ARTICLE IN PRESS

Mechanical Systems and Signal Processing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Data mining technology for mechanical engineering computer test system

Zhenjun Li [a], Xiaomo Yu [b,*]

[a] College of Computer Science & Software Engineering, Shenzhen University, Shenzhen 518060, Guangdong, China
[b] Department of Logistics Management and Engineering, Nanning Normal University, Nanning 530001, Guangxi, China

ARTICLE INFO

Article history:
Received 27 October 2019
Received in revised form 20 December 2019
Accepted 7 January 2020
Available online xxxx

Keywords:
Mechanical engineering
Test system
Apriori algorithm
Signal de-drying
Data mining technology

ABSTRACT

With the continuous development of information technology, data resources have become increasingly rich, but the knowledge contained in data resources has not been fully explored and utilized. In order to find an effective computer testing technology, the paper introduces the partial differential equation (PDE) into the denoising process of rotor vibration signal through data mining technology, and generalizes the unified model of PDE filtering. Several filtering methods are compared through simulation experiments. The effect is that the flexible rotor is balanced by different dynamic balancing methods, and satisfactory results are obtained. From the simulation results, it can be concluded that the integration method is not suitable to extract the unbalanced signal with strong noise background, but it provides a way to calculate the amplitude and phase of sinusoidal signal without noise; The processing is simple and suitable for the calculation of the dynamic balance test system with fewer sampling points; both the DFT method and the FFT method use the principle of Fourier transform spectrum analysis, but the FFT method calculates the speed much faster than the DFT method. Experiments on the classification of fault data prove that the improved Apriori algorithm is greatly improved compared with the original Apriori algorithm, and the speed of acquiring fault rules is improved.

1. Introduction

Data mining is the process of revealing hidden, previously unknown, potentially useful information from a large amount of data in a data warehouse. Data mining is becoming a new and increasingly popular research area. Using a database to store data, a machine learning method to analyze data, and mining the knowledge behind a large amount of data, the combination of the two has led to the emergence of data mining. Data mining is an indispensable part of knowledge discovery in databases. In fact, data mining is a cross-disciplinary subject involving databases, statistics, artificial intelligence and machine learning [1,2].

In recent years, with the introduction of data mining technology, new ideas have been provided for extracting useful information from these massive data. In [3], the author used data mining techniques combined with logistic regression, decision tree graphs, neural network models, and partial least squares models to determine the impact of word exposure frequency on online news based on the usability heuristic concept. In [4], the authors analyzed the acupoint selection rules of diabetic peripheral neuropathy (DPN) based on data mining technology. The results showed that the acupuncture DPN

* Corresponding author.
E-mail addresses: lizhenj@szu.edu.cn (Z. Li), yuxiaomo@nnnu.edu.cn (X. Yu).

https://doi.org/10.1016/j.ymssp.2020.106628
0888-3270/

was mainly based on the enhancement method, which promoted the circulation of blood and blood; the acupoint selection was mainly based on Yangming meridian and back-shu points. Data mining technology is a very common computer technology that has been widely used in many fields due to its superior performance. In [5], the author studied the data management technology based data management method for wireless sensor network. The author also proposed a cluster-based replication record deletion algorithm, and finally verified the accuracy of the data cleaning method. The results showed that the research method in this paper is correct and effective. In [6], in order to find out the main factors affecting the wearing comfort, and how they affect the wearing comfort of the clothing. The wearing comfort of the trousers was mainly affected by four factors: waist and hip factors, knee tibia factors, c factors and thigh calf factors. Their contributions accounted for 39.17%, 16.4%, 13.96% and 6.95%, respectively. In [7], the authors described the performance of the diesel engine using the relationship between the diesel engine load data and the fuel consumption rate at different speeds. Data mining is the process of extracting important and useful information from large amounts of data [8,9]. In [10], the author applied data mining to medicine. In the analysis of a single cancer, it was found that patients treated with angiotensin receptor blockers (ARB) had a significantly reduced risk of breast cancer, while the pancreas cancer and prostate cancer are at increased risk. In [11], the author studied the formal framework of a data mining query language for mobility data and its related implementations, illustrating its analytical capabilities in revealing the complexity of urban mobility in large urban areas. In [12], the authors proposed a data mining-based intelligent differential protection scheme for microgrids. The proposed intelligent differential relay scheme can provide effective protection measures for the safe operation of the microgrid with high reliability. In [13], the authors used data mining techniques to examine different formats of review data to predict student performance. The research results show that the method proposed by the author improves the predictive performance of final student performance. In [14], the authors proposed an enhanced cascading fault model combining data mining techniques. The article proves the effectiveness and effectiveness of the proposed enhanced cascading fault model. In [15], the author used a real-time digital simulator to develop a WAMS network physical test bench, and introduced the test bench application, simulated network capability scenarios, data set development process, and selected results.

Over the years, people have been working on the versatility and scalability of computer test systems to improve the utilization and execution efficiency of mechanical engineering automated test equipment. In [16], the author applied computer testing techniques to image recognition. In order to find an isolation scheme that can automatically find the cause of the diagnosis of damage and economically related insects, the author developed a fruit fly image recognition system called AFIS1.0. The system used Gabor surface features in automatic recognition, and the overall classification success rate obtained by independent multi-part image automatic recognition test reaches 87% of the species level. In [17], the author applied computer test technology to error measurement, and proposed a method to test the influence of computer generated hologram (CGH) manufacturing error in cylindrical interferometry system. In [18], the authors proposed a computer-aided test method based on inverse Hartmann test for high-precision testing of reflective surfaces. The author also proposed an iterative ray tracing calibration method based on computer-assisted reverse optimization, which was used for high-precision testing of reflective surfaces. In [19], the authors proposed a simulation test system for a non-contact D-point transformer for voltage measurement. The analog test system consists of a D-point transformer, a signal processing circuit and a grounded PC port. Finally, a test platform was built to simulate the performance of the entire single-phase transformer test system [20]. In [21], the authors designed a computer-based breath sound analysis system to identify normal lung sounds in children and validated the effectiveness of a computer-based breath sound analysis system. This computer-based breath sound analysis system performed 58 lung sound recognitions, correctly identified 52 times, and misidentified 6 times. The accuracy is 89.7%. In [22], the authors developed a computer adaptive test (CAT) input technique to measure the entire potential anxiety continuity. The CAT simulation showed efficient and highly accurate measurement results. In [23], the authors evaluated the parameters of the severity of traffic accidents, including speed change (delta-v) and energy equivalent speed (EES). In [24], the author proposed an effective test generation procedure using a method called "diagnosis using parallel models". In [25], the author further analyzed high-quality and appropriate-quality research in terms of content and specific teaching practices. The article discusses computer aided instruction (CAI) suggestions for future research and its implications for practice. In [26], the author systematically developed the visual literacy assessment test (VLAT), and conducted a necessity assessment on the test items in VLAT (average content effective ratio = 0.66).

Through the analysis of literature review at home and abroad, it can be seen that although many research teams at home and abroad have carried out more research and Analysis on computer technology for mechanical engineering, most of them are limited to the internal principles of mechanical engineering, but not to the technical level [27]. As a result, there is not enough powerful computer technology to carry out in-depth analysis of mechanical engineering. As an advanced computer technology, data mining technology is gradually concerned by people because of its unique algorithm settings and high recognition accuracy [28], but there is still no team that can effectively combine the principles of mechanical engineering and data mining technology. The purpose of this paper is to combine data mining technology and mechanical engineering principle effectively, which can improve the accuracy of algorithm recognition, optimize algorithm model and reduce experimental error.

The first part of the article is an introduction, introducing the overall structure of the article and the content of each part, and a review of the relevant literature on data mining technology and computer testing systems; the second part is the algorithm model established by the article; the third part is the experiment mainly data sources and experimental settings; the fourth part is the processing and analysis of experimental results; the fifth part is the conclusion.

Please cite this article as: Z. Li and X. Yu, Data mining technology for mechanical engineering computer test system, Mechanical Systems and Signal Processing, https://doi.org/10.1016/j.ymssp.2020.106628

2. Method

2.1. Theoretical basis of data mining technology

2.1.1. Components

The process of data mining system work is first composed of database, data warehouse or other database. The preprocessing of these targets, ie data cleaning and data integration, is also called data filtering. When it comes to data mining engine, it is a mining module with five functions of evolutionary calculation, deviation analysis, correlation analysis, classification induction and qualitative induction. It is indispensable in data mining system. Finally, the model evaluation module, in this step, can combine the two methods of algorithm and knowledge evaluation, which can achieve good results when mining more helpful pattern knowledge. You can then communicate with the user on a visual user interface. Users get a lot of appeals through this platform, such as obtaining the required information; consulting the definition pattern; analyzing and evaluating the data content.

2.1.2. System function

The process of concept description is to select the target and compare the two data sets, and distinguish them by the comparison description factor. This requires a precise description of the data set and an overview of its significance. Credibility, level of interest, and level of support are three metrics for association rule analysis. Apriori and FP-growth are two algorithms that are often used by association rules. They have their own advantages. The Apriori algorithm has good scalability, and the FP-growth algorithm uses a tree structure. The characteristics of the Apriori algorithm determine its application in the field of parallel computing, and the advantage of the FP-growth algorithm is that the efficiency of directly obtaining frequent sets is naturally higher than that of the Apriori algorithm. The purpose of the classification is to predict the results more accurately. It is to classify the data in the database and then infer the attributes that will appear based on these attributes.

2.1.3. Excavation process

Data cleaning: Repairing problematic data, such as eliminating abnormalities and filling in defect data, making data consistent is the job of data cleaning. Otherwise these problems will cause data mining to think in the wrong direction. Data integration: Identifying data sets with the same conceptual attributes from large data sources and unifying them into one data set is data integration. This eliminates redundancy and inconsistency, which helps to improve data mining speed and save time. And data integration is repeated. Data conversion: Data conversion is to constrain data with appropriate interval values. For different attributes, the attribute value range is different. For example, the size of the bacteria and the distance between the planets are not an order of magnitude. Generally, the three normalization methods of maximum and minimum gauge, ten base, and zero mean are used more. Data reduction: Under the premise of not affecting the accuracy of mining results, reducing the behavior of data mining tasks for the purpose of improving efficiency is data reduction. Dimensions, aggregations, and modules are all objects that can be cut. Fig. 1 is an intrusion detection process based on data mining technology.

2.2. Mechanical signal processing algorithm model based on dynamic balance test system

The signal useful in the unbalanced vibration signal from the sensor is the fundamental frequency signal [29]. In order to facilitate the theoretical analysis of the following extraction algorithm, the fundamental frequency signal is represented as a sinusoidal signal. The expression looks like this:

$$x(t) = A\sin(\omega t + \varphi) \tag{1}$$

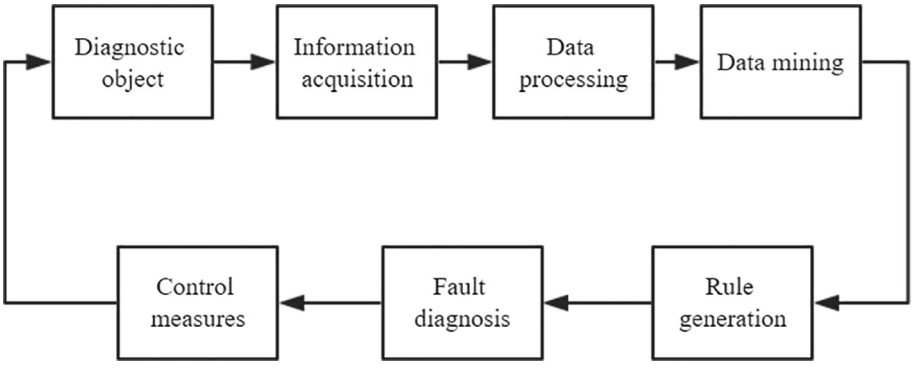


Fig. 1. Intrusion detection process based on data mining technology.

Please cite this article as: Z. Li and X. Yu, Data mining technology for mechanical engineering computer test system, Mechanical Systems and Signal Processing, https://doi.org/10.1016/j.ymssp.2020.106628

wherein A of the above formula represents the amplitude of the vibration signal; ω represents the angular frequency; φ represents the phase angle, and the purpose of the extraction algorithm is to accurately calculate the values of the amplitude A and the phase angle φ.

2.2.1. Integration method

The principle of the integral method is to integrate in two different intervals. For the two equations of 0–T/2 and T/4–3T/4 (T = 2π/ω), the following two can be obtained:

$$a_0 = \int_0^{\frac{T}{2}} A\sin(\omega t + \varphi)dt = \frac{AT\cos\varphi}{\pi}, \quad a_{90} = \int_{\frac{T}{4}}^{\frac{3T}{4}} A\sin(\omega t + \varphi)dt = -\frac{AT\sin\varphi}{\pi} \tag{2}$$

The above equation contains the amplitude and phase information of the unbalanced vibration signal, which can be obtained by solving two equations.

2.2.2. Relevant law

In order to obtain the amplitude A and the phase φ of the vibration signal, the x(t) of the Eq. (1) is cross-correlated with the standard sine signal and the cosine signal, respectively. Suppose the expressions of a standard sine wave and cosine wave in the range [0, T] are:

$$y(t) = \sin\omega t, t \in [0, T], \quad z(t) = \cos\omega t, t \in [0, T] \tag{3}$$

The above correlation analysis principle is explained by continuous signals, but in practice the computer can only process digital signals. Let the continuous signals x(t), y(t), and z(t) be sampled and become discrete data sequences:

$$x(i) = A\sin(\omega i + \varphi), \quad y(i) = \sin\omega i, \quad z(i) = \cos\omega i \tag{4}$$

2.2.3. Discrete Fourier transform (DFT)

By sampling the continuous signal x(t) in Eq. (1) with a sampling frequency fs and obtaining $\Delta = 1/f_s$, the discrete signal x (nΔ) is obtained, and x(nΔ) is equally spaced in one cycle to obtain an N-point time domain sequence. The discrete Fourier transform (DFT) definition of the N-point sequence can be expressed as:

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{nk} = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2k\pi n}{N}} \tag{5}$$

Let x(t) in Eq. (1) be a periodic continuous time function with period T, then the periodic signal can be expanded by Fourier series:

$$x(t) = \sum_{k=-\infty}^{\infty} X(2\pi k f_0)e^{j2\pi k f_0 t} \quad (k = 0, 1, \cdots, N-1) \tag{6}$$

Let t = nT_s (T_s is the sampling interval), then:

$$X(2\pi k f_0) = \frac{T_s}{T}\sum_{n=0}^{N-1} x(nT_s)e^{-j2\pi n k_0} = \frac{1}{N}\sum_{n=0}^{N-1} x(nT_s)e^{-j2\pi n k T_s} \tag{7}$$

Comparing the above formula, it can be seen that the spectral magnitude of the complex exponential form obtained by the Fourier series of the periodic signal is the corresponding spectral amplitude of the DFT transform multiplied by 1/N. The spectral amplitude of the trigonometric form of the periodic signal Fourier series is twice the amplitude of the spectral harmonic of the complex exponential form. Therefore, the spectral magnitude of the trigonometric form is the amplitude of the spectrum obtained by the DFT multiplied by 2/N.

2.2.4. Fast operation of discrete Fourier transform (FFT)

In fact, the N point X(k) is obtained by the above formula operation, and the N2 complex multiplication is required. The N (N − 1) complex addition operation includes a large number of repetition operations. Let N = 2 M, the FFT algorithm can decompose the DFT calculation of the N point into M-level iterative calculations, each consisting of N/2 butterfly elements, including a complex multiplication operation and two complexes in each dish calculation. When using the FFT algorithm, the total computational amount of the completed N-point DFT is $N/2\log_2 N$ complex multiplication and $N\log_2 N$ complex addition. The amount of calculation of the FFT is greatly reduced compared to the direct calculation of the DFT, especially when the number of transformation points N is large.

2.3. Test signal noise reduction and fundamental frequency extraction

2.3.1. Partial differential equation (PDE) denoising

The most common function in PDE is the heat transfer function, which has the following form:

Please cite this article as: Z. Li and X. Yu, Data mining technology for mechanical engineering computer test system, Mechanical Systems and Signal Processing, https://doi.org/10.1016/j.ymssp.2020.106628

$$\begin{cases} \frac{\partial u}{\partial t} - a^2 \frac{\partial^2 u}{\partial x^2} = f(x,t) \quad x \in R, t > 0 \\ u(x,0) = \varphi(x) \quad x \in R \end{cases} \tag{8}$$

where a is a normal number. Replace t with τ (evolution time), replace x with t (time variable of one-dimensional signal), and let f(x,t) = 0, the above equation can be expressed as:

$$\begin{cases} \frac{\partial u}{\partial \tau} - a^2 \frac{\partial^2 u}{\partial t^2} = f(x,t) \quad t \in R, \tau > 0 \\ u(t,0) = \varphi(t) \quad t \in R \end{cases}, \quad K(f,t) = e^{-a^2 f^2 t} \tag{9}$$

In the above formula, φ(t) represents the original signal, and u(t, τ) represents the filtered output signal. It can be seen that the Gaussian filter is determined by two variables a and τ. The amplitude-frequency characteristic of the filter becomes flat as a is small. When a = 0, the filter at this time is an all pass band filter. On the other hand, as the evolution time τ increases, the filter decays faster. When a is set to a constant, the filter is completely determined by the evolution time τ. Let fn be the cutoff frequency of the PDE filter, then the attenuation of the filter energy can be expressed as: $20 * \log\left(\frac{1}{e^{-a^2 f_n^2 \tau}}\right) = Y, \tau = \frac{0.115Y}{a^2 f_n^2}$, Y is the attenuation index of the filter. Given a cutoff frequency fn, the evolution time τ of the filter can be calculated. It shows that as the cutoff frequency increases, the evolution time gradually decreases, and vice versa. When the cutoff frequency is infinite, the input signal is equivalent to passing an all-pass filter, and the evolution time is reduced to zero.

2.3.2. Extraction of fundamental frequency amplitude and phase

The actual vibration signal collected is:

$$x(n) = A\sin(2\pi f n + \beta) + \sum_{i=1}^{m} \alpha_i \sin(2\pi \lambda_i n + \theta_i) + s(n) \tag{10}$$

where A is the amplitude of the fundamental frequency vibration signal, f is the frequency of the fundamental frequency signal, β is the phase of the fundamental frequency signal, α_i is the amplitude of the other frequency vibration signal, λ_i is the frequency of the corresponding signal, and θ_i is the phase of the corresponding signal, s(n) is a noise signal. According to the nature of the correlation function, the cross-correlation function of z(n) and v(n) with other non-co-frequency components has a value of 0, and the correlation value with the noise component also decays to 0 with the increase of n, then:

$$A = 2\sqrt{R_{xz}(r) + R_{xv}(r)}, \beta = \tan^{-1}\frac{R_{xz}(r)}{R_{xv}(r)}, \beta \in [0, 2\pi] \tag{11}$$

The equal-precision measurement sequence $x_i (i = 1, 2, \ldots, n)$ has a sample mean of $\bar{x}$, a deviation of $v_i = x_i - \bar{x}$, and a standard deviation of σ. When the deviation is greater than three standard deviations, the value at this time is considered to be a bad value and should be discarded. This principle is the 3σ criterion, namely: $|v_b| = \left|x_b - \bar{x}\right| > 3\sigma$. Where xb is the measured value that should be discarded, ie the bad value. The $x_b (1 \leqslant b \leqslant n)$ measurement sequence discards the bad value to form a new measurement sequence $x_k (k = 1, 2, \ldots, m), m \leqslant n$. The bad value is removed by using 3σ again for $x_k (k = 1, 2, \ldots, m)$ until no more bad values appear, and the deviation of the measured values at this time is within the range of 3σ.

2.4. Data classification algorithm model based on data mining technology

2.4.1. Improved decision tree - ID3 algorithm

Let S be a collection of s data samples. Suppose one of the classification attributes has k different values, and the different values in the classification attribute are divided into k different classifications C_i (i = 1, 2, …, k). Let s_i be the number of samples in the category C_i. You can find the expected information for any given sample:

$I(S_1, S_2, \ldots, S_k) = \sum_{i=1}^{k} -P_i \log_2(P_i)$ ($P_i = |C_i||S|$ is the probability that the sample belongs to the i-th class)

Let the attribute A have m different values $\{x_1, x_2, \ldots, x_m\}$. The attribute A can be used to divide S into m subsets $\{S_1, S_2, \ldots, S_m\}$. The information contained in the above S_j also has the same information in the attribute A, and the number of samples in which the S_j belongs to the i-th is S_{ij}, the entropy or information expectation of the result of partitioning by attribute A is:

$$E(A) = \sum_{j=1}^{m} \frac{|S_{1j}| + |S_{2j}| + |S_{3j}| + \ldots + |S_{mj}|}{|S|} I(S_{1j}, S_{2j}, S_{3j}, \ldots, S_n) \tag{12}$$

The core of the ID3 algorithm: Select attributes on each node in the decision tree to use information gain as a variable for partitioning attributes. In this process, the information needed for accurate classification should be reduced as much as possible. When all attributes are tested, the information gain is extracted from all the attributes, and then the attribute branches

Please cite this article as: Z. Li and X. Yu, Data mining technology for mechanical engineering computer test system, Mechanical Systems and Signal Processing, https://doi.org/10.1016/j.ymssp.2020.106628

of the decision tree are constructed according to different values, and then the subordinate sets of each branch are repeatedly judged to generate new decisions. The tree branches until the data contained in all subordinate collections belongs to the same class. This operation is terminated.

2.4.2. Apriori algorithm

First, the database is scanned to generate a "0" and "1" matrix R corresponding to the database. To facilitate subsequent calculations, the matrix behavior transaction is listed as an item set. Converts the transaction database DB containing m items, containing n items, into the transaction matrix R_{m^*n}.

$$\int : DB \to R, R = f(DB) = (d_{ij})_{m^*n}, \quad d_{ij} = \begin{cases} 1(I_j \in W_i) \\ 0(I_j \notin W_i) \end{cases}, \quad i = 1, 2, \ldots, m \quad j = 1, 2, \ldots, n \tag{13}$$

The row of things Rm*n represents the thing, the column represents the item. When the item Wi contains the item Ij, the corresponding dij = 1 in the matrix, otherwise dij = 0. Since the matrix is to be simplified, a column is added after the matrix to calculate the row total (sum_hang), that is, the number of transaction items; a row is added below the matrix to calculate the column total (lie_hang), that is, the project support number.

The structure D1 is structurally reduced to generate a new frequent k-term set. When a frequent k-term set is generated, the matrix column corresponding to the item whose number of repetitions in the frequent (k-1)-term set is less than k is removed from the matrix. If the total number of transactions is less than k, delete the row corresponding to the transaction in the matrix. If a column total is less than min_sup, the column corresponding to the element in the matrix is deleted; the simplification matrix structure is repeated according to the above steps, until the matrix Dk is obtained.

3. Experiment

3.1. Data source

The sensor outputs an analog electrical signal. The computer cannot directly process it. The continuous analog signal needs to be converted into a discrete digital signal through the data acquisition card [30]. The main performance indicators of the acquisition card include resolution, maximum sampling frequency, number of input channels, range, etc. This system uses NI 9234 capture card. The NI USB-9234 capture card has 24-bit resolution with a range of ±5 V. The four channels are simultaneously sampled. The maximum sampling frequency per channel is 51.2 k S/s. Four-channel simultaneous sampling plays a vital role in this system. It ensures the synchronism of the signals of each channel and ensures accuracy in the subsequent signal analysis.

KDDCUP99 data set:

The data set was developed based on the DARPA computer test evaluation data set of the Massachusetts Institute of Technology Lincoln Laboratory. The data is mainly from the network data generated within 9 weeks of the simulated military network, including data for testing and training. There are a total of more than 7 million network connection records, and there are dozens of frequently seen network intrusion types. There are four types of denial-of-service attacks, remote user attacks, privilege attacks, and port attacks. Each record has 42 attributes, one is a class category attribute, and the other 41 are basic feature attributes, which are divided into 4 categories.

3.2. Experimental environment and test system components

This test uses five Lenovo desktops and one notebook to build an analog small LAN, including the external network, intranet and intrusion detection analyzer. Four of the desktop units are combined into an intranet, and one desktop is used as an external network (used as a simulated attack experiment on the internal network), and that notebook is used as an analysis processor for the intrusion detection system. Equipment: 5 Lenovo desktops (P4 3.0G, 512 M RAM, 40 GB hard drive), 1 lenovo laptop (Corei3 M330 2.13 GHz, 2G RAM, 500 GB hard drive); router one: TP-LINK TL-R406; switch one Department: TP-LINK TL-SF1016S. Operating system: Windows 7 operating system. Development Tools: Visual Studio 2008.

The test system consists of the conversion, transmission, analysis, storage, and output devices of the sensor and acquisition data. There are 2 speed sensors, 5 pressure sensors and 4 flow sensors (located on DC motor, auxiliary drive motor and secondary components), 4 tension sensors and 2 torque sensors (including speed signal detection). 2 temperature sensors 4 displacement sensors for data acquisition. There is also a compactness meter that measures soil compaction and a five-wheel gauge that measures the speed of trolley travel. The test system also features a data acquisition processor (24 channels) and its corresponding processing software.

3.3. Evaluation indicators

(1) Experimental simulation was performed in a MATLAB environment. In the actual test environment, the signal picked up by the sensor is processed by filtering and other circuits to obtain the vibration signal of the rotor. The composition of the vibration signal is relatively complex, and there are DC components, fundamental frequency components of the

Please cite this article as: Z. Li and X. Yu, Data mining technology for mechanical engineering computer test system, Mechanical Systems and Signal Processing, https://doi.org/10.1016/j.ymssp.2020.106628

same frequency as the rotor, sub-frequency components, and times. Components such as frequency components and random noise of a certain bandwidth. In order to simulate the actual vibration signal more accurately, it is assumed that the general expression of the vibration signal is:

$$y(t) = A_0 + A_1\sin(\omega t + \varphi_1) + A_2\sin(2\omega t + \varphi_2) + A_3\sin(3\omega t + \varphi_3) + n(t) \quad (14)$$

In the above formula, A_0 represents a DC component; A_1 and φ_1 respectively represent the amplitude and phase of the fundamental frequency component; A_2 and φ_2 respectively represent the amplitude and phase of the 2 × frequency; A_3 and φ_3 respectively represent the amplitude and phase of the 3 × frequency; n(t) represents random noise.

(2) System transfer rate and vibration isolation performance

The force transmission rate is defined by the force transmission rate:

$$\varepsilon = (1 - T_A) \times 100\% \quad (15)$$

The closer ε is to 100%, the better the vibration isolation effect is. Generally, ε is greater than 90% to meet the requirement, that is, the amplitude transmitted to the base force is less than 1/10 of the excitation force amplitude. The force transfer rate describes the ratio of the force transmitted to the foundation and the magnitude of the excitation force, but is not easily measured in engineering. Therefore, the vibration transmission rate is more used in engineering to evaluate the effect of the vibration isolator. The acceleration rate of the vibration measurement before and after the vibration isolation is expressed in decibels as:

$$T_{dB} = 20\lg\frac{a_a}{a_p} \quad (16)$$

In the formula, a_a and a_p are the vibration accelerations of the active end and the passive end, respectively. $T_{dB} > 0$ indicates that the acceleration transmitted to the passive end is attenuated. The higher the transmission rate, the better the vibration isolation effect. $T_{dB} > 20dB$ indicates that the acceleration of the passive end after transmission to the vibration isolation is less than 1/10 of the acceleration of the active end before the vibration isolation, that is, $a_a > 10a_p$. Modeling the force isolation rate of the upper type, we can also define a more intuitive acceleration isolation:

$$\varepsilon_a = \left(1 - \frac{a_p}{a_a}\right) \times 100\% \quad (17)$$

The closer the ε_a is to 100%, the better the vibration isolation effect is. ε_a is equal to 90%, which means that the amplitude of the passive end acceleration response is equal to 1/10 of the amplitude of the active end acceleration, that is, the acceleration of the active end is reduced after the vibration is transmitted to the passive end. 90%. The acceleration transfer rate in the above formula is a commonly used evaluation index for the vibration isolation effect of the suspension system. This paper uses this index to evaluate the vibration isolation effect of the suspension system.

4. Results and discussions

4.1. Data mining results analysis

4.1.1. Efficiency analysis

The efficiency of the two algorithms is compared under different conditions of minsup_count. The results are shown in Fig. 2(a). The number of candidate sets is compared between the two algorithms under different conditions of minsup_count. The result is shown in Fig. 2(b). The execution efficiency of the original Apriori algorithm is closely related to the number of candidate sets generated by the algorithm and the number of scan databases. The improved Apriori algorithm reduces these two indicators well and greatly improves the execution efficiency.

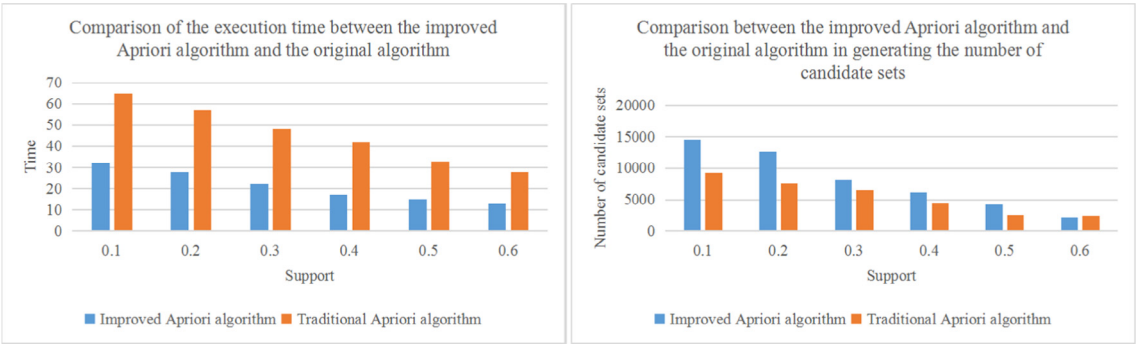
The Apriori improved algorithm based on matrix reduction only needs to scan the transaction database once; effectively avoids the generation of the same candidate set and simultaneously targets the data; as the k value of the k-item increases, the matrix becomes more and more simple, and candidates are obtained. The less time of the support number of the k-item set is; the algorithm converts the scan of the database into a scan for the memory, which greatly improves the execution efficiency of the mining association rule algorithm.

4.1.2. Acquisition of fault data

Under the condition of minimum support and minimum confidence, according to the improved algorithm, detailed fault rule classification can be obtained, which can comprehensively contain fault data and accurately classify fault data. The fault classification rules 1, 3, 6, and 11 in the fault rule can be obtained by the improved ID3 algorithm, and the remaining fault rules cannot be obtained through it. Increase the default minimum support min_sup = 20% and get the fault rule as shown in Table 1.

After the minimum support degree min_sup = 20% is increased, the number of fault classification rules is greatly reduced, indicating that the generation condition of the result is strict, and the fault data information supporting the rule occupies a

Please cite this article as: Z. Li and X. Yu, Data mining technology for mechanical engineering computer test system, Mechanical Systems and Signal Processing, https://doi.org/10.1016/j.ymssp.2020.106628



(a) Comparison in execution time (b) Comparison in the number of generation candidate sets

Fig. 2. Efficiency results analysis chart.

Table 1
min_sup = 20% fault information classification result.

	B2	A2	A6	A3	A1	A10	A4	A5	FAULTNUM
1				1					3
2					0				3
3		0					0		3
4			1	0					2
5			1					1	2
6							0	1	3

larger proportion in the original data repository, and also prove that the results obtained have a broader meaning. Due to the special circumstances, the conditions of individual faults are harsh, and the faults are discarded under higher support components, but these faults are also required. In the face of such contradictions, it is necessary to find a suitable minimum support and minimum confidence, and to determine these conditions to dig out more general fault classification rules. After several trials, the results as shown in Table 2 were obtained:

It can be seen that the fault classification rules obtained under the more appropriate support and confidence include all fault classifications. The fault rule information base composed of the classified information obtained at this time has good guiding value.

4.2. Analysis of simulation results of dynamic balance test system

In the simulation experiment, A0 = 2 mm, A1 = 8 mm, $\varphi 1 = 20°$, A2 = 2 mm, $\varphi 2 = 15°$, A3 = 1.5 mm, $\varphi 3 = 5°$, fundamental frequency f0 = 10 Hz, $\omega = 2\pi f0$, n(t) It is a Gaussian white noise with a standard deviation of 1. The simulation test is carried out under different sampling points. When the number of sampling points is N = 32 and N = 128, the time domain waveform and spectrum of the simulated vibration signal can be obtained, as shown in Figs. 3 and 4, respectively. It can be seen from the above description that the amplitude of the fundamental frequency signal is 8 mm and the phase is 20°. By analyzing the time domain diagram and the spectrogram of the vibration signal, the characteristics of the rotor imbalance signal can be obtained: the signal of the time domain diagram is a harmonic waveform. The spectrum map contains multiple lines and the line with the largest amplitude corresponds to the fundamental frequency of the rotor.

The fundamental frequency of the simulated vibration signal is 8 mm and the phase is 20°. Reasonable simulation sampling parameters should be set during the experiment: the number of sampling points N satisfies the condition of N = 2n, and the sampling frequency fs meets the conditions of the sampling theorem. The results of the four extraction methods for extracting the amplitude and phase at different sampling points are shown in Tables 3 and 4. The results of the measured amplitude and phase are listed in the table. The measurement error is calculated and each record is recorded. The time taken for the second calculation.

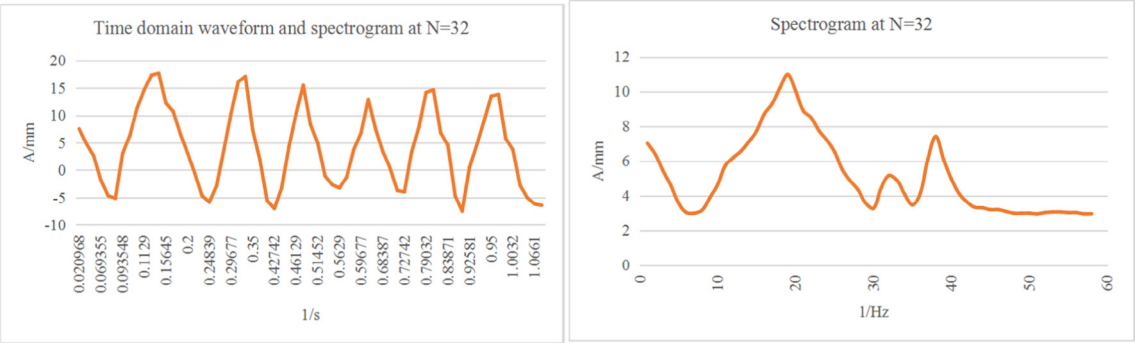
Analysis of simulation and experimental results can lead to the following conclusions:

The extraction result of the integral method has large error, which is not suitable for the extraction of unbalanced system signals with strong noise background. In addition to the integration method, the extraction results of the three extraction methods, as the number of sampling points increases, the extraction error of amplitude and phase is decreasing, that is, the extraction accuracy of amplitude and phase is improving. Among the three methods except the integral method, the results of the three experiments are the highest in the calculation accuracy of the correlation method, and the calculation time is the shortest. The calculation error is very small when the number of sampling points is 32 points. Therefore, the correlation method has the characteristics of small calculation amount and simple processing, and is suitable for calculation of a

Please cite this article as: Z. Li and X. Yu, Data mining technology for mechanical engineering computer test system, Mechanical Systems and Signal Processing, https://doi.org/10.1016/j.ymssp.2020.106628

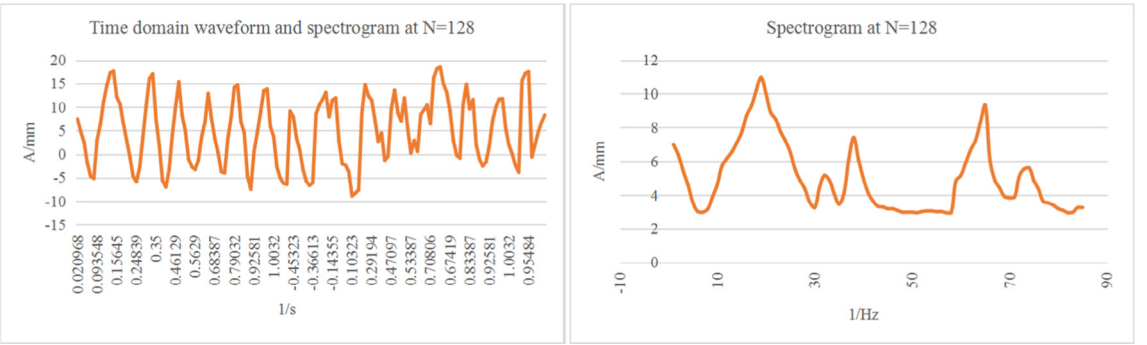
Table 2
Fault information classification results of min_sup = 15% and min_conf = 80%.

	B2	A2	A6	A3	A1	A10	A4	A5	FAULTNUM
1	0								4
2				1					3
3					0				3
4						1			2
5		0					0		3
6			1	0					2
7			1						3
8			1					0	2
9								1	3
10							0	1	3
11	1	0	1				1	0	1
12		1	0	0			1		4
13		1	0	0				1	4



(a) Time domain waveform (b) Spectrum

Fig. 3. Time domain waveform and spectrogram at N = 32.



(a) Time domain waveform (b) Spectrum

Fig. 4. Time domain waveform and spectrogram at N = 128.

Table 3
Calculation results when the number of sampling points is 32.

Extraction Method	Amplitude	Phase	Amplitude Error	Phase Error	Calculating Time
Integral Method	10.3169	−4.2631	–	–	73.456
Related Method	8.2359	20.3169	0.0031	2.0004	0.663
DFT	7.1524	33.2005	1.0357	40.1588	12.069
FFT	8.7129	31.2600	2.0639	32.1366	7.104

Please cite this article as: Z. Li and X. Yu, Data mining technology for mechanical engineering computer test system, Mechanical Systems and Signal Processing, https://doi.org/10.1016/j.ymssp.2020.106628

Table 4
Calculation results when the number of sampling points is 128.

Extraction Method	Amplitude	Phase	Amplitude Error	Phase Error	Calculating Time
Integral Method	15.3214	−6.3150	–	–	42.315
Related Method	11.2647	22.3169	0.2136	0.6489	5.399
DFT	10.9961	25.1297	0.8134	10.2316	20.345
FFT	13.5678	24.9398	0.1994	9.2467	12.334

dynamic balance test system with fewer sampling points. Comparing the calculation time of the DFT method and the FFT method of the test results, it can be seen that the calculation speed of the FFT method is significantly faster than the DFT method, which is consistent with the theoretical analysis.

When calculating the amplitude and phase by the DFT method, only the DFT of the fundamental wave X(1) needs to be calculated without calculating the N point, which can greatly reduce the calculation time and effectively suppress the high frequency interference. After the white noise interference is calculated by the N-point DFT, only the portion of the spectrum in the baseband is output with X(1), so the DFT method can increase the signal-to-noise ratio to SNR = N:1. In practical test applications, the number of sampling points is an integer multiple of the signal period, which can improve the calculation accuracy, so that the hardware requirements are greatly improved. The FFT method is very fast, because it can extract various frequency components and has a good suppression effect on harmonics. However, due to the influence of time domain truncation, energy leakage is inevitable, which makes the calculation accuracy lower. In order to reduce the spectrum leakage, it is necessary to select a reasonable sampling frequency and the number of sampling points. Generally, the number of sampling points N is an integer power of 2, and a suitable window function can be added to the signal.

4.3. Analysis of mechanical vibration test results

Firstly, two unbalances are added to the two turntables of the rotor. The masses are 1.4 g and 1.4 g respectively, and the phase angles are 0° and 90° respectively. The rotor vibration test software of the software system is used to test the vibration of the rotor. First, the rotor is raised to 6000 RPM, the sampling frequency is set to 1652 Hz, the rotor is decelerated, and the vibration signal is collected to obtain the Bode diagram of the rotor, as shown in Fig. 5.

Through the Bode diagram, we can know that the first-order critical speed of the rotor is 3480RPM. The first-order critical speed of the rotor system obtained by ANSYS modeling and analysis is 3570RPM. The actual measured critical speed has some deviation from the modeling calculation. In the modeling, the simplified error of the bearing is supported, the critical speed error caused by the estimated bearing stiffness estimation error is adjusted, and the stiffness of the model is adjusted so that the calculated critical speed is the same as the actual measurement.

The rotor is operated at 6970RPM and operated at the working speed of the rotor to monitor its vibration state. The acquired vibration time domain waveforms in the y-axis direction and the x-axis direction of the bearing housing are shown in Fig. 6. As can be seen from the figure, the vibration signal component is complex, the noise pollution is serious, and the axis trajectory is disordered. The unbalanced fault of the rotor system cannot be clearly identified. The vibration signal is denoised by the PDE denoising method. The axis trajectory of the rotor vibration is obviously elliptical. The rotor system has an unbalanced fault.

It can be seen from the Bode diagram of the balance front and rear rotors of Figs. 5 and 6 that since the equilibrium rotational speed is selected near the critical rotational speed, the vibration at the critical rotational speed after the rotor balance is significantly reduced; the critical rotational speed amplitude is reduced as shown in Table 5. The critical speed amplitude reduction ratio before and after the balance is 66.34%, which achieves a good balance effect, and the rotor can safely pass the first-order critical speed.

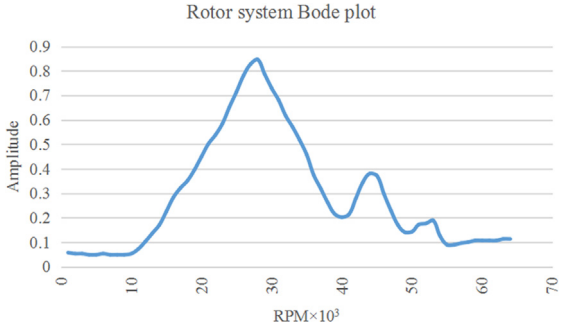


Fig. 5. Bode system Bode diagram.

Please cite this article as: Z. Li and X. Yu, Data mining technology for mechanical engineering computer test system, Mechanical Systems and Signal Processing, https://doi.org/10.1016/j.ymssp.2020.106628

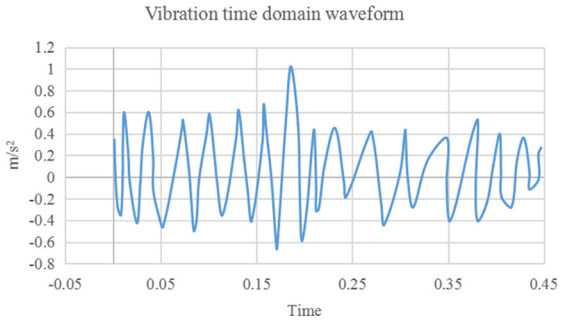


Fig. 6. Vibration time domain waveform.

Table 5
Critical speed amplitude before balance.

	Before Balance	After Balance	Reduction Ratio
Critical Speed Amplitude	1.01	0.34	66.34%

However, due to the limited balance speed selected, the vibration range of the rotor is increased in the range of 5000 RPM-7000 RPM, which is mentioned in the multi-plane influence coefficient method of the flexible rotor, in the wide range of speed. To reduce the vibration value of the rotor, it is necessary to select multiple speeds for balance.

4.4. Performance analysis of test system based on data mining technology

In order to verify the performance of the improved algorithm in the constructed system, the experiment first selects some data in KDDCup99, and discretizes it, then uses Apriori algorithm and improved Apriori algorithm to perform association rule mining on these records (minimum support). The degree is 5.13%, the minimum confidence is 96.34%, the rules they go to are stored in the knowledge base, and the time they spend processing the same training data each time is recorded. After that,

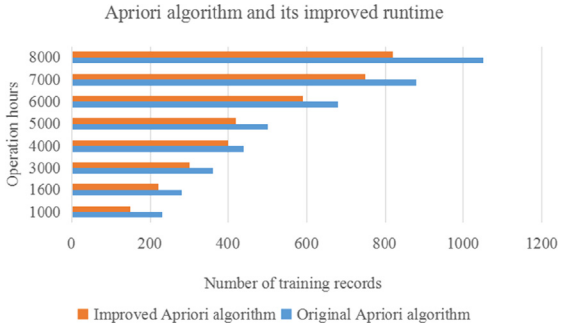


Fig. 7. System running time.

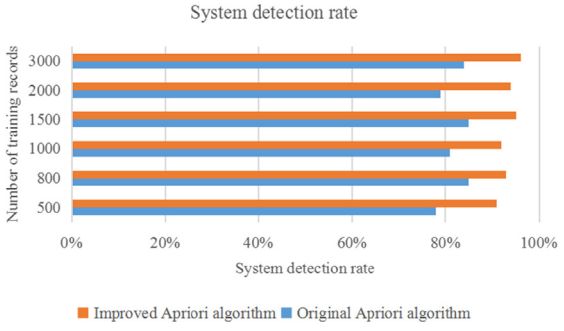


Fig. 8. System detection rate.

Please cite this article as: Z. Li and X. Yu, Data mining technology for mechanical engineering computer test system, Mechanical Systems and Signal Processing, https://doi.org/10.1016/j.ymssp.2020.106628

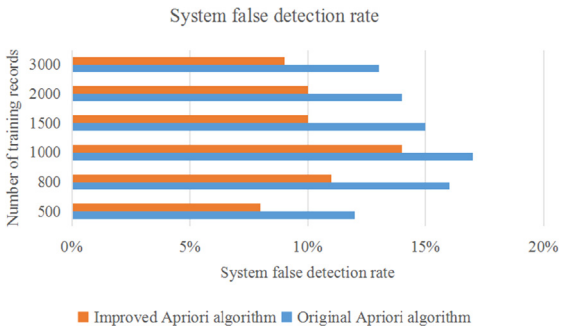


Fig. 9. System false detection rate.

the systems using the two algorithms were tested with the data in the test data set, and their detection rate and false detection rate were counted. It was found that their false detection rates were basically the same, but the detection rate was large. The system running time, detection rate and false detection rate are shown in Figs. 7–9, respectively.

According to the chart, the improved Apriori algorithm has a much higher temporal efficiency than the original algorithm, and the effect becomes more and more significant when the number of training data records is large. The improved algorithm's false detection rate is basically the same as the original algorithm. Without increasing the false detection rate, the detection rate is significantly improved. This shows that the rules generated by it are more streamlined and efficient, and can better outline the user's normal behavior. Moreover, the detection rate of the system does not decrease as the network connection record grows, because the system dynamically updates the rules in the knowledge base. It can be seen that the system established by applying the discretized data processing method and the improved algorithm has certain application value.

5. Conclusions

By introducing the partial differential equation (PDE) into the rotor vibration signal denoising process, the existing PDE filtering numerical solution algorithm is improved, and the calculation time is solved. The high-order PDE filtering theory is derived and the PDE filtering is summarized. The unified model; the effects of several filtering methods are compared by simulation experiments, and the flexible rotors are balanced by different dynamic balancing methods, and satisfactory results are obtained.

From the simulation results, it can be concluded that the integral method is not suitable for the extraction of unbalanced signals with strong noise background, but provides an idea for calculating the amplitude and phase of sinusoidal signals under noiseless conditions; The processing is simple and suitable for the calculation of the dynamic balance test system with fewer sampling points; both the DFT method and the FFT method use the principle of Fourier transform spectrum analysis, but the FFT method calculates the speed much faster than the DFT method.

The improved Apriori algorithm is described in detail and the corresponding mathematical model is established. The enforceability and timeliness of the improved algorithm are verified by an example. Experiments on the classification of fault data prove that the improved Apriori algorithm is greatly improved compared with the original Apriori algorithm, and the speed of acquiring fault rules is improved. From the simulation experiment results, the improved algorithm has higher execution efficiency and detection rate than the original algorithm, which reduces the data processing time, improves the system detection rate, and achieves the purpose of improving the quality of intrusion detection. It has certain practical and theoretical guidance.

CRediT authorship contribution statement

Zhenjun Li: Investigation, Methodology. **Xiaomo Yu:** Supervision, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (No. 61741203 and No. 61866006); Guangxi Natural Science Foundation (No. 2016GXNSFAA380243); Guangxi Higher Education Undergraduate Teaching

Reform Project in 2019 (No. 2019JGA222); Guangxi Vocational Education Teaching Reform Research Project in 2019 (GXGZJG2019A045); Guangxi innovation-driven development of special funds project (No. Gui Ke AA17204091); Research start project of Guangxi Teachers Education University (No. 0819-2017L10); Guangxi Nanning Science and Technology Development Planning Project (20181015-5); The authors gratefully acknowledge the support of the Logistics Engineering Innovation Laboratory of Nanning Normal University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymssp.2020.106628.

References

[1] C.H. Chen, F.J. Hwang, H.Y. Kung, Travel time prediction system based on data clustering for waste collection vehicles, IEICE Trans. Inf. Syst. 102 (7) (2019) 1374–1383.
[2] Q. Wang, P. Lu, Research on application of artificial intelligence in computer network technology, Int. J. Pattern Recognit. Artif. Intell. 33 (5) (2019) 1959015.
[3] E.B. Lee, J. Kim, S.G. Lee, Predicting customer churn in mobile industry using data mining technology, Ind. Manage. Data Syst. 117 (1) (2017) 90–109.
[4] H. Pan, H. Wang, Y. Wang, H. Huang, Rules of acupoint selection for diabetic peripheral neuropathy based on data mining technology, Zhongguo Zhen Jiu 36 (10) (2016) 1111–1114.
[5] Y. Bai, Data cleansing method of talent management data in wireless sensor network based on data mining technology, EURASIP J. Wireless Commun. Networking 2019 (1) (2019) 33.
[6] K. Liu, J. Wang, H. Yan, Wearing comfort analysis from aspect of numerical garment pressure using 3d virtual-reality and data mining technology, Int. J. Clothing Sci. Technol. 29 (2) (2017) 166–179.
[7] F. Sun, L.Z. Huang, Y.F. Liu, W.N. Chen, China Classification Society, D.M. University, A method of evaluating diesel engine performance by using data mining technology, J. Dalian Maritime Univ. 43 (3) (2017) 83–88.
[8] A. Chinchuluun, P. Xanthopoulos, V. Tomaino, P.M. Pardalos, Data mining techniques in agricultural and environmental sciences, Int. J. Agric. Environ. Inf. Syst. 1 (1) (2017) 8–12.
[9] X.Y. Liu, Q. Zhang, C.L. Zhang, F.L. Yuan, S.T. Jiao, Global major events in miocene and its significance: revelation from data mining, Chin. Sci. Bull. 62 (15) (2017) 1645–1654.
[10] M. Fujimoto, M. Kanou, K. Hosomi, M. Takada, Angiotensin receptor blockers and the risk of cancer: data mining of a spontaneous reporting database and a claims database, Int. J. Clin. Pharmacol. Ther. 55 (4) (2017) 295.
[11] F. Giannotti, M. Nanni, D. Pedreschi, C. Renso, R. Trasarti, A query language for mobility data mining, Int. J. Data Warehouse. Min. 7 (1) (2017) 24–45.
[12] S. Kar, S.R. Samantaray, M.D. Zadeh, Data-mining model based intelligent differential microgrid protection scheme, IEEE Syst. J. 11 (2) (2017) 1161–1169.
[13] S. Sorour, K. Goda, T. Mine, Comment data mining to estimate student performance considering consecutive lessons, Educ. Technol. Soc. 20 (1) (2017) 73–86.
[14] Q. Sun, L. Shi, Y. Ni, D. Si, J. Zhu, An enhanced cascading failure model integrating data mining technique, Prot. Control Modern Power Syst. 2 (1) (2017) 5.
[15] U. Adhikari, T. Morris, S. Pan, Wams cyber-physical test bed for power system, cybersecurity study, and data mining, IEEE Trans. Smart Grid 8 (6) (2017) 2744–2753.
[16] J. Wang, X. Chen, X. Hou, L. Zhou, C. Zhu, L. Ji, Construction, implementation and testing of an image identification system using computer vision methods for fruit flies with economic importance (diptera: tephritidae), Pest Manag. Sci. 73 (7) (2017) 1511.
[17] Q. Wang, Y. Yu, K. Mou, Testing the effect of computer-generated hologram fabrication error in a cylindrical interferometry system, J. Mod. Opt. 64 (4) (2017) 1–11.
[18] D. Wang, S. Zhang, R. Wu, C.Y. Huang, H.N. Cheng, R. Liang, Computer-aided high-accuracy testing of reflective surface with reverse hartmann test, Opt. Express 24 (17) (2016) 19671.
[19] J. Yang, J. Wang, R. Luo, C. Gao, L. Songnong, Z. Kongjun, Simulation test system of non-contact d-dot voltage transformer, Int. J. Emerg. Electr. Power Syst. 17 (2) (2016) 91–99.
[20] Z. Lv, B. Hu, H. Lv, Infrastructure monitoring and operation for smart cities based on IoT system, IEEE Trans. Ind. Inf. (2019).
[21] W.Y. Yan, L. Li, Y.G. Yang, X.L. Lin, J.Z. Wu, application of the computer-based respiratory sound analysis system based on mel-frequency cepstral coefficient and dynamic time warping in healthy children, Zhonghua Er Ke Za Zhi 54 (8) (2016) 605–609.
[22] G. Flens, N. Smits, C.B. Terwee, J. Dekker, E.D. Beurs, Development of a computer adaptive test for depression based on the dutch-flemish version of the promis item bank, Eval. Health Prof. 40 (1) (2017) 79–105.
[23] D.S. Dima, D. Covaciu, Vehicles frontal impact analysis using computer simulation and crash test, Int. J. Automot. Technol. 20 (4) (2019) 655–661.
[24] A.M. Gruzlikov, N.V. Kolesov, E.V. Lukoyanov, Test-based diagnosis of faults in data exchange addressing in computer systems using parallel model, J. Comput. Syst. Sci. Int. 57 (3) (2018) 420–433.
[25] J.R. Root, B.S. Stevenson, L.L. Davis, J. Geddes-Hall, D.W. Test, Establishing computer-assisted instruction to teach academics to students with autism as an evidence-based practice, J. Autism Dev. Disord. 47 (2) (2017) 1–10.
[26] S. Lee, S.H. Kim, B.C. Kwon, Vlat: development of a visualization literacy assessment test, IEEE Trans. Visual Comput. Graphics 23 (1) (2017) 551–560.
[27] Sang-Bing Tsai, Yu-Cheng Lee, Jiann-Jong Guo, Using modified grey forecasting models to forecast the growth trends of green materials, Proc. Inst. Mech. Eng., Part B: J. Eng. Manuf. 228 (6) (2014) 931–940.
[28] Y. Zhang, X. Xiao, L.X. Yang, Y. Xiang, S. Zhong, Secure and efficient outsourcing of PCA-based face recognition, IEEE Trans. Inf. Forensics Secur. (2019).
[29] F. Xiao, Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy, Inf. Fusion 46 (2019) 23–32.
[30] Q. Wang, Y. Li, X. Liu, Analysis of feature fatigue EEG signals based on wavelet entropy, Int. J. Pattern Recognit. Artif. Intell. 32 (08) (2018) 1854023.

Please cite this article as: Z. Li and X. Yu, Data mining technology for mechanical engineering computer test system, Mechanical Systems and Signal Processing, https://doi.org/10.1016/j.ymssp.2020.106628