

Learning of Neural Network with Reduced Interference – An Ensemble Approach

Meihua Li, Sheng-Uei Guan, Linfan Zhao, and Weifan Li

Abstract—This paper focuses on reducing the interference effect among input attributes. When training different attributes together, there may exist negative effect among them due to interference. To reduce the interference, input attributes are placed into different groups such that attributes with no interference with each other are placed in the same group. Two types of grouping strategies are examined in this paper, i.e. non-overlapping and overlapping. To further enhance the performance, multiple learners are employed to tackle different groups. Three integration methods i.e. voting, weighting and result-integration network (RIN) are examined. It turns out that the result-integration network has the best performance, followed by weighting and then voting. The ensemble approach can improve the performance of neural-network learning. Such an approach also can be employed with feature selection to further enhance the performance.

Index Terms—Interference, neural network, ensemble learning, grouping.

I. INTRODUCTION

In conventional neural network (NN) training, no partitioning is carried out in the input space, which leads to the existence of interference among attributes. For example, there may exist correlations between some attributes that can lead to overfitting, thereby reduction of the accuracy of learning. In many practical problems, some contradictory information may cost a lot of time in deciding the training direction therefore reduces the accuracy. Hence, it is crucial to keep our training away from negative interference.

Input space architecture has attracted interests of many researchers with significant work done [11-13, 16, 17]. In neuro-fuzzy network approaches [21], grid partitioning had been applied to the input/output datasets to generate an initial fuzzy inference system [27]. Sun and Peterson partitioned the input space and assigned a proper weight to each part [32]. However, none of these works considered the interference among input attributes within the network. Unfortunately, interference is one of the most important factors that affect the classification abilities of neural networks.

There exists research that focuses on reducing interference. Weaver et al. [28] mitigated the interference effect by the

reduction of a bi-objective cost function that combines the approximation error and a term that measures interference to adjust the weights of an arbitrary, non-linearly parameterized network. Kobayashi et al. [20] investigated the interference caused by training NNs when data are presented incrementally, i.e., data samples are shown sequentially. A long-term memory is incorporated into resource allocation network, i.e., the network will train all the new data and part of the old data. Although these works investigated interference, they did not deal directly with the input space partitioning as previously discussed. Currently, there exist little work using a partitioning approach to reduce interference among attributes for NN training on classification tasks.

Ang et al. [2] proposed an algorithm that works out the interference relationship among input attributes, and partitions them accordingly to this relationship. They separated the attributes that have interference with each other in different groups to avoid adverse effect, and placed the ones which don't have interference with each other in the same group to use information better. However, in Ang's work, although they partitioned the input space in the beginning, they trained all the attributes together at last. Thus they aimed to avoid interference by partitioning the input space while they still introduced interference in later training.

Ensemble learning is a method for constructing predictors or classifiers from an ensemble of weak predictors or classifiers [24]. It is used in many different areas such as semantic annotation of web data [9], information fusion [6] and classification problems [29]. There is a need of integration of results from multiple learners in ensemble learning. Many kinds of integration schemes exist, such as voting [5], Bayes voting [8], integration method based on D-S evidence theory [1, 7, 23], and integration of classifiers generated from different feature subsets [3,18,19,30,31]. We will consider these in our experiments.

In this paper, we propose an input attribute grouping scheme and an interference-less ensemble learning algorithm with ensemble learning integrated into the NN training procedure. The structure of the proposed network is described, and then some benchmark problems are employed to test our interference-less ensemble learning algorithm (ILEL).

II. THE CBP LEARNING ALGORITHM

Constructive learning algorithms include the dynamic node creation (DNC) method [4], cascade-correlation (CC) algorithm [10] and its variations [14, 25, 26], constructive single-hidden-layer network [13], and constructive back propagation (CBP) algorithm [22], etc. In this paper, CBP

Manuscript received September 18, 2012, revised November 15, 2012. This work was supported in part by the National Natural Science Foundation of China under Grant 61070085.

Meihua Li and Sheng-Uei Guan are with School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China, (e-mail: Steven.Guan@xjtu.edu.cn; lmh270341819@yahoo.com.cn)

Linfan Zhao and Weifan Li are with Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China (e-mail: Linfan.Zhao10@student.xjtu.edu.cn; Linfan.Zhao10@student.xjtu.edu.cn).

was employed as constructive neural network. CBP combines constructive algorithm and traditional BP network. The learning procedure of CBP network is shown in Figure 1. For details, please refer to [22].

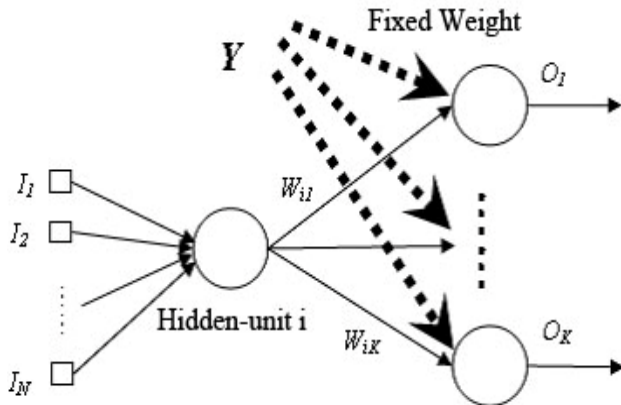


Fig. 1. Training of a Hidden-Unit in a CBP network [22]

III. INTERFERENC-LESS ALGORITHM

When input attributes are trained together, one attribute may have positive or negative effect on the other attributes within the same group. Interference is the negative effect between attributes. When determining the interference between attributes. We can refer to following formula:

$$\varphi(i, j) = \begin{cases} 0, & E_{ij} \leq \min(E_i, E_j) \\ 1, & E_{ij} > \min(E_i, E_j) \end{cases}$$

In this formula, E_{ij} stands for the error score when training i and j together. E_i is the error score gained by training i alone. If there are interference between two attributes i and j , $\varphi(i, j) = 0$, otherwise $\varphi(i, j) = 1$.

By training two data sets from UCI with focus on deriving E_i, E_j and E_{ij} , we build an interference matrix A for each benchmark. As the matrix is symmetric, it is presented as a lower triangular matrix. The first row/ column denote the attribute number, while the diagonal elements represent the error obtained by training each attribute alone. The other elements denote the errors by training different attributes together (i.e. A_{ij} is the error value obtained by training attributes i and j together).

TABLE I: INTERFERENCE MATRIX FOR GLASS DATASET

Glass dataset interference matrix									
Attribute	1	2	3	4	5	6	7	8	9
1	71.6038								
2	42.5472	55.8491							
3	48.7736	53.9623	62.0755						
4	46.3207	37.5472	36.8868	37.5472					
5	71.3387	57.7359	59.0189	42.0754	73.3019				
6	57.8302	57.2327	54.9057	37.7359	52.0755	63.0189			
7	51.6038	48.7736	43.0188	38.2076	64.9057	54.3396	56.3208		
8	58.4906	55.1887	62.2642	39.7170	57.0755	57.9515	47.6190	57.4906	
9	63.8679	53.4906	58.0190	48.8679	67.7530	57.3585	66.8868	56.6040	67.8302

TABLE II: INTERFERENCE MATRIX FOR VOWEL DATASET

Vowel dataset interference matrix										
Attribute	1	2	3	4	5	6	7	8	9	10
1	76.3765									
2	54.4130	68.8664								
3	76.2956	65.2632	88.6640							
4	62.9150	64.0081	81.4777	83.0972						
5	64.5142	63.9271	76.6802	75.3441	81.9636					
6	70.9917	67.0041	80.2429	77.3887	74.0081	84.5142				
7	68.1781	68.7450	85.8907	81.5385	78.5830	82.5709	86.9838			
8	68.7247	67.7733	81.3563	79.5547	77.2065	77.7530	81.2348	85.7895		
9	69.3320	63.4413	84.5547	78.9676	80.8502	80.5668	80.6478	83.4211	87.5911	
10	73.996	67.4696	90.0810	81.9231	79.6356	83.3603	87.0445	85.4453	85.6680	90.6883

The procedure of Interference-less grouping and ensemble approach to training is shown in Fig.2.

A. The Training Procedure Consists of Four Stages:

1) Create the interference table

Use the interference table to assemble input attributes into groups

Train each group using CBP sub-network respectively

Feed the intermediate result derived from all groups into an integration network and train it to get the final results

B. Grouping is Performed as Follows:

Step 1: Sort the input attributes by contribution;

Step 2: Choose the first attribute which has not been chosen, denote it as \underline{a} ;

Step 3: If there exists no group yet, create a new group and place \underline{a} inside. Else, test the existing group(s) one by one. If there is one or more attributes interfering with \underline{a} in a group, then proceed to test another group; if there is a group in which all the attributes do not interfere with \underline{a} , then we put \underline{a} in that group and go to Step 4. Else if \underline{a} cannot be placed in any of the existing group(s), create a new one and put \underline{a} inside, go to

Step 4;

Step 4: If there are other attributes remaining, go to Step 2; else, stop.

IV. DISCUSSION OF GROUPING STRATEGIES

A. Grouping Order

1) Sort by contribution in descending order

TABLE III: LIST OF ATTRIBUTES IN DESCENDING ORDER OF CONTRIBUTION

Name of dataset	Ordering of attributes
Glass	4,2,7,8,3,6,9,1,5
Vowel	2,1,5,4,6,8,7,9,3,10

TABLE IV: RESULTS OF NON-OVERLAPPING GROUPING IN DESCENDING ORDER OF CONTRIBUTION

Name of dataset	Results of grouping
Glass	{4,2,3}{7,8,6,1}{9,5}
Vowel	{2,1,5,4,6,8,7,9,3}{10}

2) Sort in random order

TABLE V: RESULTS OF NON-OVERLAPPING GROUPING IN RANDOM ORDER

Name of dataset	Results of grouping
Glass	{1,2,3,7}{4}{5,6,8,9}
Vowel	{1,2,3,4,5,6,7,8,9}{10}

3) Overlapping among groups

There exists no overlapping among the grouping strategies presented above, i.e. an attribute can be assigned to only one group. However, some attributes may have no interference in other groups as well. Putting attributes in multiple groups might benefit training. It motivates a grouping scheme which allows overlapping.

Unbalanced overlapping allows the deviation of occurrence frequency per attribute. Different from non-overlapping grouping, unbalanced overlapping considers all non-interfering groups. One attribute can appear in more than one non-interfering groups.

TABLE VI: RESULTS OF UNBALANCED OVERLAPPING IN DESCENDING ORDER OF CONTRIBUTION

Name of dataset	Results of grouping
Glass	{4,2,3}{7,8,6,1}{9,1,5}
Vowel	{2,1,5,4,6,8,7,9,3}{10}

Balanced overlapping only allows an equal frequency of occurrence for each attribute. To obtain an equal occurrence frequency for each attribute, we combined two strategies, sorting by contribution in descending order and sorting in random order, together. The results for two strategies were combined directly.

TABLE VII: RESULTS OF BALANCED OVERLAPPING IN DESCENDING ORDER OF CONTRIBUTION (1)

Name of dataset	Results of grouping
Glass	{4,2,3}{7,8,6,1}{9,5}{1,2,3,7}{4}{5,6,8,9}
Vowel	{2,1,5,4,6,8,7,9,3}{10}

TABLE VIII: RESULTS OF BALANCED OVERLAPPING IN DESCENDING ORDER OF CONTRIBUTION (2) - WITH DUPLICATION DELETED

Name of dataset	Results of grouping
Glass	{4,2,3}{7,8,6,1}{9,5}{1,2,3,7}{4}{5,6,8,9}
Vowel	{2,1,5,4,6,8,7,9,3}{10}

Note that the results of grouping remain the same for ‘Vowel’ remains the same under the ‘unbalanced’ and ‘balanced’ strategies.

4) Discussions of integration

During ensemble learning, there are multiple learners and each learner produces its own result. To get the final result, we need to integrate the results obtained from all learners. Three approaches for integration of results were examined in this paper i.e., voting, weighting and result integration network (RIN) methods. Voting applies majority vote for the results of different learners. The final result is the one with the most votes. Weighting method assigns different attributes with different weights according to their characteristics. The final result is obtained by a weighted average of results from different learners. RIN means training a special network to integrate results. The inputs should be the integration of outputs from each sub-network and the output of integration-network produces the final result.

In the weighting method, the determination of weights is crucial in deriving the final result. In this paper, three schemes were designed to determine the weights, i.e. equal weighting, weighting according to the contribution and weighting according to the training accuracy, i.e. precision. The equal weighting method is easy to implement. It assigns equal weight to each learner while there are differences between different learners. When the difference is significant, the integrated result might not be satisfactory.

It is expected that RIN should have the best performance. Voting integrates the final result by counting votes from each sub-learner directly. Weighting method gives different priority to each sub-learner to reduce the negative effect caused by misjudgments during decision making. However, there might exist some unknown factors that affect the accuracy of weighting method. As RIN can make use most of implicit information, it should have the best performance.

B. Grouping in Non-Interfering Datasets

For most datasets, there exist interferences between different attributes. However there exist datasets with no interference inside (e.g. Cancer dataset from UCI). In this case our ILEL algorithm will not work. Considering the effectiveness of grouping, another scheme that place together attributes with promotion to each other, is designed. To implement this algorithm, it is constrained that attributes with more promotion should be put into the same group. To identify attributes with strong promotion and attributes with weak promotion, we calculate the mean of promotion from every element in the matrix excluding diagonal elements. A pair of attributes is considered to have a strong promotion to each other if their promotion value is greater than the mean; otherwise, it is considered to have a weak promotion.

In grouping attributes, promotion degree has the first priority. First, list all pairs of attributes that has promotion to each other in descending order. Then, consider whether each attribute can be placed into the same group: For each pair, If these two attributes had been grouped, stop and consider the next pair; else if none of two attributes can be placed into an existing group, create a new group for this pair; else If one of the two attributes has been placed into an group, then if the remaining attribute also has good promotion to that group, put the remaining attribute into that group as well. Finally,

after all attributes with good promotion is grouped, create a new group for each remaining attribute.

V. BENCHMARK PROBLEMS

A. Preprocessing of Data

Before training, raw input attributes should be preprocessed. We map all input attributes to floating point number or boolean value within [0, 1]. To map data into boolean value, we refer to input attributes' original meaning. To map input data into floating point number, we introduce the following formula:

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

where

y	Value after normalization
x	Value before normalization
x_{\min}	Minimum value among all data before normalization
x_{\max}	Maximum value among all data before normalization

B. Experimental Results

TABLE VIII: RESULTS FOR GLASS DATASET (%)

Grouping Order	Voting	Weighting			RIN
		Equal	Contribution	Precision	
Random	46.69	47.08	48.96	47.92	33.11
FTS	36.98	34.81	32.92	33.67	32.16
No Grouping	39.90				

TABLE X: RESULTS FOR VOWEL DATASET (%)

Grouping Order	Voting	Weighting			RIN
		Equal	Contribution	Precision	
Random	64.33	37.91	36.55	36.59	23.78
FTS	62.83	34.97	34.25	34.31	23.09
No Grouping	37.69				

Based on the above results, we can see that voting produces the worst performance. It is even worse than no grouping. This is because subgroups do not have enough capacity to produce the final result by voting. Voting has the worst performance on Vowel dataset. As Vowel has only two groups of input, once a group made incorrect decision, the voting result from the two groups will be the same, i.e. one of the two groups will be chosen to obtain the voting result. Thus vowel dataset is more likely to produce incorrect results than the other datasets.

There is no significant difference between three types of weighting method (equal, contribution based and precision based). The result is better when weights are more related to the actual relation among the subgroups. RIN i.e. result integration by network is the best among all. This aligns with our expectation that RIN can learn the relation between subgroups well and produce better integration results.

The attributes which have more contribution have more impact for the final results. Thus putting them together with promoting attributes can make full use of the promotion. The attributes which have more contribution should be considered first for grouping.

Comparing the experimental results, contribution based grouping together with the RIN integration method produces

the best results. The aim of ensemble learning is achieved.

According to the results of non-overlapping experiments, sorting by contribution in descending order and then using network as the integration method is the best.

VI. CONCLUSION

We have proposed an interference-less neural network grouping method: 1) defined a method to measure the interference between attributes; 2) proposed a grouping algorithm to avoid the interference among attributes, this is the key step of creating the interference-less neural network; 3) applied voting, weighting or another network training to integrate the result of sub-network. This approach is shown to be efficient and has potential for future research. Interference-less neural network is also applicable on problems that require feature selection. We plan to apply such interference-less approach to regression problems and other types of machine learning in future.

ACKNOWLEDGEMENT

This research is supported by the National Natural Science Foundation of China under Grant 61070085.

REFERENCES

- [1] A. Al-Ani, M. Deriche, "A New Technique for Combining Multiple Classifiers using The Dempster-Shafer Theory of Evidence," *Journal of Artificial Intelligence Research*, vol. 17, pp. 333-361, 2002.
- [2] J.H. Ang, S.U. Guan, K.C. Tan, et al, "Interference-less neural network training," *Neurocomputing*, vol.71, pp. 3509-3524, 2008.
- [3] D. Angluin, "Computational learning theory: survey and selected bibliography," in *Proce*, the twenty-fourth annual ACM symposium on Theory of computing, pp.351-369, 1992.
- [4] T. Ash, "Dynamic node creation in back propagation networks," *Connection Sci*, vol.1, pp. 365-375, 1989.
- [5] D. Bahler, L. Navarro. *Methods for Combining Heterogeneous Sets of Classifiers*, In 17th Natl. Conf. on Artificial Intelligence (AAAI), Workshop on New Research Problems for Machine Learning, 2000;
- [6] J. Beyer, K.Heesche, et al, "Ensemble Learning for Multi-source Information Fusion," *Studies in Computational Intelligence*, 275(2010), pp. 123-141;
- [7] K. Chen, L. Wang, and H. Chi. "Methods of combining multiple classifiers with different features and their applications to text-independent speaker identification," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 11, pp. 417-445, 1997.
- [8] T. G. Dietterich, *Ensemble Methods in Machine Learning*. In *Multiple Classifier Systems*, Cagliari, Italy, 2000;
- [9] Y. Ding, Q. Li, et al, "Semantic annotation of web data based on ensemble learning and 2D Correlative-Chain conditional random fields," *Chinese Journal of Computers*, vol. 33, 267-278, 2010.
- [10] S. E. Fahlman, C. Lebiere, *The cascade-correlation learning architecture*, Advances in Neural Information Processing systems 2(1990), San Mateo, CA: Morgan Kaufmann, pp.524-532.
- [11] S. U. Guan and P. Li, "A hierarchical incremental learning approach to task decomposition," *J. Intelligent Syst*, vol.12, no.3, pp. 201-226, 2002.
- [12] S. U. Guan and P. Li, "Feature selection for modular neural network classifiers," *J. Intelligent Syst*. vol. 12, no.3, (2002) pp.173-200.
- [13] S. U. Guan and P. Li, "Incremental learning in terms of output attributes," *J. Intelligent Syst*, vol. 13, no.2, pp.95-122, 2004.
- [14] S. U. Guan and S. Li, "An approach to parallel growing and training of neural networks," in *Proc. 2000 IEEE Int. Symp. Intell. Signal Processing Commun. Syst.* (ISPACS2000), Honolulu, HI;
- [15] S. U. Guan and S. C. Li, "Parallel growing and training of neural networks using output parallelism," *IEEE Trans. Neural Networks* vol.13, pp. 542-550, 2002.
- [16] S. U. Guan and J. Liu, "Incremental ordered neural network training," *J. Intelligent Syst*, vol. 12, no.3, pp. 137-172, 2002.

- [17] S. U. Guan and J. Liu, "Incremental neural network training with an increasing input dimension," *J. Intelligent Syst.*, vol. 13, no. 1, pp. 43–69, 2004.
- [18] C. Ji and S. Ma, "Combinations of Weak Classifiers," *IEEE Transaction on Neural Networks*, vol. 8, pp. 32–42, 1997.
- [19] J. Kittler, M. Hatef, et al, "On Combining Classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 226–239, 1998.
- [20] M. Kobayashi, A. Zamani, S. Ozawa, and S. Abe, "Reducing computations in incremental learning for feedforward neural network with long-term memory," in *Proce. IEEE International Joint Conference on Neural Network*, vol. 3, pp. 1989–1994, 2001.
- [21] A. Konstantaras, M. R. Varley, F. Vallianatos, G. Collins, and P. Holifield, "A neuro-fuzzy approach to the reliable recognition of electric earthquake precursors," *Nat. Hazards Earth Sys. Sci.* vol. 4, pp.641–646, 2004.
- [22] M. Lehtokangas, "Modeling with constructive back propagation," *Neural Networks*, vol. 12, pp. 707–716, 1999.
- [23] S. Ma, X. Zhu, *Artificial Intelligence*, Beijing: Tsinghua University Press, 2004;
- [24] H. B. Mitchell, *Image fusion*, Berlin: Springer, (2010), pp.125-142;
- [25] L. Prechelt, "Investigation of the CasCor family of learning algorithms," *Neural Networks*, vol. 10, pp. 885–896, 1997.
- [26] S. Sjogaard, "Generalization in cascade-correlation networks," in *Proc. IEEE Signal Processing Workshop*, pp. 59–68, 1992.
- [27] M. M. B. R. Vellasco, M. A. C. Pacheco, L. S. R. Neto, and F. J. D. Souza, "Electric load forecasting: evaluating the novel hierarchical neuro-fuzzy BSP model," *Int. J. Electro. Power Energy Sy.*, vol. 26, pp. 131–142, 2004.
- [28] S. Weaver, L. Baird, and M. Polycarpou, "Using localizing learning to improve supervised learning algorithms," *IEEE Trans. Neural Networks*, vol. 12, pp. 1037–1046, 2001.
- [29] C. Wong, Y. Li, et al, "Ensemble learning algorithms for classification of mtDNA into haplogroups," *Briefings in Bioinformatics*, vol. 12, pp. 1–9, 2011.
- [30] Z. Xie, W. Hsu, et al, "Mode committee: a Novel Ensemble Method by Clustering and Local Learning," *16th IEEE international Conference on Tools with Artificial Intelligence*, pp.628–633, 2004.
- [31] Z. Zhou, J. Wu, et al, "Ensembling Neural Networks: Many Could Be Better than All," *Artificial Intelligence*, vol.137 (2002), pp.239–263.
- [32] R. Sun and T. Peterson, "Multi-agent reinforcement learning: weighting and partitioning," *Neural networks*, vol. 12, no. (4–5), pp. 727–753, 1999.

Meihua Lir received her M.Sc. from Xi'an Jiaotong University. She is currently an employee of Huawei.

Sheng-Wei Guan received his M.Sc. & Ph.D. from the University of North Carolina at Chapel Hill. He is currently a professor in the computer science and software engineering department at Xi'an Jiaotong-Liverpool University (XJTLU). He is also affiliated with Xi'an Jiaotong University as an adjunct faculty staff. Before joining XJTLU, he was a professor and chair in intelligent systems at Brunel University, UK.

Linfan Zhao is currently an undergraduate student at Xi'an Jiaotong-Liverpool University. He currently served in an NSFC project (61070085/F020508) as research assistant.

Weifan Liis currently an undergraduate student at Xi'an Jiaotong-Liverpool University. He currently served in an NSFC project (61070085/F020508) as research assistant.