

Application of data mining technology and wireless network sensing technology in sports training index analysis

Liqu Qian

Shandong Technology and Business University

Jiatong Liu (✉ xianquta1983p@163.com)

zhejiang wanli university

Research

Keywords: Wireless network, Data Mining, Index Parameters, Training Analysis Mechanism, Sports Training

Posted Date: March 17th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-17559/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at EURASIP Journal on Wireless Communications and Networking on June 9th, 2020. See the published version at <https://doi.org/10.1186/s13638-020-01735-z>.

Application of data mining technology and wireless network sensing technology in sports training index analysis

Liqui Qian¹, Jiatong Liu^{2*}

¹ Physical Education Department of Shandong Technology and Business
University, Yantai 264005, China

^{2*} Physical Education Department of Zhejiang Wanli University, Ningbo
315100, China

Email: xianquta1983p@163.com

Abstract: The conventional analysis method can provide a general analysis of sports training index, but its ability is relatively low when analyzing niche data. To solve this problem, this paper proposes data mining technology. First, the indicator parameter classification is determined, then the data mining technology is imported, the sports training analysis mechanism is established through this technology, and the construction of the index analysis model is completed. The model is used to analyze the process of niche data mining, and effective data of training indicators are obtained. Deep learning is a method of machine learning based on representation of data. Through the coverage test, accuracy test and immunity test, the variable parameters of the comprehensive analysis capability are determined. Further calculation of this parameter shows that the comprehensive ability of the data mining application analysis method is improved by 37.14% compared with the conventional method, which is suitable for analysis of niche sports training indicators of different data types.

Keywords: Wireless network; Data Mining; Index Parameters; Training Analysis Mechanism; Sports Training

1 Introduction

The conventional index analysis method adopts statistical method to make a general analysis of sports training indicators. When analyzing niche data, due to the small amount of statistical data, conventional methods have the disadvantage of low comprehensive analysis capabilities[1,2]. Therefore, the application of data mining technology in the analysis of sports training index is proposed in this paper. According to the characteristics of the data set, the classification of index parameters was determined, then the sports training analysis mechanism was established by importing data mining technology, and the analysis model was constructed. By analyzing the three processes of data preparation, data mining and result interpretation, the data mining results of training indicators are obtained and data analysis is completed. Finally, the data mining technology is applied in the analysis of sports training indicators. In the simulation test environment, two different methods were used for coverage test, accuracy test and immunity test to obtain variable parameters for comprehensive analysis. Through the calculation and comparison of this parameter, it can be seen that the analysis method proposed in this paper has a very high effectiveness.

The specific contributions of this paper include: 1) This paper reviews the existing algorithms and analyzes their advantages and disadvantages. 2) This paper proposes a data mining technology combined with deep learning algorithm, and constructs a model for it. 3) A case study of applying

data mining technology to sports training index is presented. 4)The performance of the proposed algorithm is analyzed and compared with other existing algorithms.

The rest of this paper is organized as follows. Section 2 discusses related work, Secondly, the third section discusses the construction process of data mining model. Section 4 discusses the application of data mining technology to sports training indexes. Section 5 presents the simulation results and summarizes the future research directions.

2 Construction of the analysis model for sports training indicators

2.1 Import of Data Mining Technology

Determining the index parameters is based on the characteristics of the data set to find the concept description of the category which represents the overall information of such data, that is, the intension description of the category. The purpose of the classification is to analyze the input data. Through the characteristics represented by the data, the accurate description is found for each type of data, such description is often expressed as a predicate, and it is used to classify the subsequent data. Although the class labels for these data are unknown, their categories can still be predicted[3].

The classification can be described as follows: Given a set T of training data, where the element record is described by several attributes, there is only one attribute as a class attribute in all the attributes. This set is represented by a vector $X=(X_1, X_2, \dots, X_n)$, where X_i ($1 \leq i \leq n$) represents a non-category attribute and may have different ranges. When the value range of the attribute is continuous, the attribute is called a Continuous Attribute; otherwise it is called the Discrete Attribute. $C=\{C_1, C_2, \dots, C_k\}$ represents the data set with k different categories of attributes. Then the T determines a mapping function from vector X to C , that is, $H : f(X) \rightarrow C$. The purpose of the classification is to use data mining techniques to express the implicit function H . The expression of the function H is as follows [4]:

$$H = pH_0 + \log(a^n + ef) \quad (1)$$

Where the H represents the implicit function, the H_0 represents the initial state of the function, the p represents the defining attribute of the function, the a represents the range of the element record, the n represents the range of the condition, the e represents the range of the sports index and the f represents the discrete index of the sports index. Index parameter classification is generally divided into two steps: The first step is to import data mining technology through known data sets. The second step is to use the obtained model for the classification operation.

First, the accuracy of model classification is estimated. If the accuracy of the model is acceptable, the model can be used for classification. The first step, as shown in Figure 1, is to use the training data set for learning. Training sets are analyzed by classification algorithms to generate classification rules.

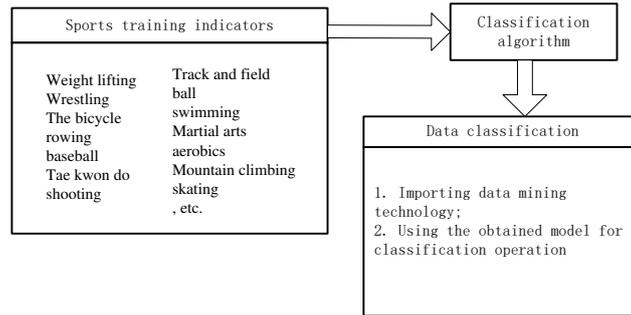


Figure 1 Learn using training set

Second, the test data is used to evaluate the model as shown in Figure 2. If the accuracy rate is acceptable, the classification rules will be used to classify the new data.

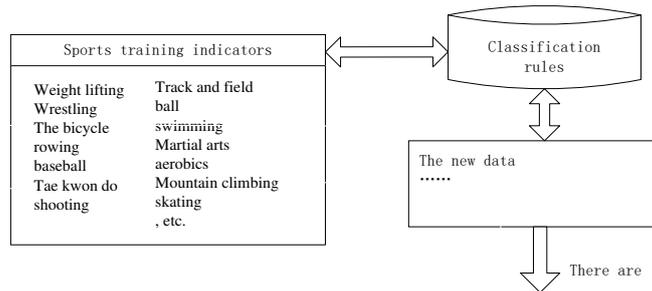


Figure 2 Evaluate the model with test data

Data mining methods are developed from artificial intelligence and machine learning methods, which combine statistical analysis methods, fuzzy mathematical methods and scientific computing visualization techniques. Data mining technology can be divided into the following six categories:

Inductive learning is currently the focus of research, and it is mainly divided into two categories: information theory and set theory. The information theory approach uses information theory to establish the decision tree. In the field of sports training analysis, the decision tree is a simple index representation method. It gradually classifies cases into different categories. This kind of method has good practical effect and great influence. Since the method finally obtains a decision tree, it is generally called a decision tree method. The more distinctive methods in information theory methods are: ID3 and IBLE method[5]. In recent years, due to the development of the rough set theory, the set theory method has been rapidly developed. This includes the coverage of positive exclusion exceptions (AQ method), the concept tree method and the rough set method. Their three relationships are shown in equation 2 [6].

$$A = \delta \sum_{i=1}^j [\psi \lambda]^2 \quad (2)$$

Typical biomimetic technology methods are neural network methods and genetic algorithms. These two methods have formed an independent research system and they have also played a huge role in data mining. The neural network method is based on the IP model and the Hebb learning rule, and the three types of neural network models are established. The neural network sports training index is a distributed matrix structure. Neural network learning is reflected in the gradual calculation of neural network weights. Using neural network techniques is particularly effective when it is difficult to obtain concepts from complex or inaccurate data. The trained neural network

is like an “expert” with some kind of specialized sports training indicators, so it can learn from experience like people [7].

Genetic algorithm is the algorithm that simulates the process of biological evolution. It consists of three basic processes of breeding, crossing and mutating. The algorithm has played a significant role in optimizing calculations and classifying machine learning.

From this it can be seen that certain mathematic operations on several data variables can get the corresponding mathematical formulas. The statistical analysis method uses statistical principles to analyze the data. It includes: common statistics, correlation analysis, regression analysis, variance analysis, cluster analysis and discriminant analysis[8].

The generation of fuzzy mathematics is due to the objective existence of ambiguity. And the higher the complexity of the system is, the stronger its ambiguity is. This is the principle of mutual grammar summarized by Zadeh. The fuzzy set theory can be used to make fuzzy judgments, fuzzy decisions, fuzzy pattern recognition, fuzzy association rules and fuzzy cluster analysis on practical problems. The expression of fuzzy mathematics method is shown in formula 3.

$$Pi(m) = \frac{P(Z_1 + Z_2)}{x_1 \cdot x_2} + \lambda \quad (3)$$

Where the $Pi(m)$ means fuzzy mathematics, P is the representation of the complexity of the system. Z_1 represents the fuzzy judgment, Z_2 is the fuzzy decision, x_1 represents the fuzzy pattern recognition, x_2 represents the fuzzy cluster, and λ represents the fuzzy association rule.

The visual data analysis technology broadened the traditional charting function and enabled users to analyze the data more clearly. For example, turning multi-dimensional data into a variety of graphics, which play a very strong role in revealing the inherent nature and regularity of the data. The purpose is to enable users to browse data and mining process alternately, and improve the effect of mining. This technology plays an important role in all stages of data mining. In the preparation phase, the source data is displayed through scatter plots and histograms, which will lay the foundation for better data selection. In the mining phase, various mining processes are described in the visual form, and the user can see from which database the data is extracted, how to extract, how to preprocess and how to mine. In the presentation phase, the technique makes the training indicators easier to understand.

2.2 Establishment of training analysis mechanism

Data mining classification techniques include decision trees, Bayesian, neural networks and rough sets[9]. This paper mainly studies the decision tree classification method based on the following considerations:

First, the decision tree method can generate easily understandable rules. Because the end users are teaching managers, they often do not have the data mining sports training indicators, so the interpretability of the mining method is very important. The decision tree represents the final classification result in a tree structure, and it can also generate If-Then rules. The theoretical expression can be written as follows:

$$E_0 = \sum_{i=1}^n (a_i - \bar{a})(f_i - \bar{f}) \cdot 1 / [\sum_{i=1}^n (e_i - \bar{e})] \quad (4)$$

Where the E_0 represents the theoretical expression function, the n is the calculated length, the a means the element record range, the f is the discrete index, and the e represents the index range.

Second, the calculation of the method is not very large. This system is mainly practical application, not algorithm research, so the work efficiency is more important. This method can greatly shorten the time of calculation and improve the system's execution efficiency. The efficiency of execution can be written as the following formula[10]:

$$E = p \sum_{i=1}^n (a_i - \bar{a})(f_i - \bar{f}) \cdot 1 / [\sum_{i=1}^n (e_i - \bar{e})] + \log(a^n + ef) \quad (5)$$

Where the E is the execution efficiency, the p means the defined attributes of function. In addition, the decision tree method can handle continuous and discrete data. The database contains more types of data, not only qualitative attributes but also quantitative attributes. Among them, qualitative attributes account for the majority. The method works better with discrete data.

Finally, the decision tree can clearly show the importance of attributes. It chooses the splitting attribute by the calculation of the information entropy, and which is the metric of the importance of the attribute. From an intuitive point of view, the higher the level of the node is, the more important the attributes represented by the node are. From an intuitive point of view, the higher the level of the decision tree node is, the more important the attributes represented by that node are. Then, the role of the nodes of same level is basically the same.

In summary, this paper chooses the decision tree method for the analysis of sports training index. Its process function can express the formula 6[11].

$$\frac{\partial T}{\partial t}(x_j, t_n) = \frac{T(x_j, t_{n+1}) - T(x_j, t_n)}{\tau} + O(\tau) \quad (6)$$

Where the T represents the set of training data, also known as the training set or training database, t is the decision tree fancier, the x_j represents the j-layer data, the O represents the split-choice attribute, the τ represents the calculation width, and n is the range of conditions.

The decision tree is the process of classifying data through a series of rules, which is the induction learning algorithm. It infers the classification rules of the decision tree representation from a set of irregular elements. It adopts the top-down recursive method to compare attribute values at internal nodes and branch downwards according to different attribute values. The leaf node is the class to be divided. The path from the root node to the leaf node corresponds to a classification rule, and the entire tree represents a set of rules.

From Figure 3, it is seen that the decision tree is a tree structure similar to a flow chart, which consists of decision nodes, branches and leaves. Each node corresponds to a non-category attribute, each branch corresponds to each possible value of the attribute, and each leaf node of the tree represents a category. The middle node of the tree is usually represented by a rectangle, while the leaf node is represented by an ellipse. At present, a variety of decision tree algorithms have been formed, such as CLS, ID3, CHAID, CART, FACT, C4.5, GINI, SEES, SLIQ, SPRINT, etc. [12]. The most famous algorithm is the ID3 algorithm proposed by Quinlan.

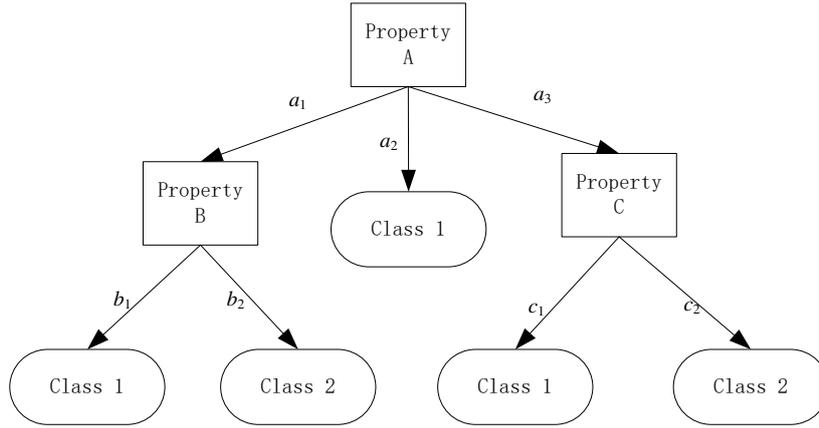


Figure 3 Decision Tree

Figure 4 describes the generation process of the decision tree, which is divided into learning and testing. The learning phase adopts a top-down recursive approach[13]. The algorithm is divided into two steps: one is the generation of the tree, and the other is the pruning of the tree, which is to remove some data that may be noise or abnormal.

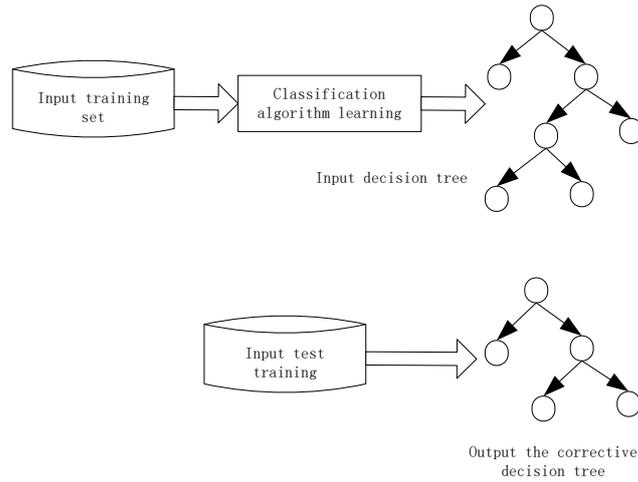


Figure 4 The decision tree generation

The formula for removing noise and abnormal data volume is as follows:

$$L_n(x) = \sum_{j=0}^n y_j \left(\prod_{\substack{i=0 \\ i \neq j}}^n \frac{x - x_i}{x_j - x_i} \right) \quad (7)$$

Where L_x represents the amount of noise removed, x represents a series, x_i is the i -th layer of the conclusion, x_j represents the j -th layer of the conclusion, C represents the decision tree, and n means the scope of the condition.

The condition that the decision tree stops splitting is that the data on one node belongs to the same category and no attribute can be reused for segmentation. Building the decision tree can be done by scanning the database several times. This means that fewer resources are required and that it is easy to handle situations where there are many predictors. Therefore, the model of the decision tree can be built very quickly and is suitable for applying to a large amount of data.

Through the determination of index parameter classification, the data mining technology is imported and the analysis mechanism is established. Finally the model is built.

3 Methods

3.1 The analysis of data mining process

Data mining is a multi-stage process, which mainly includes data preparation, data mining and result interpretation. The data mining process of sports training index is the iterative process of these three phases, as shown in Figure 5[14].

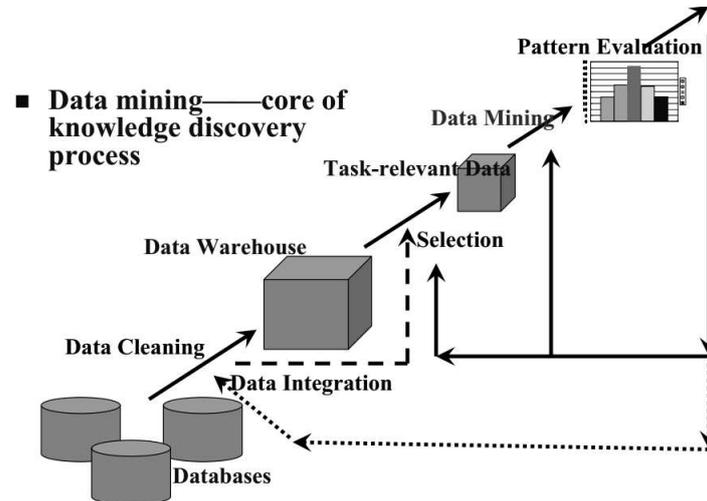


Figure 5 Data mining process model

Data preparation accounts for the largest proportion of the entire mining process, usually around 60%. This stage is divided into three steps: data selection, data processing and data transformation. Data selection mainly refers to the extraction of data from the database and the formation of target data. Preprocessing is to process the extracted data so that it meets the requirements. The main purpose of the transformation is to reduce the data dimension. According to Formula 4 and Formula 5, the expression of the initial feature function is as follows[15]:

$$m_i I_i N_i = \frac{{}^{i+1}v_i \theta_i^2}{(E - E_1 - E_2) l_i} \quad (8)$$

In which, the m represents data feature variables, I is data variability, N is target data, v means calculation magnitude, θ is spelling records, l means mining scope, E is data mining, E_1 represents the mining of initial conditions, E_2 represents the mining of working state, and i represents the data of the i -th level.

Data mining is firstly algorithmic planning, such as the discovery of data summary, classification, clustering and association rules or discovery of sequence patterns. Then, the algorithm is selected for this mining method. The choice of the algorithm directly affects the quality of the mining model. After completing the above preparations, the algorithms of data mining can be run. This stage is the phase that data mining analysts and experts are most concerned about. It can also be called data mining in the real sense and expressed by the following function:

$$D = |M(q)| = \begin{bmatrix} 0 \\ m_x f \\ 0 \\ I_i \end{bmatrix} + |G(q)| = \begin{bmatrix} E_s \\ 0 \\ E_R \\ I_i \end{bmatrix} \quad (9)$$

In which, the D represents the data mining process, M(q) represents the sum of condition vectors of the formula 5 and the formula 6, G(q) represents the sum of state mining state vectors, m_x represents the difficulty of index analysis, and f represents the frequency of mining.

I_i represents the amount of data mining in the i-period, and E_s represents the mining status. E_R represents mining conditions, where $q \in R^n$, $|M(q)| \leq d$, and E_s is a constant, M(q) is an inverse matrix, and $m_x I_i N_i - \cos v_i \theta_i^2 f / l_i (E - E_1 - E_2) = 0 \mid \forall x \in R^n$. Where, $|G(q)| \leq d \leq a$, and the integral of the value is a constant.

Data mining tasks include correlation analysis, cluster analysis, classification, prediction, timing model and deviation detection. Association analysis means that when the values of two or more data items appear repeatedly and the probability is high, there is an association between them. The association rules of these data items can be established to reflect the correlation between events. If there is an association between multiple attributes, the attribute value of one can be predicted based on the others. For example, 90% of customers who buy bread buy milk, which is an association rule. Putting them together will increase their sales. In large databases, there are many such association rules, which require them to be screened. Generally use the values of "support" and "confidence" to filter the useless rules, which can be expressed by the following formula [16]:

$$\delta = \frac{1}{L-l} \sum_{i=1}^j \frac{SF^2}{dh} \quad (10)$$

In which, the δ represents the useless association rule, L represents the value of the support, l is the value of confidence, S is data schedulability, F is the value of the data attribute, h represents the correlation coefficient, and d represents a motion index state.

The timing model of data mining refers to searching through the time series for a pattern with a high probability of recurrence. In this model, it is necessary to find out the rule that the ratio always exceeds a certain minimum percentage in a certain minimum time [17]. These rules will be adjusted as the situation changes. One of the most important methods in the model is "similar timing." Using it, the temporal event database is viewed in chronological order, from which one or more similar temporal events can be found.

$$\psi = F_s \int_{-1}^1 (P_L) \cdot \frac{Nc}{\varepsilon \xi} \quad (11)$$

In which, the Ψ is similar timing events, the Nc represents the data mining time, the F_s means the constant of the sports index, and ξ represents the data of the sports index.

Data mining clusters data into several categories. The data of the same category are similar to each other. The distances of different categories of data are relatively large and different from each other. Clustering includes statistical analysis methods, machine learning methods and neural network methods. Statistical analysis is clustering based on distance. This method is the clustering

of global comparison. It needs to examine all individuals to determine the division of the class. The distance in clustering of the machine learning is determined according to the concept description, which is called concept clustering. When the clustering objects increase, concept clustering is called concept formation. In neural networks, self-organizing neural network methods are used for clustering. Such as ART model and Kohonen model, this is the unsupervised learning method [18]. After a given distance threshold, each sample is clustered according to the threshold. The clustering formula is as follows:

$$\lambda = \frac{Ns \cdot Ns_L}{QN} + \psi \quad (12)$$

In which, the λ represents the clustering of data mining, the Q represents the mining coefficient, the N means the total amount of mining, the Ns represents the overall amount of the s-layer, the NsL represents the overall amount of the next layer of the s-layer, the Ψ is similar timing events.

Classification is most widely used in data mining. It is to find the concept description of the category and use this description to construct the model. The description represents the overall information of such data, that is, the content description. Connotation description is divided into feature description and discernment description. Feature description is the description of common features of data, and discriminatory description is the description of the difference between them. The process of classification is to analyze the input data, find an accurate description for each class by calculating the characteristics of the data, and use this description to classify the subsequent data[19].

Deviation detection is to find out the abnormal situation of data. Deviation includes many potential sports training indicators, such as anomalous instances in the classification, deviations in results from predictions, and changes in magnitude. The basic method of deviation detection is to find the difference between the result and the reference, which can be expressed by equation 13.

$$Pi(w) = \frac{C}{F \cdot E} \left| \sum_{i=1}^j Pi(m)e + \lambda \right|^2 \quad (13)$$

Where, the $pi(w)$ represents the difference between the observations and the reference, and the $Pi(m)$ is the fuzzy mathematical method.

The forecast is to use historical data to find the law of change, establish a model, and use this model to predict the types and characteristics of the data. Regression analysis is a typical prediction method, which establishes a regression equation with time as a variable. In the prediction process, entering any time value can get the status at this time. The neural network method realizes learning of nonlinear samples, which can discriminate nonlinear functions. Classification can also be used for prediction, but classification is generally used for discrete values; regression prediction is used for continuous values; neural network method prediction can be used for continuous values as well as discrete values [20].

The expression and interpretation of results is based on the user's purpose of analyzing the information and distinguishing the most valuable information.

The patterns found in the early stages are evaluated by the user, and the useless patterns are deleted. If the user's requirements cannot be met, the pattern is returned to the previous stage. In addition, the end-users faced by data mining are people. Therefore, the discovered patterns must

be visualized. For example, the decision tree is transformed into an "if...then..." rule whose process model is shown in figure 6.

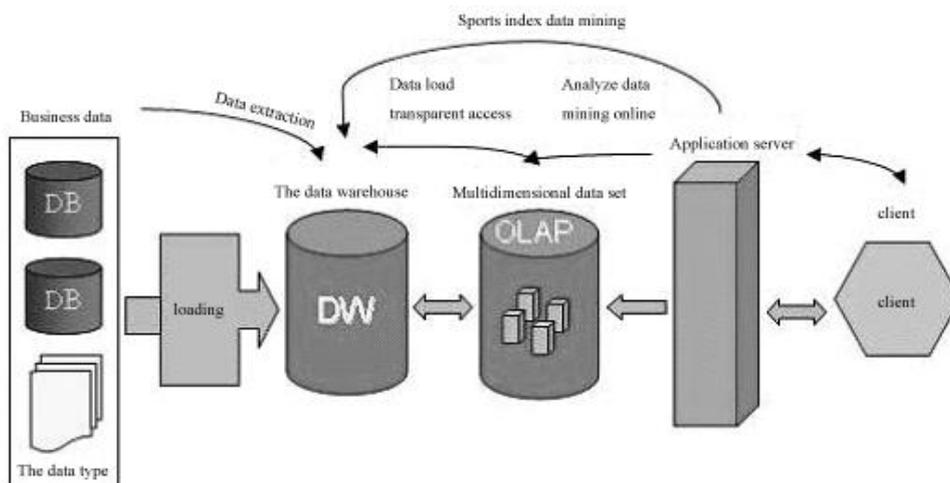


Figure 6 The model of the data mining process

3.2 Effective data analysis

If the sample belongs to the same class, this node is a leaf node and is marked with this class. Otherwise, the measure of information gain is used as heuristic information to select the attribute of the sample classification, which is the "test" or "decision" attribute of the node. Assumed that all attributes are classified, that is, take discrete values. The branch is created for each known value of the test attribute, and the sample is divided.

The algorithm uses the similar approach, recursively forming the sample decision tree on each partition. Once the attribute appears on a node, it is not necessary to consider this attribute on the descendants of the node. The entire recursion process stops when one of the following conditions is true:

- (1) All samples for the given node belong to the same class.
- (2) There are no remaining attributes that can be used to further divide,
- (3) There is no sample in the branch. In this case, the leaf is created with the majority of the training sample set.

When the decision tree is created, many branches reflect abnormalities in training. The pruning method uses statistical metrics to clip the least reliable branches to improve the ability of the decision tree to correctly classify.

The pre-pruning method prunes by stopping the construction of the tree in advance. Once stopped, the node becomes the leaf. This leaf has the most frequent classes in the subset sample. When constructing a tree, if the information gain is the equal measure, it can be used to assess the superiority of the split. If the partitioning of the sample results in a split below a predefined threshold, the partitioning stops. However, choosing a proper threshold is difficult, higher may result in an oversimplified tree, and lower will make the simplification too little. The method expression can be written as follows:

$$\delta = \frac{1}{\mu - 1} \sum_{i=1}^j [P_i(m) - t]^2 \quad (14)$$

In which, the δ means the pre-pruning method, the μ represents containing the sample of subsets, $P_i(m)$ represents the fuzzy mathematics method, and t represents the fancier of decision tree.

Post pruning is the cutting of fully grown branches. By deleting the node's branches, the nodes are cut off. In the pruning algorithm of the cost-complexity, the untrimmed nodes become leaves.

The expected error rate of pruned subtrees on non-leaf nodes is calculated. Then, combing the weighting of the branches, the error rate of each branch is used to calculate the expected error rate of no pruning. If the pruning results in a high error rate, the subtree is preserved. Using test sets to evaluate the accuracy of each tree, the decision tree with the lowest expected error rate is obtained. Post-pruning requires more calculations than pruning, but the resulting tree is more reliable, and its formula can be written as equation 15.

$$As = \frac{\partial^2(A_1 + A_2)}{\partial V^2} + g_0 F_0 / H \quad (15)$$

Where the A_s represents the marker expression, the A_1 represents the error rate of each branch, the A_2 is the weight assessment of the branches, the V means the range of element records, the g_0 represents the index of the index, the F_0 represents the index range, and the H is the concept tree method.

When extracting classification rules from the decision tree, the rules are expressed in the form of "if then". The rule is created for each path from the root to the leaves to form the conjunct of the predecessor. Leaf nodes contain class predictions that form the rule post. Based on the analysis of the three processes of data preparation, data mining and results interpretation, the model based on the analysis of sports training index obtains the data mining results of the index and realizes the analysis of the index. The application of data mining technology in the analysis of sports training index was completed.

4 Experiment

In order to ensure the effectiveness of the technology proposed in this paper, simulation test analysis is performed. The test uses different types of sports training index as objects for analysis. In order to ensure the validity of the test, the conventional index analysis method is used as comparison object. The test data is presented in the same chart and conclusions are reached through the calculation of comprehensive analysis capabilities.

4.1 Data preparation

The test parameters are set to ensure the accuracy of the test. In this paper, different types of index are used as test objects. Two kinds of analysis methods are used to conduct simulation tests and analyze the results. Because the results obtained by different methods and the analysis methods are different, it is necessary to ensure the consistency of the environment. The data set results in this paper are shown in Table 1.

Table 1 Parameter settings of the simulation test

project	Execute range/parameters	note
Simulates simulation test equipment parameters	STO simulation platform, STO-147s	The two groups used identical analog devices
Simulation test equipment platform	Windows	Windows7 64-bit flagship
Simulation device central processing unit	Intel core i7 8700	Slot type: LGA 1151 CPU main frequency: 3.2ghz

		Maximum frequency: 4.6GHz
		Core number: six cores
		Thread count: twelve threads
		Core frequency: 1569/1784mhz
Analog video card	GeForce GTX 1060-6GD5 X-Gaming OC	Display frequency: 8008MHz Video memory capacity: 6GB Display memory bit width: 192bit Power interface: 8pin

The two analysis methods are put in this environment, and the simulation data is loaded. The simulation parameters as follows:

Table 2 Simulation Parameters of Index Types

Types of sports	Sports indicators	Sports parameter range
Track and field	Speed time	0~6h
ball	Score, hit rate, time	0~150
swimming	Time, distance, action	0~30min
Martial arts	Skill, ability and standard of action	[0,1]
aerobics	Time, action	0~150
Mountain climbing	Mountain climbing	0~6h
skating	Time, distance, technique	0~200min
Weight lifting	Weight, quality, technology	0~300kg
wrestling	Ability, time, quality	[0,1]
The bicycle	Time, quality, speed	0~200min
rowing	Time, endurance, distance	0~200min
baseball	Integrity, hit rate	0~150
Tae kwon do	Hit rate, physical performance	0~200min
shooting	Hit rate, accuracy	0~10

4.2 Test design

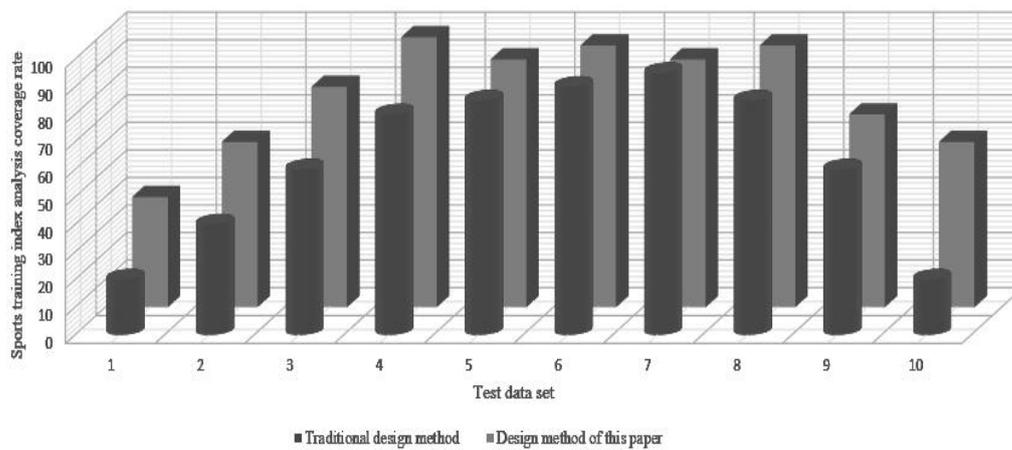
In order to verify the comprehensive analysis capabilities of the two methods, the experiments of coverage simulation, accuracy simulation and immunity simulation were carried out in this paper. The results are recorded, and the comprehensive ability of the index analysis is calculated according to the formula.

First, the data is input into the simulation system set in accordance with the requirements of Table 1, and the correlation operation is performed. Then, under the same conditions, three experiments were conducted separately.

Finally, the third-party software of analysis and recording is used to analyze the data generated, while eliminating the uncertainty caused by various factors. For different types of indicators and analysis methods, simulation tests were performed and the results were shown in the test comparison curves. The conclusion is drawn by the formula for calculating the comprehensive analysis ability.

4.3 The test of coverage simulation

The comparison curve of the results of coverage simulation test is shown in Figure



7.

Figure 7 The comparison curve of the coverage simulation test

From the comparison curve, it can be seen that the overall of the results of the method designed in this paper is 87.41%, while the coverage of the traditional method is only 79.42%.

4.4 The test of the accuracy simulation

The comparison curve of the test results is shown in Figure 8.

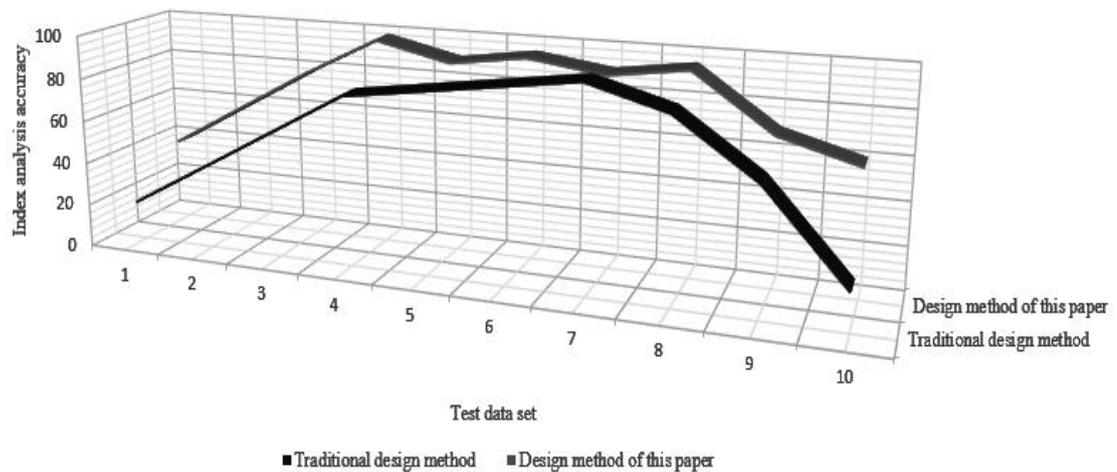


Figure 8 The comparison curve of accuracy simulation test

By comparison, the error rate of the method designed in this paper is 89.92%, while that of the traditional analysis method is 74.12%.

4.5 The test of the noise immunity simulation

The comparison curve of the noise immunity simulation test results is shown in Figure 9.

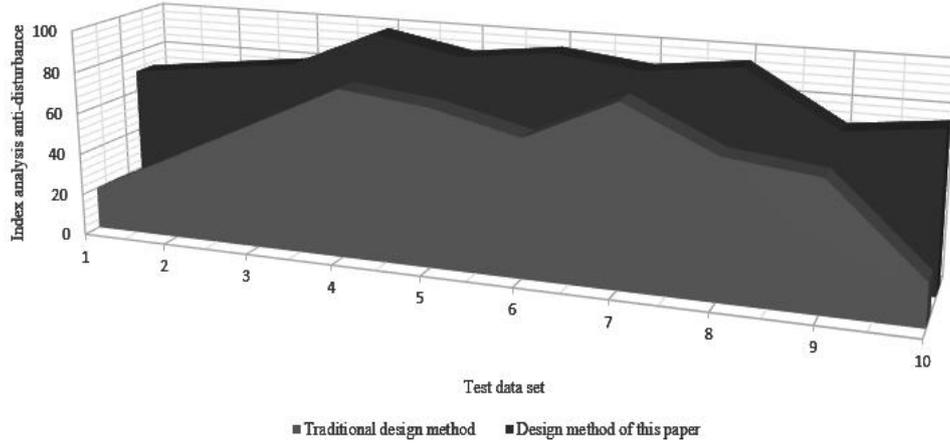


Fig. 9 The comparison curve of noise immunity simulation test

As can be seen from Figure 9, the anti-interference ability of the design method is 98.41%, while the anti-jamming capability of the traditional analysis method is 69.53%.

4.6 The calculation of comprehensive analysis ability

Substitute the test results of coverage and accuracy into the following formula:

$$\chi = \frac{1}{n} k \sum_{i=1}^n \left(\frac{Dg}{y} \right)^2 \quad (16)$$

In which, the D represents the coverage test results, the g means the results of the accuracy test, the y represents the results of the noise immunity test, and the k represents the simulation coefficient. This paper takes 0.98.

The proposed method is denoted as χ_1 , the conventional method is denoted as χ_2 , $\Delta\chi = \chi_1 - \chi_2$ is the positive number indicates that the comprehensive analysis ability is improved, and the negative number $\Delta\chi = \chi_1 - \chi_2$ represents the decrease in the comprehensive analysis ability. Then the $\Delta\chi = \chi_1 - \chi_2$ is written as following:

$$\begin{aligned} \Delta\chi &= \chi_1 - \chi_2 \\ &= \frac{1}{n} k \sum_{i=1}^n \left(\frac{D_1 g_1}{y_1} \right)^2 - \frac{1}{n} k \sum_{i=1}^n \left(\frac{D_2 g_2}{y_2} \right)^2 \\ &= 0.371409 \end{aligned}$$

Compared with the conventional method, the comprehensive ability of the data mining application is increased by 37.14%, which can be applied to the analysis of sports training index of different data types.

5 Results and Discussion

This paper puts forward the application of data mining technology in the analysis of sports training indicators. It relies on the construction of index analysis model, through the analysis of mining process and data, the analysis of sports training index is completed. The experimental data shows that the method designed in this paper has extremely high effectiveness. It is hoped that the

research in this paper can provide a theoretical basis for the analysis methods of sports training indicators.

Reference

[1]WangChong, WANGCong. Simulation of 3d visual action amplitude tracking method in sports [J]. Computer Simulation, 2017(1):245-248.

[2]ZhouPeng, WangShiwei, ZhangWuping. Simulation of high accuracy control of volley hit point on volleyball front [J]. Computer Simulation, 2017(12):246-249.

[3]Zhang L, Yang X, Sang C. Cloud computing and data mining application in enterprise profitability analysis based on the perspective of cash flow[J]. RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao, 2016, 2016:161-172.

[4]X. Ruan, G. Tao, H. Liu, et al. Application of data mining for investigating the cognition of how square dance promote community sports culture construction[J]. Boletin Tecnico/technical Bulletin, 2017, 55(13):594-600.

[5]Rodríguez - Jiménez J M, Cordero P, Enciso M, et al. Data mining algorithms to compute mixed concepts with negative attributes: an application to breast cancer data analysis[J]. Mathematical Methods in the Applied Sciences, 2016, 39(16):4829-4845.

[6]Hong H, Tsangaratos P, Iliia I, et al. Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China.[J]. Science of the Total Environment, 2018, 625:575-588.

[7]Ziatdinov M, Fujii S, Kiguchi M, et al. Data Mining Graphene: Correlative Analysis of Structure and Electronic Degrees of Freedom in Graphenic Monolayers with Defects[J]. Nanotechnology, 2016, 27(49).

[8]Hamedianfar A, Shafri H Z M. Integrated approach using data mining-based decision tree and object-based image analysis for high-resolution urban mapping of WorldView-2 satellite sensor data[J]. Journal of Applied Remote Sensing, 2016, 10(2):025001.

[9]Xue Y, Zhang X, Li S, et al. Analysis of factors influencing tunnel deformation in loess deposits by data mining: A deformation prediction model[J]. Engineering Geology, 2018, 232:94-103.

[10]Boersch I, Füssel U, Gresch C, et al. Data mining in resistance spot welding: A non-destructive method to predict the welding spot diameter by monitoring process parameters[J]. International Journal of Advanced Manufacturing Technology, 2017.

[11]Regulski K, Wilkkołodziejczyk D, Kacprzyk B, et al. Approximation of Ausferrite Content in the Compacted Graphite Iron with the Use of Combined Techniques of Data Mining[J]. Archives of Foundry Engineering, 2017, 17(3).

[12]Liu W, Rostami J, Keller E. Application of new void detection algorithm for analysis of feed pressure and rotation pressure of roof bolters[J]. International Journal of Mining Science and Technology, 2017, 27(1):77-81.

[13]Mathan K, Kumar P M, Panchatcharam P, et al. A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease[J]. Design Automation for Embedded Systems, 2018(9):1-18.

[14]Rojas J, Forero J, Gaona P, et al. Analysis of Physico-chemical Variables and their Influence on Water Quality of the Bogota River using Data Mining[J]. International Journal of High Performance Systems Architecture, 2017, In Press(In Press):In Press.

[15]Górecki J, Hofert M, Holeňa M. An approach to structure determination and estimation of hierarchical Archimedean Copulas and its application to Bayesian classification[J]. Journal of Intelligent Information Systems, 2016, 46(1):21-59.

[16]Pourtaghi Z S, Pourghasemi H R, Aretano R, et al. Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques[J]. Ecological Indicators, 2016, 64:72-84.

[17]Pinto P, Theodoro I, Arrais M, et al. Data mining and social web semantics: a case study on the use of hashtags and memes in Online Social Networks[J]. IEEE Latin America Transactions, 2017, 15(12):2276-2281.

[18]Kester Q A. Using Formal Concepts Analysis Techniques in Mining Data from Criminal Databases and Profiling Events Based on Factors to Understand Criminal Environments[J]. Lecture Notes in Computer Science, 2016, 9790:480-496.

[19]Bui D T, Tuan T A, Klempe H, et al. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree[J]. Landslides, 2016, 13(2):361-378.

[20]Han Y, Moutarde F. Analysis of Large-Scale Traffic Dynamics in an Urban Transportation Network Using Non-Negative Tensor Factorization[J]. International Journal of Intelligent Transportation Systems Research, 2016, 14(1):36-49.

List of abbreviations

(SEES) Sage Extended Enterprise Suite

Declarations

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

declares that he has no conflict of interest.

* Research involving human participants and/or animals

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

*** Informed consent**

All authors agree to submit this version and claim that no part of this manuscript has been published or submitted elsewhere.

Funding

None

Authors' contributions

Liqu Qian and Jiatong Liu wrote the entire article.

Acknowledgements

None

References

[1]WangChong, WANGCong. Simulation of 3d visual action amplitude tracking method in sports [J]. Computer Simulation, 2017(1):245-248.

[2]ZhouPeng, WangShiwei, ZhangWuping. Simulation of high accuracy control of volley hit point on volleyball front [J]. Computer Simulation, 2017(12):246-249.

[3]Zhang L, Yang X, Sang C. Cloud computing and data mining application in enterprise profitability analysis based on the perspective of cash flow[J]. RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao, 2016, 2016:161-172.

[4]X. Ruan, G. Tao, H. Liu, et al. Application of data mining for investigating the cognition of how square dance promote community sports culture construction[J]. Boletin Tecnico/technical Bulletin, 2017, 55(13):594-600.

[5]Rodríguez - Jiménez J M, Cordero P, Enciso M, et al. Data mining algorithms to compute mixed concepts with negative attributes: an application to breast cancer data analysis[J]. Mathematical Methods in the Applied Sciences, 2016, 39(16):4829-4845.

[6]Hong H, Tsangaratos P, Ilia I, et al. Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China.[J]. Science of the Total Environment, 2018, 625:575-588.

[7]Ziatdinov M, Fujii S, Kiguchi M, et al. Data Mining Graphene: Correlative Analysis of Structure and Electronic Degrees of Freedom in Graphenic Monolayers with Defects[J]. Nanotechnology, 2016, 27(49).

[8]Hamedianfar A, Shafri H Z M. Integrated approach using data mining-based decision tree and object-based image analysis for high-resolution urban mapping of WorldView-2 satellite sensor data[J]. *Journal of Applied Remote Sensing*, 2016, 10(2):025001.

[9]Xue Y, Zhang X, Li S, et al. Analysis of factors influencing tunnel deformation in loess deposits by data mining: A deformation prediction model[J]. *Engineering Geology*, 2018, 232:94-103.

[10]Boersch I, Füssel U, Gresch C, et al. Data mining in resistance spot welding: A non-destructive method to predict the welding spot diameter by monitoring process parameters[J]. *International Journal of Advanced Manufacturing Technology*, 2017.

[11]Regulski K, Wilkołodziejczyk D, Kacprzyk B, et al. Approximation of Ausferrite Content in the Compacted Graphite Iron with the Use of Combined Techniques of Data Mining[J]. *Archives of Foundry Engineering*, 2017, 17(3).

[12]Liu W, Rostami J, Keller E. Application of new void detection algorithm for analysis of feed pressure and rotation pressure of roof bolters[J]. *International Journal of Mining Science and Technology*, 2017, 27(1):77-81.

[13]Mathan K, Kumar P M, Panchatcharam P, et al. A novel Gini index decision tree data mining method with neural network classifiers for prediction of heart disease[J]. *Design Automation for Embedded Systems*, 2018(9):1-18.

[14]Rojas J, Forero J, Gaona P, et al. Analysis of Physico-chemical Variables and their Influence on Water Quality of the Bogota River using Data Mining[J]. *International Journal of High Performance Systems Architecture*, 2017, In Press(In Press):In Press.

[15]Górecki J, Hofert M, Holeña M. An approach to structure determination and estimation of hierarchical Archimedean Copulas and its application to Bayesian classification[J]. *Journal of Intelligent Information Systems*, 2016, 46(1):21-59.

[16]Pourtaghi Z S, Pourghasemi H R, Aretano R, et al. Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques[J]. *Ecological Indicators*, 2016, 64:72-84.

[17]Pinto P, Theodoro I, Arrais M, et al. Data mining and social web semantics: a case study on the use of hashtags and memes in Online Social Networks[J]. *IEEE Latin America Transactions*, 2017, 15(12):2276-2281.

[18]Kester Q A. Using Formal Concepts Analysis Techniques in Mining Data from Criminal Databases and Profiling Events Based on Factors to Understand Criminal Environments[J]. *Lecture Notes in Computer Science*, 2016, 9790:480-496.

[19]Bui D T, Tuan T A, Klempe H, et al. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree[J]. *Landslides*, 2016, 13(2):361-378.

[20]Han Y, Moutarde F. Analysis of Large-Scale Traffic Dynamics in an Urban Transportation Network Using Non-Negative Tensor Factorization[J]. *International Journal of Intelligent Transportation Systems Research*, 2016, 14(1):36-49.

Figure 1:

Title: **Learn using training set**

Legend: Figure 1 show **Learn using training set**

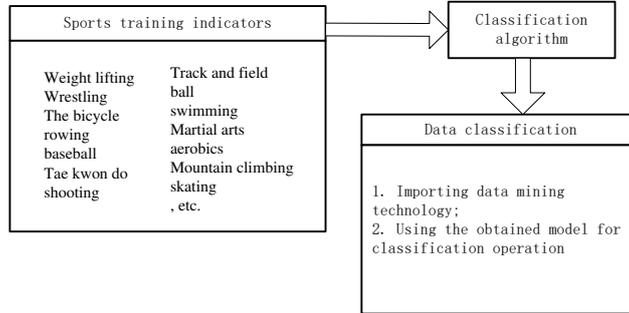


Figure 2:

Title: **Evaluate the model with test data**

Legend: Figure 2 show **Evaluate the model with test data**

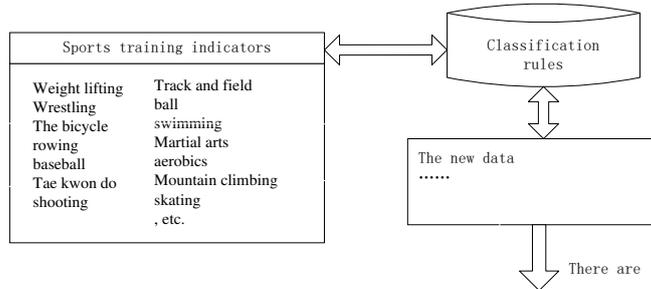


Figure 3:

Title: **Decision Tree**

Legend: Figure 3 show **Decision Tree**

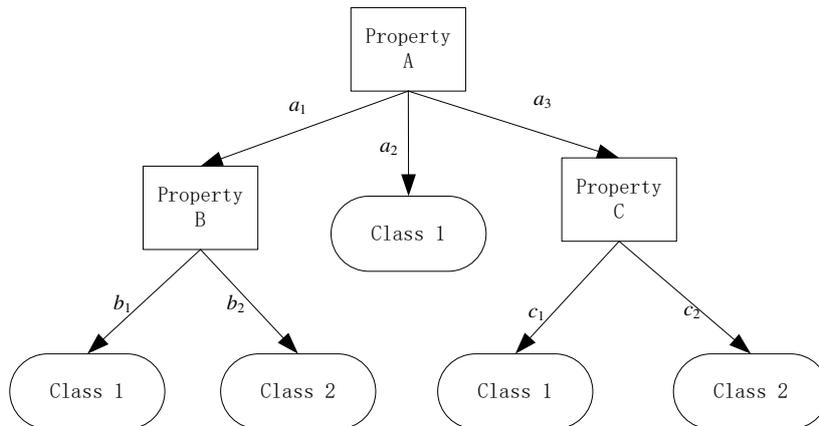


Figure 4:

Title: **The decision tree generation**

Legend: Figure 4 show **The decision tree generation**

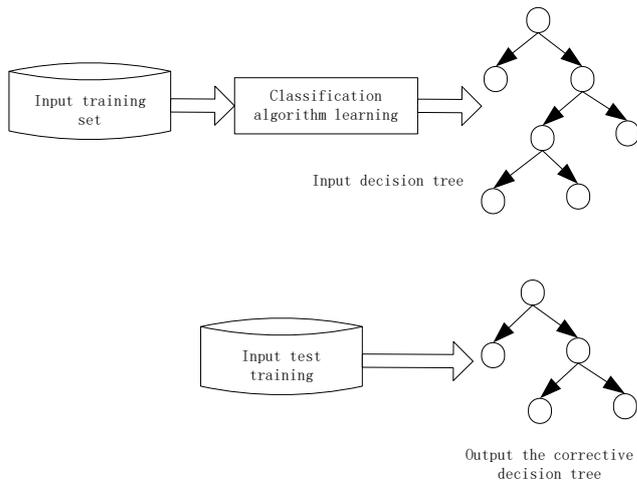


Figure 5:

Title: **Data mining process model**

Legend: Figure 5 show **Data mining process model**

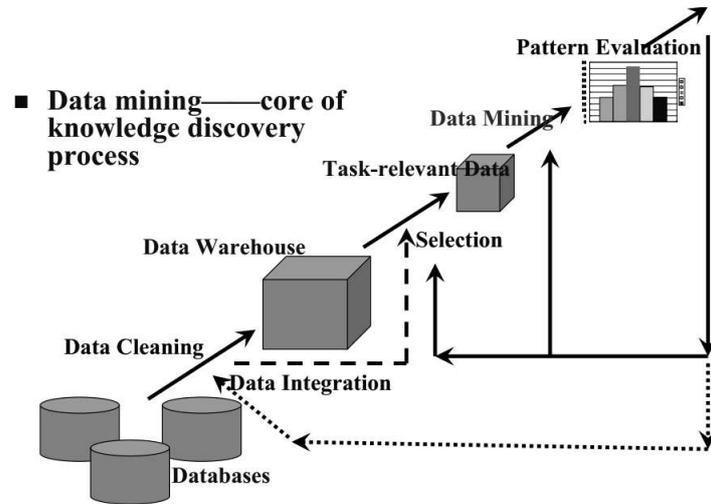


Figure 6:

Title: The model of the data mining process

Legend: Figure 6 show The model of the data mining process

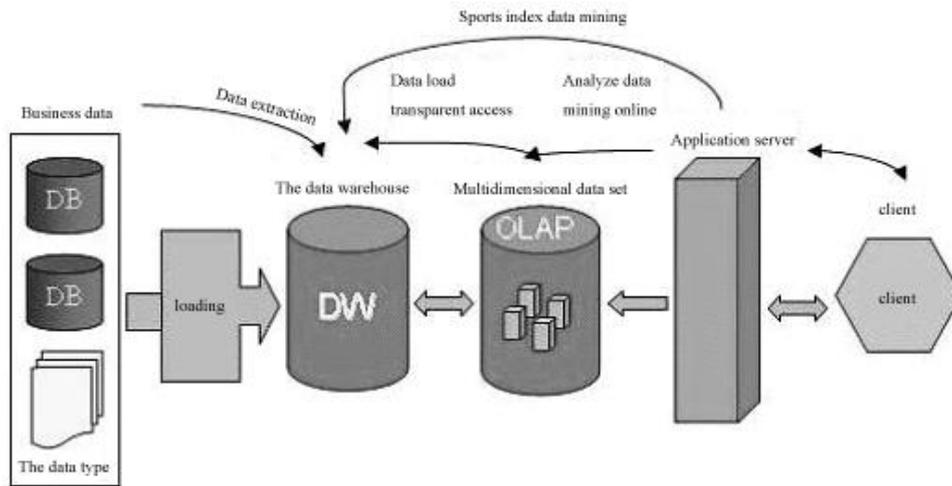


Figure 7:

Title: **The comparison curve of the coverage simulation test**

Legend: Figure 7 show **The comparison curve of the coverage simulation test**

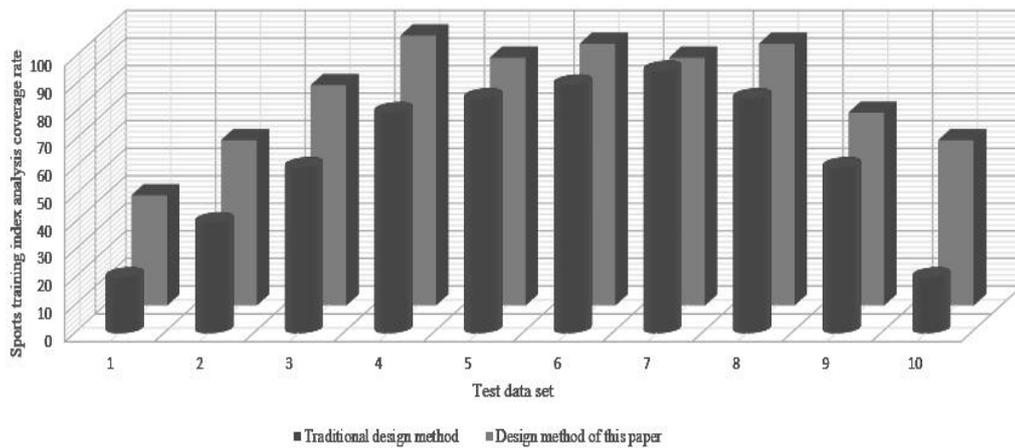


Figure 8:

Title: **The comparison curve of accuracy simulation test**

Legend: Figure 8 show **The comparison curve of accuracy simulation test**

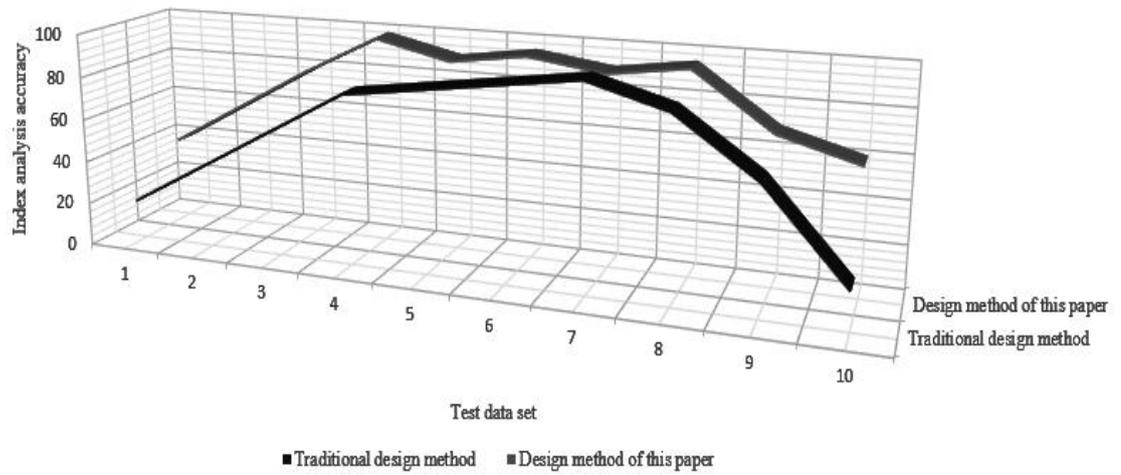
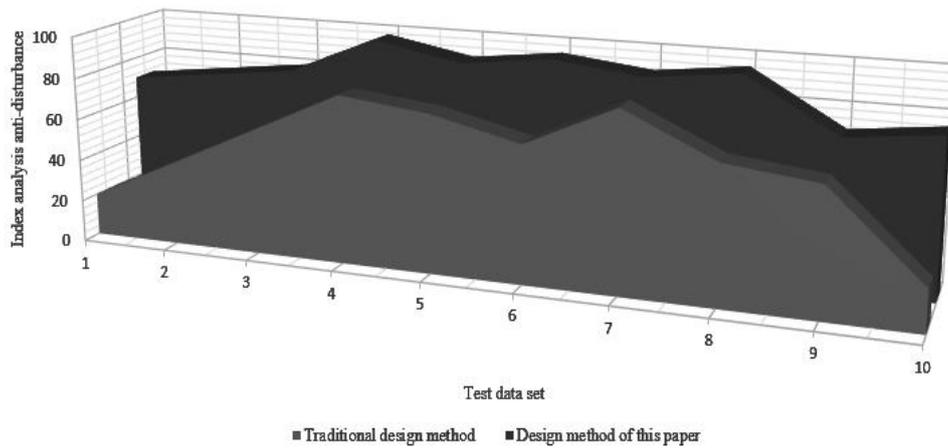


Figure 9:

Title: **The comparison curve of noise immunity simulation test**

Legend: Figure 9 show **The comparison curve of noise immunity simulation test**



Figures

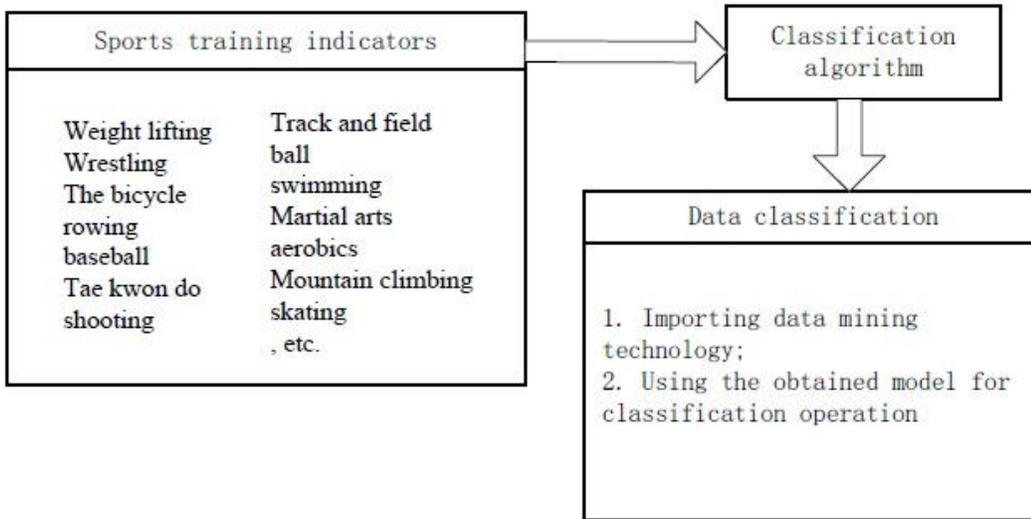


Figure 1

Learn using training set Legend: Figure 1 show Learn using training set

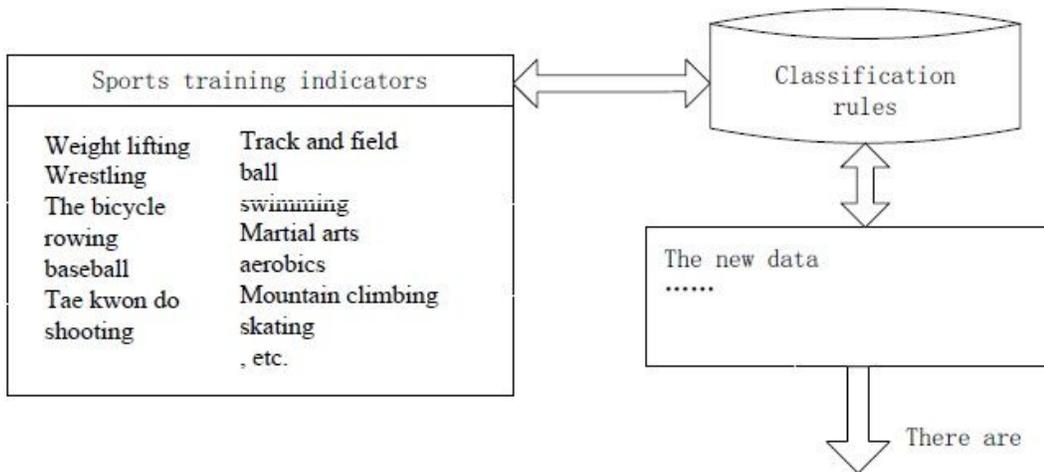


Figure 2

Evaluate the model with test data Legend: Figure 2 show Evaluate the model with test data

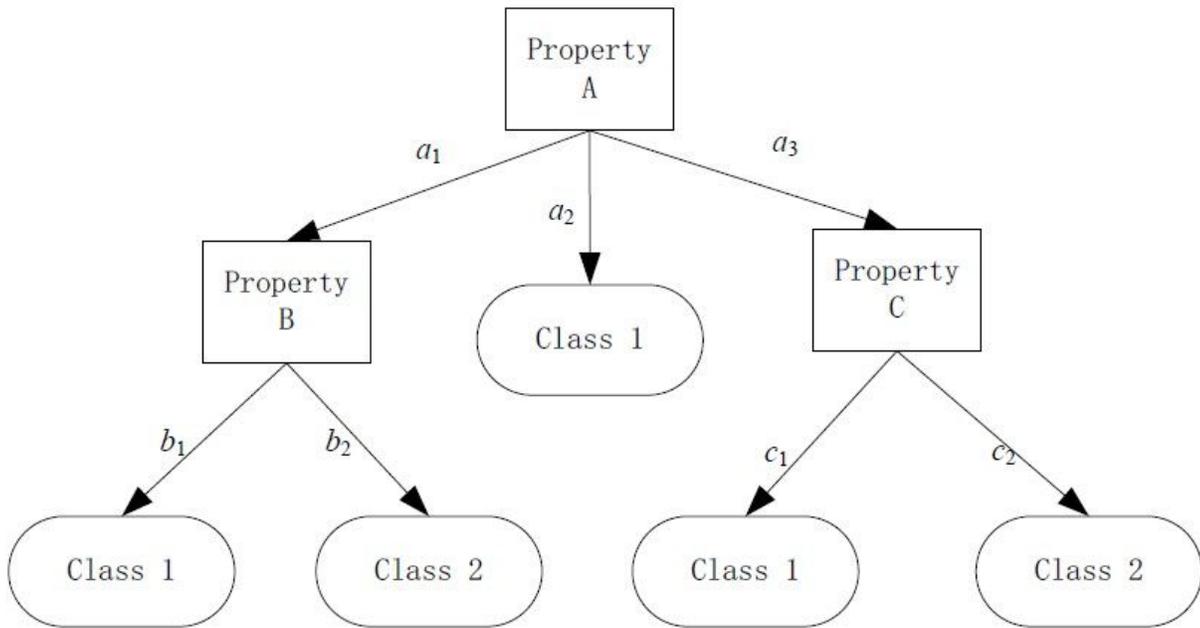


Figure 3

Decision Tree Legend: Figure 3 show Decision Tree

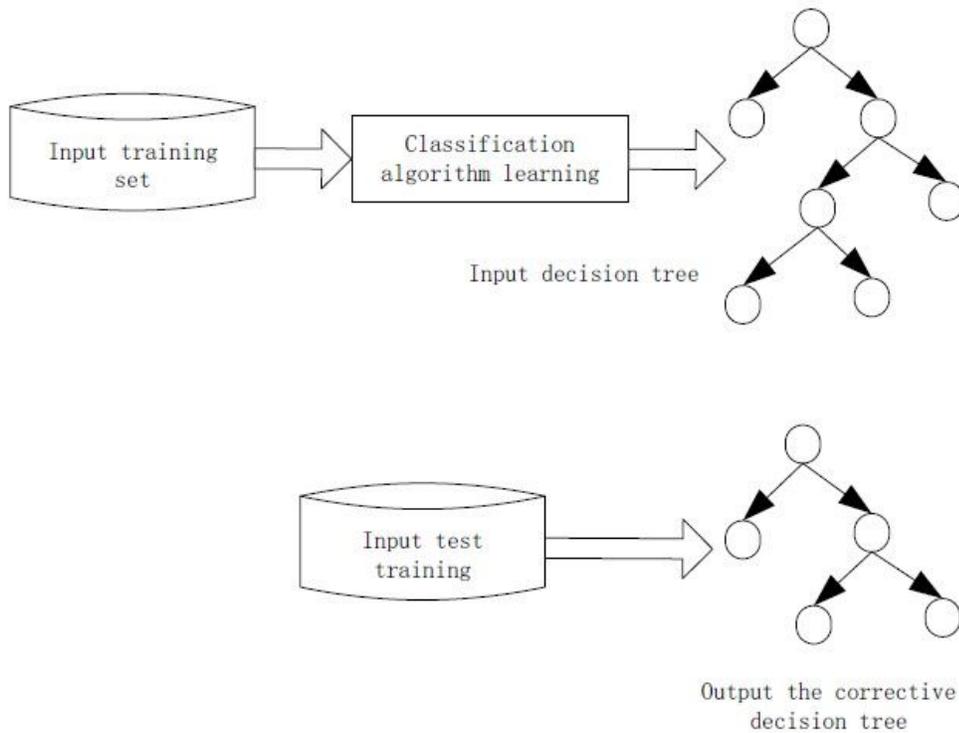


Figure 4

The decision tree generation Legend: Figure 4 show The decision tree generation

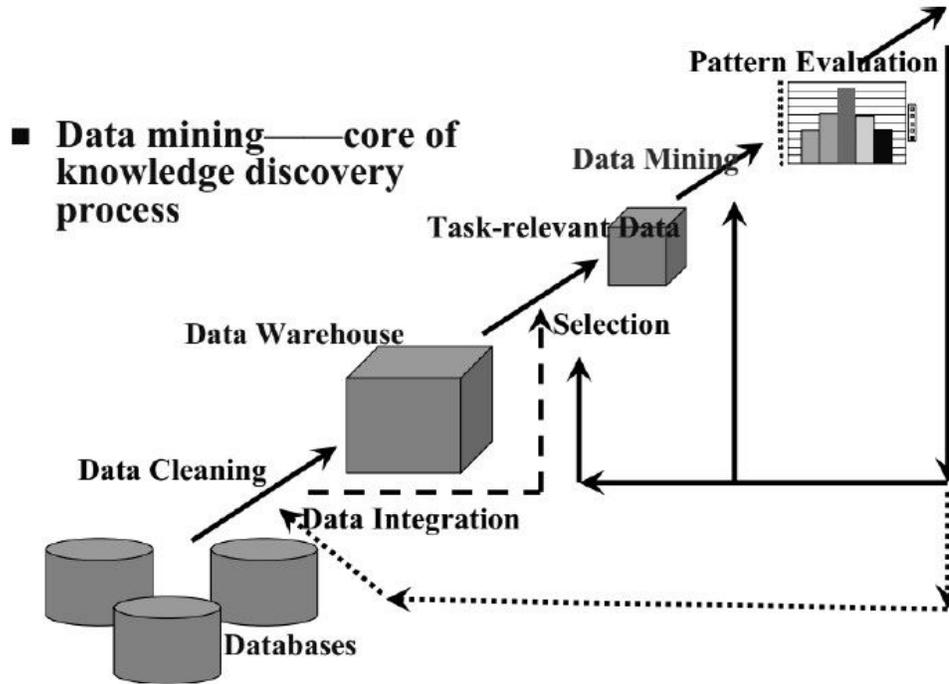


Figure 5

Data mining process model Legend: Figure 5 show Data mining process model Figure

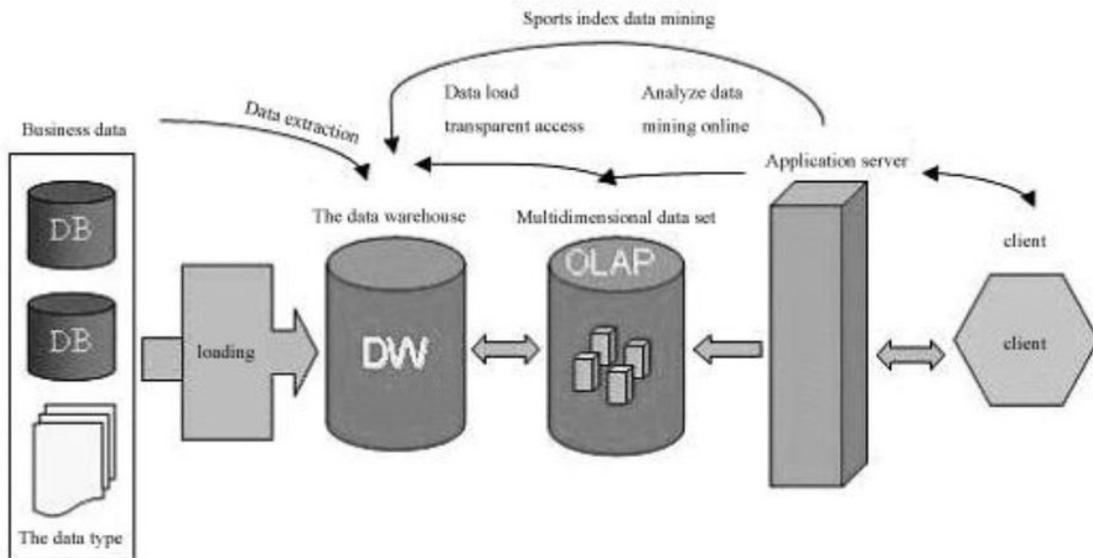


Figure 6

The model of the data mining process Legend: Figure 6 show The model of the data mining process

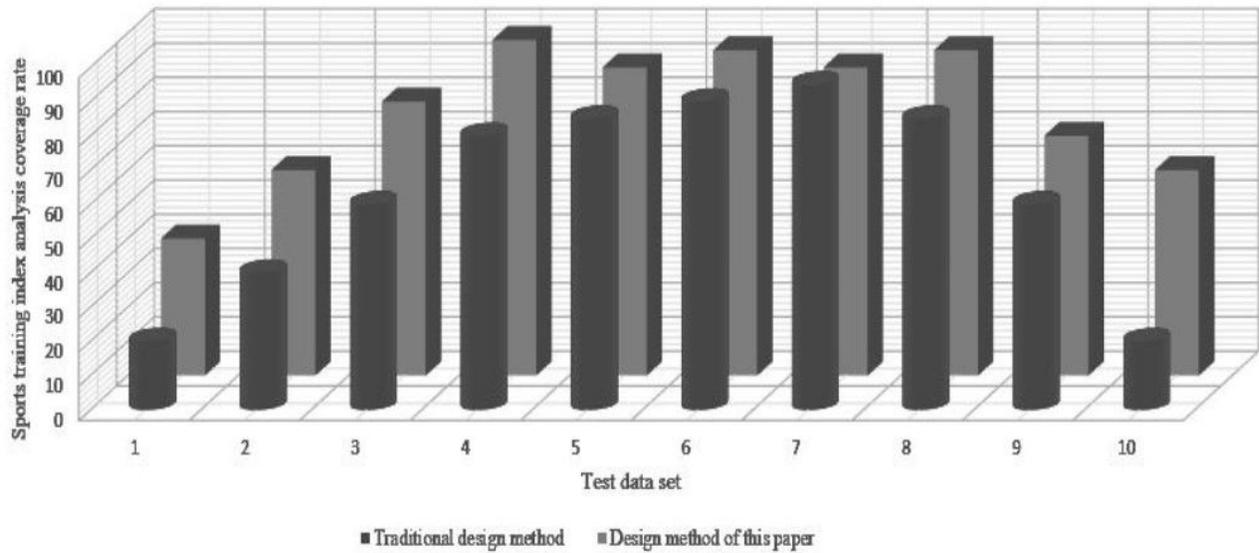


Figure 7

The comparison curve of the coverage simulation test Legend: Figure 7 show The comparison curve of the coverage simulation test Figure 8: Title: The comparison curve of accuracy simulation test Legend: Figure

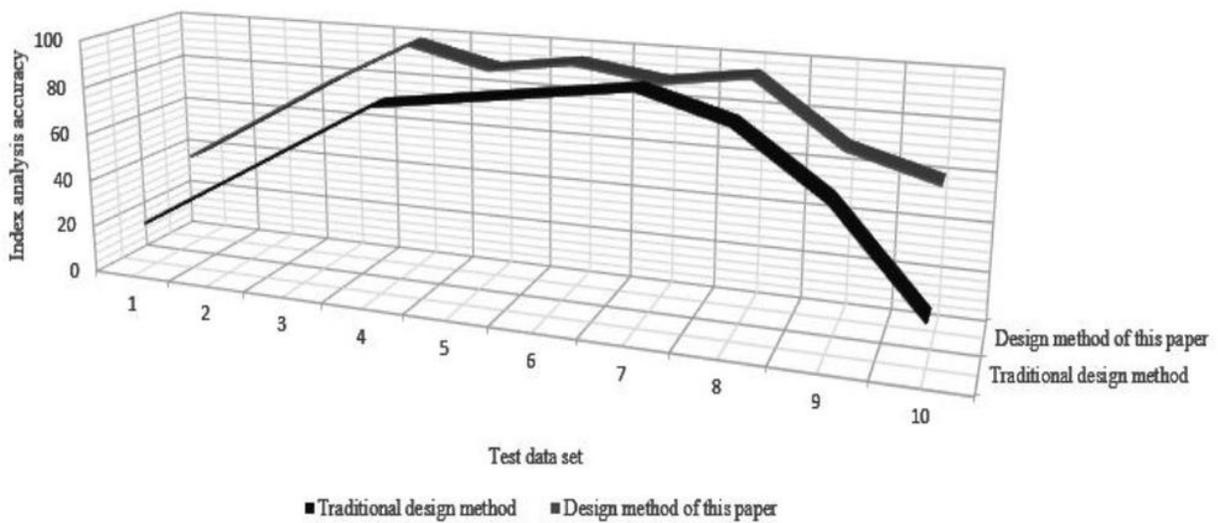


Figure 8

The comparison curve of accuracy simulation test Legend: Figure 8 show The comparison curve of accuracy simulation test

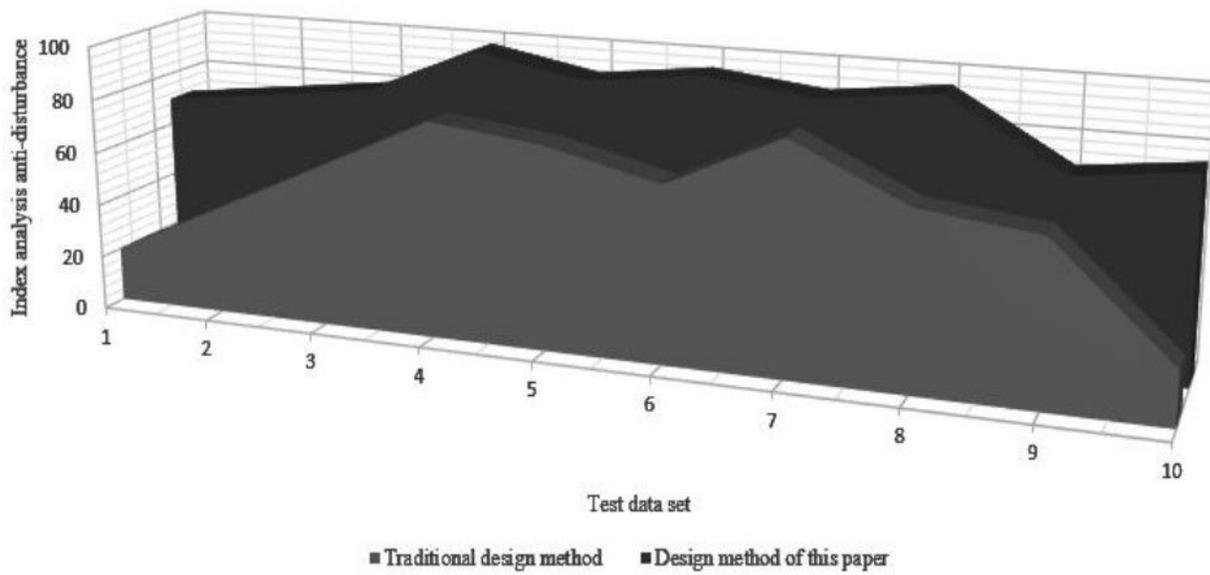


Figure 9

The comparison curve of noise immunity simulation test Legend: Figure 9 show The comparison curve of noise immunity simulation test