ESOM: An Algorithm to Evolve Self-Organizing Maps from On-line Data Streams

Da Deng, Nikola Kasabov
Dept. of Information Science, Univ. of Otago, Dunedin, New Zealand
E-mail: ddeng, nkasabov@infoscience.otago.ac.nz

Abstract

An algorithm of evolving self-organizing map (ESOM) is proposed as a dynamic version of the Kohonen self-organizing map, where network structure is evolved in an on-line adaptive mode. Experiments have been carried out on some benchmark data sets as well as on macroeconomic data. Results show that ESOM is a good tool for clustering, data analysis, and visualisation.

1 Introduction

Many real world information systems use data from on-line data streams that are updated frequently. To extract useful information hidden among these multivariate data, a number of techniques can be used, such as visualisation, dynamic clustering, and classification, as it requires.

Of growing interest are the methods of intelligent data analysis and processing, which can overcome difficulties in dealing with complex dynamics of multivariate data, noise or distortion which occurs during data collection, or lack of domain knowledge [9]. Besides traditional statistical methods, people have been approaching the problem using expert systems, machine learning techniques and artificial neural networks etc.

An ECOS paradigm is proposed in [7], setting out principles for on-line construction of intelligent information systems using connectionist-based models. The ECOS principles include: fast incremental on-line learning, evolvable network structure, and knowledge interpretation and manipulation. An evolving fuzzy neural network (EFuNN) model has been proposed as an instance of ECOS prototypes, which is found to be very effective for time series prediction and pattern classification [8].

In this paper we introduce another realisation of ECOS which focuses on data clustering and visualisation. It is called Evolving Self-Organizing Map (ESOM), as it is a variation of the Kohonen Self-organizing map (SOM) [12].

The rest of the paper is organised as follows. In Section 2, we first give a brief introduction of the ESOM computational model. In section 3, experiments are carried on a benchmark data set and a case study on international macroeconomic data are presented in an effort to generate a world macroeconomic map. Finally, conclusions are given in Section 4.

2 The Computational Model

2.1 Network structure

The network structure of ESOM is different from that of SOM. No topological constraint is given for the feature map a priori and prototype nodes are not organised onto one- or two-dimensional lattices.

The ESOM network starts without nodes. During learning, the network updates itself with on-line incoming data, creating new nodes when necessary. Each node carries a weight vector of the same dimensionality as the input data. Connections between map nodes are used to maintain the neighbourhood relationships between close nodes. The strength of the neighbourhood relation is determined by the distance between connected nodes. If the distance is too big, giving a weak strength under a threshold, the connection can be pruned. In this way the feature map can be split apart and data structures such as clusters and outliers can emerge.

2.2 The algorithm

We denote the ESOM network at time t as a triplet of a node set $\Omega \subset E^d$, an interconnection set $\mathcal{C}$, and a parameter set $\mathcal{P}$:

$$\mathcal{E}^t = (\Omega^t, \mathcal{C}^t, \mathcal{P}) \tag{1}$$

with each node $\mathbf{w}_i \in \Omega^t$ as a vector of dimension d, $i = 1...N$, and N is the current number of nodes in Ω^t.

The learning process can be summarised as follows:

1. Input a new data entry $\mathbf{x}$;
2. If none of the existing nodes matches the data vector within a small error, i.e., for all $i = 1, ..., N$,

   $$\|\mathbf{w}_i - \mathbf{x}\| > \epsilon$$

   (ϵ is an error threshold), create a new node in the network which represents exactly vector $\mathbf{x}$:

   $$\Omega^{t+1} = \Omega^t \cup \mathbf{x} \tag{2}$$

   Connect the new node with its two nearest neighbours $\mathbf{w}_{n1}$ and $\mathbf{w}_{n2}$:

   $$\mathcal{C}^{t+1} = \mathcal{C}^t \cup c(\mathbf{x}, \mathbf{w}_{n1}) \cup c(\mathbf{x}, \mathbf{w}_{n2}) \tag{3}$$

   Here $c(\cdot, \cdot)$ denotes a connection between two vectors.
3. Otherwise update the matching node and each of its neighbours, denoted as $\mathbf{w}$, according to their distances to the data vector $\mathbf{x}$, a relation represented by a function f:

   $$\Omega^{t+1} = f(\Omega^t, \mathbf{x}), \tag{4}$$

   with each node being modified as

   $$\Delta\mathbf{w} = \gamma\ e^{-\|\mathbf{w}-\mathbf{x}\|/\sigma^2}(\mathbf{x} - \mathbf{w}) \tag{5}$$

   where γ is a small learning rate and σ controls the effective neighbourhood spread.
4. Gradually prune weak connections and inactive nodes;
5. Repeat all steps above.

ESOM is aimed at achieving "life-long" learning. With a relatively small learning rate and a data sequence which is long enough, however, it can be expected that after the presentation of a certain number of data examples an optimum set of prototypes representing the data stream will be learned. A similar case is analysed using Gaussian approximation in [6].

2.3 Discussions

The weight vector update rule in Eq.(5) is similar to that of SOM, except that for the neighbourhood function the vector distance between nodes is used, rather than the grid distance in SOM [11].

Obviously ESOM learning is more localised than SOM, so it does not suffer from any *border effect* which frustrates SOM [15], i.e., prototypes staying always close to the centre of input data space. Less redundant nodes are likely to be produced in ESOM while overall quantisation error is reduced. Without fixed topology in the feature map ESOM does not need to unfold maps of ill initialisation. SOM, however, needs to start with a large neighbourhood to cope with this problem, which implies a much longer learning time.

There have been a few SOM variations [12] which support incremental learning with dynamic network structures, e.g., Blackmore's incremental grid growing algorithm (IGG) [1], and Fritzke's Growing Network

Grid (GNG) [5]. In IGG nodes and connections can be added to or deleted from the feature map, which is on a limited two-dimensional space. Unlike GNG, ESOM prototypes are assigned directly using data samples instead of applying an empirical midpoint interpolation. This suggests that whenever data of novelty appear, learning starts with a memory on the data, and continues by adapting the existed memory to the changing environment. Generally speaking, compared with SOM and its variations, ESOM has weaker geometry constraint on feature maps while better accuracy and efficiency are achieved.

3 Comparative study

3.1 Data sets

Data sets used in this study include:

1. The spiral data. This well-known benchmark data set consists of 194 data entries, each of which has a pair of X-Y coordinates and one class label (either 1 or -1).
2. Macroeconomic data. This data set is from our case study for risk analysis of European Monetary Union economy [9], which employs a number of economic and financial indicators to predict possible shocks from which EMU market is unfortunately not immune, and develop a computational system for analyzing and anticipating signals of abrupt changes of volatility in financial markets. Here we focus on the problem of generating a world macroeconomic map to evaluate performance and development in national and regional macroeconomy. Macroeconomic data in the period from 1994 to 1998 are collected for fifteen EMU countries, UK, US, and Asian countries such as Japan (JP) and Thailand(TH) (see [9]). The data is taken from [4] and from the DataStream on-line source. The data set has four attributes, namely annual change percentage of stock market (PCH), debt over GDP (DBT/GDP), deficit over GDP (DEF/GDP), and inflation rate. Each data entry carries a label composed by country code and time numbers, which will be used later for the generation of a labelled map.

3.2 Convergence and the goodness of prototype sets

Simulations on the data sets given above have shown that ESOM has a much faster convergence rate in the sense that a good set of prototypes can be quickly learned from the input data. This is also demonstrated in our experiments using two-dimensional signals randomly generated from some 2-D shapes [3].

When trained with classification data sets, the goodness of prototype sets is evaluated with its discrimination ability in the input data space. This is done by categorising each map vector with a class label using the nearest-neighbour rule operating on the distances between map vectors and data entries. Each map vector takes a class label from its best matching data entry. We show in Fig.1 how the discrimination ability develops as the map learns and converges. The spiral data set is used in both SOM and ESOM modules. In both cases we classify the data set using map vectors labeled by the nearest-neighbour rule and calculate the classification error rate during learning. Here we plot representative results from the following modules:

- Two ESOM modules. ESOM I with $\epsilon = 1.0$, $\gamma = 0.01$. After around 200 steps the network grows to 78 nodes with stable performance of an error rate around 15.46%; ESOM II with $\epsilon = 0.9$, $\gamma = 0.01$, which stabilises with 114 nodes with zero classification error rate.
- Two SOM modules. SOM I is of 12×12 size, with an initial learning rate of 0.05, and an initial neighbourhood width $N = 8$; SOM II is of 9×9 size, initial learning rate 0.05, initial neighbourhood width $N = 4$. For both modules a number of experiments are carried out with different random initial weights. Final error rates are around 22%. The discrimination ability does not increase with the map size in this case.

Similar results are obtained when experimenting the both models with other data sets. Generally speaking, ESOM learns much faster than SOM, and the discrimination ability of its feature maps is better. We have also tried the same data set with the LVQ_PAK package [12] and found that even after fine-tuning with LVQ2.1 and LVQ3, the error rates of LVQ networks of equivalent size, as given in Table 1, are bigger than their ESOM counterparts, having in mind that LVQ is a supervised learning algorithm and ESOM is an unsupervised one.

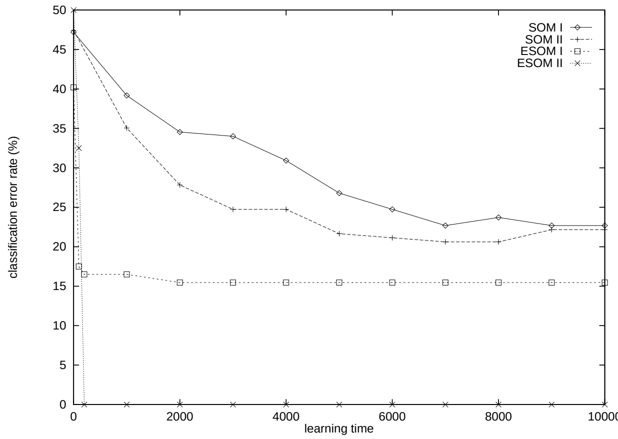


Fig. 1: Discrimination ability develops as SOM and ESOM learn the spiral data. ESOM modules learn much faster, and are more accurate.

Algorithms	SOM I	SOM II	ESOM I	ESOM II	LVQ (78 nodes)	LVQ (114 nodes)
Error Rate	22.68	22.16	15.46	0	24.23	11.86

Table 1: Classification error rate of different algorithms

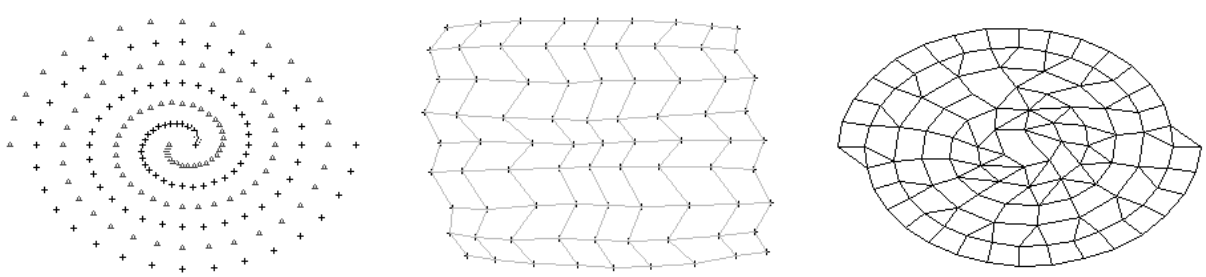
3.3 Visualisation

The spiral data of two dimensional space is visualised using map vectors obtained by SOM and ESOM, as shown in Fig.2. Obviously the ESOM result gives a more accurate representation of the spiral shape.

For feature space of high dimensionality, visualisation of the ESOM feature map is in question as the prototype space is also of high dimensionality. This problem can be solved, however, using Sammon's algorithm [14] which projects high dimensional data into a two-dimensional space while keeping the distance ordering as best as possible. The visualisation of SOM is straightforward as the feature nodes are laid usually on a two-dimensional lattice. But the lattice coordinates are not in a metric space. Distances between neighbour nodes in SOM can vary greatly. In this sense SOM may not a good choice for complex high dimensional data visualisation, although efforts have been made for improvement [13]. Opposite to SOM, ESOM creates a topology which reflects the data space rather than using a fixed pre-defined topology.

3.4 Mapping the Global Macroeconomy

SOM has been used in economic and financial data analysis in a number of studies such as [2][9][10][16]. Here a 12×12 two-dimensional map is first trained using SOM. The size of the map is selected empirically so as to obtain a well-expanded mapping space. The map learned from the annual macroeconomic data



(a) (b) (c)

Fig. 2: Visualisation of the spiral data. (a) The data set on a 2-D plane; (b) The SOM grids after training; (c) The ESOM nodes with trimmed connections.

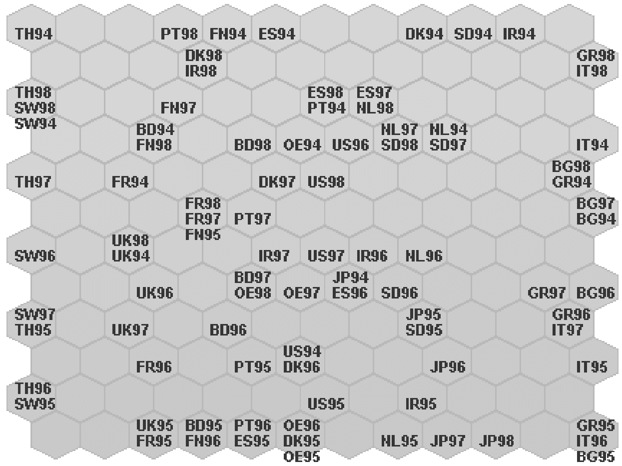


Fig. 3: The annual macroeconomic map for EMU countries etc. obtained with the use of SOM.

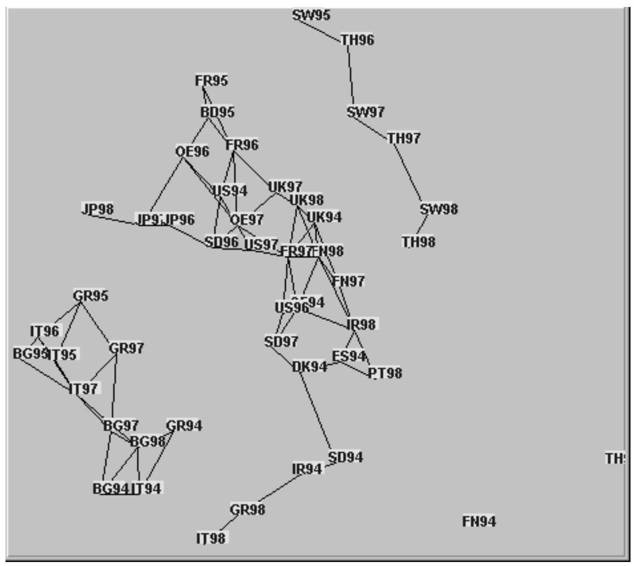


Fig. 4: The ESOM visualised with Sammon projection shows the data structure.

is shown in Fig.3, where map nodes are displayed with hexagons labelled with best-matching data entries and tinted according to the PCH component value of the corresponding feature vector. An EMU cluster is formed in the central part of the map, including the following countries: OE(Austria), NL(Netherland), DK(Denmark), IR(Ireland), SD(Sweden), BD(Germany) and FR(France). Non-EMU countries UK and US also fall into this cluster. Four EMU countries fall out apparently: IT(Italy), BG(Belgium), GR(Greece) in area of high inflation rate on the right, and SW(Switzerland) on the left side.

By using colour palette or different gray-scale for nodes of different component values, the map such as in Fig.3 presents a very useful tool to evaluate macroeconomic performance of different countries. Tracking down the movement of certain country on the map, it is also helpful to evaluate its developing trend in different years. But as we mentioned before the map distance between nodes may not match their distance in the feature space, this can be misleading. With SOM it is also difficult to find clusters visually.

Another annual map is next evolved with the same data set and shown in Fig.4. The map is first clustered using ESOM algorithm, and then projected onto a two-dimensional plane for visualisation using Sammon's algorithm. Weak connections are then clipped away. The layout of labelled nodes is quite similar to that of Fig.3, but the ESOM map gives more explicit data structure such as clusters and outliers. Here we find two major clusters, the EMU cluster with countries like FR, BD, FN, IR etc. plus UK and US, the fall-out cluster with GR, IT, and BG.

4 Conclusion

This paper introduces an evolving self-organizing map (ESOM) as an evolving variation of the Kohonen SOM, featuring its quick on-line learning ability, and a feature map of less geometric constraint, more compactness, and more prototyping accuracy.

From the comparative study on benchmark data sets and some world macroeconomic data, it is clear that ESOM module is a very useful tool for on-line data clustering, dynamic data analysis, and scientific visualisation.

Acknowledgements This work is fully supported by Grant UOO-808, funded by FRST, New Zealand.

References

[1] Blackmore J., Miikkulainen R. (1993) Incremental Grid Growing: Encoding High-Dimensional Structure into a Two-Dimensional Feature Map, Proc. ICNN'93, Int. Conf. on Neural Networks, Vol. I, 450-455, IEEE Service Center.

[2] Deboeck, G.(1999) Investment maps for emerging markets. Neuro-fuzzy techniques for intelligent information systems (N.Kasabov and R.Kozma Eds.). Physica Verlag, 373-395.

[3] Deng, D., Kasabov, N. (1999) Evolving Self-organizing Map and its Application in Generating A World Macroeconomic Map. Proceedings of the ICONIP/ANZIIS/ANNES'99 International Workshop, 7-12, University of Otago, Dunedin.

[4] European Central Bank, Montly Bulletin (1999).

[5] Fritzke, B. (1994) Growing cell structures - a self-organizing network for unsupervised and supervised learning. Neural Networks, **7**, 1441-60.

[6] Heskes, T.M., Kappen B. (1991) Learning processes in neural networks. Physical Review A, **44**, 2718-26.

[7] Kasabov, N. (1998) The ECOS framework and the ECO learning method for evolving connectionist systems. Jour. of Advanced Computational Intelligence, **2**, 1-8.

[8] Kasabov, N. (1998) Evolving fuzzy neural networks - algorithms, applications and biological motivation. in: Yamakawa and Matsumoto (Eds.) Methodologies for the Conception, Design, and Application of Soft Computing, World Scientific, 271-274.

[9] Kasabov, N., Erzegoveri, L. et. al. (2000) Hybrid intelligent decision support systems and applications for risk analysis and prediction of evolving economic clusters in Europe. To appear in N.Kasabov (ed), Future Directions for Intelligent Systems and Information Sciences, Physica Verlag (Springer Verlag).

[10] Kaski, S. (1997) Data exploration using self-organizing maps. Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series No. 82.

[11] Kohonen T. (1982). Self-organizing formation of topologically correct feature maps, Biological Cybernetics, v. 43, 59-69.

[12] Kohonen, T. (1997) Self-Organizing Maps, second edition. Springer.

[13] Mao J., Jain, A.K. (1995) Artificial neural networks for feature extraction and multivariate data projection. IEEE Transaction on Neural Networks, **6**, 296-317.

[14] Sammon, Jr., J. (1969) A nonlinear mapping for data structure analysis. IEEE Transaction on Computers, **18**, 401-09.

[15] Sarle, W. (1999) Newsgroup FAQ of comp.ai.neural-nets, part 2. URL ftp://ftp.sas.com/pub/neural/FAQ.html.

[16] Serrano-Cinca, C. (1996) Self organizing neural networks for financial diagnosis. Decision Support Systems, **17**, 227-38.