Using Stacked Generalization and Complementary Neural Networks to Predict Parkinson's Disease

Pawalai Kraipeerapun
Department of Computer Science
Faculty of Science, Ramkhamhaeng University
Bangkok, Thailand

Somkid Amornsamankul*
Department of Mathematics
Faculty of Science, Mahidol University
and Centre of Excellence in Mathematics
Bangkok, Thailand

Abstract—**This paper proposes the integration between stacked generalization and complementary neural networks to diagnose Parkinson's disease. The Parkinson speech dataset acquired from the UCI machine learning repository is used in our study. Complementary neural networks compose of the truth and the falsity neural networks which are trained to predict the truth output and the falsity output. Stacked generalization consists of two levels. They are level 0 and 1. Ten-fold cross validation is used for training complementary neural networks created in level 0. All outputs produced from each fold are merged to create new input feature. Five sets of machines are trained to create five features which are used as input used to train complementary neural networks created in level 1 of stacked generalization. It is found that the combination between stacked generalization and complementary neural networks provides better performance than using only the traditional stacked generalization or neural network in the prediction of Parkinson's disease.**

Keywords-Stacked Generalization; Stacking; Neural Network; Complementary Neural Network; Parkinson

I. INTRODUCTION

Parkinson's disease causes many movement disorders; for example, trembling in some parts of the body such as arms and legs, slowness of movement, and having trouble talking and walking [1]. In the brain of human beings, neurons which control movement normally generate dopamine. If neurons die and the level of dopamine decrease then the problems of movement occur [2]. Therefore, a number of researches have tried to detect Parkinson's disease in the earlier state. This is a binary classification problem [3,4]. There are many techniques used for classification tasks, for instance, support vector machine (SVM), neural network, and Naïve Bayes [5,6,7,8].

In order to classify patient with Parkinson's disease and healthy people, several classification techniques have been used. For example, Kaya et al. [9] applied entropy-based discretization to the dataset before classification based on SVM, C4.5, Naïve Bayes, and k-nearest neighbors. In [10], Shahbakhi et al. applied SVM and genetic algorithm to diagnose Parkinson's disease. One of approved machines applies to diagnose Parkinson's disease is neural network. For example, Astrom and Koker [11] applied parallel feed-forward neural networks in which the outputs generated from neural networks were estimated using rule-based system. In [12], neural networks were created based on the Scaled Conjugate Gradient algorithm (SCG) and the Levenberg-Marquardt algorithm (LM) to diagnose Parkinson's disease. They found that LM provided better accuracy rate than SCG. In [13], information gain was used to reduce the number of features and neural network was also trained to diagnose this kind of disease.

This paper aims to combine the complementary neural networks and stacked generalization to diagnose Parkinson's disease. In [14], the joining between these two techniques was used to solve three regression problems which are concrete compressive strength, airfoil self-noise, and water quality. It was found that this combination technique gave better results when compared to the traditional stacked generalization. In our experiment, we aim to examine this combination technique to the problem of binary classification which is the problem of diagnosing Parkinson's disease. The dataset used in this study is the Parkinson speech dataset [15] obtained from UCI machine learning repository [16]. In our study, the combination technique can give better efficiency when it is compared to the traditional stacked generalization and neural networks for solving the problem of diagnosing Parkinson's disease.

II. STACKED GENERALIZATION AND COMPLEMENTARY NEURAL NETWORKS

Stacked generalization or stacking invented by Wolpert [17] is a technique used to combine multiple machines. It consists of two learning levels. These levels are level 0 and 1. The training dataset is partitioned based on *n*-fold cross validation. All partitions in level 0 are trained and tested for every machine. For each machine, all outputs from all testing parts are integrated to form a set of a new feature. After that, all integrated outputs which are new features from all machines are combined to be used as new inputs. These inputs are used to train a machine in level 1. After a machine in level 1 is trained, all machines in level 0 are retrained using the whole dataset.

In [18], complementary neural networks were introduced to classify the binary classification problem. Complementary neural networks or CMTNN compose of truth neural network and falsity neural network trained to predict truth output and and falsity output based on truth target and falsity target. If x is the truth target then the falsity target is 1-x. The difference between regression and classification problems is that the result of the regression problem is the average between the truth output and the non-falsity output whereas the result of the classification comes from the comparison between the truth output and the falsity output. If the truth output value is higher than the falsity output value then the prediction is 1; otherwise, it is 0.

In [14], stacked generalization and complementary neural networks were combined to solve regression problems. In our experiment, we will apply this concept to solve a binary classification problem which is Parkinson's disease.

Fig. 1 shows the integration between complementary neural network and stacked generalization for diagnosing Parkinson's disease. Each machine in level 0 composes of truth neural network and falsity neural network. Therefore, *m* pairs of neural networks are created. Each pair is trained using the truth target and the falsity target using *n*-fold cross validation. The truth value and the non-falsity value are averaged to form a new output. All outputs from level 0 are used as inputs to level 1. In level 1, a pair of truth neural network and falsity neural network is trained based on those new inputs constituting of *m* features. This pair aims to predict the final output which consists of truth output and falsity output. Finally, all pairs in level 0 are retrained using the whole dataset. The final outputs from level 1 are compared in order to classify Parkinson's patient and healthy people.

III. Experimental Methodology and Results

The Parkinson speech dataset [15] obtained from UCI machine learning repository [16] is used in this study. This dataset contains multiple types of sound recording. The training set is obtained from 20 Parkinson's patients and 20 healthy people. The testing set is collected from another group of 28 Parkinson's patients. Hence, 1040 training data and 168 testing data are used in our experiment.

The training data is partitioned using 10-fold cross validation. Therefore, ten sub datasets are created. Nine sub datasets are trained and another one is tested. The training process is done ten times with different sub datasets.

Five sets of complementary neural networks are trained using five different initial weights but the same architecture. Each set composes of a pair of the truth neural network and the falsity neural network. Each pair is trained based on 10-fold cross validation. Therefore, ten sets of truth and falsity outputs from ten testing subsets are created. These ten sets of output are integrated to form complete output of 1040 training data. Hence, we have five sets of complete output. In each output set, the average between the truth output and the non-falsity output are calculated. Therefore, we have five features to be used as input to level 1.

In level 1, the truth neural network and the falsity neural network are trained based on five input features created from level 0. After that, five complementary neural networks created in level 0 are retrained based on the whole training data.

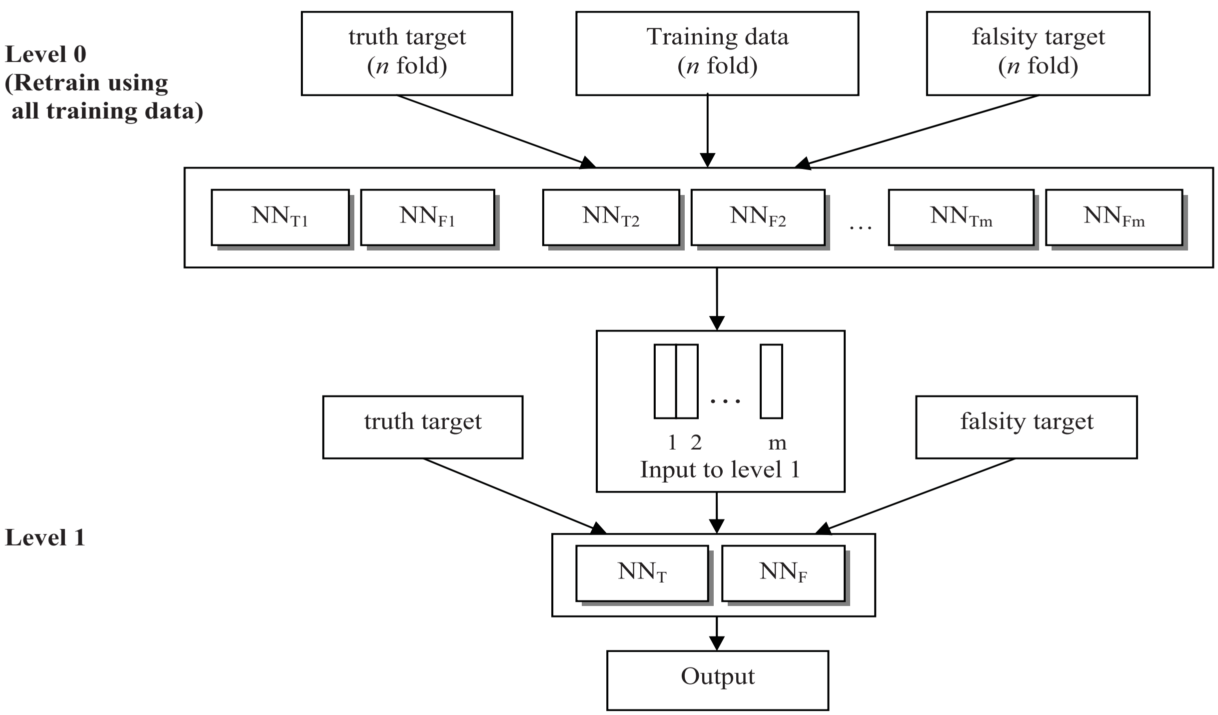

Figure 1. The combination between stacked generalization and complementary neural networks

In our experiment, the optimization is not our purpose; however, we aim to compare the proposed techniques with the traditional stacked generalization technique. Therefore, in level 1, we create ten sets of complementary neural networks with different initial weights but using the same input from level 0 in order to compare results from different neural networks.

Furthermore, we also compare the proposed techniques to the traditional feedforward backpropagation neural network (NN) and also compared to the complementary neural networks (CMTNN). Therefore, ten complementary neural networks are trained using the whole training data in order to compare results to the proposed techniques in level 1.

Fig. 2 shows results which are the percent correct from ten groups of machines in level 1 of stacked generalization. Each group contains four techniques. The first technique is the traditional stacked generalization. This technique considers only the truth neural networks in both levels. The second technique applies the truth neural networks in level 0 and using the truth neural network and the falsity neural network in level 1. The third technique applies the truth neural network and the falsity neural network in level 0 but uses the truth neural network in level 1. The fourth technique applies the truth neural network and the falsity neural network in both levels. For every technique in each group, we use the same truth neural networks in each level. Hence, the falsity neural networks are in the same environment. For the first and the third technique, if the final output is higher than 0.5 then the result is Parkinson's disease. For the second and the fourth techniques, if the truth output is higher than the falsity output then the result is Parkinson's disease.

Fig. 3 shows the average percent correct from each stacked generalization technique. It can be noted that using the opposite neural networks in stacked generalization provide better percent correct than using only the truth neural networks for diagnosing Parkinson's disease. Table 1 shows the percentage of improvement for each technique when it is compared with the traditional technique that applies the truth neural networks in stacked generalization.

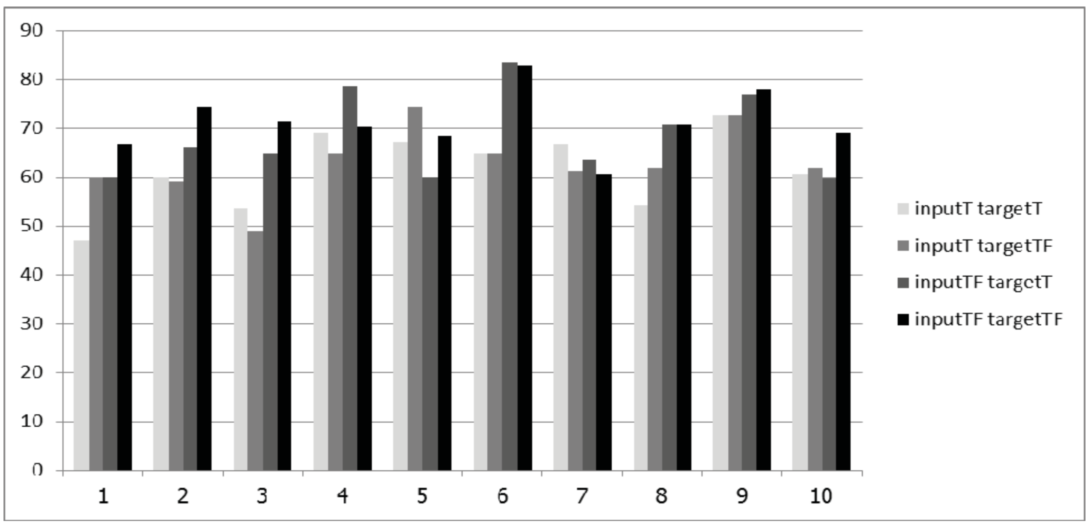

Figure 2. The percent correct obtained from each stacked generalization technique

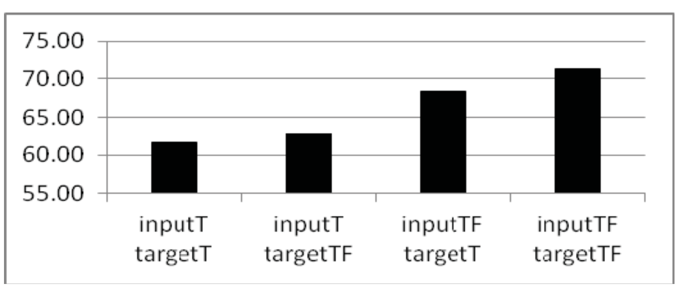

Figure 3. The average percent correct from each stacked generalization technique

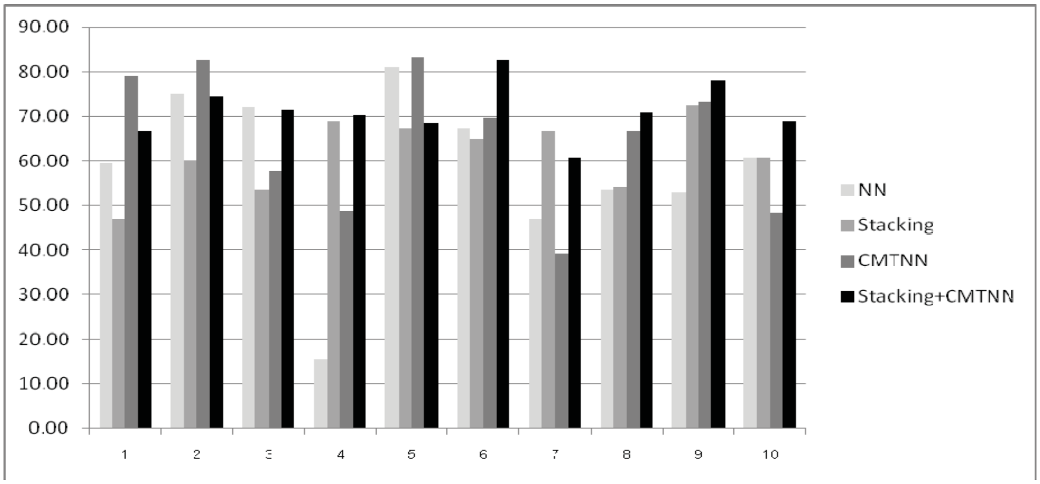

Figure 4. The percent correct obtained from each technique

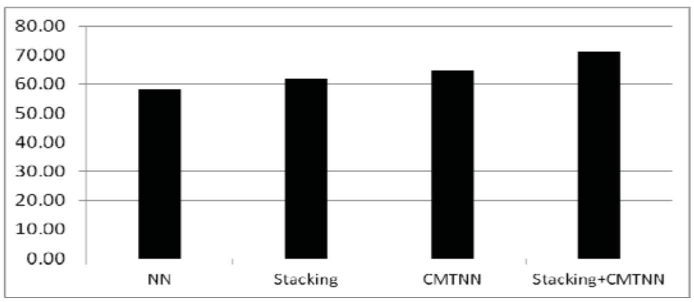

Figure 5. The average percent correct from each technique

TABLE I. THE PERCENTAGE OF IMPROVEMENT FOR EACH PROPOSED STACKED GENERALIZATION COMPARED TO THE TRADITIONAL STACKED GENERALIZATION IN LEVEL 1

Technique	%improve
inputT targetTF	2.13
inputTF targetT	11.01
inputTF targetTF	15.65

TABLE II. THE PERCENTAGE OF IMPROVEMENT OF OUR PROPOSED STACKED GENERALIZATION (INPUT TF TARGET TF) COMPARED TO OTHER TECHNIQUES

Compared to	%improve
NN	21.89
Stacked Generalization (Stacking)	15.65
CMTNN	9.82

Fig. 4 shows the percent correct obtained from different techniques which are feedforward backpropagation neural network (NN), stacked generalization or stacking (inputT targetT), complementary neural networks (CMTNN), and the proposed stacked generalization (inputTF targetTF) which is stacking+CMTNN. Fig. 5 shows the average percent correct from each technique. Table 2 shows the percentage of improvement of our technique (inputTF targetTF) compared to other techniques. It can be seen that the proposed technique provide the best result.

IV. CONCLUSION

In this paper, we use multiple types of sound recordings dataset [15] for diagnosing Parkinson's disease. This is a binary classification problem. The combination between stacked generalization and complementary neural networks is applied to separate the Parkinson's patient and the healthy people. It is found that the falsity values can be used to improve the correctness of predicting Parkinson's disease. In future, we will consider uncertainty conditions occurred in both truth and falsity neural networks on both levels. Uncertainty information will be added to this combination technique in order to enhance the prediction.

ACKNOWLEDGMENT

This research works are supported by the Centre of Excellence in Mathematics, CHE, Thailand.

REFERENCES

[1] NIH: National Institute of Neurological Disorders and Stroke, Parkinson's Disease, Retrieved 7 February 2015, from MedlinePlus: http://www.nlm.nih.gov/medlineplus/parkinsonsdisease.html

[2] W. Roth, "Small Handwriting and Other Early Signs of Parkinson's", Retrieved 7 February 2015, from Healthline: http://www.healthline.com/health-slideshow/parkinsons-warning-signs

[3] L.P. Wang and X.J. Fu, Data Mining with Computational Intelligence, Springer, Berlin, 2005.

[4] M.L. Raymer, W.F. Punch, E.D. Goodman, L.A. Kuhn and A.K. Jain, "Dimensionality reduction using genetic algorithms," IEEE Trans. Evolutionary Computation 4 (2000) 164-171.

[5] V. Kecman, Learning and Soft Computing, Support Vector machines, Neural Networks and Fuzzy Logic Models, The MIT Press, Cambridge, MA, 2001.

[6] F. Fazayeli, L.P. Wang, and W. Liu, "Back-propagation with chaos," Proc. 2008 IEEE International Conference on Neural Networks and Signal Processing (ICNNSP2008), Zhenjiang, China, June 7-11, pp.5-8,2008.

[7] L.P. Wang (Ed.): Support Vector Machines: Theory and Application. Springer, Berlin, 2005.

[8] V. Mitra, Chia-Jiu Wang, and S. Banerjee, "Lidar detection of underwater objects using a neuro-SVM-based architecture," IEEE Transactions on Neural Networks, no.17, pp.717-731, 2006.

[9] E. Kaya, O. Findik, İ. Babaoglu and A. Arslan, "Effect of Discretization Method on the Diagnosis of Parkinson's Disease," International Journal of Innovative Computing, Information and Control. vol. 7(8), pp. 4669-4678, 2011.

[10] M. Shahbakhi, D. T. Far and E. Tahami, "Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine," J. Biomedical Science and Engineering. vol. 7, pp. 147-156, 2014.

[11] F. Astrom and R. Koker, "A parallel neural network approach to prediction of Parkinson's Disease," Expert Systems with Applications. vol. 38, pp. 12470-12474, 2011.

[12] Z. A. Bakar, N. M. Tahir and I. M. Yassin, "Classification of Parkinson's Disease Based on Multilayer Perceptrons Neural Network," In Proceedings of the 2010 6th International Colloquium on Signal Processing & Its Applications (CSPA), Mallaca City, May 2010, pp. 232-235.

[13] A. Khemphila and V. Boonjing, "Parkinsons Disease Classification using Neural Network and Feature Selection," World Academy of Science, Engineering and Technology. vol. 6(4), pp. 15-18, 2012.

[14] P. Kraipeerapun and S. Amornsamankul, "Solving Regression Problems using Stacked Generalization and Complementary Neural Networks," In Proceedings of the 2014 International Conference on Earth Science and Applied Sciences (ICESAS 2014), Osaka, Japan, October 2014, pp. 335-343.

[15] B. Erdogdu Sakar, M. Isenkul, C.O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin and O. Kursun, "Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings," IEEE Journal of Biomedical and Health Informatics. vol. 17(4), pp. 828-834, 2013.

[16] M. Lichman, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2013.

[17] D. Wolpert, "Stacked Generalization." Neural Networks, vol. 5, pp. 241-259, 1992.

[18] P. Kraipeerapun and C. C. Fung, "Binary Classification using Ensemble Neural Networks and Interval Neutrosophic Sets," Neurocomputing. vol. 72(13-15), pp. 2845-2856, 2009.