JOSEIT JOURNAL OF SYSTEMS ENGINEERING AND INFORMATION TECHNOLOGY
http://jurnal.iaii.or.id/index.php/JOSEIT

Fault Detection of Mechanical Equipment Failure Detection Using Intelligent Data Analysis

Maksim Andreevich Kovito
Department of Informatics, Institute of Space and Information Technologies, Siberian Federal University, Krasnoyarsk, Russia

*krovik76@gmail.com

Abstract

Poor maintenance of machinery in manufacturing plants has always been an important link in the production process. In addition to computer technology, artificial intelligence technologies and various intelligent sensors are widely used in manufacturing industries. The amount of data generated by production machines and equipment at all stages of the production process is also growing rapidly, and it is particularly important to analyze the data generated by these devices in order to detect and even predict malfunctions. Intelligent data mining provides advanced data analysis techniques for this purpose. This article introduces the basic concepts of data mining, its processes, the main data mining technologies, and provides recommendations for applying data mining to detect failures in devices.

Keywords: Data Mining, Method, Mechanical Equipment, Sensor, System, Malfunction

1. Introduction

With the introduction of intelligent production, the industry is developing rapidly around the world [1], and today new intelligent, automatic, connected and digital production equipment are widely used. The operation of these devices is accompanied by the accumulation of large amounts of data [2]. They contain a lot of potentially valuable information. Obtaining data quickly and efficiently from the available volume and ensuring that device failures are detected in good time is an urgent task in today's reality. While the traditional method of manual data processing clearly cannot meet the needs, data mining technology can successfully cope with the task at hand.

This technology analyzes data generated by production plants, extracts the laws contained therein [3] and analyzes the operational status of production plants based on these laws in order to achieve the goal of cost savings, improved efficiency and reduced energy consumption of production [4]. In the presented article, an attempt is made to compile methods that can be used in the implementation of data mining technology in various areas of production.

Intelligent data analysis is the extraction of some data from a large database to identify implicit, previously unknown, real, novel and potential patterns. In addition, some sources refer to data mining as discovering new knowledge, deepening knowledge about certain processes, etc [5]. The concept of data mining (data analysis) was first mentioned in 1989 and was associated with the automation and optimization of queries when working with large databases. The term data analysis (data analysis), in turn, was used much earlier and meant the processing and interpretation of data collected as a result of experiments, mainly of a scientific nature. Over time, these concepts have expanded and generalized, coming very close together and are now closely related to big data and the concept of machine learning machine learning.

Intelligent data analysis is largely an applied theory, with the number of applications to real industrial problems growing every year. Currently, data mining methods and tools are used in web development, bioinformatics, computer vision systems, computer game development, marketing, medical diagnostics, optimization methods, search engine development, image, image, speech and signal recognition, etc. Data mining specialists are growing constantly, as well as the financing share for developments in this area. More and more data analysis software solutions are appearing More and more data analysis software solutions are appearing, including open code.

The notion of subject of intelligent data analysis in general is quite broad. It is predominantly a relational database, but it can also be an unstructured data store such as B. text masses (text unstructured or semi-structured data), multimedia data (including image, audio, video data), Internet data and complex types of data (mainly including distributed databases and temporal sequence databases), etc. The task of data mining of data is to Isolate patterns from datasets.

Regarding functional patterns, data analysis is mainly divided into descriptive analysis and predictive analysis. The focus of descriptive analysis is to find explicit patterns to describe the data and describe their

Accepted by editor: 25-08-2022 | Final revision: 15-09-2022 | Online publication: 25-09-2022

general properties. Predictive analysis is based on current and historical data, with an emphasis on building predictive models.

At present, the main tasks of data mining are usually divided into data description, model-based estimation, prediction, clustering, and correlation analysis according to the practical tasks. The data description is Exploratory Data Analysis (EDA), which complements the information currently available [6]. Model-based assessment consists of building a model of the object, process or phenomenon under study and estimating the difference between the target values of the indicators under study and the values predicted by the model based on the original data set. Usually, a simple linear regression or correlation is used, a neural network can also be used to estimate the model [7-9].

The classification is similar to the estimation model, except that the classification's target value is a category rather than a number. The basic methods include decision tree, neural network, logistic regression, Bayesian network, etc. and others. [10-13]. Forecasting is similar to classification and model assessment, but forecasting is mainly aimed at estimating the situation in the future. All data in the studied database is divided into several classes, objects of the same class have maximum similarity, different classes have minimum similarity. The most commonly used attributes to measure the similarity of objects are the distance between them, density, etc., usually the KNN (k-Nearest Neighbor) clustering method is used in a balanced way. Nearest Neighbor), balanced iterative reduction and clustering with (BIRCH) balanced iterative reduction and clustering using hierarchies. hierarchies) [14] etc.

Association analysis is all about finding correlations between data and determining what properties they have. Typically, the level of support and validity are used to measure the correlation of association rules. Common algorithms for finding associations are the a priori algorithm (Apriori) and the generalized rule induction algorithm [15-16].

2. Method

Intelligent data analysis consists of three stages: data pre-processing, intelligent analysis, and evaluation and presentation of the result analysis. Figure 1 shows the entire data mining process.

2.1 Data Pre-Processing

Data pre-processing is a very important part of the whole data mining process. When we integrate data from different sources, the data is often incomplete and contains a lot of noise and redundancy. Data quality has a direct impact on the reliability of the data mining model and the validity of the final decision. For the accuracy of the data mining results, the data must be pre-processed, which basically involves three steps: data cleaning, data transformation and data reduction, data dimensionality.

The content of data cleaning mainly includes dealing with missing data, identifying misclassifications and identifying outliers. In a data set with a large amount of data, missing information is quite common. Although data with missing information can be processed faster, simply skipping or discarding it can result in outliers and data loss. Some use constants or estimates based on other functions instead. Finding and correcting misclassifications can also improve the accuracy of the results. Outliers are extreme values that deviate from other trend indicators. Some data mining methods Data mining methods are severely affected by outliers, which can lead to errors in the results. Therefore, it is very important to identify outliers. The range of the different values is often very different. In some analysis algorithms, this difference in ranges causes the value with a larger range to negatively affect the result. Therefore, normalization data is essential in data mining.

Datasets used in data mining can contain millions of pieces of data and thousands of variables, not all of which are independent and uncorrelated. In data mining, it is necessary to prevent correlations between predictive variables that affect the stability of the results. Principal component analysis (PCA) [17] is a method of statistical analysis that selects a small set of important variables to describe an appropriate structure through a linear combination of multiple variables. The principal components method is only used for predicted variables and not for target variables.

2.2 Data Analysis

Data mining is the most important link in the whole data mining process, which typically involves four aspects:

a. The selection and application of appropriate data mining technology data mining technology
b. Calibrate model settings to optimize the result
c. The application of various technologies
d. In the fourth step, we get the selected analysis algorithm according to the goals of data mining and the structure of the model formed as a result of data analysis.

2.3 Evaluation and Presentation of Results

Evaluating the model requires a comprehensive review of the data mining process data analysis to determine which factors or tasks in a given can be neglected in a particular case. Model representation is a visualization of the model that allows it to be presented in a user-friendly, accessible users. This step is also important for providing the target user an easy-to-understand, visual pattern of the model.

3. Result and Discussion

Mechanical device fault detection helps to monitor, diagnose and predict the condition of these devices. This in turn ensures continuous and safe operation, which is of great importance. This, in turn, ensures uninterrupted, safe operations that are essential for efficient production and operations. For example, the mechanisms of large tunnel boring machines operating under complex conditions are prone to failure and their hydraulic system is one of the main causes of failure of the entire mechanism.

The factors that influence this process are very different. It is difficult for an operator to locate the cause and determine how to do it properly and efficiently. Due to the complexity of the hydraulic system of the road header hydraulic system, the traditional manual diagnostic method cannot meet the needs of efficient machine operation, so the intelligent analysis method should be applied in this situation. In the literature [18], a fault diagnosis method is proposed that combines the fuzzy theory with a neural network for the failure of the hydraulic system of the section shear.

VC++ and MATLAB are used as tools for implementing software for error diagnosis. The literature [29] describes the monitoring of the condition of the engine equipment based on the readings of 6 sensors. These readings are analyzed and fault diagnosis rules are created so that the system can monitor its technical condition in real time. One source [20] states that different error factors have different effects on the diagnostic process. This article proposes a weighted associative rule matrix algorithm that uses the weights of the error factors to improve the accuracy of error diagnosis. Increasing the row density of the coefficient decreases the generation scale of the candidate element sets, thereby increasing the efficiency of fault diagnosis.

The FDC method (Fault Detection and Classification) was suggested in the literature [21]. First, a principal components analysis was used to sort the training data and build a modular neural network for error identification. The D-S theory is used to classify failures in case of uncertainty (e.g. multiple failures). The FDC method, based on the modular neural network and D-S theory, has the ability to quickly identify anomalies and locate the main component of the anomaly when it occurs, so that the relevant personnel can quickly locate the faulty component and take appropriate measures can take to correct them.

An improved decision tree algorithm based on a set of variable values was proposed in the literature [22], which improved the classification accuracy and the ability to suppress noise data, and was applied to coal plant equipment and showed good results. The rotating equipment condition monitoring system described in [23] was developed based on large rotating elements, combined with the Apriori algorithm and the application of the B\S structure, has shown excellent results in practical production and application.

In the literature [24], the author proposed a visualization of the distribution of the prevalence matrix and the Riemann covariance to detect anomalies of mechanical equipment and applied it to detect gear failures of wind turbines and obtained good results. The black hole particle swarm optimization (BPS-SA) algorithm using the least squares method, the reference vector method, has been proposed in the literature [25] in combination with a glow simulation algorithm, which improved the classification speed and accuracy and reduced problems with extreme values. Table 1 shows the application of some data mining techniques to detect failures in mechanical devices.

Table 1. Application Of Data Mining Technologies for Mechanical Equipment Fault Detection

Type Of Mechanical Equipment	Data Analysis Technology
Hydraulic system of the road header road header [18]	Fuzzy set and BP neural network
Motor equipment [19]	Rough set
Hydraulic equipment for steelworks [20]	Matrix weighted rules associations
Semiconductor equipment [21]	Modular neural network and the D-S method
Machines and equipment for coal storage warehouses [22]	Rough sets and decision trees
Large rotating equipment [23]	A priori algorithm
Turbine gearbox [24]	Prevalence matrix and Riemann distribution

	covariance
Fan reducer [25]	SVM and annealing simulation algorithm

The table shows that data mining technology is currently being applied to many aspects of mechanical device failure detection. At the same time, practical application usually requires a combination of different data mining techniques to achieve the expected effect, and targeted improvements to the underlying data mining technology must be made.

At the 2019 KDD conference. [26] proposed a mobile event prediction system integrating spatiotemporal data. This framework can be generalized well and can easily be transferred to other domains. If mechanical failures can be detected and the failure can be predicted in advance, it will greatly reduce the loss of businesses. Google [27] combines features of available data to make parking predictions more efficient. Using different dates, combining different users and different seasons makes parking forecasts on Google map more productive. In the future, it will likely be possible to more effectively detect and predict mechanical device failures by combining personal information from device users and other climate and weather factors.

However, the application of data mining technology to detect failures in mechanical devices still faces many challenges. In terms of technology, the results of the study contained in this paper mainly come from the experimental phases, and the data used in the experiments are mostly manually verified and analyzed data of high quality. However, in a real production environment, there are many issues with the quality of the data. In addition, a large number of studies are based on existing data. Analysis of standard data can lead to fairly precise rules for device operation, but the amount of data must be constantly accumulated and replenished, which requires obtaining data in real time to reflect the latest changes in the state of devices is a fairly non-trivial task. At the same time, when using the model and algorithm described above in a real situation, many factors such as the versatility, flexibility and reliability of the model, as well as the real-world performance of the algorithm, should be considered.

4. Conclusion

The article provides an overview of relevant technologies and theories in the data mining process, outlines the application of data mining in mechanical device failure detection, and points out its shortcomings and development trends. In short, with the rapid development of all walks of life, the volume of data is increasing, and data mining naturally becomes a powerful decision support tool. Obviously, modern industry can no longer do without the use of data mining.

References

[1] Tao F., Qi Q. New IT driven service-oriented smart manufacturing: framework and characteristics. IEEE Transactions on Systems, Man and Cybernetics; 2017. http://dx.doi.org/10.1109/TSMC.2017.2723764
[2] Yin S., Kaynak O. Big data for modern industry: challenges and trends [point of view]. Proc IEEE. 2015; 103(2): 143-146.
[3] Harding J. et al. Data Mining in Manufacturing: A Review. VDM Verlag; 2006: 969-976.
[4] Wang K. Tong et al. Review on Application of Data Mining in Product Design and Manufacturing. International Conference on Fuzzy Systems and Knowledge Discovery; 2007: 613-618.
[5] Hand D. J. Data Mining. Encyclopedia of Environmetrics. 2006; 2.
[6] Yu C. H. Exploratory data analysis. Methods. 1977; 2: 131-160.
[7] Seber G. A. F., Lee A. J. Linear regression analysis: John Wiley & Sons; 2012.
[8] Aiken L.S., West S.G., Reno R.R. Multiple regression: Testing and interpreting interactions: Sage; 1991.
[9] Hansen L.K., Salamon P. Neural network ensembles. IEEE Transactions on Pattern Analysis & Machine Intelligence. 1990; (10): 993-1001.
[10] Keller J.M., Gray M.R., Givens J.A. A fuzzy k-nearest neighbor algorithm. IEEE Transactions on Systems, Man and Cybernetics. 1985; 15(4): 580-585.
[11] Hosmer Jr.D.W., Lemeshow S., Sturdivant R.X. Applied logistic regression: John Wiley & Sons; 2013. Информатика. Экономика. Управление/Informatics, Economics, Management 2022; 1(2) http://oajiem.com/0132
[12] Munro D.D. Forest growth models-a prognosis. Growth models for tree and stand simulation. Research Note 30. Department of Forest Yield Research. Royal College of Forestry: Stockholm. 1974; 30: 7-21.
[13] Kass R. E., Raftery A. E. Bayes factors. Journal of the American statistical association. 1995; 90(430): 773-795.
[14] Zhang T., Ramakrishnan R., Livny M. BIRCH: A new data clustering algorithm and its applications. Data Mining and Knowledge Discovery. 1997; 1(2): 141-182.
[15] Inokuchi A., Washio T., Motoda H. An apriory-based algorithm for mining frequent substructures from graph data. European conference on principles of data mining and knowledge discovery: Springer, Berlin, Heidelberg; 2000: 13-23.
[16] Freitas A.A. A genetic programming framework for two data mining tasks: classification and generalized rule induction. Genetic Programming 1997: Proc 2nd Annual Conf. Morgan Kaufmann; 1997: 96-101.

[17] Zheng X., Levine D., Shen J. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012; 28(24): 3326-3328.
[18] Hongjin Z. Fault diagnosis of hydraulic System of road header based on Fuzzy Neural network. Nanjing University of Science and Technology. 2013.
[19] Xin Xuming. Data Mining System based on process Monitoring and its application. Shanghai: Donghua University. 2004.
[20] JiaoPeng Sha. Research on intelligent device Fault diagnosis based on matrix weighted association rules. Hebei: Yanshan University. 2012.
[21] Hong S. Lim W.Y., Cheong T. et al. Fault Detection and Classification in Plasma Etch Equipment for Semiconductor Manufacturing e-Diagnostics. IEEE Transactions on Semiconductor Manufacturing. 2012; 25(1):83-93.
[22] Wang. Application of Decision Tree algorithm in mechanical equipment fault diagnosis system. Wuhan: Huazhong University of Science and Technology. 2013.
[23] XueYong Shu. Application of vibration fault diagnosis of rotating equipment based on data Mining. Jilin: Jilin University. 2013.
[24] Wang S., Sun X., Li C. Wind turbine gearbox fault diagnosis method based on Riemannian manifold. Mathematical Problems in Engineering. 2014; 16(4): 835-892.
[25] Jiao B., Xu Zh. Parameters optimization of LSSVM and application in fault diagnosis of wind power gearbox. Control Engineering of China. 2012; 19(4): 681-686. Информатика. Экономика. Управление//Informatics, Economics, Management 2022; 1(2) http://oajiem.com/0133
[26] Shen B., Liang X., Ouyang Y. et al. Stepdeep: a novel spatial-temporal mobility event prediction framework based on deep neural network. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM; 2018: 724-733.
[27]Arora N., Cook J., Kumar R. et al. Hard to Park: Estimating Parking Difficulty at Scale. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM; 2019: 2296-2304