

ONE PREDICTION MODEL BASED ON BP NEURAL NETWORK FOR NEWCASTLE DISEASE

Hongbin Wang^{*}, Duqiang Gong, Jianhua Xiao, Ru Zhang, Lin Li
College of Veterinary Medicine, NorthEast Agricultural University, Harbin, Heilongjiang Province, P. R. China, 150030

^{*} *Corresponding author, Address: College of Veterinary Medicine, NorthEast Agricultural University, Harbin, 150030, Heilongjiang Province, P. R. China, Tel: +86-451-55191940, Fax: +86-451-55190470, Email: neau1940@yahoo.com.cn*

Abstract: The purpose of this paper is to investigate the correlation between meteorological factors and Newcastle disease incidence, and to determine the key factors that affect Newcastle disease. Having built BP neural network forecasting model by Matlab 7.0 software, we tested the performance of the model according to the coefficient of determination (R^2) and absolute values of the difference between predictive value and practical incidence. The result showed that 6 kinds of meteorological factors determined, and the model's coefficient of determination is 0.760, and the performance of the model is very good. Finally, we build Newcastle disease forecasting model, and apply BP neural network theory in animal disease forecasting research firstly.

Key words: Newcastle disease; forecasting model; BP neural network ; meteorological factors

1. MATERIALS AND METHODS

1.1 Data

One province in southwest of china was selected as object of this experiment. The meteorological factors data from 1999 to 2005 obtained

Please use the following format when citing this chapter:

Wang, H., Gong, D., Xiao, J., Zhang, R. and Li, L., 2009, in IFIP International Federation for Information Processing, Volume 294, *Computer and Computing Technologies in Agriculture II, Volume 2*, eds. D. Li, Z. Chunjiang, (Boston: Springer), pp. 1209–1216.

from China Meteorological Bureau. Epidemic situation data of ND in the same period obtained from Veterinary bulletin published by China Ministry of Agriculture. The cultivation scale of domestic fowl obtained from China Stock Farming Statistics Yearbook. Because the data of cultivation scale in China Stock Farming statistic by year, to unify the calculation method, the amount of domestic fowl for one year was treated as cultivation quantity for each month. The incidence rate of ND calculates as the follow formula:

Disease Incidence for each month of ND = *morbidity amount of ND / amount of domestic fowl by end year*

1.2 Principle for BP Neural Network

BP neural network contain one input layer (Greenough G, 2001), some imply layer and one output layer, the treatment unit for data in every layer be called treatment unit. The transmission procedure of information is unidirectional transmission to input layer. The inferent information be treat in input layer, imply layer, and be transmit to output layer. The status of every layer can only be affected by next layer. If none anticipant outcome be generate in output layer, then change to back-propagation, the error between outcome and expected value be return along origin path. By modify weigh of neuron in every layer, the error be reduce gradually. This circulation would not stop until all error in net be restrained to a defined value. To analyze data by neural network, a great mount of train must be done for adjust the weigh value among neurons. The study procedure was completed by continual adjustment of weight (Li Jing, 2006).

1.3 Methods

1.3.1 Associativity analysis between meteorological factors and incidence of ND

To set meteorological data of this month as independent variable and incidence rate of ND for next month as dependent variable, associativity analysis was done by SPSS11.0, so as to provide evidence for the building of prediction model.

1.3.2 The building of prediction model of ND based on BP neural network

The data for Output layer was incidence rate of ND for each month in experiment region. The sample data make its value range in $[-1, 1]$. The formula is $X' = (X - X_{\min}) / (X_{\max} - X_{\min}) \times 2 - 1$; The data for output layer was

incidence of ND, and its formula is $Y' = (Y - Y_{min}) / (Y_{max} - Y_{min}) \times 0.8 + 0.1$, output value range in [0.1, 0.9]. the result must be treated after train and fitting for model, so as to get actual data. The data that including temperature, humidity, wind speed, cloudage, time for sunshine and amount of precipitation obtained from October 1999 to may 2004 became input layer of network. And the incidence rate for ND of this region obtained from November 1999 to June 2004 became anticipant output layer of network. The accurate of this network can be validating by check data block. One implied layer was used in this neural network. The quantity scope of implied layer units determined by $h = \sqrt{N + M} + a$ was referred in this research. Among the total, M was set as the count of output units, N was the count of input units, a was constant range from 1 to 10^[9]. After the count of input layer units-N(meteorological factors) and count of output layer units – N have been determined, the BP neural network was train and adjusted by alter the count of implied layer units h which change by the value of constant a . The train function was trainlm, the activation function was sigmoid function: $F(x) = 1 / (1 + e^{-x})$. The structures of BP neural network see fig.1.

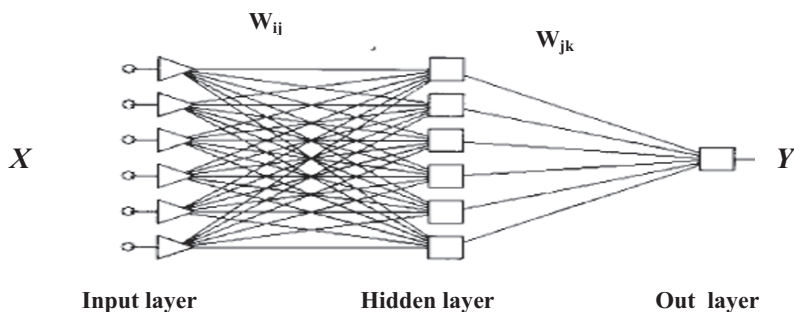


Fig.1 Structure of BP neural network Forecasting model

X means input unit(meteorological factor), Y means output unit(disease incidence); W_{ij} means the weight which connects the unit _{i} in the Input layer with the unit _{j} in the Hidden layer, W_{jk} means the weight which connects the unit _{j} in the hidden layer with the unit _{k} in the output layer, and all the connective weights can be adjusted according to Sigmoid training function.

1.3.3 The implement of BP neural network based on MATLAB7.0

1.3.3.1 The initialize of BP neural network

$\text{net}=\text{newff}(\text{minmax}(p),[m,n],\{\text{'logsig'},\text{'logsig'}\},\text{'trainlm'})$; $\text{minmax}(p)$ was the scope of input value p . before applied, the meteorological data p should be normalized, and change into $[-1,1]$. 3-layer network was adopted in this experiment, m was the count of implied units, n was the count of output units. Because output variable n was incidence of ND, $n=1$, logsig was activation function among layers in BP neural network.

1.3.3.2 The train of BP neural network

$\text{net}=\text{train}(\text{net},P,T)$; P was input vector, T was target vector. Base on the reverse transmission algorithm of error for network study, new weight of network and threshold value was obtained from last train. The min error of network was set as 1×10^{-5} , the max frequency of train was 5 000, other parameters was acquiesced by system. The end purpose of train was make the error between incidence rate simulated by network and expected output value T .

1.3.3.3 The prediction of neural network and evaluation of effects

$y=\text{sim}(\text{net},P)$; To predict the incidence of ND by network structure and input vector determined by train, and check the fitting degree between actual incidence of ND and y . Comprehensive evaluation of model can be made by error between actual value and prediction value. Coefficient of determination $R^2=(\text{corrcoef}(A,T))^2$, In numerical value, R^2 equal the square of coefficient correlation for actual incidence rate and predicted value, represent the percentage for independent variable predicted dependent variable exactly.

2. MATERIALS AND METHODS

2.1 The associatively between meteorological factors and incidence rate

By correlation analysis between incidence rate of ND each month and meteorological factors(tab.1),The incidence rate of ND positive correlated with medial humidity each month and cloudage each month significantly, and negative correlated with average wind speed each month significantly.

Tab.1 Correlation matrix for every meteorological factors and ND incidence

Coefficient correlation	Air Temperature (X1)	humidity (X2)	Cloudage (X ₃)	Wind speed (X ₄)	Precipitation (X5)	Time for sunshine (X6)	Incidence rate (Y)
air temperature (X1)	1	0.499**	0.870**	-0.173	0.822**	-0.605**	0.244
humidity (X2)	0.499**	1	0.773**	-0.849**	0.709**	-0.861**	0.431**
cloudage (X3)	0.870**	0.773**	1	-0.431**	0.922**	-0.884**	0.299*
wind speed (X4)	-0.173	-0.849**	-0.431**	1	-0.414**	0.609**	-0.485**
precipitation (X5)	0.822**	0.709**	0.922**	-0.414**	1	-0.778**	0.263
time for sunshine (X6)	-0.605**	-0.861**	-0.884**	0.609**	-0.778**	1	-0.253
incidence of a disease (Y)	0.244	0.431**	0.299*	-0.485**	0.263	-0.253	1

** means very significant difference ($P < 0.01$ at two tailed level) , * means significant difference ($P < 0.05$ at two tailed level)

2.2 The determination of BP neural network and the prediction result

By many times debugging, it was determined that 6 was the optimum count of implied layer in network. 6 units in input layer was 6 kinds of different meteorological factors, there are 6 units in implied layer, the output layer unit was incidence rate of ND, therefore, the network structure be determined as 6—6—1. The error reaches smallest: 1×10^{-5} after 68 times trains (see fig.2), the weigh between input layer and implied layers; the weigh value and threshold value between implied layers and output layers listed in table 2, the fitting for BP neural network training value and practical incidence see fig.3.

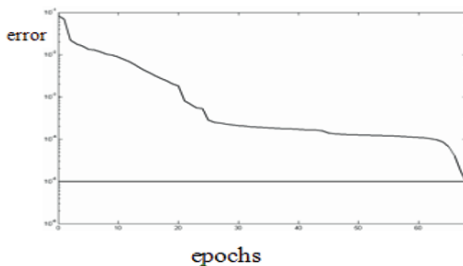


Fig.2 Training curve of BP neural network

Tab.2 Threshold and connective weight between each unit

Neuron layer	weight (W _{ij})	Shreshold
Input layer-imply layer	4.4712 -2.5107 -25.9255 31.0318 -12.6838 -24.1211	-7.3887
	-8.1328 4.2805 30.0656 12.2717 -5.3577 14.5849	-3.9163
	-1.8181 3.5385 13.1725 18.762 -16.603 2.02	6.8882
	8.1653 -16.1253 -12.2964 -9.2546 -9.9482 -34.8834	2.3161
	14.2829 16.0193 -10.8602 10.7641 1.8061 -14.7833	-25.2292
	-2.1933 3.6 -5.1103 -5.4201 5.1412 -3.4249	-6.1514
Imply layer-output layer	-1.1287 -1.2789 -10.4971 -1.2455 10.3956 -13.0465	10.7271

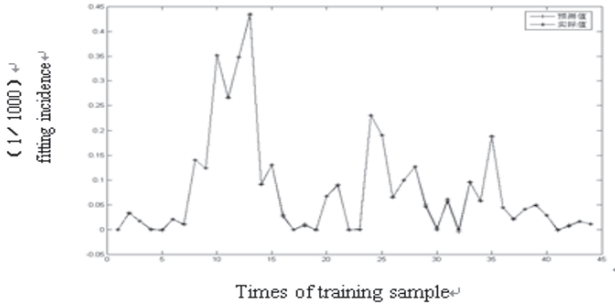


Fig.3 Fitting for BP neural network training value and practical incidence

After train of neural network by inlet input vector and expected output vector, the weigh value and threshold value among every layers can be obtained, then the value can be save and fixed, the result for train and fitting see tab. 3.

Tab.3 Value of BP neural network training value and practical incidence

Time (year.month)	Fitting value for train of BP artificial neuron network (1/1000,000)	Practical incidence rate of disease (1/1000,000)	Absolute value of error(1×10 ⁻⁶)
2003.09	187.5	187.8	0.3
2003.10	44.6	44.1	0.5
2003.11	21.6	20.7	0.9
2003.12	40.1	40.4	0.3
2004.01	49.0	49.1	0.1
2004.02	27.3	28.1	0.8
2004.03	1.1	0	1.1
2004.04	6.1	6.2	0.1
2004.05	16.2	16.1	0.1
2004.06	11.2	11.1	0.1

* Part of training data is listed in the table3

By input check data into neural network trained, the prediction result can be get after run Matlab7.0. The comparison between actual incidence and prediction was list in tab. 4.

Fig.4 Value of Absolute error and forecast and incidence

Time (year.month)	Predictive value for BP neuron artificial network (1/1000,000)	Practical incidence rate of disease (1/1000,000)	Absolute value of error (1×10 ⁻⁶)
2004.07	6.2	5.8	0.4
2004.08	5.3	4.4	0.9
2004.09	-2.4	0	2.4
2004.10	0.1	0	0.1
2004.11	3.8	4.6	0.8
2004.12	3.9	1.4	2.5

2.3 Evaluation for effects of prediction model

By train, the predicted incidence rate of ND is Correspond with actual value, the result that coefficient of determination is $R^2=1.000$ show that the effects of fitting is good, and the check can be done later.

After input the check data into well-build prediction model, the prediction results were obtained and listed in table 4. combined with the prediction result and actual incidence, the coefficient of determination of prediction model $R^2=0.760$, this result illustrate that the prediction performance are good.

3. CONCLUSION

In human medical, BP neural network was used not only in prediction of infectious disease but also in diagnosis, prognosis, etiological factor analysis, risk analysis etc(Deng wei, 2002). The mathematic model applied in those domains generally requests a high foundation. Because of complex situation in practice, the building of prediction model is very difficult. Many scholar applied the artificial neural network in epidemic of tumor. The reason for meteorological factors became main object for this study was meteorological data easy-collection, affected ND greatly. By study the non-linear relationship between meteorological and incidence was build. The structure of network has been build and can be used to predict incidence rate of ND. In this study, the main work concentrated in building model, the model need to be train and check repeatedly, and the count of implied units need be adjusted endlessly. Data in Input layer need be normalized, different activation function and algorithm be selected. Different model need be screened and check many times, the model will not be selected until the result of checking has been reach the requirement. Because the selection of coefficient for network is random, the weight obtained in every train is different from others, therefore the train need be repeated endlessly. Otherwise, the quantity of data must be very great. Consequently, to investigate more data is the one important work in future.

REFERENCES

Deng wei, Jin Peihuan. The application of artificial neural network in prophylactic medicine. Chinese Public health, 2002, 18(10):1265~1267

G Greenough, M McGeehin, S M Bernard. The potential impacts of climate variability and change on health impacts of extreme weather events in the United States. *Environ Health Perspect*, 2001, 109(Suppl 2):191~198.

Li Jing, Wang Jingfei, Wu Chunyan, et al. the development of evaluation framework for risk of high pathogenicity bird flu. *Chinese agricultural science*, 2006, 39(10):2114~2117