Overcoming limited samples and ambiguity in classification using neural networks : A Synergistic Approach

K. LAVANGNANANDA

School of Information Technology
King Mongkut's University of Technology Thonburi (KMUTT)
Bangkok, **Thailand**

Abstract : One of the factors which contributes to the success in training neural networks is sufficient number of samples in the data set. Identifying the category of the output in a classification task can be difficult when ambiguity occurs. This often happens when there is a competition among outputs with almost equal merits. Synergy of neural networks have been reported in literature under different terms and scopes, there has yet to be an overview study of such approach.

This paper presents an overview of synergistic approach to neural networks. It also systematically classify the approach according to 2 facets. It follows by presenting how a synergistic approach to classification using neural networks can overcome limited number of samples available for training and ambiguity which can occurs in the classification. The classification in this work comprises 4 categories of heart conditions. The short comings of the data set was ideal for investigating a synergistic solution as an alternative to overcome the problems mentioned.

Key-Words : ambiguity / classification / heart data set / limited samples / neural networks / synergy / synergistic neural networks

1. Introduction

Neural networks have been successfully employed in numerous applications which involve classification. One of the factors which contributes to the success in training neural networks is sufficient number of samples in the data set. Ambiguity often occurs in designing the final outcome of classification when more there is a competition among outputs with almost equal merits. Synergy of neural networks are relatively new and various terms have been used in such approach. To date, there has been no report on an overview study or a systematic classification of such works.

This paper begins with a review of related work in which adopt synergism in section 2. It then describes an overview classification of synergistic approach to neural networks in section 3. Section 4 gives brief explanation on heart conditions and the data set used. It follows description by the neural network architecture developed for the classification and shows the result from single neural network. Short comings of the data set is explained in section 5. Section 6 is concerned with the synergistic approach to overcome problems mentioned. Synergistic architectures and results are also shown in the section. This follows by findings and discussion in section 7. Finally, the paper concludes with summary and future work in section 8.

2. Related work

There are several classification techniques exist. Traditional techniques are those used in the field of statistics [1]. Statistical techniques can be effective, especially in classification of parametric tasks. However, they are inappropriate in many non-parametric tasks when variables are highly interdependent and no known mathematical model is yet established. Recent examples of other techniques in Artificial Intelligence (AI) such as rule-based systems [2] and tree classifiers [3] have been proven useful in such circumstances. However, rule-based system assumes prior knowledge in order to construct heuristics. Tree classifiers also require prior knowledge to some degrees and selecting relevant order of features to consider can be problematic in itself. Neural networks is a popular technique in AI which has been widely used in classification of non-parametric tasks as prior domain knowledge is less crucial while input features need not be ordered. However, selecting useful features among abundant of features in the data base can be crucial, its relevance to this work is discussed section 4.2.

Synergy of neural networks are relatively new, they have been employed in various applications, ranging from forecasting to diagnosis. Its application in classification is prolific and several names have been used for such strategy, such as *committee networks* [4], *modular networks* [5], *mixture of experts* [6], *hybrid systems* [7], *composite systems* [8], and *competitive networks* [9]. There are three main considerations in designing synergistic neural networks, these are type of classifiers, number of classifier units and method of combing the inputs from individual units. This is well discussed in [10].

Works in other fields in AI which adopted synergism include the attempt to capture knowledge from different experts in a decision making system [11] and the rules generation system which extracts and combines knowledge from different inductive learning units [12]. To date, literature in the area has not reported the application of synergistic neural networks in classification to overcome limited number of samples available for training and testing which is also further constrained by ambiguity in the outputs.

3. Classification of synergistic neural network systems

Synergy is a combined effect of more than one units that exceeds the sum of their individual contributions. It is liken to that of a group of experts, each with different backgrounds and experiences. This makes them produce different solutions to a given problem. Finally, another expert then examines these solutions, and if applicable synthesizes them, to produce the final group solution. Hence, in a synergistic neural network system, a *neural combiner* is therefore needed in order to determine its final output. The first systematic attempt to study synergism in neural networks was that of [10], however the work concentrated on classification methods used rather than identifying the overview of classification using synergy of neural networks. A comprehensive review of the work suggests that synergistic approach to neural networks can be classified according to *input data set* and *final output determination*.

3.1 Input data set

A synergistic neural networks can be classified by its input data set into 2 categories, namely *single data set* and *multiple data sets*.

Single data set

This category of synergistic system uses the same data set in each individual unit. Input data is not decomposed prior to training of each unit. Figure 1 depicts this category.

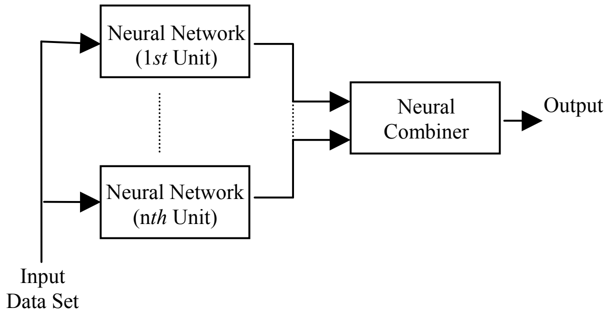

Figure 1 : Synergistic neural network (single input data set)

Multiple data sets

This category of synergistic system may use different data sets in each individual unit. Input data may also be decomposed into different constituents prior to training of each unit. Figure 2 depicts this category.

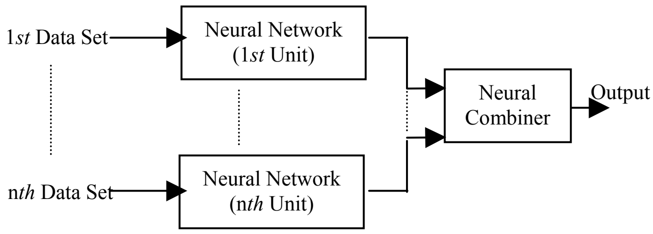

Figure 2 : Synergistic neural network (multiple input data sets)

3.2 Final output determination

A synergistic neural networks can be classified by its final out determination into 2 categories, namely *combination strategy* and *selection strategy*. Given that :

S_j : Final output value of the system

c_{ij} : j^{th} output of i^{th} network

c_{kj} : j^{th} output of k^{th} network

O_{kj} : The median output value of k^{th} network

v_{ij} : the voted value of j^{th} output of i^{th} network

m : the total number of networks.

n : the total number of outputs

Combination strategy

In this strategy, the final output is usually obtained from combining outputs from each units. Two most common methods for combining are *summation* and *product*. Their determinations can be described respectively as follows :

Summation ; $S_j = \sum_{i=1}^{m} c_{ij} \quad (1 \le j \le n)$ ……. (1)

Product ; $$S_j = \prod_{i=1}^{m} c_{ij} \quad (1 \le j \le n) \quad \ldots\ldots (2)$$

Selection strategy

In this strategy, the final output is taken to from what is believed to be the best output among individual units. Three most popular methods for determining the best value are *maximum*, *voting* and *median*. Their determinations can be described respectively as follows :

Maximum ;

$$S_j = c_{ij} \mid c_{ij} > c_{kj} \quad \forall k, k \ne 1 (1 \le i, k \le m)(1 \le j \le n) \quad \ldots\ldots\ldots\ldots (3)$$

Voting ; Output from each network can be determined from :

$$v_{ij} = \begin{cases} 1 & \text{if } c_{ij} > c_{kj} \quad \forall k, k \ne 1 (1 \le i, k \le m)(1 \le j \le n) \\ 0 & \text{Otherwise} \end{cases} \quad \ldots\ldots\ldots\ldots (4)$$

and the final output can be determined from :

$$S_j = \sum_{i=1}^{m} v_{ij} \quad (1 \le j \le n) \quad \ldots\ldots\ldots\ldots (5)$$

Median ; Outputs from each unit are arranged in ascending order. The median valued can be determined from:

$$S_j = O_{kj} \quad (1 \le j \le n) \quad \ldots\ldots\ldots (6)$$

where $k = \frac{m+1}{2}$ the total number of networks

(m)is an odd number.
(k is the mid-value position)

4. Classification of heart data using individual neural networks

This section describes the heart data set used in this work and points out the scope of classification done by previous works on this data set. This follows by description of the neural network architecture used in the classification and the result.

4.1 Heart conditions and heart data set

It must be said that there are various methods for detecting abnormality of heart conditions. Therefore, features used in this work is by no means the only method, but one of the recognized methods, another popular method is the readings from ECG graphs. Recent work classification of ECG graphs by neural networks includes [13]. This work used the heart data set which is available in the public domain [14]. To date, this data set on heart conditions is the most complete and has been used in several works in data mining as a benchmark for testing a program. Examples of such work can be found in [15] and [16]. There are 5 categories of heart conditions in the data set, normality and 4 different severity levels of abnormality. However, records in the most severe case are hardly needed and seldom obtained in practice as it is usually detected well before useful readings are taken since the patients are likely to be in intensive care unit (ICU) or demand constant attention in such condition.

297 samples were available in the data set. The data set is further constrained by limited number of samples in some categories. Previous works which used this data set limited the classification to just 2 categories (normality and abnormality). The short comings of the data set is discussed in section 5. This work is the first attempt to use the data set in training a system to classify 4 categories of heart conditions (two most severed categories are grouped together as one category).

Table 1 shows numbers of samples available in each category.

Table 1 Number of samples in each category

Category	Heart Condition	Number of samples
1	Normality	161
2	Abnormality (severity level 1)	54
3	Abnormality (severity level 2)	34
4	Abnormality (severity levels 3 and 4)	48

4.2 Neural network architecture and classification result

Each sample in the data set comprises 14 fields, 13 of these record the values of 13 features necessary for the classification, the last field records the heart condition category. Description of these 13 features can be found in [14]. The neural network architecture was designed in accordance with input features and output values. Figure 3 depicts the overview of the network architecture for this work.

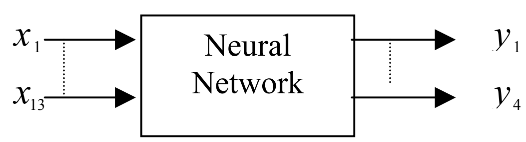

Figure 3 The neural network architecture

Inputs to the network (x_1 to x_{13}) correspond to 13 features needed for classification. Outputs (y_1 to y_4) are normalized (i.e. between 0 and 1), their values correspond to probabilities which the input sample belongs to the category 1 to 4 respectively. Several types of both feed forward and recurrent networks were implemented, trained and tested with the data set. The best result was obtained from Radial Basis Function Networks, this is shown in Table 2.

Table 2 Result from single neural network classification

Accuracy in each category (%)				Overall Accuracy (%)
Category 1	Category 2	Category 3	Category 4	
84.0	20.0	85.7	50.0	66.0

Note that the overall accuracy is not necessary the average value from 4 categories. The value in each category indicates the correct classification of samples which belongs to that class, it is possible to have rather high overall accuracy while accuracy in some categories may be quite low.

Careful analysis revealed that there can be an ambiguity in interpreting the result. Output values (i.e. y_1 to y_4) can show very little discrepancy between them. For example, it is possible for outputs to be $y_1 = 0.10$, $y_2 = 0.54$, $y_3 = 0.63$ and $y_4 = 0.20$. As values of y_2 and y_3 are so close, it may not be prudent to conclude that the input sample belongs to category 3. This ambiguity was a major factor which contributed to poor performance by single network.

5. Short comings of the data set

As shown in Table 1, apart from mediocre number of samples in the data set, it is constrained further by limited number of samples in some categories. Feature selection techniques such as principle component analysis [17] or sensitivity analysis on input features [18] had been employed in similar circumstances. However their objective was more to do with reducing the number of features necessary for the task rather than overcoming limited number of samples and ambiguity in the result. Also inter-relationship can exist among features used.

Therefore, the data set used in this work presents a challenging task in overcoming ambiguity as well as an opportunity to study yet another alternative method where training networks is constrained by limited number of samples.

6. Synergistic approach to classification

This work overcomes the problems encountered by adopting synergistic approach to the design of neural networks system. Synergy of odd number of networks is preferable as it accords with finding the median value and voting methods. Five neural network architectures were chosen for the synergy. They were chosen for their suitability and popularity in classification task and were also networks which yielded better performance in during the experiment described in section 4. These are *multi-layer perceptron* (MLP), *generalized feedforward network*, *modular network*, *radial basis function network* and *jordan network*.

Synergistic neural networks adopts in the work is of single data set type. Both *combination strategy* and *selection strategy* were designed for the final output determination. Figure 4 and Figure 5 depict the synergistic neural networks which adopt the combination strategy and selection strategy respectively. Table 3 and 4 show results of the classification in each category and the overall of synergistic neural networks for both combination type and selection type respectively.

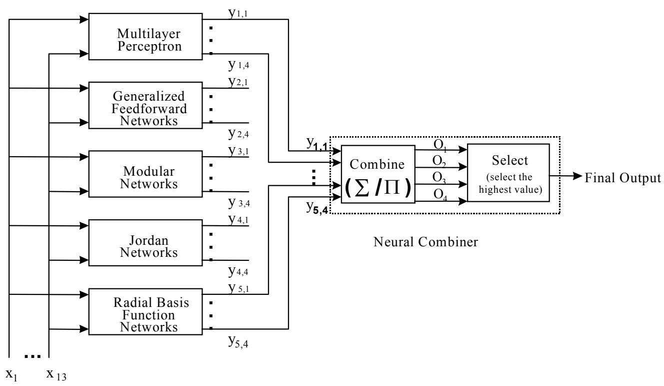

Figure 4 Synergistic neural networks (combination type)

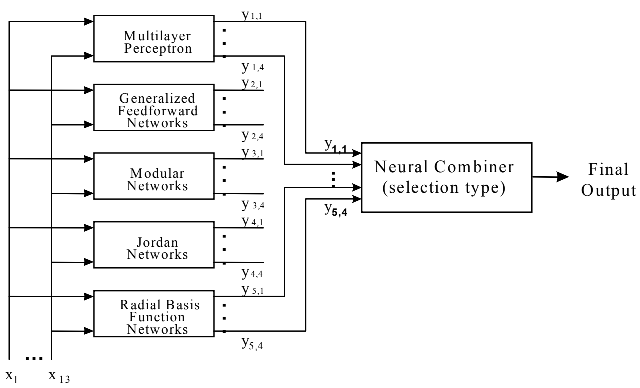

Figure 5 Synergistic neural networks (selection type)

Table 3 Classification results [synergistic neural networks (combination type)]

Combination Strategy	Accuracy in each category (%)				Overall Accuracy (%)
	1	2	3	4	
Summation	*92.0*	*20.0*	*71.4*	*62.5*	*70.0*
Product	92.0	10.0	85.7	62.5	70.0

Table 4 Classification results [synergistic neural networks (selection type)]

Selection Strategy	Accuracy in each category (%)				Overall Accuracy (%)
	1	2	3	4	
Maximum	*96.0*	*30.0*	*85.7*	*87.5*	*80.0*
Voting	92.0	10.0	85.7	87.5	74.0
Median	92.0	20.0	85.7	87.5	76.0

7. Findings and discussion

The best performance was obtained from synergistic networks which adopted the maximum strategy. Combination type of synergistic networks was not much of an improvement this is not somewhat unexpected. Combination type is, by its nature, more appropriate in applications where several outputs are constituents of the final output. While the overall accuracy of 80% cannot be considered very satisfactory in some application, it was a significant improvement from single neural network. Ambiguity cases in some categories were reduced considerably too. Figure 6 shows the comparison between the best results obtained from single network, and from both types of synergistic networks. The work affirms the benefits of synergism in neural networks in overcoming limited number of samples and ambiguity. However, appropriate architecture of each neural network selected for the synergy is crucial. The performance of the synergistic networks depend on contribution from each unit too, hence incorrect outputs from even few units can render the synergistic approach ineffective. Enough number of samples is crucial to training neural networks. Synergistic approach is an improvement to what is insufficient, it should not be considered as panacea to the problem.

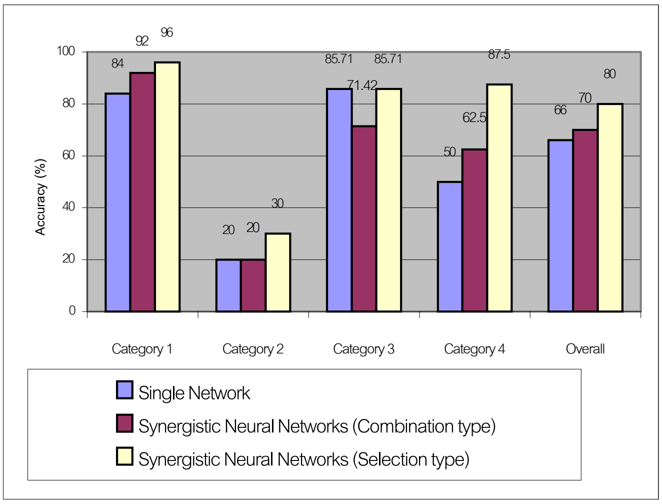

Figure 6 Comparison between the best results from single network, and from both types of synergistic networks

8. Conclusion and future work

Accurate classification can be crucial in medical diagnosis. This work is the first attempt to use neural networks technology to classify well known heart data set into 4 categories. The result may pave way to a more complete and efficient system for diagnosis of heart conditions. Different approaches to synergy of neural networks have been reported in the literature under various names. This paper classifies them systematically according to 2 facets, input data set and final output determination. A synergistic approach to classification using neural networks is suggested as an alternative to overcome, to some degrees, limited number of samples available and ambiguity in the interpreting the final result.

Future work can be carried out in the following areas:

- *Developing a synergy of more neural networks* ; A better improvement may be obtained from inclusion of more networks. Potential candidates include *learning vector quantization* and *cascade correlation network*.
- *Neural network as a neural combiner* ; Another neural network may be used as the neural combiner in the synergistic networks. Neural network may capture some characteristics which mathematical techniques cannot.
- *Introducing more samples* ; An alternative to overcoming limited number of samples is to introduce more samples. For example, duplicates may be created in categories which contain small number of samples. These duplicates can then be randomly mixed into the original to create a balance among number of samples in each category. Recent work had introduced a method known as *clamping technique* [19] where representative

values such as average are introduced into the data set may also be experimented on. A better performance ought to be possible from a synergistic networks with a more balanced data set.

9. Acknowledgement

The author wishes to express his gratitude to Miss. Unhawa Ninrutsirikun for her assistance throughout the project. The support for computing facilities offered by School of Information Technology at King Mongkut's University of Technology Thonburi is appreciated. Preparation of this paper will not be possible without the moral support from the author's wife, Dr. P. Lavangnananda.

References :

[1] Kendall, M. and Stuart, A., *The advanced theory of statistics*, 4th edition, Griffin and Co., London, U.K. 1977.

[2] Cordon, O., Del Jesus, M. J. and Herrera, F., Genetic learning of fuzzy rule-based classification systems cooperating with fuzzy reasoning methods, *Int. Journal of Intelligent Systems*, Vol. 13, 1998, pp. 1025-1053.

[3] Janku, L. et. al. Five approaches to computational personality type classification problem, *Proc. of the Int. ICSC Congress on Computational Intelligence : Methods and Applications (CIMA'2001)*, 19-22 June, University of Wales Bangor, U.K., 2001, pp. 47-53. 2001.

[4] MacKay, D. J. C., Bayesian non-linear modelling for the prediction competition. *ASHRAE Trans*., 1994, 100(2), pp. 1053-1062.

[5] Haykin, S., *Neural Networks : A comprehensive Foundation*, MacMillan, London, 1994.

[6] Jordan, M.I. and Jacobs, R.A., Hierachical mixture of experts and the EM algorithm. *Neural Computation*, Vol. 6, 1994, pp. 818-214.

[7] Goonatilake, S.and Khebbel, S., *Intelligent Hybrid Systems*, John Wiley, Chichester, U.K. 1995.

[8] Pham, D.T. and Oztemel, E., *Intelligent Quality Systems*, Springer-Verlag, Berlin, Germany, 1996.

[9] Romero, et.al., Indentifying disorder regions in proteins from amino acid sequence. *Proc. of the IEEE Conference on Neural Networks*, Vol. 1, Houston Texas, USA., 1997, pp. 90-95.

[10] Pham D.T. and Alcock R.J., Synergistic Classification Systems for Wood Defect Identification**,** *Proc. of IMech Engrs*, Vol. 213, Part E, 1999, pp. 127-133.

[11] Weigend, A.S., et. al., Nonlinear gated experts for time series : Discovering regimes and avoiding overfitting, *Int. Journal of Neural Systems*, Vol. 6, 1995, pp. 373-399.

[12] Lavangnananda, K., Synergistic Genetic Algorithm for Inductive Learning (SynGAIL), *Proc. of the Int. ICSC Congress on Computational Intelligence : Methods and Applications (CIMA'2001)*, 19-22 June, University of Wales Bangor, U.K., 2001, pp. 443-449.

[13] Silipo, R. and Marchesi, C., Artificial Neural Networks for Automatic ECG Analysis, *Transaction on Signal Processing*, IEEE Press, Italy, 1998, pp.1417-1425.

[14] Blake C, Keogh E and Merz CJ., *UCI Repository of Machine Learning Database*. Irvine, CA : University of California, Department of Information and Computer Science, [http://www.ics.uci.edu/~mlearn/MLRepository.html], 1998.

[15] Aha, D.W., et. al., Instance-based Learning Algorithms. *Machine learning*, Vol. 6, 1991, pp. 37-66.

[16] Alcock, R.J. and Manolopoulos,Y., Using Genetic Algorithms for Inductive Learning, *3rd IMACS Int. Multiconference on Circuits, Systems, Communications, and Computer (CSCC'99)*, Athens, Greece, July 4-7, 1999.

[17] Oja, E., Neural Networks, Principle Components and Subspaces, *Int. Journal of Neural Systems*, Vol. 1, 1989, pp. 61-68.

[18] Wang, W., et. al., Assessing the impact of input features in a feedforward network, *Neural Computing and Applications*, Vol. 9., 2000, pp. 101-112.

[19] Wang, W., Identifying salient risk factors by clamping neural networks, *Proc. of the IASTED Int. Conference on Artificial Intelligence & Applications*, Marbella, Spain, 4-7 Sep., 2001, pp. 62-66.