Journal of Industrial Engineering International *Islamic Azad University, South Tehran Branch*
January 2008, Vol. 4, No. 6, 32-38

Ensemble strategies to build neural network to facilitate decision making

P. Hanafizadeh*
Assistant Professor, Dep. of Industrial Management, Allameh Tabataba' i University, Tehran, Iran

E. Salahi Parvin
M.Sc. Student in Information Technology Management, Allameh Tabataba' i University, Tehran, Iran

P. Asadolahi
M.Sc. Student in Information Technology Management, Allameh Tabataba' i University, Tehran, Iran

N. Gholami
M.Sc. Student in Information Technology Management, Allameh Tabataba' i University, Tehran, Iran

Abstract

There are three major strategies to form neural network ensembles. The simplest one is the Cross Validation strategy in which all members are trained with the same training data. Bagging and boosting strategies produce perturbed sample from training data. This paper provides an ideal model based on two important factors: activation function and number of neurons in the hidden layer and based upon these factors, it compares the results of the trained single model with the cross validation one in a case which uses the presidential election data in US. The trained single model is called single best model. In this experience, the comparison shows that the cross validation ensemble leads to lower generalization error.

Keywords: Ensemble strategy; Neural networks

1. Introduction

The basic concept of the ensemble method is that diverse perspectives on a problem can be combined to produce a well rounded decision. Using changes in major parameters of learning in the neural networks (basic weights, number of hidden layer neurons and training data), decisions with lower errors comparing with having a single model can be made. In addition, if there was a special trend in training data, random basic weights and network construct, these factors would affect learning and change the final result.

By applying ensemble strategies, the effect of trends in training data can be minimized. By reaching the point in which the effect of mentioned trends is minimized, accurate decisions can be made regarding the result of training each neural network in the ensemble formed.

Hansen and Salmon [2] for the first time presented the result of training neural networks to reduce the generalization error of neural networks. They concluded that this error can be reduced using ensemble neural networks. They suggested an approach in which all members of the ensemble group are trained with the same training data. They also argued that training error can be reduced using the neural networks that are trained with the same training data. This strategy is called *cross validation* where the differences among results of each member in an ensemble are created because of various basic weights of each member [7].

The second strategy is called *bagging* where an improving ensemble is created and creates a unique training set that includes perturbed training data for

* *Corresponding author. E-mail: hanafizadeh@gmail.com*

each member of an ensemble [1].

Major characteristics of this strategy are that each member of an ensemble is trained under different training conditions and finally an algorithm is applied to the results of training of each member to make a clear decision [1].

The third strategy is *boosting* strategy which uses perturbation to improve the learning process. In the current strategy by series of continuous repetitive trains, each member of ensemble neural network enhances the significance of cases which are difficult to train or are not categorized according to previous trains.

In this paper, the authors assess the potentials of cross validation ensemble strategy to reduce generalization error and compare the performance results of cross validation strategy and a single model, which is one of members of the ensemble formed.

2. Ensemble neural networks applications

Primary applications of ensemble neural forecasting networks are categorization. Here, we illustrate several researches done using ensemble strategies.

Hu and Tesukalas [4] reported that multi-layer perception ensemble neural networks cause fewer errors in forecasting the environmental factors to choose costumers.

Sohn and Lee [5] used ensemble neural networks to increase the categorization accuracy of traffic crashes. They tested the boosting and bagging strategies and illustrated the reduction in generalization error compared to the single neural network.

West et al. [6] developed an optimized model for decision support systems to help with financial decision makings like bankruptcy forecasting and the optimization quantity of saving in banks.

Zhilkin and Somorjai [8] created an ensemble group using bagging strategy which uses multi-layer perception neural networks to categorize the brain signals and illustrated the high accuracy of this model comparing to a single model.

Zhou et al. [9] used ensembles of neural networks to identify lung cancer cells from needle biopsies.

3. Constructing single best model versus cross validation ensemble

3.1. Single best model

At the beginning two series neural networks are constructed. Both input and hidden layer have *tansig* as their activation functions but the difference is that the first series has *tansig* and the second series has *logsig* activation functions in the output layer. The structure for two series is stated below:

First Series: tansig – tansig – tansig

Second Series: tansig – tansig – logsig

The difference between members of each series comes from the number of neurons in the hidden layer which differs incrementally between 2 and the number of inputs. The number of neurons in hidden layer can be varied from 1 to 16. (16 is the number of neurons in input layer).

To reduce the calculation we have chosen the even numbers between 1 to 16. So, our operation started from network including 2 neurons in its hidden layer.

After that, among members of each series, one with lowest training error is chosen as the single best member. Then the comparison is made between the two best members of each series (which shows the optimum number of neurons in hidden layer), until the ideal activation function is determined.

Finally we have an optimum neural network which is ideal regarding to activation function and the number of neurons in hidden layer. This network is called single best model (Figure 1).

3.2. Constructing a cross validation ensemble

To construct a cross validation ensemble, 30 neural networks are grouped; each of them has a structure similar to a single best model. The only distinction between these single best models, each one is a member of the ensemble, is that the amount of basic training weights of them differs from one single best model to another (Figure 2).

3.3. Single best model versus cross validation

The results of training single best model and cross validation have been compared in an application to illustrate which of them can forecast with lower error.

Here we explain a case study to show the capability of the cross validation versus single best model.

4. Case study

4.1. Case introduction

In 1984 in U.S. before the presidential election a questionnaire designed which questioned the disagreement items between Democrat and Republican

Party Subjects which were of interest in the questionnaire included items listed below:

1. Handicapped-infants,
2. Water-project-cost-sharing,
3. Adoption-of-the-budget-resolution,
4. Physician-fee-freeze,
5. El-Salvador-aid,
6. Religious-groups-in-schools,
7. Anti-satellite-test-ban,
8. Aid-to-Nicaraguan-contras,
9. MX-missile,
10. Immigration,
11. Synfuels-corporation-cutback,
12. Education-spending,
13. Superfund-right-to-sue,
14. Crime,
15. Duty-free-exports,
16. Export-administration-act-South-Africa.

16 questions were asked in this questionnaire and they considered yes or no answers to them. In all questions the idea of followers of both Democrat and Republican were completely against each other. The above information has been extracted form Donald Bern computer and information institute in Irwin-California University.

4.2. Why is this case suitable to be solved by neural networks?

Firstly, as there is no way to ask from Americans, which party candidate they would vote, Democrat or Republican Party, a neural network has been designed to show which party candidate could be elected? The importance of this matter comes from the point that Americans agree with series of questions and disagree with the rest of them and do not obey the policy of a special party.

Secondly, the value of creation of neural networks to determine the tendency of each American to Democrat and Republican Parties. Although it is possible to answer a series of questions against the policy of a determined party.

5. The schema of neural networks

5.1. Single best model schema

Figure 1 illustrates a sample of how we have designed and created our singles models described in Section 3.1. Regarding Figure 1, the parameters M, N and O are described below:

M Number of neurons in input layer.

N Number of neurons in hidden layer.

O Number of neurons in output layer.

5.2. Cross validation ensemble schema

Figure 2 illustrates a general view of how we have designed and created our cross validation ensemble, each member of this group is exactly the same as single best model which we have made, with different basic training weight which is shown in Figure 2. Regarding Figure 2, the parameters M, N and O are described below:

M Number of neurons in the input layer.

N Number of neurons in the hidden layer.

O Number of neurons in the output layer.

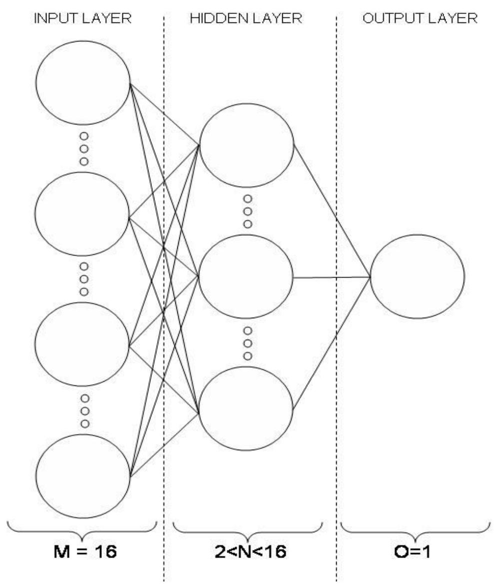

Figure 1. The schema of single best model.

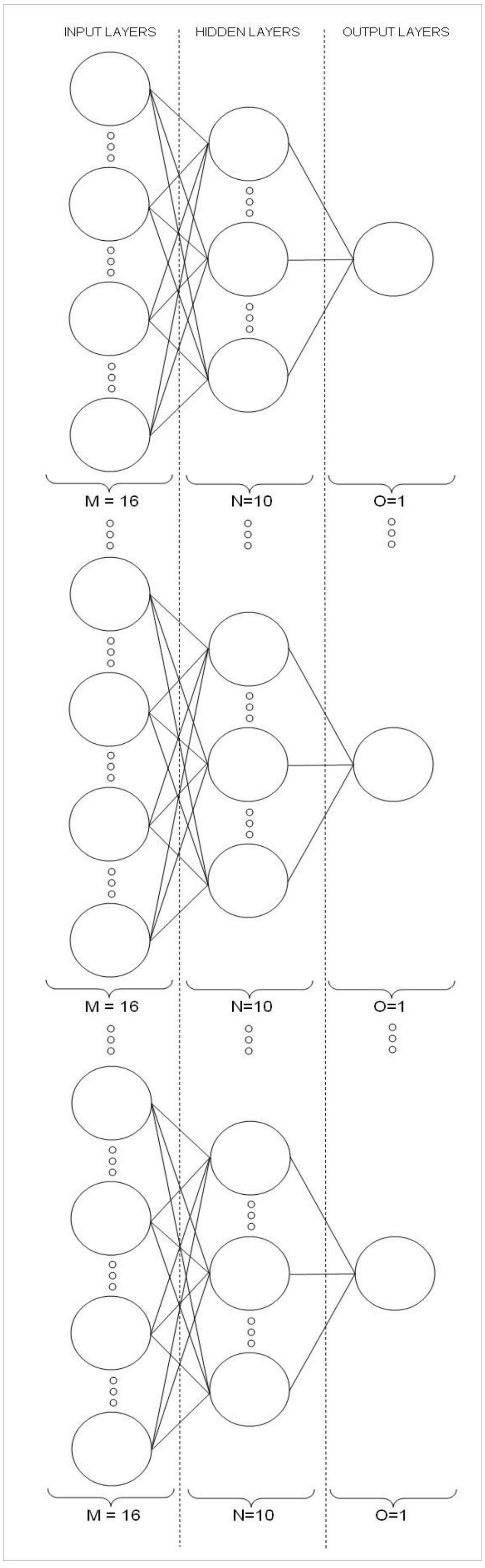

Figure 2. The schema of cross validation ensemble.

6. Finding structure of single best model

6.1. Series 1

As stated in Section 3.1, activation functions of all neurons are *tansig* and the number of neurons in the hidden layer differs between 2 to 16. Three sample diagrams from trained networks with MATLAB® software are shown in Figures 3, 4 and 5.

6.2. Series 2

The activation functions of the input and the hidden layer neurons are *tansig* and the output layers' are *logsig* and the number of neurons in the hidden layer differs between 2 to 16. Three sample diagrams from training the networks with MATLAB® software are shown in Figures 6, 7 and 8.

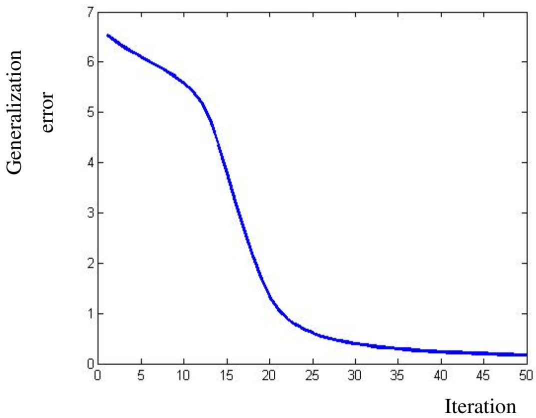

Figure 3. Number of neurons in the hidden layer is 6.

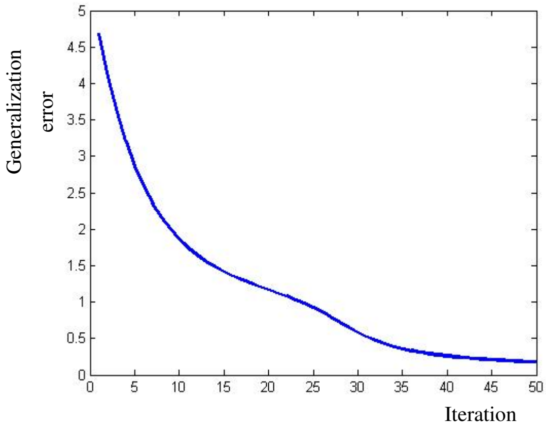

Figure 4. Number of neurons in the hidden layer is 10.

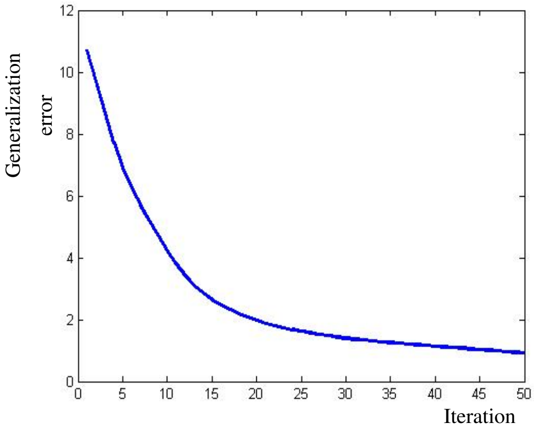

Figure 5. Number of neurons in the hidden layer is 16.

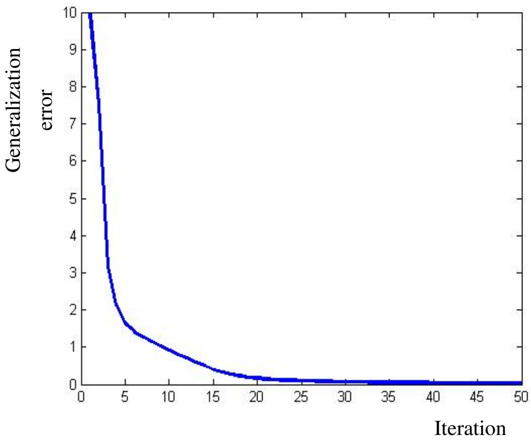

Figure 6. Number of neurons in the hidden layer is 6.

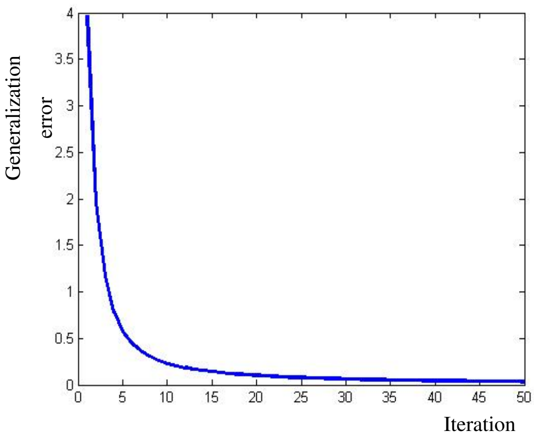

Figure 7. Number of neurons in the hidden layer is 10.

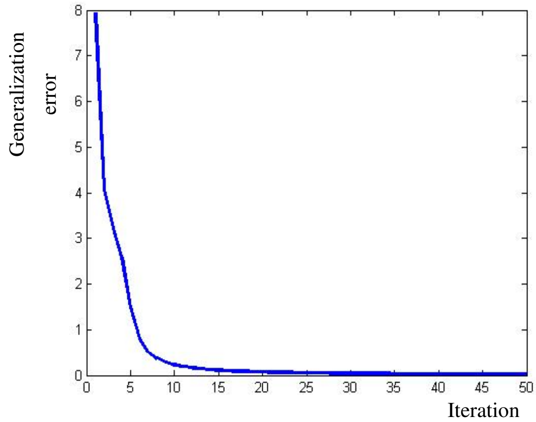

Figure 8. Number of neurons in the hidden layer is 16.

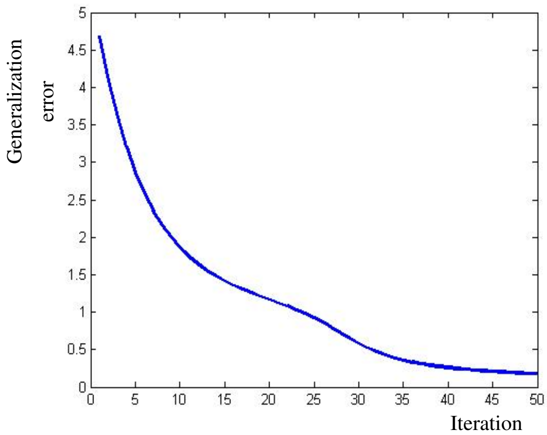

Figure 9. Number of neurons in the hidden layer is 10 and the activation function is *tansig*.

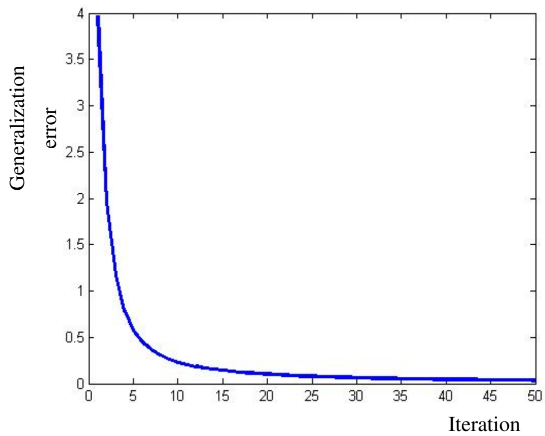

Figure 10. Number of neurons in the hidden layer is 10 and the activation function is *logsig*.

6.3. Determining best number of neurons in the hidden layer in each series

After running the program and training all members of each series individually, we have compared the result of training and determined which neural network has lowest generalization error in the series. The outcomes show that in both series, 10 neurons in the hidden layer have trained with lowest generalization error in the series.

6.4. Determining single best model structure

To recognize which activation function is more suitable for this case, the outcomes from trained networks, each network belonging to different series and having 10 neurons in their hidden layer, are compared. Finally we have found out that network which has 10 neurons in its hidden layer and has *logsig* activation function in its output layer has lowest generalization error. The outcomes of training each network with different functions in the output layer and similar number of neurons in the hidden layer are illustrated in Figures 9 and 10.

Figures show that the network with *logsig* activation function in the output layer and 10 neurons in the hidden layer can be known as single best model.

7. Creating cross validation ensemble

As mentioned in Section 3.2, the cross validation ensemble is created, with joining 30 neural networks, each of them containing 10 neurons in their hidden layer and considering *logsig* activation functions in their output layer. Using this ensemble we have solved the problem described in the case and the result achieved is illustrated in Figure 11. In the next section, the results of training of cross validation ensemble and single best model are compared.

8. The results of cross validation ensemble results versus single best model

Here, the results of training cross validation ensemble and single best model are compared in order to determine which method of creating neural networks has lower generalization error (Figures 12 and 13). By comparing Figures 12 and 13, it is evident that the rate of decreasing generalization error in the cross validation ensemble is more than that of the single best model. Therefore in deciding which party candidates the voters would vote, the cross validation ensemble can make a better decision comparing with single best model.

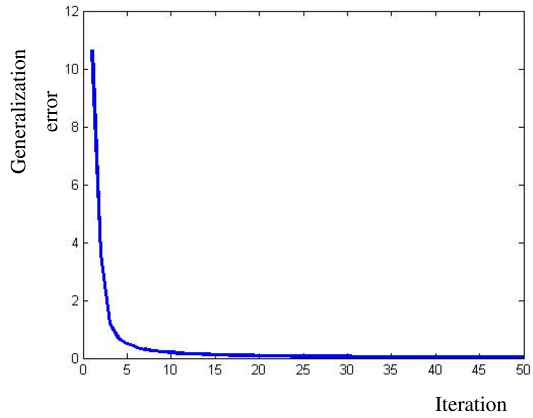

Figure 11. Generalization error in cross validation ensemble containing single best models.

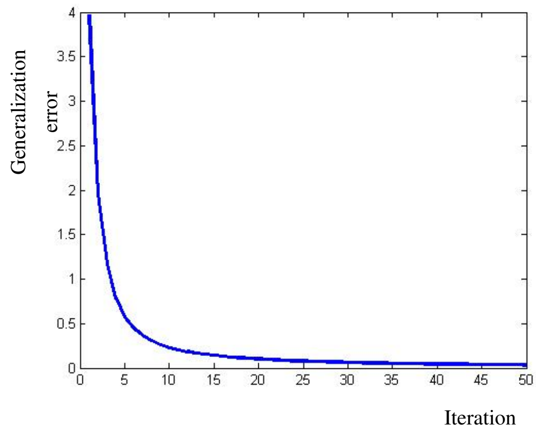

Figure 12. Generalization error in single best model with *logsig* activation function and 10 neurons in the hidden layer.

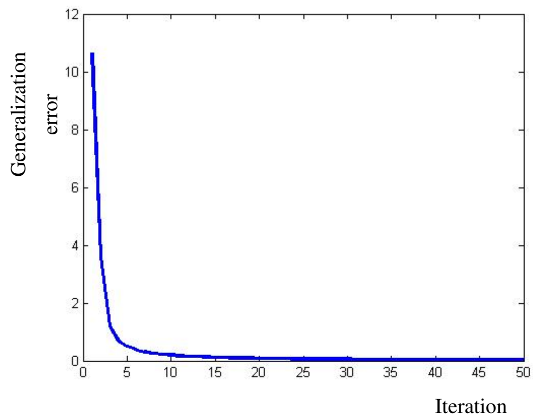

Figure 13. Generalization error in cross validation ensemble containing 30 single best models.

9. Conclusion

The ensemble strategy is suitable for prevention of trends which affect learning process and the rate of decrease in generalization error made. But using these makes problems become large in terms of size. Also it is practically impossible and the training process takes longer. Among ensemble strategies, the cross validation strategy is selected because of its simplicity. In order to form a cross validation ensemble, the authors have reached an appropriate single model called single best model. Finally to compare the results of cross validation ensemble to single best model, some experiences are made. The results show that the rate of decreasing generalization error in cross validation ensemble is more than the rate of decreasing generalization error in single best model. Thus it is less erroneous using a cross validation to decide.

References

[1] Dietterich, T., 2000, An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Journal of Machine Learning*, 40, 139-157.

[2] Hansen, L. K. and Salamon, P., 1990, Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 993-1001.

[3] Hayashi, Y. and Setiono, R., 2002, Combining neural network predictions for medical diagnosis. *Journal of Computers in Biology and Medicine*, 32, 237-46.

[4] My, H. and Tesoukalas, C., 2003, Explaining consumer choice through neural networks: The stacked generalization approach. *European Journal of Operational Research,* 146, 650-660.

[5] Sohn, S. and Lee, S., 2003, Data fusion, ensemble and clustering to improve the classification accuracy for the severity of roadtra QC accidents in Korea. *Journal of Safety Science*, 41, 1-14.

[6] West, D., Dellana, S. and Qian, J., 2004, Neural network ensemble strategies for financial decision applications. *Journal of Computers & Operations Research,* 2543-2559.

[7] West, D. and Mangiameli, P., 2004, Ensemble strategies for training neural network classifiers. *Journal of encouraging creative research,* Department of Decision Sciences College of Business Administration.

[8] Zhilkin, P. A. and Somorjai, R. L., 1996, Application of several methods of classification fusion to magnetic resonance spectra. *Journal of Connection Science*, 427-442.

[9] Zhou, Z. H., Jiang, Y., Yang, Y. B. and Chen, S. F., 2002, Lung cancer cell identification based on artificial neural network ensembles. Artificial Intelligence in Medicine, 24, 25-36.

[10] www.house-vote-84_names.html.