Individual factor analysis of wrestler's performance based on SVM

Naidan Xu¹, Linlin Zhao¹ and Zhengzhi Wu²

¹Beijing Sport University, Beijing, China,
²The National Police University for Criminal Justice, Baoding, Hebei Province, China

Corresponding author and e-mail: Linlin Zhao, 201705024106jzh@ncist.edu.cn

Abstract. Support vector machine (SVM) is a binary classification model, its algorithm is to map the sample data into a high-dimensional space. The hyperplane found in the high-dimensional space can accurately separate two kinds of data samples and maximize the interval. According to the problem of randomness and data imbalance of wrestlers' performance data, this paper proposes a new SVM model to analyze the personal factors of wrestlers' performance. Principal component analysis and chaos analysis are applied to optimize SVM model. Compared with BPNN, GM, ARIMA, LSSVM model, the contrast experiment result shows that while the number of samples increased to 300, the SVM model of the data processing time is less than 20 seconds. Operation time reduced minimum by 11% (compared with BPNN) and maximum by 60% (compared with GM). At the same time, the accuracy of sample analysis of SVM is about 90%-95%, which has a higher accuracy than other methods. Finally, the G-mean and F-value of the SVM model in the analysis of each principal component are significantly higher than the other four models. The G-mean value increases continuously when the imbalance is between 0.1 and 0.2, while the G-mean decreases gradually when the imbalance is greater than 0.324, indicating that the closer the decision boundary is, the closer the predicted results are to the true value.

1. Introduction

Wrestling is a kind of sport in which the opponent is thrown to the ground by various techniques and methods. Wrestling has an ancient history and is considered as the oldest sport. There are many factors affecting the performance of wrestlers, such as strength quality, flexibility, age, competition experience, psychological quality and other factors [2].

For the analysis of individual factors of athletes' performance, the grey system-based model or decision tree-based analysis model is usually used at home and abroad. Effective analysis can be carried out when the number of samples is small and the number of parameters is small. However, when the number of samples increases, the accuracy of the model will be greatly reduced. Aiming at this problem, the analysis model based on SVM has become a research hotspot in recent years due to its advantage of good nonlinear prediction ability. The analysis system based on SVM algorithm needs to solve the N-order matrix when solving the problem, but when the number of parameters is large, the matrix order will be too large, which greatly increases the time of the system to deal with the problem. At the same time, the accuracy of the SVM system depends on the optimization of parameters. The traditional optimization method is mainly based on cyclic verification, but this method will lead to the
increase of the time complexity of the system. Glaire b, renmin university of China, etc according to the parameter optimization in the system time complexity of the problem is put forward based on the convex hull of sparse and the genetic algorithm to optimize the SVM model, and the samples of sparse effectively reduces the training time, and get more SVM by introducing a convex hull algorithm of support vector, and finally by cross validation, and introducing the genetic algorithm to optimize further reduces the time complexity of the system. However, certain distortion will be caused after data convex hull is carried out, which will decrease the accuracy of the model [1]. For the performance analysis of wrestlers, the sample data has certain randomness and chaos, which will further affect the accuracy and reliability of the system. To solve this problem, the xinjiang medical university Zhai XiaoHe put forward a kind of chaos analysis and least squares support vector machine forecasting model, foundation Takers theorem and chaos analysis theory is introduced to reconstruct phase space data, using phase space reconstruction after data for least squares support vector machine training, finally on the basis of the optimal parameter a prediction model is established. After simulation comparison and analysis, it is found that the average accuracy of this prediction model is 93% higher than the classical algorithm, and the average modeling time is reduced by about 10% compared with ARIMA and PBNN [3]. At the same time, Wang Jianjun et al. proposed the SVM model optimized by particle swarm optimization algorithm. By taking advantage of the strong global search ability and fast convergence speed of PSO, the model could not only improve the fitting accuracy, but also enhance the generalization adaptability of the model. The experimental results show that the root mean square of residual error of PSO-SVM model is 0.683, which improves the fitting accuracy by 50% compared with the classical SVM algorithm. Moreover, it has strong environmental adaptability and is more suitable for application in actual engineering environment [4]. Wrestlers' samples also have certain imbalance. For this problem, Jiang Fei et al. proposed an unbalanced data classification method based on mixed sampling of support vector machine. At last, some data of non-support vector region of most classes are deleted randomly. The experiment found that when the sample imbalance rate was greater than 3.42, the closer the sample was to the decision boundary, the closer it was to the real minority class support vector samples [5]. In the practical application of the SVM model, Zhang Junming et al. applied the SVM algorithm in the selection of excellent soldiers, and achieved an excellent performance with an accuracy of more than 90%. No one has yet applied the SVM model to the wrestlers achievement on the analysis of the causes of the individual, this article according to the sample and sampling randomness and chaos regard the imbalance as the starting point, puts forward a new SVM model, there will be no balance samples are divided into support vector (SV), most of the district (MNSV) and support vector minority district (FNSV), support vector chaos analysis was carried out on the SV area and separate errors in the SV area as well as a minority class near the decision boundary sample repeated sampling process, test until you find the most optimal solution, and finally delete MNSV area part samples.

2. Principles and algorithms

2.1. Chaos analysis theory
According to the Takers principle, an athlete's data sample can be reconstituted to an equivalent space, whose multidimensional expression is

$$X(t) = [x(t), x(t + \tau),..., x((m-1)t + \tau)], t = 1, 2,..., M$$

(1)

M represents the embedded dimension, $\tau$ is the delay time, M is the number of points in the space.
2.2. Support vector machines

In a given training set $T=\{(X_1,Y_1),(X_2,Y_2),(X_3,Y_3),..., (X_M,Y_M)\}$, $X_i \in \mathbb{R}^n$, $Y_i \in \{-1,1\}$, represents a minority class and -1 represents a majority class, which needs to be solved in the binary problem

$$\min \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{m} \alpha_i$$

(2)

$$\sum_{i=1}^{m} \alpha_i y_i = 0$$

(3)

Among them $K(x_i, x_j)$ is the kernel function, $\alpha_i$ is the Lagrange coefficient. Solve the optimal solution in the above equation $\alpha^* = (\alpha_1^*, \alpha_2^*, ..., \alpha_m^*)^T$. According to the KKT condition and support vector, the discriminant function is:

$$f(x) = \text{sgn} \left( \sum_{i=1}^{m} \alpha_i^* y_i K(x_i, x) + b_0^* \right)$$

(4)

As shown in figure 1, the diamond blue dots represent the kind of sample points, represented by a red dot, most sample points, the area between the H1 and H2 is considered area support vector (SV), the corresponding measure on H1 area is FNSV area, middle contains noise sample points 1. Under the H2 is MNSV area, sea area includes noise sample points 2. The SVM model training process is shown in figure 2.
3. Experimental design
Firstly, data were collected from 10 wrestlers, and the factors affecting their performance were divided into six indexes: age, competition experience, strength quality, agility, mental state before competition, and flexibility. The collected data were normalized and then chaos analysis was conducted, which was embedded and the delay time was added. Then input the processed data into the trained support vector machine, divide the minority and majority samples in each data set into 4 parts, each part is tested as a testing machine, and finally get the average value of 16 tests. The analysis results of the mixed sampling SVM model were compared with the BPNN, GM, ARIMA and LSSVM models in accuracy and modeling time. The flow chart of the experiment is shown in Figure 3.

Figure 3. Full flow chart of experimental comparative analysis.

At the same time, due to the unbalance of the acquired data, two indexes, G-mean and F-value, are adopted to measure, which are usually obtained by confusion matrix.
Table 1. Confusion matrix.

| Type                        | Prediction is class | Predict negative class |
|-----------------------------|---------------------|------------------------|
| The actual is class         | TP                  | FN                     |
| The actual negative class   | FP                  | TN                     |

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (5)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (6)
\]

\[
F - \text{value} = \frac{(1 + \beta^2) \times \text{Recall} \times \text{Precision}}{\beta^2 \times \text{Recall} + \text{Precision}} \quad (7)
\]

\[
G - \text{mean} = \sqrt{\frac{TP}{TP + FN} \times \frac{TP}{TP + FP}} \quad (8)
\]

\(\beta\) is two parameters whose weight is generally 1. Finally, the G-mean and F-value indexes of the five models under the same data are compared.

4. Results and discussion
Firstly, principal component analysis was conducted on the collected samples, and the results of principal component analysis were as follows:

Table 2. Comparison table of principal component analysis parameters.

| X1     | Psychological condition |
|--------|-------------------------|
| X2     | Experience in game      |
| X3     | Age                     |
| X4     | Power quality           |
| X5     | Agility                 |
| X6     | Flexibility             |

\[
Y_1 = 0.141X_1 + 0.301X_2 + 0.189X_3 + 0.110X_4 + 0.285X_5 + 0.40X_6
\]

\[
Y_2 = 0.198X_1 - 0.076X_2 + 0.145X_3 + 0.344X_4 + 0.126X_5 + 0.317X_6 \quad (9)
\]

Then, the simplified data obtained by principal component analysis was processed with chaos, and the obtained data was input into the trained SVM model for prediction. Meanwhile, the running time of the results was compared with BPNN, GM, ARIMA, and LSSVM models. The results were shown in Figure 4:
Figure 4. Comparison results of running time of various models.

As can be seen from the data shown in Figure 4, when the number of samples increases to 300, the data processing time of the SVM model is no more than 20 seconds. Compared with the BPNN model, the running time of the SVM model decreases by 11%, and the difference between the SVM model and the GM model is the smallest, and the difference between the SVM model and the GM model is the maximum. SVM consumes the least time and has the highest efficiency in data processing. At the same time, the accuracy of the five algorithms is compared, and the results are shown in Figure 5:

Figure 5. Comparison of accuracy of sample analysis of each method.

Can be seen from figure 5, the SVM sample analysis accuracy from 90% to 95%, with the increase of sample proficiency, various methods of sample analysis accuracy are increased, that stability in increasing and under different number of features, the increase of sample size will make the system have different degrees of accuracy, the SVM and BPNN in accuracy within a large number of sample environment, at around 96%, the SVM GM and APIMA periods of low sample size accuracy comparatively large difference, difference from 8% to 12%.

Finally, the G-mean and F-value indexes of each model are compared, and the comparison results are shown in Table 3 and Table 4:
Table 3. G-mean comparison of different algorithms.

|       | Y1    | Y2    | Y3    |
|-------|-------|-------|-------|
| SVM   | 0.7338| 0.8714| 0.4904|
| BPNN  | 0.7081| 0.7469| 0.2456|
| GM    | 0.6578| 0.6482| 0.3541|
| ARIMA | 0.6847| 0.7154| 0.1565|
| LSSVM | 0.7064| 0.6971| 0.3546|

Table 4. F-value comparison of different algorithms.

|       | Y1    | Y2    | Y3    |
|-------|-------|-------|-------|
| SVM   | 0.7178| 0.9010| 0.5277|
| BPNN  | 0.7081| 0.8469| 0.4456|
| GM    | 0.6489| 0.7482| 0.3119|
| ARIMA | 0.5847| 0.7354| 0.3465|
| LSSVM | 0.7064| 0.6151| 0.3156|

By comparing the G-mean and F-value indexes of each model in Table 3 and Table 4, it can be clearly seen that the SVM algorithm is superior, and the G-mean and F-value of the SVM model in the analysis of each principal component are significantly higher than the other four models.

Figure 6. Change trend of G-mean value with data imbalance.

And can be seen from figure 6 when the unbalanced degree is between 0.1 0.2, G - mean value will increase when not balance is greater than 0.324 G - scheme will gradually decreases, and that the closer the decision boundary, the more close to the real value, that with the increase of imbalance, at this time support vector samples closer to the real decision boundary, this method can effectively improve the accuracy of the analysis of the unbalanced samples.

5. Conclusions
In this paper, a new SVM model is proposed to analyze the effect of personal factors on wrestlers' performance according to the randomness and data imbalance of wrestlers' performance data. Firstly, the obtained data is simplified by principal component, and then the simplified data is analyzed by chaos and embedded in time. Then the processed data was substituted into the trained SVM model. The running time and accuracy of SVM were compared with BPNN, GM, ARIMA and LSSVM models. The experimental results show that when the number of samples increases to 300, the data processing time of SVM model is less than 20 seconds. Compared with the BPNN model, the running time of SVM model is reduced by 11%, and the difference between SVM model and GM model is 60%. SVM consumes the least time and has the highest efficiency in data processing. At the same time,
the accuracy of SVM sample analysis is about 90%-95%, which is higher than other methods. Finally, the G-mean and F-value of SVM model were analyzed, and it was found that the G-mean and F-value of SVM model in each principal component analysis were significantly higher than those of the other four models. It is found that when the imbalance is between 0.1-0.2, the G-mean value will increase continuously, while when the imbalance is greater than 0.324, the G-mean value will decrease gradually, indicating the relationship between decision boundary and prediction accuracy. The experimental results show that this SVM has advantages over other four methods in accuracy and time complexity, offering certain practical application value.

References
[1] Qian Hongbing, Li Yanli. A SVM Algorithm Based on Convex Hull Sparse and Genetic Algorithm Optimization [J]. Journal of Chongqing University, 201, 44(01):29-36.
[2] Eer Dun Chaolu, Yin Jinfu. Analysis on Physical Fitness Characteristics of Teenage Wrestlers [A]. Editorial Department of West China Sports Research. Journal of West China Sports Research, 2015 (139) [C].
[3] Zhai Xiaohe. A prediction model of graduate employment rate based on chaos analysis and least squares support vector machine [J].Microcomputer Applications, 201, 37(02):169-172.
[4] Wang Jianjun, Zhang Chongyang. GPS height fitting method based on particle swarm optimization SVM [J]. Beijing Surveying and Mapping, 2021, 35(01):83-88.
[5] Jiang Fei, Yang Ming, Liu Yuxin. Unbalanced Data Classification Method Based on Mixed Sampling of Support Vector Machine [J]. Mathematics in Practice and Theory, 201, 51(01):88-96.
[6] Zhang Junming, Song Ying, Yu Huamou, Zheng Peisen, Song Xuexia, Li Tong. Application Research of SVM Algorithm in Selection of Outstanding Soldiers [J]. Fujian Computer, 201, 37(02):6-8.