Design of real-time data analysis system for physical training based on data mining technology

Deng Hui*, Wang Jin
Wuhan Railway Vocational College of Technology, Wuhan, China

*Corresponding author: 20120035@wrut.edu.cn

Abstract. In competitive sports, the most direct and effective way to improve athletes' competitive level is to strengthen training. At present, the training mode is still mainly based on coaches' experience and athletes' physical conditions, and lacks scientific and effective data information as the basis for planning. In this paper, the software is developed in the core construction of recessed Linux, the VIX bus technology is taken advantage of to gather sports data, and the data mining technology based on collective information characteristic extraction is designed. This paper uses big data mining method to mine the information of sports evaluation, pick up the autocorrelation characteristics of data information flow of sports training real-time data analysis system, constructs the association rule feature set combined with the prior knowledge of sports evaluation, and realizes the database construction of sports training real-time data analysis system by using the association rule feature extraction and big data information fusion processing technology. The simulation results show that the accuracy of sports assess decision-making information mining using the real-time data analysis system of sports training is high, and the dependability of the system is satisfied.

Keywords: Data mining; Physical training; Data analysis system; Sports evaluation decision

1. Introduction
With the expand of sports for all, people take note of sports, and schools take note of and invest in sports courses. For the sake of better estimate the effect of sports training quantitatively, it is necessary to optimize the design of the immediacy data analysis system of sports training, and carry out sports performance statistics, human function information collection and physical fitness prediction in the real-time data analysis system of sports training [1-2]. With the appear of the era of big data, big data analysis technology collects and processes data and information resources based on advanced software systems, which prominent its distinctive preponderance, namely, it covers a large number of data, has high timely and economic value, strong exact and timely, and variety output results, so it is favorite in the intelligent planning of training schedule.

The current sports training real-time data analysis system mainly includes neural network training method [3], SOM self-organizing map learning method and support vector machine method [4-5]. Combined with recessed integral dominate design, it realizes data mining and sports assess policy
decision, but the top of ways require quite a lot previous data information as reference training input. In the case of absence information, the exact of data mining is not well, and the devise of the system has the difficulty of low concordancy [6]. Therefore, this paper designs a immediacy data analysis system for sports training in view of big data analysis. It's not only possible bring about the immediacy transform of data analysis from handbook to computer and go through to work, but also markedly increase the timeliness and reliability of data information processing, thus providing available direct for athletes' training planning.

2. The meaning and method of data mining technology

2.1 The meaning of data mining technology
Data mining was the first time put forward in August 1989 at the 11th International Conference on Joint Artificial Intelligence. Data mining is a to step in the step of knowledge find, that is, the step of pick up secretive, unknown and potentially costly information and knowledge from a great quantity broken, fuzzy, unknown and stochastic data [7-8]. The objects of data mining mainly include relational database, data warehouse, spatio-temporal database, transaction database, multimedia database and text database.

2.2 Basic methods of data mining technology
Data mining technology, as a kind of interdisciplinary application technology, covers statistics, artificial intelligence, fuzzy analysis and other multidisciplinary contents, forming a system of obtaining knowledge after data analysis according to specific problems. According to specific problems, data mining technology can be divided into the following methods:

(1)Statistical analysis method
This method is widely used in databases, that is, using regression analysis, correlation analysis and other mathematical methods to link the potential relationships among things, so as to find the potential relationships contained in them.

(2)Genetic algorithm
This method draws lessons from the natural genetic mechanism of the biological world, and is a computational model that simulates the evolutionary process of organisms. In this method, a series of solutions are generated by simulating evolution, and the solutions with low correlation are removed by a certain screening principle, and then the optimal solution is finally obtained. Especially, it is widely used in optimization problems in industrial engineering, transportation, economy and other fields.

(3)Decision tree
This method is a kind of analysis and prediction model, which displays the existing data in dendritic form, establishes a decision tree, and makes the next prediction on this basis. This method can help decision-makers better understand and analyze problems.

(4)Visualization technology
This method is to graphically and graphically display the relationship between data in a more vivid and intuitive way, which plays a vital role in data mining.

(5) Bayes network
This method is mainly aimed at the problem of uncertain information. It establishes the interrelation among the elements nodes in a graphical way, and solves the problem when the information is incomplete and uncertain. It is widely used in the system of non-fixed information.

(6)Concept tree method
This method belongs to an abstract data processing, which can classify and arrange data, and is a data preprocessing method.
3. Design of real-time data analysis system for physical training

3.1 Model framework
MAHSUH server is introduced cross over the general framework of network program/server to enhance the connect of every module. The program also require answer to users in time, gather athletes' training demand information by means of data feedback, and the server with great speed tally and response in time, and stores the data in a large database on the server side, in order to bring about intelligent evolve focus [9]. The framework of real-time data analysis system for physical training is shown in Figure 1.

![Fig.1 Real-time data analysis system for physical training](image)

3.2 Data mining of physical training real-time data analysis system
According to the overall structure analysis of sports training immediacy data analysis system, the big data association mining of sports training immediacy data analysis system is carried out, and the self-characteristic sequence $P = \{p_1, p_2, \ldots, p_n\}$ and $Q = \{q_1, q_2, \ldots, q_m\}$ decomposition sequence of sports evaluation association rule data are set. Under the descriptive statistical analysis environment, the cloud computing technology is used for integrated processing and adaptive scheduling of sports data [10]. The characteristic state subspace modal function of sports evaluation decision-making statistical data is:

$$y(t) = \frac{1}{\pi t} \frac{x(\tau)}{t-\tau} d\tau = x(t) \times \frac{1}{\pi t} \quad (1)$$

In the above formula, $P$ is the dominant frequency characteristic of sports evaluation association rule data in statistical time scalar series, $x(t)$ is the scalar length of sports evaluation association rule data, and $\tau$ is the statistical time delay of sports training real-time data analysis system. The statistical information characteristics of sports evaluation association rule data are expressed as follows:

$$s(t) = \sum_{m=\infty}^{\infty} \sum_{n=\infty}^{\infty} a_{mn} g_{mn}(t) + [y_1, y_2, \ldots, y_N]^T \quad (2)$$

Among them, $a_{mn}$ is the relative deviation of information in the real-time data analysis system of physical training in unit statistical time period, and $g_{mn}(t)$ is the error of statistical data. Self-adaptive scheduling is carried out on the data information base of the sports training real-time data analysis system, and the envelope feature quantity of the data is extracted as $z(t)$. The Gaussian information of sports association rule data is reorganized by using spatial undersampling technology, and sports evaluation decision-making and fuzzy scheduling are carried out by combining the data sampling results. After information screening, the mining results of association rule data of sports evaluation decision-making system meet the following requirements:
\[
\int_{-\infty}^{\infty} h(t) dt = 1, h(t) = h(-t), h(t) = 0, |t| > \frac{1}{2}
\] (3)

The method of big data mining is used to mine the information of sports evaluation, and the autocorrelation characteristics of data information flow of sports training real-time data analysis system are extracted. The decision tree classification method is used to screen the characteristics and classify the statistical data \( x(t) \) of sports training real-time data analysis system, and the statistical decision function of the \( n \)-th sports evaluation decision fuzzy decision after data screening is obtained as follows:

\[
\hat{W} = \begin{cases} 
\text{sgn}(W) \|W\| - \alpha T_s & |W| \geq T_s \\
0 & |W| < T_s 
\end{cases}
\] (4)

Among them, \( \alpha \) is the adjustment coefficient of decision tree category, \( W \) is the statistical data sample feature of the physical training real-time data analysis system, and the value range is \( 0 \leq \alpha \leq 1 \). According to the above processing, the fuzzy K centroid clustering method is utilized to carry out the fuzzy weighting of sports evaluation policy decision related data, and the data optimization clustering is realized in the sports evaluation decision-making related data set [11].

3.3 Information feature extraction data

Cross over the focus of sports assess and policy decision, integrate with data mining algorithm, sports related information features are collected, which offer the basis of data input for sports resolve sustain system. Statistical information processing method is used for data collection, VXI bus data collection technology is used to survey sports related feature information, which realizes the stressing of dominate program and offer information input for sports resolve sustain system.

Pure time set of sports assess resolve information is:

\[
x = [x_1, \ldots, x_i, \ldots] \in \mathbb{R}^n
\] (5)

The dynamic divergence characteristic itemize is apply for information fusion, and the sports decision support information data to be mined is decomposed by time-frequency features. This transformation process is described as follows:

\[
y(t) = \int_{a,b} \rho(a,b) \frac{1}{\sqrt{a}} f\left(\frac{t-b}{a}\right) \frac{dadb}{a^2}
\] (6)

In which: \( f(t) \) is the unstationary transient sample worth of features data; \( \rho(a,b) \) is the articulation spread of collective information; \( a \) is the cascade matching parameter; \( b \) is the decomposed spectrum in windowed time-frequency domain.

In sufficient quantity recursive analysis is implementation in the feature space of data spread. When the time defer space of data spread include \( n \) samples, including sample \( x_i, i = 1, 2, \ldots, n \), differential evolution is carried out on the sample data, and the evolved output scalar time sequence is as follows:

\[
x_i = (x_{i1}, x_{i2}, \ldots, x_{in})
\] (7)

Set the threshold \( \xi \), and calculate the feature cluster set number of sports assess resolve data spread in the finite data set, which is expressed as:

\[
V = \{v_j \mid j = 1, 2, \ldots, c; j = 1, 2, \ldots, s\}
\] (8)

In which \( V_i \) is the \( i \)-th time-frequency spread characteristic point of the cluster core of the limited data set, and the output components of data mining are:
Δx = a + (b - a)x_n, n = 1, 2, ..., N (9)

The collective information characteristic spread center in the data characteristic spread space is obtained by using the disturbance variable differential evolution method, and the partition matrix of the cluster center is shown as below:

\[ U = \{ \mu_k | i = 1, 2, \cdots, c; k = 1, 2, \cdots, n \} (10) \]

Normalizing the data, the output mutual information features are as follows:

\[ x_i = \frac{x_i}{\|x\|} = \left( \frac{x_{i1}}{\|x\|}, \frac{x_{i2}}{\|x\|}, \cdots, \frac{x_{in}}{\|x\|} \right) (11) \]

### 4. Development, design and implementation of system software

#### 4.1 Cloud platform service

The real-time data analysis system of sports training uses Elastic Compute Service (ECS), which serves many fields, and puts computing and artificial intelligence technology into every corner of life instead of "high" technology. This product is a agile and to reach the summit computing service, which helps developers decrease costs, heighten run and maintenance efficiency, minimize the tedious work of building and maintaining servers, and enable enterprises or individual users to focus on business innovation and development to the greatest extent.

![Block diagram of server building process](image)

**Fig.2** Block diagram of server building process

Description of cloud server construction: First, choose the performance of cloud server according to
the system requirements. This software platform uses a general cloud server, which is in line with the general usage scenarios of enterprises. Secondly, open the remote connection of the server. When setting up the server, it is necessary to install the application program on the server through IP address remote connection. In addition, the network port should be opened so that other computers with client applications can communicate with the cloud server, send requests to the server, and return desired information. When data communication between client and server is used, Redis real-time database should be started as data cache to ensure efficient and accurate data transmission. The cloud server building process is shown in Figure 2.

4.2 Embedded Linux software development

The recessed Linux technology is apply to expand the system software, and the recessed Linux expand surroundings of sports resolve sustain system is constructed. The data output bus and Central part dominate module of the system delivery the complete works sports assess resolve information from PCI bus to PC cross over decipher and timing dominate, and monitor sports associated data through pwm_ioctl dominate command.

Adopting HP E1485A/B multi mode dominate chip to connect with the upper computer, the VXI bus data purchase of sports assess sustain is implementation under the dominate of MVB bus, contain Linux kernel to start, control data purchase and information supervisory clock sampling, and the code description is as below:

```cpp
class data mining technology :public vpApp
{
    public: Data transmission channel(); //Constructor
    ~myApp(); //Trigger settings
    :initialize/Documents/nfs("vp") //inputarm-linux-gcc
    virtual void executablecodedirectory(Customhpe1432_setTrigger vpAppHP E1562E::Key,int mod) //VXI bus data acquisition
    private:busybox-1.14.2.tar.bz2 //Specify the appropriate compilation tool
}
```

Fig.3 Data storage register configuration process

Use the "tar jxvf busybox-1.14.2.tar.bz2" order to reduce pressure, initialization static alternating quantity and ordinary scenario menu, implement the root file catalogue between the driver and the manipulate system, set up the root file system, implement the root file system, set up the root file system, use "free_irq ()" in Linux system to announce the interrupt, and the "module_init ()" function is called when the module is enter into the kernel, through the main Initialize related logon and RAM of CAN, run "make" and "make install" to translate and edit and install, and realize algorithm compile and program loading of sports resolve sustain system. After completion, "bin, sbin" folder and linuxrc file will be generated under the choose
path portfolio. Execute clock frequency initialize, recall initialize and break off initialize, and finish the configuration of "SPORT0_TCLKDIV" register. The allocation step of sports assess resolve data is described as display in Figure 3.

After the allocation is finish, the drivers running the A/D converter are conscientious for turning on and off the A/D translator. With Qt/Embedded 4.6, the user interface of physical training real-time data analysis system is created, and visual control is realized. "Qt/embedded" is generated under the appoint fix catalogue of the sports assess policy decision system, the top-level "Makefile" file is amend, and cross-compilation is performed to bring about the system build.

5. System debugging and experimental analysis

Selecting football players as experimental objects, simulating the training of football competition team members, as experimental samples. For the sake of guarantee the efficacious of the physical training real-time data analysis system experiment, along with the steady and exact of the experimental step, the experimental data are install equitable, as shown in Table 1.

| Frequency | Model feedback rate(%) | Intensity of training |
|-----------|------------------------|-----------------------|
| 1         | 30.25                  | 0.21                  |
| 2         | 44.21                  | 0.29                  |
| 3         | 56.82                  | 0.31                  |
| 4         | 62.74                  | 0.47                  |
| 5         | 78.96                  | 0.51                  |

Loading SO test software there is no impact on the model, nor will it affect the intelligence under the vision of big data analysis. This software estimate the exact of data acquisition in view of YUH index, and estimate the suitability of the model using Pol parameters. The experimental outcome are display in Figure 4.

In view of the data, it can get that the pol parameter of the intelligent model of sports athletes' training progress is correspondingly steady, invariably on top of 1.0, and this information can arrive a high standard of relevance, and the YUH index display a balmy upward tendency, which explain that the feedback capacity of the model is high.

6. Conclusion

With the rapid increase of industry data and the demand for potential information, data mining technology came into being. With the help of data mining technology, it is not only of high scientific research value, but also of mighty social meaning, which can very advance the expand of sports informatization. In this paper, the optimal design method of real-time data analysis system for sports training is studied, which offer data sustain for scientific sports training. The software is developed in the core construction of embedded Linux, sports data is collected by VIX bus technology, data mining
algorithm in view of collective information characteristic pick up is designed, data mining code is loaded in the program loading module, and the system design and development are realized under the development environment of embedded Linux. Through the simulation experiment of using environment simulation, the outcome show that the immediacy data analysis system of sports training from the visual angle of big data analysis has outstanding practical value, which can make training plans quickly and accurately, resolve the data wave appearance in the tendency of big data separate, and then bring about intelligent run. And the smooth and precise are very high, and it is value popularizing and use vigorously.

References
[1] Wang Pei. Design of an Instant Data Analysis System for Portable Sports Training. Microcomputer applications, vol. 035, no. 006, pp. 107-110, 2019.
[2] Zhang Y, Zhang Y, Zhao X, et al. Design and Data Analysis of Sports Information Acquisition System Based on Internet of Medical Things. IEEE Access, no. 99, pp. 1-1, 2020.
[3] Zhang J, Lv T. Design and Implementation of Interview Training System for Software Technology Major Based Docker Cluster. International Core Journal of Engineering, vol. 5, no. 12, pp. 222-226, 2019.
[4] Liu C, Wang Y. Design of abnormal data detection system for protein gene library based on data mining technology. Cellular and molecular biology (Noisy-le-Grand, France), vol. 66, no. 7, pp. 103, 2020.
[5] Liu Y, Yu Z, Yang Y. Diabetes Risk Data Mining Method Based on Electronic Medical Record Analysis. Journal of Healthcare Engineering, vol. 2021, no. 6, pp. 1-11, 2021.
[6] Zhang L. Design of a sports culture data fusion system based on a data mining algorithm. Personal and Ubiquitous Computing, vol. 24, no. 1, pp. 75-86, 2020.
[7] Ma B, Nie S, Ji M, et al. Research and Analysis of Sports Training Real-Time Monitoring System Based on Mobile Artificial Intelligence Terminal. Wireless Communications and Mobile Computing, vol. 2020, no. 6, pp. 1-10, 2020.
[8] Zhang M, R Qi. Data mining and economic forecasting in DW-based economical decision support system. International journal of reasoning-based intelligent systems, vol. 11, no. 4, pp. 300-307, 2019.
[9] Liu X. A Real-Time Detection Method for Abnormal Data of Internet of Things Sensors Based on Mobile Edge Computing. Mathematical Problems in Engineering, vol. 2021, no. 4, pp. 1-7, 2021.
[10] Yu J. Analysis and Design of Course Website for Software Testing Based on SPOC. Journal of Physics Conference Series, vol. 1187, no. 5, pp. 052015, 2019.
[11] Bi W, Wang G. Local Cultural IP Development and Cultural Creative Design Based on Big Data and Internet of Things. Mobile Information Systems, vol. 2021, no. 3, pp. 1-9, 2021.