Handwritten Digit Recognition by Neual 'Gas' Model and Population Decoding

Bai-ling Zhang [1]* Min-yue Fu[1] and Hong Yan [2]
[1]Department of Electrical and Computer Engineering
The University of Newcastle, Callaghan, NSW 2308, Australia
[2]Department of Electrical Engineering
The University of Sydney, NSW, 2006, Australia

Abstract *In this paper, we present a handwritten digit recognition scheme using a topology representation model called neural gas. Instead of applying the model only for feature extraction, we train ten separate gas models which aim to describe the data submanifolds respectively in the ten classes. A modular classification system is proposed based on the idea of population decoding, as a topographic map essentially provide a kind of population code for the input. Experimant results show a fast learning and a high recognition rate.*

1. Introduction

Topology preservation is a ubiquitous property in neural system. As one of few real neurobiology motivated models, Kohonen's self-organizing map (SOM) (1) has gained its fame not only in artificial neural networks but also in other fields such as pattern recognition, machine vision etc. As an exploratory data analysis model, SOM forms a nonlinear regression of the ordered set of reference vectors into the input space and the reference vectors form a two-dimensional "elastic network" that represents the distribution of the data.

Kohonen's feature map is a special way for conserving the topological relationships in input data, which also bears some limitations. In SOM, the neighborhood relations between neurons have to be defined in advance. In order to preserve the topological structure of the underlying data manifold and to utilize all the neurons roughly equally, it is desirable that the neighborhood relations between neurons can match the topological structure of the data manifold. In recent years, some alternative learning paradigms for introducing topology-preserving maps have attracted attention. A typical example is the topology representing network called neural 'gas' model(2–3). Martinetz and Schulten use a soft-max rule as an extension of classical k-means clustering procedure and take into account a neighborhood-ranking within the input space. To find the neighborhood-rank, each neuron compares its distance with the input vector to the distance of all the other neurons. Unlike SOM which requires a prior defined static neighborhood relations, neural gas model determines a dynamical neighborhood relations as learning proceeds.

In this paper, we apply neural gas model to handwritten digit recognition, which has become one of the attractive problems in neural network literature in recent years. Many architectures and learning algorithms have been proposed. Many previous neural network approaches for handwritten digit recognition are based on discriminative classifier, which are trained to output one of the ten classes. As a reasonable alternative, (4) addresses the advantages of fitting a separate probability density model to each class and choosing the class of the model that assigns the highest fidelity to a test input. In this paper, we applied the neural gas model to form such a separable representation, *i.e.*, we train a separate network on examples of each digit class and recognise an input digit by determining which network gives the best reconstruction of the data. Our experiment has showed promising results.

2. Neural 'Gas' Model

The idea of neural 'gas' model is partitioning the input space by competition and at the same time ordering the neurons based on a distance metric defined in the input space. Each neuron associates an reference vector to represent an input category. Approximate representation of some geometric structure hidden in the data by these reference vectors (discretizing points) and their arrangement information

*email: bailing@ee.usyd.edu.au

0-7803-4859-1/98 $10.00©1998 IEEE

(connections betwen neighboring points) is essentially modelling the submanifold of the data. Instead of a fixed topology grid as in Kohonen's model, the information about neighborhood relationship among the reference vectors can be implicitly provided by an appropriately defined distortion measure or matching score between a reference vector and the input. In the most simple case, the distortion is defined as the Euclidean distance, i.e., the set of distortions $E_{\mathbf{x}} = \{\|\mathbf{x} - \mathbf{w}(\mathrm{m})\|^2, \mathrm{m} = 1, \cdots, \mathrm{M}\}$. Each time an input $\mathbf{x}$ is presented, we first make an ordering of the elements of E_x and then determine the adjustment of reference vector $\mathbf{w}(\mathrm{m})$. The neighborhood relation is described by certain specific geometrical structure or graph. Such a graph is dynamic in the sense that it has no strict shape or size and can be adapted. In the neural gas model, the graph is Delaunary triangular (DT). As the weight vectors change, the connectivity of the underlying DT graph also changes dynamically. A connectivity matrix C is defined to represent the graph structure, with its elements $C_{ij} \geq 0$. C_{ij} can be considered as virtual lateral connections between neurons. Only if $C_{ij} > 0$ neuron i is considered as being connected with neuron j. Therefore, it is not the absolute value but the sign of C_{ij} is of prime concern in a learning algorithm for representing the neighborhood relationships among the data by C_{ij}.

Each time a data vector $\mathbf{x}$ is presented, we can construct the virtual lateral connection C_{ij} by determining the "neighborhood-ranking" $(E_{\mathbf{x}}(m_0), \mathrm{E}_{\mathbf{x}}(m_1), \cdots, \mathrm{E}_{\mathbf{x}}(m_{M-1}))$ of the distortion set, with $\mathbf{w}_{m_0}$ being closest to $\mathbf{x}$, $\mathbf{w}_{m_1}$ being second closest to $\mathbf{x}$, $\mathbf{w}_{m_k}$, $\mathrm{k} = 0, \cdots, \mathrm{M}-1$ being the reference vector for which there are k vectors $\mathbf{w}_{\mathrm{j}}$ with $\|\mathbf{x} - \mathbf{w}_{\mathrm{j}}\| < \|\mathbf{x} - \mathbf{w}_{\mathrm{m_k}}\|$. The creation of virtual lateral connection between m_0 and m_1 is described by setting $C_{m_0 m_1}$ from zero to one. At the same time, each neuron adjusts its own weight via dynamical learning rate which depends on the ranking of its representation capabilities. Denote the number k associated with each neural unit m by k_m. The following learning rule realizes the neural 'gas' algorithm in[2-3].

$$\Delta \mathbf{w}_{\mathrm{k}}(\mathrm{m}) = \mu_{\mathrm{k}} h_{\lambda}(\mathrm{k_m})(\mathbf{x} - \mathbf{w}_{\mathrm{k}}(\mathrm{m})), \ \mathrm{m} = 1, \cdots, \mathrm{M} \tag{1}$$

where $h_\lambda(k_m)$ is 1 for $k_m = 0$ and decays to zero for increasing k_m. In the simulaiton we choose the same one as in [3], $h_\lambda(k_m) = exp(-k_m/\lambda)$, with λ being a decay constant.

Each connection between nodes i and j has an age t_{ij} that is the number of adaptation steps t the connection already exists without having been refreshed. If the age of a connection exceeds its lifetime T, the connection is removed. Connections have to die out because weight vectors that are neighboring at an early stage might not be neighboring any more at a more advance stage. However, a connection does not die out if it is regualrly refreshed.

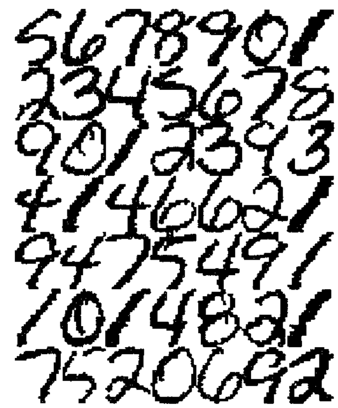

Figure 1: Some handwritten digits samples.

3. Handwritten Digit Recognition

In recent years, neural networks have often been applied to handwritten digit recognition. In most of previous works, a neural network model is mainly used as a classifier which is trained to output one of the ten classes. The shortcoming of such a paradigm has been pointed out in[4]. Another approach is to train an individual autoencoder network on examples of each digit class and then to recognize digits by deciding which autoencoder offers the best reconstruction of the data. Any network can be defined as an autoencoder provided that it has a meaningful internal representation which can be used to appropriately reconstruct input. In this sense, SOM and neural gas model can all be treated as autoencoder model as these population coding models have straightforward decoding strategy for reconstructing input.

Our experiment of classifying images of handwritten digits is proceeded as follows. A subset of 10,000 patterns from the NIST Database were used as training and another 10,000 patterns from different writers were used as testing data. The binary images in the data set are of varying sizes, so we first scaled them to lie on a 25 × 20 pixel grid. Some typical examples are shown in figure 1. The average of the 10,000 patterns was calculated beforehand and then substracted from each example during training. For

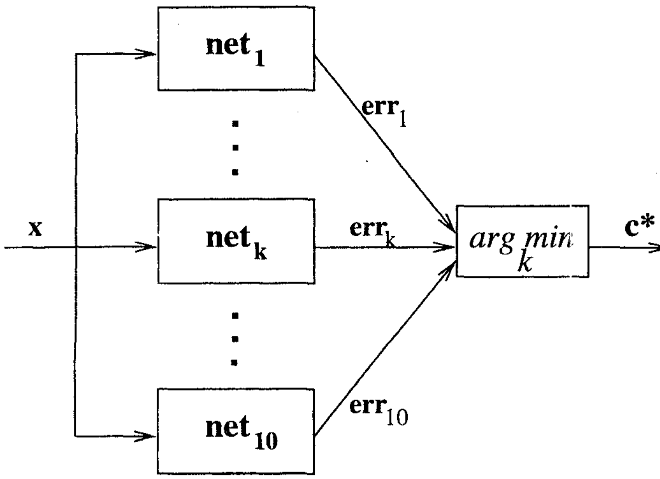

Figure 2: Our proposed modular classification scheme based on the neural 'gas' model or the SOM. In the figure, each net represents a 'gas' model (or SOM).

each class, we train a gas model with $L = 500$ input units and M output units. Different number of M has been tried. During training phase, only images of one class are presented to the corresponding model. In other words, ten networks are trained with examples of the respective class. In this way, a separate network assumes a responsibility for describing separate data manifold. Such a modular classification scheme can be illustrated in figure 2.

Learning can be proceeded in two cycles for the 10,000 training samples. The parameter μ in eqn (1) is initially set to 1 and then linearly decreases to 0.1. In figure 3, we illustrate the converged reference vectors of the ten gas models. We should note that topological (geometrical) structures are provided in the corresponding DT graph, not in the converged reference vectors.

The learned networks can then be directly used as classifiers. Topographic map provides a population coding of representations and the corresponding decoding is often simple and straightforward. In a neural gas model, mth neuron provides a reference vector $\mathbf{w}(\mathrm{m})$ as a partial description of input $\mathbf{x}$ and the most simple complete representation $\hat{\mathbf{x}}$ is the center of gravity of $\mathbf{w}(\mathrm{k})$, $k = 1, \cdots, M$, weighted by corresponding "virtual" activities. A "virtual" activity a_k can be specified as response function of the kth unit after the network has been organized, which can be taken as a Gaussian kernel function, *i.e.*,

$$a_k = \alpha_k \exp(-\frac{||\mathbf{x} - \mathbf{w}(\mathrm{k})||^2}{2\rho_k^2}) \qquad (2)$$

For a given $\mathbf{x}$, we first choose a "winner" reference vector $\mathbf{w}(\mathrm{c})$, then the DC graph C decides which reference vectors are topological neighbors of c. Denote N_x the set of those reference vectors. The height α_k and width ρ_k of kth unit's Gaussian response depends on whether k is in N_x.To the author's knowledge, previous population decoding methods do not consider the topological relationships among the reference vectors[5]. In our experiment, we found that choosing the height α_k and width ρ_k via the topology information apparently improve the model's recognition performance, especially when the size of each network is small.

For each network, the representation vector $\hat{\mathbf{x}}$ can then be written as

$$\hat{\mathbf{x}} = \frac{\sum_k a_k \mathbf{w}(\mathrm{k})}{\sum_k a_k} \qquad (3)$$

When a test image $\mathbf{x}$ is presented to all the ten networks, reconstruction errors err_l, $l = 1, \cdots, 10$, are obtained,

$$err_l = ||\mathbf{x} - \hat{\mathbf{x}}^{(l)}||^2, \quad l = 1, \cdots, 10 \qquad (4)$$

where l indicate the number of network. This squared reconstruction error is a measure of how well a module net describing a digit's bitmap, thus can be considered as an extension of the "distance" in some traditional distance-based classifiers. We execute classification by using a decision module which compare the distance or reconstruction errors (4) between the reconstructed vectors and presented pattern. A minimum operator is the most simple form for associating the class of a model with the smallest error, i.e., we assign $\mathbf{x}$ to the class c^* where

$$c^* = \arg\min_c \; err_c \qquad (5)$$

For comparison purpose, we experimented with different network size M, with different receptive field parameters for classification. The main results are shown in Table 1. More detailed discussion of designing receptive field parameters is beyond the scope of this paper. In experiment, we choose $\alpha_k = 1$ if $k \in N_x$ and 0.5 otherwise. A rough guideline for choosing width ρ_k is $\rho_k < \sqrt{M}/2$, i.e., the sharping tuning kernel functions. Different width ρ_k has been tried, as illustrated in Table 1. From the result we can also find that the larger the network, the more accurate the recognition result. However, as the number of nodes increase in each network, the learning will slow down and the improvement over the recognition rate is not obvious.

Table 1. Recognition accuracy for the neural gas models with different size M and receptive field width ρ. Both training and testing data set has 10,000 samples.

M	9	16	25	36
ρ	0.5	0.5	$\sqrt{2}/2$	2
training set	95.18%	97.06%	98.04%	98.58%
testing set	94.58%	95.6%	96.12%	96.94%

For comparison purpose, we also applied Kohonen's SOM algorithm to our modular handwritten digit recognition system. SOM develops spatial correlations in the network in order for a set of unlabelled data vectors to be approximated by the reference vectors arranged in two-dimensional sheet, i.e., every reference vector $\mathbf{w}_k$ is associated with a location k of a lattice. When a data sample $\mathbf{x}$ is given, not only the best matching reference vector $\mathbf{w}_k$ is adjusted, but also those $\mathbf{w}_l$ with l adjacent to k are updated, with a step size decreasing with the lattice distance between k and l. The learning algorithm is of the form

$$\Delta \mathbf{w}_k(m) = \mu_k h_\sigma(k,l)(\mathbf{x} - \mathbf{w}_k(m)), \ m = 1, \cdots, M \quad (6)$$

where h_σ is the neighborhhod interaction function that decreases monotonically with increasing distance between k and l, which is typically taken as a Gaussian function $h_\sigma(k,l) = \exp(-\frac{d_{kl}^2}{2\sigma^2})$. In our experiment, the parameter σ dynamically changed from maximum 5 to minimum 0.5. In Figure 4, we demonstrate the converged reference vectors of the ten SOM models. From the results, we can find that gas model provides much more variabilities in the converged weights for describing different writting styles. For classification, the population decoding scheme for SOM is same as (3), with the virtual activity a_k defined as

$$a_k = h_\sigma(k, c) \exp(-\frac{||\mathbf{x} - \mathbf{w}(k)||^2}{2\rho_k^2}) \quad (7)$$

The recognition result using the same database is illustrated in Table 2. The SOM algorithm is much slower and the recognition rate is lower. In Figure 5 we demonstrate some typical digits that have not been recognized by both the gas model and the SOM.

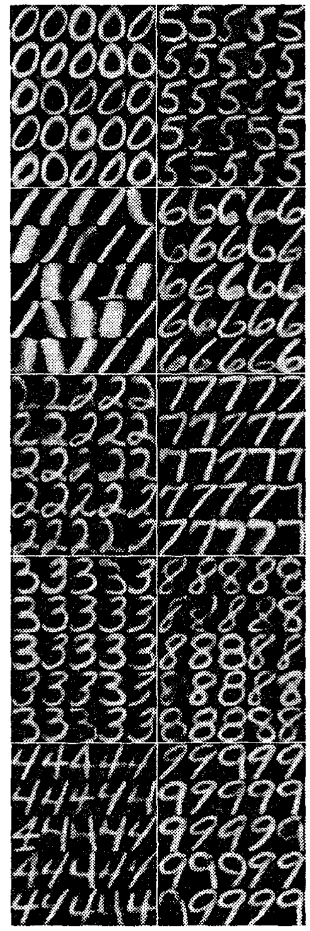

Figure 3: The converged reference vectors of ten neural gas models.

Table 2. Recognition accuracy for SOM models with different size M ($\sigma = 0.5$ in h_σ). Both training and testing data set has 10,000 samples.

M	9	16	25	36
training set	90.41%	92.36%	94.05%	94.73%
testing set	90.35%	92.32%	93.60%	94.38%

4. Conclusions

The neural gas model was previously proposed as an improvement over Kohonen's SOM model for overcoming the disadvantage of fixed topology grid. Compared with the wide applications of SOM, there are relatively few reports on practical applications of the gas model. Our experiment on handwritten digit recognition indicates that neural gas model has a number of advantages, including fast training and high recognition accuracy, provided that an appropriate reconstruction method is assumed. We proposed a population decoding scheme for reconstruction, which incorporate with the topology knowledge. Our neural 'gas' based modular classification scheme demonstrates a very promising result for handwritten digits recognition problem.

Currently, we're further improving our modular classification scheme along several directions. For example, the simplest decision by minimum operator can be replaced by a more reasonable fuzzy decision. Along the same line as in LVQ [6], we can also incorporate classification into learning phase.

References

[1] T.Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, 43, 59-69, 1982

[2] T.Martinetz, S.G.Berkovich and K.Schulten, " 'Neural Gas' network for vector quantization and its application to time-Series prediction," *IEEE Trans. Neural Networks,* 4, 558-568, 1993.

[3] T.Martinetz, K.Schulten, "A 'neural-gas' network learns topologies," in *Artificial Neural Networks,* T.Kohonen *et al* eds., Elsevier Science Publishers, 397-402, 1991

[4] G.E.Hinton,P.Dayan, and M.Revow, "Modeling the manifolds of images of handwritten digits," *IEEE Trans. Neural Networks,* 8, 65-74 1997.

[5] H.P.Snippe, "Parameter extraction from population codes: a critical assessment," *Neural Computation,* 8, 511-529, 1996.

[6] T.Kohonen, "Self-organization and associative memory," Springer-Verlag, Berlin, 1989

[7] Bai-ling Zhang, Min-yue Fu and Hong Yan, "Competitive reconstruction learning (CRL) for autoencoders with applications to handwritten digits recognition," *Proc. Pan-Sydney Area Workshop on Visual Informaiton Processing,* Ed. Hong Yan, 201-206, Dec. 1997, Australia

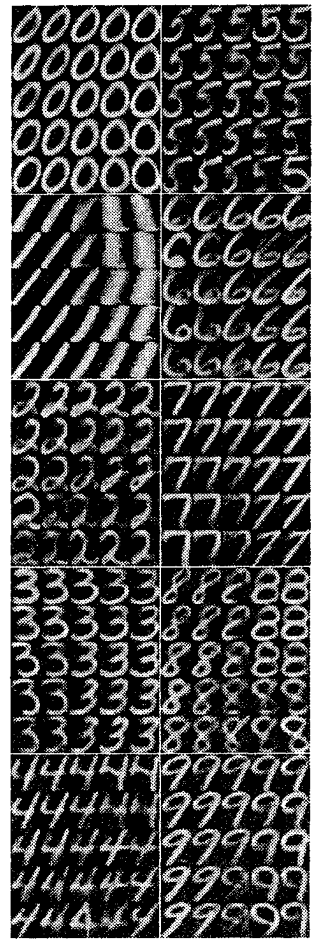

Figure 4: The converged reference vectors of ten SOM models.

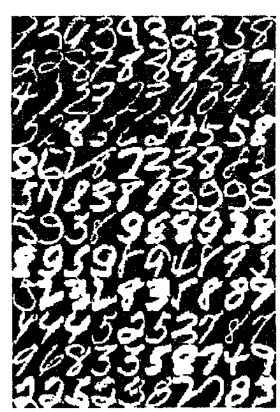

Figure 5: Some digits in the training set that have not been recognized. Some of them appear many times.