The Third AFSS(1998). 134－138

Unification of Kohonen Neural Network with the Branch-and-Bound Algorithm on Pattern Clustering

Chang-Mok Park[a] , Gi-Nam Wang[b]
a. Dept. of Industrial Engineering Ajou University
441-749, Suwon, Korea
Tel:+82-331-219-2429 Fax:+82-331-219-1610 E-mail:ken00@madang.ajou.ac.kr
b. Dept. of Industrial Engineering Ajou University
441-749, Suwon, Korea
Tel:+82-331-219-2429 Fax:+82-331-219-1610 E-mail:gnwang@madang.ajou.ac.kr)

Abstract.
Unification of Kohone SOM(Self-Organizing Maps) neural network with the branch-and-bound algorithm is presented for clustering large set of patterns. The branch-and-bound search technique is employed for designing coarse neural network learning paradaim. Those unification can be used for clustering or classfication of large patterns. For classfication purposes further usefulness is possible, since only two clusters exists in the SOM neural network of each nodes. The result of experiments show the fast learing time, the fast recognition time and the compactness of clustering.
Keywords: Neural Network, Clustering

1. INTRODUCTION

Automatic optical recognition of an object or object has a wide application in automated manufacturing systems of robotic control, parts inspection, postal code reading and automatic reading devices. In an automated manufacturing and inspection, parts and semi-assembled products are continuously moving on a conveyer line during visual inspection. It is an important and difficult problem to design an intelligent object recognition scheme. In optical object recognition, a object may have various shapes and different position, rotation and scale, therefore designing an automatic object recognition scheme is also a difficult problem..The problem is even harder when the objects have various multi-shape or their images are noisy.

This paper is concerned with a neural network-based approach to object recognition. The hierarchical search technique is employed for designing neural network architecture and learning paradigm. A new two-step identification procedure is proposed for improving network generalization performance. The presented two-step identification procedure consists of a coarse identification and a fine identification. The coarse identification is to find appropriate group in which the object is included. Once the coarse identification is completed, the fine identification is performed to identify the best probable object within a class. The presented coarse and fine procedure could be efficiently used for identifying a large set of patterns.

Neural network is based on a non-parametric classifier instead of a statistical classifier. A classical approach to pattern recognition relies on statistical classification. The Minimum Distance Classification and the Bayesian approach have served as basis for statistical pattern recognition.

The Minimum Distance Classifier computes the distance between a pattern X of unknown classification and a prototype of each class, then assigns the pattern to the class to which it is closest in distance. The Bayesian approach minimizes the average cost of misclassification as well as yields the lowest probability of error. Shortcomings of the statistical method exist, these include limitation of probability distribution, uncertainty of the sample, and difficulty in determining prototype exemplars.

2. LITERATURE REVIEW AND BACKGROUND INFORMATION

Neural Networks approach has the special objectistics that comprise learning. If a class membership is of interest, the system learns from observations of patterns that are identified by class and infers a discriminate for classification. The twin processes of generalization and specialization are all-important in neural networks.

Generalization enables a pattern-recognition system to function completely throughout pattern space, even though it has learned from observing only a limited body of examples. Specialization allows such a system to recover from error and to improve itself [6]. It can acquire the capability to generalize a specific or limited piece of input to produce an output solution. This capability is important because it allows the system to provide solution output even when it is given incomplete input information. After finishing the process of learning, pattern recognition is performed on the basis of similarity in shape between patterns. It is neither affected severely by deformation nor by changes in size, or by shift in the position of the input patterns. The learning time can be separate from the on-line computation, which results in reducing the on-line processing.

Pattern recognition tasks require the ability to match large amount of input information simultaneously and then generate categorical or generalized output. Neural networks possess these capabilities as well as the ability to learn and build unique structures for a particular problem. Fukushima [9] proposed the Neocognitron model for recognizing hand-printed objects. The local features of the input pattern are extracted by the cells of a lower stage, and they are gradually integrated into more global features. Finally, each cell of the highest stage integrates all the information of the input pattern, and responds only to one specific pattern. However, this is the most complicate network ever developed and usually requires a large number of processing elements and connections.

Carpenter and Grossberg simulated on an alphabet learning circuit based on Adaptive Resonance Theory (ART) utilizing a two-thirds rule to allow for self-stability of the network [10]. Carpenter/Grossberg algorithm can perform well with perfect input patterns but even a small amount of noise can cause problems[11]. With noise, there are problems of determining the vigilance threshold and capacity for the stored exemplars.

The incorporation of several neural memories, each coupled with a spatially filtered feature space, was presented by Richard A. Messner and Harold H. Szu [13]. They derived multiple bands of information from an input. The recognition accuracy is increased; however, it requires an excessive amount of time for processing. G. Eichmann and T. Kasparis presented a pattern classification using a linear associative memory. Hough transformation was used as a feature extraction method. Linear associative model replaces tedious clustering algorithms and similarity measure with a stored vector matrix multiplication [2].

Recently, a versatile object Recognition was developed by Rajavelu, Musavi and Shirvaiker. They designed a multishape object recognition system. The Walsh transformation was used to extract features from a object. Backpropagation [1] was applied to this problem. Recognition time and accuracy were considerably improved. However, there existed still some difficulties for recognizing a large amount of various objects (e.g. multi-shape, size, Chinese objects, and Korean objects).

Despite of several existing neural networks based approaches, few methods for determining a set of training exemplars has been suggested. Training set has a significant meaning in that it has a great effect on the networks generalization. The unsupervised clustering mechanism of ART and Kohonen is applied for choosing the exemplars. Automatic self-training procedure is suggested to improve the networks generalization performance.

3. PROPOSED METHODOLOGY

A typical pattern classification algorithm can be divided into two parts [2]; first, the patterns are extracted through preprocessing of input data. Second, these features are compared with each previously stored set of reference features (exemplars) until a match is found (see Fig 1). The proposed methodology consists of three components: preprocessing procedure for feature extraction, a training set generation method, and the two-step learning algorithm.

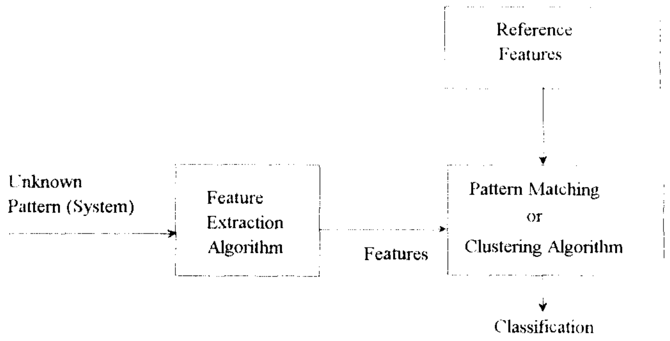

Fig. 1 General Description of Pattern Classfication

3.1 Preprocessing for Feature Extraction

Preprocessing is needed to extract a meaningful feature which categorically defines the details of the pattern. It may be used to orthogonalize the input and reduce data storage by extracting a significant feature from the image. One such preprocessing device is the Walsh transformation which constitutes a set of orthogonal function that belongs to class of piece-wise constant basis functions. It used in communication, signal processing, system analysis, and control. This function is defined as -1 or 1 over the interval [0,1] as followings: the first order of Walsh function is 1 over the interval [0,1], and the second function is defined as 1 over the [0,1/2] and -1 over the [1/2,1]. As the order of Walsh function is increased, each function generates 1 and -1 partitioning the interval [0,1] into subintervals (refer to Figure 2).

Compared with other transformation techniques, the Walsh transformation reduces computation time. The intensity distribution is defined as the number of dark image pixel defined over the subinterval. The Walsh transformation is to multiply the number of dark image pixel by the Walsh function defined over the subinterval, and integrate it. If we use eight Walsh functions, corresponding expansion coefficients can be obtained by this integration.

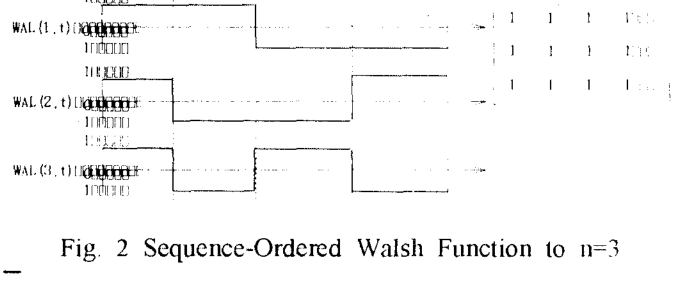

Fig. 2 Sequence-Ordered Walsh Function to n=3

Strictly speaking, the computation is dependent on the number of the Walsh functions. This number effects on the accuracy of feature representation: Figure 2 represents the first four Walsh functions, and the four Walsh functions which have the matrix form. A large number of The Walsh functions are required if a detail image is necessary. It is simpler than the existing transformation used in feature extraction [1]. Depending on the complexity of patterns, preprocessing accuracy can be controlled by the number of the expansion coefficients of the Walsh transformation; Accuracy is increased by increasing the number of The Walsh functions. However, larger numbers of the Walsh functions requires a greater computation. In general, smaller numbers of The Walsh functions is appropriate for subgroups which have low degree of similarity among patterns.

A larger number of the Walsh function is required for subgroups having a high degree of similarity. Another modification for applying the Walsh functions is the mapping procedure of intensity distribution function. It is dependent on the size of object. The reason is that they did not standardize the height of the intensity distribution function expressed as the number of dark pixel even if they normalized the X-axis interval to [0,1]. This makes the problem size dependent. To keep the dynamic range of expansion coefficients consistent for different size objects, the image plane of v(x) is mapped onto the interval defined over [0,+1]. This implies v(x) will have real values in [0,+1] range.

3.2 Two-step Learning Scheme

Kohonen self-organizing model is fundamentally based on the unsupervised learning model, which maps all possible input space to a general solution space forming a stabilized parameter space. In this study, Kohonen model is applied for a unsupervised learning process. A stabilized weight vector is obtained using the training set as an input vector. This weight vector can be utilized to give a general solution for unknown inputs. The generalization of the networks can be increased by the hierarchical decomposition of the wide parameter space into multiple stabilized space. The decision boundary region is generated forming hierarchical binary regions, which results in reducing the misclassification. Basically, hierarchical search algorithm generates all possible subnodes forming the hierarchical tree structure. Lower bound of each subnode plays an important role searching the dominant subnode eliminating the non-dominant search space. Proposed hierarchical search learning algorithm is similar to that in having hierarchical tree structure.

For example, first, 36 exemplar objects (training set) are divided into two subgroups based on the similarity. Group 1 consists of objects (A C H K L M O P S U V X Y Z 1 4 5 7 0). Objects of (B D E F G I J N Q R T W 2 3 6 8 9) are classified as group 2. Kohonen model is used to divide 36 objects into 2 groups. The next step is to divide each group into two lower-level subgroups. The objects included in group 1 are decomposed into group 3 (which the elements are K, L, M, O, P, S, U, Z, 4, 7, and 0) and group 4(which the elements are A, C, H, V, X, Y, 1 and 5). The same procedure is repeated until the variance (total Euclidean distance) within the generated subgroup is less than a specified value.

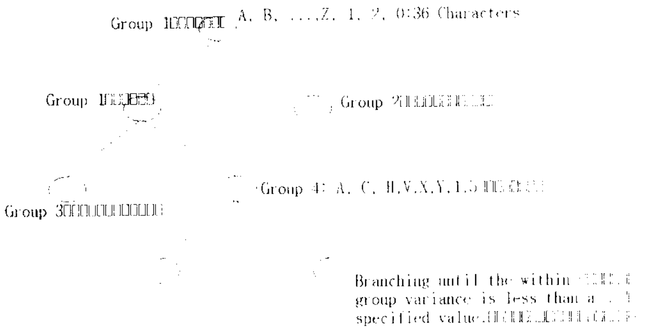

Fig. 3 The generated subgroup

The coarse learning scheme is designed to partition a group into two lower-level subgroups. In the process of generating each subset, a stabilized weight vector is obtained by the Kohonen self-organizing property, which can be interpreted as a training procedure. After generating a subgroup, branching is stopped when the total Euclidean distance within a subgroup is less than a specified value, and end-groups are found. In this case, group 3, group 4, group 5, and group 6 are end groups. They have a different number of objects in the end-groups. The unsupervised learning concept is considered after getting the hierarchically stabilized weight vector. The first step is to identify the corresponding end-group for unknown input objects, and the second step is to identify a specific pattern. In the former process, between two generated subnodes at the same hierarchical level, one node having larger lower bound is fathomed, and this is interpreted as a bounding strategy. The lower bound of each subnode is the Euclidean distance between the unknown input and the stabilized weight obtained from learning procedure. Therefore, only one hierarchical search path is found at every node and the possible search space is very limited. Once identification for an end group is finished, the identification for a object is required. In each end group, find identification procedure may be performed to identify a object. The problem is the misclassification of subgroups. If misclassification for a group occurs before reaching the end subgroup, it reduces the recognition rate and system performance. The training is designed to increase the networks generality, and the proposed twp-step learning algorithm can reduce the misclassification rate for the group by forming binary decision regions hierarchically.

3.3 Proposed Algorithm

The proposed algorithm can be divided by four main procedures: Feature extraction by the Walsh transformation, the generation method of an efficient training set, training procedure and recall procedure. Details can be described as following steps.

Step 1. Feature Extraction via the Walsh transformation. The extracted values are the inputs to neural networks model.
Step 2. Training the Branch and Bound Neural Networks.
a. Generate subgroup using the training set in step 2.
b. Branch stopping rule. Repeat (step 3. a) until the total summation of Euclidean distance within each subgroup is less than a specified value. After reaching end-subgroups, the stabilized weight vector is formed at every subnode. These weights are used to calculate the distance from an unknown object, which is used as the lower bound of the branch and bound algorithm.
Step 3. Recall procedure.
a. For a unknown object, compute the lower bounds (Euclidean distance) of the two possible subgroups. Fathom the subgroup which has the larger lower bound between the two generated subgroups.
b. Repeat (step 4. a) until reaching the corresponding end-group.

The above mentioned algorithm is based on concept of unsupervised learning. Unsupervised learning is used in the clustering of a training set. After finishing the training process, the tree structure is applied for finding subgroups.

4. EXPERIMENTAL RESULTS AND SYSTEM PERFORMANCE

4.1 Experiment I

In this experiment, three kinds of data sets are used as a test data, in which these are generated artificially. The first data sets is following

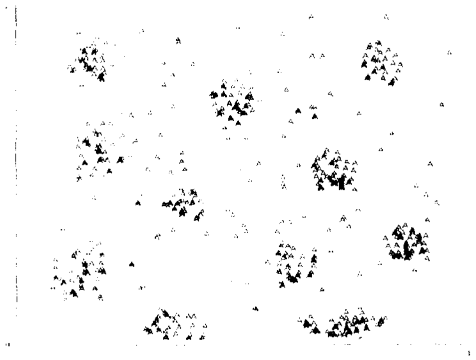

Fig. 4 A Sample data set

The following figure 5 show the clustering time and compactness of clustering according to each different cases. The number of subgroups are 4, 12 and 7 respectively. We compared the result with the clustering which is implemented with the SOM(Self-Organizing Map). The results are shown in Fig. 5, Fig 6.

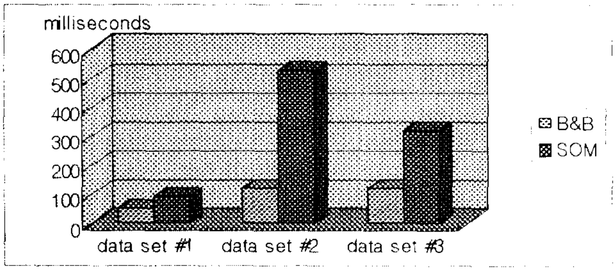

Fig. 5 Clustering time

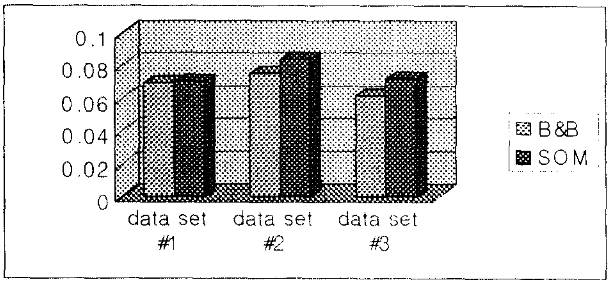

Fig. 6 Sum-of-squared-error

4.2 Experiment II

The other experiment was carried out with 640 kinds of sample images which are generated artificially. For high accuracy, we used the 16 walsh functions. After preprocessing, objects are divided into 100 end-groups using the Branch and Bound learning mechanism. Then we generated nosiy images(5%,10%,20%) and calculate the recognition time and accuracy for each type of noisy images. The result with this mechanism is better than the method of finding orginal images by using MDR(Minimum Distance Recognition). The results are shown in Fig. 7, Fig. 8.

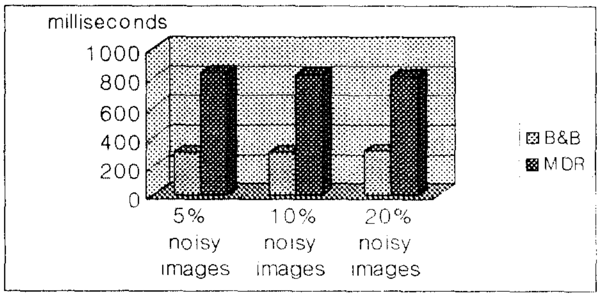

Fig. 7 Recognition Time

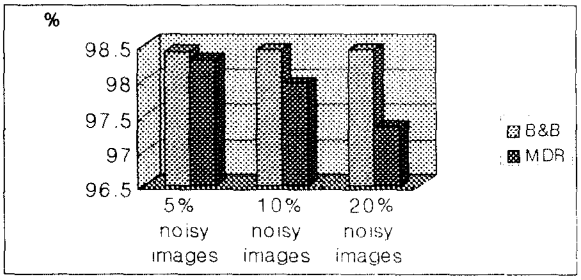

Fig. 8 Accuracy

In fact, the clustering into 100 end-groups using SOM was impossible(the images are converged in a few subgroups because of initial random weights)

5. CONCLUSION

In this paper, two major works are investigated to recognize vector sets and noisy images. In order to compare the performance, Minimum distance recognition is used as a statistical pattern classification. The two-step identification scheme is proposed to reduce computation time and to enhance the recognition accuracy. An efficient method for determining the number of The Walsh expansion coefficient is also explained to increase the system performance. Proposed algorithm might increase the recognition ration and convergence: using the two-step neural mapping, a large set of objects can be categorized into several end-groups according to the similarity.

Unsupervised learning application for neural networks is considered in this mechanism. The advantage of off-line training is that training computation time can be separate from on-line computation by training the network before entering the on-line computation. The information obtained during the training period can be stored and used in the process of recall. Actually, the real computation time may include the recall time. The further study to this project is to implement another preprocessing. Preprocessing is important to increase the recognition rate and system performance. Moment invariant feature might work well under the condition of rotation. As a statistical pattern recognition approach, Baysian classifier may be better than minimum distance classification under certain circumstances. Another further study is to solve the problem of recognizing partially occluded objects using the concept of learning in neural networks. A modification of ART can suggest a solution to this problem only when the object is exactly overlapped but not in general cases. The concept of learning in neural network is utilized to solve this problem by training all possible occlusion. But this is apparently an exhaustive computation. An efficient neural net mechanism might suggest a solution for this problem.

References

[1] A. Rajavelu, M. T. Musavi, and M. V. Shirvaikar, A Neural Network Approach to Object Recognition, *Neural Networks* vol.2, 387-393(1989).

[2] G. Eichmann and T. Kasparis, Pattern Classification using a Linear Associative memory, *Pattern Recognition*, Vol.22, No.6 pp.733-740,(1989).

[3]Laurene Fausett, Fundamentals of Neural Networks, PRENTICE HALL, 1994

[4] Rafael C. Gonzales & Richard E. Woods, Digital Image Processing, Addison-Wesley Publishing Company, 1992.

[5] Leger, James R. and Sing H. Lee, Hybrid Optical Processor for Pattern Recognition and Classification using a Generalized Set of Pattern Functions, Applied Optics, vol.21, 274-287, (1982).

[6] Yoh-Han Pao, *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley (1989).

[7] D. Casasent and P. Demetri, Position, Rotation, and Scale Invariant Optical Correlation, *Applied Optics*, Vol. 15, 1795-1799, July (1976).

[8] D. Casasent and O. Hironobu, A high-Dimensionality Pattern Recognition Feature Space, *Proceding of the SPIE*, vol. 579, pp. 245-257,Sept. (1985).

[9] K. Fukushima, Neocognitron: A Neural Network Model for a Mechanism of Visual *Pattern Recognition*, IEEEE Trans. Sys. Man and Cybern. Vol. SMC-13, No.5, 826-834 Sept/Oct. (1983).

[10] G. A. Carpenter, and S. Grossberg, Associative Learning Adaptive Pattern Recognition and Coorperative-Competitive Decision Making by Neural Networks, *to appear in SPIE Special Series Optical and Hybrid Computing* H.H.Szu ed, (1986).

[11] G. A. Carpenter, and S. Grossberg, Neural Dynamics of Category Learning and Recognition: *Attention, Memory Consolidation, and Amnesia, in J. Davis, R. Newburgh, and E. Wegman (Ed) Brain Structure, Learning, and Memory, AAAS Symposium Series*, (1986).

[12] Yuan-Neng, Hsu et al, Rotation-Invariant Digital Pattern Recognition using Circular Harmonic Expansion, *Applied Optics*, Vol.21, pp 4012-4015, (1982).

[13] R. Messner and Harold H. Szu, An Image Processing Architecture for Real Time Generation of Scale and Rotation Invariant Patterns, *Computer Vision, Graphics and Image Processing* Vol. 31, pp.50-56, July (1985).