ISSN : 0974 - 7435 Volume 10 Issue 15

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(15), 2014 [8370-8378]

The study of cancer patients hospital costs based on principal component analysis and BP neural network combination model

Wu Jianhui[1], Xue Ling[2], Hu Bo[1], Yin Sufeng[1], Wang Guoli[1]*

[1]Hebei Province Key Laboratory of Occupational Health and safety for Coal Industry, Division of Epidemiology and Health Statistics, School of Public Health, Hebei United University, Tangshan, (CHINA)

[2] Hebei Province Key Laboratory of Occupational Health and safety for Coal Industry, School of Public Health, Hebei United University, Tang Shan, (CHINA)

ABSTRACT

To collect totally 2340 cases of patients' hospital costs and related information in June 2009-March 2011 in a 3A-grade hospital's Surgical Oncology of Tangshan City. Gender, age, occupation, marital status, number of admission, admission illness, payment methods, surgical cases, secondary diagnosis, length of stay and treatment outcome are reduced dimensionality and denoising by Principal component analysis, BP neural network model was built between the selected principal component score matrix which is as input variables and hospital costs which is as output variables, and on the basis of the built model, the factors of hospital costs were analyzed by sensitivity analysis. The results showed that 8 Principal components were selected, and the cumulative contribution rate reached to 82.48%, Using a Bayesian algorithm, optimal BP neural network model was built basing on that the number of hidden layer neurons is 5, sensitivity analysis results showed that the top three influence factors on the costs of hospitalization were age, the number of days in hospital and treatment results. By this study, it was founded that using principal component analysis and BP neural network model to analyze the influence factors on the costs of hospitalization is feasible, and the hospital costs may be controlled by improving hospital efficiency, strengthening medical quality management and shortening the number of days in hospital appropriately.

KEYWORDS

Principal component analysis; BP neural network; Combined model; Hospital costs; Sensitivity analysis.

BACKGROUND

In recent years, total health expenditure around the world all showed rapid growth trend in varying degrees, especially in medical expenses, even it has exceeded the growth of gross domestic product and price index[1]. On the one hand, the growth of medical expenses increases the burden of State financial expenditure. On the other hand, the rising proportion of personal commitment increased the burden on families[2]. Therefore, it causes a series of problems about "seeing a doctor expensively, seeing a doctor difficultly" and so on. Otherwise, if the excessive growth of medical expenses isn't controlled efficiently, it will be possible that increasingly tense doctor-patient relationship continues to deteriorate[3]. To a certain extent, tense doctor-patient relationship can be eased by scientifically researching medical expenses and reasonably controlling the growth. In the form of medical expenses, hospital costs occupy a large proportion, and it constitutes main part of total health expenditure in a lot of countries. Therefore, it has become one of the popular issues which researchers concern about. How to control the excessive growth of medical expenses and how to ensure the quality of medical care have become urgent priority. Analysis of influencing factors about hospital costs is the key of its study. Therefore, and using the method which uses scientific theory as a guide to find the right idea and method of statistical analysis to analyze the relationship between hospital costs and the factors has become people's concerning issues that people want to solve them urgently. It has important practical significance for controlling unreasonable growth of medical expenses and using health resources effectively.

Foreign and domestic scholars have finished a lot of research on hospital costs and the factors, and some methods were used such as multiple linear regression, LOGISTIC regression analysis, Analysis of variance, Stepwise regression analysis and so on[4-8]. But data were required to satisfy linear, normality, independence, homogeneity of variance in most above methods. Otherwise, the information about hospital costs generally presents skewed distribution, many factors and complex nonlinear relationship, and it maybe exist multicollinearity among the factors. For the type of data and its distribution, there isn't a special requirement in the BP neural network. And it has powerful fault tolerance and self-learning ability. It can achieve complex nonlinear relationship between input and output variables. Therefore, BP neural network is applied on the research of hospital costs gradually. But, for the neural network, if input variables are excessive, it can make the network structure complex and the learning speed drop. And if input variables is insufficient, it couldn't reach the predict precision that is required[9]. At the same time, because the Selection process for the variables brinks subjectivity to some extent, it is high likely that there is a low correlation between Selected variables and output. And to some extent, it increase the possibility that network relapses into local minimum point, and decrease predictable function of network. Only in the way that a suitable set of variables is searched in many factors and it is used as input variable, the results can explain the relationship between hospital costs and the factors reasonably and accurately[10]. The factors of hospital costs are numerous.

And there exists correlation and multicollinearity among the factors. This Collinearity can have a greater impact on the analysis effect sometimes; even make the analysis out of work. Therefore during the research process, how to solve multicollinearity among factors is very necessary.

Principal component analysis is as one of the methods which can eliminate multicollinearity and achieve dimensionality reduction for multivariate, and using the thought of dimensionality, it starts from the relationships among various indicators to search the method that uses few independent integrated indicators to generalize the original index information. At the same time, it can also solve the problem of multicollinearity among factors. If principal component analysis and BP neutral network are combined to research on hospital costs, it deserves a discussion whether the combine can make up the insufficient aspects of data structure and factors and it can make the research results more objective and scientific if BP neutral network is improved and optimized.

THE PRINCIPLE OF PRINCIPAL COMPONENT ANALYSIS AND BP NEURAL NETWORK COMBINED MODEL

Establishment of single BP neural network model

One typical BP neural network is made up of input layer, output layer and hidden layer. And the hidden layer may be one layered or multiple layered. In this article, one network with 3 layers is used to analyze the hospital charges.

The basic structure of BP neural network is shown below.

The basic principle of BP neural network

BP (Back Propagation) neural network is the study process of error back propagation algorithm which is made up of information forward propagation and error back propagation[11]. The neurons in input layer are responsible for receiving the input information from outside and then deliver the information to the neurons in the hidden layer. The hidden layer is the inner information processing layer which is responsible for information conversion. The information delivered to the output layer from hidden layer can finish the forward propagation process once after the further processing. The information processing result will be output by the output layer. When the signals are forward propagated, the weight of the network will not be changed. The output of the neurons in each layer only affects the status of the neurons in the next layer. When the actual output does not correspond to the prediction, the error back propagation will be started. The errors will go through the

output layer, correct the values of weight according to the gradient descent and back propagate to the hidden layer and output layer. In the process of the error back propagation, the values of weight are continuously corrected and adjusted, shortening the distance between the practical output and the desired output. This is the network output training process. It will last till the errors output from the network are reduced to the acceptable level or the set training times are reached. Compared with the traditional statistic method, BP neural network has no hypothesis requirement on the data. The strong function approximation ability makes the network is better than the traditional statistic method.

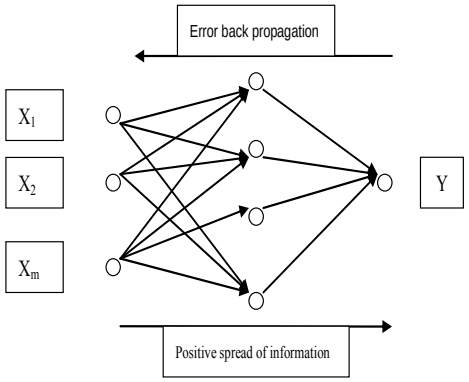

Figure 1 : The structure of BP neural network

The steps of BP neural network modeling

Data normalization

In practice, by normalizing the output and input variables, the network training will be more effective. The training speed of the network and the function of the network will be improved. The detailed normalization method will be diversified according to the data, which generally are orthonormality, change of scale and standardization. The change of scale means that after being added or deducted a constant the changed variable will be multiplied or divided by a constant. It mainly is used to change the unit of the data. The orthonormality will implement the normal conversion with the typical value 0 and the standard deviation 1 or normalize it within the scope of [-1, 1]. Standardization means we directly compress the data within the scope of [0, 1]. The formulas are as below,

Orthonormality : $S_i = \dfrac{x_i - mean}{std}$

Adjusting the normalization : $S_i = \dfrac{x_i - (\max(x) + \min(x))/2}{(\max(x) - \min(x))/2}$

Standardization : $S_i = \dfrac{x_i - \min(x)}{\max(x) - \min(x)}$

When classifying the variables, we can adjust formulas of the orthonormality and standardization. To continue the variables, we can use orthonormality and standardization formulas. Because the input and output variables are of positive values, the values of selected variables by the standardization formula are within the scope of [0,1] considering the selection rules of transmission functions and data distribution. After the network simulation, the anti-normalization is used to convert the simulation result to original data.

The selection of the transfer function

BP neural network needs differentiable non-linear functions, such as S-functions: logarithm S-function and hyperbolic tangent function. We don't have definite standard of choosing S-functions. Generally speaking, if the values of input variables are more than 0, we'd better use logarithm S-function. If the values of the output variables are less than 0, we'd better use hyperbolic tangent function. For the neurons using S-functions, the input data should be normalized to [0, 1] or [l, 1]. So the saturation region of the function and the slow speed of the network convergence are avoided. Because the variable values are all more than 0 in this research, we use S-function which is logsig function.

The partition of data sets

The partition of data sets is very important. After the study of the training set, we establish a model with good promotion function. The training sets should be well distributed, which are good representatives of the sample set. The training set and verification set are applied in the network training and model prediction. The test set is used to trace the mistakes in the training to avoid over fitting.

We adopt different strategies to improve the promotion abilities of the network. They have different requirements on the data set. If we use early stop strategy, the data sets should be divided into training sets (60%), verifying sets (20%) and test sets (20%). If we use Bayes rules BP algorithm, the data sets will be divided into training sets (80%) and test sets (20%).

Initialized weights and thresholds

When initializing the network, to initialize the weights and threshold correctly is very important. If the initialization is not proper, the training time may be longer and even the network can not be converged. If the initial value is large, the weighted value may be within the saturation region of the transfer function which results in the small descent gradient. It will lead to the stop of the training of the network. Generally, we hope that the initialized weighted values of the neurons are close to zero so that the weights of the neurons can be adjusted in the most sensitive region of change in the S-function. Therefore, the initial values of weight and threshold are randomly selected within [-1,1].

The speed of initialized learning

The learning speed can determine the changes of the weights in the process of network training. When the learning speed is slow, the change margin of the weights is small, which will lead to the slow network training and difficult convergence of the network. While the learning speed is too fast, the system will be unsteady and the weight value and error function will be different, which means the sum of squared errors of the network function can not reach a suitable value. At present, the adopted BP algorithm is not a standard one which is an improved BP algorithm, such as LM algorithm, conjugate gradient algorithm and etc. These algorithms can quicken the speed of network convergence by improving the standard BP algorithm.

The evaluation indexes of network function

For the approximation of the function, the network function evaluation can be influenced by the goodness of fit between the value of fitting of the model and the measured value. It also can be reflect by sum of squared errors (SSE), coefficient of determination (R^2), adjusted coefficient of determination (adjusted R^2), root mean square error (RMSE), mean square error (MSE) and other indexes. In this article, the coefficient of determination (R^2) is selected to evaluate the network function.

The selection of training algorithm

When using BP neural network to implement the function approximation, the convergence speed of LM algorithm (regarded as trainlm in MATLAB) of early stop strategy is the fastest and the prediction is very accurate. BR algorithm (regarded as trainscg in MATALB) of Bayes rules is one method to increase the prediction accuracy of the network. Mutative scale conjugate gradient algorithm (regarded as trainbr in MATLAB) is comparatively good which performs well in prediction and function approximation. Quasi-Newton algorithm (regarded as trainoss in MATLAB) is faster in the convergence than the mutative scale conjugate gradient algorithm.

The set of number of layers of the neural network and neurons in each layer

The determination of number of layers in the network: one 3-layer neural network can implement the approximation of any nonlinear function with the desired accuracy. Generally, one-layer or two-layer hidden layer is enough to solve the practical problems. The adding of the layers will make the network complicated, which will influence the convergence speed of the network.

The determination of the number of neurons of each layer: the number of the neurons in the input and output layers should be determined according to the practice. They are just like the independent variables and dependent variables. At present, there is no definite formula to determine the number of the neurons in the hidden layer. If we calculate by the experienced formula, the result may be quite different. When using trial-and-error method, m refers to the number of neurons in input layer; n refers to the number of the neurons in the hidden layer; we have $n=\log 2^m$. If the convergence is not ideal, the number of the neurons will be added. If the errors can't be reduced and the convergence speed is slow, we should consider the stop the adding; or we can suppose $n=2m+1$ according to Komlogorov theorem.

The principal component analysis

The basic principle of principal component analysis

Suppose there are p indexes X_1, X_2, …, X_p. We should find the independent comprehensive indexes Z_1, Z_2, Z_p, which can summarize the information of p indexes. From the perspective of mathematics, we should find a group of constants a_{i1}, a_{i2} …, a_{ip} (i=1, 2…, p) and linear combine p indexes.

$$\begin{cases} Z_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\ Z_2 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{2p}X_p \\ \vdots \\ Z_p = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p \end{cases}$$

In order to summarize the main information of p original indexes X_1, X_2…, Xp, we introduce the following matrix,

$$Z=\begin{Bmatrix}Z_1\\Z_2\\\vdots\\Z_P\end{Bmatrix}\ A=\begin{Bmatrix}a_{11} & a_{12} & \cdots a_{1p}\\ a_{21} & a_{22} & \cdots a_{2p}\\ \vdots & & \\ a_{p1} & a_{p12} & \cdots a_{pp}\end{Bmatrix}\underline{\Delta}\equiv\begin{pmatrix}a_1^{'}\\a_2^{'}\\\vdots\\a_p^{'}\end{pmatrix}\ X=\begin{pmatrix}X_1\\X_2\\\vdots\\X_p\end{pmatrix}$$

Formula （1-1） can be illustrated as

Z=AX

$$\text{Or}\begin{cases}Z_1=a_1^{'}X\\Z_2=a_2^{'}X\\\vdots\\Z_p=a_p^{'}X\end{cases}$$

If $Z_1=a_1^{'}X$ satisfies $a_1^{'}a_1=1$ and $Var(Z_1)=\underset{a_1^{'}a_1=1}{Max}\{Var(a^{'}X)\}$, Z_1 is the first principal component of original indexes X_1, X_2…,X_p. When $Z_i\neq Z_j$, Z_i and Z_j have no relation and Z_1 is the maximum variance of all the linear combinations of X_1, X2…, Xp. And Z_2 ranks the second. The rest may be deduced by analogy.

The basic steps of principal component analysis

1) The orthonormality of original data. We use Z-score method to change the formula into

$$z_{ij}=\frac{x_{ij}-\bar{x}_j}{s_j}$$

In the formula, $\bar{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij}$

$$s_j=[\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2]\ \ \text{i=1, 2…, n j=1, 2…, p}$$

After the transmission, the typical value is zero and the variance is one.

The reason of orthonormality of the original data: when solving the principal component by the related coefficient matrix R, we usually pay more attention to the larger variable of variance δ_j^2, which means that it will be influenced by the measurement scale of the variables and sometimes we will get unreasonable result. In order to illustrate the connotation of the principal component more objectively, we must normalize the original data and eliminate the impacts of the unit of measurements and order of magnitude.

2) Solving the related matrix R of indexes

The variable related matrix R is the starting point of principal component analysis. The measurement formula is

$$r_{ik}=\frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_{ij}-\bar{x}_i)}{s_j}\frac{(x_{ik}-\bar{x}_k)}{s_k}$$

$$\text{Or}\ r_{ik}=\frac{1}{n-1}\sum_{i=1}^{n}Z_{ij}Z_{jk}$$

And $R_{ii}=1$ $r_{ik}=r_{kj}$

3) Solving the latent root, eigenvector and contribution rate of matrix R

The secular equation of R is $\left|\lambda \mathrm{I}_{\mathrm{P}} - R\right|$=0; λ_g (g=1, 2…, P) is the latent root of the equation and the variance of the principal component Z; the amount shows the information ability of the principal component comprehensive original indexes. We use L to represent the dimensional real vector. The vector $\mathrm{L_g}$ obtained from equation $\left[\lambda_g \mathrm{I}_{\mathrm{P}} - R\right]L_g$ =0 is the corresponding eigenvector of the latent root λ_g , which is the sub-vector coefficients in the coordinate system of standardized vector $Z_j\left[Z_i = \begin{pmatrix} Z_{1j} \\ \vdots \\ Z_{pj} \end{pmatrix}\right]$. $\mathrm{a_g}$= λ_g / $\sum_{g=1}^{p} \lambda_g$ shows the each principal component can reflect the information amount of the original variable. That is the variance contribution rate.

4) The selection of amount of the principal components

Theoretically speaking, the maximum number of the principal components equals to the number of the original variables, which can reflect all the information provided by the all the original indexes. Because the aim of the analysis is to use less comprehensive indexes to reflect the main information of all the original indexes, the total number of the principal components is less than the number of the original indexes. There are a lot of principles in determining the number of components, which are shown below,

Cumulative contribution rate guideline

How many principal components will be kept is determined by the percentage of the cumulative variance in the sum of the variances (cumulative contribution rate). It shows the how much information the previous principal components have. Generally when the cumulative contribution rate of k principal components reaches 80%, k principal components[16] should be kept.

Latent root guideline

First the typical value $\overline{\lambda}$ of latent root λ_g should be calculated. Select k sub-vectors which are $\lambda_g > \overline{\lambda}$ as principal sub-vectors. We can get $\overline{\lambda}$ =1 from the standardized data related matrix R. We needn't do any calculations and just select forward k principal sub-vectors which λ_g>l.

The final determination of principal components

The principal components got from cumulative contribution rate are always a lot, while the components got from latent root method are rare. So we will consider the combination of cumulative contribution rate and latent root to determine the final principal components and considering the professional significance of variables to determine if necessary.

The establishment of the combination model of principal components analysis and BP neural network

First we should implement the principal components analysis to all the prediction factors and save the selected principal component score matrix[12].

Sensitivity analysis

Sensitivity analysis refers to that after building BP neutral network model, it is used to observe the corresponding changes in network output variables by changing a part of network input variables, and it can decide the importance of this variable in output prediction in this way. Changing each recorded variables in the sample one by one is as the implementation process. For the categorical variables, it will be tested for all possible combinations of classes which is the value after normalizing in general. For the continuous variables, if value is in the range of [0,1], these values including o,0.25,0.5,0.75 and 1 are used to cut the range of values into quartered. When the value changes, The maximum and minimum output will be recorded, and the proportion of the maximum value and minimum value of the output value relative to the ratio of the maximum is calculated, and in the end, all of the averages recording this proportion is the sensitivity of the variable which indicates the factor influence for the hospital costs.

THE ANALYSIS OF HOSPITAL COSTS INFLUENCING FACTORS IN CANCER PATIENTS

According to the Hospital Information System in Tangshan City, a 3A-grade hospital and International Classification of Diseases ICD10 coding, totally 2340 cases of patients' hospital costs and related information including number of admission (x1), age (x2), length of stay (x3), gender (x4), marital status (x5), surgical cases (x6), secondary diagnosis (x7), admission illness (x8), treatment outcome (x9), payment methods (x10), occupation (x11) and so on.

Principal component analysis of factors for the hospital costs

When the principal component analysis is used to research, KMO and Bartlett's sphericity test must be adopted to fit test before factor analysis and to analyze the suitability of principal component analysis. KMO is an index that is used to observe correlation coefficient value and partial correlation coefficient value. And if the value of KMO is larger and, the correlation of factor analysis is stronger, KMO is more suitable to carry on factor analysis The results about KMO and Bartlett's sphericity test are showed in TABLE 1.

TABLE 1 : KMO and Bartlett's Test

Test methods	Statistics	*P*
KMO	0.581	
Bartlett	1515.517	0.000

The results in TABLE 1 showed that the value of KMO is 0.581, which is larger than 0.5, and the test value of Bartlett's is 1515.517, $P<0.05$,which indicates that there exists multicollinearity among data, and it is suitable to adopt principal component analysis for processing.

Principal component analysis is adopted to analyze the factors for the hospital costs. The results are showed in TABLE 2.

TABLE 2 : The results of principal component analysis

Principal component	Characteristic roots	Cumulative contribution rate
Z_1	1.86291676	0.1694
Z_2	1.33563879	0.2908
Z_3	1.22350943	0.4020
Z_4	1.14249897	0.5059
Z_5	1.01961435	0.5986
Z_6	0.91989590	0.6822
Z_7	0.80578298	0.7554
Z_8	0.80254847	0.8284
Z_9	0.67013812	0.8893
Z_{10}	0.64012858	0.9475
Z_{11}	0.57732768	1.0000

If the absolute value of loading factor is larger, the correlation between principal component and loading factor is larger. After analysis for twiddle factor loading matrix, it is founded that the first principal represents the main information of age, length of stay and treatment outcome, and the second principal component represents the main information of whether surgery, and the third principal component represents the main information of number of admission, and the forth principal component represents the main information of marital status, and the fifth principal component represents the main information of payment methods, and the sixth principal component represents the main information of gender and secondary diagnosis, and the seventh principal component represents the main information of admission illness, and the eighth principal component represents the main information of occupation.

According to considering for the characteristic roots and cumulative contribution rate, the previous 8 principal components are selected, and the expression equation of principal component is inferred according to the factor loading matrix. The equation is as follows:

$$Z_1=0.3306X_1+0.5951X_2+0.4896X_3-0.2915X_4-0.2807X_5-0.4673X_6+0.5218X_7-0.1536X_8+0.4409X_9+0.1995X_{10}+0.4977X_{11}$$

$$Z_2=0.3111X_1+0.3550X_2-0.4461X_3+0.0344X_4-0.4419X_5+0.5038X_6+0.1539X_7+0.3654X_8+0.3233X_9-0.1594X_{10}+0.3162X_{11}$$

$$Z_3=0.5875X_1-0.4010X_2+0.2495X_3+0.5495X_4+0.1401X_5-0.1855X_6-0.1471X_7+0.3817X_8+0.3578X_9-0.0455X_{10}+0.0437X_{11}$$

$$Z_4=-0.2902X_1+0.1047X_2+0.3248X_3+0.3034X_4-0.4835X_5+0.0919X_6-0.1626X_7+0.4643X_8-0.4374X_9+0.2290\ X_{10}+0.3491\ X_{11}$$

$$Z_5=0.1889X_1-0.1553X_2-0.1976X_3-0.2045X_4+0.1889X_5+0.2320X_6-0.0886X_7+0.0235X_8+0.0408X_9+0.8649\ X_{10}+0.1771X_{11}$$

$$Z_6=0.0085X_1-0.0317X_2+0.0063X_3+0.5692X_4+0.0614X_5+0.2819X_6+0.5752X_7+0.0235X_8+0.0408X_9+0.8649X_{10}+0.1771X_{11}$$

$$Z_7=0.0140X_1+0.0492X_2-0.0598X_3-0.1814X_4+0.4159X_5-0.0175X_6+0.4656X_7+0.5554X_8-0.2379X_9-0.1007X_{10}+0.0370X_{11}$$

$$Z_8=0.2950X_1+0.0432X_2-0.2080X_3-0.0605X_4+0.0728X_5+0.2763X_6-0.1725X_7-0.1218X_8-0.1454X_9-0.3455X_{10}+0.6325X_{11}$$

Training results of BP neutral network

Principal component score matrix selected is as input variable,and the total cost of hospitalization is as output variable. Then BP neural network model based on principal component analysis is built.

Four different algorithms including LM, BR, OSS and SCG and four different hidden layer neurons network including 5,10,15 and 25 are carried on random testing respectively. And in the end, the best of BR algorithm and 5 hidden layer neurons are selected to build neutral network model. The testing results are showed in Figure 2, and the parameters of the index about the model are showed in TABLE 3.

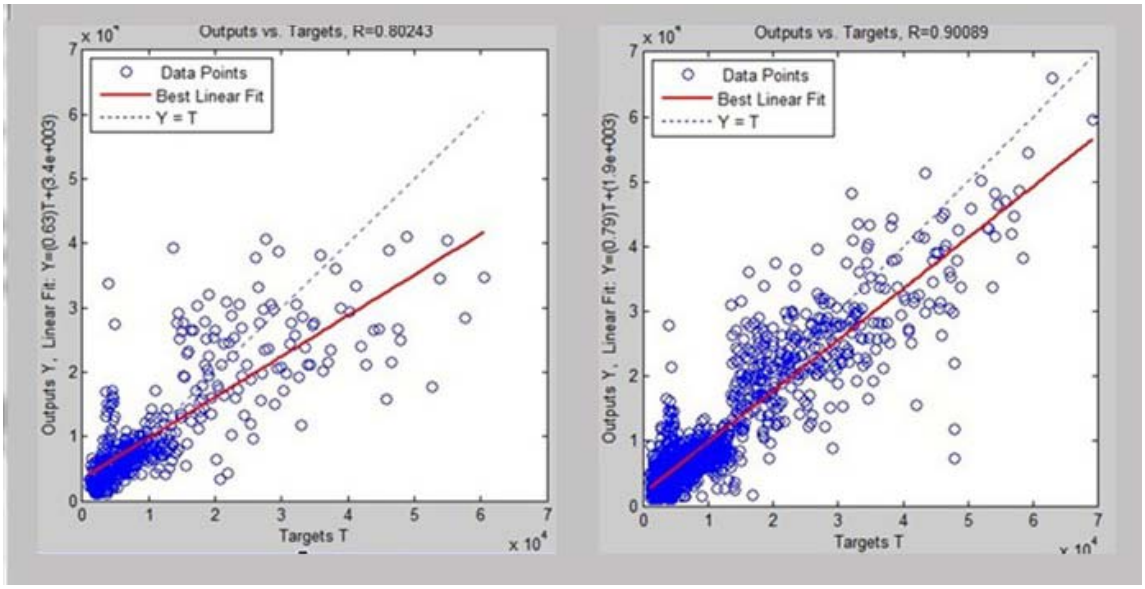

Figure 2 : training fitting figure of BP neutral network

TABLE 3 : BP neural network model parameter table about hospital costs of surgical oncology patients

Network structure parameters	Network training parameters	Test set simulation results	Training set fitting results
Hidden layers: 1 layer	Training algorithms: BR algorithm	R=0.8269	R=0.8931
Hidden layer neurons: 15	Total iteration stop training Views: 103	R^2=0.6838	R^2=0.7976
input layer neurons: 8	Learning speed: 0.01	RMSE=6422.11	RMSE=6246.53
Output layer neutrons:1	Performance function: SSE	SSE=4.6940e+10	SSE=3.0314e+10
	Stop training SSE = 3.2246	MSE=4.1243e+07	MSE=3.9019e+007

***SEE In terms of the data which is normalized. RMSE In terms of the data which is anti-normalized.**

The sensitivity analysis results of influence factors about hospital costs

The sensitivity analysis is adopted to analyze the sensitivity of the influence factors about hospital costs. And in this way, the influence degree of Each factor on hospital costs is reflected. The results is showed in TABLE 4.

TABLE 4 : The ordering of influence degree on influence factor about hospital costs

The ordering of influence degree	Influence factors	sensitivity
1	Z_1	3.2909
2	Z_2	2.4050
3	Z_3	2.1783
4	Z_5	1.8208
5	Z_7	1.2912
6	Z_8	1.1887
7	Z_6	1.0046
8	Z_4	0.9541

Sensitivity analysis shows that the descending order of each main component's sensitivity is as follows:Z_1(3.2909), Z_2(2.4050), Z_3(2.1783), Z_5(1.8208), Z_7(1.2912), Z_4(1.1887), Z_6(1.0046), Z_8(0.9541). After the data is transformed into the influence factor information which is contained, the descending order is as follows: 1. age, length of stay and treatment outcome, 2. surgical cases, 3. number of admission, 4. payment methods, 5. admission condition, 6. occupation, 7. gender and secondary diagnosis, 8. marital status.

CONCLUSION

After analyzing the influence factor about hospital costs in the method of sensitivity analysis,it is founded that the main influence factors including age, length of stay and treatment outcome is the same as many related literature. And it illustrates that using principal component analysis and BP neutral network model to analyze the influence factor of hospital costs is feasible. At the same time, there is no special requests for the data in this model, and it can reduce the data in dimensionality. It can solve the multicollinearity among the factors. It can optimize the neutral network structure and improve fitting precision of the model. Therefore, this method has the larger scope of application, and it deserves to sue widely in the analysis of multifactor.

ACKNOWLEDGMENTS

This work is supported by the accented term of Health Department of Hebei Province (20130055).

REFRENCES

[1] D.F.Seott, R.L.Grayson, E.A.Metz; "Disease and illness in U.S. mines 1983-2001", Occupation and Environmental Medicine, **46**(12),1272-1277, **(2004)**.

[2] Susan Megison, Lynn Westbrook; "Hospital Dissemination of Information Resources on Intimate Partner Violence: Statewide Analysis of Texas Emergency Room Staff", Journal of Hospital Librarianship, **9**, 231-248, **(2009)**.

[3] T.John Schousboe, L.Misti Paudel, C.Brent Taylor, et al; "Estimation of Standardized Hospital Costs from Medicare Claims That Reflect Resource Requirements for Care: Impact for Cohort Studies Linked to Medicare Claims", Health Services Research, **49(3)**, 929–949, **(2014)**.

[4] A.Margaret Olsen, Sorawuth Chu-Ongsakul, E.Keith Brandt; "Hospital-Associated Costs Due to Surgical Site Infection after Breast Surgery", Arch Surg, **143**, 53-60, **(2008)**.

[5] Young Sun Rhee, Young Ho Yun, Sohee Park; "Depression in Family Caregivers of Cancer Patients: The Feeling of Burden as a Predictor of Depression", Journal of Clinical Oncology, **26**, 5890-5895, **(2008)**.

[6] Julie Ann Sosa, T.Charles Tuggle, S.Tracy Wang; "Clinical and Economic Outcomes of Thyroid and Parathyroid Surgery in Children", J Clin Endocrinol Metab, **93**, 3058–3065, **(2008)**.

[7] X.F.Song, L.Z.Xu, X.Z.Wang; "Analysis on Affecting Factors of Unnecessary Hospitalization Expenses for Two Simplex Diseases", Chinese Health Economics; **26**, 30-33, **(2011)**.

[8] H.Chen, J.G.Chen, X.C.Deng; "Research of Affecting Factors of Hospitalization Expenses for a Hospital Attached to a Military Medical University", Northwest Medical Education, **16**, 423-426, **(2011)**.

[9] E.Bartfay, W.J.Mackillop, J.L.Peter; "Comparing the Predictive Value of Neural Network Models to Logistic Regression Models on the Risk of Death for Small-Cell Lung Cancer Patients", Eur J Cancer Care (Engl), **15(2)**, 115-124, **(2006)**.

[10] S.Fatih Erol, Hadi Uysal, Ucman Ergiun, et al; "Prediction of Minor Head Injured Patients Using Logistic Regression and MLP Neural Network", Journal of Medical Systems, **29**, 205-215, **(2005)**.

[11] R.F.Yu; "Tinking about Medicine Profession and Medicine Profession", World Clinical Drugs, **28**, 198-203, **(2011)**.

[12] Bu-Qing Cao, Jian-Xun Liu, Bin Wen; "Currency Characteristic Extraction and Identification Research Based on PCA and BP Neural Network", JCIT, **7(2)**, 38-44, **(2012)**.