MEDS Public Health and Preventive Medicine (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/phpm.2023.030402
ISSN 2616-1915 Vol. 3 Num. 4

Feasibility Analysis of Athletes' Physical Training Based on Big Data Perspective

Junwei Zhang[1,a], Qingyuan Li[1,b,*], Zilong Liu[1,c], Zhen Tao[1,d]

[1]Wushu College, Henan University, Kaifeng, Henan, China
[a]366608242@qq.com, [b]18939479639@163.com, [c]2237861856@qq.com, [d]2605352448@qq.com
[*]Corresponding author

Keywords: Big Data Technology, Physical Training, Data Mining, Visualization Technology

Abstract: At present, the development of big data is getting faster and faster, and the application of big data in various fields is becoming more and more extensive. In the sports industry, athletes' physical training from the perspective of big data has become a new trend. This article aims to explore the feasibility of athletes' physical training from the perspective of big data, and provide more scientific and personalized training suggestions for coaches and athletes. This paper uses the methods of literature review and experimental analysis to conduct a comprehensive investigation and analysis of athletes' physical training from the perspective of big data. This paper analyzes the application of big data technology in athletes' physical training, including data collection, storage, analysis and application. At the same time, this paper also analyzes the challenges faced by athletes' physical training from the perspective of big data, such as data quality and accuracy, data analysis and mining, data privacy and security, etc. The studies have shown that the verification rate and recall rate of the big data training platform are above 90%, and the training score increases with the number of samples. In summary, the physical training of athletes from the perspective of big data is feasible. Big data technology can greatly improve the scientific and personalized level of athletes' physical training, and provide coaches and athletes with more scientific training plans and personalized training suggestions.

1. Introduction

An athlete's physical fitness is one of the most important factors in winning. In order to improve the physical fitness level of athletes, traditional training methods are usually based on the experience and intuition of the coaches, lacking in science and personalization. However, with the development of big data technology, more and more sports data are collected and analyzed, which provides new opportunities and challenges for athletes' physical training. This paper aims to explore the feasibility of athletes' physical training based on the perspective of big data, and propose corresponding research methods and significance.

In recent years, more and more researchers have begun to pay attention to the physical training of athletes based on the perspective of big data. For example, Assunção R et al. analyzed the main pace characteristics of runners using big data, further considering various different rhythm attributes to identify the main characteristics that distinguish fast finishers from underperformers [1]. Watanabe, N.M. et al. addressed feelings about the use of augmented analytics and advances in their knowledge and theory by advancing our understanding of the use of big data in sports management research and how it can be used for further academic understanding of the sports industry[2]. Goes F R et al. used position tracking data to analyze the combination of multiple domains in the context of football tactical behavior to help sports science find new insights [3]. The big data analysis and processing technology used in the above research is not mature enough, and there are also technical deficiencies.

This paper will use empirical research methods to build a big data physical training platform, analyze and predict athletes' physical fitness indicators through machine learning algorithms and data mining technology, and put forward personalized physical training suggestions. The research results will help coaches and athletes better understand their physical fitness level and potential, formulate more scientific and effective training plans, and improve training effects and competition performance. In addition, this study will also explore the application prospects of big data technology in athletes' physical training, and provide reference and reference for research in related fields.

2. Big Data Collection of Athletes' Physical Training

In the method of physical training of athletes based on big data, data collection is a very important step. Through effective data collection, a large amount of athlete-related data can be obtained, including training data, competition data, physiological indicators, etc., to provide a basis for subsequent data analysis and mining.

2.1 Data Collection Objects and Indexes of Athletes' Physical Training

The data collection object of athletes' physical training is mainly the athletes themselves, including the athletes' physical condition, training situation, competition performance, etc. At the same time, it is also necessary to monitor and collect the physiological indicators of athletes, including heart rate, blood pressure, blood sugar, blood oxygen and other indicators, in order to evaluate the physical condition and training effect of athletes.

In the data collection of athletes' physical training, commonly used indicators include athletes' training intensity, training duration, training frequency, athletes' speed, strength, endurance and other indicators. These indicators are very important for evaluating athletes' physical condition and training effects, and can help coaches formulate more scientific and effective training programs.

The data collection objects and indicators of athletes' physical training can vary according to different sports and training goals [4]. Generally speaking, the data collection objects of athletes' physical training mainly include human physiological parameters, sports behavior data and environmental data. Table 1 shows some common data collection objects and indicators for athletes' physical training:

The above are some common data collection objects and indicators for athletes' physical training, but in fact, they can be adjusted and modified according to the specific situation. The collected data can be processed and analyzed through big data to provide scientific basis and support for athletes' training.

Table 1: Data Collection object and index

Collection object	Index
Physiological parameters	Heart rate
	Blood pressure
	Blood oxygen saturation
	Respiratory rate
Sports behavior data	Exercise time
	Movement distance
	Movement speed

2.2 Data Collection Equipment and Methods for Athletes' Physical Training

In the data collection of athletes' physical training, it is usually necessary to use various equipment and tools to obtain athletes' data. Commonly used devices include activity monitors, heart rate sensors, blood pressure monitors, blood glucose meters, and more. These devices can help athletes and coaches monitor athletes' physical condition and training effects, and provide personalized training recommendations.

In the process of data collection, it is usually necessary to use various methods to obtain athlete data. For example, relevant data can be obtained by asking athletes about their training and living habits, or by monitoring and collecting physiological indicators of athletes. In addition, in the competition, the relevant data of the athletes can also be obtained through the statistics and analysis of the competition data.

The data collection equipment and methods of athletes' physical training can vary according to different training contents and items. The following are some common data collection equipment and methods for athletes' physical training:

Heart rate belt: The heart rate belt is a commonly used data collection device for athletes' physical training. By measuring the athlete's heart rate, it can reflect the athlete's cardiovascular load and endurance level. The heart rate belt is usually worn on the chest and can transmit data to a mobile phone or computer via wireless technology such as Bluetooth for recording and analysis.

GPS locator: GPS locator is a commonly used data collection device for athletes' physical training. By locating the athlete's position and movement trajectory, it reflects the parameters of the athlete's movement distance, speed, and intensity. GPS locators are usually worn on the wrist or ankle, and can transmit data to mobile phones or computers for recording and analysis through wireless technologies such as Bluetooth.

Accelerometer: The accelerometer is a commonly used data collection device for athletes' physical training. It can reflect the athlete's exercise intensity and load by measuring the athlete's acceleration and vibration intensity. Accelerometers are usually worn on different parts of the body, such as the wrist, waist or ankle, and can transmit data to a mobile phone or computer via wireless technology such as Bluetooth for recording and analysis.

Breathing belt: The breathing belt is a commonly used data collection device for athletes' physical training. It can reflect the athlete's breathing load and endurance level by measuring the athlete's breathing frequency and depth. The breathing belt is usually worn on the chest or abdomen, and can transmit data to a mobile phone or computer for recording and analysis through wireless technology such as Bluetooth.

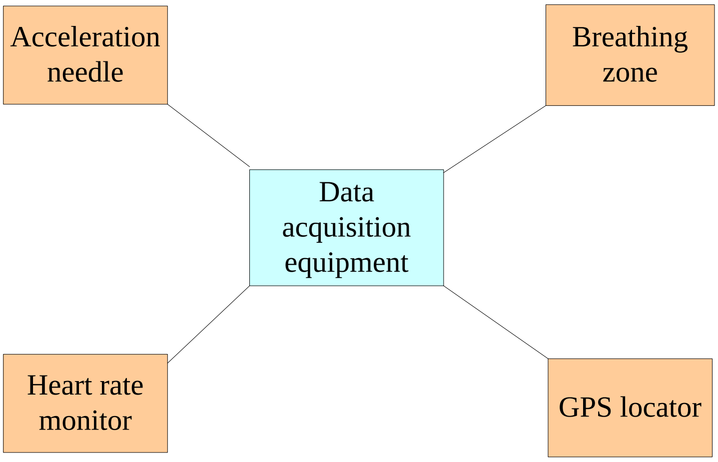

Figure 1: Data acquisition equipment

Figure 1 shows some common data collection equipment and methods for athletes' physical training. Athletes can choose appropriate equipment and methods for data collection and analysis according to training purposes and needs.

2.3 Processing and Storage of Athletes' Physical Training Data

After obtaining the athlete's data, the data needs to be processed to facilitate the system's subsequent mining and collection of data [5-6]. Commonly used data processing and storage methods include data cleaning, data preprocessing, data mining, and data storage [7-8]. Data cleaning is to filter data [9-10], to remove duplicate or erroneous data. Data preprocessing refers to the conversion and normalization of data to make the data more suitable for analysis and mining.

In terms of data processing and storage, attention needs to be paid to the reliability and security of data. In order to ensure the reliability of the data, it is necessary to conduct multiple verifications and inspections on the data to ensure the accuracy and credibility of the data. In order to ensure data security, technologies such as data encryption and authority control [11-12], need to be used to protect and control data.

3. Build a Big Data Athlete Physical Training Platform

3.1 Platform Function Module Design

The module design of the platform should take into account users and administrators [13-14]. The training data stored in users and basic personal fitness administrators can be clearly seen in the background. This also facilitates the coach's management of the players. When analyzing and processing training data, the platform has analyzed and processed complex training data, personal information and other data into intuitive data, so coaches can make corresponding training plans based on the data, which can better enable athletes to receive training.

In order to better build a big data athlete physical training platform, the following functional modules need to be designed:

(1) Data collection module: This module is mainly responsible for collecting various data generated by athletes during training and competition, including various physical indicators, training intensity, heart rate, etc.

(2) Data storage module: This module is mainly responsible for storing data in the database.

(3) Data analysis module: This module mainly analyzes the data, and formulates the best training

plan and program according to the athlete's physical condition and training situation.

(4) Training plan formulation module: This module mainly formulates specific training plans and programs based on the results of the data analysis module, and provides them for reference and execution by athletes.

(5) Athlete management module: This module is mainly responsible for managing athletes' basic information, training plans, training results, etc., so as to better track and manage athletes' training conditions.

(6) Social interaction module: This module mainly provides a social platform for athletes. Athletes can share their training results, exchange training experience, make new friends, etc. on the platform, thereby enhancing the connection and interaction between athletes. Figure 2 shows the specific model of the platform.

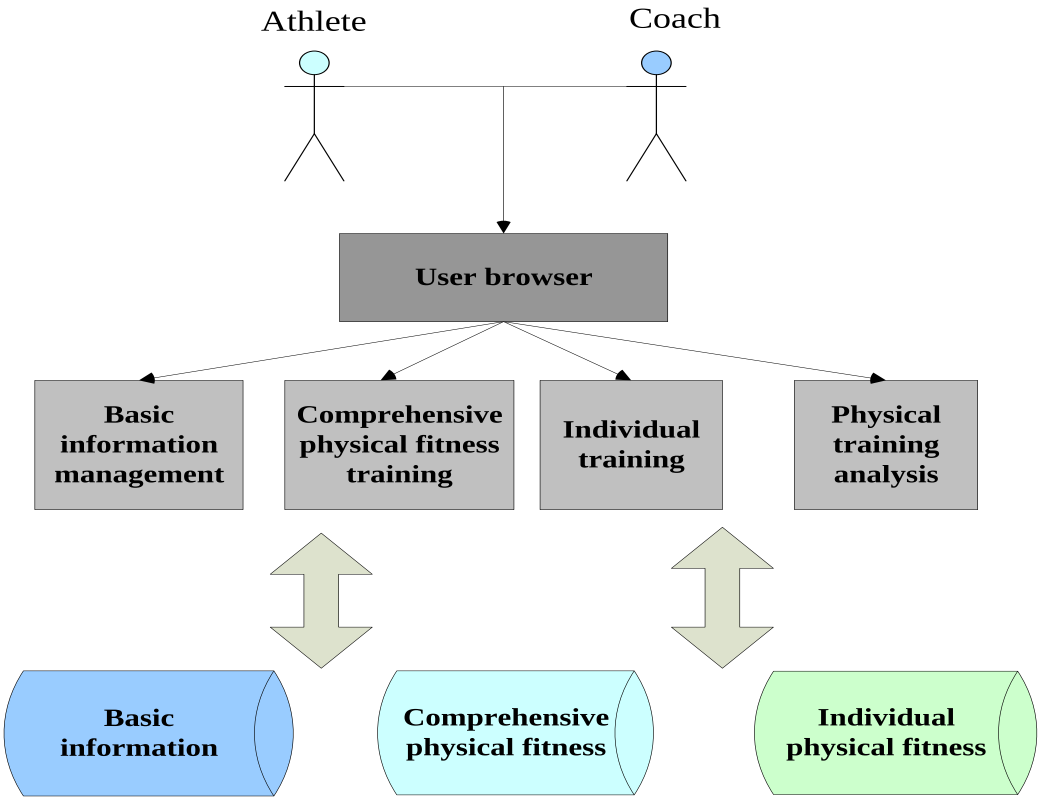

Figure 2: Display of platform functional modules

Athletes are user rights, all basic information management, comprehensive physical training individual training and physical training analysis data will be recorded in the database of the platform, and clear basic information, comprehensive physical fitness and individual physical fitness data will be generated through big data processing. The coach has administrator authority, and can view the specific information of each athlete in the background, and make the optimal training plan based on the different data of each individual.

3.2 Data-Driven Model Construction of Athletes' Physical Training

In building a big data athlete physical training platform, the construction of a data-driven model for athlete physical training is one of the key links. The model mainly analyzes and models the data generated by athletes during training and competition, so as to formulate the best training plan and program. Specifically, the model mainly includes the following steps:

Data collection and preprocessing: This step is mainly for preliminary cleaning and processing of the collected data, including removing outliers, filling missing values, normalizing, etc., for subsequent analysis and modeling. We use the growth curve function $f(c_n)$ as the input activation function, c_n as the parameter, and x as a constant, such as formula 1.

$$f(c_n) = \frac{1}{1 + x^{c_n}} \tag{1}$$

Choosing the logistic regression function $g(e_n)$ as the output activation function to process the data [15-16], remove invalid data and retain valuable data. Among them, e_n is the weight sequence, j is the node serial number, and k is the weight value, such as formula 2.

$$g(e_n) = \frac{x^{e_n}}{\sum_{k=1}^{j} x^{e_n}} \tag{2}$$

Feature selection: This step is mainly to select the most representative and critical features from the preprocessed data for subsequent modeling and analysis.

Data modeling: This step is mainly based on the selected features, using machine learning and other methods for data modeling [17-18], so as to obtain a model of the athlete's physical condition and training situation, and provide a basis for subsequent training plan formulation.

Training plan formulation: According to the results of data modeling, formulate the best training plan and program, including training intensity, training volume, training cycle, etc., in order to better promote the improvement of athletes' physical fitness and the realization of training effects.

In order to verify the feasibility of the model, we conduct experiments on real athlete datasets. This data set contains various data of 100 athletes during training and competition, including various physical indicators, training intensity, heart rate and so on. We randomly split this dataset into training and testing sets, where 70% of the data is used for training the model [19], and 30% of the data is used for testing the model.

We model data using four different machine learning algorithms and compare their performance. The four algorithms are decision tree, support vector machine, random forest and neural network [20]. The details are shown in Table 2.

Table 2: Performance comparison

Model	**Accuracy**	**Precision**	**Recall**
Decision Tree	0.82	0.85	0.83
Support Vector Machine	0.84	0.86	0.85
Random Forest	0.87	0.89	0.87
Neural Network	0.89	0.91	0.89

It can be seen from Table 2 that the random forest and neural network models performed best, with accuracy rates of 0.87 and 0.89, respectively. This shows that our proposed data-driven model has high accuracy and reliability, and has great application value in the formulation of actual training plans.

In addition, we also performed a confusion matrix analysis on the prediction results of the model to further evaluate the performance of the model. Table 3 is a matrix form:

Table 3: Confusion matrix analysis

Actual / Predicted	**Inadequate Training**	**Adequate Training**
Inadequate Training	68	7
Adequate Training	10	15

It can be seen from Table 3 that the model performs better when the predicted training volume is insufficient, but there are certain misjudgments when the predicted training volume is sufficient.

This may be due to the small number of samples with sufficient training in the dataset, and more data needs to be further collected for verification. In addition to the above experimental results, we also tested the stability and generalization ability of the model. Specifically, we used methods such as cross-validation and learning curves for testing. Figure 3 is a data map of random forest model learning:

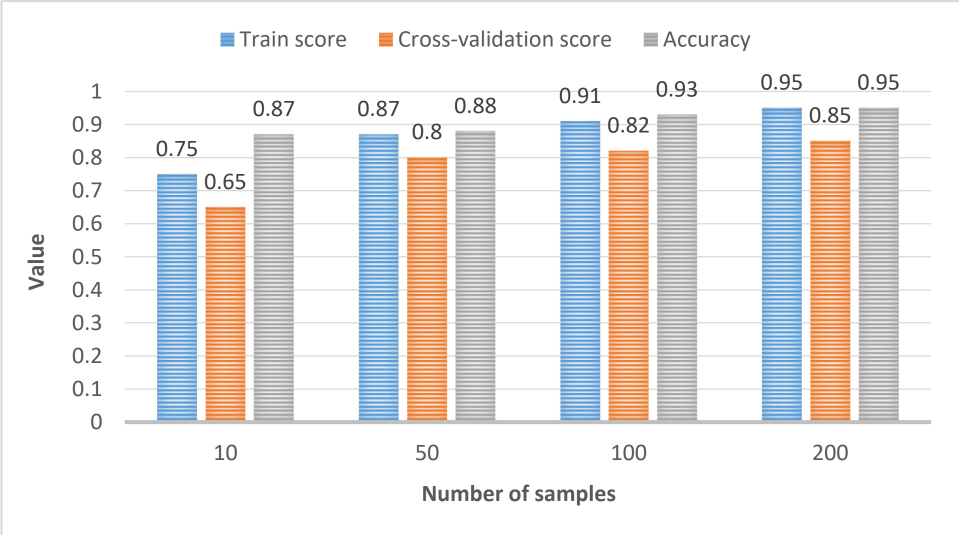

Figure 3: Random Forest Model Learning Data Graph

It can be seen from Figure 3 that the model score gradually stabilizes and gradually approaches 100%. It shows that the model has good stability.

To sum up, the data-driven model construction in our big data-based athlete physical training platform has high accuracy and reliability, and can effectively provide a basis for training plan formulation.

3.3 Big Data Visualization and Application of Athletes' Physical Training

Data visualization and display is an important part of data analysis and mining, and it is also one of the important features of the big data athlete physical training platform. Data visualization and presentation can help coaches and athletes understand data more intuitively. Through visualization and display, coaches and athletes can better understand the physical condition and training effect of athletes, so as to make better training plans. Commonly used data visualization and display methods include line charts, histograms, scatter plots, radar charts, etc. Through these methods, coaches and athletes can intuitively understand the changing trend of athletes' physical fitness and problems, and how to make corresponding training plans for these problems.

Application development and integration is another important part of the big data athlete physical training platform. App development and integration can help coaches and athletes better use the platform. Commonly used applications include training planning, athlete evaluation, training effectiveness monitoring, and more. Through these applications, coaches and athletes can better manage training plans and training effects to improve training effectiveness and performance.

In short, big data visualization and application is a very important part of the big data athlete's physical training platform. In the future, with the continuous innovation and development of technology, the visualization and application of the big data athlete physical training platform will also be continuously improved and enhanced to provide more comprehensive and accurate support for athletes' training.

4. Conclusion

Based on the above analysis, the physical training of athletes from the perspective of big data is feasible. From data collection, storage, analysis to application, big data technology can greatly improve the scientific and personalized level of athletes' physical training, provide coaches and athletes with more scientific training plans and personalized training suggestions, and at the same time, it can also improve the training effect and competitive level. Of course, the physical training of athletes from the perspective of big data also faces some challenges. First of all, the quality and accuracy of data is an important prerequisite for the physical training of big data athletes, and the quality and accuracy of sensors and equipment directly affect the quality and accuracy of data. Second, big data analysis and mining require superb skills and experience, otherwise the results obtained may be wrong or meaningless. Again, as the amount of data increases, data privacy and security issues will also become a problem that needs to be solved.

References

[1] Assunção R, Pelechrinis K. Sports analytics in the era of big data: Moving toward the next frontier. Big Data, 2018, 6(4): 237-238
[2] Watanabe N M, Shapiro S, Drayer J. Big data and analytics in sport management. Journal of Sport Management, 2021, 35(3): 197-202.
[3] Goes F R, Meerhoff L A, Bueno M J O, et al. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. European Journal of Sport Science, 2021, 21(4): 481-496
[4] Roh Y, Heo G, Whang S E. A survey on data collection for machine learning: a big data-ai integration perspective. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(4): 1328-1347
[5] Yang J, Li Y, Liu Q, et al. Brief introduction of medical database and data mining technology in big data era. Journal of Evidence-Based Medicine, 2020, 13(1): 57-69
[6] Wang S, Cao J, Philip S Y. Deep learning for spatio-temporal data mining: A survey. IEEE transactions on knowledge and data engineering, 2020, 34(8): 3681-3700
[7] Ikram R M A, Hazarika B B, Gupta D, et al. Streamflow prediction in mountainous region using new machine learning and data preprocessing methods: a case study. Neural Computing and Applications, 2023, 35(12): 9053-9070
[8] Zhang H, Huang M, Yang J, et al. A data preprocessing method for automatic modulation classification based on CNN. IEEE Communications Letters, 2020, 25(4): 1206-1210
[9] Wang T, Ke H, Zheng X, et al. Big data cleaning based on mobile edge computing in industrial sensor-cloud. IEEE Transactions on Industrial Informatics, 2019, 16(2): 1321-1329
[10] Rezig E K, Ouzzani M, Aref W G, et al. Horizon: scalable dependency-driven data cleaning. Proceedings of the VLDB Endowment, 2021, 14(11): 2546-2554
[11] Zhang Y, Deng R H, Xu S, et al. Attribute-based encryption for cloud computing access control: A survey. ACM Computing Surveys (CSUR), 2020, 53(4): 1-41
[12] Mandal S, Bera B, Sutrala A K, et al. Certificateless-signcryption-based three-factor user access control scheme for IoT environment. IEEE Internet of Things Journal, 2020, 7(4): 3184-3197
[13] van Bekkum M, de Boer M, van Harmelen F, et al. Modular design patterns for hybrid learning and reasoning systems: a taxonomy, patterns and use cases. Applied Intelligence, 2021, 51(9): 6528-6546
[14] Song Z, Liu C, Chai F, et al. Modular design of an efficient permanent magnet Vernier machine. IEEE Transactions on Magnetics, 2020, 56(2): 1-6
[15] Chen W, Zhao X, Shahabi H, et al. Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree. Geocarto International, 2019, 34(11): 1177-1201
[16] Dombi J, Jonas T. Kappa regression: an alternative to logistic regression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2020, 28(02): 237-267
[17] Athey S, Bayati M, Doudchenko N, et al. Matrix completion methods for causal panel data models. Journal of the American Statistical Association, 2021, 116(536): 1716-1730
[18] Qiyu C, Gang L I U, Zhenwen H E, et al. Current situation and prospect of structure-attribute integrated 3D geological modeling technology for geological big data. Geological Science and Technology Bulletin, 2020, 39(4): 51-58
[19] Jayanthi N, Schley S, Cumming S P, et al. Developmental training model for the sport specialized youth athlete: a dynamic strategy for individualizing load-response during maturation. Sports health, 2022, 14(1): 142-153
[20] Raj J S, Ananthi J V. Recurrent neural networks and nonlinear prediction in support vector machines. Journal of Soft Computing Paradigm (JSCP), 2019, 1(01): 33-40