

Research Article

An Improved Logistic Regression Method for Assessing the Performance of Track and Field Sports

Songling Zheng  and Xi Man

Institute of Physical Education, Inner Mongolia Normal University, Hohhot 010020, China

Correspondence should be addressed to Songling Zheng; zhsl2020@imnu.edu.cn

Received 1 June 2022; Revised 22 June 2022; Accepted 6 July 2022; Published 2 August 2022

Academic Editor: Le Sun

Copyright © 2022 Songling Zheng and Xi Man. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Track and field is an important part of sports. Track and field athletes are an important reserve force for the development of national sports. An accurate assessment of track and field athletes' performance can help them develop more appropriate training programs and improve their performance. In order to assess the performance of track and field athletes better, this paper proposes an improved logistic regression method. Firstly, this method uses factor analysis to reduce the data dimensions of the factors that affect the performance of track and field athletes, and uses the principal component analysis to select common factors and their corresponding values. Then, according to the common factors, a binary logistic regression model is established to evaluate the performance of track and field athletes. Experiments show that the method can effectively evaluate the performance of track and field athletes and is suitable for athletes of different track and field sports. It has high accuracy, fast evaluation efficiency, and good universality of performance evaluation. For different numbers of athletes, the proposed method has a lower error evaluation index, higher evaluation accuracy, and better evaluation quality. Compared with the other two methods, the proposed method has the shortest evaluation time and is more effective for the performance evaluation of track and field athletes.

1. Introduction

Athletes are an important reserve force for the development of national sports, and the accurate assessment of athletes' performance can develop more applicable training plans for them and improve their performance [1]. In addition to a complete training system, an objective and fair assessment evaluation system is particularly important for training athletic sports talents. An objective evaluation system of track and field sports performance aims to explore the potential of track and field athletes and is conducive to the national selection of more suitable track and field sports talents [2]. Constructing an evaluation model reflects the objective training effects of track and field athletes, finds the strengths and weaknesses of track and field athletes themselves, and then promotes track and field training reform and maximizes the effects and benefits of the sports training reform. At the same time, the feedback information from the model can also promote track and field athletes to clearly

recognize their training situation in the future training process and continuously adjust their training status to achieve the highest training efficiency. This can also serve as a guideline for track and field athletes' career planning.

Factors such as training intensity and track and field athletes' own physical quality can directly affect their performance. Accurately understanding the changing characteristics of track and field athletes' performance can ensure their better performance [3]. This makes it very important to assess track and field athletes' performance. The assessment of track and field athletes' performance is an important part of the athletes' training activities. This work plays a role in diagnosing, regulating, and strengthening the training process of track and field athletes, as well as making value judgments about the effectiveness of their training [4]. The evaluation of the training effectiveness of track and field athletes should be an evaluation of the training effect and training process. The evaluation should not only emphasize the function of screening and selection, but also strengthen

the function of motivation and development [5]. What should we do in the evaluating the training performance of track and field athletes to achieve the purpose of cultivating the training interest of track and field athletes, stimulating the subjective initiative of track and field athletes, and meeting the psychological needs of athletes is directly related to the functional orientation of track and field athletes' training and the realization of training goals. This is a problem that needs to be solved urgently at present.

Athletic athlete performance assessment is not only a test of sports training effects, but also a comprehensive judgment of athletes' sports ability. Whether the assessment is comprehensive, objective and fair, and truly reflects an athlete's actual level in sports [6] is often an concern of the athletes. Therefore, it is particularly important to construct a diversified performance assessment system. The diversification of the assessment system is reflected in the diversification of the assessment content. Athletic performance assessment should not be limited to physical fitness and motor skills, but also include training attitude, physical exercise, training participation, and competition winning together. The assessment should cover various factors such as cognition, emotion, cooperation, learning, and practice of the athletes [7]. The diversification of the assessment system is also reflected in the setting of dual subjects of assessment. They are the summative assessment made by the coach as the main body relying on the assessment results of training programs and the formative assessment made by the athlete as the main body with training activities as the main content [8]. By constructing a diversified track and field athletes' performance assessment system, it further broadens the dimensions and connotations of track and field athletes' performance assessment, which is of practical significance to improve the fairness and comprehensiveness of track and field athletes' performance assessment, enhance track and field athletes' participation and dominance in the process of performance formation, and help track and field athletes understand themselves, discover themselves, and transform themselves more objectively.

The regression model is a predictive model that studies the dependent and independent variables and integrates various possible influencing factors to assess athletes' performance and training effects through multiple regression models [9]. The research methodology in this paper takes the factors affecting the training performance of track and field athletes as the object of study, selects the factors affecting the assessment of track and field athletes' performance as the target variable, and establishes a logistic regression model. In this paper, the historical performance of track and field athletes was selected as the dataset. Among the assessment variables were competition ranking, competition time, age, gender, training duration, BMI, and blood pressure. First, factor analysis is carried out on the evaluation indicators to reduce the dimension of the data, eliminate the correlation between the data, and determine the final indicators. Then, a logistic regression model was established based on the final indicators. Finally, the assessment effects of the models were compared. Compared with the other methods, the method in this paper can achieve high-quality assessment of track

and field athletes' performance, which is very important for their training planning. Accurate assessment of track and field athletes' performance can help them understand themselves and training planning, which is good to improve their performance and make them better and better.

This paper has the following innovative points.

- (1) The factors affecting track and field athletes' performance are multiple. In order to effectively conduct track and field athletes' performance assessment, this paper simplifies the data and influencing factors by the factor analysis method. Discarding secondary factors and selecting primary factors as evaluation variables allows for a more simplified and efficient operation of the algorithm.
- (2) The common factors affecting the performance of track and field athletes were selected using the principal component analysis, and classified and assigned different weight values according to the degree of influence, which can improve the accuracy of the evaluation.

This paper mainly consists of five parts; the first is the introduction, the second is the state of the art, the third is the methodology, the fourth is the experiment and analysis, and the fifth is the conclusion.

2. State of the Art

2.1. Research Status. At present, with the deepening of the concept of "Internet Plus," information technology has been widely used in sports training activities. A large number of scholars have conducted in-depth research on sports performance assessment models and constructed many assessment models. Under the guiding principles of advancement and comprehensiveness, the literature [10] established indicators such as training hours to improve the quality of sports training and to promote further the internalization of athletes' knowledge. The literature [11] established an evaluation model from three aspects of the basic needs theory. The model used hierarchical analysis to analyse the indicator weights and found that the greatest weight was given to the autonomy needs and the least weight to the competence needs. When summarizing the methodological studies on the quantification of performance evaluation in universities, hierarchical analysis was found to be the most representative, but it is very difficult to test whether the judgment matrix is consistent when studying real-world problems and it is difficult to truly reflect the fuzzy nature of human evaluation [12]. Therefore, the literature [13] addresses the shortcomings of expert scoring in the hierarchical analysis method and integrates the principles of fuzzy mathematics to establish a mathematical model to evaluate the training quality more objectively. The literature [14] established an evaluation system from three aspects: training platform, coaches, and athletes. The method is based on AHP to determine the index weights and introduces a fuzzy comprehensive evaluation model for the differences that exist between the consistency of judgment matrix and the consistency of human brain thinking. This

provides a new perspective for athlete training quality assessment. With the continuous improvement of the fuzzy complementary judgment matrix theory, the literature [15] established athlete satisfaction indicators. The theory indicates that the influence of personal factors on the index system is the highest and the influence of gymnasium factors is the lowest, which provides a more scientific and reasonable reference basis for athlete training strategies. The literature [16] investigated the athlete performance prediction method integrating knowledge mapping and collaborative filtering to establish a training knowledge map depicting training information. The algorithm calculates the similarity of training at the knowledge level by the neighbour node-based method and the knowledge graph-based learning method, and integrates the obtained similarity into the collaborative filtering performance prediction framework to obtain athlete performance prediction results. The literature [17] investigates a two-way attention-based mechanism for athlete performance prediction model. The model obtains the attention scores of different attribute features on the first stage and second stage competition performance through two attention calculations, and combines the multi-feature fusion approach to obtain the competition performance prediction results. The historical data-driven prediction method is implemented by historical data. There are many historical data-driven prediction methods such as hidden Markov models, chaotic prediction, and support vector machines [18]. Support vector machines have the advantage of small sample learning and high learning ability in prediction, and hence they are also used to study the historical data-driven athlete performance estimation method. This method uses the KNN algorithm to pre-process the historical performance of athletes to remove the effect of distracting data and classify the data accurately. It uses support vector institutions to build regression prediction models and introduces Lagrangian functions for data transformation to avoid data operations from getting localized [19]. The support vector regression prediction model parameters are optimized using the particle swarm algorithm to reduce the interference of input quantity noise and reduce the complexity of the computation. These evaluation models are built according to different application scenarios and are important for promoting the scientific training of athletes.

2.2. Factor Analysis Method. Factor analysis is a technique to reduce the dimensionality and simplify the data. It explores the underlying structure of the observed data by examining the internal dependencies among many variables and represents the underlying data structure with a few “abstract” variables. These abstract variables are called “factors.” By discarding secondary factors and selecting primary factors as evaluation variables, the model is made more simplified and the efficiency of the algorithm operation is improved. This reflects the main information of the original set of variables. The original variables are observable explicit variables, while the factors are generally unobservable latent variables. The common factors in factor analysis are common influences

that are not directly observable but exist objectively. Each variable can be expressed as a linear function of the common factor and the sum of the special factors [20]. Its mathematical model can be expressed as

$$\begin{pmatrix} i_1 \\ i_2 \\ \vdots \\ i_u \end{pmatrix} = \begin{pmatrix} g_{11} & L & g_{1u} \\ g_{21} & \cdots & g_{2u} \\ \vdots & & \vdots \\ g_{u1} & \cdots & g_{uu} \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_w \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_w \end{pmatrix}. \quad (1)$$

That is, $I = GF + \varepsilon$, where $I = (i_1, i_2, \dots, i_u)^N$ is an observable u -dimensional random vector. Each component represents an indicator or vector. F in $F = (F_1, F_2, \dots, F_w)^N$ is an w -common factor variable. w is less than or equal to u . It is the factor that appears in the expressions of each original observed variable, which are mutually independent unobservable theoretical variables. Matrix G is called the factor loading matrix. g_{xy} is called the factor loading. It represents the correlation coefficient between the x -th original variable and the y -th public factor variable. The larger g_{xy} indicates the stronger correlation between the public factor F_y and the original variable I . ε is a special factor. It represents the part of the original variables that cannot be explained by the common factor variables, which is equivalent to the residual part in the multiple linear regression analysis.

Factor analysis utilizes the idea of dimensionality reduction, starting from the study of the dependencies within the correlation matrix of the original variables, and groups the original variables according to the magnitude of the correlation, making the correlation between variables within the same group high and the correlation between variables in different groups low [21]. Each group of variables represents a basic structure and is represented by an unobservable composite variable. This underlying structure is called the common factor. Capturing these main factors can help us analyse and interpret complex problems.

3. Methodology

3.1. Regression Models

3.1.1. Regression Model Where the Dependent Variable Is a Qualitative Variable. (1) *Qualitative variables* The dependent variable takes only two outcomes. $j=0$ means that the event did not occur. $j=1$ means that the event occurred. Consider the following expression for a simple linear regression model.

$$\begin{aligned} j_x &= \beta_0 + \beta_1 i_x + \varepsilon_x, \\ E(j_x) &= \beta_0 + \beta_1 i_x. \end{aligned} \quad (2)$$

Since j_x is a Bernoulli random variable of type 0 to 1, the following probability expression is obtained:

$$\begin{aligned} U(j_x = 1) &= \pi_x, \\ U(j_x = 0) &= 1 - \pi_x. \end{aligned} \quad (3)$$

According to the definition of discrete random variable expectation, the following function is obtained:

$$E(j_x) = 1(\pi_x) + 0(1 - \pi_x) = \pi_x. \quad (4)$$

Thus, $E(j_x) = \pi_x = \beta_0 + \beta_1 i_x$.

(2) *Error term* The error term $\varepsilon_x = j_x - (\beta_0 + \beta_1 i_x)$ can only take two values for a dependent variable, i.e., 0 or 1. Its expression is as follows:

$$\begin{aligned} j_x = 1, \varepsilon_x &= 1 - (\beta_0 + \beta_1 i_x) = 1 - \pi_x, \\ j_x = 0, \varepsilon_x &= -(\beta_0 + \beta_1 i_x) = -\pi_x. \end{aligned} \quad (5)$$

The error term is a two-point discrete distribution, and thus it cannot be assumed to be a normal error regression model.

Zero-mean heteroskedasticity means that the error term is zero-mean and its variances are not equal, and the expressions are as follows:

$$D(\varepsilon_x) = D(j_x) = \pi_x(1 - \pi_x) = (\beta_0 + \beta_1 i_x)(1 - \beta_0 - \beta_1 i_x). \quad (6)$$

If a multiple linear regression equation is used to analyse the quantitative relationship between the dependent variable and the independent variable, the relationship function is expressed as follows:

$$j = \beta_0 + \beta_1 i_1 + \beta_2 i_2 + \dots + \beta_w i_w. \quad (7)$$

(3) The left side of the equation j takes 0 or 1, and the right side of the equation can take any real number; the left and right sides do not correspond to each other in terms of the range of values. Therefore, multiple linear regression cannot be used for fitting the dependent variable as a qualitative variable.

3.1.2. Logistic Regression Model. The logistic function has the form [22].

$$f(i) = \frac{e^i}{1 + e^i} = \frac{1}{1 + e^{-i}}. \quad (8)$$

The range of values of its independent variable is $(-\infty, +\infty)$ and the range of values of the function is $(0, 1)$.

The dependent variable j itself takes only two discrete values of 0 or 1. It is not suitable as the dependent variable in the regression model, such that

$$\begin{aligned} \pi_x &= f(i_x) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 i_x))}, \\ \ln\left(\frac{\pi_x}{1 - \pi_x}\right) &= \beta_0 + \beta_1 i_x, \end{aligned} \quad (9)$$

where π_x is the probability that the random variable j takes 1, and its value varies continuously in the interval $[0, 1]$; thus, π_x can be used as the dependent variable instead of j .

Let j be a variable of type 0 to 1, and t sets of observations be $(i_{x1}, \dots, i_{xu}, j_x)$, where j_1, j_2, \dots, j_t is a random variable that takes the value 0 or 1. The expression for the expected value is as follows:

$$E(j_x) = \pi_x = f(\beta_0 + \beta_1 i_{x1} + \dots + \beta_u i_{xu}). \quad (10)$$

The expression of the function for the logistic regression model [23] is as follows:

$$\pi_x = f(i_x) = \frac{e^{\beta_0 + \beta_1 i_{x1} + \dots + \beta_u i_{xu}}}{1 + e^{\beta_0 + \beta_1 i_{x1} + \dots + \beta_u i_{xu}}}. \quad (11)$$

Thus, j_x is a random variable of type 0 to 1 with mean $\pi_x = f(\beta_0 + \beta_1 i_{x1} + \dots + \beta_u i_{xu})$; and the probability function is

$$\begin{aligned} U(j_x = 1) &= \pi_x, \\ U(j_x = 0) &= 1 - \pi_x. \end{aligned} \quad (12)$$

The random probability of j_x can be defined as:

$$U(j_x) = \pi_x^{j_x} (1 - \pi_x)^{1 - j_x}, \quad j_x = 0, 1; \quad x = 1, \dots, t. \quad (13)$$

The likelihood function of j_1, j_2, \dots, j_t is thus

$$L = \prod_{x=1}^t U(j_x) = \prod_{x=1}^t \pi_x^{j_x} (1 - \pi_x)^{1 - j_x}. \quad (14)$$

The likelihood function is taken logarithmically and the following expression can be obtained:

$$\begin{aligned} \ln L &= \sum_{x=1}^t [j_x \ln \pi_x + (1 - j_x) \ln(1 - \pi_x)] \\ &= \sum_{x=1}^t \left[j_x \ln \frac{\pi_x}{1 - \pi_x} + \ln(1 - \pi_x) \right]. \end{aligned} \quad (15)$$

Bringing equation (2.14) into the equation gives the expression

$$\ln L = \sum_{x=1}^t \left[j_x (\beta_0 + \beta_1 i_{x1} + \dots + \beta_u i_{xu}) - \ln(1 + \exp(\beta_0 + \beta_1 i_{x1} + \dots + \beta_u i_{xu})) \right]. \quad (16)$$

The maximum likelihood estimation yields the estimate $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_u$ of $\beta_0, \beta_1, \dots, \beta_u$.

3.2. Prediction Model

3.2.1. Data Processing. The data taken in this paper are from the historical performance of athletes in a sports school. They mainly contain factors such as competition ranking, competition sports time, age, gender, training hours, and physical fitness at all levels of events. The dataset contains data related to 100 athletes. The dataset consists of training predictor variables and one target variable for track and field sports. The predictor variables include athletic athletes' competition ranking, competition time, age, gender, training time, BMI, and blood pressure, and the variable descriptions are shown in Table 1. The goal of the dataset is to predict athletic athletes' performance based on certain parameter measures contained in the dataset.

The process of data cleaning requires the consideration of the following effects.

- (1) Duplicate or irrelevant data.
- (2) Mislabelled data or multiple occurrences of the same label.

TABLE 1: Variable declaration.

I_1	Competition ranking
I_2	Race time
I_3	Age
I_4	Gender
I_5	The training time
I_6	BMI
I_7	Blood pressure

(3) Missing or empty data points.

(4) Outlier values.

The data are a standard database, and hence there is no duplicate or irrelevant data and no vacant data points have been checked. Since blood pressure, age, and body mass index cannot be 0 in general, and 0 is an abnormal data point, the rows with 0 values in each feature of blood pressure, age, and body mass index were filtered out. There were 724 valid data left after processing.

3.2.2. Factor Analysis

(1) *Applicability Test of Factor Analysis.* The results of the KMO and Bartlett's sphericity tests on the data of track and field athletes using SPSS software are shown in Table 2. It is generally considered that if the KMO measure is greater than 0.5, then factor analysis can be performed. The significance of $p=0$ indicates that there is a certain correlation between the original variables, and the conditions for factor analysis are available.

(2) *Extraction of Common Factors.* Factor analysis was performed on the data, and the extraction of principal components was performed by principal component analysis. Classifying them according to the degree of influence and assigning different weight values improves the assessment accuracy. Under the principle of eigenvalue of 1, three principal factors were retained, i.e., the seven variables were grouped into three categories. This reduces the amount of operations, but categorization causes information loss, and the amount of information retained is 64.49%, and the amount of information lost is large; thus, a common factor is added to make the amount of information lost reside within an acceptable range. The following variance interpretation Table 3 shows that each principal component contains the total variance of each original variable, and the improved retained information is 77.08%.

(3) *Public Factor Naming.* The original factor loading matrix was rotated by extracting the four public factors and performing maximum variance orthogonal rotation to obtain the variance maximum orthogonal rotation matrix, as shown in Table 4.

Based on the rotated component matrix, the four common factors can be named. The first factor Z1 has large loadings on the gender and age indicators. The second factor Z2 has larger loadings on the race ranking and race time indicators. The third factor Z3 has a large loading on BMI and

training duration. The fourth factor Z4 has a larger loading on blood pressure. It can be found that the evaluation indexes corresponding to Z1 are indirect influence data. The evaluation indexes corresponding to Z2 are race performance-related data. The evaluation indexes corresponding to Z3 are other physical data. Z4 represents blood pressure. They are named as indirect factors, competition performance, physical quality, and blood pressure, respectively.

3.2.3. Binary Logistic Regression

(1) *Hosmer-Lemeshaw test.* The original hypothesis H0: the model fits well with the observations. The results are shown in Table 5 $p=0.279 > 0.05$; the original hypothesis is accepted and the regression model can fit the data well.

$$\text{Logit}U = -0.82 + 0.65k_1 + 0.86k_2 + 0.67k_3 + 0.38k_4, \quad (17)$$

(2) As shown in Table 7, the significant p -values are all 0, indicating that BMI, age, gender, and training duration have highly significant effects on the performance of track and field athletes. The effects were ranked from the highest to the lowest: BMI > age > gender > training duration.

(3) As shown in Table 7, the significant p -values are all 0, indicating that BMI, age, gender, and training duration have highly significant effects on the performance of track and field athletes. The effects were ranked from the highest to the lowest: BMI > age > gender > training duration.

(4) The accuracy is shown in Table 6, with an accuracy of 74.9%, which indicates that the model predicts more accurately.

(5) From the regression analysis of multiple factors, a binary logistic regression equation was established. where $U = \beta_0 + \beta_1 i_1 + \beta_2 i_2 + \dots + \beta_w i_w$.

4. Result Analysis and Discussion

Using track and field athletes of a sports school as the experimental subjects, 10 groups of 200m sprinters were randomly selected as the research subjects. The evaluation results are shown in Figure 1. According to Figure 1, the method of this paper can effectively evaluate the performance of 200m sprinters, and the estimated value is very close to the actual value. The experiment proves that the method in this paper can accurately estimate the performance of track and field athletes and has a high accuracy of track and field athletes' performance assessment results.

The athletes of 10 types of track and field sports were randomly selected in this sports school to verify the generality of this paper's method. Using the method of this paper, the athletes' performance of these 10 types of track and field sports was evaluated and compared with the actual values, and the evaluation accuracy of the 10 types of track and field sports is shown in Figure 2. According to Figure 2, it can be seen that for different types of track and field sports, the method in this paper can accurately assess the performance of track and field athletes, and the estimation accuracy is basically maintained at more than 96%. The experiment proves that the method in this paper has good generality and a high estimation accuracy for different types of track and field sports.

TABLE 2: KMO and Bartlett's test.

KMO sampling suitability quantity	0.592
Bartlett sphericity test	The approximate chi-square Degrees of freedom Significant
	952.82 26 0

TABLE 3: Total variance explained.

Composition	Total	Percentage of variance of initial eigenvalue	Cumulate %	Total	Load square and percent variance	Cumulate %	Total	Rotational load square and percent variance	Cumulate %
1	1.98	28.28	28.28	1.98	28.28	28.28	1.70	24.41	24.41
2	1.53	21.90	50.19	1.53	21.90	50.19	1.35	19.35	43.77
3	1.00	14.29	64.49	1.00	14.29	64.49	1.30	18.70	62.48
4	0.88	12.58	77.08	0.88	12.58	77.08	1.02	14.60	77.08
5	0.65	9.33	86.41						
6	0.54	7.77	94.19						
7	0.40	5.80	100.00						

TABLE 4: Rotated factor matrix and score matrix.

Variable	Indicators	Rotation factor matrix				Factor scoring matrix			
		1	2	3	4	1	2	3	4
I_1	Competition ranking	0.846	-0.018	0.008	-0.001	0.519	-0.028	-0.116	0.044
I_2	Race time	0.868	0.077	0.113	0.011	0.516	0.029	-0.048	0.035
I_3	Age	-0.155	0.855	-0.035	0.123	-0.096	0.674	-0.149	-0.031
I_4	Gender	0.278	0.748	0.231	0.001	0.124	0.561	0.032	-0.129
I_5	The training time	-0.146	0.207	0.798	0.173	-0.205	0.010	0.650	0.090
I_6	BMI	0.334	-0.044	0.773	-0.108	0.083	-0.1465	0.612	-0.127

TABLE 5: Hosmer-Lemeshaw test.

Chi-square	Degrees of freedom	Significant
9.806	8	0.278

TABLE 6: Prediction accuracy.

	Assessment of conformity		Accuracy rate	
	0	1		
Actual qualification	0	412	64	86.6
	1	118	132	52.8
Overall percentage			74.9	

Comparing the method of this paper with the methods in the literature [16] and literature [17] clearly indicates that the assessment of athletes' performance of the above-mentioned 10 types of track and field sports was implemented at the same time. The evaluation accuracy and evaluation efficiency of the three methods were tested by comparing, among which literature [16] is a student performance prediction method integrating knowledge mapping and collaborative filtering, and literature [17] is a student performance prediction model based on a two-way attention mechanism. The performance of 100 athletes in each type of track and field sports was selected for testing and the average value was taken to enhance the credibility of the experiment. The accuracy and assessment efficiency of the three methods for

assessing athletes' performance in the 10 types of track and field sports are shown in Figures 3 and 4. According to Figure 3, it can be seen that for athletes of different types of track and field sports, the assessment accuracy of athletes' performance of this paper's method is significantly higher than the remaining two methods, and the average assessment accuracy of this paper's method is 97.7%, the average assessment accuracy of literature [16] is 81.8%, and the average assessment accuracy of literature [17] is 86.5%. The experiment proves that when assessing the performance of athletes in different types of track and field sports, the method in this paper has the highest assessment accuracy, which significantly reduces the estimation error of athletes' performance and increases the credibility of the assessment results at the same time.

According to Figure 4, for athletes of different types of track and field sports, the evaluation time of this paper's method is significantly lower than the remaining two methods, and the evaluation time of this paper's method is always maintained within 20s with less variation, while the estimation time of the remaining two methods is more variable and less stable. This is due to the fact that the algorithm in this paper introduces factor analysis to optimize the parameters affecting the evaluation, which reduces the computational parameters and decreases the computational effort. The experiment proves that the evaluation time of this paper's method is the least and the athlete's performance evaluation is more efficient.

TABLE 7: Logistic regression analysis.

	B	Standard error	Wald	Degree of freedom	Significance	Exp (B)	95% confidence interval of EXP (B)	
							Lower limit	Upper limit
Gender	0.65	0.09	51.66	1	0	1.92	1.61	2.3
BMI	0.86	0.1	75.36	1	0	2.38	1.96	2.9
Age	0.67	0.09	49.18	1	0	1.97	1.63	2.38
Training time	0.38	0.09	17.32	1	0	1.47	1.22	1.76
Constants	-0.82	0.09	76.33	1	0	0.43		

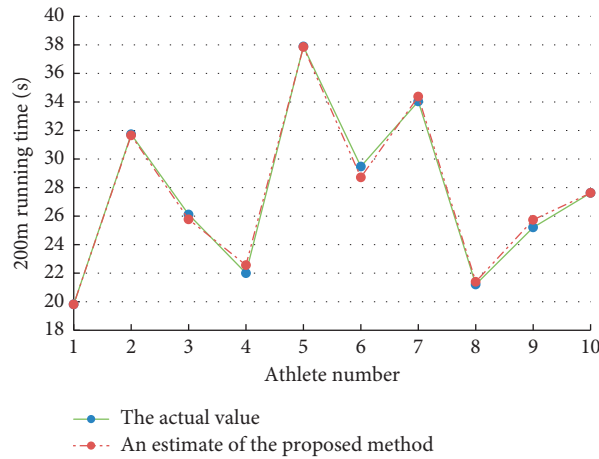


FIGURE 1: Performance evaluation results of the 10 groups of 200m sprinters.

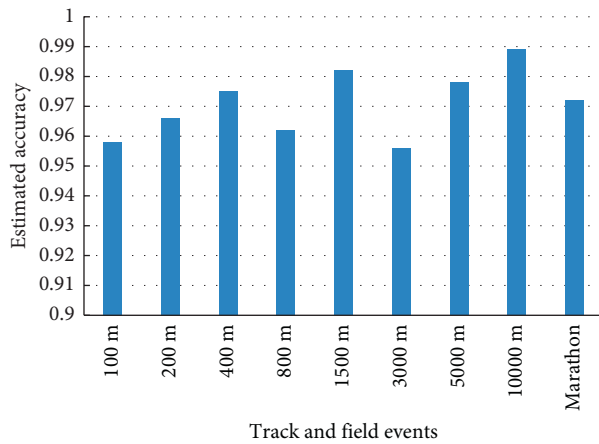


FIGURE 2: Estimation accuracy of the performance of the 10 track and field sports.

Taking 200m sprinters’ performance as an example, the accuracy of the performance assessment of the three methods was tested with different numbers of athletes. The accuracy of the three methods was evaluated by the Mean Absolute Percentage Error (MAPE), an error evaluation index. The results of the error evaluation index tests for the three methods with different numbers of athletes are shown in Figure 5. According to Figure 5, the MAPE values of all three methods increased with the increasing number of athletes. Generally, if the MAPE value is lower than 10, it indicates that the evaluation accuracy of the evaluation methods is higher. With different numbers of athletes, the

MAPE values of this paper are significantly lower than those of the remaining two methods. The MAPE value of this method always stayed within 6, and the MAPE value of the other two methods was lower than 10 only when the number of athletes was less than 200. When the number of athletes was more than 200, the MAPE values of the other two methods were greater than 10.

The experiment proves that the MAPE value of this paper method is the lowest when the number of athletes is different. This indicates that the evaluation value of this paper’s method is closest to the actual value, with a higher evaluation accuracy and high evaluation quality.

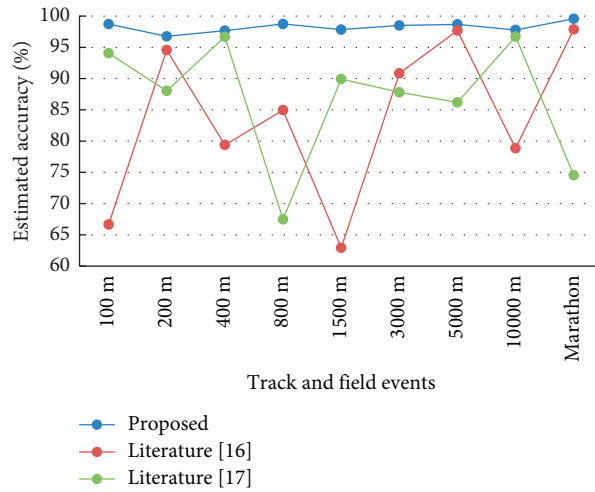


FIGURE 3: Evaluation accuracy of the three methods.

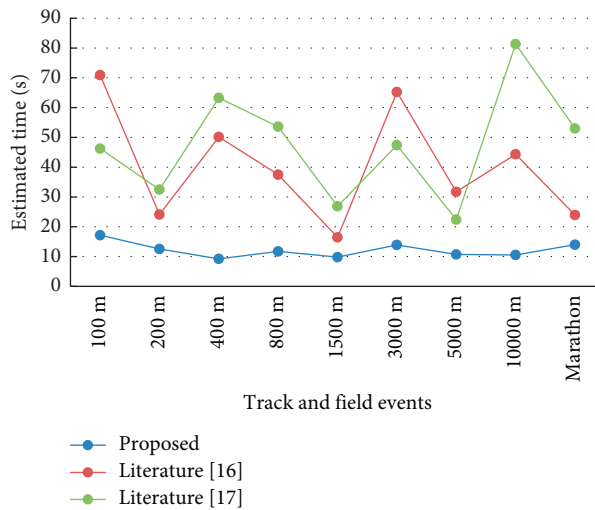


FIGURE 4: Evaluation efficiency of the three methods.

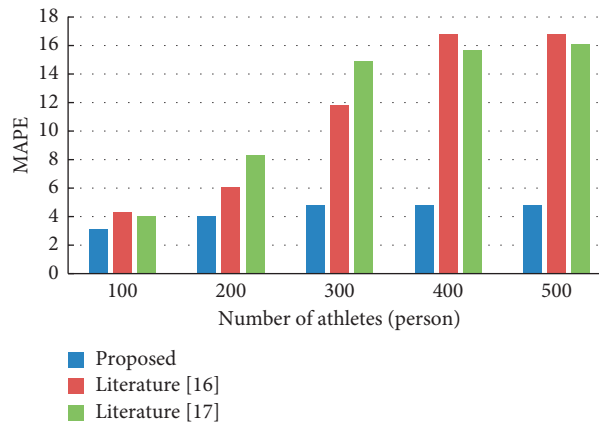


FIGURE 5: Comparison of the error evaluation indexes of the three methods.

5. Conclusion

Based on the model obtained in this paper, the performance of track and field athletes can be effectively evaluated. The main objective of this paper is to evaluate the performance of track and field athletes using a logistic regression model. The method adopts the idea of factor analysis, reduces and simplifies the data, and improves the evaluation effect. The experiments show that the method can accurately estimate the performance of track and field athletes and has a high accuracy of track and field athletes' performance assessment. At the same time, the method has good generality for track and field athletes' performance evaluation, less evaluation time, and higher evaluation efficiency. Overall, the method in this paper can achieve high-quality athlete performance assessment, which is very important for athletes' training planning. The accurate estimation of athletes' performance can help improve their performance and make them better by understanding the training planning they need. Therefore, we study the historical data-driven athlete performance estimation method to improve the accuracy and estimation efficiency of athlete performance estimation, provide more valuable information for athlete training planning, and develop better athletes for the country. However, the good experience of using this method requires a large amount of track and field athletes' historical competition performance data, and the effectiveness of the evaluation for individual track and field athletes' performance is yet to be verified. The experimental data are all track and field athletes' performance, and the generality of the prediction for other non-track and field sports is also yet to be verified. The next step will be to further explore the effectiveness of the model in evaluating the performance of athletes in a wider range of sports, to verify the generality of the model in evaluating athletes' performance, and to expand the scope of the application.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The author declares no competing interests.

Acknowledgments

This study was sponsored by Inner Mongolia Natural Science Foundation of China, Grant 2021MS03018, & High-Level talent Scientific Research Start-Up Foundation of Inner Mongolia Normal University, Grant 2019YJRC060.

References

- [1] K. Lindberg, P. Solberg, B. R. Rønnestad et al., "Should we individualize training based on force-velocity profiling to improve physical performance in athletes?" *Scandinavian Journal of Medicine & Science in Sports*, vol. 31, no. 12, pp. 2198–2210, 2021.
- [2] C. E. Thomas, P. B. Gastin, G. Abbott, and L. C. Main, "Impact of the talent development environment on the wellbeing and burnout of Caribbean youth track and field athletes," *European Journal of Sport Science*, vol. 21, no. 4, pp. 590–603, 2021.
- [3] R. Rico, C. B. Gibson, M. Sánchez-Manzanares, and M. A. Clark, "Building team effectiveness through adaptation: team knowledge and implicit and explicit coordination," *Organizational Psychology Review*, vol. 9, no. 2-3, pp. 71–98, 2019.
- [4] N. Pollock, S. Kelly, J. Lee et al., "A 4-year study of hamstring injury outcomes in elite track and field using the British Athletics rehabilitation approach," *British Journal of Sports Medicine*, vol. 56, no. 5, pp. 257–263, 2022.
- [5] M. Paais, J. R. Pattiruhu, "Effect of motivation, leadership, and organizational culture on satisfaction and employee performance," *The Journal of Asian Finance, Economics and Business*, vol. 7, no. 8, pp. 577–588, 2020.
- [6] A. C. Jeffries, L. Wallace, A. J. Coutts, S. J. McLaren, A. McCall, and F. M. Impellizzeri, "Athlete-reported outcome measures for monitoring training responses: a systematic review of risk of bias and measurement property quality according to the COSMIN guidelines," *International Journal of Sports Physiology and Performance*, vol. 15, no. 9, pp. 1203–1215, 2020.
- [7] S. Costa, G. Santi, S. di Fronso et al., "Athletes and adversities: athletic identity and emotional regulation in time of COVID-19," *Sport Sciences for Health*, vol. 16, no. 4, pp. 609–618, 2020.
- [8] A. Tomiak, H. Braund, R. Egan et al., "Exploring how the new entrustable professional activity assessment tools affect the quality of feedback given to medical oncology residents," *Journal of Cancer Education*, vol. 35, no. 1, pp. 165–177, 2020.
- [9] S. Del Rosso, D. Pinho Souza, F. Muñoz, D. G. Behm, C. Foster, and D. Boulosa, "10 km performance prediction by metabolic and mechanical variables: influence of performance level and post-submaximal running jump potentiation," *Journal of Sports Sciences*, vol. 39, no. 10, pp. 1114–1126, 2021.
- [10] A. J. Sturges, Z. Huysmans, W. Way, and A. Goodson, "Examining the role of high school athletic directors in promoting leadership development in high school student-athletes," *Journal for the Study of Sports and Athletes in Education*, vol. 14, no. 1, pp. 58–81, 2020.
- [11] M. Vansteenkiste, R. M. Ryan, and B. Soenens, "Basic psychological need theory: advancements, critical themes, and future directions," *Motivation and Emotion*, vol. 44, no. 1, pp. 1–31, 2020.
- [12] H. Zhang, X. He, and H. Mitri, "Fuzzy comprehensive evaluation of virtual reality mine safety training system," *Safety Science*, vol. 120, pp. 341–351, 2019.
- [13] W. Li, G. Xu, Q. Xing, and M. Lyu, "Application of improved AHP-BP neural network in CSR performance evaluation model," *Wireless Personal Communications*, vol. 111, no. 4, pp. 2215–2230, 2020.
- [14] A. Rizvandi, M. Taghipour Gharbi, and M. Esmaeili, "The evaluation of performance indicators of coaches in football development[J]," *Journal of Humanities Insights*, vol. 3, no. 04, pp. 246–252, 2019.
- [15] X. Huang, X. Ma, and Q. Zhang, "Effect of building interface form on thermal comfort in gymnasiums in hot and humid climates," *Frontiers of Architectural Research*, vol. 8, no. 1, pp. 32–43, 2019.
- [16] I. Fernández-Tobías, I. Cantador, P. Tomeo, V. W. Anelli, and T. Di Noia, "Addressing the user cold start with cross-domain collaborative filtering: exploiting item metadata in matrix

- factorization,” *User Modeling and User-Adapted Interaction*, vol. 29, no. 2, pp. 443–486, 2019.
- [17] M. Qiu, B. Thuraisingham, M. Daneshmand, H. Ning, and P. Barnaghi, “Special issue on robustness and efficiency in the convergence of artificial intelligence and IoT,” *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9460–9462, 2021.
- [18] J. S. Raj and J. V. Ananthi, “Recurrent neural networks and nonlinear prediction in support vector machines,” *Journal of Soft Computing Paradigm (JSCP)*, vol. 1, no. 1, pp. 33–40, 2019.
- [19] Z. Tian, Y. Ren, and G. Wang, “An application of backtracking search optimization-based least squares support vector machine for prediction of short-term wind speed,” *Wind Engineering*, vol. 44, no. 3, pp. 266–281, 2020.
- [20] X. Wang, Y. Liu, F. Wang, J. Wang, L. Liu, and J. Wang, “Feature extraction and dynamic identification of drivers’ emotions,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 62, pp. 175–191, 2019.
- [21] M. Espadoto, R. M. Martins, A. Kerren, N. S. T. Hirata, and A. C. Telea, “Toward a quantitative survey of dimension reduction techniques,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 3, pp. 2153–2173, 2021.
- [22] R. Gupta, R. Pachauri, and A. K. Singh, “Image encryption method using dependable multiple chaotic logistic functions,” *International Journal of Information Security and Privacy*, vol. 13, no. 4, pp. 53–67, 2019.
- [23] T. Denoeux, “Logistic regression, neural networks and Dempster-Shafer theory: a new perspective,” *Knowledge-Based Systems*, vol. 176, pp. 54–67, 2019.