Evaluation model and simulation of basketball teaching quality based on maximum entropy neural network

Zhang Xiaodong *

Department of sports training and competition, Shandong Sport University, Jinan 250102, China

Received 15 October 2014, www.cmnt.lv

Abstract

Establishment and implementation of models of teaching quality evaluation are important tasks for teaching management in colleges. In this thesis, we use a neural network (NN) model optimized by maximum entropy principle, conduct fitting for complicated nonlinear relationship between many indices and evaluation results to evaluate basketball teaching quality, and implement simulated residual comparison with back propagation (BP) NN model that has not been optimized. Results demonstrate that the evaluation results of maximum entropy NN model are better than those of the BP NN model that has not been optimized in experimental function simulation and example verification, thereby indicating that the optimized model has strong generalization ability and high degree of confidence. This optimization algorithm is feasible in establishing evaluation models for basketball teaching quality.

Keywords: Maximum entropy, Neural network model, Optimization algorithm, Evaluation model, Basketball teaching quality

1 Introduction

Back propagation (BP) neural network (NN) is the most represented among all NN models [1]. This network model is a multilayer mapping NN [2]. The multi-index comprehensive evaluation method based on the BP neural network may establish comprehensive evaluation models that approach thinking modes of people extensively through high self-adaptive ability, self-learning ability, and strong fault-tolerance ability of artificial NN to satisfy evaluation requirements [3-4].

Entropy is used to express the degree in which any kind of energy is distributed in space evenly. The entropy is increased when the energy distribution is significantly uniform [5]. When the energy of a system is distributed uniformly, the entropy of this system reaches its maximum [6]. Maximum entropy principle means that existing information must be used fully and a group of probability distribution must be chosen with maximum information entropy as results of estimation or prediction under the situation that information is limited or probability space is imperfect [7]. Maximum entropy estimation is an estimation method with minimum influence from subjective factors and maximum randomness, and the principle of maximum entropy ensures that the maximum entropy model has good generalization effect [8].

Teaching quality is the core content of educational competitiveness and lifeline of higher education, particularly for vocational colleges, as well as the foundation of sustainable development in education [9]. Recently, the Ministry of Education in China has evaluated the teaching quality of many colleges. Thus, establishment and implementation of teaching quality evaluation are important tasks for teaching management of colleges [10]. Teaching quality evaluation refers to the process in which effective technical methods are used to judge teaching value under the situation that many indices of teaching quality are evaluated comprehensively to improve teaching activities and teaching quality [11]. Teaching quality evaluation of colleges is a complicated problem that involves numerous factors, such as teaching conditions, course difficulty, teaching strategy, and learning effect. Such factors interact with one another [10]. The relationship among evaluation factors is complicated, and teaching quality is affected by many external factors. Consequently, a universally acknowledged and ideal teaching quality evaluation system does not exist. In recent years, several scholars have conducted substantial work on how to evaluate teaching quality accurately [12]. However, the aforementioned methods consider evaluation models of teaching quality as linear regression models and ignore the complicated nonlinear relationship between evaluation objects and evaluation elements. Yu Liang recognized that the teaching quality evaluation of colleges is a multivariable, multifactor, and fuzzy nonlinear process. This evaluation also employs BP NN to build an evaluation model of teaching quality.

The maximum entropy model is a universal frame of machine learning. This model has been successfully applied to many fields, including space physics, computer vision, and natural language processing. However, the maximum entropy model is rarely used in teaching evaluation. Based on the theory and method of BP NN, this thesis uses the optimization algorithm of the maximum entropy model and constructs an evaluation model that is appropriate for basketball teaching quality in colleges.

2 Maximum entropy NN model

2.1 ENTROPY

Entropy describes the uncertainty of random variables. The entropy for a discrete random variable is defined as

* ***Corresponding author's*** e-mail: zhangxiaodong1234@yeah.net

$$H(X) = -\sum_i P(X = a_i)\log P(X = a_i), \quad (1)$$

where a_i is the value of X and $P(X = a_i)$ represents the probability when the value of X is a_i.

2.2 MAXIMUM ENTROPY PRINCIPLE

According to the maximum entropy principle, Jiann Horng Lin et al. confirmed probability distribution in a system, considering that the probability of distribution is presented as follows in a multidimensional output system in accordance with the maximum entropy principle:

$$\mathrm{P}_m = \frac{e^{-\mu(y_m - d_m)^2}}{\sum_m e^{-\mu(y_m - d_m)^2}} \quad (2)$$

where μ is a parameter, m refers to the m th output neural node in the NN model, d_m represents desired output, and y_m is actual output. According to Formula (2), entropy of the system is expected output, and y_m is practical output. According to Formula (2), the system entropy is expressed as

$$H = -\sum_m P_m \ln P_m = -\sum_m \left[\frac{e^{-\mu(y_m - d_m)^2}}{\sum_m e^{-\mu(y_m - d_m)^2}} \ln \frac{e^{-\mu(y_m - d_m)^2}}{\sum_m e^{-\mu(y_m - d_m)^2}}\right] \quad (3)$$

2.3 MAXIMUM ENTROPY MODEL

In this study, the maximum entropy optimization algorithm is used based on the BP NN model to establish an evaluation model of teaching quality.

Only one output node exists in the evaluation model of teaching quality. Thus, m in Formula (1) is usually 1. The NN model is illustrated in Figure 1.

Error evaluation function of the traditional BP NN model usually involves error of mean square, which is expressed as follows:

$$E = \sum_{P=1}^{P} E_P = \frac{1}{2}\sum_{P=1}^{P}\sum_{j=1}^{m}(t_{pj} - y_{pj})^2 \quad (4)$$

where E_p is the error of the p th sample, and t_{pj} and y_{pj} are the actual output and model prediction output, respectively.

Wei Zhengyuan et al. concluded that methods optimizing BP NN models based on the maximum entropy principle finally replay the error evaluation function of traditional BP NN models with Formula (3) and deduce the error function of the maximum entropy NN model as follows:

$$J = -\alpha\sum_n\sum_m\left[\frac{e^{-\mu(y_m - d_m)^2}}{\sum_m e^{-\mu(y_m - d_m)^2}} \ln \frac{e^{-\mu(y_m - d_m)^2}}{\sum_m e^{-\mu(y_m - d_m)^2}}\right] + \beta\sum_n\sum_m[(y_m - d_m)^2] \quad (5)$$

where α and β are constant coefficients larger than 0. They are set as 1 in the following context.

Excitation function of the BP network output node should vary according to different applications. If the BP network is applied to classification, then the output nodes will utilize the sigmoid function or hard limiting function. If the BP network is applied to functional approximation, then the output nodes should adopt a linear function. Thus, the linear activation and sigmoid activation functions are selected for the output and hidden layers of BP NN in this study, respectively. In accordance with the error function J, the NN model improved the algorithm under the maximum entropy error criterion that can be obtained via gradient descent.

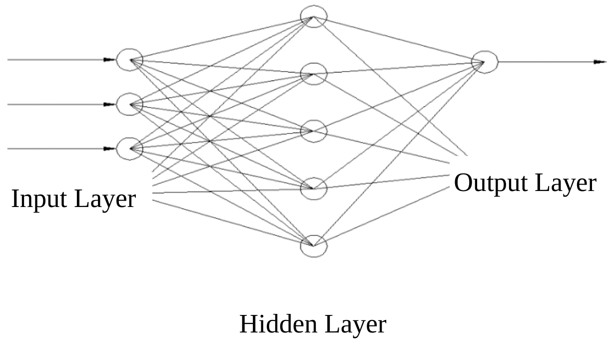

FIGURE 1 Structure of NN model

2.4 GRADIENT DESCENT

Gradient descent is also called steepest descent method, which is a commonly used learning method in NN models. The basic principle of this method is that the target of nerve cell weight correction is assumed to be minimization scalar function $J(W)$. If current weight of the nerve cell is $W(k)$, then the weight modifier formula at the next moment is assumed to be

$$W(k+1) = W(k) + \Delta W(k) \quad (6)$$

where $\Delta W(k)$ refers to weight correction at the current moment. Under a general condition, each correction should achieve the following:

$$J[W(k+1)] < J[W(k)] \quad (7)$$

Taylor series expansion is performed at k for $J[W(k+1)]$. Then, we may obtain

$$J[W(k+1)] = J[W(k) + \Delta W(k)] \approx J[W(k)] + g^T(k)\Delta W(k), \quad (8)$$

where $g(k) = \nabla J(W(k))_{W=W(k)}$ represents gradient vector of $J(W)$ at $W(k)$.

If $\Delta W(k) = -\eta g(k)$ (η refers to learning rate and stays on [0,1]) and minimum value is taken for weight correction along the negative gradient direction, then the second item on the right side of Formula (8) is less than 0. At this moment, Formula (8) is valid undoubtedly. This concept is the basic principle of gradient descent.

3. Algorithm simulation and examples of maximum entropy model

3.1 ALGORITHM SIMULATION

3.1.1 Multivariate nonlinear function fitting

Following the method of Wei Zhengyuan et al., we use the BP NN model that does not adopt the optimization algorithm and optimized maximum entropy model to perform fitting for multivariate nonlinear function $y = \frac{1}{x_1} + x_2^{\ 2} + x_3^{\ 3}$, respectively. Fitted residual value is used as a standard. Advantages and disadvantages of two models are judged based on this value.

According to expression of functions that need fitting, the number of input and output modes of the NN model is 3 and 1, respectively. The number of nodes at the hidden layer is decided as 10 by performing cut-and-try testing. Sigmoid activation function and linear activation function are applied to the hidden and output layers of the model, respectively. Based on the model, load maximum entropy optimization and traditional non-optimized algorithms implement fitting to objective functions. Specific operation procedures are as follows:

a) Three groups of value are selected in the interval [−1,1] randomly (each group of value is selected 10,000 times). Three independent variables, including x_1, x_2, and x_3, are provided, and a matrix of 10,000 × 3 is formed as input data of model simulation.
b) A total of 9,900 lines are selected in the input matrix randomly as training samples, and the BP NN model that does not use optimization algorithm and optimized maximum entropy model is acted on.
c) The remaining data of 100 lines act as verification data. These data are keyed into two models for fitting, and predicted value is obtained. The predicted value is compared with function value, and model prediction residual errors are obtained.

The distribution of the residual error of the predicted value in the two NN models is illustrated in Figure 2 with respect to the finally obtained 100 groups of data in the experiment.

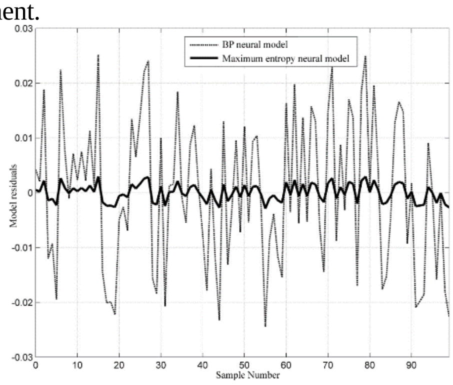

FIGURE 2 Fitted residual distribution of multivariate nonlinear model

Figure 2 indicates that the difference in fitted residual advantages and disadvantages between the two models is evident. For the maximum entropy model that uses the optimization algorithm, the absolute value of most fitted residual errors is lower than 0.003, whereas that of the non-optimized BP NN model is higher than 0.01. Difference in fitting precision of the two models is about one order of magnitude. Residual value of the maximum entropy model is considerably lower than that of the non-optimized BP NN model. Thus, the simulation experiment proves that the maximum entropy model that employs the optimization algorithm is clearly superior to the BP NN model that does not use the optimization algorithm.

3.1.2 Multiply linear function fitting

Usually, multivariate linear models are applied to practical problems except to multivariate nonlinear ones. In this study, we adopt a method similar to that in the preceding section. In detail, we use the BP NN model that does not utilize the optimization algorithm and a maximum entropy model after being optimized to perform fitting for the multivariate linear function $y = x_1 + x_2 + x_3$ and consider fitted residual value as a standard by which advantages and disadvantages of the two model algorithms are judged.

A similar NN model, load maximum entropy optimization algorithm, and traditional non-optimized optimization algorithm are applied. This model includes three input nodes and one output node. The number of nodes on the hidden layer is decided by performing cut-and-try testing. Sigmoid activation function is used for the hidden layer of the model, whereas linear activation function is adopted for its output layer. Moreover, fitting for target function is implemented. Thus, fitted value and residual distribution are obtained (Figure 3).

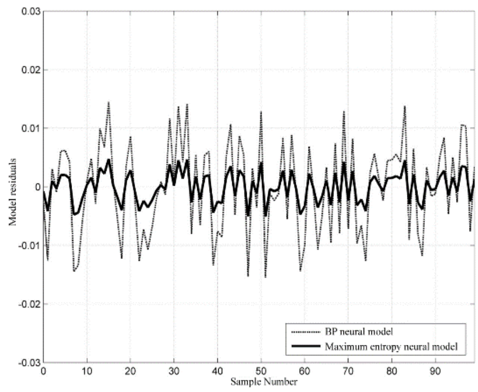

FIGURE 3 Fitted residual distribution of the multivariate linear model

Figure 3 indicates that the difference in advantages and disadvantages between the fitted residual error of the two models and the target function is similar to that of the nonlinear model. Moreover, most fitted residual values of the maximum entropy model adopting optimization algorithm are lower than those of the BP NN model that has not been optimized. However, the dissimilarity is that the difference in the fitting precision of the two models is less for the multivariate linear function than for the multivariate nonlinear function. The reason for this finding may be the simple expression of the multivariate linear function. In addition, the ability of this function to identify the advantages and disadvantages of model algorithms is not as efficient as the complicated multivariate nonlinear function.

3.2 EXAMPLES OF APPLICATIONS OF ALGORITHMS

We consult the methods of Sheng Feng et al. to verify the application effect of the maximum entropy model in practical problems. Moreover, we select 40 students from Class 1302 who are majoring in basketball at a physical culture college. Then, we evaluate each student based on ratings provided by teachers, combine our system with the evaluation system of experts, and establish a system of evaluation for basketball teaching of PE majors at colleges.

All kinds of indices for performance appraisal in the basketball teaching evaluation system are key qualities that basketball students must have. These indices are also important factors for teaching basketball as a major. The contents of these indices include sports skill, social adaptiveness, mental health, and sport consciousness. These indices can be further divided into 19 secondary indices. The indices can comprehensively reflect all aspects of student quality. The evaluation of teachers involves division of each secondary index into five evaluation grades (i.e., 5, 4, 3, 2, 1). Meanwhile, the evaluation obtained according to the performance of each index is used as original input data samples. Normalization processing must be performed on the original data after these original data have been obtained from teacher evaluation. The formula for the normalization algorithm is as follows:

$$X = (X_i - X_{\min}) / (X_{\max} - X_{\min}), \quad (9)$$

where X is normalization coefficient of the i th index data, and $X_{\max}$ and $X_{\min}$ represent maximum and minimum values in the data set X_i.

In addition, the results obtained after the expert group has examined the 40 students in the experiment are used as expected output values of the NN model. The input data of the model after normalization are obtained by summarizing previous work. Some input data of the model are shown in Table 1.

TABLE 1 Some input data of NN model after normalization

Sample	Evaluation index							Evaluation objective
	X_1	X_2	X_3	$X_4 \ldots X_{16}$	X_{17}	X_{18}	X_{19}	
1	5	5	4	…	3	4	5	88
2	4	5	3	…	5	2	3	79
3	2	5	3	…	2	4	3	62
4	5	2	3	…	4	3	3	70
5	5	5	5	…	4	5	4	92
6	4	4	3	…	5	2	3	85
7	3	4	2	…	5	3	3	72
8	1	5	3	…	2	4	2	60
…	…	…	…	…	…	…	…	…
32	3	4	3	…	4	4	4	80
33	3	3	3	…	5	2	1	69
34	2	3	5	…	3	3	3	64
35	4	4	4	…	4	4	2	82
36	4	5	5	…	5	2	1	80
37	4	5	5	…	5	5	3	87
38	4	4	2	…	2	5	5	78
39	5	4	3	…	5	5	5	91
40	5	5	2	…	2	5	2	78

In this study, the BP NN model that has not been optimized and the maximum entropy NN model are used to establish teaching evaluation models for the 40 students, respectively. Details are presented as follows:

a) With respect to evaluation indices, the number of nodes on the input layer and output layer is 19 and 1, respectively. In addition, cut-and-try testing is performed to decide the number of nodes on the hidden layer.
b) A total of 35 sets of data are randomly selected as training samples. The BP NN model that has not been optimized and the maximum entropy NN model are trained, and teaching evaluation models are established.
c) The remaining five sets of samples are selected as test samples. Two models are used for prediction, and the predicted value and residual error of the five sets of samples are obtained.

Relative errors between predicted rating results of the five sets of samples and prediction of models are shown in Table 2 in accordance with preceding samples. Distribution of predicted values of models and target values are illustrated in Figure 4.

TABLE 2 Predicted results of tested samples

Sample		36	37	38	39	40
Target value		80	87	78	91	78
Predicted value	Maximum entropy	79	88	78	90	79
	BP	75	83	84	92	75
Relative error (%)	Maximum entropy	−1.25	1.15	0.00	−1.10	1.28
	BP	−6.25	−4.60	7.69	1.10	−3.85

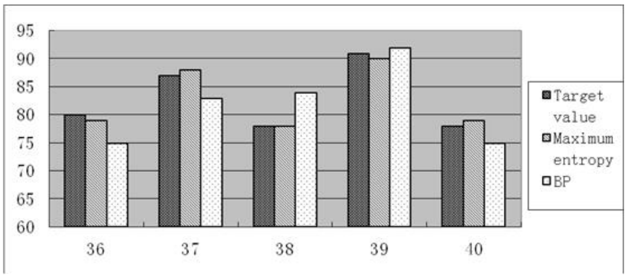

FIGURE 4 Distribution of predicted values of models and target values

A number of sample data are needed at the training stage of the NN models. The data size of input samples obtained by the study is negligible because of limited research conditions and approaches by which samples are obtained. This result may restrict the prediction accuracy of selected models. However, the difference in robustness and universality of model algorithms is extensively shown under a rigorous objective condition. This experiment is feasible considering that the study emphasizes evaluating the advantages and disadvantages of algorithms.

According to Table 2 and Figure 4, the maximum absolute value of the relative error between the predicted value of the maximum entropy network model and the result examined by experts is 1.28%, whereas the maximum absolute value of the relative error for the BP NN model that has not been optimized reaches 7.69%. The difference between the two maximum absolute values produces an order of magnitudes. We may detect relative errors of the BP NN model that are larger than those of the maximum entropy network model when a general survey of the entire data set is performed. However, the relative error of the two models in the 39th set of data is equal. The difference between the predicted results of the maximum entropy network model and the results obtained by experts is not significant. In addition, the overall relative error of the maximum entropy network model is relatively less than that of the BP NN model that has not been optimized. This finding indicates that the evaluation model established by maximum entropy network model has high prediction accuracy. Moreover, this evaluation model is more appropriate for basketball teaching evaluation than the BP NN model that has not been optimized.

4 Conclusion

This thesis investigates the application of the maximum entropy model to the evaluation model of basketball teaching quality and uses the BP NN model that has not been optimized as a control sample to examine the practical appellation effect of the optimization algorithm. Therefore, evaluation results of the maximum entropy NN model are better than those of the non-optimized BP NN model in both experimental function simulation and example verification. This finding implies that the optimized model possesses a strong generalization ability and degree of confidence. This optimization algorithm is feasible in establishing evaluation models related to basketball teaching quality.

According to the fitted results of multivariate linear function and nonlinear function, the difference between the maximum entropy NN model and non-optimized BP NN model is minimal in linear problems. However, the evaluation results of the former are better than those of the latter. Therefore, the degree of maximum entropy NN model dependency on and the identification of advantages and disadvantages of algorithms are not high for simple linear problems.

References

[1] Ozan Kocadağlı, Barış Aşıkgil. 2014 Nonlinear time series forecasting with Bayesian neural networks. *Expert Systems with Applications*, **41**(15), 6596-6610

[2] Danilo Costarell. 2014 Interpolation by neural network operators activated by ramp functions. *Journal of Mathematical Analysis and Applications*, **419**(1), 574-582

[3] Ailong Wu, Zhigang Zeng. 2015 Improved conditions for global exponential stability of a general class of memristive neural networks. *Communications in Nonlinear Science and Numerical Simulation*, **20**(3), 975-985

[4] Sunita Panda, Archana Sarangi, Siba Prasada Panigrahi. 2014 A new training strategy for neural network using shuffled frog-leaping algorithm and application to channel equalization. *AEU - International Journal of Electronics and Communications*, **68**(11), 1031-1036

[5] *Deleted by CMNT Editor*

[6] Harte J, Newman E A 2014 Maximum information entropy: a foundation for ecological theory *Trends in Ecology & Evolution* **29**(7) 384-389

[7] R.A.W. Bradford. 2013 An investigation into the maximum entropy production principle in chaotic Rayleigh–Bénard convection. *Physica A: Statistical Mechanics and its Applications*, **392**(24), 6273-6283

[8] A. Coluccia, A. D'Alconzo, F. Ricciato. 2013 Distribution-based anomaly detection via generalized likelihood ratio test: A general Maximum Entropy approach. *Computer Networks*, **57**(17), 3446-3462

[9] Richard F. Bieker. 2014 Does AACSB accreditation provide quality assurance and foster quality improvement for limited resource business schools whose missions are primarily teaching?. *The International Journal of Management Education*, **12**(3), 283-292

[10] *Deleted by CMNT Editor*

[11] *Deleted by CMNT Editor*

[12] Natasha R. Johnson, Jie Chen. 2006 Medical student evaluation of teaching quality between obstetrics and gynecology residents and faculty as clinical preceptors in ambulatory gynecology. American *Journal of Obstetrics and Gynecology*, **195**(5), 1479-1483

Author

Xiaodong Zhang, 1971.2, Shandong Province, P.R. China

Current position, grades: Associate Professor of Department of sports training and competition, Shandong Sport University, China.
University studies: received his Master Degree from Shandong Sport University, China.
Scientific interest: His research interest fields include exercise training and physical education of basketball.
Publications: more than 8 papers published in various journals.
Experience: He has teaching and research experiences of 27 years, has completed 5 scientific research projects.