IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 25, 2018, accepted November 30, 2018, date of publication December 6, 2018,
date of current version January 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2885310

Effective Information Filtering Mining of Internet of Brain Things Based on Support Vector Machine

XIAOFENG ZHAO[1], WEIHUA ZANG[2], RULAN LV[3], AND WEI CUI[1]

[1]School of Management Engineering and Business, Hebei University of Engineering, Handan 056038, China
[2]Electronic and Information Engineering Experiment and Training Center, Handan University, Handan 056005, China
[3]Hebei Branch of Construction and Administration Bureau of South-of-North Water Diversion Middle Route Project, Handan 056006, China

Corresponding author: Weihua Zang (hdxyzwh@hotmail.com)

This work was supported by the Project of the Hebei Social Science Foundation under Grant hb15gl120.

ABSTRACT The traditional Internet of Brain Things' big data information filtering method ignores the extraction of big data features, the filtering effect, and the effect of denoising processing on the simulation results, resulting in low filtering accuracy and poor performance. An effective information filtering and mining algorithm for the Internet of Brain Things based on support vector machine (SVM) is proposed. First, the model construction and feature extraction of the Internet of Brain Things' big data system are carried out. The correlation feature extraction is performed on the effective information features; the correlation factors of the effective information data are sorted; the main feature quantity of the relevance degree is extracted; and the filter non-association is designed. The information is reasonably filtered; all data are processed, converted to the same interval for processing; data protocol is implemented; and data effective information feature mining is implemented based on the SVM algorithm. The simulation results show that the algorithm is effective for filtering big data and has high precision and superior performance, which shows good application value.

INDEX TERMS Internet of Brain Things, data mining, support vector machine.

I. INTRODUCTION

With the development of stochastic networks and data information science, humans have entered the era of networks and big data. Among them, the Internet of Things technology is based on the network structure of the Internet based on the development of the Internet. The emergence of the Internet of brain Things has set off the third wave of the world information industry [1]–[2]. The Internet of brain Things (IOBT) uses data structure to pass items through data-aware technology and identity recognition technology, and adopts data mining methods to realize information sharing and network connection, thereby realizing intelligent control and information exchange and sharing of data information. At the same time, with the development of big data information technology, data mining has received people's attention. Through data mining, effective information features are obtained from massive amounts of big data for my use, combined with Internet of brain Things technology for data collection and data management. It can be seen that in the Internet of brain Things environment and effective information filtering of big data is needed to improve the ability to mine and identify useful data. It is important to study the big data effective information filtering and mining algorithm in the IOBT environment to improve the data control and recognition ability [3]–[5]. Users are not completely disgusted or indifferent to advertising information on social networks. According to the survey, nearly 50% of Weibo users access the website through the website link in Weibo; when Weibo users interact with Weibo ads, 40.4% of users will participate in forwarding; 60.7% of Weibo users are Discounts, sweepstakes attracted and clicks on Weibo image ads; and 44.4% of users who shop on Weibo because of advertising and demand. Therefore, what the user really dislikes is not the advertisement itself, but the existing mobile social network lacks a formal and friendly product promotion mechanism and a false and inferior commodity information screening mechanism, which leads to various advertisements in the mobile social network being pervasive, and the advertisement information is difficult to be true. In order to solve the above problem, it can be solved in two directions. On the one hand, it researches spam filtering technology and implements filtering of spam advertising information on mobile social networks. On the other hand, the establishment of a product recommendation mechanism based on mobile social networks not only provides a good merchandise display platform for regular merchants, but also provides users with friendly and intimate recommendations is a suitable choice.

At present, the information filtering and mining algorithms for the Internet of brain Things big data mainly include particle filter algorithm, neural network control algorithm and Support Vector Machine (SVM) algorithm [6]–[8]. The fundamental realization of these methods is that they need to be solved. Big data noise background suppression and IOBT big data space dimensionality reduction, an effective method is to filter the preprocessing of big data to achieve large data reduction. With regard to the research on effective information filtering and mining of big data in the Internet of brain Things environment, related literatures have been carried out and elaborated, and some research results have been obtained. Among them, the literature [9], [10] proposes a data filtering algorithm for data cross-linking big data filtering. Mining algorithm, the literatureuses the genetic algorithm to filter the effective information of the Internet of brain Things big data, through the effective preprocessing of the Internet of brain Things big data, to achieve large data reduction, improve data mining performance, but the algorithm has calculation Complex problems are difficult to implement. In addition, PSO particle swarm clustering algorithm and association rule discovery discriminant algorithm are proposed in the further research to realize effective information filtering mining of big data, which can avoid different density areas in the data set. When the edges are close together, the edge points are misjudged. However, this kind of algorithm is more difficult to find the characteristics of big data [11], [12]. The traditional particle filter algorithm requires extremely high initial trajectory of particle filtering, and the mining performance is not good when the data noise is large.

Aiming at the above problems, this paper proposes an effective information filtering algorithm for big data based on support vector machine (SVM) to improve the filtering and mining performance of big data effective information in the Internet of brain Things environment. Firstly, the data feature model is constructed. The correlation feature extraction is performed on the effective information features. The filter non-associated information is designed for reasonable filtering. The support vector machine SVM algorithm is used to realize the effective information feature mining. Finally, the simulation experiment is carried out [13]. The two main research points of this paper are related research on information filtering and product recommendation. In the aspect of spam filtering, the status of filtering technology was investigated, and the main filtering technologies were classified and the advantages and disadvantages were compared. In the aspect of commodity recommendation, three main commodity recommendation technologies—collaborative filtering recommendation, content-based recommendation, and social recommendation—are analyzed in detail, and the advantages and limitations of each method are analyzed, which lays a theoretical foundation for the research of this topic. Based on the research of spam filtering, this paper studies the SVM incremental learning method based on the characteristics and existing problems of SVM in information filtering [14]–[16]. An improved SVM incremental learning algorithm is proposed by comparing the incremental learning methods in SVM algorithm. The algorithm will violate the KTT condition as the classification basis of the incremental data set, and optimize the improved algorithm to make it more efficient to classify the SVM incremental set. Finally, the improved SVM incremental learning algorithm is applied to the spam filtering system to obtain better filtering results.

II. RELATED WORK

A. SPAM FILTERING TECHNOLOGY

Keyword filtering is simple to use and easy to understand, but it requires constant maintenance of updated keyword lists and requires specialized domain knowledge. For mobile social networks, where information is all-encompassing and topic updates are developing very fast, it is not suitable for spam filtering using keyword filtering methods. The social network-based filtering method uses the difference in social relations and behavior between the sender of the spam and the sender of the normal information to filter the sender of the spam; or to filter the garbage by using the difference between the spam and the propagation characteristics of the normal information [17]–[20]. However, the accuracy and stability of this filtering method are still in doubt, so this paper does not use this method to filter spam in mobile social networks. Based on the machine learning classification filtering method, the filtering accuracy rate is high and the speed is fast, and the research on machine learning is relatively mature. Therefore, this paper uses the filtering method based on machine learning classification to filter spam in mobile social networks. In the literature research, many articles have pointed out that the support vector machine algorithm has excellent performance in engineering application and academic research. Therefore, this paper will focus on the spam filtering method based on SVM classification. Considering that the training time faced by the SVM classification filtering method is too long, and the retraining training time becomes very long when the model needs to be changed, the mobile social network information update is developing very fast, and the incremental SVM algorithm is deeply studied. Use it in a spam filtering system [21].

1) COMMODITY EVALUATION VALUE AND PRODUCT RECOMMENDATION DEGREE

Because social networks define the connections between users, social networks are often defined using diagrams. As shown in Figure 1, define a social network with graphs G (V, E, w), where V is a set of vertices, each vertex represents a user, and E is a set of edges. If users v_a and v_b have social network relationships, then there is a side $e(v_a, v_b)$, connecting these two users, $w(v_a, v_b)$ defines the weight of the edge.

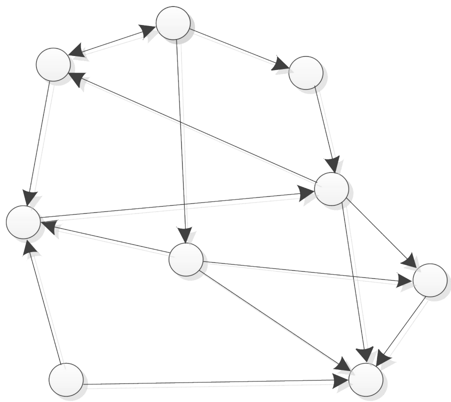


FIGURE 1. Social network relationship model.

Similarly, this article defines User A's n-layer focus on friends u: Assume that user v is user A's n-1 layer followers, user v pays attention to user u and user A and its n-2 layer of friends are not directly concerned User u defines user u as user A's n-layer attention friends (as shown in Figure 2).

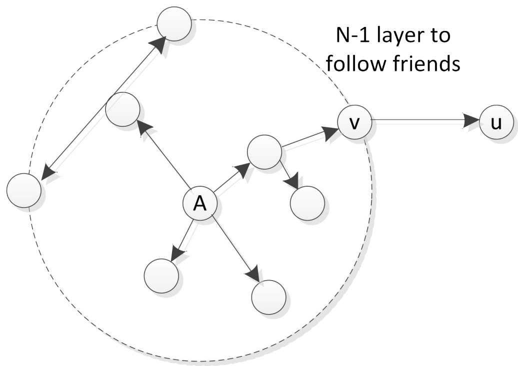


FIGURE 2. n layer attention friends map.

After obtaining the N users with the highest similarity of the target users, it is possible to recommend the collection of items that the similar users like to the target users. Then how to judge the user's favorite products, this article measures the user's preference for the product through the user's operation behavior of the product, that is, the user's evaluation value of the product.

Such user operations on the merchandise include purchasing merchandise, rating merchandise, and collecting merchandise. Now users score a product on a five-point scale, and users can type an integer from 1 to 5. In addition to this explicit rating, the user may have purchased the item but did not score it (this phenomenon often occurs), but this often indicates that the user recognizes the item, but the item does not meet the user's 100% satisfaction, so this article defines the user. The item not rated for purchase is defaulted to a score of 4 points. The behavior of collecting goods often reflects the user's interest in the product, but there is still hesitation. The definition of the collected goods in this article defaults to 4 points for the product. This unifies the user's interest/satisfaction as reflected by the user's operational behavior.

This article defines user c's overall evaluation of the product as:

$$evaluation(u, i) = 0.5 + \frac{score(u, i) - 3}{5} \tag{1}$$

where, $score(u, i)$ is the rating of user u on item i.

At this point, N users with the highest similarity of the target users and the favorite products of each user and the evaluation value of the products can be obtained, as shown in Figure3, that is, a batch of products to be recommended are obtained.

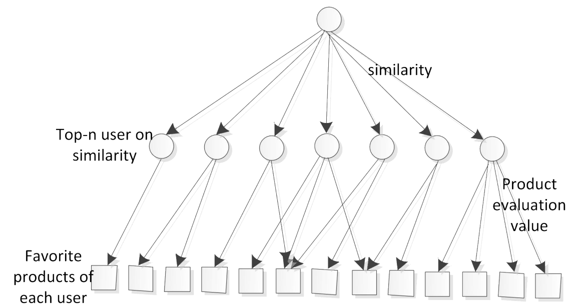


FIGURE 3. Schematic diagram of the product to be recommended.

After obtaining the list of items to be recommended, we define the recommended degree for the user u in the list of items to be recommended:

$$\text{recommendation}(u, i) = \sum_{v \in In(i)} [\chi \cdot similarity(u, v) + \delta \cdot evaluation(v, i)] \tag{2}$$

where $In(i)$ represents a set of users having an evaluation relationship with the product i among the N users with the highest similarity to the user u; $similarity(u, v)$ represents the similarity between the user u and the user v; e valuation (v, i) indicates the overall evaluation value of user v for item i; ψ and δ are correction values used to correct the user similarity and the proportion of product evaluation in the recommendation.

At this point, the recommended product list is sorted according to the product recommendation degree, and the K products with the highest recommendation are recommended to the target user, and the recommendation for the target user is completed.

B. SVM INCREMENTAL LEARNING CLASSIFICATION ALGORITHM

The filtering method based on machine learning classification has high filtering accuracy and fast speed. Especially the support vector machine algorithm shows extremely high classification accuracy in engineering application and academic research. However, the SVM classification filtering method still faces the following problems: First, the training time is too long; secondly, the training data often changes with time, the original model may not be able to identify new data well, and then it needs to be changed. The model adapts it to new changes. If the method of retraining all the training data is taken at this time, the training time of the classification model will become very long, which will seriously reduce the efficiency. In response to this serious and serious problem, the researchers proposed an incremental support vector machine algorithm [22]–[24] . The incremental support vector machine algorithm is different from the retraining method. It does not completely abandon the previous learning results. Instead, it uses the previous learning to analyze the representative data and combines the new data to form continuous training.

1) BASIC PRINCIPLES OF SUPPORT VECTOR MACHINE ALGORITHM

First understand the linear classifier from the case of linear separability. In order to separate the two different types of points, there can be many classification planes. And there is such a classification surface, on the basis of separating the two types of sample points, at the same time, the distance between the two types of classifications is maximized, that is, the closest point among all the points in the optimal classification plane has the largest spacing, which is called Optimal classification surface.

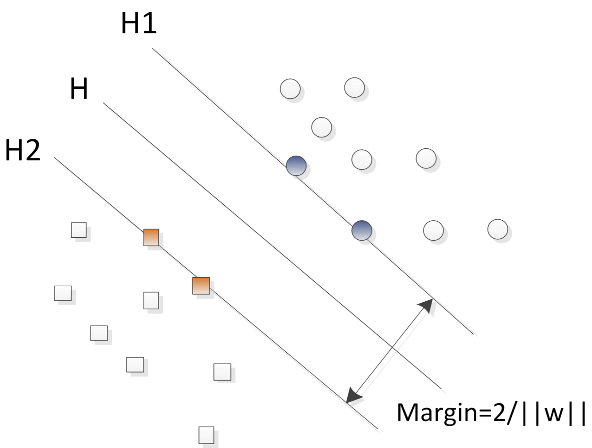


FIGURE 4. **Classification hyperplane.**

As shown in Figure 4, the dots and squares are two types of sample points, which are linearly separable and formally expressed as:

$$(x_1, y_1), \cdots, (x_n, y_n)\ x \in R^n, \quad y \in \{+1, -1\} \tag{3}$$

Because the classification face correctly separates the two types of sample points, it satisfies:

$$y_i(w \cdot x_i + b) \geq 1 \quad i = 1, 2, \cdots, n \tag{4}$$

All the dot product operations in the optimization problem are replaced by kernel function operations, so that the most general form of support vector machine algorithm can be obtained:

$$\begin{aligned} &\max \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_j \phi(x_i) \cdot \phi(x_j) \\ =\ &\max \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_j K(x_i, x_j) \end{aligned} \tag{5}$$

$$s.t.\ 0 \leq a_i \leq C \quad i = 1, 2, \cdots, n \sum_{i=1}^{n} a_i y_i = 0 \tag{6}$$

The kernel function generally has a polynomial kernel, a Gaussian radial basis kernel, an exponential radial basis kernel, a multi-hidden layer perceptual kernel, a Fourier series kernel, and a spline kernel. But how to choose the kernel function and its parameters, there is no definite method yet.

2) AN IMPROVED SVM INCREMENTAL LEARNING ALGORITHM

From the basic principle of the above support vector machine, the calculation problem of the optimal classification hyperplane can be described as a conditional extreme value problem, and the optimal solution can be transformed into the saddle point of the following Lagrange function:

$$\begin{aligned} L_p = \frac{1}{2} \|w\|^2 + C \sum_{i=1} \xi_i - \sum_i a\{y_i(\Phi(x_i) \cdot w + b) \\ - 1 + \xi_i\} - \sum_{i=1} u_i \xi_i \end{aligned} \tag{7}$$

Where w and b are the normal vectors and thresholds of the optimal hyperplane of the attribute space, respectively, u_i and ξ_i is non-negative Lagrange multipliers respectively, and C is a non-negative error control parameter, which is a function of mapping the input space to the attribute space.

According to the Karush-Kuhn-Tucker (KKT) theorem, the optimal solution satisfies the following conditions:

$$w = \sum_i a y_i \phi(x_i) \quad 0 \leq a \leq C, \quad \sum_i a y_i = 0 \tag{8}$$

In most cases, the number of SVs in the training set is only a small fraction of the training sample set, $N_{sv} << N_{train}$. Therefore, the SV set can be used instead of the training sample set for classification learning, so that the training time can be greatly reduced without affecting the classification accuracy.

III. EXPERIME MODEL CONSTRUCTION AND FEATURE EXTRACTION OF BIG DATA SYSTEM IN INTERNET OF THINGS

With the development of random network and data information science, mankind has entered the era of Internet and big data. Among them, the Internet of Things (IOBT) technology is an interconnected network structure based on the development of computer Internet. The emergence of the Internet of brain Things has set off the third wave of the world information industry. The Internet of Things (IOBT) realizes information sharing and network connection through data sensing technology and identity recognition technology in the form of network structure, thus realizing intelligent control and information exchange and sharing of data information. At the same time, with the development of big data information technology, data mining has attracted people's attention. Through data mining, effective information features from large amounts of data can be used by us. Data acquisition and data management are carried out in combination with Internet of Things technology [25], [26]. It can be seen that in the environment of the Internet of brain Things, it is necessary to filter large data effectively and improve the ability of mining and identifying useful data. It is of great significance to study the filtering and mining algorithm of large data in the environment of Internet of brain Things for improving the control and recognition ability of data.

At present, there are mainly particle filters, neural network control and support vector machine (SVM) algorithms for information filtering and mining of large data in the Internet of brain Things. These methods need to solve the problem of large data noise background suppression and dimensionality reduction of large data space in the Internet of brain Things. An effective method is to perform filtering preprocessing on big data to achieve dimensionality reduction for big data. With regard to the research on filtering and mining of large data effective information under the environment of the Internet of brain Things, the relevant literature has been elaborated and some research results have been achieved. Among them, a particle filter algorithm is proposed in the literature, which is based on genetic algorithm to filter out the effective information of large data in the Internet brain of Things. Through the effective preprocessing of large data in the Internet of brain Things, the dimensionality of large data is reduced and the performance of data mining is improved. However, the algorithm has complex computational complexity and is difficult to implement. In addition, the PSO particle swarm optimization algorithm and the discovery and discrimination algorithm of association rules are proposed to realize effective information filtering and mining for large data of the Internet of brain Things [27], [28]. This method can avoid the misjudgement of edge points when regions with different densities are close to each other. However, this kind of algorithm is difficult to find the characteristics of large data. The traditional particle filter algorithm requires very high initial trajectory of particle filter and has poor mining performance in the case of large data noise.

In order to solve the above problems, this paper proposes an effective information filtering and mining algorithm based on support vector machine (SVM) for large data in the Internet of brain Things (IOBT), which can improve the filtering and mining performance of large data effective information features in the environment of the Internet of Things. Firstly, the data feature model is constructed, the correlation dimension feature extraction of the effective information feature is preprocessed, the filter non-correlation information is designed to filter reasonably, and the data effective information feature mining is realized based on the support vector machine SVM algorithm. Finally, the simulation experiment verifies the superiority of the algorithm.

A. OVERALL FRAMEWORK OF SYSTEM MODEL

The model of large data information feature control system under the environment of Internet of brain things is shown in Figure 5.

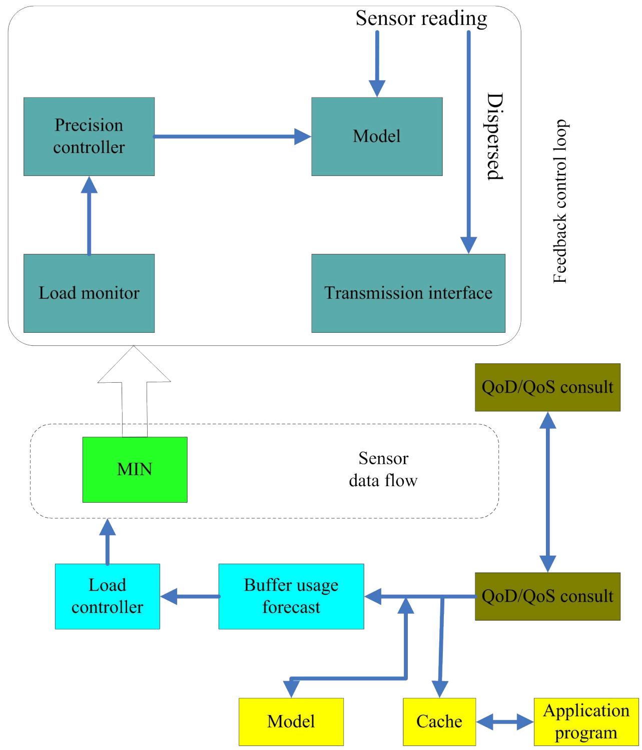


FIGURE 5. Model of large data information feature control system under Internet of brain things environment

The basic purpose of data mining is to extract the valid information features of data stream or time series in the networked environment, and to filter the useless features reasonably. In the framework of the Internet of brain Things, large data sources collect real-time data with data information and logical sensors released by many publishers. Firstly, the data generation model is constructed for network control model. Different application platforms can be monitored by multi-source information resource cloud retrieval mechanism. Each publisher collects a set of lower-level sensor data and publishes related data streams to a subset. The data management back-end can distribute and count the data collected by the lower sensors. There is $C_x \cap C_y = \phi$, which encodes the collected information and gets the coding characteristics of the large data information in the Internet of brain Things.

$$X = 2\left(\frac{P}{N_0}\left(9+\frac{3}{\alpha-1}+\frac{6}{\alpha-2}\right)\left(\frac{P}{\beta N_0}-r^{\alpha}\right)^{-1}\right)^{1/\alpha} \tag{9}$$

On the basis of the above system model, we design and process the extraction algorithm of information characteristics of large data in the Internet of brain Things. Considering the discreteness of large data spatial information in the Internet of brain Things, a large number of discrete points are dominant, and feature dimension reduction design is needed. Feature data includes various types of data models. Assuming that the starting time of feature partition of large data in the Internet of brain Things is t_0, and the effective information of large data in the Internet of brain Things is distributed at level i, the return state x_0 (t_k) is as follows:

$$\dot{Y} = AY + B\left[f\left(Y\right) + u\right] \tag{10}$$

The set S is encoded as a bit string B with m bits. At the beginning, all bits are set to 0. Referring to MapReduce data processing process, differential evolution is carried out based on excellent gene loci. Data vector X is divided into V non-overlapping subsequences $\{X_v, v = 1, 2, ..., V\}$. The SVM algorithm is used to reduce the dimension of the feature space, and the data feature can be expressed by two dimensional expressions.

$$g_{mn}\left(t\right) = g\left(t - mT\right) e^{j2\pi(nF)t} \quad m, n = 0, \pm 1, \pm 2, \ldots \tag{11}$$

Assuming that the observed value of the Internet of things sensor S $_k$ is θ, the data feature segmentation satisfies:

$$x_{id}^{t+1} = \omega x_{id}^{t} + c_1 r_1 \left(p_{id} - x_{id}^{t}\right) + c_2 r_2 \left(p_{gd} - x_{id}^{t}\right) \tag{12}$$

When data transmission is carried out, other nodes are not allowed to send data in the same time. The definition of monitoring data correlation is that the probability of producing a layer of aggregate tree frequent pattern set X in behavior set t is recorded as $P\left(X, T_d\right)$, and the feature extraction of correlation dimension is defined as:

$$P\left(X, T_d\right) = \prod\nolimits_{i_u \in X \wedge X \subset T_d} p\left(i_u, T_d\right) \tag{13}$$

The cumulative interference of the node v of the Internet of brain Things is calculated, and the interference filter is designed to filter the noise interference of the characteristic data. The system function of the filter is as follows:

$$\begin{aligned} \Re\left(\rho, \theta\right) &= \frac{p\left(z_t|x_t\right) p\left(x_t|u_{t-1}, \ldots, z_0\right)}{p\left(z_t|u_{t-1}, d_{0,\ldots,t-1}\right)} \\ &= \eta p\left(z_t|x_t\right) \int p\left(x_t|x_{t-1}, u_{t-1}\right) Bel\left(x_{t-1}\right) dx_{t-1} \end{aligned} \tag{14}$$

Assuming that $x\left(t\right), t = 0, 1, \ldots, n-1$ is training sequences of samples, t=0 is set to filter the non-related information of large data in the Internet of brain Things reasonably through the filter designed above. By sorting the data association factors and extracting the main feature of association degree, the probability density of indirect effective feature mining of node I to node j at t-Time is obtained as follows:

$$I_{i,j}\left(t\right) = \frac{\sum D_{i,k}^{n}\left(t\right) D_{k,j}^{n}\left(t\right)}{\sum D_{i,k}^{n}\left(t\right)} \tag{15}$$

Through the above analysis, the correlation dimension feature extraction and preprocessing of the effective information features are carried out, and the filter non-correlation information is designed to filter reasonably, which improves the performance of filtering and mining the effective information of large data in the Internet of Things.

B. IMPLEMENTATION OF DATA MINING ALGORITHM

On the basis of the above system model construction and feature extraction of correlation dimension, the algorithm design of effective information filtering and mining for large data in the Internet of brain Things is carried out. The traditional filtering method of large data in the Internet of brain Things uses particle filter algorithm for information filtering mining. The algorithm requires very high initial trajectory of particle filter and has poor mining performance in the case of large data noise. An effective information filtering algorithm for big data in Internet of things based on support vector machine (SVM) is proposed.

The idea of algorithm improvement is described below.

Firstly, SVM support vector machine algorithm is used to clean up the large data of the Internet of brain Things, and to clean up the abnormal data collected, such as noise data, unrelated data and so on. Then, data integration is carried out, that is, data collected by different devices are classified and aggregated according to their sources, physical characteristics and logical characteristics. Then, data conversion is carried out. On the basis of data integration, all data are processed regularly and converted to the same interval for processing. Finally, the data specification, the formulation of certain data mining rules, the design of filter non-associated information for reasonable filtering, based on support vector machine SVM algorithm to achieve data mining features of effective information, improve the efficiency of data mining. Based on the above improvement ideas, the key technology of the algorithm is described.

Define the distortion sensitive parameter $\left\{S_j^{(n)}, j = 0, 1, \ldots, N-1\right\}$ of the SVM support vector machine training algorithm, and find out the minimum distance of the SVM node

$$N_{j*}, d_{j*} = \min_{0 \le j \le N-1} \left\{d_j\right\}$$

Data aggregation tree association is a kind of data dependence. In the behavior set D of filtering effective information data, the expected support number of data aggregation tree correlation strength X is recorded as $\exp SN\left(X\right)$ and the unbiased risk estimate value based on aggregation tree is defined as:

$$\exp SN\left(X\right) = \sum\nolimits_{T_d \supseteq X \wedge T_d \in D} P\left(X, T_d\right) \tag{16}$$

In the formula, T_d is the sampling interval for extracting data correlation information in one observation.

In the process of adjusting weight vector α_c, support vector machine is used for data mining, and n valid information data samples are obtained in set S_s. The association rule matrix of valid information data is as follows:

$$Q^{'} = \begin{bmatrix} 0 & y_1 & \cdots & y_n \\ y_1 & Q_{11} & \cdots & Q_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & Q_{n1} & \cdots & Q_{nn} \end{bmatrix} \overset{def}{=} \begin{bmatrix} 0 & y^T \\ y & Q \end{bmatrix} \tag{17}$$

In the formula, the matrix Q is positive definite.

By gradually increasing or deleting the effective information weight vectors in the samples, the effective information measure features of the large data of the Internet of Things are obtained as follows:

$$\det\left(Q^{'}\right) = \det(Q) \cdot \left(-y^T Q^{-1}\right) \neq 0 \tag{18}$$

The planar region of the Internet of brain Things nodes is divided into several blocks with $K \cdot l$ edges and no overlap, and the minimum unbiased delay estimation of data mining is obtained.

$$\tau = \alpha\left(1+2^{-\alpha/2}\right)(\alpha-1)^{-1}+\pi 2^{-\alpha/2}(\alpha-2)^{-1}/2 \tag{19}$$

Considering the data transmission between dominant nodes, it takes R to transmit the results to sink through the shortest path in the process of filtering and mining large data valid information in each round of Internet of brain Things (IOBT). Through the above processing, an improved algorithm of filtering and mining large data valid information in Internet of brain Things based on Support Vector Machine (SVM) is realized. Finally, a simulation experiment is carried out to verify the performance.

C. SIMULATION ANALYSIS

In order to test the performance of the algorithm in the effective data filtering of big data in the Internet of brain Things, a simulation experiment is carried out. The hardware environment of the simulation experiment is Intel (R) 2.3 GHz CPU, 2 GB memory and 32 bit Windows 7 PC. Under the environment of MyEclipse 8.5, it is based on Matlab2010 programming platform. The model of Internet of brain Things (IOBT) is constructed. The structure model of IOT is set up by sensor network, and the continuous supply of energy is needed when data acquisition is carried out by sensors. The working voltage of the sensor network chip is 0.81 V, the frequency is 917 MHz, the DieSize size is 828∗1016 um2, and the different thresholds are marked as high-Vt and low-Vt. RTL code is transplanted to floorplan by Design Compiler, and data structure is characterized and measured by 32-bit embedded CPU of low-Vt unit. Real-time monitoring data is persistent. In the design of large data filter, the row number of support vector machine data is 20 × 1000 = 20000, the column number is 20, the user parameter is L = 500, and the user parameter is L = 500. The weight vector $W = \{\omega_s, \omega_t, \omega_u\}^T$ of ontology features and the information parameters of standard data sets in effective information filtering mining are shown in Table 1 and Figure 6.

TABLE 1. Standard data sets needed for experiment acquisition.

Data set name	Data set size	Attribute set size
Breast cancer	223	65
Ship radiated noise	437	36
Sonar	823	44

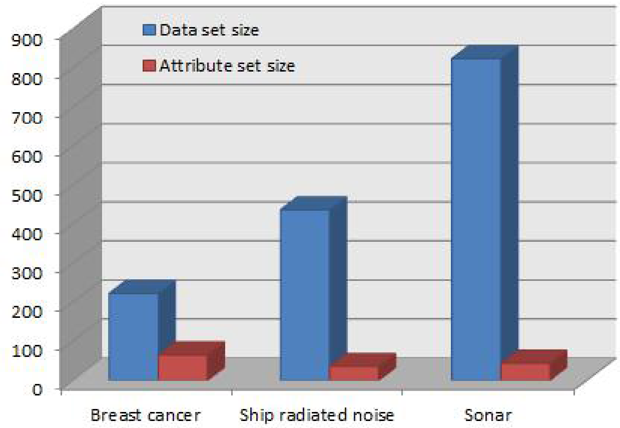


FIGURE 6. Comparison of standard datasets.

According to the above simulation environment, the valid information filtering and mining simulation experiments of large data in the Internet of brain Things are carried out, and the waveforms in the data information domain and frequency domain are obtained, as shown in Figure 7.

The data samples given in Figure 7 are filtered and mined to extract the features of correlation dimension, and the results of filtering and mining of large data in the Internet of brain Things are obtained, as shown in Figure 8. The X, Y and Z coordinates in Figure 8 represent the three-dimensional spatial position of the data, and the red area represents the result of filtering.

It can be seen from the graph that the proposed algorithm can effectively filter out the interference information, realize the filtering and mining of the effective information, and improve the feature extraction performance of the large data of the Internet of brain Things, in order to compare the performance of the algorithm. Compared with the classical heritage algorithm and particle swarm optimization algorithm in the literature, Monte Carlo experiment is used to obtain the detection probability results of different algorithms for accurate data mining, as shown in Figure 9.

From Figure 9, we can see that the accurate mining rate of the data calculated in this paper is 96.7%, while compared with the traditional algorithm, and the accuracy of data mining is improved by 17.5%. This shows the superiority of this algorithm in the realization of effective data filtering and mining.

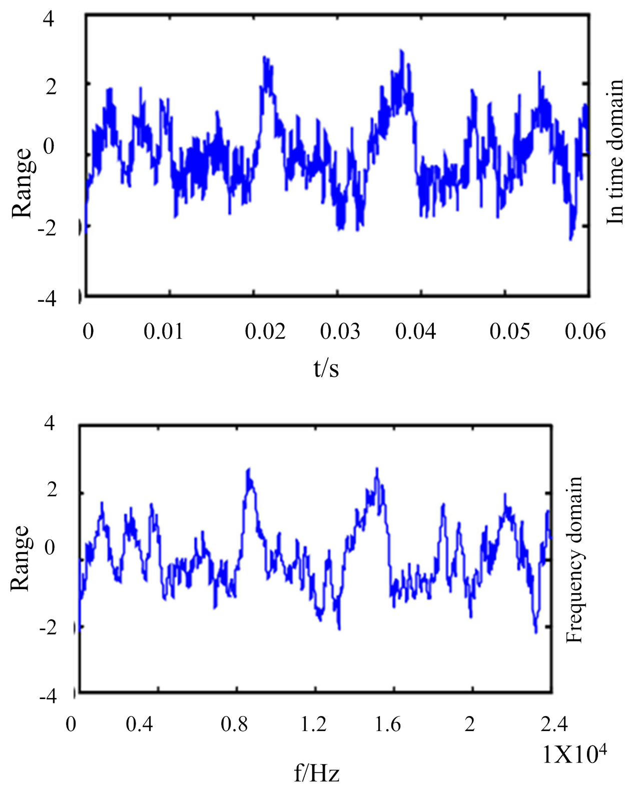


FIGURE 7. Data sample time and frequency domain information.

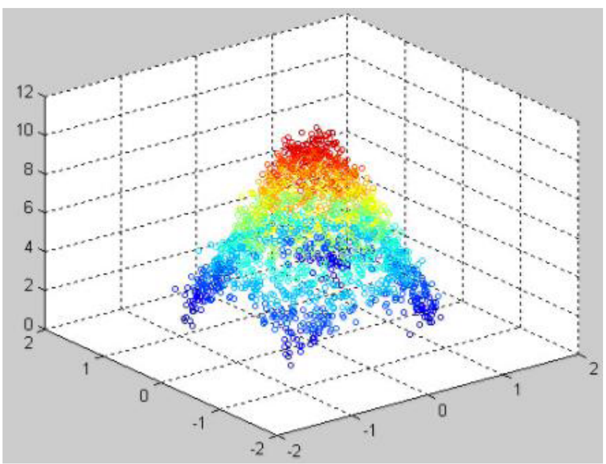


FIGURE 8. Effective information extraction, filtering and mining results.

IV. EXPERIMENTAL DESIGN AND ANALYSIS

A. EXPERIMENTAL DATA AND EVALUATION INDICATORS

The data set used in this experiment is the internal test data of the Friends of the system. A total of 889 users, including 659 normal users and 130 spammers, crawled 80,000 of these microblog posts. Then select the microblog information for marking, mark the microblogging as spam or normal information, a total of 60,000 marking microblogs, as a standard data set. In order to evaluate the performance of classification filtering, this experiment is mainly verified by two large-direction information experiments. The first big direction is the accuracy of the classification filter. Since spam filtering is a two-point classification problem, there are four cases as shown in Table 2.

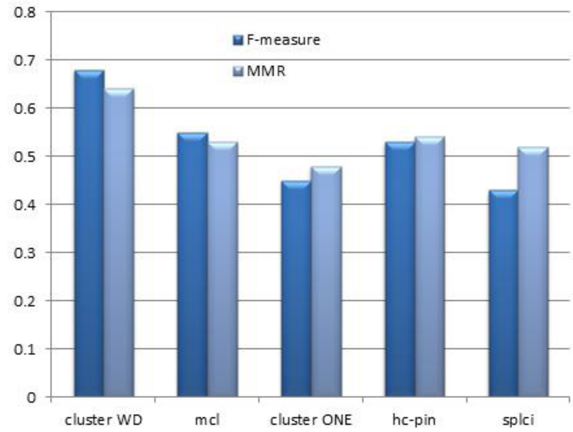


FIGURE 9. Performance comparison test.

TABLE 2. Confusion matrix for two types of problems.

		Predicted category	
		+	-
Actual category	+	TP	FN
	-	FP	TN

As can be seen from Table 2, the correct positive case TP and the correct counterexample TN are the correct classification samples, while the wrong counterexample FN and the wrong positive case FP are samples of the classification error, so the classification accuracy is usually defined as follows, which is the proportion of the correct classification of samples to all samples, can be expressed as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Error rate, which is the proportion of the sample with the wrong classification to the total number of samples, can be formally expressed as:

$$Errorate = \frac{FN + FP}{TP + FP + TN + FN}$$

B. EXPERIMENTAL PROCESS

The experimental data is published for the labeled social network users, and is divided into spam and normal information. After the word segmentation and feature extraction, the feature vector is obtained, and a standard data set is generated. Taking Test 1 as an example, the specific experimental results are as follows: (Number of samples: 800)

(1) Establish initial data set, sample increment set, performance test set: 200 of them as the initial data set do.txt,

400 samples as the sample increment set di. Txt, the remaining 200 samples are used as the final classifier performance test set dt.txt. There are no duplicate samples in the training set and the test set, and there is no intersection between the training set and the test set.

(2) The initial data set, the sample delta set, and the performance test set are scaled to obtain the respective sample files do.scale, di.scale and dt.scale; and the initial data set do.txt is used to train the initial The classification model model1.model tests the initial classification model with the test data set dt.txt.

(3) For the SVM-based retraining algorithm, merge the original data set do.txt and the incremental data set di.txt into the data set do1.txt, and train the new classification model model1.model with do1.txt. The data set dt.txt tests the classification model.

(4) For the standard SVM incremental learning algorithm, the situation after iteration is mainly considered. The incremental dataset di.txt is first classified using the original classification model model2.model. Then, according to the correctness of the classification, the erroneous data in di.txt and the SV set of the initial prediction model can be combined into the data set do2.txt, and the new classification model model22.model is trained with do2.txt, and the test data set dt can be used.

(5) For the improved SVM incremental learning algorithm, follow the steps similar to the standard incremental learning algorithm, and filter the data in di.txt one by one according to whether the KTT condition is violated, and the data in the di.txt that violates the KTT condition and The SV set of the initial prediction model is the data set do3.txt, the new classification model model3.model is trained with do3.txt, and the classification model is tested with the test data set dt.txt. A total of three tests were performed, the number of samples being: 800 Samples (increment of 400), 2500 samples (increment of 1500), and 50,000 samples (increment 25,000). The test results of the three algorithms under these three experimental conditions are shown in Table 3, Figure 10 and Figure 11:

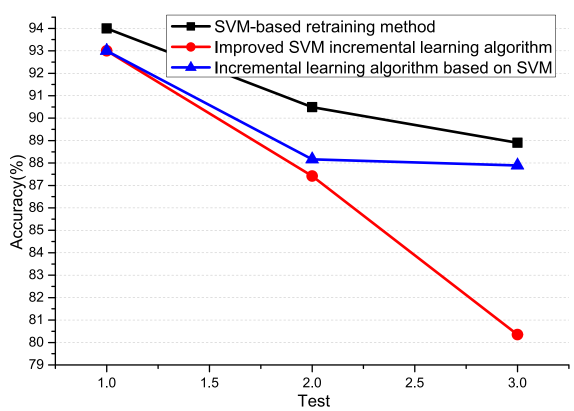


FIGURE 10. Algorithm test accuracy comparison.

TABLE 3. Algorithm performance comparison.

		Number of iterations	Numbe r of SV	Test accuracy (%)
SVM-based retraining method	Test1	263	142	94%
	Test2	1027	892	90.4881%
	Test3	21470	12567	88.9027%
Incremental learning algorithm based on SVM	Test1	184	108	93%
	Test2	796	578	87.4196%
	Test3	1006	835	80.3529%
Improved SVM incremental learning algorithm	Test1	158	83	93%
	Test2	503	378	88.1646%
	Test3	619	399	87.8903%

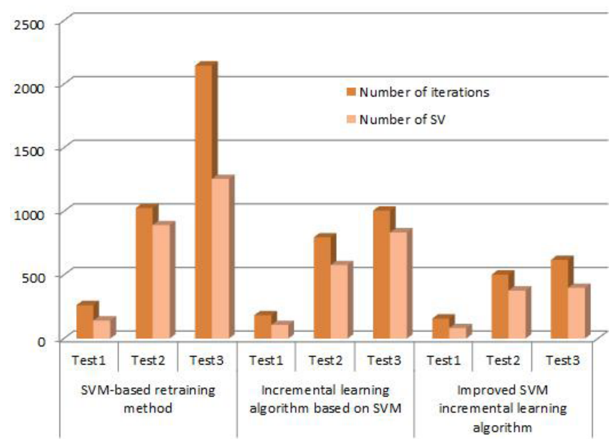


FIGURE 11. Comparison of algorithm test cases.

C. FILTER PERFORMANCE ANALYSIS

Firstly, from the accuracy analysis of classification filtering, in order to clearly and clearly discover the data change law and the performance comparison between different algorithms, according to Table 3, the accuracy ratio is compared with Figure 12, where algorithm 1 is a SVM-based retraining algorithm, and algorithm 2 is a standard. SVM incremental learning algorithm, algorithm three is an improved SVM incremental learning algorithm.

Combined with Figure 5 and Table 3, it can be found that when the test sample point is 800, the accuracy of the three algorithms is good, 94% and 93%; as the number of samples increases, the correct rate of the standard SVM incremental learning algorithm Gradually worse than the other two algorithms; until the number of samples increases to 50000, the accuracy of the standard SVM incremental learning algorithm is lower, and the accuracy of the improved SVM incremental learning algorithm is not as good as the retraining

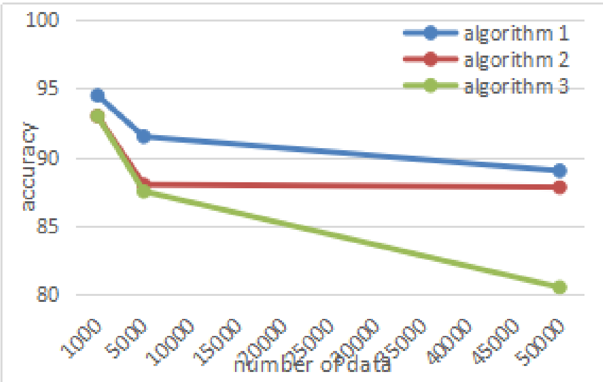


FIGURE 12. Comparison of algorithm accuracy.

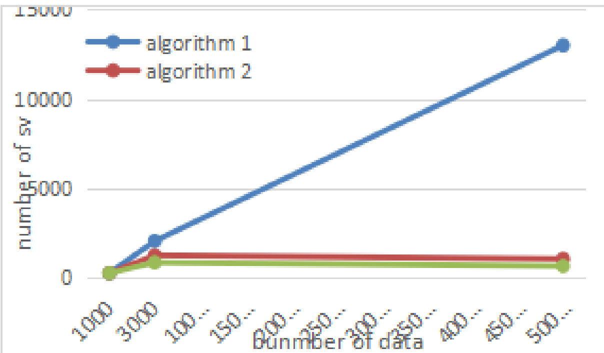


FIGURE 13. Comparison of algorithm and sv quantity.

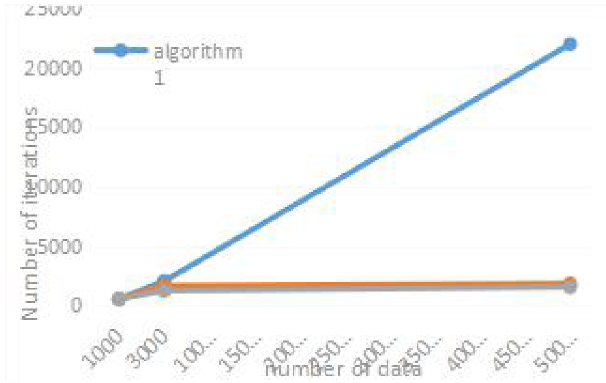


FIGURE 14. Iteration number comparison chart.

algorithm, but it is similar, keep A certain correct rate. This shows that the improved algorithm is suitable for spam filtering of many users and high Weibo concurrent. Secondly, from the performance analysis of classification filtering, draw the sv quantity comparison Figure 13 according to Table 3.

From the number of support vector SVs, the SVM-based retraining algorithm yields a much larger SV than the other two algorithms, which means that the retraining algorithm takes more time and space; the improved SVM incremental learning algorithm is based on The SVM retraining algorithm and the standard SVM incremental learning algorithm significantly reduce the number of SVs, which saves time and space.

Then compare the number of iterations according to Table 3 to Figure 14.

It can be found that the number of iterations of SVM-based retraining increases sharply with the increase of the number of samples, while the standard SVM incremental algorithm increases with the increase of the number of samples, but the increase is not fast, and the improved SVM incremental learning algorithm increases most. Slow, not subject to a sharp increase in the number of test set samples, a lot of changes, so it is very suitable for the filtering of a large number of users and microblog information.

In summary, it can be concluded that the spam filtering system based on the improved SVM incremental learning classification proposed in this paper has better classification accuracy, and the number of SV and the number of iterations of the key classification algorithm grows slowly. Good time and space complexity. At the same time, it should be noted that as the information to be filtered increases, the performance of the algorithm is more prominent. Therefore, the spam filtering system based on the improved SVM incremental learning classification is very suitable for spam filtering with super-user and high information concurrent.

V. CONCLUSION

Based on the analysis of existing information filtering technology, this paper chooses the filtering method based on machine learning according to the application environment and needs. In the process of machine learning, the model changes need to be retrained, and the method of SVM incremental learning is studied. An improved SVM incremental learning algorithm is proposed by comparing the incremental learning methods in SVM algorithm. The algorithm will violate the KTT condition as the classification basis of the incremental data set, and optimize the improved algorithm to make it more efficient to classify the SVM incremental set. Finally, the improved SVM incremental learning algorithm is applied to the spam filtering system to obtain better filtering results. By analyzing the actual needs and advantages of social network commodity recommendation, a hybrid recommendation system based on social network is proposed. Based on the user's social network, N users with the highest similarity to the recommended users are mined. At the same time, considering the user's possible commodity demand contained in the user message of the social network, the user's message is mined, and the mined information is applied to the recommendation to obtain a better recommendation effect.

In this paper, an improved algorithm for mining and filtering large data in the Internet of brain Things is studied, which can get effective information features from massive data, and data acquisition and data management are carried out in combination with the Internet of brain Things technology. In the environment of the Internet of brain Things, it is necessary to filter large data effectively and improve the ability of mining and identifying useful data. It is of great significance to study the filtering and mining algorithm of large data in the environment of Internet of Things for improving the control and recognition ability of data. In this paper, an effective information filtering algorithm for big data in Internet of brain Things is proposed based on support vector machine. Firstly, the large data system model of the Internet of brain Things is constructed and feature extraction is carried out. The correlation dimension feature extraction of the effective information feature is preprocessed, and the filter non-correlation information is designed to filter reasonably. Support Vector Machine (SVM) algorithm is used to mine the valid information features of data. The simulation results show that the proposed algorithm has high precision and superior performance in filtering large data valid information.

REFERENCES

[1] F. Ali, K.-S. Kwak, and Y.-G. Kim, "Opinion mining based on fuzzy domain ontology and support vector machine: A proposal to automate online review classification," *Appl. Soft Comput.*, vol. 47, no. 1, pp. 235–250, Oct. 2016.

[2] L. Li, K. Wu, and D.-W. Zhou, "Extraction algorithm of mining subsidence information on water area based on support vector machine," *Environ. Earth Sci.*, vol. 72, no. 10, pp. 3991–4000, 2014.

[3] B. Zheng, S. W. Yoon, and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1476–1482, 2014.

[4] P. Liu, J. Sun, L. Han, and B. Wang, "Research on the construction of macro assets price index based on support vector machine," *Procedia Comput. Sci.*, vol. 2014, pp. 1801–1815, 2014.

[5] S. Ding, Y. Han, J. Yu, and Y. Gu, "A fast fuzzy support vector machine based on information granulation.," *Neural Comput. Appl.*, vol. 23, no. 1, pp. S139–S144, Dec. 2013.

[6] S. Liu, L. Xu, L. Longqin, and L. Zeng, "Online prediction for dissolved oxygen of water quality based on support vector machine with time series similar data," *Trans. Chin. Soc. Agricult. Eng.*, vol. 30, no. 3, pp. 155–162, 2014.

[7] S. Das and S. Saha "Data mining and soft computing using support vector machine: A survey," *Int. J. Comput. Appl.*, vol. 77, no. 14, pp. 40–47, 2013.

[8] J. Shu, S. Liu, L. Liu, L. Zhan, and G. Hu, "Research on link quality estimation mechanism for wireless sensor networks based on support vector machine," *Chin. J. Electron.*, vol. 26, no. 2, pp. 377–384, Mar. 2017.

[9] H. Gao, W. Chen, and L. Dou, "Image classification based on support vector machine and the fusion of complementary features," *Comput. Sci.*, vol. 34, no. 9, pp. 1600–1607, 2015.

[10] C. C. Yang, A. Chung, C. Y. Ku, L. M. Brill, R. Williams, and D. A. Wolf, "Research and application of wind short term power based on support vector machine algorithm," *J. Shanghai Univ. Electr. Power*, vol. 3, no. 1, p. 115, 2013.

[11] "Support vector regression based on grid-search method for short-term wind power forecasting," *J. Appl. Math.*, vol. 2014, nos. 1–4, pp. 1–11, 2014.

[12] Y. Chen, Y. Yang, C. Liu, C. Li, and L. Li, "A hybrid application algorithm based on the support vector machine and artificial intelligence: An example of electric load forecasting," *Appl. Math. Model.*, vol. 39, no. 9, pp. 2617–2632, 2015.

[13] N. S. Raghavendra and P. C. Deka "Support vector machine applications in the field of hydrology: A review," *Appl. Soft Comput.*, vol. 19, no. 6, pp. 372–386, 2014.

[14] S. Liping, W. Panpan, H. Yongjun, and H. Li, "Broken rotor bar fault diagnosis of induction motors based on bare-bone particle swarm optimization and support vector machine," *Trans. China Electrotech. Soc.*, vol. 29, no. 1, pp. 147–155, 2014.

[15] M. E. Poorazizi, A. J. S. Hunter, and S. A. Steiniger, "A volunteered geographic information framework to enable bottom-up disaster management platforms," *ISPRS Int. J. Geo-Inf.*, vol. 4, no. 3, pp. 1389–1422, 2015.

[16] T. A. N. Kun, Y. E. Yuan-Yuan, and D. U. Pei-Jun, "Identification and classification of near-infrared spectrum of adulterated wine based on support vector machine," *Acta Photon. Sin.*, vol. 42, no. 1, pp. 69–73, 2013.

[17] Y. Wang, G. Ding, and L. Liu, "A regression forecasting model of carbon dioxide concentrations based-on principal component analysis-support vector machine," *Commun. Comput. Inf. Sci.*, vol. 2015, no. 482, pp. 447–457, 2015.

[18] X. F. Du, S. C. H. Leung, J. L. Zhang, and K. K. Lai, "Demand forecasting of perishable farm products using support vector machine," *Int. J. Syst. Sci.*, vol. 44, no. 3, pp. 556–567, 2013.

[19] H. Van Nguyen and F. Porikli, "Support vector shape: A classifier-based shape representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 970–982, Apr. 2013.

[20] M. A. Bo, H. Y. Dong, and R. Wei, "Status assessment of uhv substation secondary equipment based on fuzzy comprehensive support vector machine method," *Comput. Syst. Appl.*, vol. 12, no. 2, pp. 1561–1569, 2014.

[21] J. K. Laurila *et al.*, "From big smartphone data to worldwide research: The mobile data challenge," *Pervasive Mobile Comput.*, vol. 9, no. 6, pp. 752–771, 2013.

[22] Y.-F. Y. Chan *et al.*, "The asthma mobile health study, a large-scale clinical observational study using ResearchKit," *Nature Biotechnol.*, vol. 35, no. 4, pp. 354–362, 2017.

[23] K. A. Cook, B. D. Modena, and R. A. Simon "Improvement in asthma control using a minimally burdensome and proactive smartphone application," *J. Allergy Clin. Immunol., Pract.*, vol. 4, no. 4, pp. 730–737, 2016.

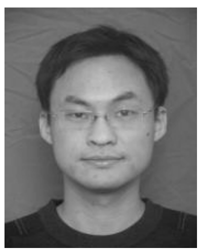
[24] M.-Y. Kim *et al.*, "Feasibility of a smartphone application based action plan and monitoring in asthma," *Asia Pacific Allergy*, vol. 6, no. 3, pp. 174–180, 2016.

[25] K. A. Haze and J. Lynaugh, "Building patient relationships: A smartphone application supporting communication between teenagers with asthma and the RN care coordinator," *Comput. Inf. Nurs.*, vol. 31, no. 6, pp. 266–271, 2013.

[26] G. Larkou, M. Mintzis, P. G. Andreou, A. Konstantinidis, and D. Zeinalipour-Yazti, "Managing big data experiments on smartphones," *Distrib. Parallel Databases*, vol. 34, no. 1, pp. 33–64, 2016.

[27] S. Morozov and H. Saiedian "An empirical study of the recursive input generation algorithm for memory-based collaborative filtering recommender systems," *Int. J. Inf. Decis. Sci.*, vol. 5, no. 1, pp. 36–49, 2013.

[28] D. Nelson, L. Macks, G. Paine, F. Poursoltan, T. Mekonnen, and M. L. Hartzler, "Efficacy of smartphone technology on improving asthma control and compliance," *PLoS Pathogens*, vol. 7, no. 10, pp. 515–534, 2012.



XIAOFENG ZHAO was born in Hebei, China, in 1978. He received the bachelor's degree from the Hebei University of Technology, in 2000, and the master's degree from the Hefei University of Technology, in 2008. He was an Engineer with Hansteel, from 2000 to 2004. Since 2008, he has been a Lecturer with the School of Management Engineering and Business, Hebei University of Engineering. He has published 12 papers. His research interests include artificial intelligence, information management, data mining, and distributed computing.

WEIHUA ZANG was born in Hebei, China, in 1977. He received the bachelor's degree from Hebei Normal University, in 2000, and the master's degree from the Hebei University of Engineering, in 2011. He was a Teacher with Handan University. From 2007 to 2008, he was a Visiting Scholar with the College of Computer Science and Technology, Zhejiang University. He has published several papers on computer networks, computer applications, and embedded systems.

RULAN LV was born in Hebei, China, in 1992. She received the bachelor's degree from the Huaxin College, Shijiazhuang University of Economics, in 2014, and the master's degree from the Hebei University of Engineering, in 2018. Since 2014, she has been an Assistant Engineer with the Hebei Branch of Construction and Administration Bureau of South-of-North Water Diversion Middle Route Project. Her research interests include the system of information service, video monitoring, and SOA.

WEI CUI was born in Hebei, China, in 1985. She received the bachelor's degree from Vancouver Island University, in 2010, and the master's degree from the Hebei University of Engineering, in 2015. She has been a Lecturer with the Hebei University of Engineering. Her research interests include business management, marketing management, and logistics management.