Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3916383, 11 pages
https://doi.org/10.1155/2022/3916383

Research Article

Sports Performance Prediction Based on Chaos Theory and Machine Learning

Wei Sun

Sports and Military Education Department, Zhejiang University of Water Resources and Electric Power, Hangzhou, 310018 Zhejiang, China

Correspondence should be addressed to Wei Sun; sunwei917@zjweu.edu.cn

Received 26 March 2022; Revised 31 May 2022; Accepted 9 June 2022; Published 26 June 2022

Academic Editor: Chia-Huei Wu

In order to combine chaos theory and machine learning technology to predict sports performance, a research on sports performance prediction based on chaos theory and machine learning is proposed. This paper takes the sports performance as the goal to predict the data; introduce the chaos theory algorithm, and combine the neural network system and particle swarm optimization algorithm to actively train sports results and ensure the quality of performance prediction. The comparison between shot put data prediction and real data shows that the prediction results given by the model have little deviation and can provide technical services for performance prediction for special sports training. When predicting the sports performance of college students, the prediction accuracy of sports performance of each subject is no less than 90%, which proves that this system can be used in college sports management. The delay time of the model data is short, which is mainly related to the correlation coefficient. When the coefficient is determined to be 0.02, the prediction delay time is 5 s, which can effectively complete the prediction and analysis of sports performance. The combined model has various technical advantages such as chaos theory, neural system, and particle optimization. It has strong sports performance prediction ability and can provide technical support for performance prediction for athletes' training, college sports management, and other related sports industries.

1. Introduction

In the environment of economic development, the national physical quality has been enhanced, and the students' sports performance is a key index to judge the development level of a country's sports career [1]. Therefore, the prediction and analysis of students' sports performance have been highly valued by all countries. Students' physical performance can effectively describe students' physical ability, and the prediction method of students' physical performance is helpful for physical education teachers and physical education management departments to master the dynamic characteristics of students' performance [2]. According to the specific characteristics of students, formulate a more scientific and reasonable physical education curriculum plan, significantly improve students' physical exercise ability, and ensure that students' physical education subjects achieve better results. In order to predict the performance demand of physical education colleges at all levels, researchers have deeply explored the change characteristics of physical education performance, built a multilevel physical education performance analysis system, and effectively built a prediction model [3, 4]. It is found that there are two key nodes in the prediction process of sports data of sports students at all levels. Node 1 is the linear model. This process adds a variety of algorithms, such as regression analysis and grey detection. This prediction form is based on the sports information of a single sports student to make a special prediction of performance data. If students' sports information changes little, this analysis process can be used to feed back the changes of sports information of sports students. However, when most sports students exercise, their sports performance will be disturbed by many factors, such as psychological state, age, and sports environment. This kind of sports interference problem will increase the variability of students' sports information and make it show nonlinear characteristics. Therefore, the single linear prediction method used in the initial stage cannot fully show the sports situation of sports students, will increase the deviation of prediction results, cannot guarantee the availability of the prediction

system, and reduces the practicability of students' sports operation data. In order to ensure the optimization effect of the linear model and actively expand the availability of model prediction, introduce the nonlinear idea and carry out a new modeling and analysis research. The new analysis techniques introduced are as follows: support vector, neural algorithm, and so on. The new modeling system mainly relies on the idea of linear analysis; multilevel analysis of the variation characteristics of students' sports performance has strong prediction ability and stronger applicability of the model [5].

The method of constructing the prediction model with a new vision has become the mainstream way of sports achievement prediction. The neural network system contains three subjects [6]. With the help of three units, we can deeply analyze the internal correlation of students' sports data, which is more accurate than the previous prediction results. The use of the neural network needs to determine the initial parameters. The effective design of initial parameters is the key factor to ensure the accuracy of prediction. At present, in most cases, the empirical method is used to assume the initial parameters of the algorithm, which increases the randomness of the results of the model algorithm. If the initial parameters are given irregularly, the prediction quality of the network model will be weakened and the accuracy of prediction information cannot be guaranteed [7, 8]. Taking the initial parameters as the optimization starting point, the chaos theory is introduced to combine it with the machine learning system to build a new prediction model of sports results. In the early stage, the students' sports results are analyzed, and the chaotic analysis method is introduced to obtain the change law of sports results [9]. Then, use a neural algorithm to deal with sports results efficiently. The particle swarm optimization method is used to train the processed motion data online to accurately obtain the initial parameters required by the neural system. Through practice, it is found that this modeling method can effectively measure the future trend of motion data and has high prediction accuracy of motion information and high research value [10].

2. Literature Review

Chaos theory is a theoretical problem that puzzles people for a long time. In short, the universe itself has a certain degree of chaos. In a specific environment, there will be conflicts between the subjects of all kinds of unrelated things, which will have an unpredictable impact on the things that do not conflict in the universe. This theory shows that a single small amount of motion will form a large motion influence after systematic amplification, and its influence degree will exceed the motion itself. The formation of butterfly effect is a representative phenomenon based on chaos theory. In the documentary, Professor Jim takes the birth of the universe as the starting point to study the breeding process of intelligent life. Chaos theory can be used to interpret various disordered phenomena in the universe and explain the beautiful and exquisite natural science laws in the natural environment [11]. The research process of chaos theory can integrate various complex factors of the universe and efficiently sort out repeated algorithms and simple rules. The problem worthy of attention is that all kinds of calculation processes show the huge amount of data, and there is no predictive ability between the data. Although the theoretical application gives a judgment basis, it is still unable to truly obtain the future situation. The main task of machine learning is to complete classification. Specifically, it includes the target variables that can be predicted in the invisible data in the early stage. Classification processing is to effectively predict target variables with the help of the partition model of the training data set. With the help of this model, the results of predicted data variables are accurately obtained [12–14]. This data processing method can be called supervised learning, because the data processing link is used as the modeling variable. Figure 1 shows the flow chart of supervised learning and unsupervised learning of machine learning.

In Figure 1, during the sports performance prediction, the system will collect a large number of performance characteristic parameters, such as the team's previous performance, game results, and player data. With the help of a variety of sports information, various stakeholders can understand the possibility of winning or losing. Generally, during sports competition, if the predicted competition results are obtained, the winning or losing situation of each supporting subject will be judged. In addition, sports managers strive to create appropriate schemes to evaluate the sports performance of potential opponents in the competition. Therefore, the process of predicting sports performance has always been the research direction of all kinds of stakeholders, including media and sports training organizations. Various sports data related to sports can be obtained through open electronic channels. This increases the possibility for people to develop intelligent models and prediction systems to predict sports performance [15]. In order to carry out in-depth research on sports performance prediction in the machine learning mechanism, an artificial neural network should be a method with high frequency [16]. Therefore, taking the research of the artificial neural network as the starting point, the artificial neural network usually contains interconnected components (neurons), which convert a set of inputs into desired outputs. The structure of the artificial neural network is shown in Figure 2. In Figure 2, there are four input nodes in the input layer, five hidden nodes in the hidden layer, and one output node in the output layer. The main implementation of the neural network focuses on the nonlinearity of hidden neurons when adjusting weights. The output of the neural network mainly depends on the input characteristics and other components related to the network, such as these weights. The neural network model is constructed after processing the training data set, which contains the features used to construct the neural network classification model. In other words, the weights associated with interrelated components are constantly changing to improve the prediction accuracy. These changes are performed by the artificial neural network algorithm to achieve the accuracy of the expected model previously given by the user [17]. In most cases, it may lead to the problem of overfitting, such as the waste of system operation caused by computing resources such as training time and memory.

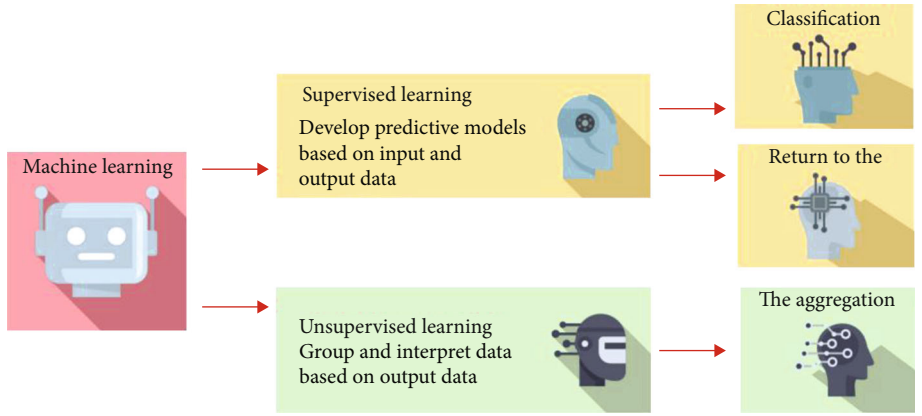


Figure 1: Flow chart of supervised learning and unsupervised learning.

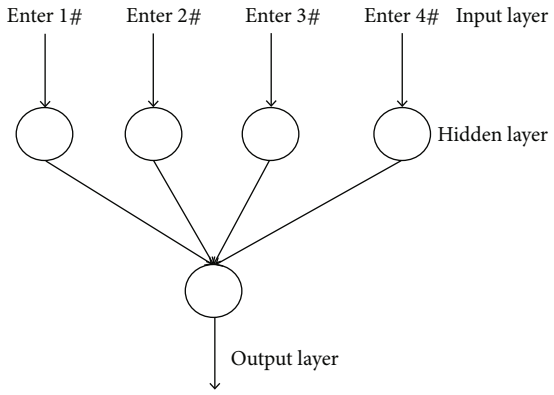


Figure 2: Flow chart of the artificial neural network.

3. Research Methods

3.1. Solutions to Research Problems

3.1.1. Machine Learning Intelligent Framework. Try to measure sports performance from the perspective of structured experiment, so as to ensure the accuracy of sports performance prediction and obtain a more comprehensive set of sports data. Figure 3 shows the modeling process of machine learning. Figure 4 shows the prediction chart of sports performance.

Figure 3 is a modeling specification of machine learning, which has multiple links such as business understanding and data preparation, to ensure the accuracy of sports performance prediction and analysis. The functions and usage of each process in Figure 4 are as follows.

3.1.2. Business Understanding. The business understanding unit contains multiple data features, such as modeling concept and motion category. The rational use of data understanding can carry out performance prediction for specific sports subjects, excavate the interference conditions of sports performance, and comprehensively sort out the interference factors by means of sports performance and expert analysis, so as to gradually deal with the interference problems and ensure the prediction quality, for example, sports performance result prediction, expert prediction, and online competition [18–20]. If the prediction model is used for game prediction, it is necessary to correlate the winning and losing amount factors. For example, when predicting the expected results of sports events, the introduced odds parameters are small, and the system return parameters do not match the real input parameters. For example, a sports organization obtains an investment of 1.01 yuan, which means that after obtaining 100 yuan, if the sports performance reaches the expected goal, only 1 yuan will be returned.

3.1.3. Data Understanding. All kinds of data used in sports prediction can be obtained from online public channels in most cases [21]. Part of the research work completed in the early stage can automatically collect motion data. With the help of script writing program, extract the sports data of public resources online, and import the extracted data into the data unit of specific format. Some researchers have created user operation terminals. Users can use matching to obtain loading data and effectively predict sports results. In this link, we need to deeply investigate the factors of data granularity and level. The preliminary research work uses all kinds of training information of competition and team. The sports ability information of sports members is generally stored in a separate data unit to effectively convert the data in the game and obtain the data set of sports performance of members in each game [22].

3.1.4. Data Preparation. Adjust the characteristic value of sports performance. The motion result data contains a variety of characteristic parameters, which contain multiple difference subsets. The eigenvalue matching and division can be carried out in combination with the development state of actual sports. If gambling factors need to be added, a mixed analysis model can be established. When selecting some eigenvalue algorithms, it is necessary to select the type with relatively complete feature set or select the combination of initial feature data. Ideally, forecasters should classify and select data in combination with the performance prediction process. Performance prediction selects various characteristics to ensure the accuracy of prediction results. Data processing will distinguish between data matching and various external characteristics of data. For example, the performance

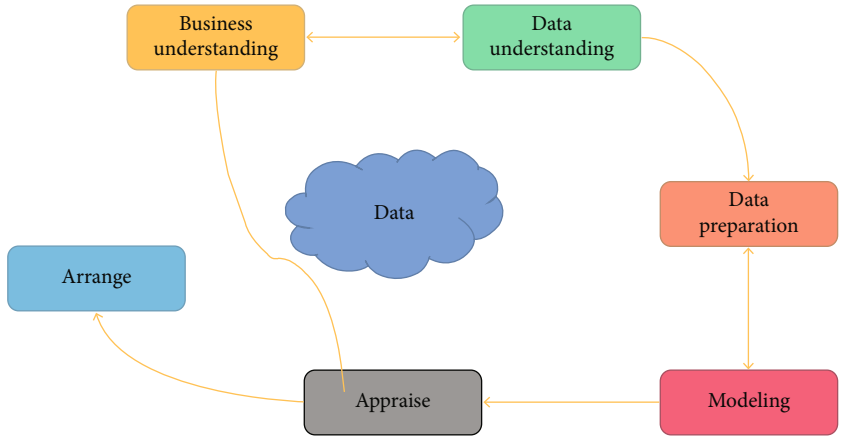


Figure 3: Machine learning modeling process diagram.

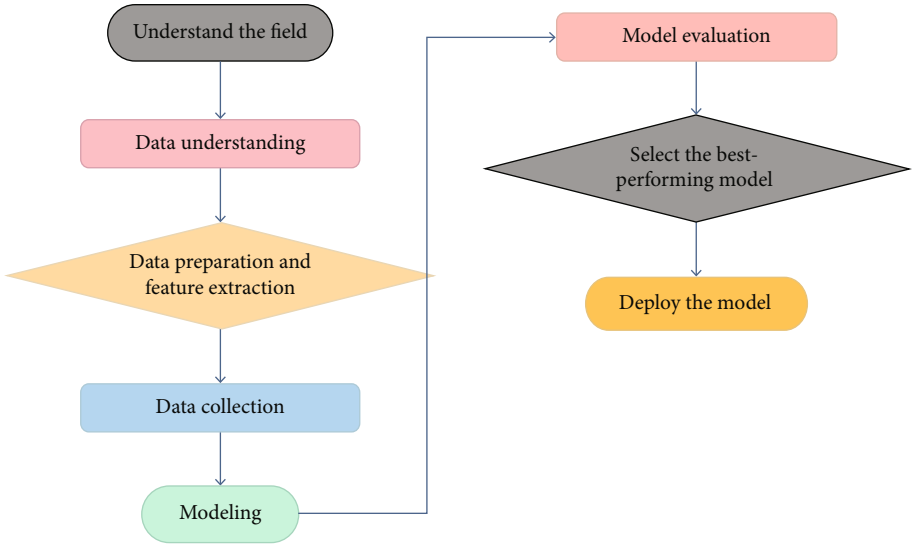


Figure 4: Sports performance prediction model.

prediction model in football sports needs to add factors such as unit "meter" and passing speed per hour. The external characteristic data is not related to the competition results, and such factors have a certain correlation with the competition itself, for example, the optional reserve player in the game [23, 24]. This kind of data processing has certain differences, which can ensure the efficiency of data preprocessing. In the early stage of the competition, the characteristics of external data are clear, for example, the distance between the sports personnel of the two teams and the state of the sports members of the two teams in the game.

3.1.5. Modeling. The preliminary modeling process is to select the available models from the experimental data, review the modeling process in the past, and select the models with high prediction success. Secondly, each model contains characteristic parameters, the corresponding algorithm is selected, and the data test is carried out in the direction of a subset. The feature selection and category division of such differences can be used for performance prediction as the optimal scheme of feature selection.

3.1.6. Performance Prediction. In order to ensure the effectiveness of the prediction model, the game results can be divided effectively, such as home victory and draw. Then, the standard is used to divide the data types, and the data matrix is used to check the correctness of the model, so as to effectively identify the number of competitions. The assignment process of data set does not need to be highly balanced. Therefore, the accuracy of classification will become the key factor to ensure the rationality of evaluation. When the data is highly unbalanced, the results of performance prediction and evaluation may be more applicable [25, 26]. In the process of sports prediction, it is key to save training data efficiently. Compared with the past sports prediction of future results, the cross validation method can be used, as shown in Figure 5. It can be seen from Figure 5 that the order of various data will be disrupted during data

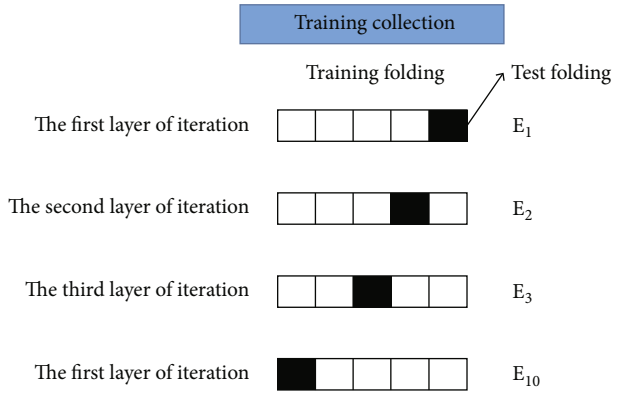


Figure 5: Diagram of the cross validation method.

prediction. It is found that cross validation has high applicability for the prediction process of sports performance.

3.1.7. Model Allocation. Generally, model matching is carried out independently, various sports data are obtained from the network structure, the sports data are efficiently distributed to the data unit, and then the forecasters independently add information to the data unit, effectively adjust the training information and detection data, use a new data training method, efficiently obtain the data training model, obtain a new performance matching result, and feed back the matching result to the user as prediction information. The learning unit introduced by this architecture can efficiently obtain sports information in each period of sports. The training data set is constantly updated, and the training information is significantly increased. This classification unit has dynamic variability and can comprehensively reflect the learning process of sports members.

3.2. Experimental Method for Verifying the Scheme

3.2.1. Chaotic Algorithm. The performance of sports members is related to many factors, such as the physical quality of members, the quality of coach curriculum design, sports training methods, training tools, relevant policies, and national economy, which increase the dynamic variability of sports personnel's performance. In a specific time range, the sports performance of sports members is mainly related to the sports information in the current period. Therefore, chaos theory can be introduced to deeply excavate the changing characteristics of performance information of sports members and find out the hidden characteristics of various sports achievements. Assuming that the score of sports personnel available for a period of time is $\{x(t_j), j = 1, 2, \cdots n\}$, n in the score expression represents the number of nodes collecting data. With the help of chaos theory, the motion information can be processed efficiently and the changing data results can be obtained, as shown in

$$X(t)\{x(t), x(t+T), \cdots x(t+(m-1)T)\}, \quad t = 1, 2, \cdots, M. \tag{1}$$

In formula (1), T represents the delay time of specific sports member information and M represents the embedding dimension of specific sports members. Among them, the parameter represented by m can be used to feed back the internal relationship between each data and the current sports information. Using C-C, "Cao," and other algorithms, the T and m values corresponding to the sports information of sports members are given one by one, so as to feed back the change law of sports members' performance information hiding [27].

3.2.2. C-C Algorithm Flow. Suppose that the two information sample nodes of sports performance of a single sports member are $X(i) = \{x(i), x(i+T), \cdots x(i+(m-1)T)\}$ and $X(j) = \{x(j), x(j+T), \cdots x(j+(m-1)T)\}$; use formula (2) to obtain the distance between the two score information.

$$r_{ij} = \|X(i) - X(j)\|. \tag{2}$$

The critical radius parameter r is added to formula (2), and the correlation integral is used to describe the data proportion of each point in this critical area. The specific description method is shown in

$$C(m, N, r, T) = \frac{2}{M(M-1)} \sum_{1 \le i \le j \le M} H(r - \|X(i) - X(j)\|). \tag{3}$$

In formula (3), N represents the data specification and the expression of $H(*)$ is a Heaviside function. The calculation method of this function is shown in

$$H(x)\begin{bmatrix} 0 & x < 0 \\ 1 & x > 0 \end{bmatrix}. \tag{4}$$

With the help of the above algorithm, the sports performance of sports members can be effectively processed and divided into t branch sequences. The obtained expression is shown in

$$S(m, r, T) = \frac{1}{t} \sum_{i=1}^{t} \{C_t(m\,r\,t) - [C_t(m\,r\,t)]^m\}. \tag{5}$$

The calculation method of the difference between the maximum value and the minimum value of the branch sequence is shown in

$$\triangle S(m, t) = \max\,[S(m, r, T)] - \min\,[S(m, r, T)]. \tag{6}$$

Assuming that the sports performance m value of sports members is $1 - k$, the expression processing is shown in

$$\triangle \bar{S}(t) = \frac{1}{4} \sum_{m=1}^{k} \triangle S(m, t). \tag{7}$$

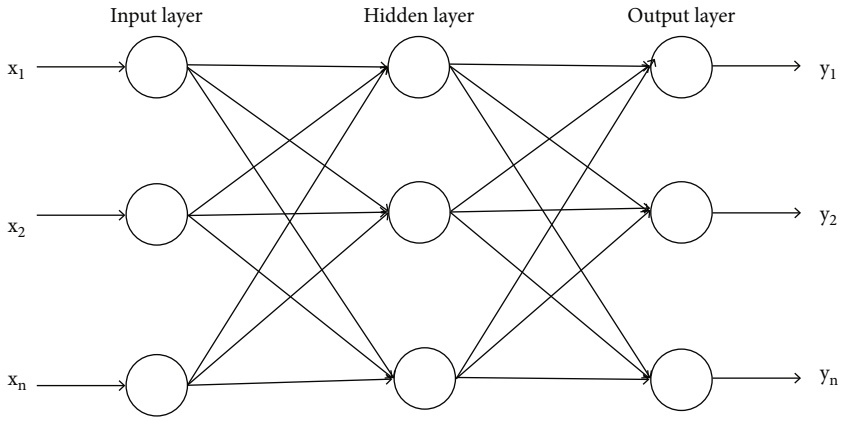


Figure 6: Assignment diagram of the neural network.

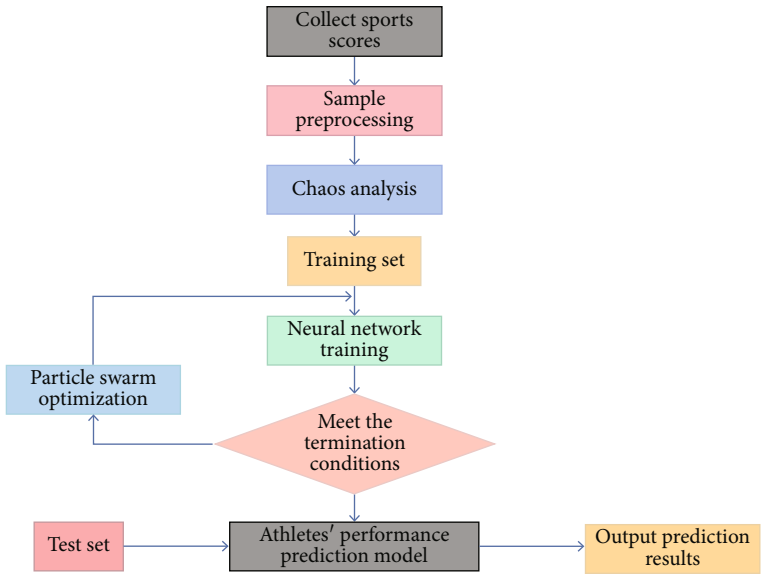


Figure 7: Analysis flow chart of the sports performance prediction model.

When the $\triangle\bar{S}(t)$ value is the smallest fraction, the T value at this time can be obtained as the most reasonable result of sports performance prediction of sports members.

3.2.3. Cao Algorithm Flow. The score data set of a single sports member is obtained. The reconstruction vector parameter of the ith position is designed as $X_i\ (M = 1)$, and the vector parameter with the smallest interval is $X_{m(i+m)}(m = 1)$, as shown in

$$\alpha(i, m)\frac{\left\|X_i(m+1) - X_{n(i+m)}(m+1)\right\|}{\left\|X_i(m) - X_{n(i+m)}(m)\right\|}. \tag{8}$$

When the result of formula (9) is the minimum parameter and the variation range of the parameter is small, the m value obtained at this time can be used as a reasonable value for the performance prediction of sports personnel. The algorithm of formula (9) is as follows:

$$E(m) = \frac{1}{N - mT}\sum_{i=1}^{N-mT}\alpha(i, m). \tag{9}$$

3.3. Machine Learning Algorithm. The neural network has strong advantages of machine computing, which can effectively transmit signals, share the analysis data from the transmission layer to the hidden layer, and then share it to the data output layer through the hidden layer to analyze the error of the output results. If the deviation rate is high, it is necessary to return the data. The hidden layer receives the prediction result with large deviation and then returns to the value transmission layer to eliminate the error until the prediction result with high accuracy is obtained. The result is shown in Figure 6.

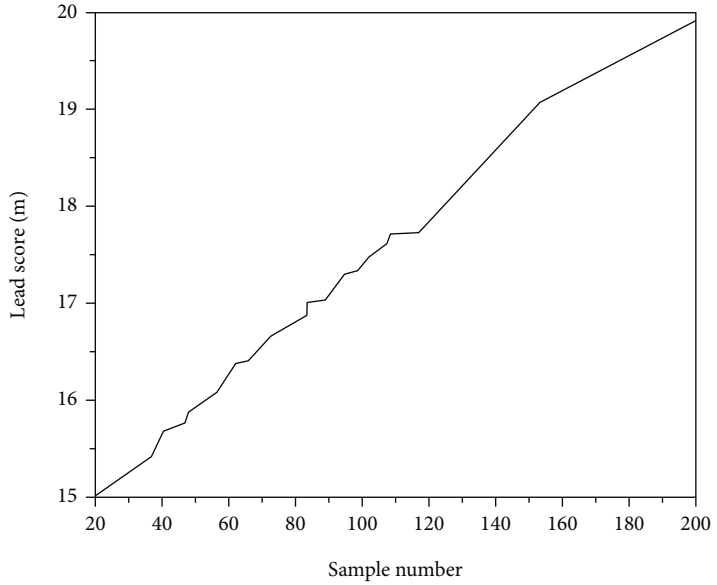


Figure 8: Prediction results of shot put results.

Table 1: Prediction results of physical education achievements of college students.

Sports events	Boys' triple jump	Boys' 300 m running	Girls' 1000 m running	Girl standing long jump
Prediction accuracy	93.84%	94.90%	96.07%	95.15%

Table 2: Statistical results of delay time of the prediction model of sports training organization.

Correlation coefficient	1	0.8	0.4	0.2
Delay time (s)	1	2	4	5

It can be seen from Figure 6 that x adds parameters to the network and y represents the output result of the hidden layer.

Output of layer: d is the expected output; $V = (v_1, v_2, \cdots, v_j, \cdots, v_m)$ indicates the connection weight between the input layer and the hidden layer; $W = (w_1, w_2, \cdots, w_k, \cdots, w_p)$ represents the connection weight between the hidden layer and the output layer. For the kth node, the calculation formula of its output value is shown in

$$o_k = f\left(\sum_{j=0}^{m} w_{jk} y_j\right). \tag{10}$$

The value of K in formula (10) is $1, 2, \cdots p$. p represents the number of neuron nodes in the output layer. The calculation formula of the jth node output of the hidden layer is shown in

$$y_i = f\left(\sum_{i=0}^{n} v_y x_i\right). \tag{11}$$

In formula (11), $J = 1, 2, \cdots, m$ and m represents the number of neuron nodes in the hidden layer. The error between the expected value and the output value is shown in

$$E = \frac{1}{2}\sum_{k=1}^{p} (d_k - o_k)^2. \tag{12}$$

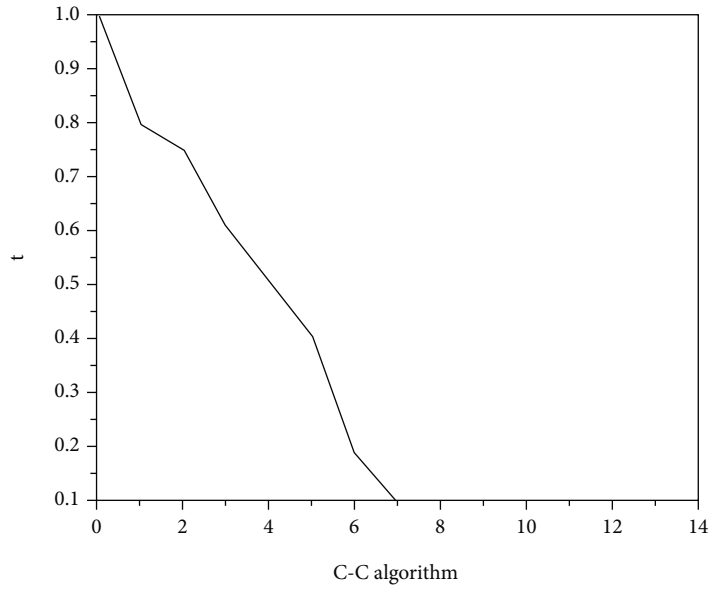
Continue to optimize the connection weights w_{jk} and v_{ij} until the prediction results have high accuracy; then, formula (13) is as follows:

$$\begin{aligned} \triangle w_{ik} &= -\eta \times \frac{\varepsilon E}{\varepsilon w_{ik}}, \\ \triangle V_{ik} &= -\eta \times \frac{\varepsilon E}{\varepsilon V_{ik}}. \end{aligned} \tag{13}$$

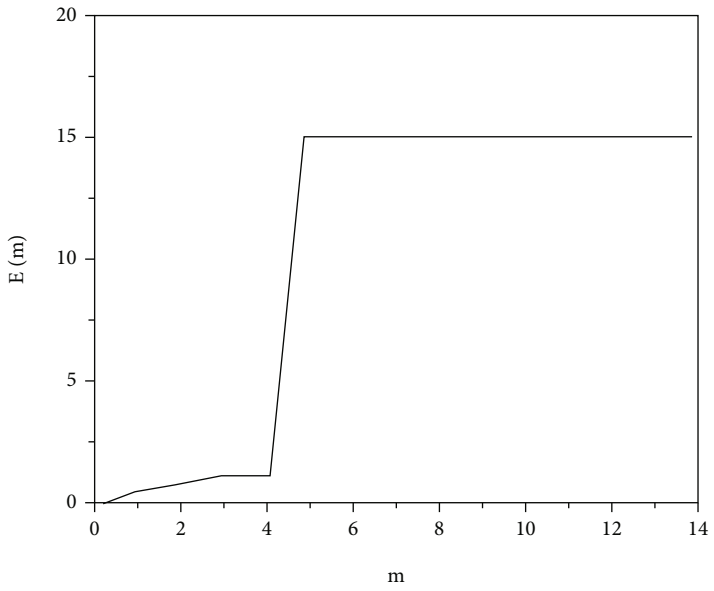
The learning speed is expressed in formula (13).

In the process of athlete performance modeling, the determination of the initial connection weight of the BP neural network is very important, and the empirical method has a certain blindness. Therefore, the particle swarm optimization algorithm is used to determine the most reasonable connection weight.

3.4. "Chaos+Machine learning" Joint Prediction Model. (1) Obtain the sports data of a sports member in a specific time, and screen out the invalid and large deviation data. (2) C-C and Cao algorithms are used to accurately obtain the T and m best results of the target members. Combined with the



(a) C-C algorithm



(b) Cao algorithm

Figure 9: Analysis diagram of T and m values of shot put results determined by the joint model of the chaotic algorithm and machine learning.

numerical analysis results, the motion information is processed again. (3) Combined with the predicted performance data of sports athletes, the structure of the neural network is determined, and the initial calculation parameters of the neural network are set synchronously. (4) Particle swarm optimization is used to effectively deal with the initial calculation parameters. Synchronously update the data location. (5) After the optimal initial calculation parameters are obtained, the prediction model of sports members' performance is created. (6) The same sports type data of other sports members are introduced to analyze the usability of the results given by the prediction model. If the accuracy of the prediction results is high, it proves that the prediction model has practical applicability. If the sports performance is not accurate, the prediction system needs to be rebuilt until the expected effect is achieved. The model analysis process is shown in Figure 7.

Figure 7 shows that the performance prediction has gone through the process of chaotic analysis and data training of mechanical learning, and the neural algorithm is used for multiple calculations to ensure the accuracy of the prediction results.

4. Result Discussion

4.1. Data Results of Different Dimensions

4.1.1. Prediction and Analysis of Shot Put Performance of Sports Members. In order to verify the accuracy of sports

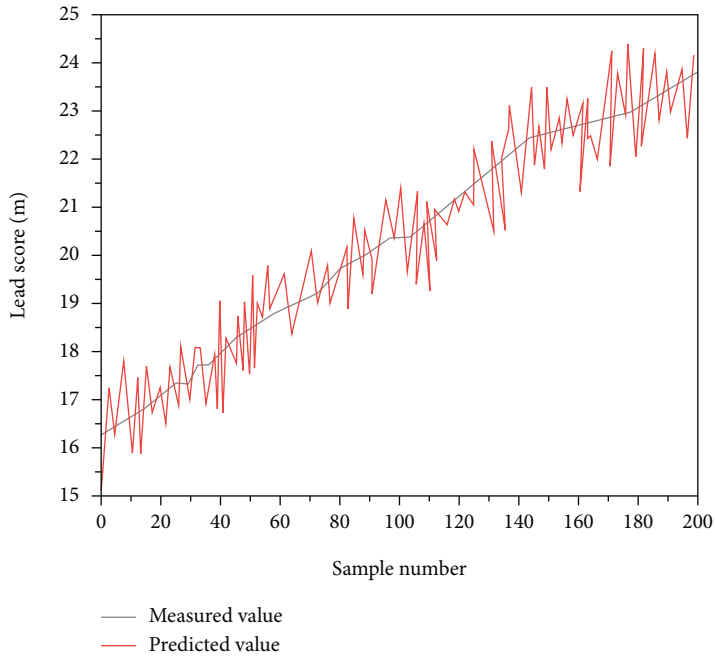


Figure 10: Shot put simulation prediction results.

performance prediction of the joint algorithm, the sports performance of a shot putter is transferred for performance prediction simulation. The prediction data of shot put sports performance is shown in Figure 8.

It can be seen from Figure 8 that when the number of shot put exercises of this PE student increases, the shot put performance is in an increasing state.

4.1.2. Prediction Results of Physical Education Achievements of College Students. A university used chaos theory combined with the algorithm system of mechanical learning to create a performance prediction model to predict the physical performance of 200 boys and 150 girls. The prediction accuracy is shown in Table 1.

4.1.3. Prediction Model of Sports Training Organization Delay Time. Table 2 shows the relationship between the delay time and the correlation coefficient when the prediction model judges the sports performance. Generally, the correlation coefficient is 2 and the delay time is 5 s.

4.2. Discussion of Data Results

4.2.1. Chaotic Analysis of Shot Put Results. The C-C and Cao algorithms are used to effectively process the shot put performance information in Figure 8 and obtain the best results of T and m. The processing results are shown in Figure 9. The prediction results in Figure 9 are as follows: when $\triangle\bar{S}(t)$ is the minimum value, $T = 8$, $m = 5$, and the change speed of $E(m)$ results is relatively slow.

The prediction results of shot put performance are shown in Figure 10. This figure shows that the deviation between the prediction results and the actual value is small, which can fully show the change law of shot put performance and has high accuracy of performance prediction.

4.2.2. Prediction and Analysis of College Students' Physical Education Achievements. In Table 1, colleges and universities make early prediction on students' achievements in long jump and long-distance running. The deviation between the measured and predicted results is small, and the prediction accuracy of various sports achievements is greater than 90%, which proves that this model is available.

4.2.3. Delay Time Analysis. In Table 2, when the correlation coefficient of the model is fixed at 0.02, the time for the model to give the prediction results is 5 s, which proves that the system has strong computing ability and can predict sports results efficiently.

5. Conclusion

(1) The sports performance of sports members is the key index to evaluate the athletes' competition and students' sports ability. The sports performance results of athletes and students not only are limited to the athletes' own physique but also interfere with the training means, methods, and external factors. Various factors will change sports performance to varying degrees, and the floating range of performance is not regular. In order to accurately obtain the changing characteristics of sports performance of personnel at all levels, a performance analysis model combining chaos theory and machine learning is proposed. Firstly, a complete collection of sports performance information of a single sports member in a specific time is carried out, and it is set that there is an internal correlation between sports performance of personnel in different time. With the help of chaos theory, this paper deeply analyzes the performance characteristics of this person at a specific time,

obtains the change trend of sports performance, comprehensively predicts the change characteristics of this person's sports performance by using the internal BP neural network of machine learning algorithm, and constructs the prediction process of sports performance of specific personnel. The active use of particle swarm optimization can effectively solve the difficulty of measuring the connection weight and threshold of the neural network and significantly enhance the efficiency of neural network performance prediction. Finally, the accuracy of the prediction simulation is verified by the real performance of athletes, and the available value of this prediction model is analyzed

(2) Through the comparison and verification of shot put data, it is found that the prediction results are accurate. This model can be used in athlete training and student sports performance management

(3) Colleges and universities predict the results of students' long jump and long-distance running in the early stage. The deviation between the measured and predicted results is small, and the prediction accuracy of various sports results is more than 90%. It is proven that this prediction model can be used in college physical education teaching management

(4) When the correlation coefficient of the model is fixed at 0.02, the time for the model to give the prediction result is 5 s, which proves that the system has strong computing ability and can predict sports performance efficiently

In order to better describe the situation of athlete performance changes, more factors need to describe the characteristics of athlete performance changes and provide more information, which is the direction of future research.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

References

[1] M. Moghadasi, H. A. Ozgoli, and F. Farhani, "A machine learning-based operational control framework for reducing energy consumption of an amine-based gas sweetening process," *International Journal of Energy Research*, vol. 45, no. 1, pp. 1055–1068, 2021.

[2] S. Chen, S. Ma, and Z. P. Liu, "Zirconia-supported zno single layer for syngas conversion revealed from machine-learning atomic simulation," *Journal of Physical Chemistry Letters*, vol. 12, no. 13, pp. 3328–3334, 2021.

[3] G. Fonseca, I. Poltavsky, V. Vassilev-Galindo, and A. Tkatchenko, "Improving molecular force fields across configurational space by combining supervised and unsupervised machine learning," *The Journal of Chemical Physics*, vol. 154, no. 12, pp. 124102–124106, 2021.

[4] P. Ziolkowski, M. Niedostatkiewicz, and S. B. Kang, "Model-based adaptive machine learning approach in concrete mix design," *Materials*, vol. 14, no. 7, pp. 1661–1665, 2021.

[5] J. Chatterjee, R. Bhattacharyya, A. Maulik, and K. Palodhi, "Application of machine learning on sequential deconvolution and convolution techniques for analysis of images from nuclear track detectors (ntds)," *Radiation Measurements*, vol. 144, no. 11, pp. 106581–106587, 2021.

[6] T.-Y. Kim, S.-H. Kim, and H. Ko, "Design and implementation of BCI-based intelligent upper limb rehabilitation robot system," *ACM Transactions on Internet Technology*, vol. 21, no. 3, pp. 1–17, 2021.

[7] P. K. Gupta, L. S. Schadler, and R. Sundararaman, "Dielectric properties of polymer nanocomposite interphases from electrostatic force microscopy using machine learning," *Materials Characterization*, vol. 173, no. 16, pp. 110909–110913, 2021.

[8] B. Ren, Z. Long, and R. Deng, "A new criterion for predicting the glass-forming ability of alloys based on machine learning," *Computational Materials Science*, vol. 189, no. 3, pp. 110259–110263, 2021.

[9] O. Altay, T. Gurgenc, M. Ulas, and C. Özel, "Prediction of wear loss quantities of ferro-alloy coating using different machine learning algorithms," *Friction*, vol. 8, no. 1, pp. 107–114, 2020.

[10] P. Tuo, X. B. Ye, and B. C. Pan, "A machine learning based deep potential for seeking the low-lying candidates of al clusters," *The Journal of Chemical Physics*, vol. 152, no. 11, pp. 114105–114109, 2020.

[11] Q. V. Vu, V. H. Truong, and H. T. Thai, "Machine learning-based prediction of cfst columns using gradient tree boosting algorithm," *Composite Structures*, vol. 259, no. 4, pp. 113505–113509, 2021.

[12] J. Adeoye, Y. T. Jia, S. W. Choi, and P. Thomson, "Prediction models applying machine learning to oral cavity cancer outcomes: a systematic review," *International Journal of Medical Informatics*, vol. 154, no. 6, pp. 104557–104560, 2021.

[13] A. A. Dhiraj, A. Agrawal, V. Meruane, and K. S. Sangwan, "Development of a machine learning based model for damage detection, localization and quantification to extend structure life," *Procedia CIRP*, vol. 98, no. 1, pp. 199–204, 2021.

[14] A. Sentuna, A. Alsadoon, P. Prasad, M. Saadeh, and O. H. Alsadoon, "A novel enhanced Naïve bayes posterior probability (enbpp) using machine learning: cyber threat analysis," *Neural Processing Letters*, vol. 53, no. 1, pp. 177–209, 2021.

[15] Y. Qiu, Z. Song, and Z. Chen, "Short-term stock trends prediction based on sentiment analysis and machine learning," *Soft Computing*, vol. 26, no. 5, pp. 2209–2224, 2022.

[16] S. P. Samyuktha, P. Kavitha, V. A. Kshaya, P. Shalini, and R. Ramya, "A survey on cyber security meets artificial intelligence: AI–driven cyber security," *Journal of Cognitive Human-Computer Interaction*, vol. 2, no. 2, pp. 50–55, 2022.

[17] K. Magudia, C. P. Bridge, K. P. Andriole, and M. H. Rosenthal, "The trials and tribulations of assembling large medical imaging datasets for machine learning applications," *Journal of Digital Imaging*, vol. 34, no. 6, pp. 1424–1429, 2021.

[18] C. Deb, Z. Dai, and A. Schlueter, "A machine learning-based framework for cost-optimal building retrofit," *Applied Energy*, vol. 294, no. 6, pp. 116990–116995, 2021.

[19] L. Yin, F. Tao, Y. Chen, F. Liu, and J. Hu, "Improving terrestrial evapotranspiration estimation across China during 2000-2018 with machine learning methods," *Journal of Hydrology*, vol. 600, no. 4, pp. 126538–126541, 2021.

[20] H. Song and M. Brandt-Pearce, "A 2-D discrete-time model of physical impairments in wavelength-division multiplexing systems," *Journal of Lightwave Technology*, vol. 30, no. 5, pp. 713–726, 2012.

[21] B. B. Gupta, K. Yadav, I. Razzak, K. Psannis, and X. Chang, "A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment," *Computer Communications*, vol. 175, no. 3, pp. 47–57, 2021.

[22] J. Li, E. Garshick, J. E. Hart, L. Li, and P. Koutrakis, "Estimation of ambient $PM_{2.5}$ in Iraq and Kuwait from 2001 to 2018 using machine learning and remote sensing," *Environment International*, vol. 151, no. 151, pp. 106445–106448, 2021.

[23] M. Rana, A. Rahman, J. Dabrowski, S. Arnold, and B. Pais, "Machine learning approach to investigate the influence of water quality on aquatic livestock in freshwater ponds," *Biosystems Engineering*, vol. 208, no. 1, pp. 164–175, 2021.

[24] I. Miyazato, T. N. Nguyen, L. Takahashi, T. Taniike, and K. Takahashi, "Representing catalytic and processing space in methane oxidation reaction via multioutput machine learning," *The Journal of Physical Chemistry Letters*, vol. 12, no. 2, pp. 808–814, 2021.

[25] N. Costa, M. Costa, and R. Barbosa, "A review on the application of chemometrics and machine learning algorithms to evaluate beer authentication," *Food Analytical Methods*, vol. 14, no. 1, pp. 136–155, 2021.

[26] K. Purwandari, J. Sigalingging, T. W. Cenggoro, and B. Pardamean, "Multi-class weather forecasting from twitter using machine learning aprroaches," *Procedia Computer Science*, vol. 179, no. 4, pp. 47–54, 2021.

[27] A. Vitui and T. Chen, "MLASP: machine learning assisted capacity planning," *Empirical Software Engineering*, vol. 26, no. 5, pp. 1–27, 2021.