

# Exploration and Practice of Using Data Analysis to Improve the Effectiveness of Sports Training

Jianjie Gao<sup>1, a</sup>

<sup>1</sup>Shanghai Yanyu Technology Co., LTD, Shanghai, China

<sup>a</sup>runtech007@qq.com

## Abstract

With the development of technology, the application of data analysis in the field of sports is becoming increasingly widespread, and using data analysis to improve the effectiveness of sports training has become a research hotspot in the current sports industry. This article explores and practices data analysis methods in sports training, collection and processing of sports training data, evaluation and analysis of sports training effects, and data-driven optimization strategies for sports training. The research aims to provide useful reference and inspiration for the development of sports in China, promote the scientific and intelligent process of sports training, and improve the training effectiveness and competitive level of athletes.

## Keywords

Data analysis; Sports training; Training effect.

## 1. Introduction

With the development of technology and the arrival of the big data era, the application of data analysis in various fields is becoming increasingly widespread, and the sports field is no exception. Sports training, as a key link in improving the competitive level of athletes, is of great significance for the development of the sports industry. However, traditional sports training mainly relies on the experience of coaches and is easily constrained by individual differences, single training methods, and other factors. Therefore, how to use data analysis methods to scientifically and effectively improve the effectiveness of sports training has become a focus of attention in the current sports industry.

## 2. Data Analysis Methods in Sports Training

### 2.1. Basic Concepts and Technologies

#### 2.1.1. Data Collection

Data collection is the foundation of sports training data analysis, which involves the collection of various data during the athlete training process. These data include physical fitness indicators, technical parameters, biomechanical data, psychological status, etc. The methods of data collection are divided into two categories: manual collection and automated collection. Manual collection mainly relies on coaches, researchers, or athletes themselves to record and report data during the training process, such as speed, strength, endurance, and other indicators. Automated collection is the use of modern technological means, such as sensors, cameras, motion analysis systems, etc., to capture and record real-time data during the training process. During the data collection process, attention should be paid to the accuracy and completeness of the data to ensure the quality and effectiveness of subsequent data analysis[1]. In sports training, the main purpose of data collection is to understand the training status and development trends of athletes, providing a basis for training plan formulation, training effect

evaluation, and training optimization. Therefore, the scope of data collection covers various indicators such as the physiological, psychological, and technical aspects of athletes. In addition, data collection should also focus on individual differences and be customized according to the characteristics of different athletes. The comprehensive use of manual and automated collection methods can comprehensively and objectively grasp the training situation of athletes, providing strong support for improving training effectiveness. With the development of technology, the application of data collection technology in sports training will become increasingly widespread, which is of great significance for improving the competitive level of Chinese athletes.

### **2.1.2. Data preprocessing**

Data preprocessing is an important part of sports training data analysis, which cleans, integrates, and standardizes the original data for subsequent data analysis and application. In sports training, data preprocessing mainly includes the following aspects:

Data cleaning is the first step in preprocessing, aimed at removing invalid data and outliers, and improving the quality of data. The methods of data cleaning include deleting duplicate records, filling in missing values, and correcting data errors. This step is crucial for ensuring the accuracy of data analysis results[2].

Data integration is the process of unifying and integrating data from different sources, formats, and types for subsequent analysis. In sports training, data integration mainly includes the integration of athlete basic information, training data, competition data, and other aspects.

Data normalization is the process of converting data into a unified data format and unit for subsequent data mining and analysis. Data normalization includes standardizing and normalizing data to eliminate the impact of dimensional and numerical differences on data analysis.

The quality of data preprocessing is directly related to the effectiveness of data analysis. Therefore, in sports training, researchers need to pay full attention to data preprocessing work. Through data preprocessing, data can be made cleaner, more standardized, and more unified, providing strong support for subsequent data analysis. With the continuous increase in the amount of sports training data, the application of data preprocessing technology in the field of sports will become increasingly important, which will help improve the scientificity and effectiveness of sports training in China.

### **2.1.3. Data visualization**

Data visualization is a crucial step in the process of analyzing sports training data. It presents abstract data in an intuitive and visual way, making it easier for coaches, athletes, and researchers to understand and analyze the data. In sports training, data visualization can be applied to multiple aspects, such as athlete technical analysis, training plan evaluation, and competition strategy formulation.

Data visualization techniques include various forms such as line charts, bar charts, scatter plots, radar charts, etc. Each form has its specific application scenario. For example, a line chart is suitable for displaying the trend of athletes over a period of time, a bar chart is suitable for comparing various indicators of different athletes or training stages, a scatter plot is suitable for analyzing the relationship between athletes' technical movements and performance, and a radar chart is suitable for displaying the comprehensive performance of athletes in multiple aspects[3].

### **2.1.4. Data Mining and Machine Learning**

Data mining and machine learning are important methods for analyzing sports training data. By mining and analyzing a large amount of data, they discover the potential patterns and advantages of athletes, providing a basis for formulating more scientific training plans and

improving training effectiveness. Data mining mainly includes techniques such as classification, clustering, and association rule mining, which are used to deeply analyze various data of athletes and mine valuable information. Machine learning, on the other hand, learns and predicts training data by constructing algorithmic models, thereby achieving evaluation and prediction of athlete training effectiveness[4].

In sports training, data mining and machine learning can be applied in multiple aspects, such as athlete technical analysis, performance prediction, and training plan optimization. Through data mining and machine learning, the patterns of athletes at different training stages can be discovered, providing support for developing more personalized training plans. Meanwhile, data mining and machine learning can also predict the future performance of athletes, providing reference for competition strategy formulation and athlete selection.

## **2.2. Common Data Analysis Methods**

### **2.2.1. Descriptive statistical analysis**

Descriptive statistical analysis is an important method in sports training data analysis. It summarizes, describes, and analyzes various data of athletes to objectively and comprehensively understand their training status, strength level, and potential. Descriptive statistical analysis mainly includes the statistical summary of data, the drawing of statistical charts, and the extraction of data features[5].

In sports training, descriptive statistical analysis can be applied to multiple aspects, such as athlete technical analysis, performance evaluation, and training plan formulation. Through descriptive statistical analysis, coaches, athletes, and researchers can quickly understand the training effectiveness, performance changes, and competitive status of athletes, providing a basis for developing more scientific training plans. Meanwhile, descriptive statistical analysis can also help athletes understand their strengths and weaknesses, thereby adjusting training methods and improving training effectiveness.

### **2.2.2. Correlation analysis**

Correlation analysis is an important method in sports training data analysis, which is used to measure the degree of correlation between different variables and reveal their intrinsic connections. In sports training, correlation analysis can be applied to multiple aspects, such as athlete technical analysis, training plan evaluation, and competition strategy formulation. Through correlation analysis, coaches, athletes, and researchers can understand the correlation between different variables, thereby identifying key factors that affect athlete training effectiveness and providing a basis for developing more scientific training plans.

In correlation analysis, commonly used correlation coefficients include Pearson correlation coefficient, Spearman correlation coefficient, and judgment coefficient. These methods can be used to measure the strength of linear, monotonic, or overall relationships between variables. Pearson correlation coefficient is a method used to measure the linear relationship between two continuous variables, with values ranging from -1 to 1, where -1 represents complete negative correlation, 1 represents complete positive correlation, and 0 represents no linear relationship.

### **2.2.3. Causal relationship analysis**

Causal analysis is an important method in sports training data analysis, which is used to explore the causal relationships between different variables and reveal their intrinsic connections. In sports training, causal analysis can be applied to multiple aspects, such as athlete technical analysis, training plan evaluation, and competition strategy formulation. Through causal analysis, coaches, athletes, and researchers can understand the causal relationships between different variables, thereby identifying key factors that affect athlete training effectiveness and providing a basis for developing more scientific training plans.

Causal relationship analysis mainly includes experimental design and causal inference methods. The experimental design is used to explore the causal relationship between different variables, analyze the size and direction of the causal relationship by comparing the differences between the experimental group and the control group. The causal inference method is used to infer causal relationships based on existing data, providing a basis for training plan formulation and athlete adjustment.

#### **2.2.4. Prediction analysis**

Predictive analysis is a key method in sports training data analysis. Through mining and analyzing historical data, it predicts the future performance and status of athletes, providing a basis for formulating more scientific training plans and competition strategies. In sports training, predictive analysis can be applied to multiple aspects, such as athlete performance prediction, training plan optimization, opponent analysis, and so on.

Predictive analysis methods mainly include regression analysis, time series analysis, machine learning, etc. Regression analysis is used to explore the linear relationship between independent and dependent variables, in order to predict future development trends. Time series analysis is used to analyze time series data, identify its inherent patterns, and predict future development trends. Machine learning evaluates and predicts the training effectiveness of athletes by constructing algorithmic models to learn and predict training data.

### **3. Collection and Processing of Sports Training Data**

#### **3.1. Data sources**

The sources of sports training data are very extensive, covering various physiological, psychological, technical, and competition data of athletes. These data come from multiple channels, such as training sites, medical monitoring, competition sites, athlete self reports, etc. The source of data is crucial for ensuring the effectiveness and accuracy of data analysis. Therefore, before conducting data analysis, it is necessary to fully understand the source and collection method of data to ensure its quality and reliability.

In sports training, the sources of data mainly include the following aspects: basic information of athletes, training data, competition data, physiological data, psychological data, and biomechanical data. These data can be collected through various methods, such as sensors, cameras, motion analysis systems, etc. In addition, data from different sources can be integrated through a data integration platform to facilitate subsequent data analysis and application[6].

#### **3.2. Data collection methods**

There are various methods for collecting sports training data, including manual and automated collection. Manual collection mainly relies on coaches, researchers, or athletes themselves to record and report data during the training process, such as speed, strength, endurance, and other indicators. Automated collection is the use of modern technological means, such as sensors, cameras, motion analysis systems, etc., to capture and record real-time data during the training process.

In sports training, the method of data collection needs to be selected according to the actual situation. For example, automated collection methods can be used for physiological data that requires real-time monitoring, such as heart rate, blood pressure, etc. For some data that requires long-term tracking, such as athlete's technical movements, competition results, etc., a combination of manual and automated collection methods is needed. In addition, for some special data, such as the psychological state of athletes, biomechanical data, etc., specialized data collection equipment and instruments need to be used for collection.

### 3.3. Data storage and transmission

Data storage and transmission are crucial steps in the analysis process of sports training data, which involve the security, reliability, and real-time performance of the data. In sports training, the storage and transmission of data need to consider the following aspects:

**Data storage.** The amount of sports training data is huge and complex, so efficient and reliable data storage methods are needed. Common data storage methods include relational databases, non relational databases, and distributed storage. These storage methods can ensure the security, integrity, and traceability of data[7].

**Data transmission.** The real-time and speed of data transmission are crucial for the analysis of sports training data. Common data transmission methods include wired transmission, wireless transmission, and cloud transmission. These transmission methods can ensure the real-time and speed of data, making it easier for coaches, athletes, and researchers to obtain and analyze data in a timely manner.

**Data security and privacy.** Data security and privacy protection are important issues in the storage and transmission of sports training data. It is necessary to use encryption technology, identity authentication and other means to ensure the security and privacy of data.

### 3.4. Data preprocessing

Data preprocessing is an important part of sports training data analysis, which includes steps such as data cleaning, data integration, data transformation, and data reduction. Data cleaning mainly involves removing outliers and filling in missing values from raw data to improve the quality and reliability of the data. Data integration is the process of unifying and integrating data from different sources and formats for subsequent analysis. Data transformation is the process of standardizing and normalizing data to eliminate the impact of dimensional and numerical differences on data analysis. Data reduction is the process of dimensionality reduction and feature selection on data to reduce the amount of data and improve the efficiency and effectiveness of data analysis.

In sports training, data preprocessing needs to be selected and applied according to the actual situation. For example, for some physiological data that requires real-time monitoring, data cleaning and transformation methods can be used. For some data that requires long-term tracking, such as athlete's technical movements, competition results, etc., data integration and data reduction methods need to be adopted.

## 4. Evaluation and Analysis of Sports Training Effectiveness

### 4.1. Training effectiveness evaluation indicators

Training effectiveness evaluation indicators are important tools for measuring the results of sports training, and they can objectively reflect the changes and progress of athletes during the training process. The evaluation indicators for training effectiveness need to have characteristics such as objectivity, quantifiability, and ease of operation, which can accurately reflect the changes and progress of athletes during the training process. The evaluation indicators for training effectiveness need to be comprehensive and targeted, covering multiple aspects of athletes such as physiology, technology, psychology, etc., in order to comprehensively evaluate the training effectiveness of athletes. The evaluation indicators for training effectiveness need to be comparable and able to compare and analyze between different athletes at different times[8].

### 4.2. Application of Common Analysis Methods

In the evaluation and analysis of sports training effectiveness, common analysis methods include descriptive statistical analysis, correlation analysis, causal analysis, and predictive



analysis. Descriptive statistical analysis is used to summarize, describe, and analyze various data of athletes, in order to objectively and comprehensively understand their training status, strength level, and potential. Correlation analysis is used to measure the degree of correlation between different variables, in order to reveal their intrinsic connections. Causal analysis is used to explore the causal relationships between different variables, in order to identify key factors that affect the effectiveness of athlete training. Predictive analysis is used to predict the future performance and state of athletes, providing a basis for developing more scientific training plans and competition strategies.

In the evaluation and analysis of sports training effectiveness, the application of these common analysis methods needs to be selected and applied according to the actual situation. For example, descriptive statistical analysis can be used to summarize and describe the training data of athletes; Correlation analysis can be used to analyze athlete training data and identify key factors that affect training effectiveness; Causal analysis can be used to explore the relationship between athlete training effectiveness and factors such as training methods and diet; Predictive analysis can be used to predict the future performance and condition of athletes.

### 4.3. Practical Analysis

In track and field events, training effectiveness evaluation and analysis are mainly applied to the technical movements, speed, strength, and endurance of athletes. By conducting descriptive statistical analysis, correlation analysis, and predictive analysis on the training data of athletes, coaches can understand their technical characteristics, identify key factors that affect training effectiveness, and predict their future performance and status.

In ball games, training effectiveness evaluation and analysis are mainly applied to the technical, tactical, speed, and psychological aspects of athletes. By conducting descriptive statistical analysis, correlation analysis, and causal analysis on the training data of athletes, coaches can understand the technical characteristics, tactical application, speed, and psychological changes of athletes, identify key factors that affect training effectiveness, and predict the future performance and state of athletes[9].

In water sports, training effectiveness evaluation and analysis are mainly applied to athletes in terms of speed, strength, endurance, and technique. By conducting descriptive statistical analysis, correlation analysis, and predictive analysis on the training data of athletes, coaches can understand the changes in their speed, strength, endurance, and skills, identify key factors that affect training effectiveness, and predict their future performance and status.

## 5. Data-driven Optimization Strategies for Sports Training

### 5.1. Training Plan Optimization

The optimization of training plans is a crucial step, which analyzes the physiological, psychological, technical, and competition data of athletes to identify key factors that affect training effectiveness, providing a basis for developing more scientific and personalized training plans.

In the optimization of training plans, coaches, athletes, and researchers need to fully understand the actual situation of athletes, including age, physical fitness, technical characteristics, etc. At the same time, they also need to pay attention to factors such as the psychological state and injury situation of athletes to ensure the effectiveness and safety of the training plan. In addition, they also need to develop targeted training plans based on training objectives, competition tasks, and the actual situation of athletes to improve training effectiveness.

## 5.2. Monitoring and adjustment of athlete status

Athlete status monitoring and adjustment are important components of data-driven sports training optimization strategies. By monitoring and analyzing the physiological, psychological, technical, and competition data of athletes in real-time, coaches, athletes, and researchers can understand the training status, fatigue level, and competitive level of athletes, thereby adjusting training plans, optimizing training methods, and improving their condition in a timely manner. In the monitoring and adjustment of athlete status, coaches, athletes, and researchers need to pay attention to the physiological indicators of athletes, such as heart rate, blood pressure, lactate, etc., in order to understand the fatigue level and physical condition of athletes. At the same time, they also need to pay attention to the psychological state of athletes, such as emotions, confidence, etc., in order to understand their psychological condition and competitive level. In addition, they also need to comprehensively evaluate the athlete's condition based on factors such as their technical movements and competition results, in order to develop targeted adjustment plans.

## 5.3. Tactical Analysis and Optimization

Tactical analysis and optimization are important components of data-driven sports training optimization strategies. By analyzing the technical, tactical, speed, and psychological aspects of athletes and opponents, coaches, athletes, and researchers can understand the tactical characteristics of athletes, identify key factors that affect competition results, and thus develop more scientific and effective competition strategies.

In tactical analysis and optimization, coaches, athletes, and researchers need to collect match data from athletes and opponents, including technical movements, match results, and number of fouls. Then, they need to conduct descriptive statistical analysis, correlation analysis, and causal analysis on these data to identify the key factors that affect the competition results. In addition, they also need to develop targeted competition strategies based on the actual situation of the athletes to improve the effectiveness of the competition[10].

## 5.4. Training equipment and venue optimization

Training equipment and venue optimization are important components of data-driven sports training optimization strategies. By collecting and analyzing data on the equipment, venue, and related environment used by athletes during training, coaches, athletes, and researchers can identify key factors that affect training effectiveness, optimize training equipment, venue, and environment, and improve training effectiveness.

In the optimization of training equipment and venues, coaches, athletes, and researchers need to pay attention to the equipment used by athletes during the training process, such as treadmills, strength training equipment, etc., in order to understand the usage, performance, and safety of the equipment. At the same time, they also need to pay attention to training venues, such as track and field courts, basketball courts, etc., to understand the quality, flatness, and safety of the venues. In addition, they also need to pay attention to the training environment, such as temperature, humidity, air quality, etc., to understand the impact of the environment on athlete training.

## 5.5. Coach Training and Guidance

Coaches, as guides for athletes and managers of teams, have significant professional qualities and guidance levels for the training effectiveness of athletes. Through data-driven coach training and guidance, the professional quality and guidance level of coaches can be improved, thereby enhancing the training effectiveness of athletes.

In coach training and guidance, coaches need to learn advanced educational concepts and methods, such as personalized teaching, heuristic teaching, etc., to improve teaching

effectiveness. At the same time, coaches also need to learn modern technological means, such as data analysis, sports physiology, etc., to improve the scientificity of training. In addition, coaches also need to learn knowledge in team management, psychological counseling, and other aspects to improve team collaboration and the psychological qualities of athletes.

## 6. Conclusion

With the development of technology, data-driven sports training will become more comprehensive and diversified, which will help improve the scientificity and effectiveness of sports training in China. In the future, China should increase investment in talent cultivation and popularization, strengthen the protection of privacy and information security, improve the professional quality and skills of coaches, athletes, and researchers to cope with increasingly serious challenges and promote the sustainable development of China's sports industry.

## References

- [1] Zhao Wenjian, Yu Zijia, Lu Biao Research and design of a data collection and monitoring system for physical fitness indicators in sports training [J] Journal of Ezhou University, 2023, 30 (05): 102-104+108 DOI: 10.16732/j.cnki.jeu.2023.05.037.
- [2] Dilimurati Abudu Rousu Li A New Era of Sports Training Technology under the Integration of Diversity [J] Cultural and Sports Goods and Technology, 2023, (18): 187-189.
- [3] Yang Haopeng Research on the Application Strategies of Information Technology in Sports Training - Review of Sports Information Technology [J] Chinese Science and Technology Paper, 2023, 18 (08): 950.
- [4] Zhou Xi, Yang Yong Research on Empowering Physical Training Teaching System in Universities with Artificial Intelligence Technology [C]//Sports Training Branch of the Chinese Sports Science Society Collection of abstracts (special report) from the 2023 National Academic Symposium on Sports Training School of Physical Education, Yunnan Normal University;, 2023: 2.
- [5] Tang Beibei, Liu Haoqi, Jiang Xianxin Research on the Development Status of Artificial Intelligence Technology and Its Application in Sports Training [C]//China Bandi Association, Macau Physical Fitness Association, Guangdong Provincial Physical Fitness Association Proceedings of the 8th China Physical Training Science Conference Research Institute of Capital Sports Institute;, 2023: 4.
- [6] Tang Chuande Research on a personalized sports training effectiveness evaluation model based on GA-SVM [J] Journal of Xichang University (Natural Science Edition), 2023, 37 (01): 105-109+119.
- [7] Hedemei Research on the Precision Guidance of Artificial Intelligence Technology in the Process of Physical Education Training in Universities [J] Journal of Wuhan Shipbuilding Vocational and Technical College, 2023, 22 (01): 77-81.
- [8] Wan Zhiyin How to make "technology empowerment" add wings like a tiger in middle school physical education classrooms Research on Innovation in Ice and Snow Sports, 2023, (03): 102-105.
- [9] Cai Xiaofei A Logistic model-based method for evaluating the completion of sports training actions [J] Journal of Shangqiu Normal University, 2022, 38 (09): 80-83.
- [10] Liu Yipeng Research on Informatization of Sports Training [J] Cultural and Sports Goods and Technology, 2022, (15): 193-195.