ISSN : 0974 - 7435 Volume 10 Issue 8

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(8), 2014 [2731-2739]

BP neural network-based world men decathlon performance development trend research

Zhendong He, Ruimin Hu

Physical Education Department, Hebei University of Science and Technology, Shijiazhuang 050018, Hebei, (CHINA)

ABSTRACT

The paper makes statistics of world men decathlon world annual best performances in ten years during 2004~2013, applies principal component analysis and BP neural network combinative method to establish prediction model. The paper carries on principal component analysis of statistical data, and solves four principal components. On the basis of principal component analysis result, take previous year men decathlon four principal components as input, and the second year men decathlon performances as output, it establishes BP neural network model. The paper takes neural network training with performances during 2004~2011, uses performances from 2011 to 2013 to detect established prediction model accuracy, applies 2013 principal component data into predicting, and gets men decathlon annual best performance in 2014 is 8750.5 points. Take principal component analysis' data handling result as input, it solves men decathlon individual event many data and each individual event performance stronger correlations bad impacts on prediction model's prediction accuracy, enhances prediction model's prediction accuracy, and expands BP neural network predication model application.

KEYWORDS

Men decathlon; Principal component analysis; BP neural network; Performance prediction.

INTRODUCTION

Men decathlon is a kind of comprehensive, all-round synthesis men competition event, is one kind of all-round events in athletics events. It is composed of 100m, long jump, shot put, high jump, 400m, 110m hurdle, discus, pole vault, javelin throw and 1500m ten sports events. Its complex scoring system lets athlete to essentially possess abundant physical ability, all-round techniques and strong will, and meanwhile develop their advantageous events, is all-round test on athletes, it is named of "ironman" event.

Sports' performance prediction accuracy extent directly affects athletes training plan making and training schedule, and has bigger impacts on athlete performance enhancement. By far, in the aspect of applying BP neural network prediction model to predict sports performance, domestic and foreign lots of scholars have made relative studies. BP neural network can make full use of existing information, and predicts future sports performance by previous sports performances. Principal component analysis method can utilize thought of dimension reduction to reduce indicators numbers, and solve inconvenience that is brought to model establishment by indicators amount and correlations in problems to be solved. Principal component analysis approach and BP neural network model have been widely used in relative sports field problems, and the combination of the two will also provide new ideas to solve sports problems. For principal component analysis and BP neural network application, from which Li Jian-Jun (2002) applied documents literature, questionnaire survey, logical reasoning and others to analyze and research on principal component analysis method application in sports teaching evaluation, and provided references for systematical evaluating education and teaching[1]; Zhang Yu etc. (2013) researched on the 30th Olympic Games men five events athletics champions performances, with the 23~29th sessions performances as evidence, applied BP neural network prediction model to predict their performances[2]; Li Ping (2014) on the basis of Chinese and foreign seventeen metropolitans sports tourism development, applied analytic hierarchy process and principal component analysis method to study on Chinese sports tourism and foreign countries gap, it provided references for promoting Chinese cities soft power[3]; Wu Jin-Zhou etc.(2014) applied principle component analysis method and BP neural network combination, they analyzed wine data, and verified established model effectiveness by simulating data[4]; Huang Xiang-Jun (2008) utilized principal component analysis method to reduce dimension of data in problems that to be solved, and based on processed data, established BP neural network prediction model, and solved accuracy declining problems when BP neural network prediction model predicted on multiple dimensions data[5].

The paper makes improvement on BP neural network prediction model, applies principal component analysis method to handle with statistical data, reduces BP neural network input dimensions, and finally on the basis of BP neural network prediction model, it establishes men decathlon performance prediction model, promotes men decathlon performance prediction accuracy while at the same time analyzes and researches on principal component method and BP neural network prediction model combinative model's application in sports prediction.

DATA STATISTICS AND PREPROCESSING

Men decathlon annual highest performance data statistics

In order to promote prediction accuracy on world men decathlon annual highest performance, its statistical data not only is required to make statistics of previous world men decathlon annual highest performances athletes total performances, but also is required to make statistics of their each individual event performance so as to make more effective mining on their total performances change rules.

The paper makes statistics of world men decathlon annual highest performances in ten years during 2004~2013, and based on this, it establishes prediction model. Year2004~2013 world men decathlon annual highest performances statistical data is as TABLE 1 shows.

TABLE 1 : World men decathlon annual highest performance

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Total performance	8893	8732	8677	8697	8832	8790	8483	8729	9039	8809
100m	10.85	10.43	10.42	10.94	10.39	10.45	10.35	10.33	10.21	10.35
Long jump	7.84	7.54	7.67	7.84	7.39	7.83	7.51	7.8	8.23	7.73
Shot put	16.36	16.25	15.56	16.47	15.17	15.33	15.38	14.14	14.20	14.39
High jump	2.12	2.00	2.06	2.12	2.08	1.99	2.06	2.05	2.05	1.93
400m	48.36	47.78	48.87	48.99	48.41	48.13	49.66	46.35	46.70	46.02
110m hurdle	14.05	14.43	13.74	14.39	13.75	13.86	14.08	13.52	13.70	13.72
Discus	48.72	53.68	52.21	47.66	52.74	48.08	49.85	41.58	42.81	45.00
Pole vault	5.00	4.90	5.00	4.80	5.00	5.20	4.60	5.05	5.30	5.20
Javelin throw	70.52	72.00	66.47	68.87	70.55	68.00	66.1	56.19	58.87	64.83
1500m running	300.0	303.8	313.5	280.4	290.9	288.0	296.4	264.1	254.5	269.8

Statistical data processing

The paper's prediction on world men decathlon event annual highest performance is prediction on their total performances, in data prediction, prediction methods that often use are traditional regression prediction, time sequence prediction and else, while also recent years newly-developed grey prediction, fuzzy prediction and neural network predication and so on. For traditional prediction model and grey prediction model and else, their data changes should follow certain change rules, and each prediction model's prediction on different change rules data also have their own advantages and disadvantages.

In order to easier analyzing world men decathlon annual highest performances change trend during 2004~2013 so as to easy selecting different prediction models, the paper processes with TABLE 1 total performances data, in TABLE 1 data total performances changing curve is as Figure 1 shows.

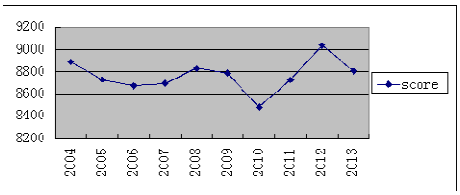

Figure 1 : 2004~2013 world men decathlon annual highest performances changing graph

By Figure 1, it is clear during ten years from 2004 to 2013, men decathlon world annual highest performances are in fluctuation changes, from which performances in 2010 and 2012 are respectively the lowest performance and highest performance in the ten years. Due to performance changing trend and changing rules are not obvious, traditional statistics methods are hard to establish prediction model on them, so the paper selects BP neural network model to establish prediction model on them.

Due to TABLE 1 statistical data dimensions are different, each individual event performance and total performance statistical data values changes are bigger, and it is bad for its principal component analysis and BP neural network prediction model establishment. For the convenience of the paper's men decathlon performance prediction model establishment, the paper carries on normalization processing on them, for total performance, long jump, shot put, high jump, discus, pole vault and javelin throw these seven events performances, statistical data gets bigger, it shows performances are better, and other four individual events performances are just the opposite. So for total performance and long jump so on seven events, take 2004 total performance and each individual event performance as denominator, and take 2004~2013 corresponding total performance and each individual event performance as numerator, carry on processing. For 100m, 400m, 110m hurdle and 1500m these four individual events, it takes 2004 total performance and each individual event performance as numerator, and takes 2004~2013

corresponding total performance and each individual event performance as denominator, carries on processing. Their processing result is as TABLE 2 shows.

TABLE 2: World men decathlon annual highest performance preprocessing result

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Total performance	1.000	0.982	0.976	0.978	0.993	0.988	0.954	0.982	1.016	0.991
100m	1.000	1.040	1.041	0.992	1.044	1.038	1.048	1.050	1.063	1.048
Long jump	1.000	0.962	0.978	1.000	0.943	0.999	0.958	0.995	1.050	0.986
Shot put	1.000	0.993	0.951	1.007	0.927	0.937	0.940	0.864	0.868	0.880
High jump	1.000	0.943	0.972	1.000	0.981	0.939	0.972	0.967	0.967	0.910
400m	1.000	1.012	0.990	0.987	0.999	1.005	0.974	1.043	1.036	1.051
110m hurdle	1.000	0.974	1.023	0.976	1.022	1.013	0.998	1.039	1.026	1.024
Discus	1.000	1.102	1.072	0.978	1.083	0.987	1.023	0.853	0.879	0.924
Pole vault	1.000	0.980	1.000	0.960	1.000	1.040	0.920	1.010	1.060	1.040
Javelin throw	1.000	1.021	0.943	0.977	1.001	0.964	0.937	0.797	0.835	0.919
1500m	1.000	0.988	0.957	1.070	1.031	1.042	1.012	1.136	1.179	1.112

PREDICTION SYSTEM ESTABLISHMENT

BP neural network prediction model is a kind of simple and steepest descent optimizing method, it can fully excavate data relationship, it has good effects on sports performance prediction problems with fewer data information amount, but BP prediction system itself has certain drawbacks, as when input space is seriously self-correlated and network dimensions are higher, its prediction effects will go bad[5]. In order to promote men decathlon performances prediction accuracy, the paper needs to apply all effective data that are sports competitions' total performances and individual events performances. Due to ten individual events partial events correlations are higher, it surely causes established BP neural network model input data dimensions are higher and correlations are big, it affects model prediction accuracy. The paper applies principal component analysis to handle with data, and on the basis of handled data, it establishes BP neural network prediction model.

Sketch principal component analysis method

The purpose of principal component analysis is solving variables information highly overlapping and highly correlation impacts on problems to be solved, it can cut down problem's variables amount and also reserve variables information to largest extent, solves information loss and incomplete problems due to cut down variables. Its calculation steps are as following:

According to problems, select corresponding indicator data, if data statistical criterion and order of magnitudes are different, it needs to first carry on dimensions and order of magnitudes differences elimination, and gets standard matrix.

By formula computation selected elements covariance matrix, covariance matrix R is composed of element r_{ij}, its element R_{ij} computational formula is as following:

$$r_{ij} = \frac{\sum_{k=1}^{n}(X_{kj} - X_i)(X_{kj} - X_j)}{\sqrt{\sum_{k=1}^{n}(X_{kj} - X_i)^2 (X_{kj} - X_j)^2}}$$

Among them, R_{ij} is correlation coefficient after original variable or dimensions unifying, $R_{ij} = R_{ji}$, that is covariance matrix R real symmetric matrix.

Compute covariance matrix R feature value, feature value computational formula is as following

$$\left|\lambda E - R\right| = 0$$

By above formula, it can compute covariance matrix R feature value λ_i, from which i is the number of variables. Among them, λ_i is positive number, rank it with sizes, its value gets bigger then corresponding principal components included information amount would be more, and its impacts would be bigger. Each principal component can be expressed by z_i, its contribution rate can use W_i every principal component corresponding contribution rate that is solved by following formula

$$W_i = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$$

Among them, n is number of original variable or variable through dimension or order of magnitudes unifying. According to W_i size order, it calculates solved principle components accumulation rate, then previous m pieces of principal components accumulative contribution rate can be solved by formula $\sum_{j=1}^{m} \lambda_j / \sum_{j=1}^{n} \lambda_j$, finally select accumulation contribution rate as 85%~95% previous m pieces of principal components.

In order to reflect solved principal component z_i and original variables or variables after dimension or order of magnitudes unifying correlation degree, it needs to do principle component loading l_{ij} solution on principle component z_i, compute feature value λ_i corresponding feature vector a_i, then principal component loading solution can be solved by following formula :

$$l(Z_i, X_j) = \sqrt{\lambda_i} a_{ij}$$

Similarly it solves principal component expression as following:

$$z_i = l_{1i} X_1 + l_{2i} X_2 + \cdots + l_{ni} X_{\mathrm{n}}$$

Sketch BP neural network

BP neural network is a kind of feedforward neural network with monitoring learning algorithm, BP neural network its computation process is as following:

Set initial parameters of BP neural network model that is required to establish, according to input value and output value of problems to be solved, set neural network model input layer, output layer and hidden layer nerve cells numbers, and set network initial weight matrix learning factor η and momentum factor a and other parameters values.

Provide training mode, train network by corresponding training modes, and excavate data relationship, till established BP neural network model to meet requirements.

According to training mode and neural network weight, compute network output mode, its computational formula is as following :

$$a_{pj} = \sum_{j=1}^{N} W_{ji} o_{pi}$$

$$o_{pj} = f(a_{pj}) = \frac{1}{1 + e^{-a_{pj}}}$$

Among them, p represents the p training mode, j represents neural network cell the j nerve cell, a_{pj} is one nerve cell input totality, o_{pj} represents the nerve cell output. Compare the j nerve cell output amount o_{pj} with expected value, if output amount doesn't meet requirements, then execute (4), if it meets demand, then return to (2). Among them, neural network error computational formula is as following:

$$E = \sum_{p} E_p$$

$$E_p = \frac{1}{2}\sum_{i}(d_{pj} - o_{pj})^2$$

Among them, E is global error.

(4) Apply counter propagation and maximum gradient descent to train neural network, let it to get satisfied effects, excavate data relationship. If computational accuracy doesn't meet demands, it can correct neural network weights by corresponding functions, correction formula is as following:

$$W_{ij}(t+1) = W_{ij}(t) + \eta E o_{pj} + a[W_{ij}(t) - W_{ij}(t-1)]$$

After weights correcting, then return to (2). That is select mode again, train neural network, till neural network system meets demands.

BP neural network prediction model establishment

The paper established BP neural network prediction model has input layer, output layer and hidden layer, from which hidden layer has one layer nerve cell, model selects S type tangent function transfer function and pure linear function, gradient descent momentum BP gain function, gradient descent momentum weight learning function and mean square error normalization property function. Among them, parameter a and η are respectively 0.5and 0.2, neural network maximum training times are 10000times, training accuracy is 0.005, training indication interval is 1000times. Output layer nerve cell number is 1, input layer nerve cell number is number of principal components after principal component analysis handling, hidden layer nerve cells amount is solved by formula $i = \sqrt{n+m} + a$, from which a value is between 0 and 1.

MODEL RESULT AND RESULT ANALYSIS

Principal component analysis result

Compute TABLE 2's ten individual sports performances normalization data composed matrix, it can solve correlation coefficient matrix R each element value that is composed of r_{ij} as following :

$$R = \begin{bmatrix}
1.00 & -0.01 & -0.80 & -0.58 & 0.46 & 0.62 & -0.21 & 0.36 & -0.55 & 0.35 \\
-0.01 & 1.00 & -0.30 & 0.05 & 0.44 & 0.22 & -0.73 & 0.60 & -0.55 & 0.66 \\
-0.80 & -0.30 & 1.00 & 0.47 & -0.73 & -0.86 & 0.68 & -0.57 & 0.84 & -0.74 \\
-0.58 & 0.05 & 0.47 & 1.00 & -0.59 & -0.24 & 0.14 & -0.45 & 0.14 & -0.21 \\
0.46 & 0.44 & -0.73 & -0.59 & 1.00 & 0.55 & -0.69 & 0.75 & -0.62 & 0.74 \\
0.62 & 0.22 & -0.86 & -0.24 & 0.55 & 1.00 & -0.52 & 0.61 & -0.70 & 0.47 \\
-0.21 & -0.73 & 0.68 & 0.14 & -0.69 & -0.52 & 1.00 & -0.48 & 0.86 & -0.90 \\
0.36 & 0.60 & -0.57 & -0.45 & 0.75 & 0.61 & -0.48 & 1.00 & -0.39 & 0.54 \\
-0.55 & -0.55 & 0.83 & 0.14 & -0.62 & -0.70 & 0.86 & -0.39 & 1.00 & -0.77 \\
0.35 & 0.66 & -0.74 & -0.21 & 0.74 & 0.47 & -0.90 & 0.54 & -0.77 & 1.00
\end{bmatrix}$$

Compute correlation coefficient matrix R, solve its feature values ten feature values λ_i as 5.9049, 1.7516, 1.0259, 0.6261, 0.3481, 0.1969, 0.1236, 0.0156, 0.0073 and 0.0001, from which $i = 1,2,\cdots,10$. When several principal components total contribution rates arrive at 85%~95%, it can use these principle components to replace other variables. By λ_i value, it can solve $\lambda_1 = 5.9049$, $\lambda_2 = 1.7516$, $\lambda_3 = 1.0259$ and $\lambda_4 = 0.6261$ four feature values corresponding principal component contribution rates arrive at 93.09%, so select the four principal components. Four feature values corresponding feature vectors are respectively:

$$W_1 = [0.260 \quad 0.244 \quad -0.381 \quad -0.185 \quad 0.355 \quad 0.318 \quad -0.347 \quad 0.305 \quad -0.357 \quad 0.354]^T$$
$$W_2 = [-0.486 \quad 0.506 \quad 0.198 \quad 0.488 \quad -0.059 \quad -0.182 \quad -0.342 \quad -0.016 \quad -0.128 \quad 0.248]^T$$
$$W_3 = [-0.228 \quad 0.238 \quad 0.244 \quad -0.487 \quad 0.334 \quad -0.316 \quad 0.065 \quad 0.479 \quad 0.387 \quad 0.002]^T$$
$$W_4 = [-0.059 \quad 0.149 \quad 0.015 \quad 0.363 \quad -0.143 \quad 0.559 \quad 0.247 \quad 0.573 \quad 0.151 \quad -0.312]^T$$

Handle with four feature vectors, solve their corresponding unit vectors, according to normalization processing result, it can solve four principal components loadings as :

$$l_1 = [0.631 \quad 0.594 \quad -0.926 \quad -0.450 \quad 0.862 \quad 0.774 \quad -0.842 \quad 0.742 \quad -0.868 \quad 0.861]$$
$$l_2 = [-0.643 \quad 0.669 \quad 0.262 \quad 0.646 \quad -0.078 \quad -0.240 \quad -0.452 \quad -0.022 \quad -0.169 \quad 0.329]$$
$$l_3 = [-0.231 \quad 0.241 \quad 0.247 \quad -0.494 \quad 0.338 \quad -0.320 \quad 0.066 \quad 0.485 \quad 0.392 \quad 0.002]$$
$$l_4 = [-0.047 \quad 0.118 \quad 0.012 \quad 0.287 \quad 0.454 \quad 0.442 \quad 0.196 \quad -0.113 \quad 0.120 \quad -0.247]$$

By above analysis, it is clear that in the first principal component, the larger loading values components are shot put, discus, javelin throw and 1500m four events. Among them, shot put, discus and javelin throw all belong to throwing type event, 1500m belongs to endurance type event, which shows the two types of events have largest impacts on world men decathlon annual highest performance. In the second principal components, larger loading components are 100m, long jump and high jump three events, the three is related to athletes lower limbs strength and explosive power. In the third principal component, larger loading components are high jump and pole vault. In the fourth principal component, larger loading events are 400m and 110m hurdle. By above analysis, it is clear that solved four principal components contain world men decathlon annual highest performances' each event information to very high extent. Established principal component analysis model analysis effects on men decathlon world annual best performance is good, it can proceed with BP neural network prediction model analysis.

By above solved loading amount, it can solve 2004~2013 four principal components values as TABLE 3 shows.

TABLE 3: Solved four principal components values

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
First principal component	1.376	1.272	1.408	1.404	1.418	1.607	1.412	2.037	2.077	1.872
Second principal component	0.300	0.163	0.186	0.352	0.170	0.241	0.216	0.340	0.368	0.245
Third principal component	0.727	0.753	0.685	0.704	0.693	0.735	0.635	0.615	0.669	0.720
The fourth principal component	1.221	1.202	1.237	1.270	1.224	1.207	1.165	1.137	1.160	1.161

Neural network prediction model result

The paper established neural network prediction model uses previous year four principal components values to predict next year total performance. By above analysis, it is clear, established BP neural network prediction model system input has four principal components values that input layer has

four nerve cells, set neural network output and four principal components corresponding next year men decathlon total performance value that output layer nerve cell unit number is one. According to input layer and output layer nerve cell number, it defines hidden layer nerve cell number is three. The paper takes year 2004~2011 data as training mode, uses year 2012 and 2013 two years data to detect established model, and apply year 2013 data to predict year 2014 annual highest performance. Then training mode input matrix P1 and output matrix T1 are respectively as following:

$$P1=\begin{bmatrix} 1.376 & 1.272 & 1.408 & 1.404 & 1.418 & 1.607 & 1.412 \\ 0.300 & 0.163 & 0.186 & 0.352 & 0.170 & 0.241 & 0.216 \\ 0.727 & 0.753 & 0.685 & 0.704 & 0.693 & 0.725 & 0.635 \\ 1.221 & 1.202 & 1.237 & 1.270 & 1.224 & 1.207 & 1.165 \end{bmatrix}$$

$$T1=\begin{bmatrix} 0.982 & 0.976 & 0.978 & 0.993 & 0.988 & 0.954 & 0.982 \end{bmatrix}$$

Utilize set training mode to train BP neural network model, its training process curve graph is as Figure 2 shows.

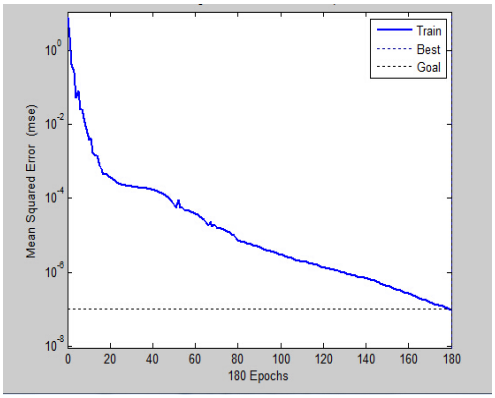

Figure 2 : Training process curve

As Figure 2, neural network model through 180 times training, its model accuracy then arrives at set accuracy, its convergent effects are also good. Respectively input year 2011 and year 2012 data into established BP neural network prediction model, its prediction results are respectively 1.0063 and 1.0000, errors with the two years corresponding actual values 1.0164 and 0.9906 are respectively 0.99%and 0.95%, prediction accuracy is very high. Input year 2013 data into prediction model, it gets its predicted value is 0.9840, multiply it by year 2004 highest performance, then it get year 2014 annual highest performance prediction model is 8750.7.

CONCLUSION

Men decathlon world annual highest prediction has important significances in world men decathlon development, but due to its performance change trend fluctuation is stronger, change rules are not obvious, traditional prediction model is hard to make prediction model on it. The paper applies BP neural network prediction model to model it, but men decathlon each event performance correlation is bigger and individual event data is numerous and other problems surely will bring bad impacts to BP neural network modeling.

To solve above problems bad impacts on modeling, the paper applies principal component analysis to process with statistical data; take established principal components as input, and total performance as output. It solves above problem, and meanwhile, it promotes established prediction model’s prediction accuracy.

REFERENCES

[1] Wu Shao-ming, Zhou Zhi-jie; The research on fatigue and recovery in sports training[J]. Journal of Wuhan Institute of Physical Education, **35(4)**, 93-94 **(2001)**.

[2] Chen Xue-mei; Comprehensive Evaluation and Recovery on Exercise Fatigue of Athlete[J]. Bulletin of Sport Science & Technology, **15(6)**, 22-23, 43 **(2007)**.

[3] Lan Zheng; The Analysis on the Main Influencing Factors of the Football Players' Running Ability[J]. Sport Science and Technology, **29(1)**, 58-60 **(2008)**.

[4] Zhong Da peng, Zhang Xin yun; Occurrence and elimination of sports fatigue[J]. Journal of Wuhan Institute of Physical Education, **36(5)**, 50-51 **(2002)**.

[5] Hu Qi; The Producing of the Fatigue and Recovery[J]. Journal of Physical Education Institute of Shanxi Teachers University, **19(1)**, 90-92 **(2004)**.

[6] Zhao Zhong-hua; Research Progress of Neuro-muscular Fatigue in Sport[J]. Journal of Shenyang Sport University, **3**, 44-46 **(2003)**.