

Evaluation of Eigen Value Extraction and Automatic Computational Classification of Corpus in Scene Scheduling of Sports and Health IoT Platform

Yanjun Zhang

Jinan University, Guangzhou 510000, China

Abstract

This study constructed a corpus by collecting samples of applications in exercise health information. Based on the extracted 8 eigen values, this study used the Fisher score algorithm to score the eigen values and then formed a combination of 8 groups of eigen values according to the score ranking. This study performed computerized automatic classification using cross-checking based on three methods, support vector machines, neural networks, and Naive Bayes model. By analyzing the experimental results, the optimal eigen value combination and the optimal classification algorithm were finally derived in this study.

Keywords: Exercise health, eigen value, support vector machine, neural network, naive bayes model

I. Introduction

Active exercise and rehabilitation are an important part of today's sports medicine prescription health care research. It is a scientific and feasible exercise program based on the actual health condition of the human body, aiming at the ultimate health care and therapeutic purposes. However, those current equipment and data management methods for national fitness, amateur sports fans, and patients with chronic diseases have the problem of single parameter and poor accuracy; while the equipment and data management methods for those used by professional athletes are highly accurate, but these methods are loaded and costly and cannot be popularized. Therefore, the development of sports health equipment and data management for the general public and the provision of effective market-oriented services have an urgent social demand. This also provides an opportunity for the industrialization of sports health. Especially in the more economically developed regions such as the Pearl River Delta, the sports health industry has a broad space for innovation and entrepreneurship [1-3].

Nowadays, the fast pace of modern social life, the lack of exercise in the urban population, coupled with people's incorrect lifestyles are the main causes of various urban populations in a state of the high incidence of sub-health or occupational diseases. For people, the lack of an appropriate amount of moderate exercise may not only lead to obesity but also has a close cause of diabetes, osteoporosis, and metabolic disorders (see Table 1).

Table 1 Condition of the associated disease and means of sports and exercise intervention

Disease Type	Disease Name	Basic information on chronic diseases	Current exercise assessment and intervention techniques
Metabolic diseases	Obesity	The mortality rate for 20% overweight is 128%, and for 35-40% overweight is 151%. In China, the proportion of overweight and obesity has reached more than 20% among adults and more than 30% among urban residents, and the proportion of obesity among children is also climbing sharply.	There are no clear prevention techniques but basically rely on a controlled diet and physical exercise.
	Diabetes	The prevalence of diabetes is about 5%. In China, the number of people with diabetes is about 40 million, and	Proper exercise, regular checkups, and a sensible diet; using specialized shoes or insoles for intervention.

		the number of people with diabetes increases by more than 1.2 million each year. Of these, about 15% of diabetic patients may develop diabetic foot disease, and about 80% of diabetic patients may face amputation.	
Cardiovascular diseases	High blood pressure	The prevalence of high blood pressure is about 18.8%. At present, there are 150 million hypertensive patients in China, and the prevalence of the patients is increasing, among which the young population is more obvious than the elderly; among the people aged 35 to 44, the prevalence rate of high blood pressure is 74% for men and 62% for women.	Regular blood pressure measurements and diet control are performed. Exercise therapy can significantly delay and improve the organ damage caused by high blood pressure.
Orthopedic diseases	Cervical spine lumbar spine osteoporosis	At present, the number of cervical spondylosis patients in China has reached 50 million, and the number of patients is increased by 1 million every year. The incidence of cervical spondylosis is on the trend of low age, with more than half of the patients under 40 years old, mostly in white-collar workers who are absorbed in a work; the incidence of osteoporosis in people over 60 years old is 22.6%.	Patients should frequently move their cervical spine, avoid prolonged sitting and absorbed in a work, avoid incorrect sitting and sleeping posture, and normally carry out exercise and physical exercise.
Neurological diseases	Parkinson's disease	In China, the prevalence of Alzheimer's disease is 1%, and there are about 12.9 million people with the disease. Alzheimer's disease has the fourth-highest number of deaths among all diseases. Among them, there are about 1.7 million patients with Parkinson's disease.	There are no clear prevention techniques, but the main focus is on fall prevention, follow-up, and exercise in the company of others.

II. Extraction of Eigen Value

In this paper, virtual reality technology will be used as a carrier to conduct a comprehensive analysis of users' exercise physiology and potential diseases by combining body networks and cloud-based big data. At the same time, a virtual training scenario is designed to intervene in advance for the user. The specific research plan is as follows [4]:

(1) Multimodal health information collection

The sports health assessment platform is proposed to implement multimodal information such as motion, physiology, and images. The motion information acquisition mainly includes plantar pressure distribution measurement module and motion capture module of optical marker points. The physiological information acquisition mainly includes the exercise ECG measurement model, the blood oxygen module, and the respiratory motion measurement module. In order to accurately measure the small changes in pressure, the pressure sensing system designed in this paper will use PVDF piezoelectric membrane as the sensor for plantar pressure collection. And in the motion monitoring link, this paper uses a lumbar support device for fixing the waist. When the user wears the device, the motion will be limited to a certain range of activity. At the same time, there is a circle of infrared transmitting and receiving sensors around the chassis under the feet, forming a light array. When the user moves in it, the infrared light at the corresponding location will be received or blocked so as to achieve the purpose

of motion detection. Moreover, a virtual reality headset was used in this study to obtain the user's steering data and further obtain the direction of the user's body movement. In this study, we used electronic fabrics for non-contact continuous ECG monitoring, a multi-channel bioelectrical impedance technique to collect respiratory signals from different parts of the body, and a cuffless approach based on pulse conduction time for continuous blood pressure monitoring [5].

(2) Individualized exercise assessment and intervention

In this paper, we adopt the training scenario constructed by VR technology to realize movement assessment and intervention, study ergonomic human-computer interaction training scenarios, explore the scientific training method of movement gait based on VR, and evaluate the method and scenario. In terms of the specific implementation, we conduct balance training for people prone to fall, personalized sports gait guidance for people with deformed plantar pressure distribution, and proposed biofeedback intervention for people with motor function caused by cardiovascular diseases, such as heart rate variability feedback and respiratory feedback [6].

Virtual reality interaction technology can turn virtual objects into something we can touch that you can push, grab, or even squeeze them. It captures real objects in mutual space and uses 3D modeling to achieve high-quality "virtual clones" that become accessible to users, reducing the distance between the people, places, and activities that matter in life. With the help of this technology, we can achieve more amazing experiences in the virtual world, such as synchronizing realistic walking and other movements to the virtual space. Device sensors can quickly and accurately identify the direction and speed of the user walking. An interactive cervical spine motion intervention system is developed by integrating inertial sensors and computer graphics technology, and its focus is on the design of ergonomic interaction links and the evaluation of intervention effects. Additionally, a nonlinear heart rate variability approach was used to assess visceral neurological and cardiovascular status, and a personalized fuzzy expert feedback algorithm was designed for the study of respiratory exercise intervention techniques that focused on the setting of respiratory intervention patterns and the evaluation of intervention effects [7].

(3) Cloud-based technology for big data services

The plantar pressure data, ECG, respiratory, and potential disease analysis generated by the user during use will be automatically collected and permanently stored on the cloud platform, which can be accessed by the user and his/her health care practitioner at any time. Furthermore, by building computing models and databases on the cloud platform, dynamically analyzing historical data, and continuously training and optimizing the algorithm models, the platform can push health reports and trend analysis to users periodically and output better exercise intervention programs and evaluation mechanisms [8].

(4) Integrated storage of exercise and health information

Construct a sports health platform that integrates multimodal health information collection, personalized sports assessment and intervention, and a big data service platform with cloud technology as one. And build a cloud data platform to be applied on Ali cloud. The platform collects user data to the cloud platform through cell phone APP, WeChat, and other mobile terminals for subsequent analysis and processing, and service provision.

This study ultimately defines the combination of key parameters of health information to be recorded. The rich text stored in the daily exercise was built into a corpus, and then the corpus was computed. A total of 8 items, including body temperature, blood oxygen, heart rate, pulse rate, muscle strength, and parameters containing the corresponding body temperature, blood oxygen, heart rate, pulse rate, and muscle strength variables, extracted after the computation are used as eigen values to construct feature vectors.

III. Eigen Value Assessment

3.1 Optimal assessment method

The optimization of eigen value has been a topical issue. The main approach is to select the eigen value that has the greatest impact on the regression results and classification results from the many characteristics of a sample, etc. At present, there are few studies on this method in China, while the major studies on it globally are as follows:

(1) Fast correlation based filter (FCBF)

The fast correlation filter developed based on the study of correlation and redundancy of eigen values is a

symmetrical, uncertain value used to evaluate the redundancy relationship between two eigen values and two categorical labels, which the definition refers to as SU (Symmetrical Uncertainty). This symmetric, uncertain value is based on the entropy value and the nonlinear correlation measure. Suppose there are two variables X and Y. Then the value of SU is calculated on the basis of the variables X and Y. X and Y are any two eigen values, and the value of SU is the information gain of the variable X after the observation of the variable Y.

(2) MRMR algorithm

The mRMR algorithm is an eigen value extraction algorithm and belongs to an algorithm that selects the highest relevant eigen values and tries to minimize their redundancy from the selected eigen values. The mRMR is used to estimate the scores between eigen value and eigen value and between eigen value and classification labels with common information.

(3) Fisher score algorithm (FS)

Fisher score is an effective and simple method to quickly find the specificity between eigen values. The idea is to transform the data in a high-dimensional space into another low-dimensional space so that the variance is as large as possible while the variance is as small as possible, and the distance to the eigen value can be projected farther depending on the label. Fisher score is to evaluate the ability of eigen value to maintain the data structure, and Fisher score is supervised feature selection.

(4) T-test algorithm

The t-test is one of the widely used methods to determine whether the means between two samples or eigen values are dissimilar.

(5) ReliefF algorithm (RF)

ReliefF is mainly derived from the Relief algorithm, which extends on two problems, so it can handle different classes of problems. ReliefF belongs to an eigen value weight filtering model, which is self-supervised and can estimate its weight based on the eigen value, and when the value reaches a critical threshold, the eigen value is removed.

3.2 Fisher score for optimal sorting

In this paper, after sorting the original data and eigen value extraction, the dataset has 10,000 samples with 8 feature attributes, and the eigen value combinations from serial number 1 to 8 are formed according to the gradual increment from 1.

The Fisher score algorithm was chosen to rank the 10000×8 eigen value matrix based on the study of the international general optimal eigen value selection algorithm. The results show that most of the top-ranked eigen values are related to the data of heart rate and blood oxygen within the normal atmosphere, which also indirectly proves that heart rate and blood oxygen within the normal range is the physiological manifestations of healthy exercise in exercise health monitoring in Table 2.

Table 2 Optimal eigen value ranking

Eigen value No.	Score	Ranking
NO.1	5.88	1
NO.4	4.56	2
NO.2	2.15	3
NO.6	2.03	4
NO.3	1.98	5
NO.8	1.06	6
NO.7	0.95	7
NO.5	0.32	8

IV. Automatic Classification of Sports Health Information

ISSN: 0010-8189

© CONVERTER 2021

www.converter-magazine.info

In the experiments of this paper, 10,000 sports health information are extracted, and the information is extracted by the corpus calculation method. Data with a series of keywords are extracted, and the keywords are body temperature, blood oxygen, pulse, upper limb muscle strength, lower limb muscle strength, etc. Three different classification algorithms are applied in the experiments for classification detection of health information: support vector machines, neural networks, and Bayesian.

4.1 Support vector machines

Support vector machines (SVM) is a machine learning method proposed by Vapnik in 1955. SVM is based on the VC dimensional theory of statistical learning theory and the theory of structural risk minimization, and its learning strategy is interval maximization, so SVM is also called the maximum edge algorithm. Since the SVM has strong generalization ability, it has many advantages in many small samples and high-dimensional pattern recognition and can be applied to other machine learning fields such as function fitting (see Figure 1).

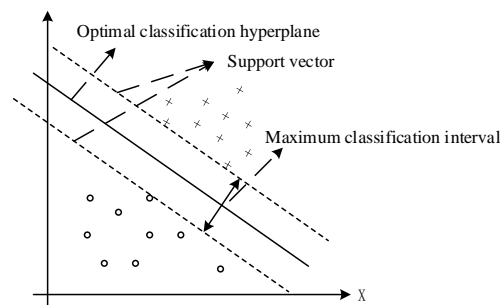


Figure 1 Mathematical model of SVM classifier

The main approach of support vector machines is to solve the separated hyperplane that divides the classified data set with maximum geometric interval. The basic concept of support vector machines is to solve the separated hyperplane that can correctly partition the categorical dataset with maximum geometric separation. Interval maximization means classifying the training data with sufficiently large confidence, that is, separating not only the positive and negative instance points but also the instance points closer to the hyperplane with sufficiently large confidence to separate them. Such a hyperplane also has better classification prediction capability for unknown new instances.

Since the sample data set in this paper is nonlinear and does not belong to the range of linear separability, it is necessary to transfer the low-dimensional sample points into the high-dimensional space by means of kernel function mapping to generate several different optimal classification hyperplanes for multi-dimensional plane classification.

In this paper, we use support vector machines of grid search method, (C, g) (assumed to be the two-parameter settings in libsvm), and the value of (C, g) can not be determined if no training or testing is done on the sample data. The grid search method is to make the number of parameters C taken as N and the number of g as M . Then, $M \times N$ combinations of (C, g) are trained by different support vector machines so as to compare their learning accuracy and finally to obtain the highest learning accuracy among all the learning accuracy results.

The biggest advantage of the grid search method is that it can be trained for different SVMs since the values taken for each (C, g) are independent of each other and the SVM mechanism, which ensures efficiency and accuracy. However, at the same time, the grid search method is very computationally intensive, so a grid search is very time-consuming.

The support vector machines applied in this paper add a 5-fold cross-check. By partitioning the data into K copies, and other $K-1$ copies as training sample data and 1 copy as test sample data, and the error results of the learning

process can be evaluated to obtain the average value of the error occurrence, the division of the training sample data set and the test sample data set generally use equal average distribution or random division method.

In the early stage, after the selection of the eigen value most optimal choice, 8 new eigen value sample groups will be formed, and each combination is sequentially increasing. The 8 eigen values are sorted according to the scoring results to form 8 sets of sample data sets, and the grid search support vector machines are applied to classify the 8 sets of sample data sets and construct the error recognition number matrix. The algorithm implementation steps are as follows [9,10]:

- (1) Read the sample dataset and perform cross-testing, with 4/5 of the dataset as training samples and 1/5 as test samples.
- (2) Normalize the values of all eigen values for the interval [0-1].
- (3) Roughly select the parameters of the kernel function of support vector machines and select the rbf kernel function.
- (4) Perform another optimal parameter selection based on the results of the rough selection.
- (5) Use SVM for training and testing.
- (6) Output classification accuracy and error recognition matrix.

Group a was cross-validated five times with an average classification accuracy of 66.23%. The average accuracy of each of the eight selected groups of eigen values is shown in Table 3.

4.2 Neural Networks

Neural network algorithms mimic biological neural networks and are a class of pattern matching algorithms that are commonly used to solve classification and regression problems. There are hundreds of different algorithms for neural networks, among which the important ones are: perceptron neural network (PNN), backpropagation (BP), self-organizing mapping (SOM), learning vector quantization (LVQ), etc. In this paper, a backpropagation (BP) neural network is used for malware classification detection. The concept of back propagation neural network (BP) is to use the error after the output to estimate the error of the previous layer of the output layer, and then use the error of this layer to estimate the error of the previous layer, so as to obtain all the error estimates. The connection weights of each layer are adjusted according to the error, and then the output error is recalculated with the adjusted connection weights until the output error meets the requirement or the number of iterations overflows the set value. BP neural network basic structure diagram is as in Figure 2:

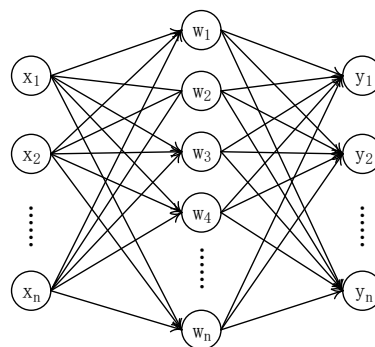


Figure 2 BP input layer, implicit layer, and output layer

As in support vector machines, it was divided into 15 data sets, and based on the neural network classification experience, more than 1/2 samples were selected as neurons to train the classifier, and the remaining data were test samples, and the results of the BP neural network algorithm for malicious application classification detection are shown in Table 3.

Group a was cross-validated five times and the average classification accuracy shown was 60.04%. The average

accuracy of each of the eight groups selected for eigen value is shown in Table 3.

4.3 Bayesian Networks

Bayesian network (also known as Belief Network) is a probabilistic network based on probabilistic inference, which is proposed to solve the problem of uncertainty and incompleteness. A Bayesian network is a directed acyclic graph consisting of nodes and directed edges connecting the nodes, where each node represents a feature, and the directed edges between nodes represent dependencies or causal relationships between features. The direction of the directed edges represents the directionality of the dependency or causality, and the unconnected features are conditionally independent of each other. The structure of the Bayesian network classifier is more complex than the Naive Bayes model, not only considering the conditional independence between features but also considering the dependency relationship existing between features. Therefore, once the Bayesian network classifier is trained, it can achieve better results than the Naive Bayes model classifier.

Bayesian network basic structure diagram is as in Figure 3:

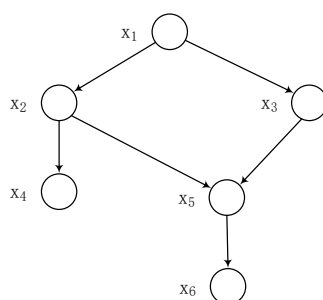


Figure 3 Bayesian network

Results of Bayesian network algorithm for malicious application classification detection: group a was cross-validated five times, and the average classification accuracy shown was 58.01%. The average accuracy of each of the selected 8 groups of eigen value is shown in Table 3.

V. Analysis of Experimental Results

By applying the above 3 classifiers, experiments were conducted in the order from a~h (8 eigen value combinations), 5 cross-validations were performed for each experiment, and the average values were taken, and the experimental results are shown in Table 3 below.

Table 3 Classification results of each classifier corresponding to the combination of eigen value

Group No.	Grid Support vector machines classifier	BP neural network classifier	Naive Bayes model classifier
a	66.23%	60.04%	58.01%
b	67.54%	62.35%	65.02%
c	70.01%	58.03%	63.02%
d	66.35%	63.25%	65.36%
e	70.22%	72.35%	57.98%
f	71.35%	65.32%	66.35%
g	70.22%	65.50%	67.01%
h	73.01%	66.01%	65.35%
Mean value	69.37%	64.11%	63.51%

(1) For the sample dataset of 10000×8 in this paper, the highest classification accuracy is obtained for the grid support vector machines classifier with an accuracy of 69.37%, and the lowest is for the Naive Bayes model

classifier with an accuracy of 63.51%.

(2) For the 8 eigen values, after the weight update calculation, it was found that the eigen value with the highest impact and the highest staging correlation after sorting was column 1.

(3) The two eigen values that have the greatest impact on the staging of the grid support vector machines classifier and the BP neural network classifier are the eigen values of the 1st and 4th scores, while the eigen value of the 3rd score has an impact on the Naive Bayes model classifier, and the eigen value of the 8th score has no impact on it. The eigen value of the 8th score has no effect on it.

(4) After forming 8 sets of sample data sets a~h from 8 eigen values after optimal sorting, it is found that the best sample set for grid support vector machines is set h, the best sample set for BP neural network is set e, and the best sample set for Naive Bayes model is set g.

VI. Conclusion

In this paper, three different machine learning methods, support vector machines, neural networks, and bayesian networks, were used to classify and detect sports health information, and the classification results were evaluated in terms of detection rate, false alarm rate, and classification accuracy. It could be seen that the support vector machines method has some advantages in the classification detection of sports health. In future research work, various machine learning algorithms will be further improved to better apply to the classification detection of sports health.

Acknowledgments

This paper is funded by the project of Guangzhou Education Bureau that the construction of innovation and entrepreneurship platform for Sports Health Industry under IAB vision, project No. is 2020PT203.

Reference

- [1] B. Elhoussaine. "A health remote monitoring application based on wireless body area network," International Conference on Intelligent System and Computer Vision, 2018.
- [2] K. Prathamesh "Wearable smart health monitoring system for animals," 2nd International Conference on Trends in Electronics and Information, 2018.
- [3] O.T. Emre, "An android application for geolocation based health monitoring," Consultancy and Alarm System, IEEE 42nd Annual Computer Software and Applications Conference, 2018.
- [4] "World Health Organization," The Word Health Report 2003: Shaping the Future, http://www.who.int/whr/2003/en/overview_en.pdf, 2010.
- [5] C.F. Lin, S.D. Wang, "Fuzzy support vector machine," IEEE Ransacion on Neural Networks, vol. 13, pp. 464-471, 2002.
- [6] C.F. Lin, S.D. Wang, "Fuzzy support vector machines with automatic membership setting," Studies in Fuzziness and Soft Computing, vol. 177, pp. 233-254, 2005.
- [7] M.J. Li, M.K. Ng, Y.M. Cheung, J.Z. Huang, "Agglomerative fuzzy k-means clustering algorithm with selection of number of clusters," IEEE Trans Knowledge and Data Engineering, vol. 20, pp. 1519-1534, 2008.
- [8] H.C. Huang, Y.Y. Chuang, C.S. Chen, "Multiple kernel fuzzy clustering," IEEE Transactions on Fuzzy Systems, vol. 20, no. 1, pp. 120-134, 2012.
- [9] S.B. Haitham, H. Sabah, "Hand posture and gesture recognition technology," Neural Computing and Applications. vol. 2014, no. 3-4.
- [10] L. Wang, J.L. Kenneth, "Wearable carbon nanotube-based fabric sensors for monitoring human physiological performance," Smart Materials and Structures, vol. 2017, no. 5.