DICTA 2002: Digital Image Computing Techniques and Applications, 21—22 January 2002, Melbourne, Australia

A New Pseudo-Hierarchical Neural Network for Pattern Classification

K.Anand, San Murugesan, Athula Ginige
School of Computing and Information Technology
University of Western Sydney Bldg 17, Campbelltown Campus
Locked Bag 179 PENRITH SOUTH DC NSW 1797 Australia
{a.kuppuswami,s.murugesan,a.ginige}@uws.edu.au

Abstract

Pattern recognition and pattern classification have been the active research area for past few decades. Neural networks have evolved into a fine pattern recognition and classification tool. Complexity in training, over-fitting and inability of simple network to classify hard problems are some of the key issues, which hinder their application to real world problems. We propose a new neural network model called pseudo-hierarchical neural network (PHNN), with simple networks as the building blocks. It provides an optimum solution for the problems faced in neural network arena. We propose two new concepts called cropping and positive inheritance, which helps PHNN in forming non-convex decision surfaces. It has been applied to the task of face detection and results are comparable with standard systems available in literature.

1. INTRODUCTION

There is growing interest in image content analysis, given a large number of applications such as image retrieval from image database, face recognition and content-based image coding. The human face could provide useful information about person. Depending on a person's facial expression for instance we could easily guess whether he is happy, sad or angry. But for a computer to do this is an extremely difficult task. Researchers [1] from different branch of sciences such as psychophysicists, neuroscientist, and engineers are trying to figure out how the vision system works in human brain, so that they could apply similar principles in computer-based system.

In the past few decades a number of techniques have been proposed for pattern recognition (PR) and pattern classification (PC). Probabilistic visual learning, based on density estimation in high dimensional space has been proposed by Moghaddam *et.al* [2][18] for PR. In knowledge-based system, the apriory knowledge about the pattern is used for PR. Govindaraju *et.al* [3] have developed a PR tool for frontal face detection based on the knowledge of a human face. Higher order models which include biologically inspired and mathematical model, have also been developed for PR. Heung *et.al* [4] have developed a new PC system for face detection using the concepts of genetic algorithm (GA) and knowledge based rules. Biologically inspired models [5] and hidden markov model [6] is also found in literature for PR and PC. Techniques involving dimensionality reduction and image coding [7][17] have also been proposed for PR task. Neural networks have also been actively used for PR and PC task. Multi-layer perceptron (MLP), bi-directional associative memory (BAM) developed by Kosko [8], adaptive resonance theory (ART) developed by Carpenter and Grossberg [9] have become the most common tools in NN arena for PC and PR. These networks implement major pattern classification paradigms: MLP runs a supervised parameter learning algorithm, Kohonen network performs vector quantisation and ART is motivated by biological relevance and runs on unsupervised learning. Since all networks reflect different principles, commensurate comparison of their performance would be fallacious. Complexity in training, over-fitting [10] and inability of simple networks to form non-convex decision surfaces are the major limitations of these NNs.

We propose a new NN architecture called pseudo- hierarchical neural network (PHNN) for pattern classification. It is a multi-layer architecture built using single hidden layer feed-forward network (SLFN). It has inherent ability to form convex decision surface, overcomes the problem of over-fitting and forms an easy learning class. We also propose new techniques called *cropping* and *positive inheritance*, which helps PHNN in forming non-convex decision surfaces using SLFN.

In this paper, we describe the new PHNN and its application to face detection. In section 2 we describe the pattern classification ability of feed-forward neural networks (FFN). In section 3 we describe in detail the new model proposed. In section 4, an application of PHNN to face detection (FD), the experimental setup has been discussed. Finally conclusions are drawn in section 5. In the appendix we present a detailed discussion about the pattern space formed by SLFN.

2. FEEDFORWARD NETWORKS

FFN, a sub-class of MLP, form the simplest tool for PC and PR. It ranges from a simple linear classifier to a highly complex non-linear classifier. A multi-layer architecture with more than 3 active layers forms non-convex region and can classify any *hard problem* [11]. SLFN, with its inherent inability to classify non-convex pattern space, is classified under the *easy learning class.* A network with more neurons has a higher degree of freedom and assists the network in forming an enhanced estimation function. It has been showed by Jiri Sima [12] that efficient back-propagation algorithm (BPA) is less efficient compared to smaller networks. It has been shown by Guang-bin *et.al* [13] [14] that SLFN can form disjoint decision surfaces, provided the training has been adequate and has enough hidden neurons. Such SLFN is termed as complete single hidden layer feed forward neural network (CSLFN) and networks not satisfying this criteria is called incomplete single hidden layer feed forward neural network (ISLFN). Increasing the number of hidden neurons in CSLFN indicts the problem of over fitting [10] [15] and increases the complexity of training the network.

Assume any arbitrary architecture formed using multiple ISLFNs. It could be any linear or non-linear combination of these networks, which can be replicated using logical operators. For AND logic we form the intersection of convex regions formed by each network. This requires that all the subspaces belonging to class A (for details see the appendix) should be inside the convex region formed by each network. We have proved that the smallest convex region X (for details see the appendix) formed has non-zero error, thus ANDing these regions will also give non-zero error

We could form networks which are trained to classify only m of the d sub-regions belonging to class A, $\forall$ m= 1,2, … d. We could form the union of these regions to get the full pattern space. For all cases except m=1, we have non-zero error (for details see the appendix). For m=1, we have the pattern space to be convex. The case where the distribution of classes in the non-convex pattern space is unknown, it is very difficult to find an algorithm to convert the pattern space to piece wise convex. The lowest case is when m=2, and since error is non-zero for this case, summing these network will also give non-zero error, which is higher than the error given by a MLP, as a MLP can form non-convex decision surfaces. Hence any logical combination of ISLFN cannot give an error lower than a MLP

3. PSEUDO-HIERARCHICAL NEURAL NETWORK

A logical combination of an ISLFN can classify perfectly, if the pattern space can be converted to piecewise convex regions and trained with appropriate set. The process of converting the non-convex pattern space to piecewise convex is called *cropping*. All real world pattern classification problems involve non-convex regions, with no or little knowledge about the pattern space distribution (PSD) and it is very difficult to find a cropping algorithm for such pattern space. A typical example would be a face pattern space, wherein face class forms a non-convex region in between non-face class. Without prior knowledge of PSD, a set of face samples can be collected, but to classify them to appropriate region in pattern space for cropping would be a highly intricate task

We developed a new neural network model, PHNN, working on the principle of *positive inheritance*, where the parent layer does the cropping and the child layer tries to learn in the new subspace created. Since we need to form only convex decision surfaces in the new subspace, an ISLFN would suffice.

3.1. PHNN ARCHITECTURE

PHNN is a multi-layer structure with each layer being an ISLFN, with varying number of hidden nodes. Parameters like number of input-nodes, number of output-nodes, activation function and training procedure being same for all the layers, the whole system depends on the new training technique, a modified form of the original bootstrapping technique developed by Henry *et.al.* [16]

3.2. CROPPING IN PNHH

The process of converting a non-convex pattern space to piece-wise convex is called cropping. A N-layer PHNN is shown in figure 1. In PHNN the first layer forms a convex region consisting of all sub-regions of class A (face, in this case). The second layer is trained from samples collected from this sub-region only and not the whole space. In this new space class B (non-face patterns) lies in-between sub-regions belonging to class A Objective of all subsequent layer is to remove these regions belonging to class B and the resulting space will contain only class A patterns.

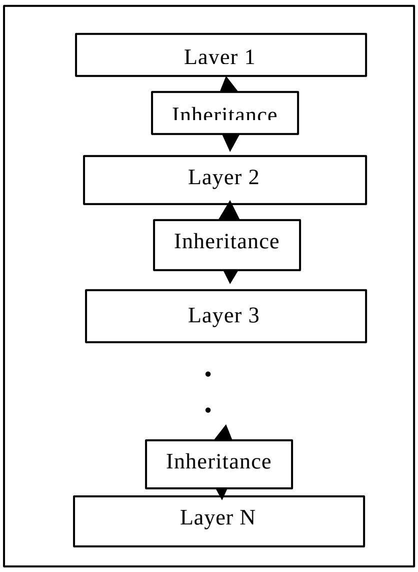

Figure 1. A N-Layer PHNN

Let $\mathbf{P_i}$ be the space from which the training patterns for layer i is selected and $\mathbf{L_i}$ be the convex region formed by layer i in a N-layer PHNN, where i = 1,2…N. The layer 1 after being trained with this set forms a convex region in the pattern space. Let U be universal set obtained by the union of class A and class B samples. The training samples of layer 1 is selected manually from the whole space, hence we have

$$P_1=U \tag{1}$$

$\mathbf{L_1}$ becomes the new pattern space and samples are selected from this space only.

$$P_2 = L_1 \tag{2}$$

Now we form a new training set consisting of large number of samples of class B, samples of class A from the $\mathbf{P_2}$ and append the initial samples of class A to it. Layer 2 is trained with this new set. In this new region $\mathbf{L_1}$, class B occupies region in-between the class A samples and objective of layer 2 and subsequent layer is to remove these regions from the $\mathbf{L_1}$. Layer 2 tries to form a convex region around one of the class B sub-region. This sub-space is removed from the convex region formed and a new pattern space is formed given by

$$P_3=P_2-L_2 \tag{3}$$

A similar procedure is followed for other layers, the pattern space for layer **N** (N > 1) given by

$$P_N = P_{N-1}-L_{N-1} \tag{4}$$

The above equation can be decomposed and expressed as follows

$$P_N = \begin{cases} L_1 - \sum_{i=2}^{N-1} L_i, & \forall N > 2 \\ L_1, & \forall N=2 \\ U, & \forall N=1 \end{cases} \tag{5}$$

Since the first layer detects all the face patterns it will be having the highest detection rate and highest number of false detections. All the subsequent layers remove regions belonging to non-face, the number of false detection decrease. In the process of removal there is also a decrease in the detection rate also.

3.3. MODIFIED BOOT STRAPPING AND POSITIVE INHERITANCE IN PHNN

Let T_i be the training set for layer i, ∀ i=1,2…n. The set T_1 is a manually generated set and we have complete control over it. The layer 1 is trained with this set till it has the highest classification rate, irrespective of high false classification. Then layer 1 is passed through the test phase where, it is fed with test data and a new training set comprising the wrong classification by layer 1 and initial samples of class A is formed. Layer 2 is trained with this set and a similar procedure is followed for subsequent layers.

The whole architecture of PHNN works on the principle of positive inheritance. Each layer except the first and last layer has a parent and child. The first layer has only child and last layer has only parent layer. The inheritance is where each layer learns the properties of parent and appends its own properties and passes this new set of properties to its child layer.

The positive inheritance is a new principle, wherein the child layer inherits only the positive properties or the correct classification of the parent and appends its own properties and passes it to the child layer. The positive inheritance gives a pseudo-hierarchical structure to the network and performs the process of cropping.

3.4. INTERNAL DETAILS OF PHNN

The training set is a growing set from layer 1 to layer 3 or layer 4 depending on the pattern space. After layer 4 most of the class B region is removed, hence after this layer, the training set starts shrinking. The initial samp les of class A is present in all training set, the only variation is the number of class B samples. The training set grows from layer 1 to layer 4, the hidden nodes (required for optimal performance i.e. training time and detection rate) too grow accordingly, but after layer 4 the training set starts shrinking and so does the hidden nodes. This expanding and shrinking phenomenon carves a diamond shape to the whole architecture of PHNN. This process is called as *diamond formation* in PHNN and gives a filtering property to the network.

Key advantages of PHNN are:

- It facilitates easy learning since it uses ISLFN as the building blocks, which comes under the easy learning class.

- It requires lesser training samples as it works on the positive inheritance and cropping principle
- It has an open architecture and offers a better performance with an increase in number of layers.

4. APPLICATION OF PHNN TO FACE DETECTION

Face detection (FD) is a highly intricate task where we try to automate the process of locating the position and size of human face in an arbitrary image. FD is a typical non-linear, non-separable, non-convex, two-class problem and hence it forms an ideal application area for PHNN. Face forms the class A, giving a positive output and non-face forms the class B, giving a negative output.

4.1. EXPERIMENTAL SETUP & RESULTS

In the first stage, principle component analysis (PCA) [17] is performed on the face samples and a modular eigenspace [2] is developed. In the second stage, the image is scanned with a peep window and is projected into the eigenspace and the new coordinates are evaluated [7]. This forms the input to the PHNN, which is trained using the modified bootstrapping and positive inheritance techniques. The whole system was simulated using MATLAB 6.0. The system gives 97 % detection rate and an average of 13.5 false detections per image. Comparison of system with that developed by Henry *et.al* [16], is given in table 1.

Table 1. Comparison of PHNN with the system developed by Henry *et.al* (Std system)

	Std System	PHNN
Detection rate	42/89.9%	45/97.2%
False detection per image	16.5	13.5
Scanning factor	1.2	1.2

5. CONCLUSION

We presented a new PHNN for the pattern classification and it is expected that it will have widespread applications in areas such as face detection. It provides an optimum solution to the problems like over-fitting and complex training procedures faced in NN. It tries to form non-convex decision surfaces with ISLFN, using modified bootstrapping and positive inheritance technique developed. It tries to covert the non-convex pattern space to piecewise convex, using a pseudo hierarchical structure. As the whole model is built with ISLFN blocks, it can be trained with a higher efficiency. PHNN was applied for face detection task yielding better results. Further work on this new model is under progress.

References

[1] S.Grossberg, R.D.S. Rizada, "Contrast – Sensitive perceptual grouping and object based attention in the laminar circuits of primary visual cortex". *Journal of Vision Research 1999*

[2] A.Pentland, B.Moghaddam, T.Starner, "View-based and modular eigenspace for face recognition", *Proc of CVPR '94., 1994 IEEE Computer Society*,Page(s):84 -91

[3] Venu Govindaraju, David B.Sher, Rohini K.Srihri and Sargur N.Srihari, " Locating

human faces in newspaper photograph", *Proc of Computer Vision and Pattern Recognition, 89., IEEE Computer Society* Pp:549 –554

[4] Heung Soo Kim, Eun Yi Kim, Sang Won Hwang, Hang Joon Kim, "Object-based human face detection", *ICCE. 2000 Digest of Technical Papers. International Conference on Consumer Electronics, 2000* Pp: 354 -355

[5] Takshi Hamada, Kazunori kato, Kowichiro Kawakmi, "Extracting facial featrures as in infants ", *Pattern Recognition letters 2000*

[6] Ara V.Nefian and Monson H.Hayes, "Face detection and recognition using hidden markov models', *Proc of Intl Conf on Image Processing, 1998,*Vol: 1 Pp:141-145

[7] R.Gonsalves and R.E.Woods, "Digital Image Processing", Addison Wesley Publishing Company, 1992 Pp: 148-157

[8] Bart Kosko, "Bidirectional associative memory", *IEEE Transactions on Systems, Man and Cybernetics*, Volume: 18 Issue: 1, Jan.-Feb. 1988, Pp: 49 –60

[9] Carpenter, G.A.; Grossberg, S,"A self-organising neural network for supervised learning, recognition, and prediction ", *EEE Comm Magazine*, Vol: 30 Issue: 9 , Sept. 1992, Pp:38 –49

[10] Steve Lawrence and C.Lee Giles, "Over fitting and neural networks: Conjugate gradient and back propagation", *Proc of the IEEE-INNS-ENNS,* Vol: 1, 2000, Pp: 114 -119

[11] B.Yegnanarayana, " *Artificial Neural Network"*, Prentice Hall India, 1999

[12] Jiri Sima, " Back propagation is not efficient", *Academy of Sciences of Czech Republic,* Institute of Computer Science, Page(s):

[13] Guang-bin huang, Yan-Qiu Chen, Haroon.A.Babri, "Classification ability of single hidden layer feed forward neural network", *IEEE Transactions on Neural Networks*, Vol: 11, Issue: 3, May 2000, Pp: 799 –801

[14] Guang-bin huang, Haroon.A.Babri, "General approximation theorem on feed forward networks", *Intl Conf on Information, Communication and SignalProcessing,* 97,Page(s): 698-702

[15] R.Caruana, S.Lawerence and L Giles, "Over-fitting in neural nets: Back propagation, conjugate gradient, and early stopping", *Proc of Neural Information Processing Systems-2000,* Pp:

[16] Rowley, H.A.; Baluja, S.; Kanade, T "Neural network-based face detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol: 20 Issue: 1, Jan. 1998, Pp:23 –38

[17] Sang-Jean Lee, Sang-Bong Jung, Jang-Woo Kwon, Seung-Hong Hong, "Face detection and recognition using PCA". *Proc of the TENCON 99*, Vol:1, 1999 Pp:84 –87

[18] B.Moghaddam, A.Pentland, "Probabilistic matching for face recognition", *IEEE Southwest Symposium on Image Analysis and Interpretation,* 1998 Pp: 186 –191

[19] K.Anand, S.Murugasen, A.Ginige, "A pseudo-hierarchical neural network for face detection",(Submitted to IEEE Transaction on Neural Networks)

Appendix

Let us assume a 2-dimensional pattern space with patterns belonging to either class A or class B. The union of class A and class B gives the universal set. Class A form region in-between class B samples. We need to extract class A from the pattern space. Let there be **J** number of class A sub-regions in the universal space. Let $\mathbf{e_S}$ be the error of a ISLFN when classifying the pattern space having **S** number of class A sub-regions. Then we try to prove that higher the value of **S,** higher will be the value of error.

PATTERN SPACE ANALYSIS OF FFNs

Consider a ISLFN with **D** input-nodes, **H** hidden-nodes and one output-node, $\mathbf{w_{ij}}$ the weight between the i^{th} unit in input layer and j^{th} unit in the layer under consideration and $\mathbf{a_j}$ being the output from

j^{th} unit in input layer . A single perceptron can form a decision surface governed by

$$\Upsilon_j = \sum_{i=1}^{i=D} w_{ij} a_i \quad \forall\, j=1,2,\ldots H \qquad (1)$$

Multiple perceptrons can form multiple decision surfaces. When their output is combined with set of perceptrons forming another layer, forming ISLFN architecture, it forms convex regions. A convex region is one in which a line joining any two points is entirely confined to the region itself. A ISLFN can form convex decision surf aces and can separate convex regions in the pattern space.

Let **U** be the universal space formed by the union of class A and B. Let us divide the regions belonging to class A into **J** disjoint sets denoted by A_i such that

1. Their union gives the region belonging to class A
2. There is no common boundary between neighbouring A_is

Then we have

$$A = A_1 \cup A_2 \cup \ldots A_j \qquad (2)$$

$$B = U\text{-}A \qquad (3)$$

Since all A_i are disjoint, union can be replaced by addition, giving

$$A = A_1 + A_2 + \ldots + A_n \qquad (4)$$

The ISLFN tries to form an open or closed convex region around A or B. Since B is not bounded, it tries to form open or closed convex region $\mathbf{W}$ around class A. Let **f** be the area of the region $\mathbf{W}$, **F** the total area belonging to A and F_i, the area of each A_i, then we have

$$F = \sum_{i=1}^{i=j} F_i \qquad (5)$$

The area covered by region B is given by

$$NF = C\text{-}F \qquad (6)$$

Lets assume that we can form a convex region covering all Ai by following these steps

1. Take two leftmost points L_1 and L_2 in the sub-region belonging to class A
2. Take two rightmost points R_1 and R_2 in the sub-region belonging to class A
3. Take two topmost points T_1 and T_2 in the sub-region belonging to class A
4. Take two bottommost points Q_1 and Q_2 in the sub-region belonging to class A

Assuming the training to be perfect, we can form a closed region with the smallest area, by the intersection of hyper surfaces passing through points L_1L_2, R1R2, T_1T_2, and Q_1Q_2, assuring the region to be convex. If we shrink the region, we can find two points breaking the convexity of the region, proving $\mathbf{W}$ to be the smallest convex region that can be formed by ISLFN. Since we have already proved that $\mathbf{W}$ is the smallest area, that can be formed and F is constant, the value of NF is smallest when the convex region formed is $\mathbf{W}$.

ANALYSIS OF CONVEX REGION W

Let us analyse the convex region Ω formed by ISLFN case wise.

Case J = 1:
There is only one subregion and hence forms only convex pattern in the universal space. So ISLFN can perfectly classify the pattern space. For such space we have the error ε_1 is given by

$$\Omega = F \qquad (7)$$

$$NF = \Phi \qquad (8)$$

$$\varepsilon_1 = 0 \qquad (9)$$

Case J = 2:
Following the steps explained in section 6 we can form a convex region Ω_2, of area ρ_2, such that this region covers the two disjoint subspaces. Then there must be a region, which is not spanned by both subspaces; hence the error ε_2 is non-zero.

$$\varepsilon_2 = \rho_2 - (F_1 + F_2) \qquad (10)$$

Case J = 3:
Following the steps explained in section 6 we can form a convex region Ω_3, of area $\rho 3$, such that this region covers three disjoint and distinct subspaces. Hence there must be a region, which is not spanned by any of the subspaces and the error ε_3 is non-zero.

$$\varepsilon_3 = \rho_3 - (F_1 + F_2 + F_3) \qquad (11)$$

Now $\mathbf{r}_3$ can be considered as the sum of $\mathbf{r}_2$, one more class F_3 and some area between region $\mathbf{r}_2$ and $\mathbf{F_3}$. This gives

$$\rho_3 = \rho_2 + F_3 + \text{Area between } \rho_2 \text{ and } F_3 \qquad (12)$$

$$\Rightarrow \rho_3 - \rho_2 = F_l + \text{Area between } \rho_2 \text{ and } F_l \qquad (13)$$

Hence $\rho_3 > \rho_2$

Similarly for cases **J** > 2, by the mathematical principle of induction, we can prove

$$\rho_k > \rho_i \quad \forall\, k = 1,2\ldots J; \forall\, i = 1,2,..J \ \& \ k > i \qquad (14)$$

As the **J** increases the ISLFN adds more of class B patterns into the convex region formed. Hence the error $\mathbf{e_J}$ is directly proportional to **J** and as **J** increases so does the classification error.