European Journal of Scientific Research
ISSN 1450-216X Vol.36 No.4 (2009), pp.639-648

http://www.eurojournals.com/ejsr.htm

Cooperative Neural Network Generalization Model Incorporating Classification and Association

Waleed, A.J. Rasheed
Software Engineering Department, Faculty of Computer Science
Al- Isra Private University, P.O. Box 22, Code 11622, Amman, Jordan
E-mail: Rasheed_wal@ipu.edu.jo

Hamza A. Ali
Software Engineering Department, Faculty of Computer Science
Al- Isra Private University, P.O. Box 22, Code 11622, Amman, Jordan
E-mail: alsewadi@hotmail.com

Abstract

Problem statement: Generalization ability in mapping inputs to rational outputs occupies the primary interest in most of neural networks researches. The main trends for generalization denote training modulation, pattern analysis and structural design. Generalization feature recognition enhancement of neural networks especially for feed forward structural model has limited progress.
Approach: A new approach to promote the generalization ability of neural networks is presented. It is based on the strategy of incorporating classification and association behaviour of neural networks. This approach is achieved through the involvement of two networks, an auxiliary net and main. The auxiliary net with traditional architecture performs dynamic adjustment to the threshold value of the main net, which is constructed in dual-layer architecture.
Results: Experiment results and analysis of the proposed multiple network technique shows effective and acceptable level of recognition although it is achieved at the price of network complexity and computation time.
Conclusions: Despite all of the presented concepts and methodologies in NN literature, it could be deduced that the problem of neural networks generalization still needs much attention and efforts to enhance the performance and upgrade these artificial issues for imitating the natural intelligence of human conduct in which generalization is a trainable conduct rather than an instinct.

Keywords: Neural Network Modeling, Generalization, Psychology Development, Pattern Recognition

1. Introduction

Artificial neural networks (ANNs) serve to process, learn, and predict information using layers of interconnected computational units. For these tasks, an ANN performance depends on its generalization ability, or its ability to recognize trends from training data and employ what it has learned to make predictions on new test data. Nonetheless, ANNs often perform poorly when applied to new cases dissimilar to those they have encountered, a flaw possibly attributed to data anomalies that adversely affect the training process. Despite all of the concepts and methodologies presented in NN literature, it could be deduced that the problem of NN’s generalization is still in need for much attention and efforts in order to enhance the performance and upgrade these artificial issues for simulating the natural intelligence of human conduct. Therefore, it is important to develop methods to improve their generalization ability, since the quality of future predictions on a comprehensive set of all possible data is the ultimate determinant of a network's proficiency. Besides, various parameters that needs to be chosen carefully in order to produce good performance such as network type, size and architecture, besides training step size, stop criterion, learning algorithms, and data representation. The main concerned trends for generalization denote three approaches to feature enhancement of neural networks, they are *training modulation*, *pattern analysis* and *structural design.* Much research has been reported on these three aspects.

With regards to the *training modulation* approach, Caudill and Butler [1], claim that "A neural network is able to generalize", but they provide no justification for this claim, but they completely neglect the complex issues involved in getting good generalization. However, artificial neural networks do not automatically generalize, because generalization requires prior knowledge, as pointed out by Hume [2], Russell [3], and Goodman [4] and rigorously proved by Wolpert [5-7]. In response to this claim the authors incorporated an integrating phase of generalization training in addition to the classical association phase instead of being part of it.

Intensive efforts to promote NN generalization ability has been made by many researchers, for example, Chuanyi and Sheng [8] reported a learning method based on combinations of weak linear classifiers or perceptron. It can do a little more than making random guesses, then combined through a majority vote, resulted into fast training and good generalization performance. Also to improve generalization ability of neural network, Weigand et al. [9] used weight elimination for forecasting application.

As for the *pattern analysis,* Sarle [10] reported some remedies for over fitting, Ishibuchi and Nii [11] used fuzzification of input vector to avoid over fitting while Opitz and Maclin [12] suggested neural network ensemble methods. Recently Wu and Wang [13] improved neural networks learning performance through the use of result-feedback algorithm. For *structural design* on the other hand, Feng et. al. [14] improved generalization ability of NNs by suggested an approach that appropriately shrinks or magnifies input vector. This algorithm is called “Shrinking-Magnifying Approach” (SMA) that finds the appropriate shrinking-magnifying factor (SMF) and obtains a new neural network having better generalization ability, while Ganchev et. al. [15] tackled generalized locally recurrent probabilistic neural networks GLRPNN, for text independent speaker verification. It is contrasted with that of locally recurrent probabilistic neural networks, diagonal recurrent neural networks, Infinite impulse response and finite impulse response MLP-based structures, as well as with Gaussian Mixture Models-based classifier.

Although NNs generalization ability was improved to some extent, however the problem of generalization is generally still not completely solved due to the fact that the principle behavior of artificial neural networks is of instance-based learning. A neural network should learn a relation from limited data and properly respond to unseen input, therefore Inohira et al [16] stated that, it is impossible for NNs to solve all the problems by learning from limited examples, and hence developing new methods for improving NNs’ generalization ability is highly sought. Waleed and Ali [17] suggested a structure based on both Pavlov and Piaget theorems [18] in order to enhance the generalization capability of Feed Forward neural networks. Basically the structure incorporates an extra layer attached to the output layer of a traditional NN with the capability of dynamically adjustable neuronal threshold during both of training and testing phases. The procedure involves two learning cycles; the first cycle stands for Pavlov learning assimilating capability and the second cycle substantiates Piaget arguing through the accommodating capability.

This paper extends the latter work of the structural design approach presenting a novel model of a dual network scheme or cooperative network. This model features a full dynamic layering response instead of single layer characterizing the foregoing work of Waleed and Ali [17]. Two cooperative networks are involved and termed as a main and an auxiliary. The first emerges as the main association net that accepts the inputs and correspondingly generates the related output. The second is utilized as the driving net for the threshold settings of the first net. The implementation of this two subnet structure model embeds two conjugated application tasks of NN; pattern association and pattern classification. The main net stands for the association task whereas the auxiliary net carries out the classification task supplementing the main net. On feeding a pattern, convenient threshold values are determined along the auxiliary net and are applied as bias settings to the main net that drift the activation functions of the scheme and hence properly tuning the resultant output to be generated in agreement with the pattern class threshold space.

2. The Proposed Model

The threshold value of the neuron activation function for any neural network can be interpreted as a pseudo weight adjusted factor connected to a fixed bias that is implied on the network structural scheme. This factor is managed with the other original input connections adaptively in training phases in order to verify the required association between inputs and outputs. Although such consideration facilitates the computation task, especially to what concern the data representation, it keeps threshold arguments out of the scene of the adaptation capabilities that neural networks possess in its major behavior. Hence thresholds of neural activation emerge as variable arguments in the training phases and as fixed constants in testing and normal operations during data retrieval phases.

Patterns in general are mostly described as standard samples corrupted and/or supplemented with noise resulting into creation of a set of classes. Wide range of studies has emphasized the important effects of the proper selection of samples aiming to improve the performance of neural networks in general and its generalization ability in particular. Resultant discussions of literature in this context suggested various techniques and methodologies for pattern selection purposes. It is believed that none of such proposals give any correlation between the structure and patterns, or in other words the effect of patterns selection on the structure as parametric dependency.

In this proposed model, it is envisaged that threshold values can be utilized to adjust the outputs corresponding to the input pattern. Hence for a successful implementation, patterns have to be categorized into their primary classes represented by their typical samples. Furthermore, other samples have to be attributed to the primary classes and are tagged with a distinct threshold values. This mechanism forces the net response to follow input patterns in adapting the threshold in terms of a drifted threshold value instead of being a constant value during the data retrieval phase. With these performance characteristics, the model is made capable of involving two significant properties. The first one is to support input patterns with structural classification parameters while the second one is to make net operation dynamically responsive to the inputs instead of being static scheme. This process will enable the net to work dynamically to generate outputs enhancing the generalization ability.

Figure 1: Overall layout of the proposed cooperative network model.

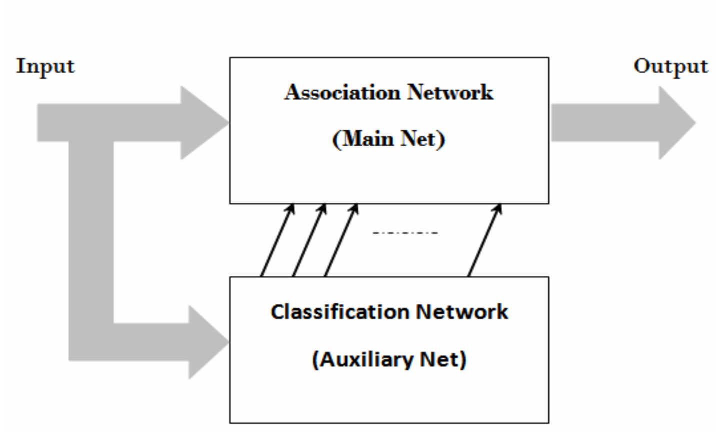
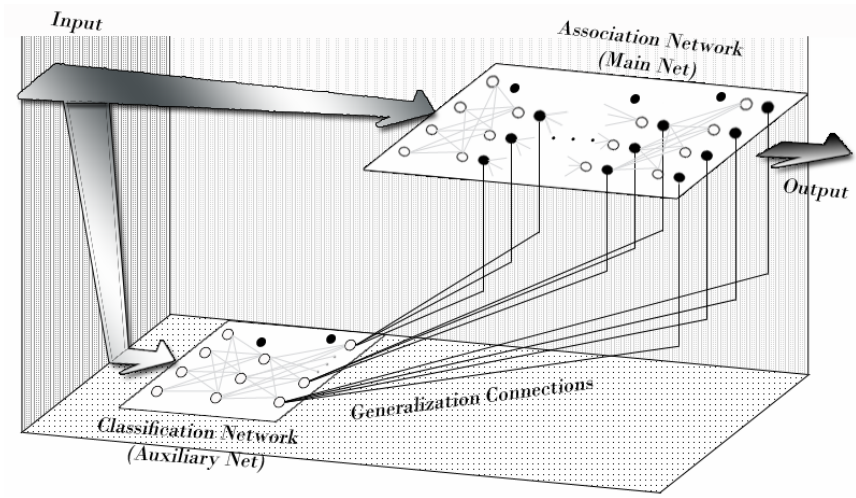


Fig 1 illustrates a block diagram for the overall layout of the proposed two nets model which integrates the above mentioned properties.

3. Model Structure Characteristics

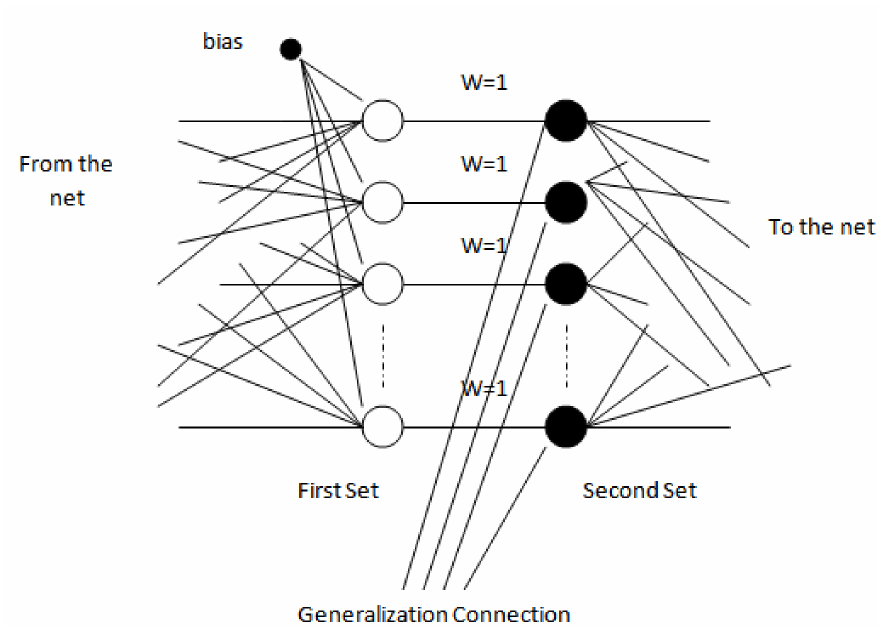
The presented model is composed of two neural networks; main and auxiliary networks. The main network is a feed forward structure that accepts inputs and generates corresponding outputs. Therefore the inputs and the outputs are determined by the applied application dimensions. The major difference of this net from the traditional ones is that its layers are not single set of neurons but they are organized as dual-layer sets. Each dual-layer consists of two sets of neurons except for the input layer which consists of one set only. The neurons of the first set are connected to the neurons of the second set through a full weight link with one to one configuration. Moreover, the first set adjusts its neuron thresholds by standard bias structure where a unity bias feeder is mounted and related connections are reticulated to each neuron. Whereas neurons of the second set are made such that their thresholds are adjusted by bias resources taken from the second network (auxiliary net) and related connections weights are extended from the auxiliary to each neuron. Furthermore, the activation function of the second set in the dual-layer is a linear function whereas for the first set, any activation function can be implemented depending on the design requirements as the case for the traditional networks. This configuration in fact offers the needed compensation to keep the responses within two modes, idle mode and variant mode depending on the type of the input being fed as will be detailed in later section. The general structure of the proposed model is outlined in Fig.2 and details of the dual layering scheme for one layer are illustrated in Fig.3.

Figure 2: The general structure of the cooperative network model



The auxiliary net, as Fig 2 shows, is a traditional scheme. The input signals to this net are the same inputs of the main net, i.e. input patterns are simultaneously fed to the main and the auxiliary networks. The outputs of this net constitute a set of neurons, which are fed successively as bias resources to the second set in successive layers of the main net. Hence the number of output neurons in the auxiliary net equals to the number of the dual-layers excluding the input layer.

Figure 3: The dual layering scheme.



Functionally as the model shows, there are two different responses characterizing each network independently. The main net works as associating network as it relates input-output patterns, while the auxiliary net generates a related output (set of thresholds) to the input patterns, which can be described as a classifier. Therefore, the complete model integrates both, association and classification functions cooperatively. This model facilitates the required training effort that incorporate generalization as training stage supplementing the traditional simple association stage, which might support Caudill and Butler [1] claim.

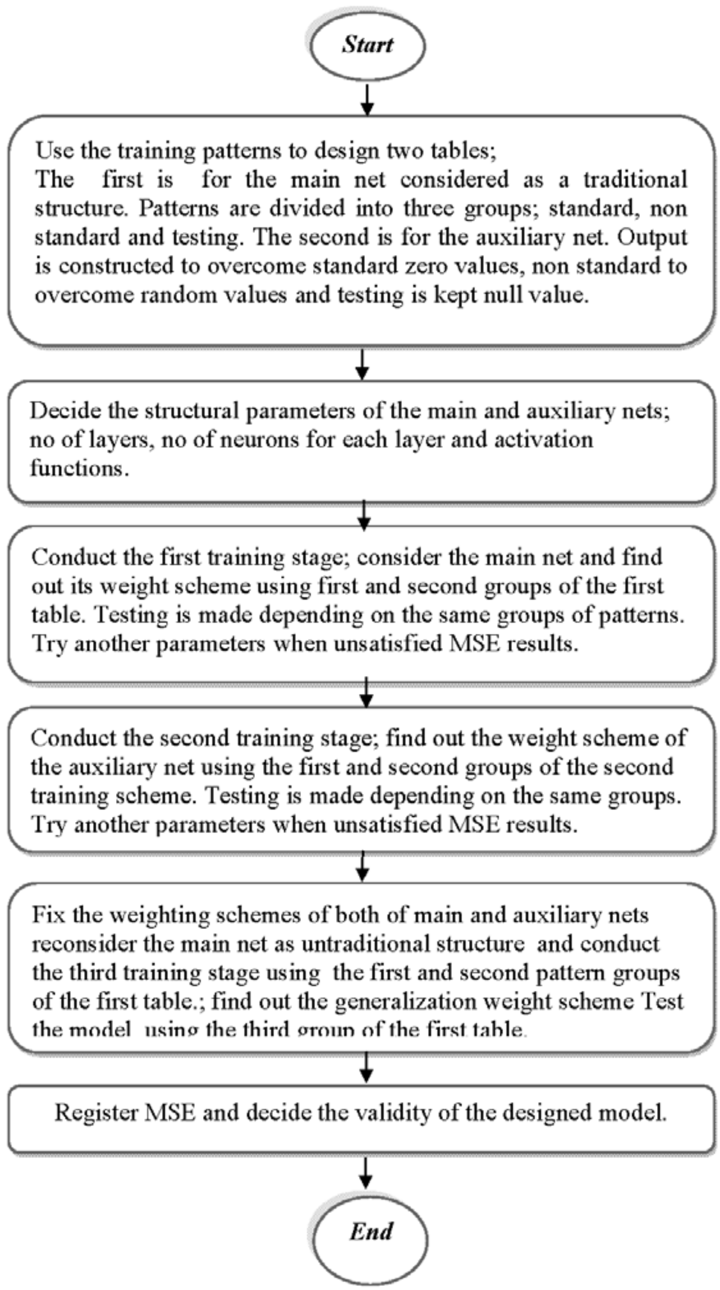
4. Model Structure Adaptation

The complete flow scheme of the proposed model structure adaptation is depicted in fig 4. It gives an account of the sequence of all required processes prior to get the system ready for implementation. The cooperative networks model is designed to implement two different modes of responses, classical response for simple association and a non classical for higher level of generalization response. In the first mode, normal propagation of signals from the input to the output is activated along the reticular formation of the main net. The first set of each dual-layer organization in the main net sums up its inputs applies the activation function and generates the outputs correspondingly. Meanwhile, the second set is switched into an idle state because no drifting for threshold is stimulated due to the full connection weighting linking each neuron to the preceding one of the first set. Therefore the output on the second set is simply identical to the output of the first set in each of the dual-layer configuration. Obviously, the main net in this mode behaves as single layer traditional network.

In the second mode, tuning process for the activation function of the second set is stimulated to show anomalous mode of operation. When the first set of the dual-layer configuration sums up the inputs and applies its activation functions, it transfers the output to the second set. The second set is no longer being in idle mode, because it will drift the threshold in accordance with the generated bias injected from the auxiliary net via the generalization connections. Here, the model acts as an adaptive structure rather than being static. i.e. the threshold values are adaptively changing in accordance with input patterns. This dynamic threshold modification enhances adaptation of the designed network to any input drift away from the standard patterns.

Before initiating the training phase, two separate tables are needed. Obviously the first table fulfills the main network training and the second is needed for the training of the auxiliary net. For the first table, pattern association (inputs and outputs) is divided into three sections. The first stands for standard pattern group, the second for non-standard pattern group and the third for performance measurement as a testing group. The second table is constructed independently in order to adjust the output of the association. In this table, the input patterns have the same number and take the same values of those of the first table, while the outputs number is assigned equals to the number of the dual-layers in the main net. Moreover, the values are set such that; for the standard input patterns, zero outputs are given, for the non standard inputs random numbers are generated whereas null output values are kept for the testing patterns group.

Figure 4: Flow scheme for the proposed model structure adaptation



Training phase of the model is conducted based on these two tables. It is designed to follow three stages using genetic algorithm training techniques.

i). Stage one: Simple Association Training

This stage is concerned with training of the main net, which means computing the weights of the connections of the neurons between the second set of any dual-layer to the first set of the next dual-layer with its many to many configuration. The connection between the first set and the second set in each dual-layer is not involved in the calculations at this stage since it has no influence due to full weight links between these two sets.

To simplify programming task, main net in this stage can be regarded as traditional net with single layer structure. The first two groups (standard and non-standard patterns) of the first table are considered as the needed data for the training purposes. The number of layers selection in the proposed model may follow the same criteria as the case for any traditional application. Furthermore, testing of main net at this stage is done using only these two groups of patterns and the MSE value are calculated. It is worth to mention that minimizing MSE values is not necessary at this stage because they will be dealt with later on during the course of further stages.

ii). Stage two: Threshold Classification training

This stage is concerned with training of the auxiliary net, which means computing the weights of the connections of the neurons between the successive layers of the structure. The first two groups of patterns (standards and non-standard) of table two are used for the training in this stage. Number of layers can be decided with the satisfaction of error reduction (i.e. MSE values) bounded limits used in traditional design of neural networks.

iii). Stage three: Generalization Training

This training stage encompasses the two networks together after fixing their connection weights due to their computation in the previous two stages. The main net here appears with its distinct dual-layer configuration. The only variables sought for are the weights of the connections extended from the output neurons of the auxiliary net to the neurons of the second set of the dual-layers in the main net. Each output feeds all the neurons with separated weight link at the second set.

The computed weights in the preceding two earlier stages together with the association patterns of the first table are implemented. The first two groups of this table again are used for the training in this stage. Whereas, the third group of patterns are implemented to test the validity of the overall model design.

5. Results and Discussion

The proposed cooperative neural network is tested using different combinations of network elements such as various numbers of neurons for input and output terminals, number of hidden layers and number of neurons in each hidden layer. Experimentations with the proposed model have implemented the standard data given by Probin [19].

Genetic algorithm technique is adopted for the network training with different mutations and number of employed generations. It is intended not to exaggerate the number of generations in order to notice the effect of generalization training and to what extent it can compensate for the error. This technique of intentional low adaptation of weights is followed by the authors in order to be able to magnify the effect of the generalization improvements caused by the support of auxiliary net despite the reduced number of generations. It is noticed that to gain the same results with traditional structure and training scheme the number of generations exceeds this number of generations, drastically (i.e. > 3000). The technique of testing and comparing the performance of the embedded traditional network (main net) with cooperative self structural cooperative network is adopted in order to avoid variations in modalities of structure, components, activation functions and training techniques when using different schemes. A number of 300 generations with 100 population size for the traditional net learning is chosen as initial training in all tests understudy while performance measurement of generalization is embodied via the integrating weight adaptation of auxiliary to main connecting weights (or generalization weights). And 20% mutation ratio is used to moderate the searching technique of variable space. Table 1 presents a summary for the conducted experiments, in which different architectural setups were used with different numbers of training patterns.

Table 1: Test parameters for the Cooperative neural network model.

Test	Number of input neurons	Number of hidden neurons	Number of output neurons	Number of generalization generations	Number of patterns used
1	2	40	4	42	80
2	2	23	1	143	55
3	2	29	1	219	52
4	2	35	2	250	60
5	2	50	1	219	52

For each test example, the mean square error, MSE is computed for two different stages. MSE is calculated first for all the test experiments understudy using only the traditional network (main net) for 300 iterations (generations), namely MSE_t, and that for the complete Cooperative network including both main and auxiliary nets, namely MSE_p, as listed in Table 2. Obviously, great improvement in the incorporation of the generalization training has been noticed. Hence, it can be deduced that when the cooperative model is used, a performance gain in the range of (10% - 56.38%) is achieved.

Table 2: Mean Square Error Measurements

Test	MSE_t	MSE_p	Recognition Improvement
1	2.7781	1.8529	33.33%
2	0.1048	0.051	51.34%
3	0.084	0.0756	10. 00 %
4	1.064	0.5318	50.02%
5	2.9203	1.2152	56.38%

As the required number of iteration for reported traditional neural network schemes [20] is in order of 1000s, it can be claimed here that total number of iterations in the model proposed here is far more less. It is clear that the improvements obtained by the implementation of the cooperative neural network model are attributed to the involvement of the classification notion of the auxiliary net that positively affects the association action of the main net. This process exhibits an added value of intelligence boosting the traditional association action, which is in agreement with Claims stated in the introduction section by Caudill and Butler [1].

6. Conclusions

This research work focuses on designing a neural network structure based on high level learning theorems for behavioral development. The procedure involves both assimilating capability of Pavlov and the accommodating capability of Piaget. Therefore, it may be seen as a novel approach that merges psychological learning concept with artificial neural network capabilities in order to improve intelligence representation and performance. Conventional neural network designs goes through learning phase that sets the weights and threshold values, then they stay static afterwards. These designs do not incorporate the intelligent generalization behavior inherently exist in biological neural cells that involve dynamic adjustment of knowledge during the process of recognition. The proposed model has incorporated an auxiliary net that facilitates a continuous adjustment of threshold in order to accommodate variations in the input patterns away from the standard patterns. The resulted generalization improvement reported earlier in this paper can be attributed to the fact that this model is developed based on both functional and behavioral philosophies. It also gives the notion of interpreting both of facts and rules as a structural frame of neural networks, where the generalization connections stands for rule impacts while fact have been modeled within the traditional connection scheme.

References

[1] Caudill, M. and Butler, C., 1990. "Naturally Intelligent Systems", *MIT Press*: Cambridge, Massachusetts.

[2] Hume, D., 1978. "A Treatise of Human Nature", *Selby-Bigge*, L.A., and Nidditch.

[3] Russell, B., 1948. "Human Knowledge: *its Scope and Limits", London*: Routledge.

[4] Goodman, N. (1954/1983), Fact, Fiction, and Forecast, 1st/4th ed., Cambridge, MA: *Harvard University Press*.

[5] Wolpert, D.H. 1995. "The Mathematics of Generalization", *Proceedings of the SFI/CNLS Workshop on Formal Approaches to Supervised Learning,* Santa Fe Institute Studies in the Sciences of Complexity, Volume XX, Reading, MA: Addison-Wesley.

[6] Wolpert, D.H., 1996a. "The lack of a priori distinctions between learning algorithms", *Neural Computation*, 8, pp. 1341-1390.

[7] Wolpert, D.H., 1996b. "The existence of a priori distinctions between learning algorithms", *Neural Computation*, 8, pp. 1391-1420.

[8] Chuanyi J. and Ma, 1997. "Combinations of Weak Classifiers", *IEEE Transaction on Neural Networks*, 8 (1), pp. 32-42. DOI: 10.1109/72.554189.

[9] Weigand A. S., D. E. Rumelhart, and B. A. Huberman, 1991. "Generalization by weight elimination with application to forecasting", *In Advances in Neural Information Processing Systems*. 3, R. P. Lippman, J. E. Moody and D. J. Touretzky, San Mateo, CA: Morgan Kaufmann, pp. 575-582. http://www.icsi.berkeley.edu/~moody/moodypubs.htm.

[10] Sarle W. S., 1995. "Stopped training and other remedies for over fitting", Proceedings of the 27th Symposium on the Interface. http://mars.elcom.nitech.ac.jp/~anto/NEURO/inter95.ps.gz. http://www.d.umn.edu/~rmaclin/publications/opitz-jair99.pdf.

[11] Ishibuchi H., M. Nii, 1998. "Fuzzification of input vector for improving the generalization ability of neural networks", *The International Joint Conference on Neural Networks*, Anchorage, 4th-8th May, Alaska, vol. 2, pp. 1153-1158. DOI: 10.1109/FUZZY.1998.686281.

[12] Opitz O. and R. Maclin, 1999. "Popular Ensemble Methods: An Empirical Study", *Journal of Artificial Intelligence Research*, 11, pp. 169-198.

[13] Yan Wu and S. Wang, 2004. "A New Algorithm to Improve the Learning Performance of Neural Network through Result-Feedback", *Journal of Computer Research and Development* (in Chinese), 41 (9), pp. 488-492. DOI: 10.1109/WCICA.2004.1341928.

[14] Feng N., F. Wang and Y. Qiu, 2006. "Novel Approach for Promoting the Generalization Ability of Neural Networks", *International Journal of Signal Processing*, Waset, 2 (2), pp. 131-135. www.waset.org/ijsp/v2/v2-2-20.pdf.

[15] Ganchev T, D. K. Tasoulis, M. N. Vrahatis, and N. Fakotakis, 2007. "Generalized locally recurrent probabilistic neural networks for text independent speaker verification", *Proceeding of the EuroSpeech*, 3, pp. 1673–1676. DOI: 10.1016/j.neucom.2006.05.012.

[16] Inohira E, T. Uoi and H. Yokoi, 2008. "Generalization Capability of Neural Networks for Generation of Coordinated Motion of Hybrid Prosthesis with a Healthy Arm", *International Journal of Innovative Computing, Information and Control*, 4 (2), pp. 471- 484. www.ijicic.org/07-044-1.pdf.

[17] Rasheed W. A. J. and H. A. Ali, 2009. “Generalization Aspect of Neural Networks on Upgrading Assimilation Structure into Accommodating Scheme”, *Journal of Computer Science* 5 (3), pp. 177-183, 2009. http://www.scipub.org/fulltext/jcs/jcs53177-183.pdf.

[18] Piaget, J., 1979. "The Origin of Intelligence in the Child", Routledge and Kegan Paul, *Great Britain*.

[19] Proben1, neural network bench mark Database.Via FTP from ftp.cs.cmu.edu [128.2.206.173] as /afs/cs/project/connect/bench/contrib./prechelt/probwn1.tar.gz or ftp.era.uka.de [129.13.10.90] as /pub/neuron/proben.tar.gz,1996.

[20] Judith Dayhoff, "Neural Network Architectures: An Introduction", *Van Nostrand Reinhold, New York*, 1990, ISBN 0-442-20744-1.