The Classic Swine Fever Morbidity Forecasting Research Based on Combined Model

Yi Liang, Shihong Liu

The Classic Swine Fever Morbidity Forecasting Research Based on Combined Model

Yi Liang[1], Shihong Liu[1,1]

[1]Key Laboratory of Digital Agricultural Early-warning Technology,
Ministry of Agriculture, Beijing, P. R. China 100081
ryuui@126.com, lius@mail.caas.net.cn

Abstract. This paper proposes the combined forecasting model which study on the classic swine fever (CSF) morbidity, using the forecasting results of ARIMA and GM (1, 1) model as the inputs of the majorizing BP neural network. Analyzing the monthly data from 2000 to 2009 and the accuracy of the forecasting results is 97.379%, more accurate and more steady than traditional methods. This research provides efficient Analytical tools for animals' diseases forecasting work, and can provide reference to other animal diseases

Keywords: Combined Model, ARIMA, GM (1, 1), GA, BP neural network

1. Introduction

Classic Swine Fever(CSF) as one of the A level legal animal diseases of OIE, being the main control object to swine epidemics, causes a huge impact on the aquaculture industry in China[1]. Animal diseases prediction plays an important role in human health protection mechanism, and animal diseases outbreaks have complexity, therefore it has higher demands for modern society's forecasting methods.

Traditional forecasting methods have a relatively large difference in prediction accuracy, and this paper proposes a combined model which takes the optimized BP neural network by genetic algorithm (GA) as the carrier and the forecasting results of ARIMA mode and GM(1,1) model as the inputs. ARIMA model and grey forecasting model have good accuracy in data prediction, and the forecasting results after initial process by these two models is the input of optimized BP neural network, then construct the CSF morbidity forecasting model based on combined model, making empirical study aiming to the gathered data about morbidity and livestock on hand from January 2000 to June 2009(Data Source: official veterinary bulletin from MOA).

Defined the amount breeding stock at end of year as N, monthly occurrence number as M, so the formula of morbidity defined as follow:

[1] Corresponding Author, Address: Agricultural Information Institute, Chinese Academy of Agricultural Sciences, 100081, Beijing, P. R. China, Tel: +86-10-82103075, Fax: +86-10-82103075.

$$Morbidity = \frac{M}{N} \times 10000\% \qquad (1)$$

10000% means the occurrence number per ten thousand, meanwhile it facilitates calculation.

2. ARIMA model

ARIMA model that Autoregressive Integrated Moving Average, the Model regards the time series of predicted object as the variable depends on time T, and describes this time series approximately by mathematical model. We will describe the predicted object's development continuity from past value and present value after the autocorrelation of this set of random variables being identified, and then we can forecast the practical data [2].

Because the predicted model' s time series of ARIMA model is steady random series whose means is zero, so original data must be steady by first, second or natural logarithm according to their liner relation.

Pick the January 2000 to May 2008 as the model constructive data, and June 2008 to June 2009 as the verification data. Verifying and selecting the fittest model by AIC and SC criterion, and finally confirmed the model is ARIMA (1, 0, 1).

The expression of model is:

$$y(t) = 0.001022 + 0.812276 y(t-1) + u(t) - 0.164883 u(t) \qquad (2)$$

3. Grey model GM (1, 1)

Gray forecast model is a predicted method which builds mathematical model using little and incomplete information [3]. When we solve the practical problem using operation knowledge, formulate development strategies and policies, or make decision of vital problem, we should predict the future in a scientific way. Gray model is systemic theory based on gray information (information is incomplete, inadequacy, non-unique) in small sample, and it can provide a new solution to the problems which have complicated factors and ambiguous operative mechanism.

Give a macro forecasting data series:

$$x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(N)\} \qquad (3)$$

Use AGO operator after time cumulative:

$$x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(N)\} \qquad (4)$$

If $x^{(1)}$ satisfied the first order ordinary differential equation:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \tag{5}$$

a is called develop gray number and it reflects the develop trend of original data and predictive data; u is called Endogenous control grey number, which is constant input to system and reflects the relations of data variation. This equation satisfied the initial conditions:

$$x^{(1)} = x^{(1)}(t_0) \text{ when } t = t_0 \tag{6}$$

and the answer is:

$$x^{(1)}(t) = \left[x^{(1)}(t_0) - \frac{u}{a} \right] e^{-a(t-t_0)} + \frac{u}{a} \tag{7}$$

And OLS is:

$$\hat{U} = \begin{bmatrix} \hat{a} \\ \hat{u} \end{bmatrix} = (B^T B)^{-1} B^T y \tag{8}$$

Put estimation value into the response time equation:

$$\hat{x}^{(1)}(k+1) = \left[x^{(1)}(1) - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}} \tag{9}$$

And finally get the response time equation:

$$X(k+1) = 0.631815e^{0.001776k} - 0.631671 \tag{10}$$

Calculate the fitted value $\hat{x}^{(1)}(i)$, and restore the calculated result by reverse subtraction

$$x^{(0)}(i) = x(i) - \hat{x}^{(1)}(i-1)(i = 2,3,...,N) \tag{11}$$

4. The optimized BP neural network based on GA

Genetic Algorithm (GA) is applied to neural network widely, mainly due to the GA having a strong global search capability in a complex, polymorphism and continuous space, which can help neural network optimize its network structure and parameters. The most important operations in GA are: selection, crossover and recombination, mutation. Selection is to fix individuals for crossover and recombination and decide how many generations will be produced by these individuals, and selection always includes roulette methods, tournament methods and so on; crossover and recombination is a significant link to improve the population's qualities, which can fix new individuals combining the parents' genetic mating information, and crossover and recombination always includes real value reorganization and binary crossover; mutation is a change that individuals after crossover and recombination affects population by small probability transformation, and individuals can revolute via mutation, getting higher quality individuals, and it always includes real value mutation and binary mutation[4].

The process of GA is as follow: 1.generate the initial population and code them; 2. Analyze the fitness of individuals, and if individuals satisfies optimized principle then output the best individual with its parameters and end; else go next step; 3.select individuals by fitness and save the highest fitness one; 4.apply the cross operator on whole individuals to produce new generation; 5.apply the mutation operator on individuals of population to adjust the structure and build new individuals; 6.the individuals after selection, crossover, mutation constitute next generation and repeat 2.

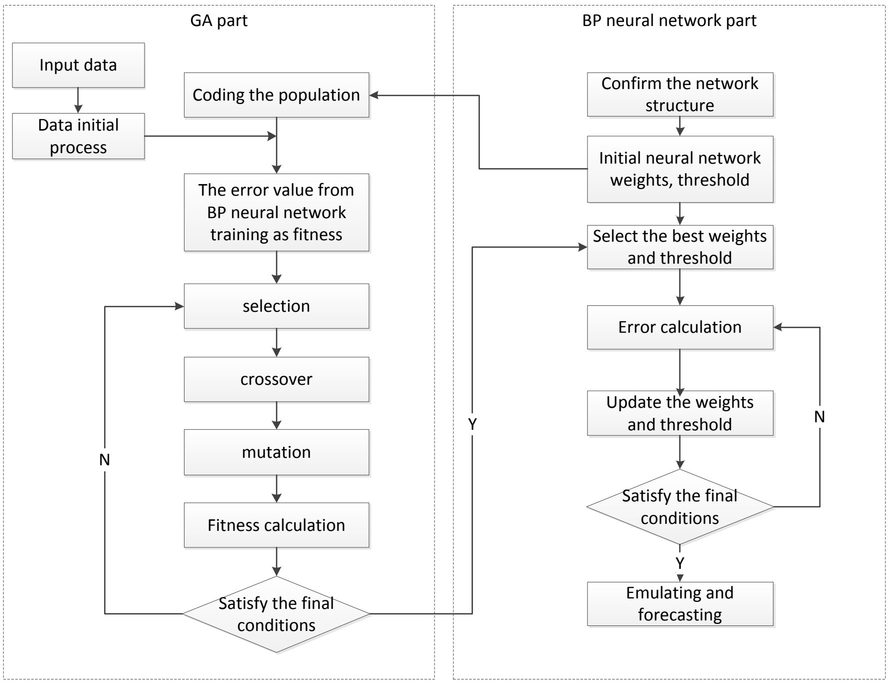

Fig. 1. Optimize the BP neural network by Genetic Algorithm

In the optimizing process of neural network, we can confirm the network structure and parameters via the global search capability of GA. We can rough adjust the network at the initial construction and get an approximate optimal solution. Meanwhile using the LM algorithm to adjust the network meticulously, in order to avoid GA algorithm sinking into a complicated iteration, which cost more time and space. GA is regarded as a decision algorithm not optimization algorithm and we can achieve the goal using rapidity and high efficiency of LM algorithm. This way balances the complexity, high efficiency and generalization of whole neural network [5].

5. Combined model

In the practical prediction, due to the model mechanism and starting point being different from others, there are different forecasting methods upon the same problem. Different methods provide different information and forecasting accuracy. If we abandon some low accuracy methods, we will lose some useful information. Therefore, we propose a more scientific way: combining different methods in a proper way, and form the combined model methods and it is propitious to synthesize useful information from all kinds of methods and improve forecast accuracy [6].

The model in this paper contains ARIMA model, grey model and optimized BP neural network model. ARIMA model based on the time series, and analyze the data using statistic, so that we can dig internal statistic relation from morbidity data. Grey model has a good capacity to incomplete information. Because the original data has data hollow space, then grey model can simulate the whole incidence trend well and form a creditable forecast process. BP neural network is the most popular A.I. model, and has strong functions at pattern recognition and approximation of function. Take the BP neural network as the carrier of combined model can achieve a higher accuracy and save more time and space cost.

Owing to the data quality, even though the forecast result of single model has directive significance, it still can't up to the scratch [7]. Take optimized BP neural network as the carrier of combined model, and change the inputs of neural network. The forecast result of ARIMA and grey model will be the inputs and improve the inputs' quality, which improves the efficiency, quality and produce a better result.

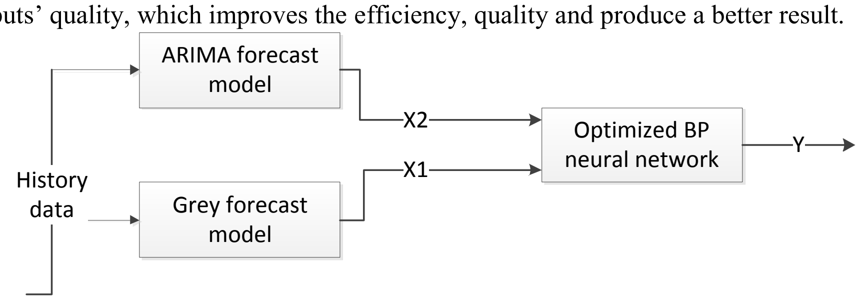

Fig. 2. Combined model structure. The results of ARIMA and GM (1, 1) as the input of BP neural network

Combined model process is as follow:

1). Select data from January 2000 to May 2008 as the model constructive data, June 2008to June 2009 as the model verification data. Build the GM (1, 1) model and define the forecast value as X1; build the ARIMA (1, 0, 1) model and define the forecast value as X2.

2). X1 and X2 are the inputs of neural network, the real data X is the output, and construct the input-output set ((X1, X2), X). The hidden layer of neural network has

10 neurons. Inputs layer has 2 neurons includes ARIMA forecast value and grey model forecast value. Output layer has 1 neuron of real data. The configuration of GA is as follow: the size of population is 10, evolution generation is 15, cross probability is 0.5, and the training process cost 14 iterations.

3).Optimize the neural network structure and configuration by GA, and achieve the forecast data Y, and verify the Y.

Compare the Y and real data X, broken line is the real data, and solid line is forecasting data. We can find that there is a little difference between real data and forecast data in the beginning of the graph, but at the posterior segment, the data error is almost negligible.

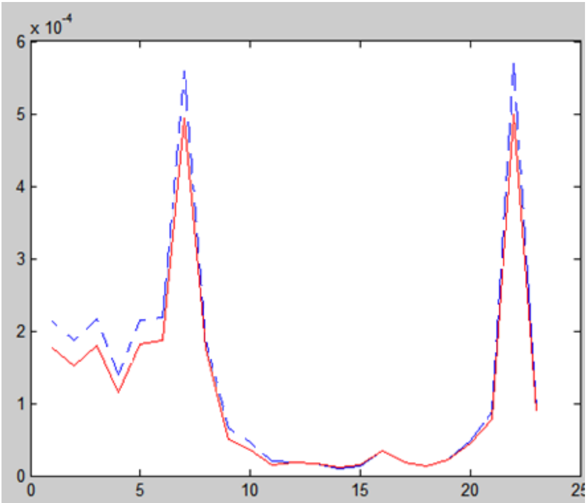

Fig. 3. Forecasting results analysis. The red solid line is the forecasting data and blue broken line is real data. The date of the result data range from June 2008to June 2009

Table 1. Different model comparison

Model	Accuracy (%)	Mean error (%)
Optimized BP model	97.379	2.261
BP model	96.191	3.809
ARIMA	91.91	8.09
GM(1,1)	93.88	6.12

Compare the forecast value with the real value, and the forecast range is the morbidity from June 2008 to June 2009. The mean error of forecast value without optimization is 3.809%, and optimized model is 2.621%, decrease 1.188% than non-optimization model.

6. Conclusion

The combined model of this paper takes the BP neural network as the carrier, integrates the vantage of ARIMA model and GM (1, 1), and increases the accuracy of data process. Due to the optimizing by GA, strengthen the global search capability of BP neural network, avoiding the problem of traditional BP network being easily sinking into local minima, and can play the role of combined model well.

Using the combined model for animal disease forecasting application, the result proves that it is feasible. The initial data process makes use of ARIMA model's steady process capability to non-stationary, and the long term forecasting trend of grey model, and can try to apply it to the practical case. With the data being richer, there is still a large room for us to adjust the model, improve the accuracy so as to get more precise forecast value and have stronger directive significance.

Reference

1. Hongming, X.: Severe Animal Diseases and Their Risk Analysis. Science Press. Beijing(2005)
2. Mengquan, W., Kai, Zh.: Hand-Foot-Mouth Disease Analysis and Prediction in Shandong in 2009 Based on the ARIMA Model. Ludong University Journal (Natural Science Edition). 71--75(2011)
3. Daocai, Ch., Yanfang, T., Tuo, G., Miao, Y., Zheng, L., Zongzheng, M.: Predication of Irrigation Water Use Using Parallel Gray Neural Network. Transactions of the Chinese Society of Agricultural Engineering. 26--29(2009)
4. MATLAB Chinese Forum, 30 Case Studies of MATLAB Neural Network. pp.286. BeiHang University Press. Beijing (2010)
5. Dong, Zh., Kaiyuan, C.: A Genetic-Algorithm-based Two-Stage Learning Scheme for Neural Networks. Journal of System Simulation. 1088-1090(2003)
6. Kaiyi, M.: Combination Forecast in Weight Determination and Application. Chengdu: Chengdu University of Technology. 2007
7. Ning, L.: Traffic Flow Forecasting Method Based on Combination Mode, Shanghai: East China University of Science and Technology. 2010