ are filtered and cleaned, we calculate MIC between X_{raw} and Y as follows:

$$MIC\,[x_i; y_j] = \max_{|X_{raw}||Y|<B} \frac{I[x_i; y_j]}{\log_2\left(\min\left(|X_{raw}|, |Y|\right)\right)} \quad (1)$$

where $I[x_i; y_j]$ represents the mutual information between the i-th process variable x_i and the j-th quality index y_i

for $i \in 1, 2, ..., n, j \in 1, 2, ..., m$. Thus, we effectively extract industrial process data and proceed to visual analysis.

3.2 Visualization

iHPPVis is developed to support multi-level exploration and comparison of different conditions, which consists of five parts: a control panel (Fig. 3A) to present the configuration data, select the automatic algorithm for subsequent analysis and promote interactively changing the configuration (**R4**); a condition overview (Fig. 3B) to exhibit the time distribution of the products under different operating conditions, and troubleshoot inefficiencies of production capacity on the plant-wide production process (**R2**); an embedding view (Fig. 3C) reveals process data for corresponding operating condition to identify clusters and outliers (**R2**, **R3**); a diagnosis view (Fig. 3D) to diagnose the crucial stage that leads to the abnormality of product quality (**R1**, **R3**); a key stage view(Fig. 3E) to exhibit the data distribution of heterogeneous process variables in the key stage, further analysing the detailed causes of problems with the quality of products (**R3**). It combines automatic algorithms and domain knowledge to enable three kinds of analysis: the monitoring of operating conditions, analysis of the relationship between process variables and product quality, and root cause of quality-related faults, thereby promoting product quality enhancement and guiding production operations. The proposed visualization system aims at five requirements which could be divided into two levels. In this section, we first describe each component of the iHPPVis. Then, we illustrate the

user interactions provided to enable the exploration and analysis of heavy plate production.

4. VISUAL DESIGN

4.1 Control Panel

The control panel (Fig. 3A) aims at supporting visual operation on the system configuration of product parameters and automatic algorithm types to diagnose product quality (**R4**). It shows several types of system configuration data: the detailed parameters of each product, the key stage variables recommended based on number of user clicks, and the automatic algorithms for subsequent analysis including the dimension reduction algorithm in embedding view (Fig. 3C), the multivariate statistical analysis algorithm in diagnosis view (Fig. 3D). Moreover, the control panel supports two types of visual encoding, namely quality-based and category-based. Due to the interactions provided by the control panel, on-site operators could utilize their domain knowledge, including the in-depth understanding of production technology and extensive on-site operational experience, to explore the cause of abnormality on the product quality (**R1**).

After modifying the configuration data, the timeline (Fig. 3F) will then update its range to the selected days and display the number of products of different quality in a finer resolution. By brushing the corresponding range on the timeline, users can click the search button to invoke the product quality analysis algorithm, and the returned result will be added to the condition overview and the embedding view.

4.2 Condition Overview

The condition overview (Fig. 3B) exhibits the time distribution of the products under different operating conditions (**R2**). The essential target of this view is to analyze the impact of multi-operating condition conversion for product quality. The condition overview is inspired by Marey's graph (Tufte, 1986) and VIDX (Xu et al., 2017) system. Marey's graph is a traditional method for describing bus or train schedules. It adopts a parallel layout of time axes. Each time axis is used to represent a bus or train stop. The polylines between time points on the axes indicates when the vehicles arrived at the stop based on the schedules. This visual encoding can be directly applied to plant-wide production process data if we represent each sub-stage station on the whole production line as a bus or train stop, and the production process of a product as a polyline. The angle of the line segments between the axes implies its time on each sub-stage. In addition, the condition overview supports simultaneous highlighting with the embedding view and provides interaction for subsequent analysis (**R5**).

By observing the condition overview experiment (Fig. 4), it illustrates the different types of visual patterns:

a. Visual pattern a (Fig. 4a) implies that the no abnormal operating condition has occurred and the production line works without abnormality.

b. Visual pattern b (Fig. 4b) illustrates that the production line is affected by part of stages, which leads to a slight decrease in production capacity.

c. Visual pattern (Fig. 4c) indicates that the production line has been completely suspended for a short period of time, which may be due to the change of the type of products.
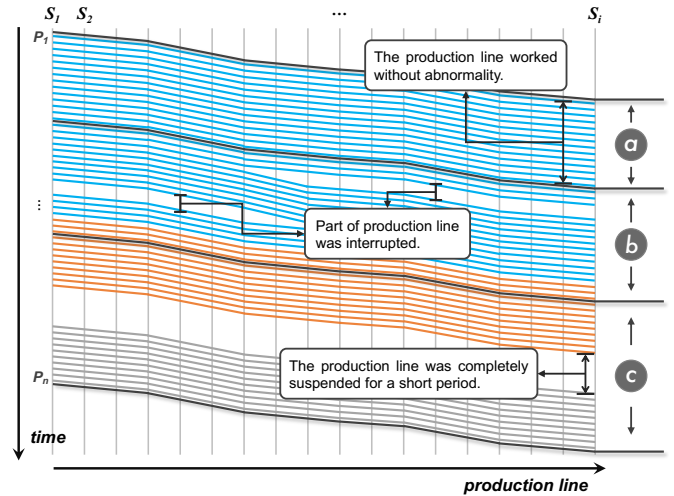

Fig. 4. Visual patterns of condition overview.

4.3 Embedding View

Abnormalities are not only closely related to production conditions, but also to high-dimensional process variables. We explore the underlying relationship by embedding view (Fig. 3C), which projects high-dimensional process data X onto a 2D space, and their relative similarities are reflected through their placements to help users discover clusters and outliers (**R2**). Many dimension reduction algorithms, such as principal components analysis (Wold et al., 1987) and multidimensional scaling (Borg and Groenen, 2003), could be used for this target. In embedding view, we adopt the t-distributed stochastic neighbour embedding (t-SNE) algorithm because t-SNE repels dissimilar points strongly to form more obvious clusters (Maaten and Hinton, 2008) (**R3**). Therefore, the embedding view is visualized so that similar products are placed nearby while dissimilar products are placed faraway.

$$C = tSNE(X) \tag{2}$$

where C is the embedded data. Following that, we could distinctly observe the abnormal batches in the outlier area. In embedding view, we provide other dimension reduction algorithms, which can be selected in the control panel. Furthermore, the view not only supports two types of user interactions including panning and zooming, but also provides brushing the outlier area to interact with the diagnosis view in real time (**R5**).

4.4 Diagnosis View

The diagnosis view (Fig. 5) is designed to diagnose the crucial stages that lead to the abnormality of product quality (**R1**, **R3**). First of all, normalize process variables for the standard products and the product selected in the embedding view. Then calculate the upper and lower limits of the process variables for standard product. Finally, show them in the area chart which is the top part of diagnosis
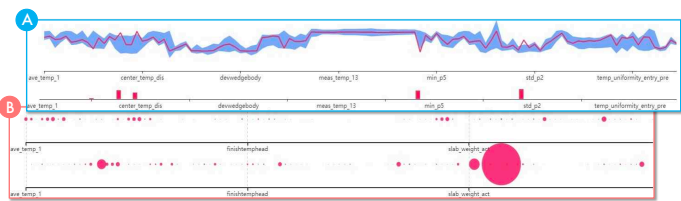
Fig. 5. Diagnosis view.

view (Fig. 5a). Because of the noise from industrial data, this view utilize exponentially weighted moving average (EWMA) (Izadi et al., 2009) to reduce noise and extract features (**R4**). An EWMA filter is defined as:

$$z(k) = x(k) + (1 - \alpha)z(k - n) \qquad (3)$$

where $x(k)$ is the sequence of original raw process data, $z(k)$ is the sequence of filtered data, α is filter parameter and n is the order of filter. However, exclusively considering unidimensional anomaly detection is not comprehensive enough. We adopt multivariate statistical analysis to troubleshoot quality-related faults. The variable contributions are displayed in the bottom part of diagnosis view (Fig. 5b). Users are provided with conjoint analysis through the aforementioned two methods. Besides, iHPPVis supports alternative algorithms based on multivariate statistical analysis, e.g., total projection to latent structures (T-PLS) (Gang et al., 2009), concurrent canonical correlation analysis (Zhu et al., 2016) (**R5**).

4.5 Key Stage View

The key stage (Fig. 3E) exhibits the data distribution of heterogeneous process variables for key stage, further analysing the root cause of abnormality on the product quality (**R1**, **R3**). In this view, each variable of selected key stage is sorted in descending order of deviation. It's effective to assist users in taking more accurate actions in production, such as calibrating a specific sensor, prioritizing maintenance of a specific piece of equipment, etc. Moreover, every view of data distribution supports two types of user interactions such as drag and drop.

5. CASE STUDY

Affected by space limitation, only one case study is given here. The ultimate goal of iHPPVis is to provide domain experts with an efficient and effective tool to capture the reason for the fluctuation of the plate shape. Based on the purpose of this study case, We applied a sampled dataset from the real production of large-scaled steel factory in eastern China. It contains the 30-day data of 9589 plates. In addition, we had trained four aforementioned domain experts in basic knowledge and use skills before this experiment. The results are shown in the (Fig. 6).

In the scenario, the domain experts would like to explore and analyze the shape quality of heavy plates and to investigate the case of abnormality on product quality. First of all, the experts chose to look over the capacity of heavy plates throughout the month by timeline (Fig. 6A), and they focused on the production data of heavy plates from September 3 to September 6. Because the number of abnormal quality plates on this day is 762, but the

plates of normal quality account for less than half of the total, which is 527 (**R2**). Under our suggestion, the experts selected the dimension reduction algorithm t-SNE and the diagnostic algorithm T-PLS in the control panel and set the system parameters (**R4**, **R5**). Then they viewed the operating conditions of production through the condition overview (Fig. 6B). After examination, experts discovered that the types of plates produced would change constantly, and consider that the reason is consistent with their expertise (**R2**). Afterwards, they explored the embedding view (Fig. 6C) for identifying the abnormal batches in the outlier area. Combining interaction with the condition overview, experts found that the data distribution of abnormal plates is very similar (**R3**). They wanted to further explored how each stage affects quality of these plates by the diagnosis view (Fig. 6D), then domain experts discovered that the cooling stage leads to the wave shape on edge of plates (**R1**, **R3**). Therefore they excluded the other unrelated stages on control panel. After observing the key stage view (Fig. 6E), the excessive flux of cooling water leaded to the lower temperature of plates(**R3**). By querying the historical maintenance records, maintainers found that the 15th flow sensor failed to detect at 3 pm on September 5. The domain experts consider that the analysis of iHPPVis was not only consistent with the results of on-site maintenance, but also practical and efficient.

Through the aforementioned case study, we demonstrate the effectiveness and usability of iHPPVis, which enables situation-aware exploration and analysis of heavy plates production. Overall, our cooperative experts are satisfied with the strong analytical capability of our system. Especially the visualization is helpful to on-site operators (**R5**). A whole picture of the heavy plates production is presented in conjunction with the condition overview and the embedding view (**R1**). Smooth interactions with prompt visual feedback allow users to explore from overall-production-level situation to detailed diagnosis analysis (**R2**). The diagnosis view accurately locates the crucial stage that leads to the abnormality of product quality. The root cause is further displayed in the key stage view (**R1**, **R3**). When compared with standard data analysis software in the process industry, iHPPVis has already processed raw industrial big data and provided a comprehensive analysis of the whole heavy plates production process and specific product quality with well-coordinated views. To the best of our knowledge, iHPPVis is the first visualization system analyzing the hot rolling production of heavy plates, which closely follows domain requirements. The experts with basic knowledge can effectively analyze complicated industrial data in heavy plate production.

However, our research is still in progress. We will some improvements for the foregoing shortcomings. Firstly, the visual representation of the diagnosis view is not explicit enough. It is necessary to integrate multiple analysis of quality-related fault diagnosis in one view. Secondly, the extensive experience accumulated from on-site operators also needs to be further analyzed. Mining knowledge from their operation can effectively improve the quality of heavy plates. Finally, as the more industrial sensors are installed in the heavy plates production, our system will be confronted with the issue of computation efficiency when more real-time massive data need to be analyzed.
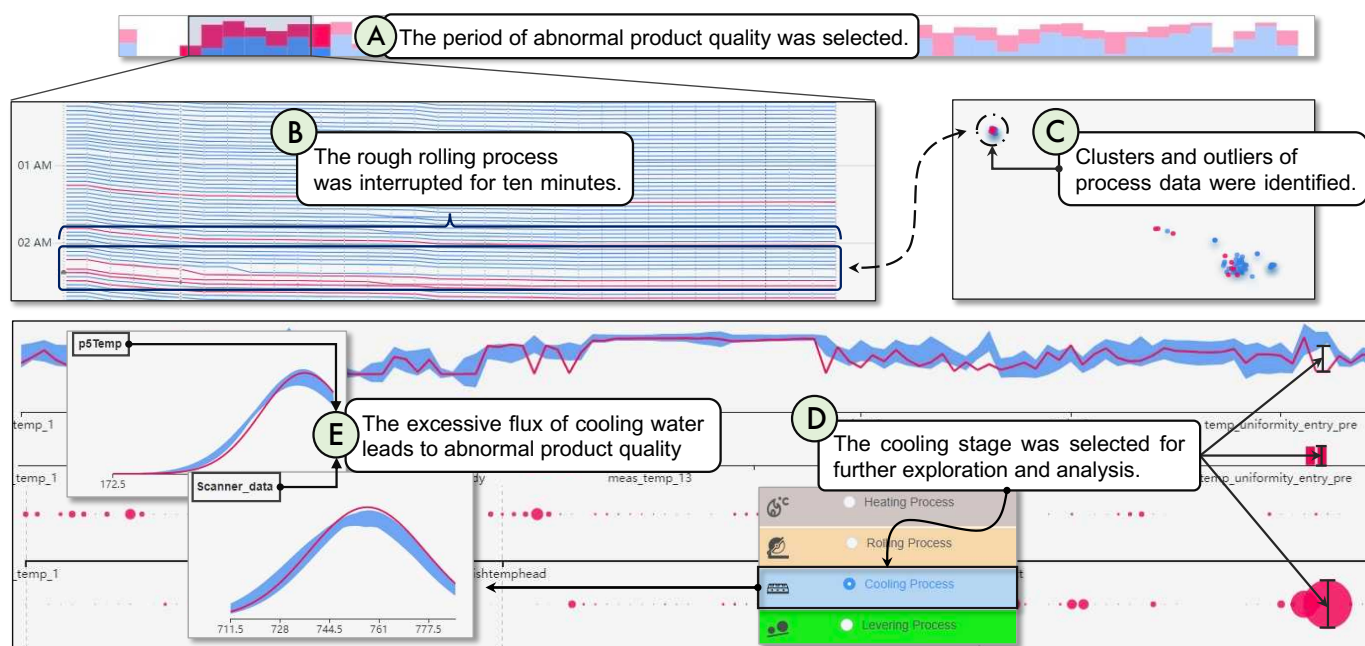
Fig. 6. Results of domain experts experiment.

6. CONCLUSION

In this paper, we develop a highly interactive visualization system called iHPPVis, which assists on-site operators in analyzing industrial big data: a condition overview presenting the time distribution of the products under different operating conditions; a embedding view visualizing the process data for corresponding conditions to identify clusters and outliers; a diagnosis view identifying the crucial stage that leads to the abnormality of product quality; a key stage view exhibiting the data distribution of heterogeneous process variables in the crucial stage. By integrating multiple algorithms with interactive visual analysis, iHPPVis can facilitate the improvement of product quality in production. A case study demonstrates its effectiveness and exhibits a pilot application of visual analytics for industrial data of heavy plates production.

REFERENCES

Borg, I. and Groenen, P. (2003). Modern multidimensional scaling: Theory and applications. *Journal of Educational Measurement*, 40(3), 277–280.

Chen, W., Guo, F., and Wang, F. (2015). A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems*, 16(6), 2970–2984.

Chodorow, K. (2013). *MongoDB: The Definitive Guide.* O'Reilly Media, Inc.

Gang, L., Si-Zhao, Q., Yin-Dong, J., and Dong-Hua, Z. (2009). Total pls based contribution plots for fault diagnosis. *Acta Automatica Sinica*, 35(6), 759–765.

Izadi, I., Shah, S.L., Shook, D.S., and Chen, T. (2009). An introduction to alarm analysis and design. *IFAC Proceedings Volumes*, 42(8), 645–650.

Maaten, L.v.d. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov), 2579–2605.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M., and Sabeti, P.C. (2011). Detecting novel associations in large data sets. *Science*, 334(6062), 1518–1524.

Shvachko, K., Kuang, H., Radia, S., and Chansler, R. (2010). The hadoop distributed file system. In *Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, MSST '10, 1–10. IEEE Computer Society, Washington, DC, USA.

Tang, L. and Wang, X. (2011). A scatter search algorithm for a multistage production scheduling problem with blocking and semi-continuous batching machine. *IEEE Transactions on Control Systems Technology*, 19(5), 976–989.

Tufte, E.R. (1986). *The Visual Display of Quantitative Information.* Graphics Press, Cheshire, CT, USA.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3), 37–52.

Wu, W., Zheng, Y., Chen, K., Wang, X., and Cao, N. (2018a). A visual analytics approach for equipment condition monitoring in smart factories of process industry. In *2018 IEEE Pacific Visualization Symposium (PacificVis)*, 140–149.

Wu, Y., Lan, J., Shu, X., Ji, C., Zhao, K., Wang, J., and Zhang, H. (2018b). ittvis: Interactive visualization of table tennis data. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 709–718.

Xu, P., Mei, H., Ren, L., and Chen, W. (2017). Vidx: Visual diagnostics of assembly line performance in smart factories. *IEEE Transactions on Visualization and Computer Graphics*, 23(1), 291–300.

Yue, X., Bai, J., Liu, Q., Tang, Y., Puri, A., Li, K., and Qu, H. (2020). sportfolio: Stratified visual analysis of stock portfolios. *IEEE Transactions on Visualization and Computer Graphics*, 26, 601–610.

Zhu, Q., Liu, Q., and Qin, S.J. (2016). Concurrent canonical correlation analysis modeling for quality-relevant monitoring. *IFAC-PapersOnLine*, 49(7), 1044–1049.