The Backpropagation Multilayer Feedforward Neural Network Based Competition Task Solution

Serhiy Shtovba, Yakiv Mashnitskiy
Vinnitsa State Technical University
Khmelnitskoe Shosse, 95, 21021, Vinnitsa, Ukraine
Phone: +380-432-440430, Fax: +380-432-465772
e-mail: shtovba@svitonline.com

ABSTRACT: The competition task solution, which was based on the use of the neural network and analysis of results are presented in the paper.

KEYWORDS: Prediction, Neural Network.

1. INTRODUCTION

The competition task lies in the prediction the index of the caravan insurance policyholders. This task relates with decision making tasks.
Analyzing the data we can not point out the exact structure of the data and derive the analytical formula for prediction. So for the solution of the given task it is reasonably to use the neural network, since it is not necessary to us to generalize the rule from the data and derive the analytical formula of the data approximation. It is a main argument for the application of the neural network for the task solution of such class. Neural network is the general approximator, which consists of simple elements operating in parallel.

2. THE SOLUTION

The solution of the task we divided into two stages. The first stage is the training of the neural network with the training data set and the testing of the one after the training. The second stage is the prediction and selection a number of the potential customers of the insurance license.
The task is solved by next algorithm:

1. The creation of the neural network.
2. The division of the training data set into five date set at random (the number of the one data set is 1200).
3. The training of neural network with each data set and in result we receive the five prediction models .
4. The testing each prediction model by whole set of experimental dada and the selection of the model, which has the maximum level of correctness.
5. Prediction of the potential customers from testing data set by the best model.

The steps 1-4 consist of the first stage, the step 5 is the second stage.
For solving the task we have created the neural network, which consists of three layers. The input (first) layer consists of 85 neurons according to the structure of the information about the customers. The hidden layer consists of 20 neurons, and the last layer has one neuron since we have only one parameter, which we have to determine. For the neurons of the input and hidden layers we have used the sigmoid transfer function. The neuron of the last layer has the linear transfer function.
For training the neural network we have used the resilient backpropagation training algorithm. The number of iterations is 350. The performance function is mean square error, which has been used to calculate the differences between the target output and the network output. It took 35 minutes to exceed the number of iterations for training the neural network with each data set. The value of the performance function we have received after the last training is 0.0008.
To solve the task we have used the architecture of the network with the backpropagation algorithm - the multilayer feedforward network. An elementary neuron has *R* inputs and each input is weighted with an appropriate w. The sum of the weighted inputs and the bias forms the input to the transfer function. Neurons may use any differentiable transfer

function to generate their output. This network can be used as a general function approximator. It can approximate any function with a finite number of discontinuities, arbitrarily well, given sufficient neurons in the hidden layer.
After the solving the task we receive the value of probability of having the caravan insurance policy by the customers. After that we choose the first 800 customers with the value of probability above the boundary probability so the ones are the policyholders and for the rest we can mail the advertising bulletins.

3.CONCLUSION

We haven't got the best score. The possible ways for the improvement of the model are

- increase of the number of the neurons in the hidden layer what let us to consider a greater number of the relationships among the data but it leads to the increase time of the training too;
- use of the genetic algorithm for the neural network learning since one is more effective in global extremum finding.