A Heterogeneous Ensemble Network Using Machine Learning Techniques

Tarik Rashid

†Faculty of Computing, College of Computer Training (CCT),
102-103, Amiens Street, Dublin 1, Ireland

Summary

The process of the ensemble network can be defined as grouping several networks where each independent network from that group is trained independently and then outputs are combined to obtain the overall output. This is called ensemble network. Usually, the output of ensemble network is often more accurate than outputs of any independent network. The ensemble network can be either homogenous or heterogeneous. The homogenous ensemble network obtains the overall output from different independent network structures of the same type, whereas the heterogeneous ensemble network obtains the overall output from identical independent network structures, but each independent network is trained with different training sets. In this paper, we introduce a heterogeneous ensemble network that can combine the outputs of several networks of different types. Heterogeneous ensemble network is created from recurrent neural network namely is called multi context recurrent neural network, variants of this network is used with variants functions of support vector machines to improve accuracy of the prediction task. The ensemble network is applied here to solve the problem of forecasting for electricity load energy. The network shows the ability to avoid an under or over estimated prediction of electricity load.

Key words:

Keywords: Ensemble Network, Recurrent Neural Networks, Support Vector Machines, Prediction.

1. Introduction

The process of the ensemble neural network can be defined as grouping several neural networks where each independent network from that group is trained independently and then outputs are combined to obtain the overall output. The ensemble network can be either homogenous or heterogeneous. The homogenous ensemble network obtains the overall output from different independent network structures, whereas the heterogeneous ensemble network obtains the overall output from identical independent network structures, but each independent network is trained with different training sets.

Both Hansen and Salamon demonstrated that the generalization ability of learning systems based on neural networks can be considerably improved through ensemble artificial neural networks [1], i.e. training multiple artificial neural networks and combining their predictions [2]. Consequently, different types of ensemble neural networks [3] have already been successfully applied to various fields such as optical character recognition [4, 5, 6], face recognition [7, 8], image analysis [7], signals classification [9], etc. More importantly, in the recent years many research works are increasingly paying a great deal of attention to ensemble neural networks [3, 7, 10, 11]. This is because these ensemble networks are made up of a linear combination of several networks and are also better in terms of less training time and likelihood of falling into over fitting than an independent large network [11].

An independent network that is within the ensemble network can have a potentially different weight in the output of the ensemble. For most regression and classification problems, combining the outputs of several predictors improves the performance of a single generic one [3].

Formal support to this property is provided by the so-called bias/variance dilemma [11, 12, 13] based on a suitable decomposition of the prediction error. Ensemble members must be both accurate and diverse, which poses the problem of generating a set of predictors with reasonably good individual performance and independently distributed predictions for the test. Different independent predictors can be achieved in different conducts. These include; Learn from various data sets or subsets of the data set. Apply various techniques to learn from the data (in various machine learning techniques). Apply different structures of a given technique (for instance, number of input/hidden neurons, using different architectures or different activation functions in various machine learning techniques) and learn from various learning algorithms.

Manuscript received August 20, 2009
Manuscript revised August 30, 2009

The most commonly used method for combining the networks is the average of the outputs of the networks, with a weight vector that measures the confidence in the prediction of each network. The problem of obtaining the weight vector is not an easy task. In this work we have considered the weights values are identical for each network. The aim is to produce estimators with lower prediction error, even though the justification of this constraint is just intuitive [12, 13]. When the method of majority voting is applied, the vote of each network is weighted before it is counted. Obtaining the optimal weight vector may not be simple and can be approached in different ways, nonetheless, the basic ensemble method (BEM), as it is called by Perrone and Cooper [14], consists of weighting all the networks similarly.

This paper introduces the heterogeneous ensemble network that is made of different types of machine learning technique namely; multi context recurrent network[15] and support vector machines [16] to handle the under or over estimated prediction for a power station. Next the architecture of the ensemble network is explained, then after, different simulations are explained, then the results are explained and finally, the main points are concluded.

2. Theoretical Consideration

The process of ensemble network is designed from the multi context recurrent neural network which is made of 4 layers namely; input, hidden, context (memory) and output layers [15]. At initial time, the input layer receives an external input pattern. At this time, the context layer neurons have zero values; the neurons in the hidden layers receive activations from both input and context neurons, the neurons from output layer receive inputs from hidden and context layers, and then the hidden neurons simultaneously feed forward and backward to the output neurons and context neurons respectively. The output in the output layer is compared with the target, and the error back propagates through the network layers to modify the connection weights. The recurrent connection weights from the hidden layer to the context layer are not exposed to any modification. The same process is repeated at subsequent time steps. The multi context recurrent neural network can be trained with Back Propagation (BP), Back Propagation Through Time (BPTT) and Real Time Recurrent Learning (RTRL) [15, 17, 18].

Support vector machines (SVMs) [19, 21, 21] are another type of machine learning technique, used to handle the classification and regression problems. This technique displays input data in two sets of vectors in an n-dimensional space. The techniques build a separating hyper plane in that space, to maximize the margin between the two data sets. In order to compute the margin, the technique builds two parallel hyper planes, each of which on each set of separating hyper plane, to separate the two data sets. Generally speaking, a fine separation can be obtained by the hyper plane that has the major distance to the neighboring data points of both classes, and this is because the major margin will reduce the generalization error of the classifier [19, 21, 21].

A few kernels are used with SVMs regularly. The most commonly used are the radial basis functions, Gaussian, Polynomial kernels (homogenous and non-homogenous), Sigmoid functions (logistic and hyperbolic tangent) [19, 21, 21]. Generally any function can be used as long as it represents a dot product in some vector space. All valid functions must satisfy a condition called Mercer's condition [19, 21, 21]. However, finding the best function is very difficult (this is another reason why we have used ensemble networks). A general model is used to derive a function based on the information from the given training data. Different kernels are used in this work, namely Gaussian, Polynomial with zero, first, second, third and forth degree. The ensemble network is made of variants of both multi context recurrent neural network and support vector machines; the multi context

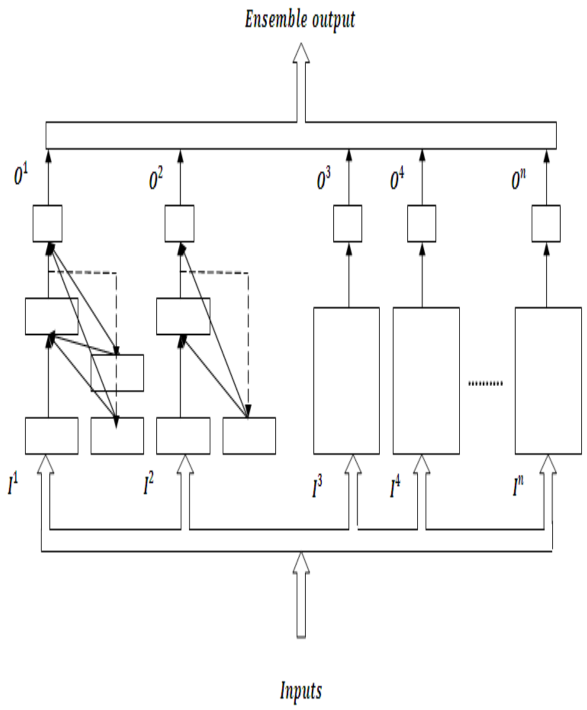

Fig. 1: shows Heterogeneous Ensemble network composed of variants MC-RNNs and SVMs.

recurrent network (MC-RNN) is trained with RTRL and support vector machines used with Gaussian and different kernel functions (see Fig. 1) to solve the problem of electric load forecasting

3. Experimental Consideration

3.1 Simulation Experiment

Load forecasting task depends on the past and current information of the load for a period of time. Our forecasting system can be processed as follows: obtain and analyze the historical data; pre-processing and normalizing information; choosing the training and testing set; choosing the type of the network and network's parameters; choosing a suitable learning algorithm; and finally implementation. The inputs and outputs are vectors of either discrete attributes of real values as for regression task. We have used the same input/output variables for our MC-RNN variants, as input/output variables to SVMs. The inputs are selected from a combination or four types of variables; the time and seasonal variables, absolute and/or change in weather component variables, the status of the day (social events) and historical absolute and/or change in load. The output of an SVM is the current change of the daily peak load, which is the difference between the forecast daily peak and the previous daily peak loads. We have different training data sets [15] [16]; therefore, we had various selections of input variables applied to the different structures and kernel function of ensemble network.

The network consists of multi context recurrent network and support vector machines. Each of which has five variant structures. For example a variant of MC-RNN has different input, context layer, hidden layers and neurons and different training sets from the other structures. On the other hand, a variant of support vector machine has different kernel functions (Gaussian, different polynomials).

The same network structure is used for the different periods (classes of data sets). Each network structure uses relatively different parameter values. These parameters rely heavily on the size of the training and testing sets. Learning rates and momentum are varied from 0.009 to 0.3 and from 0.3 to 0.7 respectively. The training cycles are also varied between 18000 and 60000 cycles. Fig. 2 displays prediction results of 5 different structures of MC-RNN for January 1999 that has not included in the training sets against the target data.

3.2 Results

It can be seen from above Fig. 2, that MC-RNN (2) over estimated the prediction results, whereas MC-RNN (5) under estimated the prediction results. MC-RNN ensemble produced more accurate and closer to the target values.

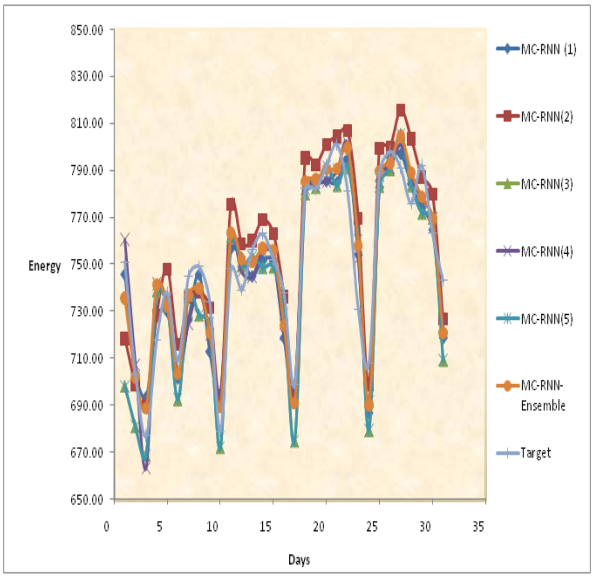

Fig. 2: Shows the forecasting results for 5 different structures of multi context recurrent network makes MC-RNN ensemble network.

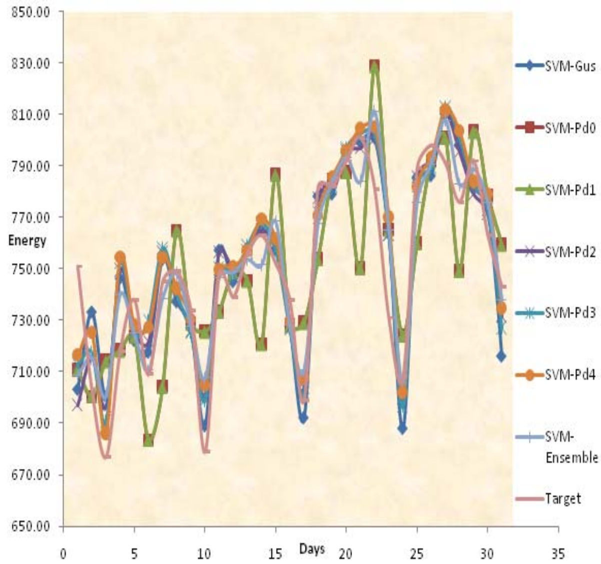

Fig. 3: shows the forecasting results of 5 different kernel functions use to form SVM- ensemble network.

Fig. 3 shows the forecasting results of 5 different kernel functions use to form SVM- ensemble network. Support vector machine using polynomial zero degree function produced prediction values that are relatively away from the target values, whereas SVM with polynomial forth, third and second degree were closer to the target. The SVM ensemble marginalized the difference and it was more accurate.

Fig. 4 shows MC-RNN ensemble performed slightly better than SVM ensemble. Nevertheless the ensemble network narrowed the difference between the actual prediction and the target values.

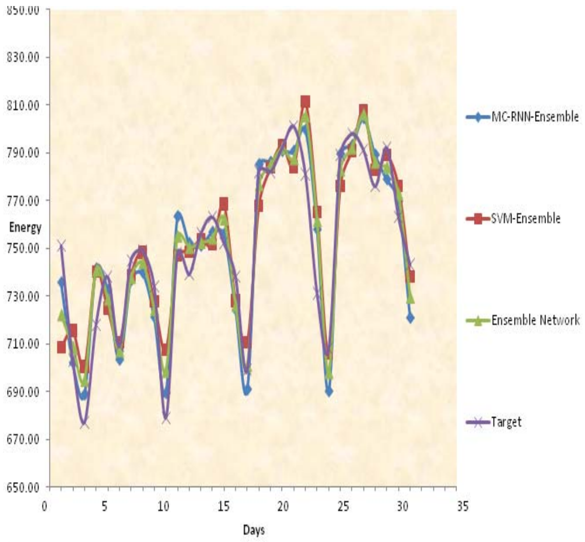

Fig4. The forecasting results of ensemble network.

4. Conclusion

This paper introduces heterogeneous ensemble network that is made of variants of MC-RNN structure and support vector machines using different function. As it is very difficult to select the structure in the network as it depends mainly on trail and error, also it is very difficult to determine the best kernel function to use for support vector machine therefore ensemble network is used to reduce the over and under estimation of the prediction results. The ensemble network shows a safer prediction than any individual network to solve the problem of load energy forecasting. Although some research was done concerning to basic homogenous ensemble network, very little or no research work was done on heterogonous ensemble network that involves combining different network structure within different type or family of networks. Therefore we believe more research work should be done on this topic on different sections of ensemble network, most importantly, techniques on distribution weights for individual networks and selecting best kernel functions on such as weight distribution on individual networks.

Acknowledgment

The author would like to express his cordial thanks to Faculty of Computer Science, College of Computer Training.

References

[1] L. K. Hansen and P. Salamon. Neural network ensembles. IEEE Transactions Pattern Analysis and Machine Intelligence, vol.12(10):993–1001, 1990.
[2] Z. Hua Zhou and Y. Jiang. Medical diagnosis with c4.5 rule preceded by artificial neural network ensemble. IEEE Transactions on Information Technology in Biomedicine, 2004.
[3] D. Sharkey. Combining artificial neural nets: Ensemble and modular multi–net systems. Springer–Verlag, London, 1999.
[4] H. Drucker, R. Schapire, and P. Simard. Improving Performance in Neural Networks Using a Boosting Algorithm, chapter Advances in Neural Information Processing Systems, pages 42–49. CA: Morgan Kaufmann, 1993.
[5] J. Mao. A case study on bagging, boosting and basic ensembles of neural networks for ocr. In Proceeding of IEEE International Journal Conference, Neural Networks, pages 1828–1833, Anchorage, AK, 1998.
[6] L. K. Hansen, L. Liisberg, and P. Salamon. Ensemble methods for handwritten digit recognition. In Proceeding IEEE Workshop on Neural Networks for Signal Processing, pages 333–342., Helsingoer, Denmark, 2002.
[7] F. J. Huang, Z. H. Zhou, H. J. Zhang, and T. H. Chen. Pose invariant face recognition. In Proceeding of the 4th IEEE International Conference Automatic Face and Gesture Recognition, pages 245–250, Grenoble, France, 2000.
[8] S. Gutta and H. Wechsler. Face recognition using hybrid classifier systems. In Proc. IEEE Int. Conf. Neural Networks,pages 1017–1022, Washington, DC, 2002.
[9] Y. Shimshoni and N. Intrator. Classification of seismic signals by integrating ensembles of neural networks. In IEEETransactions Signal Processing, volume 4, pages 1194–1201, 1998.
[10] Z. Hua Zhou and Y. Jiang. Medical diagnosis with c4.5 rule preceded by artificial neural network ensemble. IEEE Transactions on Information Technology in Biomedicine, 2004.
[11] P. M. Granitto, P. F. Verdes, and H. A. Ceccatto. Neural network ensembles: Evaluation of aggregation algorithms. arXiv:cs.AI/0502006 v1 1, 1,
[12] L. Breiman. Bagging predictors. Machine Learning, 24:123–140, 1996.
[13] D. Ortiz-Boyer Dortiz, C. Hervis-Martknez Chervas, and N. Garcka-Pedrajas Npedrajas. CIXL2: A crossover operator for evolutionary algorithms based on population features.
[14] M.P. Perrone and L.N. Cooper. When Networks Disagree: Ensemble Method for Neural Networks, chapter Artificial.

[15] Rashid, T., Gleeson, B. and Kechadi, M-T. . (2006) Neural Network Based Methods for Electric Load forecasting. The 3rd International Conference on Cybernetics and Information Technologies, Systems and Applications (CITSA'06), Orlando, Florida, USA (July 20-23).

Tarik Rashid received the B.Sc. degree from Mosul University in 1990 and his Pg.D. degree from Griffith College Dublin (GCD) in 2001.

He received his Ph.D. degree from University College Dublin (UCD) in 2006. Dr Tarik Rashid is postdoctoral follow and senior member of Parallel Computational Research Group (PRCG) at University College Dublin, and currently is a director of academic affairs at the College of Computer Training.