Accepted Manuscript

Data and Knowledge Mining with Big Data towards Smart Production

Ying Cheng , Ken Chen , Hemeng Sun , Yongping Zhang ,
Fei Tao

PII: S2452-414X(17)30058-4
DOI: 10.1016/j.jii.2017.08.001
Reference: JII 45

To appear in: *Journal of Industrial Information Integration*

Received date: 18 August 2017
Accepted date: 24 August 2017

Please cite this article as: Ying Cheng , Ken Chen , Hemeng Sun , Yongping Zhang , Fei Tao , Data and Knowledge Mining with Big Data towards Smart Production, *Journal of Industrial Information Integration* (2017), doi: 10.1016/j.jii.2017.08.001

Data and Knowledge Mining with Big Data towards Smart Production

Ying Cheng, Ken Chen, Hemeng Sun, Yongping Zhang, Fei Tao*

School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, P. R. China

(Corresponding Author: Fei Tao, E-Mail: ftao@buaa.edu.cn)

Acknowledgement

This work is financially supported in part by Beijing Nova Program (Z161100004916063) in China, National Natural Science Foundation of China (NSFC) under Grant 51522501, and National Key Research and Development Program of China (No.2016YFB1101700).

Abstract: Driven by the innovative improvement of information and communication technologies (ICTs) and their applications into manufacturing industry, the big data era in manufacturing is correspondingly arising, and the developing data mining techniques (DMTs) pave the way for pursuing the aims of smart production with the real-time, dynamic, self-adaptive and precise control. However, lots of factors in the ever-changing environment of manufacturing industry, such as, various of complex production processes, larger scale and uncertainties, more complicated constrains, coupling of operational performance, and so on, make production management face with more and more big challenges. The dynamic inflow of a large number of raw data which is collected from the physical manufacturing sites or generated in various related information systems, caused the heavy information overload problems. Indeed, most of traditional DMTs are not yet sufficient to process such big data for smart production management. Therefore, this paper reviews the development of DMTs in the big data era, and makes discussion on the applications of DMTs in production management, by selecting and analyzing the relevant papers since 2010. In the meantime, we point out limitations and put forward some suggestions about the smartness and further applications of DMTs used in production management.

Keywords: Big data, data mining techniques (DMTs), production management, smart manufacturing, statistical analysis, knowledge discovery

1. Introduction

We are in a big data era with explosively growing data. In May 2011, McKinsey Global Institute firstly published a research report, named *Big Data: The next frontier for innovation, competition, and productivity*. This report notes that big data has penetrated into all respects of life, and gradually becomes an important factor in production. The application of massive data indicates development of productivity and surplus of consumers[1]. The arrival of the big data era inevitably brings up challenges for the ability of data controlling. To exploit value of data, to escape from the plight of 'data grave', to seek for second and third or even more utilization by massive data, and to form a sustainable competitive advantage, are not only the challenges that all manufacturing enterprises have to face with, but also the opportunities they obtain. Knowledge discovery in database was defined in 1989, as a nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns from a data set[2]. Data mining is a key step of knowledge discovery, which is based on the technology of database and usually refers to the process of searching for the hidden information in a large amount of data. The big data era makes it possible to find valuable information and to make the best decision more easily with some advanced data mining techniques (DMTs).

Manufacturing industry is impacted enormously due to the arrival of the big data era. Firstly, with the development of some typical information and communication technologies (ICTs), the mode of production has changed greatly. The actual production or manufacturing process is becoming more and more complex with large-scale tasks, complex constraints, coupled operating performance and uncertain environment. Therefore, quite a few of techniques and methods used in the traditional mode of production management are no longer applicable. They are extremely time-consuming and rigorous to rely on the engineers' senses and experience to tackle with the problems that cannot be solved by the traditional methods. Secondly, the demand of the real-time, dynamic, self-adaptive and precise production management has brought new challenges to the traditional methods. Production managers need to give accurate prediction of product quality, production and processing time by new effective techniques, within shorter computation time to control the continuous real-time production systems and to identify faults, defects and some other abnormal situations. With the completion of various kinds of information systems deployed in manufacturing enterprises, abundant data related to production management is accumulated, by using enterprise resource planning (ERP), manufacturing execution system (MES), computer-aided process planning (CAPP), and other information systems. Naturally, it is desirable to extract and mine some useful information or even knowledge from relevant online and offline data to improve the effectiveness and efficiency of production management decision-making[3].

After some internet of things (IoT) related technologies applied into manufacturing, it is possible that enough structured and unstructured data covering all of production-related elements/flows/businesses could be collected. Then, it also promotes the development and applications of DMTs into manufacturing industry. In the big data era, DMTs have made considerable progress. In the actual production processes, some industrial automation systems, control systems, and other enterprise information systems (EISs), all involve and gather a large number of images and information, which present the data characteristics of large scale, multi-source, diverse structure, strong dynamicity, low value density, and so on. Specially, as for the overall production management, e.g., production planning and scheduling, total quality management, process monitoring, fault detection, etc., all refer to the gathered complex information and data. In detail, the collected or gathered data during production process includes online and offline data. The online data includes the status of machines, work-in-processes (WIP), environmental factors, and so on. The offline data includes the basic information, function, structure or assembly information of products, and process parameters such as maintenance information, line or equipment layout, production plan, and so on. As a result, the DMTs could be utilized into production management based on the aforementioned data, in order to (a) find problems and unknown mutation operations, (b) mine useful and efficient patterns or rules, (c) adjust the production plan timely, (d) improve the intelligence and automation of production management, and (e) improve production efficiency and product quality step by step. All in all, that efficient, accurate and adaptive production management can be achieved through DMTs.

Therefore, in view of the inevitability and necessity mentioned above, we select related papers referring to DMTs and the typical applications in production management since 2010, and carry on detailed investigation and analysis. In this paper, the existing applications are mainly reflected in production planning and scheduling, quality improvement, fault diagnosis, defect analysis and others. The main contributions of this paper are highlighted as follows:

(1) Both the development of DMTs and the problems of production management solved using DMTs are summarized.

(2) Relevant discussion and limitation investigation on the existing applications of DMTs in production management are brought out.

(3) Directions and suggestions about the future DMTs-based smart production management are pointed out.

The rest of this paper is organized as follows. Section 2 simply introduces the functions and classifications of DMTs. Section 3 reviews the existing applications of DMTs in production management. Section 4 makes conclusion about the current applications and limitations of DMTs in production management, and makes expectation about the future directions and further development of DMTs utilized in production management. Finally, Section 5 concludes the full paper.

2. Data mining techniques (DMTs)

2.1 Brief overview of the development processes of DMTs

The word *data mining* appeared on the base of both the emergence of ultra-large-scale databases and the development of advanced ICTs. The DMTs is the techniques used in data mining process to search for the hidden information in a large amount of data. The main progresses of DMTs can be summarized as shown in Fig. 1. The development of data mining is mainly influenced by database technology, machine learning, statistics and artificial intelligence. Some theories and methods of DMTs are developed and extended based on the statistical theory. Artificial intelligence is used to generate the process of human thinking, which enables computer the ability of learning without precise programming and facilitates new techniques used in the data mining processes. While database technology provides the basis of data storage, organization and other functions for data mining and this is the basis of data mining. In the 1960s, people began to collect and store data on computers, tapes, and CDs. By the late 1960s, the database and information technologies evolved from the original file processing system to a sophisticated and powerful database system. In 1970, the researcher E. F. Codd in San Jose lab of United States International Business Machines (IBM) Corporation first proposed a database system relational model. The invention and use of relational database which is a database system to support the relational model, enables people to collect, store, and process large amounts of data. At the same time, with the application of computer network, people gradually used online transaction processing (OLTP) on the information data for timely and efficient storage and processing. In the 1980s, it came into the golden age of relational database development. Some new data models were proposed and the corresponding database technology was also continually developing. The application-oriented database management system evolved as well. In particular, the standardization of structured query languages (SQL) enables data to provide dynamic data information at the record level [4]. The concept of knowledge discovery in Database (KDD) was presented in1989. The Data Warehouse was proposed in 1993, which was defined as a thematic, integrated, stable, and time-varying set of data that supports management decision-making processes [5]. Multidimensional database and online analytical processing (OLAP) contributed to the growth of data warehouses. As the core step in KDD, data mining was presented by the American Computer Society (ACM) in 1995. Since the 1990s, people have contributed to more complex data mining work benefited from the popularity of data warehouse technology and the development of DMTs. The rapid growth of the scale and size of data indicates the arrival of the big data age. Moreover, there is also rapid development of technology. Theses have led more people to explore how to use DMTs and more advanced algorithms to discover valuable knowledge in massive databases and more fields as well as more complicated conditions, so as to support them to make the best decisions.

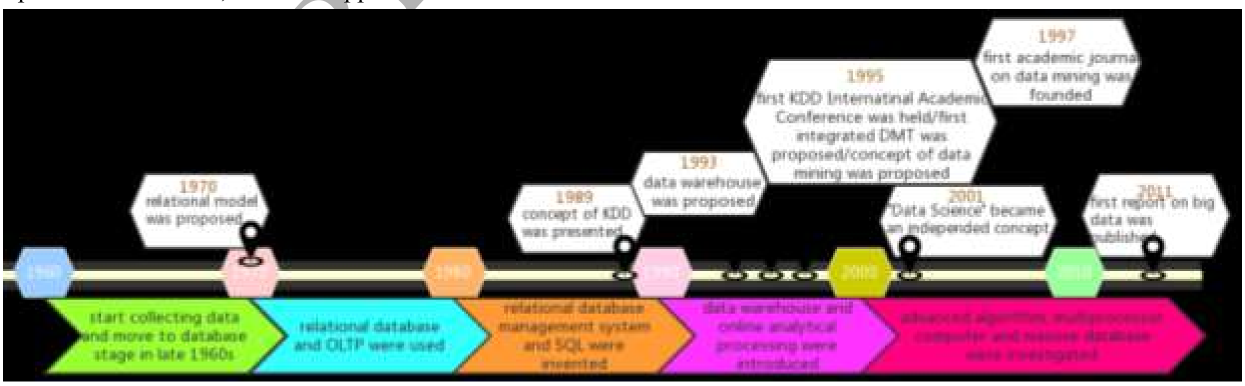


Fig. 1 The brief development process of DMTs

2.2 Major functions of DMTs

DMTs can be classified according to different standards. DMTs are used to extract knowledge from massive data, and the output knowledge can be divided into different kinds based on different functions. Considering different aims and requirements of DMTs, they can be divided into the descriptive functions and the predictive functions, as shown in Fig. 2. In this section, we introduce the related concepts of data mining functions and compare the differences between similar functions.

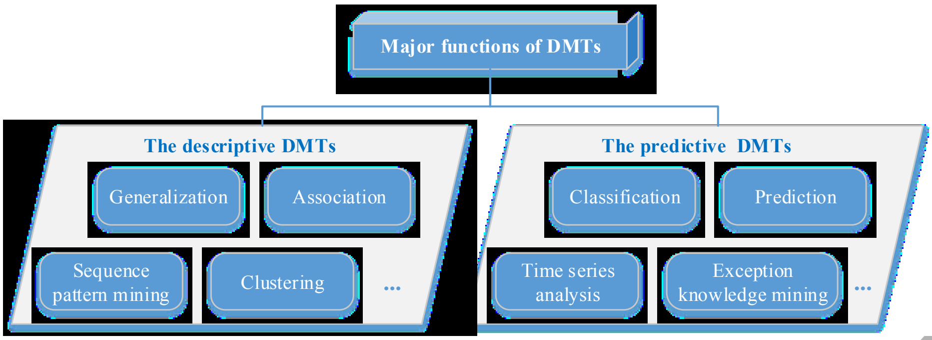


Fig.2 Classification of DMTs by different functions

2.2.1 The descriptive functions of DMTs

The descriptive DMTs mainly aim to explore the potential or recessive rules, characteristics and relationships (such as dependency, similarity, etc.) that exist in the data, such as generalization, association, sequence pattern mining, and clustering, etc. These extracted knowledge can be used to recognize critical parameters contributing to construct efficient solutions and universal rules, account for some phenomena, thus to realize better utilization of existed conclusions, better management and exact prediction.

Generalization refers to generalize knowledge that describes the characteristics of a sort of data. Source data (such as the database) generally stores the detailed data, and the purpose of generalization is based on the microscopic characteristics of these data to find the universal or higher-level concept of knowledge. The generalized knowledge can be used as a basis for excavation for other kinds of knowledge (such as classification and prediction). For example, it can be applied in quality description to recognize the critical attributes. The existing methods of generalization mainly include concept description method, multidimensional data analysis, and concept stratification technology.

Association reflects the dependency or association among an event and with other events. The purpose of association is to find the interdependence between features or data from a large number of data. Based on different association rules, it can be divided into simple association, timing association, causal association, etc. Association rules mining is the most common method for association. The most famous is Apriori proposed by Agrawal *et al.*[6] and its improved algorithm. The association rules are mainly used in business practice, including analysis of customer behavior, market segmentation and selection of target customers.

Mining sequential pattern refers to the discovery of high frequency time-related or other sequence-related sub-sequences from the sequence database. In other words, it can be described as the pattern excavation of sequences that is often presented in a set of well-organized data sequence. It is similar to association. However, its object is a sequence with a certain order and it is predictive. The algorithm, AprioriAll[7] and its improved algorithm, are mainly used for this function. Mining sequential pattern are usually used in forecasting merchandise sales, forecasting highest electricity load, exploring frequent gene combinations appeared in a class of patients, and so on.

Clustering divide a group of individuals into a number of categories in accordance with the similarity. Its purpose is to figure out both the differences between individuals in the same category as small as possible and the differences between different categories of individuals as large as possible. The most representative one of clustering methods is based on geometric distance measurement. But in many cases, the result of clustering is to form a concept, so some problems need to be solved based on the description of the concept. At present, there are some other methods used for clustering, such as neural network (NN), rough, fuzzy sets, and so on. The applications of clustering distribute in marketing, urban planning, psychology, archaeological research, earthquake research, meteorology research, and so on.

2.2.2 The predictive functions of DMTs

As to the predictive DMTs, they are usually selected to analyze the relevant trends of the data to or the relevant laws to predict the future state. It includes classification, prediction, time series analysis, and exception, etc. This kind of DMTs concludes the characters or tendency from existed data with an aim to predict the future classification or continuous value according to the inferred future data trends.

The purpose of *classification* is to construct a classification function or a classification model (i.e., classifier) according to the characteristics of the data set, and to classify each object in the data set into a known object class. The data classification process consists of two steps. The first step is at the model establishment phase, or the training phase. The purpose of this step is to describe the classifier of the predefined data class or concept set. The second step is to use the classifier from the first step, so as to evaluate the prediction accuracy of the classifier. Classification has a wide range of applications, such as medical diagnosis, credit card rating, image pattern recognition, target market positioning, fault detection, effectiveness analysis, graphics processing, insurance fraud analysis, and so on.

Prediction is to mine the knowledge that is generated by historical and current data and can infer future data trends. Although classification is used to predict class labels, people often want to predict some missing or unknown data values rather than class labels. That is, the desired prediction result is the numerical data. Therefore, we think using historical and current data to generate and output continuous curves and other issues as the main form of prediction function. Prediction is also widely used in many fields, such as the identification of objects in large image databases, credit assessment, medical diagnosis, performance prediction and marketing, etc.

Time series analysis is the extraction of information and knowledge from a large number of time series data with one or more time attributes that people are not known in advance but are potentially useful for short, medium and long term prediction. The time series is a special form of data, and the past values of the sequence affect future values. Time series analysis is mainly used to solve two types of problems. One is to summarize the data sequence or trend, such as studying the purchase behavior of supermarket customers, forecasting stock, futures trading, forecasting page click sequence records, etc. The other is to monitor the periodic change of data.

Exception knowledge mining is the extreme special case contained in the source data or the knowledge description that is distinct from the other data, which reveals the answers why object deviates from the normal law. Exception knowledge mining, including outlier analysis, sequence anomaly analysis, specific rule discovery, etc., is a process to find out behaviors that are different from the expected objects. It can be combined with other DMTs to further acquire exception knowledge while digging common knowledge, for example, the anomaly cases in the classification, special cases without meeting the general rules, data clustering outliers, etc. It is of great value in some areas, such as insurance and credit card fraud, loan approval, medical analysis, network security, intrusion detection, text mining in the new theme discovery, and so on.

2.2.3 Functions comparison

Known from the above description, the differences between some similar functions of DMTs are summarized as shown in Table 1, including the following five groups of comparisons.

- *Firstly*, with training instance and pre-defined class identifier, classification always helps the new element find the belonging class under the specific class identifier, while clustering generates new class identifiers by comparing and analyzing the data.
- *Secondly*, the main difference between prediction and classification is that the classification focuses on classification label (i.e., discrete value) prediction, while the prediction creates a continuous-value function model, which focus more on the future state.
- *Thirdly*, time series analysis describes the time changing data value and trends of the same object, while prediction predicts the data values of another objects through some parameters with sequential information ignored.
- *Fourthly*, association is looking for an association relationship between two events, without predicting whether a future event occurs or not, so association is a descriptive function while prediction is a predictive function.
- *Finally*, association can only dig the relationship between items in the same transaction, while mining sequential

pattern can also dig it in different transactions and both the objects and results of sequence pattern mining are ordered in time or space.

Table 1 Comparison on some similar functions of DMTs

The specific characteristics of the two compared similar functions	***Classification*** v.s. ***Clustering***	***Classification*** v.s. ***Prediction***	***Time Series Analysis*** v.s. ***Prediction***	***Association*** v.s. ***Prediction***	***Association*** v.s. ***Mining sequential pattern***
The former one	*Classification* is predictive and supervised	*Classification* is to predict classification label which is a discrete value	*Time Series Analysis* is to predict changing value or trend of the same object	*Association* is descriptive and reveals '*if A, then B*'	*Association* reveals different events relationship in the same transaction
The latter one	*Clustering* is descriptive and unsupervised	*Prediction* is to predict data state which is a continuous value	*Prediction* is to predict value of different data objects	*Prediction* is predictive and reveals '*if A or/and B, how will C be*'	*Mining sequential pattern* reveals different events relationship in the same or different transactions, both objects and results are ordered

2.3 Major categories of DMTs

The general process of data mining includes problem clarification, data collection, preprocessing, data mining in the narrow sense, and interpretation and evaluation of results. Whereas, data mining in the narrow sense refers only to a step to generate a specific pattern using a particular algorithm within an acceptable computational efficiency limit. In this section, we introduce two categories of DMTs, i.e., the enable techniques and some specific mining techniques, and summarize the functions they can realize.

2.3.1 The enable techniques

We refer to the techniques that support the implementation of data mining process in the narrow sense as the enable techniques, that is, the techniques supporting the process that raw data change to available data for mining algorithms. Before data mining, data index and query techniques are needed for effective management of massive data, so that data can be quickly accessed. There are three aspects of challenges for these techniques, i.e., the use of lower complexity search algorithm, the multi-dimensional data index techniques, and the temporal data indexing techniques. Multi-dimensional data index refers to the data in the multi-dimensional data space, and the return value is a row of data. Each data is associated with many other data, and a multi-dimensional data index is created based on the associated data, enabling index and query of multiple dimensions. Since the data in the actual life has a lot of dimensions, the dimension reduction is required before the multi-dimensional index processing. Not only current data but past and future multi-dimension data are needed for decision-making, which makes the index technique for tense data a key technique. Since structure and storage of the tense data are special, it is need to use the composite tense data structure index.

Data preprocessing is an important step in ensuring the successful achievement of data mining. It can improve data quality, which helps to improve the accuracy and performance of the subsequent mining process. The process of data preprocessing mainly includes data cleaning, data integration, data transformation and data reduction. Data cleaning involves padding missing data, smoothing of noise data, identifying or removing isolated points, and resolving data inconsistencies. Data integration is to store all the data in a database, data warehouse or file to form a complete data set with eliminating redundant data. Data transformation mainly converts original data to the required format to meet the mining needs, such as limiting the data value in a specific range. Data reduction removes the attributes that cannot characterize the key features of the system to reduce the amount of data under the premise that the reduced representation of the data set close to the original data, as well as obtain the same or similar analysis results. In particular, discretization of data is a very important part of data reduction, especially for numerical data. According to different characteristics or quality of the data object and based on the different requirements of the chosen mining algorithms, the corresponding data preprocessing operation is selected and not every preprocessing operation is necessary. Some techniques or methods are used in data preprocessing stages. For example, regression and clustering analysis techniques for noise data processing, decision tree and NN method for data cleaning process, etc.

2.3.2 The mining techniques

The mining techniques are the algorithms or methods used in data mining in the narrow sense to explore useful knowledge from massive data. The mining techniques can be further divided into two categories: statistical analysis (SA)-oriented DMTs and knowledge discovery (KD)-oriented DMTs. SA-oriented DMTs make assumptions about data distribution and relationship between variables based on prior knowledge in advance and verify or deny the assumptions then. On the contrary, KD-oriented DMTs search the relationship automatically under no clear assumption. In addition, we can define them according to whether there are clear expressions. SA-oriented DMTs are often expressed as one or a group of expressions while KD-oriented DMTs focus more on the results without clear expressions. What's more, SA-oriented DMTs remove unnecessary information under assumptions while the amount of information of KD-oriented DMTs remain constant or increase. The specific mining techniques of each category and their available functions are summarized in Table 2.

Table 2 Summary on the specific mining techniques of each category and their available functions

Techniques \ Functions		Genera-lization	Classi-fication	Clustering	Association	Prediction	Exception Mining	Time Series Analysis	Mining Sequential pattern
SA-oriented DMTs	Regression		√			√	√	√	
	Bayesian		√	√			√	√	
	Clustering			√			√		
	KNN		√	√					
	…								
KD-oriented DMTs	SVM		√			√	√	√	
	Decision tree		√						
	Attribute-oriented induction	√					√		
	Case-based reasoning		√				√		

	(CBR)								
	GA		√	√			√		
	Cobweb			√			√		
	NN		√	√		√	√	√	
	Apriori				√		√		
	AprioriAII								√
	GSP								√
	FP-Growth				√		√		
	Rough set	√	√	√	√		√	√	
	Fuzzy set	√	√	√	√		√	√	
	…								

(1) SA-oriented DMTs

SA-oriented DMTs use the mathematical models to induce and analyze existing data as well as infer and predict unknown information. Usually they can be used in the derivation from history to future, from local to overall, etc. They can also check out and explain those abnormal forms of data. It includes linear analysis, nonlinear analysis, regression analysis, logistic regression analysis, nearest neighbor algorithm and clustering analysis, etc.

Regression is an analysis method to estimate and predict the trend and total mean of the dependent variables based on known independent variables. It is commonly used in classification, prediction, time series analysis and other functions. The K-nearest neighbor (KNN) method can detect the closest matching sample, which are often used in classification, clustering, prediction, and other functions. Clustering analysis divides data into multiple classes or clusters according to the similarity between them. It is suitable for exploring the internal relationship between the samples and also useful for the detection of isolated points. Bayesian is based on the Bayesian theorem and developed for the systematic interpretation and resolution of statistical problems, which is mainly used in classification, clustering, prediction and other functions.

(2) KD-oriented DMTs

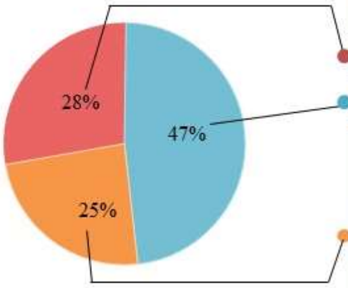
Unlike statistical analysis, the category of KD-oriented DMTs is data-driven, without need to put forward assumptions and problems in advance. They can screen information as well as find a possible model and unknown rules from a large number of data in data warehouse. This category mainly includes decision tree, rough set, NN, genetic algorithm (GA), association rule, support vector machine (SVM), generalized sequential pattern (GSP), etc.

Decision tree can find valuable and potential information by sorting large amounts of data with purpose. GA is a random search algorithm based on biological natural selection and genetic mechanism. It can be used for clustering analysis and classification and can also be used to evaluate the fitness of other algorithms because of its advantages in dealing with combinatorial optimization problems. Different types of NNs can be used for different data mining functions and the weight and topology of NN determine the type of pattern it can identify. For example, Back Propagation (BP) network and Radial Basis Function (RBF) network are used in the function of classification and prediction while

Adaptive Resonance Theory (ART) network is commonly used in clustering, prediction and exception knowledge mining. Rough set judges the importance of attribute in data to get and analyze rules by deleting redundant attributes and exception data values. It is commonly used in classification, attribute reduction and other problems. The form of association rule mining is usually expressed as "IF-THEN", which aims to find the correlation between two events by finding frequent item sets. In addition to implement the association function, it can also be used in prediction, time series analysis and other functions.

3. Applications of DMTs in production management

In this section, we investigate a total of 47 related literatures, including 13 (28%) using only one single SA-oriented DMT, 22 (47%) using only one single KD-oriented DMT, and another 12 (25%) using a combination of multiple DMTs. It can be seen from Fig.3 that we classify the applications of DMTs in production management into five fields, i.e., production scheduling, quality improvement, defect analysis, fault diagnosis, and other applications. The main task of production scheduling is to create and modify the scheduling rules. Quality improvement focuses on parameter optimization, quality classification, process monitoring, quality description and prediction. Fault diagnosis and defect analysis are mainly for identifying and classifying fault or defect state. In addition, there are other applications including yield prediction, time-related prediction covering flow time prediction, lot cycle time prediction and lead time prediction, and wear lever and remaining life of tools. So that we can carry out effective control of factors influencing yield and production time or causing abnormity. We conclude the detailed applications of these literatures in Table 3.



Categories of DMTs		Applications				
		Production scheduling	Quality Improvement	Defect Analysis	Fault Diagnosis	Other Applications
Only one SA-oriented DMT		2	8	1	0	2
Only one KD-oriented DMT		5	6	2	6	3
Multiple DMTs Combination	Multiple SA-oriented DMTs	0	0	0	0	0
	Multiple KD-oriented DMTs	0	8	0	1	2
	Both SA-oriented DMTs and KD-oriented DMTs	0	1	0	0	0

Fig. 3 Applications distribution of DMTs in production management

Table 3 The existing applications of DMTs in production management

Problems	DMTs	Categories	Functions	Applications	Utilities (+: Increase, -: Decrease)	References
Production scheduling	Clustering	SA-oriented	Clustering	Extract new scheduling rules	Performance^{+}, time or samples^{-}	8
	Regression		Prediction		Performance^{+}, time or samples^{-}, simplicity^{+}	9
	CBR	KD-oriented	Classification	Extract new scheduling rules	Performance^{+}, time or samples^{-}	10
	Decision tree				Performance^{+}, simplicity^{+}	11
	IOA				Performance^{+}, time or samples^{-}	12
	Decision tree			Choose optimal scheduling rule dynamically	Performance^{+}, simplicity^{+}	13
	NN				Time or samples^{-}	14
Quality Improvement	Bayesian	SA-oriented	Time sequence analysis	Parameter optimization	Accuracy^{+}, time or samples^{-}	15, 17, 18
	Regression				Accuracy^{+}	16
	KNN		Classification		Time or samples^{-}	19
	NN	KD-oriented	Classification	Parameter optimization	Accuracy^{+}	20
	IOA		Prediction			21
	Clustering	SA-oriented	Clustering	Classification of quality	Time or samples^{-}	22
	Regression & boosted trees	SA-oriented & KD-oriented	Classification	Classification of quality	Accuracy^{+}	23
	NN	KD-oriented	Classification	Classification of quality	Accuracy^{+}	24
	SVM				Accuracy^{+}, time or samples^{-}	25
	Bayesian	SA-oriented	Prediction	Process monitoring	Accuracy^{+}, time or samples^{-}	26, 27
	SVM & IOA	KD-oriented	Time sequence analysis/ classification	Process monitoring	Accuracy^{+}、time or samples^{-}	28, 32, 35
	NN & IOA		Classification			29, 30
	SVM&NN					31
	Decision tree				Time or samples^{-}	33

	NN					34
	PageRank		Generalization	Description or characterization of product or process	Accuracy+, time or samples-	36
	SVM		Prediction	Predicting quality	Accuracy+	37
Defect Analysis	Clustering	SA-oriented	Clustering	Defect classification	Time or samples-, accuracy+	38
	NN	KD-oriented	Classification		Accuracy+	39
	SVM					40
Fault Diagnosis	NN	KD-oriented	Prediction/classification	Fault pattern recognization	Accuracy+	41, 43, 44, 46
	SVM		Classification			42
	Decision tree					45
	Association rule		Sequence pattern mining			47
Other Applications	Regression	SA-oriented	Prediction	Flow time prediction	Accuracy+, time or samples-, simplicity+	48
	Bayesian		Classification	Lot cycle time prediction	Accuracy+, time or samples-	49
	SVM	KD-oriented	Prediction	Tool wear and remainder life prediction	Accuracy+ ,Time or samples-	50, 51
	Fuzzy set & NN			Yield prediction	Accuracy+	52
	Association rule					53
	Decision tree & NN			Flow time prediction		54

3.1 Applications in production scheduling

Applications in production scheduling can be divided into two sides. One is extracting new scheduling rules, that is extracting new scheduling rules or optimizing the existing scheduling rules from the obtained production data. The other is dynamically choosing optimal scheduling rules, that is establishing the mapping model of the current state of the production scheduling processes to the desired optimal scheduling rules, so that we can choose optimal scheduling rules on the basis of the current state dynamically.

3.1.1 Extracting new scheduling rules

(1) SA-oriented DMTs based applications

Tamura *et al*.[8] proposed a novel reactive scheduling method using local clustering organization, modifying the predetermined schedule to cope with variable situations without suspending the proceeding of the processes on the schedule. It can improve the schedule with fast computational time. Heger *et al*.[9] suggested Gaussian Process Regression to be applied to predict dispatching rule performance so as to select and switch dispatching rules depending on current system conditions timely. Despite better scheduling performance (e.g., reduced mean tardiness) compared to other techniques, a single Gaussian Process model can easily provide a measure of prediction quality, which make it a promising approach.

(2) KD-oriented DMTs based applications

Lim *et al*.[10] presented a scheduling method for semiconductor manufacturing systems through utilizing a CBR with the help of Petri nets to significantly outperforming a meta-heuristic algorithm in terms of computation time as well as requiring much less sacrifice in a performance metric than the competitive dispatching rules. Olafsson *et al*.[11] proposed a two-step method based on data mining classification and GA optimization to extract new scheduling rules from historical scheduling data. Firstly, the GA is used to extract the optimal training from the historical scheduling data. Then the decision tree classification technique is used to extract the new scheduling rules from the extracted training data sets. It proved that the extracted scheduling rules can produce scheduling schemes superior to historical ones. Shahzad *et al*.[12] used the data mining method to extract new scheduling knowledge from the job scheduling problem optimization solutions obtained by using the tabu search algorithm. The extracted scheduling knowledge can be used as a new scheduling rule and well approach the scheduling performance of the ones obtained from tabu search algorithm.

3.1.2 Dynamically choosing optimal scheduling rules

(1) KD-oriented DMTs based applications

Metan *et al*.[13] developed a new scheduling system combining the techniques of simulation, data mining, and statistical process control charts, which can select dispatching rules in real time. The proposed scheduling system selects a dispatching rule periodically from a decision tree constructed by extracting knowledge from data coming from the manufacturing environment and has the capacity to adapt itself to changes by dynamically updating the decision tree whenever the manufacturing conditions change. Azadeha *et al*.[14] proposed a new algorithm based on computer simulation and artificial NNs (ANNs) to select the optimal dispatching rule for each machine from a set of rules. It can provide higher precision and is capable of finding the optimal solution of SJSSPs since it evaluates all possible solutions.

3.2 Applications in quality improvement

Applications in quality improvement are divided into five parts, parameter optimization, classification of quality, process monitoring, description or characterization of product or process, and predicting quality.

3.2.1 Parameter optimization

Parameter optimization determines the optimal level of process or product parameters that produce target quality based on learned characteristics of high quality, and these parameters can be used as the bounds of the control charts or the assist of establishing an exact model.

(1) SA-oriented DMTs based applications

Celik *et al*. [15] introduced a sequential Monte Carlo method (sequential Bayesian inference technique) and embedded it into the simulation to enable its ideal fidelity selection for given massive datasets. The proposed method incorporates available dynamic data and steers the measurement process for selective data update and can employ both discrete and continuous variables directly to formulate the dynamics of systems under uncertainties and perform fidelity selection in a rigorous but efficient manner. Tao *et al*.[16] quantitatively characterized the height variation of the Carbon nanotube(CNT) array by constructing a spatial–temporal model with regression method to improve the quality and stability of the CNT array. The effectiveness in terms of goodness of fit and prediction accuracy of the proposed models is shown by a practical case study. Wang *et al*.[17] proposed a Bayesian framework for parameter estimation when only low-resolution information on categorical scale is available. The proposed method incorporates low-resolution information recursively and updates parameter estimates related to quality characteristics in statistical process adjustment (SPA) in real time, so that it is possible to use low-resolution quality characteristic information and realize accurate and quick parameter estimation in the quality control of the production process. Du *et al*.[18] proposed an engineering model-based Bayesian approach to estimate the process control parameters used to establish control limits for cause-selecting chart(CSC) to monitor the transition phase of MMPs. This approach can reduce the samples needed to complete the estimation due to the combination of the engineering model and historical data, which is very useful when scarce measurement information is presented. Since that, this method can be used to monitor and find quality problems as early as possible. Ranjit *et al*.[19] addressed fault diagnosis under a circumstance that only few samples of training data are available and lack fault samples and propose a fault detection technique using human machine co-construct intelligence (FD-HMCCI), to optimize parameters (e.g. number of neighbors and confidence level) tuned to enhance accuracy of the fault detection. Both the distribution of process variables and the expert's domain knowledge are taken into account in the proposed approach, which are represented as a form of envelopes and use KNN method to evaluate the results.

(2) KD-oriented DMTs based applications

Dong *et al*.[20] established the relationship between the single-bead geometry and the welding process parameters based on ANN model. The proposed model is used to predict the combination of the optimal welding process parameters (i.e. the wire feed rate and the travel speed), which will produce the desired bead geometry. Kumar *et al*.[21] mapped extensible markup language (XML) data from feature-based models and developed process parameters based on experimental research and optimization using GA to result in the corresponding process parameters required for the manufacture of micro parts.

3.2.2 Classification of quality

Classification of quality classifies the quality of the products into several different levers and adopts effective measures according to the lever predicted ahead of time to improve the quality of the products.

(1) SA-oriented DMTs based applications

Sahebjamnia *et al*.[22] used process mootness factor based sub-assemble products (SAP) predefined clustering algorithms to cluster homogenous SAPs. The proposed method contributes to the reduction of the sub-standard SAPs cost by dividing these SAPs to inhomogeneous and defect SAPs. Weiss *et al*.[23] invoked a linear regression method with boosted trees to update the estimation of final characteristic of a product after each manufacturing operation in order to continually predict manufactured product quality prior to final testing. The initial application is for the speed prediction of microprocessors to perform early corrections and it has been proved that the proposed method can provide a good prediction accuracy.

(2) KD-oriented DMTs based applications

Podrˇzaj *et al*.[24] designed an artificial intelligence monitoring system based on linear vector quantization NN, determining the quality of each weld based on the measurements performed in real time. The displacement of the upper electrode and variable welding force consist of the input vector for linear vector quantization NN in view of the relationship between various parameters of the displacement curve and the weld strength and disturbances in air supply.

And the output signal classifies a certain measurement into a good quality weld or a bad quality weld. Du *et al.*[25] proposeed a novel classification approach for workpiece surfaces based on the extracted features by combining dual-tree complex wavelet transform and selective ensemble classifiers called modified matching pursuit optimization with multiclass support vector machines ensemble (MPO-SVME). SVM is adopted as basic classifiers. A case study demonstrated MPO-SVME can increase the classification accuracy without sacrificing much computational expense.

3.2.3 Process monitoring

Process monitoring monitors the manufacturing process and recognize the unusual patterns. Some may try to identify the source of the out-of-control signal or the change point of the quality quickly. So that the engineer can repair the abnormity as soon as possible.

(1) SA-oriented DMTs based applications

Miloa *et al.*[26] developed Bayesian Posteriors Updated Sequentially and Hierarchically (BPUSH) to detect the anomalies in a system by monitoring the parameters of a relevant measurement (e.g., temperature, voltage) in how they affect the shape of posterior Cumulative Distribution Functions. Other than rapid and reliable detection of anomalies at real-time processing speeds, the proposed method outperforms previous approaches in low computational cost and low false alarm rates. Rao *et al.*[27] suggested an advanced Bayesian nonparametric to analyze in situ heterogeneous sensor data. This sensor data-driven defect detection approach optimizes process conditions for obtaining the best surface roughness and facilitates real-time identification and correction of fused filament fabrication (FFF) process drifts with a high accuracy and precision. Azadeh *et al.*[28] developed a support vector regression (SVR) model integrated with an imperialist competitive algorithm aiming to improve the performance of failures and reliability estimation of components and systems, which is applied to optimize the selection of the SVR parameters. The experimental results indicate that the proposed model can achieve high estimation accuracy.

(2) KD-oriented DMTs based applications

Du *et al.*[29] developed an improved particle swarm optimization with simulated annealing based selective NN ensemble approach to identify the variation sources of out-of-control signals in the aircraft horizontal stabilizer assembly processes. Yang *et al.*[30] introduced an effective multivariate statistical process control model enabled by two-level discrete particle swarm optimization-based selective ensemble of learning vector quantization networks for monitoring and diagnosing of mean shifts in multivariate manufacturing processes. In this model, one network is developed for detecting out-of-control signals in process mean, while the other network is developed for further classifying the detected out-of-control signals as one of the specific mean shift types. Those ensemble techniques provide system with better generalization capability and makes it easier to be understood and modified, and perform more complex tasks than any of its components (i.e., component NNs in the ensemble). Salehi *et al.*[31] proposed a hybrid learning-based model consisting of two modules for on-line analysis of out-of-control signals in multivariate manufacturing processes. In the first module a support vector machine-classifier is used to recognize the type of unnatural pattern, while in the second module three NNs for shift mean, trend and cycle are used to recognize magnitude of mean shift, slope of trend and cycle amplitude for each variable simultaneously. Du *et al.*[32] suggested an improved particle swarm optimization and simulated annealing based selective multi-class support vector machine set method, in which some selective multi-class SVMs are used in the classification of source of the average offset of multivariable control graphs. He *et al.*[33] presented a bivariate mean shift monitoring and fault recognition model based on decision tree under the assumption of constant variance-covariance matrix. Two decision tree classifiers based on C5.0 algorithm are constructed, one for monitoring and another for fault identification. The simulation results show that the proposed model can not only detect the mean offset, but also give information about the subset of variables or variables that cause the automatic control signal and its departure direction. Ahmadzadeh *et al.*[34] used the multivariate index-weighted moving average control chart and NN to identify the runaway signals of the production process more efficiently and more quickly. The method can predict the change point and speed up the search for the parameter that caused the failure, thus improving the detection quality while reducing the total cost of detecting the fault. Hu *et al*.[35] presented an intelligent ensemble model for quality fluctuation analysis to estimate the variance change point in multivariable process to accelerate the location of assignable causes causing the quality fluctuation of process and make measures for process adjustment. There are two crucial techniques included in the model: moving window analysis for process decomposed and multi-kernel SVM model formed based on particle swarm optimization. The particle swarm optimization is considered to search the optimized multi-kernel parameters.

3.2.4 Description or characterization of product or process

Description or characterization of product or process identifies attributes or variables that affect quality obviously and sort these attributes or variables according to the significance. In addition, it can determine how low, medium and high-quality products are naturally grouped in the data and identify the most likely causal relationship between low-quality products and high-quality products.

(1) KD-oriented DMTs based applications

Xu *et al*.[36] proposed a QCD-NodeRank indicator and an evaluation method based on improvement of PageRank method, aiming to evaluate the critical degree of a manufacturing resource node (MRN) of quality control network constructed by using unqualified batches' manufacturing information. The MRN stands for the set of physical elements of a workstation, including the manufacturing machine and its affiliated cutting tools, fixtures, gauges, workers and so on. The MRN with high QCD-NodeRank values is considered to be the critical one and needs to have priority attention when doing quality control and improvement activities of the shopfloor.

3.2.5 Predicting quality

Predicting quality develops a model that associates the quality input characteristics with the output when the quality output is a real value variable and uses this model to predict the quality characteristics of a given set of input parameter values.

(1) KD-oriented DMTs based applications

Diao *et al*.[37] presented a dynamic quality control approach by improving dominant factors (DFs) based on improved principal component analysis (iPCA). Using iPCA to identify the DFs which lead to quality problems, a quality prediction model for improving DFs is proposed based on modifiedSVM. The modification in SVM is realized by introducing an incremental weight to improve its sparsity, so that increasing the accuracy of quality prediction.

3.3 Applications in defect analysis

Defect analysis problems recognize and classify defect problems into different kinds. It can be realized by clustering defect problems with different levels of accuracy, which may produce new defect kinds, or classifying defect problems into known defect kinds.

(1) SA-oriented DMTs based applications

Tao *et al*.[38] proposed a multistep defect analysis approach based on K-nearest neighbor noise removal technique and two kinds of cluster analysis techniques, which provides clustering results with different levels of accuracy. The proposed approach which estimate the number of defect clusters and the spatial pattern of each cluster a priori has superiority both in computational speed and detection accuracy when analyze general defect patterns generated during the IC fabrication process.

(2) KD-oriented DMTs based applications

Weimer *et al*.[39] proposed a new approach based on deep convolution NN to perform visual defect detection. The proposed method is used to measure 12 different categories of the data sets in the context of severe texturing. The proposed method has a low false alarm rate for defect detection results. Motivated by the aim to improve the classification capability of automatic optical inspection in CMOS image sensor manufacturing, Chen *et al*.[40] developed a manufacturing intelligence framework combining defect detection, feature extraction, SVM classifier and similar matching methods to reduce false alarms of defect classification while improve capture rates.

3.4 Applications in fault diagnosis

Fault diagnosis problems can be described as working states and fault types identification and prediction of machines or parts. And the results can be used to predict the maintenance times and give the optimal maintenance solutions.

(1) KD-oriented DMTs based applications

Hsu *et al*.[41] introduced an intelligent maintenance prediction system using the radial basis function NN and variability of the working attributes identified by feature extraction to predict the maintenance times and aging of the LED wafer test machines. Wu *et al*.[42] proposed a non-invasive state monitoring method for fused deposition modeling (FDM) machines based on acoustic emission（AE）sensing technology. The different extruder operating conditions of the FDM machine are identified using the time domain characteristics collected by AE technique and classified by SVM applying the standard deviation of the energy of the acrylonitrile butadiene ester. Ding *et al*.[43] investigated the application of DMTs in fault diagnosis and maintenance operations of turbine. They use artificial NN to identify the fault type of the manufacturing grid and combine CBR and fuzzy set to give the optimal maintenance solution to ensure a timely response to the failure of the turbine no matter how far the maintenance department is. Kane *et al*.[44] tried to put the psychoacoustic features and statistical indices extracted in the data as the input of the artificial NN to identify the state of the gearbox. It was found that the two characteristic parameters could detect the gear box with good precision without depending on the experience and ability of the operator. Unal *et al*.[45] used DMTs to show the potential of sound analysis in fault diagnosis of rolling bearings. The fault diagnosis is done in three parts and finally the classification is done according to five class values, normal and four different types of faults. Decision tree algorithms, SVM and boosting used with the decision tree algorithms are used to build classifier models. Although decision tree and SVM show comparable performance, the rest method outperforms all the other classifiers at smaller defect sizes at different loads and speeds. Khazaee *et al*.[46] presented a novel approach based on vibrational signal and an ANN classifier for detection of the three operating conditions, normal operating conditions, high-load operating conditions and high-temperature operating conditions, for a timing belt in an internal-combustion engine to help to preserve the engine from eventual failures. Rashid *et al*.[47] proposed a data mining-based framework for fault identification and anomaly detection from machine vibration data. In this framework, they introduce associated frequency patterns and a mining algorithm called SAFP (sliding window associated frequency pattern). Their SAFP algorithm can mine associated frequency patterns (i.e., fault frequency signatures) and use them to identify faults in the bearing data.

3.5 Other applications

There are prediction problems solved by using DMTs including yield prediction, flow time prediction, lot cycle time prediction and lead time prediction, and wear lever and remaining life of tools. Effective control of attributes influencing yield and production time or causing abnormity can be realized according to the exact predictions.

(1) SA-oriented DMTs based applications

To provide responsive and high-quality prediction of a new job's flow time through the system and ensure accurate and real-time lead time quotation, Li *et al*.[48] attempted to design thoroughly simulation experiments and estimate regression models to quantify the relationship between the characteristics (e.g., mean, variance, etc.) of the flow time distribution and the predictor variables. Meidan *et al*.[49] suggested a data-driven approach using selective naive Bayesian classifier to select a minimal, most discriminative key-factor set for CT prediction. The proposed method dramatically improves the accuracy of the prediction and has an advantage on simplicity and interpretability, as well as speedy and efficient model training.

(2) KD-oriented DMTs based applications

Benkedjouh *et al*.[50] introduced a method based on nonlinear feature reduction and support vector regression. The number of original features extracted from the monitoring signals is first reduced and these features are then used to learn nonlinear regression models to estimate the wear progression and predict the remaining useful life of the cutting tools.

Zhang *et al*. [51]suggested the least squares SVM to be applied in estimation of tool wear for milling process. Both machining parameters and position parameter of ball-end cutter are considered as input in the established model, while the output of the proposed model is tool wear of cutting edge position. After analysis and comparison of predicted performance given by taking different tuning parameters and data regularization, it shows the proposed method has a better prediction performance at a certain range of cutting conditions. Wu *et al*.[52] proposed a yield prediction model based on fuzzy NN, taking the impact factors of yield and critical electrical test parameters into account simultaneously and they are taken as independent variables. This method has improved the prediction accuracy compared with the traditional method, which is beneficial to improve the production and reduce the cost. Lee *et al*.[53]proposed a predictive association rule considering the event sequence algorithm that creates a set of rules, in which each rule estimates the yield for a sequence of events including alarm events, change events, and maintenance events. These rules are used to predict the final yield of a printed circuit board and the predictive accuracy is improved compared to those of the regression models that did not consider the event sequences. Tirkel *et al*. [54]developed a flow time prediction model using decision tree method and NN method respectively for individual production steps, line segments or complete lines to generate flow time prediction.

4. Discussions, limitations, and suggestion

4.1 Discussions

In this section, we discuss DMTs used in production management and compare the traditional methods with DMTs, and then point out their connections and differences.

4.1.1 Summary on DMTs used in production management

DMTs can be used in different fields of production management, including production scheduling, quality improvement, defect analysis, fault diagnosis, and other applications. They are used to search useful and efficient patterns or rules and find problems and unknown mutation operations, to improve production efficiency and product quality more intelligently and exactly and adjust the production plan timely. According to Fig. 4, we can use DMTs to dig out useful knowledge from the data and make the definition and correction of the various parameters of the model in realtime. Using DMTs to extract knowledge from data expressing and implementing the schedules, making them as input conditions and constraints, and even the definition of the models, it will be of significance for improvement of decision-making and prediction performance. Besides, DMTs can be used to establish the mapping knowledge model from the current state to the desired optimal target state. The pattern matching rules can effective guide the production management processes.

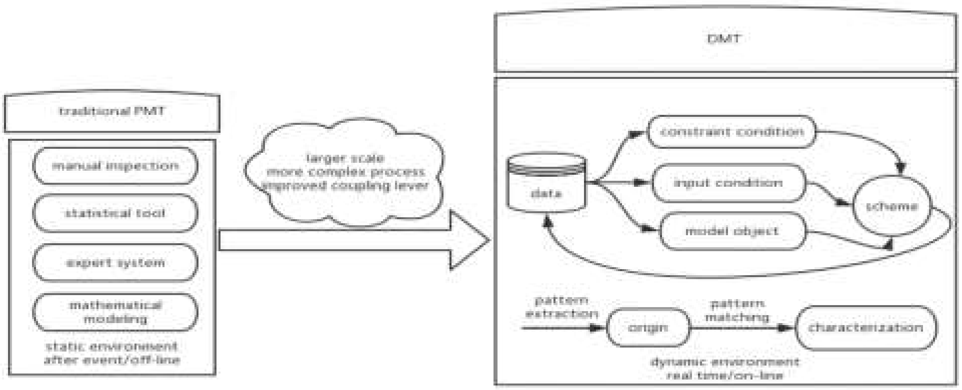


Fig. 4 the comparison between traditional methods and DMT

The category of SA-oriented DMTs can model the relationship between quality characteristics or other target value of products or production processes and different parameters to optimize the production processes and improve products quality. When analyzing data using SA-oriented DMTs, we must know what variables are and what we need to analyze before the analysis begins. While the category of KD-oriented DMTs do not have to put forward assumptions or questions, but it can still find those unexpected information, representing the relationships and patterns between different data elements. By using this category of DMTs, the computers can automatically discover the characteristics related to the machining processes and automatically extract the fault discrimination rules so that the working status of the devices can be identified and classified, find the source of abnormal information fast and accurate, and achieve the purpose to control the products quality dynamically and effectively.

In addition, we can learn the following seven conclusions by comparing the two categories of DMTs from Fig.3 and Table.3.

(1) Although the two categories of DMTs both have relevant literature in extracting new scheduling rules, the applications are different in detail. SA-oriented DMTs predict the performance of scheduling rules through modeling while KD-oriented DMTs extract rules from a series of data or scheduling rules established before to generate better scheduling rules.

(2) SA-oriented DMTs are never used to choose optimal scheduling rules in the investigated literature, because KD-oriented DMTs have advantage in pattern matching while SA-oriented DMTs are not available to complete this task.

(3) Although the two categories of DMTs both have applications in process monitoring, KD-oriented DMTs also focus on the source of out-of-control signal recognition except the abnormal recognition.

(4) KD-oriented DMTs have two more parts of applications in quality improvement, i.e., description or characterization of product or process and predicting quality. According to the definition or characterization of product or process, the premise of completing the task is to recognize the critical attributes impacting quality, but SA-oriented DMTs cannot do this. Because there are too many variables, SA-oriented DMTs have no ability to output quantized quality characteristic value through modeling, so these DMTs can only realize classification of quality.

(5) There are no relevant literatures about using SA-orientedDMTs in fault diagnosis. We can describe the applications of DMTs in fault diagnosis as fault pattern matching. In detail, DMTs are used to search the pattern matching relationships between characterization of a fault and rules of a fault from abnormal production information of machines and then guide the location of faults. And SA-oriented DMTs cannot be competent in this task.

(6) The function of using SA-oriented DMTs in defect analysis is clustering while that category of KD-oriented DMTs is classification. It can be seen from this that SA-oriented DMTs can divide defects into different parts under different classification accuracy levers and we cannot know what and how many the category will be ahead of time. But KD-oriented DMTs are used to classify the defects into the categories defined before and this may cause some defects which cannot be classified into proper categories.

(7) We can see from the pie that there are 25% of literatures using a combination of different DMTs, and the single one of KD-oriented DMTs are used more than the single one of SA-oriented DMTs. In addition, there are general three kinds of typical combinations. It proves that the combination application of different DMTs can generally improve the performance, the generalization ability, the ability to escape from the local optimal value of a single DMT, and even finish the complicate tasks that a single DMT is unable to complete.

Under the background that the traditional methods cannot meet the actual needs of production, the use of DMT can be a good response to the challenges faced by traditional methods, which are introduced in next section. Specifically, there are the following four advantages of using DMTs.

(1) As the comprehensive consideration of various state information of the production environment is carried out, the generated program performance is better and the prediction results are more accurate.

(2) It is available to complete tasks in a short period of time to meet the real-time requirements of production management.

(3) There is almost no need to rely on a priori knowledge and the experience of the operator to avoid the uncertainty of subjective factors.

(4) Compared with the traditional methods of expost facto inspection, DMTs can complete the prediction of faults, defects or abnormal modes, and production status assessment ahead of time with good accuracy to take measures to deal with the uncertainty of production management.

However, it cannot be said that DMTs can replace traditional methods completely. The relationship between them can be concluded that traditional methods provide study samples for DMTs and DMTs help identify and modify the parameters of traditional methods to improve the performance of schemes.

4.1.2 The traditional methods and challenges faced

The traditional methods, including manual inspection, statistical tools, expert systems and mathematical modeling, are for static production management environment. The four methods are defined as follows, respectively.

- Manual inspection: the operator uses the senses or testing tools, relying on their own experience and knowledge to determine the production status of the product, so as to control the production process.
- Statistical tools: use frequency distribution, control charts, significant test and other statistical techniques to proceed production management, which is characterized by finding the key factors affecting the production process and then taking measures to control production, reduce the fluctuation of production process and product quality to reach the aim to improve the quality of production management.
- Expert systems: accumulate knowledge through the manual inspection, statistical tools and mathematical modeling and other means, form an expert system through the relationship between characterization and the cause and use the formation of the expert system to guide the production management tasks.
- Mathematical modeling: Under the premise of a specific hypothesis, the variables determining the target variables are used as input and the relevant indicators are used as output under constraints.

The scale of modern manufacturing industry is increasingly large, the manufacturing processes are more complicated, and the coupling degree between the manufacturing links is increasing, and the probability of dynamic uncertainties in the manufacturing processes also increase. All of these pose a great challenge to the traditional methods. It mainly reflects in the following four aspects correspondingly.

For the manual inspection methods, they rely on the operator's senses and experience to judge, it is time-consuming and labor-consuming and contains the uncertainty caused by subjective factors. Moreover, sampling detection for some products is destructive and a batch of product quality cannot replace the quality of all products. In addition, because it is off-line after the test, when adjusting the product quality process after the quality problems, it has caused a lot of production waste.

When using statistical tools to carry out production management, the statistical analysis is weak in terms of implementation and reliability due to the decentralization, disorder and completeness of manufacturing quality data. In actual production activities, it is often used as a purely post hoc statistical tool to statistical analyze the situation after the production problem.

For the expert systems, they has been abstracted from the relevant expert knowledge that has been accumulated for a long time, usually uses only limited information to complete the task. So its performance is often poor. The results are not the same got from different experts, so there is no good inheritance and only can be applied to a fixed range.

As to mathematical modeling of complex production processes, it is difficult to establish accurate mathematical model due to difficult mechanism modeling. And usually human assumptions reduce the actual complexity in terms of establishing the model under the assumptions, so it is difficult to guarantee the accuracy of the model built. At the same time, due to lack of accurate and timely model parameters, further causing the scheme obtained by the mathematical modeling is relatively low reliability and making it difficult to directly guide the actual production. Moreover, this approach requires a lot of computing time and execution time, which make it difficult to meet the real-time demand of actual production, and at the same time led to the model not easy to modify and cannot make a quick and effective response to dynamic environment.

4.2 Limitations and suggestions

DMTs has undergone many years of development. Every year new data methods and models come out, the existing DMTs are also improved, and the applications of the fields are also increasingly widespread. However, in the big data era, DMTs still have limitations, such as the efficiency of data mining methods need to improve, especially large-scale data set mining problems. The following is a discussion of the limitations of data mining in production management in view of the characteristics of data in production management.

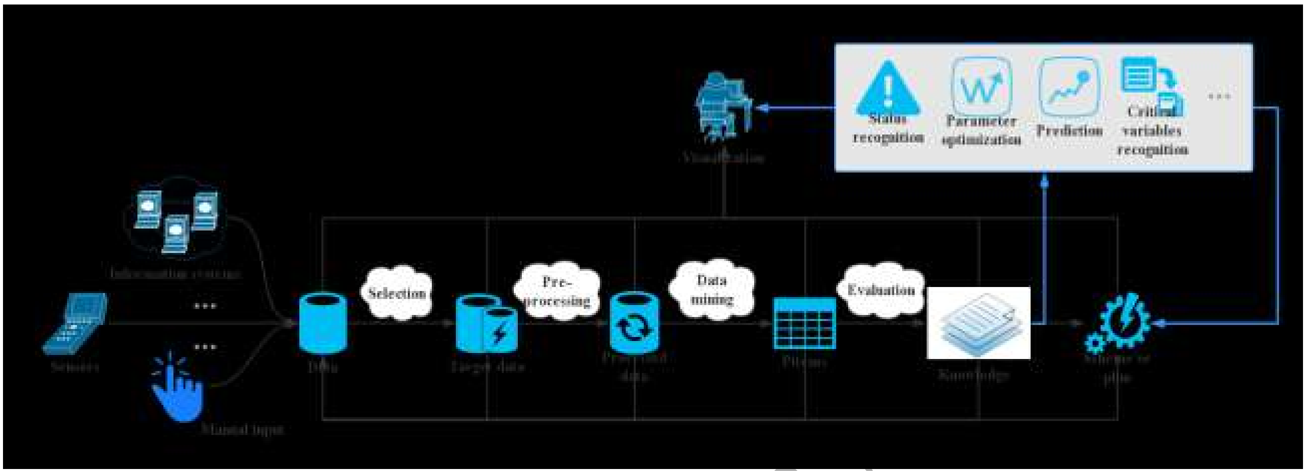


Fig. 5 The general data mining process in production management

Fig. 5 tells us the general process of data mining in production management. Considering the main steps in the general data mining process, this part of limitations and suggestions is divided into the following three parts to discuss.

4.2.1 The processing techniques required before data mining process

DMTs for complex data types. In the big data era, massive amounts of data contain a variety of types of data, while production management contains some special data types, including time series data, data streams, non-structural data, sparseness data, multi-scale data, and so on. The existing DMTs are difficult to deal with some of these data types directly. Although DMTs have been developed to deal with the random uncertainty of data or data ambiguity, there is almost no way to better consider both the random uncertainty and the ambiguity. Therefore, it is necessary to further study DMTs for complex data types. Since the manufacturing systems are essentially non-linear, the combination of appropriate kernel functions can provide the possibility for more DMTs to be applied to production management.

Data preprocessing technologies. Due to the complexity of the production processes and the impact of noise, the success of the data mining process depends in a large extent on the success of the data preprocessing. More data processing techniques that can effectively improve the incompletion and inconsistency of data and improve the signal-to-noise ratio of the signal should be developed and can even exploring the appropriate DMTs for data preprocessing. For outlier data, it may contain important exception information, we cannot simply clear it. Therefore, the development of successful filters is very important, which can not only reduce the noise, but also improve the abnormal information capture rate and reduce false positives. In the case of anomalies that may exist in the collected data, in order to solve the problem that the curse of dimensionality suffered by the classification-based method, the frequent pattern can be detected by the DMTs pattern mining function to detect the abnormality. For the problems that the extraction of feature patterns are limited to statistical variables, the future work can be considered to introduce the data mining frequent sub-sequence mining, symbolization and other technologies for more in-depth researches and applications.

4.2.2 The data mining process

High-performance integrated DMTs and parameter setting practices. If appropriate, the integrated or combined DMTs can achieve better performance than a single DMT and perform more complex tasks than components. But there still a worry about how to select components and set proper parameters to form high-performance integrated DMTs.

Many DMTs (such as SVM) are deeply dependent on the success of the setting of optimal value of the parameters and people are concerned about what are the key parameters obviously affecting production process. Practice has proved that the meta-heuristic algorithm has a good performance when selecting components of DMT. At the same time, the efficient meta-heuristic algorithm can also improve the rationality of DMT parameter setting, and solve the limitations in identifying parameters according to experience or a lot of experiments.

The scalability and generalization capabilities of DMTs. An algorithm is scalable, which means that if the available system resources such as memory and disk space are available, the run time should increase linearly with the database size. This is very important for the efficiency of the calculations, especially when there is a large number of data. Besides, many of the existing DMTs are only for a particular or a certain type of specific issues and improving the generalization capability means removing this limitation to be adopted to solve more problems.

Intelligence of data mining process. Intelligence here includes the intelligence for choosing effective DMTs and self-learning ability. With the advent of more and more improved and effective prediction techniques, a lot of time wasted in testing each of the predictive techniques to find the right solution for the best needs. Intelligence requires DMTs to automatically select the appropriate DMT based on the characteristics of the data and require seldom or a small amount of manual intervention in the process. In order to achieve this purpose, we need to make full use of background knowledge, or expert knowledge in related fields to guide the discovery process to help speed up data mining process to assess the degree of interest in the discovery model. Many DMTs in defect analysis, fault diagnosis and process detection applications are used to achieve the classification function. Although good classification accuracy has been achieved, they are not always applicable for new or different problem sets, and new feature descriptions must be added manually. In addition, study for identification of out of control signal source is also based on classification, that is, only to determine which set of variables led to an exception but cannot directly determine which variable led to an exception. So we should focus on developing DMTs that automatically extract and feedback new patterns of defects from other types of defects from time to time, and remove the limitation in recognition for the out of control signal**.**

4.2.3 The expression, evaluation and update of knowledge

Data mining quality evaluation and validation technology. The results of the mining may find thousands of patterns, some of which are wrong, or may not be of interest to a particular user. But now, there is no definite system or method to evaluate the issues including the extent of mining and how much are the benefits. And because data mining uses specific methods of analysis or logic to discover knowledge, the system may not be able to interactively identify discovered knowledge, making the discovered knowledge not universally adaptive so that cannot be a useful knowledge. Some DMTs (such as Gaussian regression) can determine the quality of their excavation and the uncertainty of the results. So it is important to develop more such DMTs and remove the limitation for interactively confirmation of the general adaptability of the discovered knowledge.

Knowledge update and maintenance. Due to the uncertainty of the manufacturing environment, the dynamic changes in the data often lead to the existence of instability in the rules and even make the previously discovered model no longer valid. In the meantime, the new data accumulation may lead to the failure of previously discovered knowledge. Therefore, it requires DMTs to have a good update and maintenance ability in real time. At the same time, constantly enriching the knowledge base through cases accumulation and learning to achieve self-optimization.

The expression and interpretation of knowledge. If the knowledge digged out by DMTs cannot use a simple and easy way to show and understand, that knowledge cannot be effectively evaluated and well applied. We need more concise rules. Therefore, the development of DMTs needs to be combined with graphics, natural language and visualization techniques. In addition, some of DMTs (such as decision tree) are expressed in a concise way for the knowledge they are excavated, so that they can be better understood and accepted by the people. Some DMTs are difficult to express knowledge which they extracted can be combined with these DMTs to solve the limitations of knowledge difficult to understood and to improve the interpretability of knowledge.

The interactivity and visualization. Data mining is dedicated to retrieving knowledge from a large number of data sets by using automated algorithms. However, due to the characteristics of the automation processes, current data mining method does not allow the user to visually understand, explore and optimize the data set and the calculation process. Because it is difficult to accurately foresee what can be found in the database, it needs improvement for visualization and more people's participation in data mining process to improve the efficiency of data mining and find the rules of interest as soon as possible.

All in all, based on the general data mining process shown in Fig. 5 and the above discussions, Table.4 shows the summary on the limitations and suggestions from the following five aspects, i.e., data, data mining, excavation quality evaluation, knowledge, and visualization.

Table 4 Summary on limitations and suggestions

Aspects	Limitations	Suggestions
Data	- Processing of dirty, abnormal and complex data. - The extraction of feature patterns is limited to statistical variables.	- Develop DMTs for complex data types. - Develop efficient data preprocessing technology.
Data mining	- Parameter setting is not reasonable. - Scalability and generalization capabilities of DMTs need to be improved. - Intelligence for choosing proper DMTs. - DMTs based on classification cannot intelligently update the number of categories. - The study of integrated DMT is still in its infancy.	- Improve the scalability and generalization capabilities of data mining. - Explore high-performance integrated DMTs and use efficient meta-heuristic algorithms for parameter setting practices. - Improve the intelligence of DMT.
Mining quality evaluation	- Data mining quality evaluation and limitations of validation technology.	- Establishment of data mining quality evaluation system, search DMTs can determine the quality of mining.
Knowledge	- Knowledge update and maintenance. - The expression and interpretation of knowledge.	- Improve the ability to update the knowledge and environmental adaptability. - Improve the interpretability of knowledge.
Visualization	- Improve the "human - system" interaction.	- Improve interaction and visualization of system.

5. Conclusions

In modern production management, data that characterizes the manufacturing processes is collected and stored in database. As a result, data mining tools can be used to automatically discover interesting and valuable knowledge and patterns in the manufacturing processes. Then these knowledge and models can be used to promote the entire manufacturing process in areas such as defect prevention and detection, quality improvement, reduced flow time, increased safety, and so on. The arrival of the big data era has brought about more challenges to DMTs and also promoted the development of DMTs. Due to the particularity of manufacturing data, there are special requirements for DMTs. In this paper, the main concern is the applications and challenges of DMTs in production management in the era of big data. Driven by this purpose, this paper investigates the applications of DMTs in production management, including production scheduling, quality improvement, defect analysis, fault diagnosis and so on. The conclusion is drawn through discussion that DMTs are more efficient, accurate and independent compared with the traditional production management methods. In addition, we summarize the studies on the current situation of DMTs in the field of production management. In view of the existing studies, this paper points out the limitations of the current research including the processing and mining of complex data, the intelligentization and scientificization of mining process, the evaluation of excavation quality, the expression and maintenance of knowledge, interaction, and so on. This paper also puts forward some suggestions on the future research directions of DMTs in production management corresponding to the concluded limitations.

References

[1] Manyika J., Chui M., Brown B., et al., Big Data: The Next Frontier For Innovation, Competition, And Productivity, Analytics, 2011.

[2] Fayyad U., Piatetskyshapiro G., Smyth P., Knowledge Discovery and Data Mining: Towards a Unifying Framework, Shapiro, 1996, pp. 82-88.

[3] Li L., Sun Z., Ni J., et al., Data-based scheduling framework and adaptive dispatching rule of complex manufacturing systems, Int. J. Adv. Manuf. Tech., 2013, 66(9) 1891-1905.

[4] J. Han, M. Kamber ,J. Pei ,et al., Data Mining: Concepts and Techniques, third ed., 2011.

[5] Inmon W. H., Building the Data Warehouse, third ed., 2002.

[6] Agrawal R., Imielinski T., Swami A.N. Mining association rules between sets of items in large databases, ACM SIGMOD International Conference on Management of Data, Washington D.C., 1993, 22 (2) 207-216.

[7] Agrawal R., Srikant R.. Mining sequential patterns, IEEE Computer Society, 1995.

[8] Tamura Y., Iizuka H., Yamamoto M., et al., Application of local clustering organization to reactive job-shop scheduling, Soft Comput., 2015, 19(4) 891-899.

[9] Heger J., Hildebrandt T., Scholz-Reiter B., Dispatching rule selection with Gaussian processes, Cent. Europ. J. Oper. Re., 2015, 23(1) 235-249..

[10] Lim J., Chae M. J., Yang Y., et al., Fast Scheduling of Semiconductor Manufacturing Facilities Using Case-Based Reasoning, IEEE Trans. Semiconduct. Manufact., 2016, 29(1) 22-32.

[11] Olafsson S., Li X., Learning effective new single machine dispatching rules from optimal scheduling data, Int. J. Prod. Econ., 2010, 128(1) 118-126.

[12] Shahzad A., Mebarki N., Data mining based job dispatching using hybrid simulation-optimization approach for shop scheduling problem, Eng. Appl. Artif. Intel., 2012, 25(6) 1173-1181.

[13] Gokhan Metan, Ihsan Sabuncuoglu, Henri Pierreval, Real time selection of scheduling rules and knowledge extraction via dynamically controlled data mining, Int. J. Prod. Res., 2010, 48 (23) 6909-6938.

[14] A. Azadeh, A. Negahban, M. Moghaddam, A hybrid computer simulation-artificial neural network algorithm for optimisation of dispatching rule selection in stochastic job shop scheduling problems, Int. J. Prod. Res., 2012, 50(2) 551-566.

[15] Celik N., Son Y. J., Sequential Monte Carlo-based fidelity selection in dynamic-data-driven adaptive multi-scale simulations (DDDAMmemetS), IEEE Simul. Conference, 2009, 2281-2293.

[16] Tao J., Wang K., Li B., et al., Hierarchical models for the spatial–temporal carbon nanotube height variations, Int. J. Prod. Res., 2016, 54(21) 1-20.

[17] Kaibo Wang, Fugee Tsung, Recursive parameter estimation for categorical process control, Int. J. Prod. Res., 2010, 48(5) 1381-1394.

[18] Shichang Du, Xufeng Yao, Delin Huang, Engineering model-based Bayesian monitoring of ramp-up phase of multistage manufacturing process, Int. J. Prod. Res., 2015, 53(15) 4594-4613.

[19] Ranjit M., Gazula H., Hsiang S. M., et al., Fault Detection Using Human–Machine Co-Construct Intelligence in Semiconductor Manufacturing Processes, IEEE Trans. Semiconduct. Manufact., 2015, 28(3) 297-305.

[20] Ding D., Pan Z., Cuiuri D., et al., Bead modelling and implementation of adaptive MAT path in wire and arc additive manufacturing, Robot. Com.-Int. Manuf., 2016, 39 32-42.

[21] Kumar S. P. L., Jerald J., Kumanan S., Feature-based modelling and process parameters selection in a CAPP system for prismatic micro parts, Int. J. Comput. Integ. M., 2014, 28(10) 1-17.

[22] Sahebjamnia N., Mahdavi I., Cho N., Designing a new model of distributed quality control for sub-assemble products based on the intelligent web information system, J. Intell. Manuf., 2010, 21(4) 511-523.

[23] Weiss S. M., Dhurandhar A., Baseman R. J., et al., Continuous prediction of manufacturing performance throughout the production lifecycle, J. Intell. Manuf., 2016, 27(4) 751-763.

[24] P. Podržaj, A. Čebular, The Application of LVQ Neural Network for Weld Strength Evaluation of RF-Welded Plastic Materials, IEEE/ASME Trans. Mechatron., 2016, 21(2) 1063-1071.

[25] Du S., Liu C., Xi L., A Selective Multiclass Support Vector Machine Ensemble Classifier for Engineering Surface Classification Using High Definition Metrology, J. Manuf. Sci. Eng., 2015, 137(1) 011003-1-15.

[26] Milo M. W., Roan M., Harris B., A new statistical approach to automated quality control in manufacturing processes, J. Manuf. Syst., 2015, 36 159-167.

[27] Rao P., Liu J., Roberson D., et al., Online Real-time Quality Monitoring in Additive Manufacturing Processes using Heterogeneous Sensors, J. Manuf. Sci. Eng., 2015.

[28] A. Azadeh, J. Seif, M. Sheikhalishahi, M. Yazdani, An integrated support vector regression–imperialist competitive algorithm for reliability estimation of a shearing machine, Int. J. Comput. Integ. M., 2016, 29(1) 16-24.

[29] S. Du, L. Xi, J. Yu, J. Sun, Online intelligent monitoring and diagnosis of aircraft horizontal stabilizer assemble processes, Int. J. Adv. Manuf. Tech., 2010, 50(1) 377-389.

[30] Yang W. A., Monitoring and diagnosing of mean shifts in multivariate manufacturing processes using two-level selective ensemble of learning vector quantization neural networks, J. Intell. Manuf., 2015, 26(4) 1-15.

[31] M. Salehi, A. Bahreininejad, I. Nakhai, On-line analysis of out-of-control signals in multivariate manufacturing processes using a hybrid learning-based model, Neurocomputing, 2011 , 74 (12–13) 2083-2095.

[32] Shichang Du, Jun Lv, Lifeng Xi, On-line classifying process mean shifts in multivariate control charts based on multiclass support vector machines, Int. J. Prod. Res., 2012, 50(22) 6288-6310.

[33] S.G. He, Z. He, G.A. Wang. Online monitoring and fault identification of mean shifts in bivariate processes using decision tree learning techniques, J. Intell. Manuf., 2013, 24(1) 25-34.

[34] Ahmadzadeh F., Lundberg J., Strömberg T., Multivariate process parameter change identification by neural network, Int. J. Adv. Manuf. Tech., 2013, 69(9) 2261-2268.

[35] Hu S., Zhao L., Yao Y., et al., A variance change point estimation method based on intelligent ensemble model for quality fluctuation analysis, Int. J. Prod. Res., 2016(19) 1-15.

[36] Xu A.M., Gao J.M., Chen K., Excavation of critical resource node for quality control of multi-variety mixed production shopfloor based on complex network property, J. Sci. Proc. Inst. Mech. Eng., B J. Eng. Manuf., 2016, 230(1).

[37] Diao G.Z., Zhao L.P., Yao Y.Y., A dynamic quality control approach by improving dominant factors based on improved principal component analysis, Int. J. Prod. Res., 2015, 53(14) 4287-4303.

[38] Yuan T., Kuo W., Bae S.J., Detection of Spatial Defect Patterns Generated in Semiconductor Fabrication Processes, IEEE Trans. Semiconduct. Manufact., 2011, 24(3) 392-403.

[39] Weimer D., Scholz-Reiter B., Shpitalni M., Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection, J. Sci. CIRP Ann.-Manuf. Techn., 2016, 65(1) 417-420.

[40] Chen Y.J., Fan C.Y., Chang K.H., Manufacturing intelligence for reducing false alarm of defect classification by integrating similarity matching approach in CMOS image sensor manufacturing, Comput. Ind. Eng. 2016;99 465–73.

[41] Hsu C.C., Chen M.S., Intelligent maintenance prediction system for LED wafer testing machine, J. Intell. Manuf., 2016, 27(2) 335-342.

[42] Wu H., Wang Y., Yu Z., In situ monitoring of FDM machine condition via acoustic emission, Int. J. Adv. Manuf. Tech., 2016, 84(5) 1483-1495.

[43] Ding Y.F., Sheng B.Y., Study on steam turbine fault diagnosis and maintenance service grid system, Proc. Inst. Mech. Eng., B J. Eng. Manuf., 2010, 224(3) 517-530.

[44] Kane P.V., Andhare A.B., Application of psychoacoustics for gear fault diagnosis using artificial neural network, J. Low Freq. Noise V. A., 2016, 35.

[45] M. Unal, Y. Sahin, M. Onat, M. Demetgul, H. Kucuk, Fault Diagnosis of Rolling Bearings Using Data Mining Techniques and Boosting, J. Dyn. Syst. Meas. Contr., 2017, 139 (2)

[46] Khazaee M., Banakar A., Ghobadian B., et al., Detection of inappropriate working conditions for the timing belt in internal-combustion engines using vibration signals and data mining, Proc. Inst. Mech. Eng. D J. Automob. Eng., 2016.

[47] Rashid M.M., Amar M., Gondal I., et al., A data mining approach for machine fault diagnosis based on associated frequency patterns, Appl. Intell., 2016, 45(3) 1-14.

[48] Li M., Yang F., Wan H., et al., Simulation-based experimental design and statistical modeling for lead time quotation, J. Manuf. Syst., 2015, 37 362-374.

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

[50] Benkedjouh T., Medjaher K., Zerhouni N., et al., Health assessment and life prediction of cutting tools based on support vector regression, J. Intell. Manuf., 2015, 26(2) 213-223.

[51] Zhang H., Modelling and prediction of tool wear using LS-SVM in milling operation, Int. J. Comput. Integ. M., 2016, 29(1) 76-91.
[52] Lihui Wu, Jie Zhang, Fuzzy neural network based yield prediction model for semiconductor manufacturing system, Int. J. Prod. Res., 2010, 48(11) 3225-3243.
[53] Lee H., Chang O.K., Ko H.H., et al., Yield Prediction Through the Event Sequence Analysis of the Die Attach Process, IEEE Trans. Semiconduct. Manufact., 2015, 28(4) 563-570.
[54] Israel Tirkel, Forecasting flow time in semiconductor manufacturing using knowledge discovery in databases, Int. J. Prod. Res., 2013, 51(18) 5536-5548.