A HYBRID ARTIFICIAL NEURAL NETWORK MODEL FOR DATA VISUALISATION, CLASSIFICATION, AND CLUSTERING

TEH CHEE SIONG

UNIVERSITI SAINS MALAYSIA

2006

A HYBRID ARTIFICIAL NEURAL NETWORK MODEL FOR DATA VISUALISATION, CLASSIFICATION, AND CLUSTERING

by

TEH CHEE SIONG

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

January 2006

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Associate Professor Dr Lim Chee Peng for his insightful guidance and encouragement throughout this work. Without his incisive advise, I would not have been able to proceed and bring the research to a satisfactory completion. My appreciation also goes to my co-supervisor, Associate Professor Dr Norhashimah Morad of the School of Industrial Technology, USM, for her constant motivation and invaluable assistance.

I am grateful to Dr Simon Huang, a gastroenterologist of Timberland Medical Centre, Kuching, Sarawak, Malaysia for allowing the use of medical data. Sincere appreciation also goes to the students of Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak for their participation in this research work.

I would also like to thank my family members who provide continuous support and encouragement during the past three years while working on this thesis. My special thanks go to my wife, Chwen Jen, and my children for their patience and greatest company.

Last by not least, I wish to thank my employer, Universiti Malaysia Sarawak, for sponsoring this study, without which this work would have never begun.

TABLE OF CONTENTS

CHAPTER 7 : CONCLUSIONS AND FUTURE WORK

LIST OF TABLES

LIST OF FIGURES

LIST OF NOTATIONS

A	Discrete lattice consisting of N map unit
C_i	Index of class label i
$G^{\mu}(t)$	Total number of active neurons at time t
h_{ij}	Neighbourhood kernel centred on unit i and evaluated at unit j
i, j	Index of neuron on the map grid
i^*	Index of the best-matching unit for a input vector $\mathbf{x}$
I_{max}	Entropy maximisation
$K(.)$	Kernel output function
M	Number of input vectors
N	Number of reference vector (map units)
no_act	Total number of active neurons
OV^j	Overlap Variability at run j
$p(\mathbf{x})$	Input density distribution
$p(\mathbf{w})$	Density estimate at the prototype
$p(\mathbf{v})$	Input density
$p(\mathbf{x} \mid w_i)$	Probability density function of input $\mathbf{x}$
$p(w_i \mid \mathbf{x})$	Posterior probability of input $\mathbf{x}$
$\hat{p}^*(\mathbf{x})$	Fixed kernel estimate
$\hat{P}_{p_S}(\mathbf{x})$	Variable kernel estimate parameterised by parameter ps
P_s	Parameter to control the degree of smoothness of the density estimate

$P_{s_{opt}}$	Optimal P_s value
R^d	Input space with *d*-dimensional Euclidean space
$\mathbf{r}_i$	Location of neuron *i* on the map grid
S_i	Receptive field region at neuron *i*
Sgn(.)	Sign function taken componentwise
t	Discrete time index
$\mathbf{w}_j(t)$	Prototype vector of unit *j* at time *t*
$\mathbf{x}$	A vector in the input space
$\mathbf{x}^{\mu}$	All sample vectors from the input data set
$\sigma_{\wedge 0}$	Initial neighbourhood range
$\sigma_{\wedge}(t)$	Neighbourhood range spanned at time *t*
σ_i	Kernel radius at neuron *i*
$\sigma_{\wedge}^{j}(t)$	Neighbourhood range at run *j* and time *t*
$\eta(t)$	Learning rate at time *t*
μ	Index of the input data set for batch mode training
τ_i	Cross section of the kernel at neuron *i*
ξ_i	Binary code membership of receptive field region activation function at neuron *i*
Ξ_i	Fuzzy code membership of receptive field region activation function at neuron *i*
ρ, ρ_r	Topographic map scale factor
$\Delta\mathbf{w}_i$	Update of kernel centre of neuron *i*
$\Delta\sigma_i$	Update of kernel radius of neuron *i*
$\Lambda(.)$	Neighbourhood function

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
Alb	Albumin
ALP	Alkaline Phosphataste
ALT	Alanine Aminotransferase
ANN	Artificial Neural Network
ART	Adaptive Resonance Theory
AST	Aspartate Aminotransferase
BM2	Boltzmann Machine with binary-coded input vectors
BMU	Best Matching Unit
BP	Back-Propagation
CART	Classification and Regression Tree
CW	Circulating Water
DSS	Decision Support System
EC	Evolutionary Computation
FAM	Fuzzy ARTMAP
FTL	Fine Training Length
GA	Genetic Algorithm
Glb	Globulin
iDSS	intelligent Decision Support System
IGA	Interactive Genetic Algorithm
KE	Kansei Engineering
kMER	kernel-based Maximum Entropy learning Rule
LFTs	Liver Function Tests

LMDT	Linear Machine Decision Tree
LVQ	Learning Vector Quantization
MER	Maximum Entropy learning Rule
minEuC	minimum Euclidean distance
MLPs	Multi-Layer Perceptrons
MNN	Minimum-error Neural Network
MSE	Mean Squared Error
NN	Neural Network
NNb	Nearest Neighbour
OC	Oblique Classifier
OV	Overlap Variability
PCA	Principal Component Analysis
pdf	probability density function
PFAM	Probabilistic Fuzzy ARTMAP
PNN	Probabilistic Neural Network
pSOM-kMER	probabilistic Self-Organising Map - kernel-based Maximum Entropy learning Rule
RBF	Radial Basis Function
RF	Receptive Field
RGB	Red Green Blue
RMSE	Root Mean Squared Error
RTL	Rough Training Length
SOM	Self-Organising Map
SOM-kMER	Self-Organising Map - kernel-based Maximum Entropy learning Rule
TBil	Total Bilirubin
TPro	Total Protein
TL	Training Length
UCI	University of California, Irvine

UCL	Unsupervised Competitive Learning
vIGA	visualised Interactive Genetic Algorithm
viSOM	visualisation induced Self-Organising Map
VK	Variable Kernel
VQ	Vector Quantization
VRML	Virtual Reality Modelling Language
WTA	Winner-Take-All

SATU MODEL HIBRID RANGKAIAN NEURAL BUATAN UNTUK VISUALISASI, KLASIFIKASI DAN PENGKLUSTERAN DATA

ABSTRAK

Tesis ini mempersembahkan penyelidikan tentang satu model hibrid rangkaian neural buatan yang boleh menghasilkan satu peta pengekalan-topologi, serupa dengan penerangan teori bagi peta otak, untuk visualisasi, klasifikasi dan pengklusteran data. Model rangkaian neural buatan yang dicadangkan mengintegrasikan *Self-Organizing Map* (SOM) dengan *kernel-based Maximum Entropy learning rule* (kMER) ke dalam satu gabungan rangkakerja dan diistilahkan sebagai SOM-kMER. Satu siri kajian empirikal yang melibatkan masalah piawai dan masalah dunia sebenar digunakan untuk menilai keberkesanan SOM-kMER. Keputusan eksperimen menunjukkan SOM-kMER berupaya untuk mencapai kadar penumpuan yang lebih cepat apabila dibandingkan dengan kMER dan menghasilkan visualisasi dengan bilangan unit mati yang lebih kecil apabila dibandingkan dengan SOM. Ia juga mampu membentuk peta kebarangkalian setara pada akhir proses pembelajaran. Penyelidikan ini juga mencadangkan satu variasi SOM-kMER, iaitu *probabilistic* SOM-kMER (pSOM-kMER) untuk visualisasi data dan klasifikasi. Model pSOM-kMER ini boleh beroperasi dalam persekitaran kebarangkalian dan mengimplementasikan prinsip dari teori keputusan statistik dalam menangani masalah klasifikasi. Selain daripada klasifikasi, satu ciri istimewa pSOM-kMER ialah keupayaannya untuk menghasilkan visualisasi struktur data. Penilaian prestasi dengan menggunakan set data piawai menunjukkan hasilan pSOM-kMER adalah setanding dengan beberapa sistem pembelajaran pintar yang lain. Berdasarkan SOM-kMER, penyelidikan yang setakat ini bertumpu kepada klasifikasi data, diperluaskan untuk merangkumi pengklusteran data dalam usaha untuk menangani masalah yang melibatkan data tidak berlabel. Satu algoritma

pengawasan penyahkusutan *lattice* digunakan bersama SOM-kMER bagi tujuan pengklusteran berdasarkan ketumpatan. SOM-kMER bersama algoritma pengawasan yang baru ini telah ditunjukkan secara empirikal berupaya untuk mempercepatkan pembentukan peta topografik apabila dibandingkan dengan pendekatan kMER yang asal. Dengan mempergunakan keberkesanan SOM-kMER untuk klasifikasi dan pengklusteran data, penggunaan SOM-kMER (dan variasinya) dalam masalah sokongan keputusan didemonstrasikan. Hasil yang diperolehi menunjukkan keupayaan pendekatan yang dicadangkan untuk mengintegrasikan (i) pengetahuan, pengalaman, dan/atau penilaian subjektif manusia dan (ii) keupayaan sistem komputer untuk memproses data dan maklumat secara objektif ke dalam satu gabungan rangkakerja bagi menangani tugas pembuatan keputusan.

A HYBRID ARTIFICIAL NEURAL NETWORK MODEL FOR DATA VISUALISATION, CLASSIFICATION, AND CLUSTERING

ABSTRACT

In this thesis, the research of a hybrid Artificial Neural Network (ANN) model that is able to produce a topology-preserving map, which is akin to the theoretical explanation of the brain map, for data visualisation, classification, and clustering is presented. The proposed hybrid ANN model integrates the Self-Organising Map (SOM) and the kernel-based Maximum Entropy learning rule (kMER) into a unified framework, and is termed as SOM-kMER. A series of empirical studies comprising benchmark and real-world problems is employed to evaluate the effectiveness of SOM-kMER. The experimental results demonstrate that SOM-kMER is able to achieve a faster convergence rate when compared with kMER, and to produce visualisation with fewer dead units when compared with SOM. It is also able to form an equiprobabilistic map at the end of its learning process. This research has also proposed a variant of SOM-kMER, i.e., probabilistic SOM-kMER (pSOM-kMER) for data classification. The pSOM-kMER model is able to operate in a probabilistic environment and to implement the principles of statistical decision theory in undertaking classification problems. In addition to performing classification, a distinctive feature of pSOM-kMER is its ability to generate visualisation for the underlying data structures. Performance evaluation using benchmark datasets has shown that the results of pSOM-kMER compare favourably with those from a number of machine learning systems. Based on SOM-kMER, this research has further expanded from data classification to data clustering in tackling problems using unlabelled data samples. A new lattice disentangling monitoring algorithm is coupled with SOM-kMER for density-based clustering. The empirical results show that SOM-kMER with the new lattice disentangling monitoring algorithm is

able to accelerate the formation of the topographic map when compared with kMER. By capitalising on the efficacy of SOM-kMER in data classification and clustering, the applicability of SOM-kMER (and its variants) to decision support problems is demonstrated. The results obtained reveal that the proposed approach is able to integrate (i) human's knowledge, experience, and/or subjective judgements and (ii) the capability of the computer in processing data and information objectively into a unified framework for undertaking decision-making tasks.

CHAPTER 1
INTRODUCTION

1.1 Preliminaries

The extraordinarily rapid development of the electronic computer has invigorated human curiosity about the working of the brain and the nature of the human mind. The availability of the computer as a research tool has tremendously accelerated scientific progress in many fields which are important for a better understanding of the brain, such as neuroscience, psychology, cognitive science, and computer science. AI is concerned with making the computer behaves like a human and focusing on creating computer systems that can engage on behaviours that humans consider intelligent. Indeed, the field of AI plays a major role in the theoretical and practical studies on human intelligence.

Conventional computers use an algorithmic approach where the computer follows a set of instructions in order to solve a problem. With such an approach, the computer must know the specific problem solving steps. An ANN, on the other hand, depicts a different paradigm for computing than that of conventional computers. It is inspired by the way biological nervous systems, such as the brain, process information. The distinctive element of this paradigm is that the network comprises a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problem. ANNs, like humans, learn by example. An ANN is configured for a specific application, such as classification or regression, through a learning process. They do not need to be programmed to perform a specific task. Indeed, unlike the algorithmic approach, this ANN computing paradigm depicts its potential in developing intelligent machine that could solve multifaceted problem situations.

In the early days, Rosenblatt (1958) developed the ‘perceptron’, which is an artificial neuron that is capable of performing learning and classification of patterns using simple connections called weights. This is one of the first 'artificial brain models' that had successfully demonstrated the ability to 'learn' from an input data. Subsequently, a series of important developments in the area of ANNs has arisen, such as the discovery of associative memory (Taylor, 1956), model of self-organization of feature detectors (von der Malsburg, 1973), and ordered neural connections (Willshaw & von der Malsburg, 1976). Later, a number of pioneering studies concerning various properties of different ANN models have been published. These include the Hopfield Network (Hopfield, 1982), SOM (Kohonen, 1982), field theory of self-organising neural nets (Amari, 1983), Backpropagation Learning (Rumelhart et al., 1986), and ART (Carpenter & Grossberg, 1987). All these models provide a much more refined depiction of the brain function than what have been anticipated a few decades ago.

It has been known for quite some time that the various areas of the brain, especially the cerebral cortex, are organised according to different sensory modalities (Kohonen, 1984), and the neighbouring neurons or nerve cells in a given area are projected to neighbouring neurons in the next area. The pattern of connections establishes a *neighbourhood-preserved* or *topology-preserved* map, which is similar to repeatedly mapped out of the two-dimensional retinal image in the visual cortex (Van Essen et al., 1981).

Kohonen (1982) proposes an artificial brain model, known as the SOM neural network, that mimics the abovementioned biological phenomena with the assumption that the organisation encountered in many regions of the brain are spatially ordered and in the form of two-dimensional neuron layers. The mathematical developments of SOM that utilise the competitive learning and self-organising process are proposed and have been successfully implemented in a broad spectrum of applications (Kohonen, 1997). Nevertheless, the original SOM model has a number of shortcomings. These shortcomings motivate several researchers to develop modified versions of SOM. Some of the modifications rooted in the basic principles and assumption employed in the original model, i.e., the WTA approach for selecting an active neuron. Later, a number of extended researches that utilise heuristic to adjust the definition of the “winner” based-on “conscience” or recent neuron’s activation history are proposed (Bauer et al., 1996; Van Hulle, 1995). Among others, Van Hulle (1996) proposes an information-based approach that aims at maximising the information-theoretic entropy of the map in order to produce an equaprobabilistic map.

The following section provides an introduction to and definitions of ANNs and topographic map. The problems of the current topographic map formation are presented, and how these problems motivate this research in which a new hybrid ANN model along with a number of its variants is discussed. This is followed by a description of the research scope, the specific research objectives, and the research methodology. An overview of the organisation of this thesis is included at the end of the chapter.

1.2 Artificial Neural Networks

ANNs are computational networks that attempt to simulate, in a gross manner, the networks of nerve cell of human or animal biological central nervous system. Two important aspects of ANNs are (Graupe, 1997):

a. it allows the use of very simple computational operations to solve complex, mathematically ill-defined, non-linear and stochastic problems, and
b. it has a self-organising features and “learning” ability, allowing it to solve for a wide range of problems.

These aspects are very similar to the ability of the human brain to resolve simple problems, such as movement and vision. Several computational formalisms of ANNs are developed to cope with real-world situations. They are mainly considered in the situations such as ill-defined and noisy natural data. Under these conditions, ANN computing methods are more effective and economical than the traditional computation methods. In order to deal with non-stationary data, the properties of the ANNs should be made adaptive. In this case, the performance of the system should improve with use, and the system should be able to capture and store information, or to perform learning. In addition, ANNs should be able to perform generalisation, that is, able to deal with subsets of the problem domain that are not yet to be encountered. All these properties must be appreciated and inherited in the context of ANNs. Without these properties, ANNs are similar to mere 'look-up' tables.

Kohonen (1997) categorises the numerous ANN models into three major categories:

a. Signal-transfer networks

   The output signal values depend uniquely on the input signals. The mapping is parametric and is defined by fixed 'basis functions' that depend on the available unit of neurons. Typical representatives are layered feed-forward networks such as MLPs (Rumelhart et al., 1986), Madaline (Widrow & Winter, 1988), and Radial Basis Functions networks (Broomhead & Lowe, 1988).

b. State-transfer networks

   The feedback and non-linearity are so strong that the activity state very quickly converges to one of its stable values. The initial activity states are set by the input information and the final state represents the result of computation. Examples of such network include Hopfield network (Hopfield, 1982) and Boltzmann machine (Ackley et al., 1985).

c. Competitive learning

   The neurons in the competitive learning or self-organising networks receive information from the input signals. Then, using the lateral interactions in the networks structure, these neurons compete in their activities by selecting the best matching neurons and update the neurons to match the current input signal. Each neuron or group of neurons is sensitised to a different domain of input signals and acts as a decoder of the domain. Typical examples of this type of networks are SOM (Kohonen, 1984) and ART models (Carpenter & Grossberg, 1987). During the self-organisation (or competitive learning) stage, the neighbouring neurons in the network cooperate and code for similar events in the input space whereas more distant neurons compete and code for dissimilar events. This learning scheme, i.e., cooperative/competitive learning, serves as an important rule in the formation of topographic maps.

The following section provides an introduction to the topographic map.

1.3 The topographic map

Today, about 80 cortical areas are known in the human cortex (Ritter et al., 1992). Each of these cortical areas represents a highly parallel "special purpose" module for a specific task. For example, the visual cortex areas are for the analysis of edge orientation, colour, and etc., while other cortical areas host modules for speech comprehension, recognition, spatial orientation and so on (Ritter et al., 1992). Each of these cortical areas is also connected to and interacted with numerous additional cortical areas as well as the brain and nerve structures outside the cortex. These cortical areas basically consist of six layers, which are circuitry connected with one another using a common "topographic" organisational principle: adjacent neurons of an output field are almost always connected to adjacent neurons in the target field (Ritter et al., 1992). Due to the preservation of adjacency and neighbourhood relationships, this mapping can be regarded as a topographic map.

These observations have led researchers to model an element of self-organisation implicated in topographic map formation, driven by correlated neural activity and intended to improve the precision of the existing but coarse topographic ordering (Willshaw & von der Malsburg, 1976). This specification occurs in the same topological order that describes the similar relations of the input signal patterns. As the map is often in a 2-dimensional space, it implies that the topology-preserved map is also performing dimensionality reduction of the representation space.

The SOM network and some of the related models (e.g. vector quantisation) are inspired by these biological effects and the abstract self-organising processes, in which maps resembling the brain maps are formed using mathematical processes and ANN concepts. The first computer simulations to demonstrate a self-organising process that involves synaptic learning for the local ordering of feature-selective cortical cells were conducted by Von der Malsburg (1973). Subsequently, his model serves as a source of inspiration for many other orientation selectivity models (Grossberg, 1976; Amari & Takeuchi, 1978).

One of the architectures worth mentioning is the Kohonen's model that explains the criterion of self-organisation using the competitive learning scheme in topographic map formation. The SOM model encompasses two stages of operation: the competitive stage and the cooperative stage. In the first stage, the "winner" (competition) of the neurons is selected, and in the second stage, the weights of the winning neuron are adapted as well as those of its immediate lattice neighbour (cooperation). The neighbourhood function plays an important role in the cooperative stage. It is essential for the formation of a topology-preserved mapping, which can be interpreted as a statistical kernel smoother (Van Hulle, 2000). However, topological defects occur owing to the rapid diminution of the neighbourhood range during the topographic map formation process. Besides, if non-square distribution of input data is used to map on the 2D square lattice of SOM, then topological mismatches occur, which can result in "dead" neurons (neurons that have low probability to be active) in the model.

1.4 Problems and motivation

From the theoretical point of view, neural models which produce a topology-preserved map are more alike to the theoretical explanation of the brain map than traditional ANN models. According to Kohonen (1997), the relevant items and their categories of the topographic map are located spatially close to each other. This shortens the communication link for information sharing. In addition, the weights/prototype data of the neurons in the topographic map are segregated and clustered spatially, in order to avoid cross talk. This self-clustering feature is essential toward the formation of cluster regions which could be used for feature extraction, dimensionality reduction, and visualisation. Therefore, the converged topographic map are localised, clustered, and ordered, and it is very useful for data modelling purposes. In this case, depending on the applications, the topographic map can be used a variety of tasks such as regression analysis, feature extraction, dimensionality reduction, data visualisation, as well as classification and clustering analysis.

The topographic map produced by SOM has two important characteristics. First, the weights of SOM are regarded as a representative sample of the data. Since the map grid is an ordered representation of the data, the neighbouring regions on the map are similar to each other but dissimilar with other far away regions. Second, the weight formed by SOM is a model of the data. The weights can be used to determine the probability density estimation of the input data. However, topological mismatch could occur in the SOM model. On the other hand, an essential ingredient of SOM is the neighbourhood function, which is responsible for generating a topology-preserved quantisation region. The decreasing range of the neighbourhood function results in a time-varying characteristic of the weight distribution of the neurons. This also produces neurons that are out of the distribution support region and have zero (or very low) probability to be active, which are known as ‘dead’ units (Van Hulle, 2000).

The weight density achieved by SOM at convergence is not a linear function of the input density (Kohonen, 1995). In addition, SOM tends to undersample high probability regions and oversample low probability ones (Van Hulle, 2000). For applications such as density-based clustering and Bayesian classification, a model for estimating the probability density function underlying the training samples is needed. Nevertheless, SOM is not intended to model the fine structure of the input density (Kohonen, 1995). Owing to this limitation, SOM is not able to provide a "faithful" representation of the probability distribution that underlies the input data (Van Hulle, 2000).

The information preservation model (Linsker, 1988) is based on a learning procedure that is founded on the maximum information preservation principle for topographic map formation. Based on this principle, a probabilistic WTA network that maximises the Shannon information rate (average mutual information) between the output and the input signal can be obtained. Deriving from the similar idea, Van Hulle (1995) proposes a more intuitive approach to build a topographic learning rule that maximises the information-theoretic entropy directly. Several learning rules are proposed, such as Maximum Entropy learning Rule (Van Hulle, 1995) and Vectorial Boundary Adaptation Rule (Van Hulle, 1996) that use the lattice quadrilaterals as the RFs. However, for this type of RFs, the lattice topology is rectangular and its dimensionality is the same as that of the input space. As a result, it cannot be used for non-parametric regression and dimensionality reduction purposes as the definition of the quantisation region is too complicated, thus impractical (Van Hulle, 1999).

A new learning procedure known as the kMER (Van Hulle, 1998; Van Hulle & Gautama, 2004) model is proposed. The RFs of this kind are defined using an individually adapted kernel that performs local smoothing of the interpolation function, which is defined by the sum of all RF kernels. The limitation of lattice-based RFs (i.e., Maximum Entropy learning Rule and Vectorial Boundary Adaptation Rule) is relaxed; hence kMER can be used for non-parametric regression and dimensionality reduction purposes (Van Hulle & Gautama, 2004).

The kMER model, on the other hand, exhibits limitation in terms of its computational efficiency. In the self-organising learning process, especially in the competitive learning stage, the neighbourhood relation is essential in the formation of the topology-preserved mapping (Kohonen, 1995). In kMER, the non-uniform RFs of neighbouring neurons overlap during the initial learning stage, and more overlapping occurs at the early stage of the learning process. This leads to computational inefficiency and slows down the formation of the topographic map, and subsequently increases the processing time. Indeed, this problem is significant in practice as real-world datasets are often huge and complex.

On the other hand, the learning rate of kMER is normally set to a small value (Van Hulle, 2000). This is to ensure that the average RF centres obtained at convergence represent the (weighted) medians of the input samples that activate the respective neurons. It also allows the average RF radii obtained in an equiprobabilistic manner to activate the neurons. Owing to the small learning rate, the map formation using kMER needs a long training time, and requires many training epochs.

On the contrary, SOM does not define any RF region during the topographic map formation. The local enhancement of the winning neuron’s activity is usually the WTA operation, which is a global one and allows for only one “winner” at a time when an input pattern is given. The batch map SOM (Kohonen & Somervuo, 2002) is a variant of SOM that uses a fixed-point iteration process to accelerate the topographic map formation. In this way, instead of using a single data vector at a time, the whole dataset (batch learning) is presented to the map before any adjustment is made. It provides a considerable speed-up to the original SOM training procedure by replacing the incremental weight updates with an iterative scheme that sets the weight vector of each neuron to a weighted mean of the training data.

In this research, a hybrid approach is adopted to formulate a new topology-preserving map model that is able to harness the benefits offered by both SOM and kMER, while mitigating their limitations. The converged topographic map is then applied to various problem domains such as data visualisation, non-parametric estimation of the input probability density, classification and clustering analysis, as well as decision support.

Many structured and semi-structure problems are so complex that they require expertise for their solutions. In this aspect, ANN techniques can be deployed as inference engines and decision support tools in many applications. These applications are able to demonstrate the possibility of combining the best capabilities of both humans and computers in the decision-making process. However, computerised intelligent systems and humans are viewed as two entities that work separately in the whole decision-making process. This leads to the research conducted in this work whereby a novel system capitalising on the advantages of both SOM and kMER models that allow the integration of human and the computer into a cooperative and interactive platform is proposed.

1.5 Research objectives

The multidimensional reduction technique resulted from SOM is used to produce a two-dimensional topology-preserved map that enables the visualisation of the input data. Nevertheless, the constraint of SOM is that topological mismatches occur in the resulted map, and the density estimation of SOM is not proportional to the input density. The kMER approach that utilises the kernel-based maximum entropy learning rule to produce a faithful representation of the probability distribution of the input data is able to overcome this constraint. However, owing to the highly overlap RFs regions that occur at the early stage of kMER training, it causes inefficient computation that makes kMER impractical when voluminous data samples are available. Based on this motivation, this research undertakes the design and development of a hybrid model in an attempt to accelerate the equiprobabilistic map formation process and to improve the applicability of the model to real-world problems.

In addition, by utilising the Bayes' decision theory and the RFs of the equiprobabilistic map, the probabilistic density function of the input data can be obtained. This statistical approach offers strong theoretical as well as practical foundations for the implementation of classification systems. On the other hand, if the input data are unlabeled (i.e., clustering), the density-based clustering method can be used together with the topographic map to produce cluster regions for data visualisation and data exploration purposes. Such data modelling features, which include data visualisation, classification, and clustering, points to the worthiness to design and develop a computing system that is able to supplement humans' decision-making abilities.

There are three main components of this research work. The first component is centred on developing of an ANN hybrid model that can improve the visualisation as well as the convergence rate of a topographic map when compared with the original stand-alone models. The second component is based on devising appropriate strategies based on the hybrid approach to support data visualisation, classification, and clustering. The third component focuses on demonstrating the applicability of the hybrid model to interactive intelligent decision support.

The specific objectives of this research are as follows

- to develop a novel hybrid ANN model for topographic map formation;
- to study the feasibility of the hybrid model as a probabilistic classifier;
- to devise a new lattice disentangling monitoring algorithm for topographic map formation and density-based clustering
- to demonstrate the applicability of the hybrid model to decision-making
- to assess the performances of the hybrid model (and its variants) in data visualisation, classification, clustering, and decision support using simulated and benchmark datasets as well as real-world case studies in the areas of engineering, design, and medical diagnosis

1.6 Research scope

The scope of this research is confined to the design and development of models and algorithms to improve the current methods of topographic map formation. Particular focus is placed on the convergence rate and the visualisation of the resulted map. A novel hybrid neural network model that is founded on the topographic map as well as other statistical data analysis methods to support visualisation, classification, as well as clustering are investigated. The effectiveness of the proposed model is examined using empirical approaches with simulated as well as benchmark datasets. Applicability of the proposed system to data visualisation, classification, clustering, as well as decision support is demonstrated empirically, which include real-world case studies in the domains of engineering, design, as well as medical diagnosis.

1.7 Research methodology

This research begins with a thorough literature review on various methods of topographic map formation in an effort to comprehend the limitations that exist in these methods. It then proposes a new ANN hybrid model, founded on the topographic map, which is able to overcome some of the identified limitations. This proposed model is then implemented using Microsoft Visual Basic 6.0 and MATLAB® 6.0 software packages. The research then devises appropriate strategies based on this hybrid model to support data visualisation, classification, and clustering. The performance of the hybrid model in data visualisation and as a probabilistic classifier is evaluated empirically using both simulated and benchmark datasets. These include the Gaussian source separation dataset, waveform classification dataset, Ionosphere dataset, and Pima Indian dataset.

This research also proposes a novel lattice disentangling monitoring algorithm to be integrated into the hybrid model to alleviate topological defects that occur during the topographic map formation process for improved density-based clustering analysis. The principal curve distribution and the Gaussian dataset are used to demonstrate the effectiveness of this new monitoring algorithm in accelerating the topographic map formation.

A thorough literature review on human's cognitive processing model of decision-making is also undertaken to identify the main cognitive processes that occur during the decision-making process. These cognitive processes lead to the use of the hybrid model that combines human's cognitive processes into a cooperative framework. Finally, the applicability of the proposed hybrid model is demonstrated using real-world case studies in the domains of engineering, design, and medical diagnosis.

1.8 Thesis outline

The organisation of this thesis is as follows.

Chapter 2 presents a literature review of three important areas, namely data visualisation, classification, and clustering, which are related to this work. Classical as well as advanced methods in the statistical methods and ANNs in the relevant areas are reviewed. Advantages and limitations of these methods and the related research studies that are conducted to overcome these shortcomings are presented. This chapter also provides the definition of DSS and describes several categories of DSSs. The iDSS literature and some applications that utilise neural computing techniques for decision support are highlighted.

Chapter 3 described the importance of data visualisation in intelligent data exploration. It explains in details of the SOM and kMER models for multivariate data projection. It then explains the limitations of both models. A detailed description of the proposed novel hybrid model, termed as SOM-kMER, for topographic map formation is presented. Several experiments to evaluate the proposed hybrid model in terms of convergence rate, the resulted visualisation and formation of equiprobabilistic map are described. The experimental results are analysed, compared, and discussed.

Chapter 4 presents a novel classifier design that takes a hybrid approach by integrating the probabilistic estimation procedure of Bayes' theorem with the SOM-kMER model. It provides studies of the decision making process in statistical pattern classification, explains the Bayes decision criterion, and discusses the general characteristics of fixed and variable kernel density estimation. This is followed by a description of the proposed hybrid classifier model and the results of a series of simulation studies using benchmark datasets. How the proposed classifier is employed to tackle a real-world problem related to fault detection and diagnosis in a power generation plant is described.

Chapter 5 introduces a novel lattice disentangling algorithm to overcome topological defects occurred in SOM and kMER owing to a rapid diminution of the neighbourhood range during the topographic map formation process. It then demonstrates the ability of the new monitoring algorithm in SOM-kMER to accelerate the formation of the topographic map. A comparison between the results obtained by the proposed monitoring algorithm and with those from the original kMER monitoring algorithm using simulated datasets is conducted. The applicability of the hybrid SOM-kMER model with the new monitoring algorithm for data clustering in two real-world applications are demonstrated: (i) pen-based handwritten digit recognition and (ii) bedroom colour scheme design based on Kansei Engineering. The results are analysed and discussed in terms of visualisation of the data structure and feature extraction of the cluster regions for data mining purposes.

Chapter 6 highlights the applicability of the hybrid model that combines the capabilities of the computer system and the cognitive capabilities of humans into a cooperative framework. A liver diseases diagnosis and a pen-based handwritten digit recognition problem are used to assess its effectiveness in tackling decision support problems.

Finally, Chapter 7 draws the conclusions and contributions of this research. A number of areas to be pursued as future work are suggested at the end of this chapter.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

Today, the rapidly growing databases and other computational resources have generated huge amount of data. Numerous advanced methods of machine learning, pattern recognition, data analysis, and visualisation are devised to uncover salient structures and interesting correlations in data (Jain et al., 2000) with the intention to generate useful, meaningful, and interesting information out of this flood of data.

This chapter reviews three important areas of intelligent systems that are related to this research: visualisation, pattern recognition (classification and clustering), and decision support systems. Various visualisation techniques, as well as models for supervised classification and clustering are studied. This chapter also provides a review of the human decision-making process and various existing computer-based decision support systems.

2.2 Visualisation

Visualisation techniques are becoming increasingly important in data mining and exploration of huge high-dimensional datasets. A major advantage of visualisation techniques when compared with other non-visual data mining techniques is that visualisation allows direct user interaction that could then provide immediate feedback (Ankerst, 2000). Data visualisation enables the user to speculate the properties of a dataset based on his/her own intuition and domain knowledge. It is meant to convey hidden information about the data to the user. Ware (2000) and Tufte (1983) provide an excellent overview of modern as well as classical techniques for data and information visualisation.

There are many types of data visualisation approaches. Geometric approaches (Huber, 1985), for example, are based on statistical methods like factor analysis, principal component analysis, and multidimensional scaling. Icon-based approaches map the attribute values of each input data onto small graphical primitives known as icons (Wong & Bergeron, 1997). Hierarchical approaches employ a hierarchical partitioning into subspaces (Tipping & Bishop, 1997). The following section highlights a number of visualisation techniques based on classical multi-dimensional scaling and artificial neural networks, particularly for data projection.

2.2.1 Principal Component Analysis

The PCA is a well-known statistical method for data projection (Fukunaga, 1990), and is widely used in pattern recognition, data analysis, as well as signal and image processing. It is a linear orthogonal transform from a *d*-dimensional input space to an *m*-dimensional space, $m \leq d$, such that the coordinates of the data in the new *m*-dimensional space are uncorrelated and maximal amount of variance of the original data is preserved by only a small number of coordinates.

Although PCA is a proven and widely used robust method for data projection, it cannot cope with certain situation of dataset such as the one shown in Figure 2.1, due to its inherently linear approach. Such datasets require the computation of non-linear principal or curvi-linear components (König, 2000). Nevertheless, alternative method such as Sammon's non-linear mapping is able to overcome this problem.

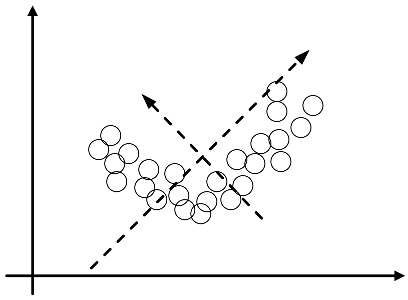


Figure 2.1: Non-linear principal components

2.2.2 Sammon's non-linear mapping

Sammon (1969) proposes a non-linear projection technique that attempts to maximally preserve the inter-pattern distances in the original space and inter-pattern distances in the projected space. For a given M input vectors in a d-dimensional space, say $\mathbf{x}_i$, where $i = 1,2,...,M$. These M vectors are to be mapped from the d-dimensional space to either a 2D or 3D space. The M vectors in the lower dimensional space are denoted by $\mathbf{y}_i$, where $i = 1,2,...,M$. The $\mathbf{x}_i$ and $\mathbf{y}_i$ are expressed as follows:

$$\mathbf{x}_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1d} \end{bmatrix} \mathbf{x}_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2d} \end{bmatrix} \cdots \quad \mathbf{x}_M = \begin{bmatrix} x_{M1} \\ x_{M2} \\ \vdots \\ x_{Md} \end{bmatrix}$$

$$\mathbf{y}_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1D} \end{bmatrix} \mathbf{y}_2 = \begin{bmatrix} y_{21} \\ y_{22} \\ \vdots \\ y_{2D} \end{bmatrix} \cdots \quad \mathbf{y}_M = \begin{bmatrix} y_{M1} \\ y_{M2} \\ \vdots \\ y_{MD} \end{bmatrix}$$

Initially, a set of M vectors (or data points) is generated randomly in the 2D (or 3D) space. The error in mapping is determined by comparing these data points with the corresponding points in the d-dimensional space, and it has to be (locally) minimum for a perfect mapping. Let $E(t)$ be the mapping error after the t-th iteration. This is expressed mathematically as follows:

$$E(t) = \frac{1}{C} \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{[d_{ij}^{*} - d_{ij}(t)]^2}{d_{ij}^{*}}, \quad i < j \tag{2.1}$$

$$C = \sum_{i=1}^{M}\sum_{j=1}^{M} d_{ij}^{*} \tag{2.2}$$

$$d_{ij}(t) = \left[\sum_{k=1}^{D} [y_{ik}(t) - y_{jk}(t)]^2 \right]^{1/2} \tag{2.3}$$

where d_{ij}^{*} is the distance between two points in the *d*-dimensional space and $d_{ij}(t)$ is the distance between the corresponding two mapped points in *D* dimension (2D or 3D) space. Sammon uses a diagonal Newton method to locally optimise *E(t)* and reduce the step length by a factor value of 0.3 to 0.4. The details of the algorithm can be found in Sammon (1969).

Sammon mapping is shown to be more superior than PCA for data structure analysis. However, the Sammon algorithm is a point-to-point mapping which does not provide the explicit mapping function and cannot accommodate new data points (Mao & Jain, 1995). For any additional data, the projection has to be re-calculated from scratch based on all data points. This is impractical as many real-world applications have typically a large amount of data, and data are added sequentially. Moreover, the Sammon mapping approach has drawback in terms of its generalisation ability when compared with some of the ANN-based projection techniques (Mao & Jain, 1995).

2.2.3 Self Organisation Map

The most promising mapping method in terms of structure preservation uses either topology or distance preservation for the non-linear mapping process (König, 1998). The SOM (Kohonen, 1982), one of the ANN techniques that uses unsupervised non-linear projection methods, has a very desirable topology preserved property, i.e. close points in the input space are mapped to the nearby neurons in the map space. Such property enables visualisation of the relative or quantitative mutual relationships among the input samples (Yin, 2002a).

In SOM, a topological structure is defined on the competitive layer. The neurons of the competitive layer are organised in the shape of a straight line, a rectangle, or a hyper-cuboid. As the map is often arranged in a low dimensional grid and the inputs are drawn from a high dimensional space, SOM is used as a visualisation tool for dimensionality reduction.

As pointed out by Vesanto (2002), visualisation produced by SOM plays two important roles. First, the prototype of SOM is regarded as a representative sample of the data. It is assumed that the properties seen from the visualisation of the prototypes also hold for the original data. The prototypes is an ordered representation of the data, with the neighbouring prototypes on the map are similar to each other but dissimilar with the far away prototypes. Second, the prototype formed by SOM is a model of the data. The prototype can be used to determine the probability density estimation of the input data. As the density of the map prototype follows roughly the density of the data, the map grid provides visualisation based on the input data density (Vesanto, 2002).

However, the inter-neuron distance on the map grids are not directly visible or measurable on the map. One has to use a colour scheme such as U-matrix (Ultsch, 1993) or interpolation (Yin & Allinson, 1999) to visualise the relative distance between the prototype neighbouring neurons. Yin (2001) proposes a visualisation induced self-organising map (viSOM) by considering the latent contraction force between the neuron that helps to regularise the inter-neuron distance.

Topological mismatch also occurs in the SOM map. This is shown in the experiments conducted by Van Hulle (2000): a mapping of a circular and L-shaped distribution produced by SOM (see Figure 2.2). The circular distribution shows that the prototype (weights) distribution is non-uniform. As for the L-shaped distribution, there are several neurons outside the support of the distribution, and some of them have zero (or very low) probability to be active, which are known as ‘dead’ units.

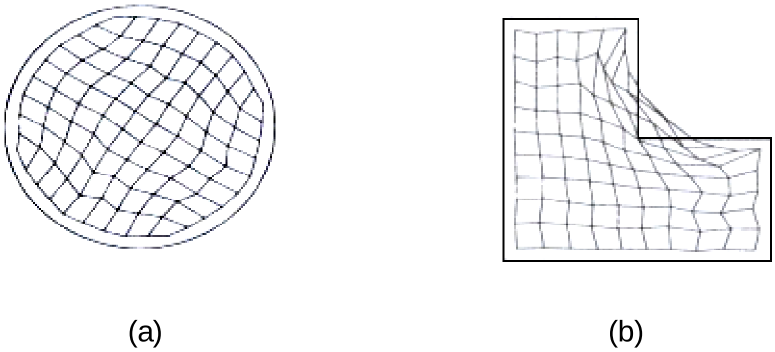


Figure 2.2: (a): Mapping of a 20x20 lattice onto a circular distribution.
(b): an L-shaped uniform distribution. (Adapted: Van Hulle, 2000)

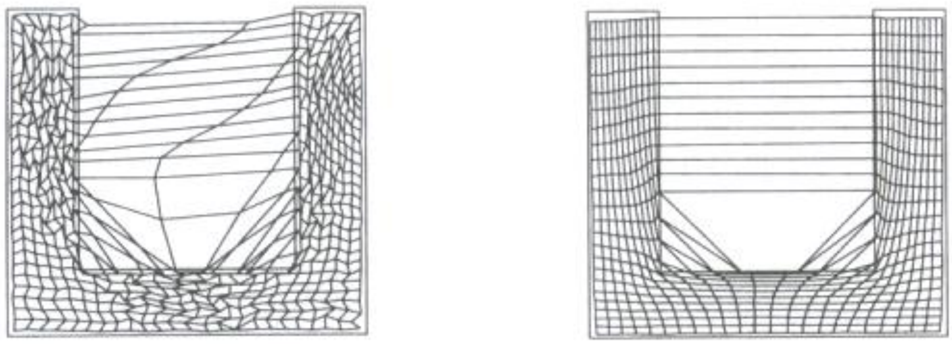
The weight density $p(\mathbf{w})$ achieved by SOM at convergence is not a linear function of the input density $p(\mathbf{v})$. In the case of discrete grids, when the neighbourhood range vanishes in SOM training, $p(\mathbf{w})$ is proportional to $p^q(\mathbf{v})$ where $q = 1/(1+(2/d))$ (Kohonen, 1995). In addition, SOM tends to undersample high probability regions and oversample low probability ones. Due to this limitation, it is unable to provide a “faithful” representation of the probability distribution of the underlying input data (Van Hulle, 2000).

The following section provides a review of the equiprobabilistic map, which is able to transfer the maximum amount of information available about the input distribution to the formed topographic map (Van Hulle, 1995; 1996).

2.2.4 Equiprobabilistic map

Standard unsupervised competitive learning models (such as SOM) are unable to produce similar density estimate as in the input space and yield neurons that are never active (“dead” units) in the projected map. There are several attempts to build equaprobabilistic maps. This includes using strategy to maximise the information-theoretic entropy of the map output. As a result, the map transfers the maximum amount of information available about the (stationary) distribution from which it receives inputs. The weight density $p(\mathbf{w})$ is proportional to the input density $p(\mathbf{v})$, or all neurons have equal probability to be active, hence producing an equiprobabilistic map (Van Hulle, 2000).

There are two categories of equiprobabilistic map formation rules: the lattice-based learning rule and the kernel-based learning rule. Examples of the first category include the Maximum Entropy learning Rule (MER) (Van Hulle, 1995) and the Vectorial Boundary Adaptation Rule (Van Hulle, 1996). The map grid units (neurons) have the same dimensionality as that of the input space. The quantisation regions (or RFs) correspond to the lattice quadrilaterals and do not assume any Voronoi tessellation. Van Hulle (2000) demonstrated that MER, as opposed to SOM, does not produce any dead unit in one of his experiments that uses an U-shaped uniform distribution. Figure 2.3 depicts the lattice obtained in the experiment.



(a) (b)

Figure 2.3: Lattice obtained for a U-shaped uniform distribution. **(a):** SOM training **(b):** MER training (Adapted: Van Hulle, 2000)

However, in MER, the lattice topology is rectangular and its dimensionality is the same as that of the input space in which the lattice is developed. As a result, it cannot be used for non-parametric regression and dimensionality reduction purposes as the definition of the quantisation region is too complicated (Van Hulle, 1999).

In the second category, the RFs are kernel-based. The individually adapted kernel performs local smoothing of the interpolation function, which is defined by the sum of all RF kernels (Van Hulle, 2000). An example of this type of equiprobabilistic map formation model is kMER (Van Hulle, 1998). The requirement of equidimensionality of MER is relax with different RF definitions. In kMER, the quantisation regions are non-uniform, and the neurons can turn active using graded magnitude. Since the RFs regions may overlap on the map, multiple winners can be selected during competitive learning. Depending on the neural activation, the weights are adapted so as to produce a topology-preserved mapping. The lattices formed by kMER not only produce few dead units but can also be used for non-parametric density estimation and clustering. Chapter 3 describes the kMER model in detail.

2.3 Data classification

Pattern recognition and classification have been extensively studied (Fukunaga, 1972; Duda et al., 2000), and Jain et al. (2000) provide a comprehensive review of statistical pattern recognition. These studies mainly focus on how machine can observe the environment, learn to distinguish patterns of interest from their background and able to classify patterns into different categories (Jain et al., 2000). The process of pattern recognition can be partitioned into several sub-processes, i.e., sensing, segmentation, feature extraction, classification, post processing, and decision (Duda et al. 2000). However, this research focuses primarily on the pattern or data classification component. The following section provides a review on some of the well-known pattern classification approaches.

2.3.1 Statistical approach

In the statistical decision theoretic approach, the decision boundaries are determined by the probability distributions of the patterns for each class, which is either specified or learned (Duda et al., 2000). There are a number of well-known decision rules, such as Bayes decision, maximum likelihood, minimum-error-rate classification, and Neyman-Pearson criterion, to define the decision boundaries (Duda et al., 2000). Generally, most recognition systems of statistical decision theoretic approach operate in two modes: training (learning) and classification (testing). In the training mode, the feature extraction module finds the appropriate features to represent the input patterns, and the classifier is trained to partition the feature space. In the classification mode, the trained classifier assigns the input pattern to one of the pattern classes under consideration based on the measured features (Jain et al., 2000).

2.3.2 Neural networks approach

In recent years, a large part of connectionist research is devoted to the development and theoretical analysis of pattern classifiers having ANN-like structure and learning capabilities. The main characteristic of ANN-based classification approach is its ability to learn complex, non-linear input-output relationships. ANNs are some kind of massively parallel computing system with a large number of interconnected simple processors that are able to adapt themselves to the data. Although this enthusiasm is first considered with some scepticism by researchers in mainstream statistical pattern recognition (Duin, 1994), ANNs can be generally seen as a type of statistical pattern classifier (Werbos, 1991). Some of the commonly used models include feed-forward networks, basis and kernel function networks, and self-organising and competitive networks.