A Survey on Ensemble Combination Schemes of Neural Network

Varuna Tyagi
Department of Information Technology
Amity University,
Noida, India

Anju Mishra
Department of Information Technology,
Amity University,
Noida, India

ABSTRACT
The Neural network ensembles are the most effective approach to improve the neural network system. The combination of neural networks can provide more accurate result than a single network. The simple averaging, weighted averaging, majority voting and ranking are commonly used combination strategies, and from these strategies each method has its limitations like for which application area particular is suited .This paper present a survey on different ensemble combination schemes as invented in literature.

1. INTRODUCTION
DNA microarray technology can represent expression level of thousands of genes [1], using several statically methods and machine learning [2], these genes information can be analyzed rapidly and precisely by managing all this information at one time. The cancer classification can be incomplete or misleading. For this reason DNA microarray technology has been used to the field of accurate prediction of cancer. Accurate classification is necessary issue for the treatment of cancer. Gene expression data consist of huge amount of genes, and several researchers have been working on the problem of cancer classification using data mining methods, machine learning algorithm and statically methods [3, 4]. Many researcher have worked on the ensemble of multiple classifier to improve the performance of classification, it's not only increase the accuracy of classification but also gives more accurate results.
The representative ensemble methods such as average combination, voting, weighted voting and Bayesian approach have been applied to many fields. Ensemble inspired by stacking[5] uses cross validation technique, Reliability based ensemble[6] uses several steps, that is discussed in this paper, Bagging[7] uses bootstrap sampling, optimal method of ensemble[8] uses EDA algorithm, ensemble method using weights[9] works on weights, majority voting ,weighted voting and Baysian combination [10] also uses weights for ensembling the classifiers.

2. STACKING
Stacking [11] constructs two areas of ensemble preparing data and ensemble combination. It select training data for ensemble units by cross validation technique, for exploring second level generalizers, it combine the results of the first level generalizers (ensemble units). By stacked generalization the information supplied to the first-level ensemble units comes from multiple partitioning of the original datasets, which divides that datasets into two subsets. Every ensemble unit is trained by one part of the partitions, and the rest of the part is used to generate the outputs of the ensemble units (to be used as the second space generalizers(units) input). Then second level generalizers are trained with the original ensembles outputs that are treated as the correct guess. Infect, stacked generalization works by combined classifiers with weights according to the individual classifier performance, to find a best combination of ensemble outputs. Based on the idea of the stacking combination, the next method [1] is proposed.

2.1 ENSEMBLE INSPIRED BY STACKING
This method [12] inspired by stacking, by using a single neural network model as a combiner to combine the ensemble units results. This method is more efficient because it provides effective generalization compared with majority voting.

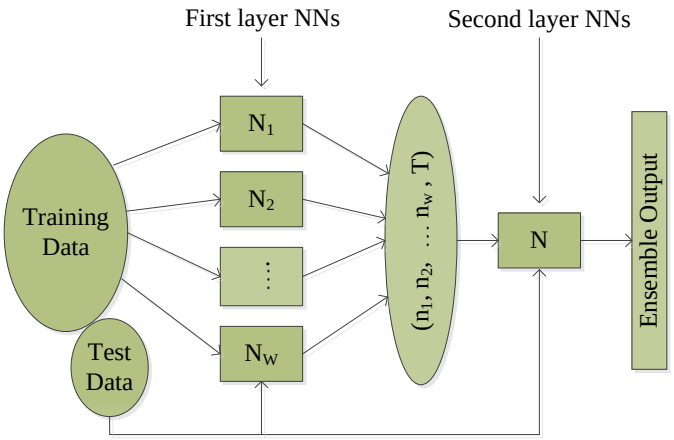

Fig 1: A demonstrate multistage ensemble neural network.

The experiment on multistage neural network ensembles depends on a well trained group of diverse single neural networks. To combine these well trained neural networks, a single neural network is trained. This well trained neural nets results by concatenating their outputs together as its input. It's depend on the capability of another neural network that they may be employing or not. By adjusting the connection weights, a neural network can be trained to perform complex functions. Majority voting doesn't adopt weights while combining than other approaches. For this some automatic method should introduce to assign weights to those ensemble units instead of using some traditional mathematical method manually. A neural network can automatically adjust the connection weights. Hence neural network is used to combine the results.
In Fig 1 , suppose there is a source data set A{a1,a1,……an} and its corresponding target dataset D{d1,d2,…..dn}.This target data set are partitioned into test data and training data. The training data should be preprocessed by applying some method for generating results before it being applied to the first layer's of neural network models N1, N2, ….. Nw. Bagging, Boosting etc, the preprocessing methods on training data set.

After training, the test data set will be applied and each first layer neural networks' corresponding results (n1, n2,…..nw) are used as the second layer neural network model's inputs. The second layer neural networks, was trained by using the first layers generated results on the whole training data as inputs combined with their target data set . As early as 1993, some experiments were done in digit recognition [13] by using a single layer network to combine ensemble classifiers. In 1995, Partridge and Griffith presented a selector-net approach [14]. The selector-net was defined as a network which used the outputs from a group of different trained nets as its input. More recently, Kittler [15] stated that:"it is possible to train the output classifier separately using the outputs of the input classifiers as new features".

Very recently, Zeng [16] used a single neural network as an approximator for voting classifiers. It was claimed that storage and computation could be saved, at the cost of a little less accuracy.

3. MULTISTAGE RELIABILITY BASED NEURAL NETWORK ENSEMBLE

In this method, a bagging sampling approach is first used to generate different training sets for enough training data. In terms of different training data set multiple individual neural classifiers are trained. A decorrelation maximization algorithm is used to select the ensemble members from the multiple trained neural classifiers. After this on some bases ensemble members are aggregated, and their generated results are output based upon reliability measure. The final result is called the ensemble output. The architecture of this method is shown in Fig. 2.

3.1 Partitioning Original Data Set

Bagging [17] is used for creating samples by varying the data subsets selected. The bagging algorithm is very efficient in constructing a reasonable size of training set due to the feature of its random sampling with replacement. This algorithm uses the bagging algorithm to generate training data subsets during the scarcity of original data.

3.2 Creating Different Neural Network Classifiers

Several methods have been introduced for the generation of ensemble members making different errors [18]. The main methods include the following steps.

3.2.1 Varying Initial conditions:

By varying some initial conditions like initial random weights, different ensemble member can be created.

3.2.2 Various network architecture

Changing the number of hidden layers and the number of nodes in every layer, different neural networks with different architectures can be created.

3.2.3 Various training data:

By re-sampling and preprocessing data, we can obtain different training sets for making different network generations [19]. The techniques for obtaining different training sets are bagging [20], noise injection [21], cross-validation [22].

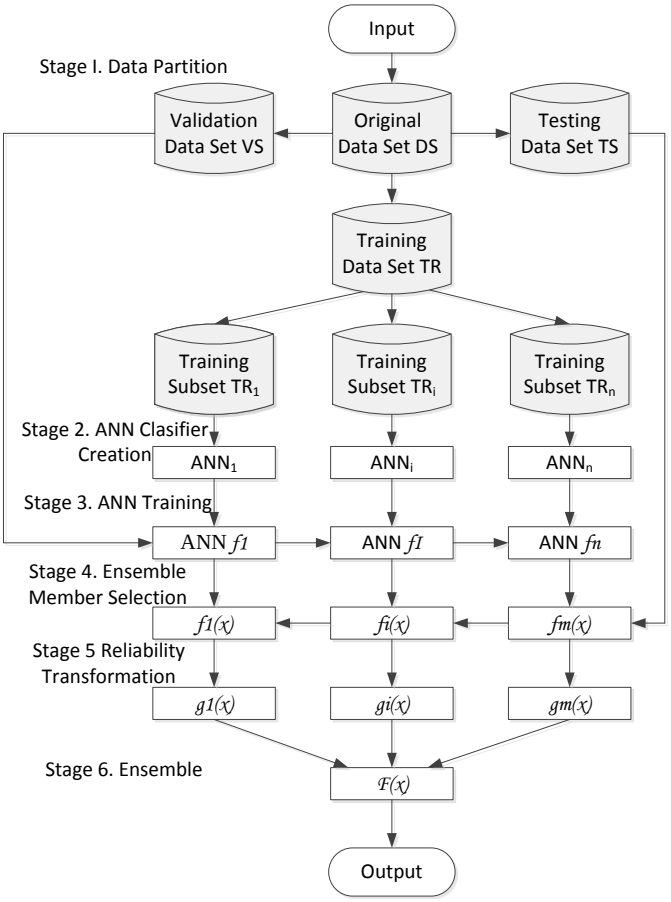

Fig 2: The process of multistage neural network ensemble learning model.

3.2.4 Various training algorithm

Diverse ensemble member can also be generated by selecting different core learning algorithms. For example, a multilayer feed forward network can use the steep-descent algorithm [23, 24], Levenberg-marquardt algorithm[25] and other training algorithms.

In this method, the third way is selected and the three layer backpropagation nueral networks (BPNN) is selected because a sufficient amount of middle-layer units.

3.3 Neural Network Learning and Confidence Value Generation

After creating different neural network classifiers, Different training datasets are used to train the neural network. In this method, for the class of supervised error back-propagation learning method, the BPNN is used in the form of the neural network associative memory. Backpropagation learning mechanism has two stages forward stages and backward stages.

3.4 Selecting Right Ensemble Units

After learning each neural network classifier has generated its own result. From the great number of individual members, it's necessary to select a subset performer in order to increase ensemble efficiency. The more number of classifier is not better for the ensemble but it should be performer and informative as mentioned by Yu, Wang, and Lai[26]. Different neural network classifier is required for ensemble learning. In this method, a decorrelation maximization technique [21] is used to decide the appropriate number of neural network ensemble members.

3.5 Reliability Value Transformation

In the previous step, the outputs of neural networks are used as the measure of realiability. It is not an issue that reliability value held into the interval (-∞, +∞).The main disadvantage of this confidence value is that the classifier who have the large absolute value mostly dominate the final decision of the ensemble architecture.

To overcome this limitation, the strategies are to re-scale the normalized the output value into zero and unit standard deviation, i.e.

$$g_i^+(x) = \frac{f_i(x)-\mu}{\sigma} \quad (14)$$

Where μ and σ are the mean and standard deviation of the pooled classifier outputs, respectively. For the ease of use to convert the confidence value into the unit interval [0, 1] is a nice solution [27]. In neural network different functions is used for reliability transformation such as scaling function, i.e.,

$$g_i^+(x) = \frac{1}{1+e^{-f_i(x)}} \quad (15)$$

3.6 Integrating Multiple Classifiers Into Ensemble Output

This step is related to the previous several steps. Based on the previous steps a set of right number of ensemble units can be collected. To make aggregated classifier architecture, it is necessary to combine these selected units. There are some ensemble strategies in the literature like ranking, majority voting and weighted averaging .majority voting is widely used ensemble strategy for classification problems because of its ease of use. Ensemble member voting determines the final decision.

4. ENSEMBLE BY USING BAGGING AND BOOSTING

In general a neural network ensemble is constructed in two steps, i.e. training number of units of neural networks and combined these units predictions. As for training the units of neural networks, the mostly used approach are Bagging and boosting. The origin of bagging is bootstrap sampling was proposed by Breiman[23].Many learning sets is generated from the original training set and then train each of unit neural network by each of those training sets. Boosting was proposed by Schapire[24] and improved by Freund et al.[25]. It produces a sequence of unit neural networks. Unit neural network training sets are produce by the former steps. Wrongly predicted training sets by former networks are more important in training of later networks. Speed of neural network training always changes slowly.

5. OPTIMAL CLASSIFIERS DESIGN METHOD FOR CONSTRUCTING ENSEMBLE CLASSIFIERS

Rather than selecting all classifier, it is better to select many classifiers for constructing the committee [26].So it is necessary to select the appropriate classifiers to form the classification committee. In literature many approaches can do this task such as greedy hill climbing. It can observe all the possible local changes of current set, for example adding one classifier to the set or removing one. It chooses the best for improving the performance of subset. Once a change is made for a subset, It is never reconsidered but cannot find the optimist solution. In this selection method EDA algorithm is used.

The EDA was first introduced by Larranaga, P. and Lozano, J. A [27]. It is a search algorithm that removes crossover and mutation from the Genetic Algorithm(GA). It produces the next generation based on probability distribution of N superior population samples. EDA provides probability distribution that generates more superior units at each stage.

Suppose A base classifiers are generated after trained by the feature subsets. They expressed as N1, N2, N3,…….,Nk. D is the subset of {N1, N2, N3,……Nk}. Binary vectors are introduced to denote D. If Ni is selected, the ith position of the vector is 1. While Ni is not selected, the ith position is 0.The Binary vectors are used as a chromosome of units and they can be evolved by EDA algorithm.

In order to measure units, the fitness function must be introduced. We first generate the validation set S and then calculate the error Rs of each individual on S.1/Rs is the fitness. R_s is depicted as follows:

$$R_{si=\sum_{j=1}^{A} p_{ij} \times classifier_j}$$

Here R_{si} is error of the ith unit. A is the total number of base classifiers. p_{ij} is the binary number of chromosome at the j_{th} position.$classifier_j$ is the error of the jth base classifier on S.

6. ENSEMBLE CLASSIFIERS BY VOTING

Complementarily correlated features are used to classification problem. Given k*n is the features-classifier combinations. There were Nm possible ensemble classifiers when m feature-classifier combination was ensemble classifiers. Complementary correlated features were used to train the ensemble classifiers and finally for the output a combining module was used. After the classifiers train independently with some features to produce their on outputs, the final answer can be judged by a combining module, where the majority voting, weighted voting or Bayesian combination can be adopted.

The networks that are learned from negative correlated gene subsets can also be combined. Since combining the heterogeneous classifiers helps in increasing the performance of the classification. In this method the Bayesian approach is used. The Tie-break between the ensemble classifiers can be solved by using the Bayesian approach with priori knowledge of each combined classifier.

7.1 Majority Voting

In this method the class that is most favored by the base classifiers is used. Majority voting does not require any previous knowledge and complex computation to decide. Where Ci is the class i(i=1,…….m), and Si(classifier j) is 1 if the output of the jth classifier j equals the class i otherwise 0, majority voting is defined as follows:

$$C_{ensemble} = \underset{1\le i\le m}{\arg max}\left\{\sum_{j=1}^{k} S_i(classifier_j)\right\}$$

Weighted Voting

In majority voting a poor classifier can affect the result of the ensemble. The effect of poor classifier is reduced by weighted voting by giving a different weight to a classifier based on the performance of each classifier. The accuracy of training dataset determines the weights of the classifiers. In weighted voting the weight of the jth classifier is defined as follows:

$$C_{ensemble} = \underset{1\le i\le m}{\arg max}\left\{\sum_{j=1}^{k} w_j S_i\, classifier_j)\right\},\ w_i = \frac{1-E_i}{\sum_k (1-E_k)}$$

Bayesian Combination: When classifiers are combined with the help of majority voting with their results, the Bayesian

combination makes the error possibility of each classifier affect the final result. The method combines the classifiers with different weights with the help of previous knowledge of each classifier. Where k classifiers are combined, C_i, i=1,…..,m, is the class of a sample, C($classifier_j$) is the class of a sample, c($classifier_{j)}$ is the class of the j_{th} classifier, and w_i is the a priori possibility of class C_i, the Bayesian combination is defined as follows:

$$C_{ensemble} = \underset{1 \le i \le m}{\arg\max} \left\{ \prod_{j=1}^{k} p(C_i / C(classifier_j)) \right\}$$

7. CONCLUSION

This paper has described that, on a huge datasets, multistage neural network ensembles provide improved performance. In this survey several method of ensembles are represented, which shows how classifier can be ensemble in a better way for providing the improved performance. In this paper several methods of ensemble like Ensemble inspired by stacking, reliability based neural network ensemble, Bagging and Boosting, Majority Voting, Weighted
Voting and Baysian Classification have explained. In future the performance of multistage neural networks can be enhanced by using more ensemble members in the layers, choice of training, validation and test datasets and choice of neural network for the next layer combiner.

8. REFERENCE

[1] M.B. Eisen, P.O. Brown, DNA Arrays For Analysis Of Gene Expression, Methods Enzymol. 303 (1999) 179–205.

[2] C.A. Harrington, C. Rosenow, J. Retief, Monitoring Gene Expression Using DNA Microarrays, Curr. Opin. Microbiol. 3 (2000) 285–291.

[3] A. Ben-Dor, L. Bruhn, N. Friedman, I. Nachman, M. Schummer, N. Yakhini. 2000. Tissue Classification With Gene Expression Profiles, J. Comput. Biol. 7 pp. 559–584.

[4] S. Dudoit, J. Fridlyand, T.P. Speed. 2000 Comparison Of Discrimination Methods For The Classification Of Tumors Using Gene Expression Data, Technical Report 576, Department Of Statistics, University Of California, Berkeley.

[5] Shuang Yang, Antony Browne, And Philip D. Picton 2002. Multistage Neural Network Ensembles. LNCS 2364, pp. 91–97, Springer-Verlag Berlin Heidelberg.

[6] Krogh, A., & Vedelsby, J. 1995. Neural Network Ensembles Cross Validation And Active Learning. In G. Tesauro, D. Touretzky, & T. Leen (Eds.), Advances In Neural Information Processing Systems (Pp. 231–238). Cambridge, MA: MIT Press.

[7] Breiman, L. 1996. Bagging Predictors. Machine Learning, Vol. 24, pp. 123-140.

[8] Freund, Y. ,Schapire, R.E. 1997. A Decision-Theoretic Generalization Of On-Line Learning And An Application To Boosting. Journal Of Computer And System Sciences, Vol. 55, pp. 119-139.

[9] Zeng, X., Martinez, T. R. 2000. Using A Neural Network To Approximate An Ensemble Of Classifiers. Neural Processing Letters, 12, pp. 225-237.

[10] Lean Yu, Shouyang Wang, Kin Keung Lai. 2008 Credit Risk Assessment with a Multistage Neural Network Ensemble Learning Approach, Expert Systems With Applications 34 pp.1434–1444, Elsevier.

[11] Lee, D. S., Srihari, S. N. 1993 Handprinted Digit Recognition: A Comparison Of Algorithms. Pre-Proc. 3RD International Workshop On Frontiers In Handwriting Recognition, Buffalo, USA, pp. 153-162.

[12] Wolpert, D. H. 1992 Stacked Generalization. Neural Networks, 5, pp. 241-259.

[13] Partridge, D., Griffith, N. 1995 Strategies For Improving Neural Net Generalisation. Neural Computing And Applications, 3, 27-37.

[14] Kittler, J. 1998 Combining Classifiers: A Theoretical Framework. Pattern Analysis And Applications, 1, pp. 18-27.

[15] Breiman, L. 1996. Bagging Predictors. Machine Learning, 26, 123–140. Chen, M. C., & Huang, S. H. (2003). Credit Scoring And Rejected Instances Reassigning Through Evolutionary Computation Techniques. Expert Systems With Applications, 24, pp. 433–441.

[16] Sharkey, A. J. C. 1996. On Combining Artificial Neural Nets. Connection Science, 8, 299–314.

[17] Yang, S., & Browne, A. 2004. Neural Network Ensembles: Combining Multiple Models For Enhanced Performance Using A Multistage Approach. Expert Systems, 21, pp. 279–288.

[18] Raviv, Y., & Intrator, N. 1996. Bootstrapping With Noise: An Effective Regularization Technique. Connection Science, 8, pp. 355–372.

[19] Tumer, K., & Ghosh, J. 1996. Error Correlations And Error Reduction In Ensemble Classifiers. Connection Science, 8, pp. 385–404.

[20] Hornik, K., Stinchocombe, M., & White, H. 1989. Multilayer Feedforward Networks Are Universal Approximators. Neural Networks, 2, pp. 359–366.

[21] White, H. 1990. Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings. Neural Networks, 3, pp. 535–549.

[22] Yu, L., Wang, S. Y., & Lai, K. K. 2005. A Novel Nonlinear Ensemble Forecasting Model Incorporating GLAR And ANN For Foreign Exchange Rates. Computers And Operations Research, 32, pp.2523–2541.

[23] Lai, K. K., Yu, L., Wang, S. Y., & Zhou, L. G. 2006. Credit Risk Analysis Using A Reliability-Based Neural Network Ensemble Model. Lecture Notes In Computer Science, 4132, pp. 682–690. Huijuan Lu, Jinxiang Zhang, And Lei Zhang. 2006. Tissue Classification Using Gene Expression Data And Artificial Neural Network Ensembles, Pp. 792 – 800, Springer-Verlag Berlin Heidelberg. LNBI 4115.

[24] Schapire, R.E. 1990. The Strength Of Weak Learnability. Machine Learning, Vol.5, pp. 197–227

[25] Larranaga, P., Lozano, J.A. 2001 Estimation Of Distribution Algorithms: A New Tool For Evolutionary Computation. Kluwer Academic Publishers.

[26] Lars Kai Hansen, Peter Salamon. 1990. Neural Network Ensembles , IEEE Transactions On Pattern Analysis And Machine Intelligence, VOL. 12, NO. IO, Pp 993-1001.

[27] Kyung-Joong Kim, Sung-Bae Cho. 2006. Ensemble Classifiers Based On Correlation Analysis for DNA Microarray Classification, Elsevier. Neurocomputing 70 pp. 187–199.