ICONIP’2002, vol.2, pp. 787-791, Nov. 18~22, 2002, Singapore

ECONOMIC STATES ON NEURONIC MAPS

Cheng-Yuan Liou[1] , Yen-Ting Kuo

Dept. of Computer Science and Information Engineering, National Taiwan University.
Supported by National Science Council under Project NSC 90-2213-E-002-092
[1]correspondence , Email: cyliou@csie.ntu.edu.tw

ABSTRACT

We test the idea of visualizing economic statistics data on self-organization [1][2] related maps, which are the LLE[3], ISOMAP[4] and GTM[5] maps. We report initial results of this work. These three maps all have distinguished theoretical foundations. The statistic data usually span high-dimensional space, sometimes more than 10 dimensions. To perceive these data as a whole and to foresee future trends, perspective visualization assistance is an important issue. We use economic statistics (see Table 2) for the United States over the past 26 years (1977 to 2002) and apply them on the maps. The results from these three maps display historic events along with their trends and significance.

1. INTRODUCTION

It is hard to perceive the economic state of a nation as a whole based on large amounts of multivariate economic statistics. In Fig.1, we plot economic statistics for the United States over the past 25 five years. Such a chart may not help us to perceive intricate economic situations. One may wonder wether if the data can be represented on a 2-D surface [2] rather than in a high dimensional form, one will be able to see it clearly. This issue involves the fundamental problem of dimensionality reduction: how to discover compact representations of high-dimensional data. The manifold concept is introduced to obtain such a representation. By assuming that the statistic data are intrinsically low dimensional, one can apply the manifold methods, such as LLE, ISOMAP and GTM, to reduce the dimension of the data from high to low. In the following section, we briefly review these three methods and apply them to visualize these economic data.

2. METHOD

We show in the following how to process raw statistics, and we briefly review the three manifold methods.

2.1. Data preparation

We used economic data (Table. 2) for the United States over the past 26 years (1977 to 2002) in all our simulations, which included eleven statistics: the averaged gross national product (GNP), consumer price index (CPI) growth rate, unemployment rate, foreign currency reserve, foreign exchange rate, prime rate, export growth rate, import growth rate, economic growth rate, balance of international payments and money supply growth rate.

Some of these data were in the percentage format, and some of them were in the numerical format spread over various large ranges. We normalized each range so that each statistic made an equal contribution. Three processing steps, as listed below, were used to construct 2-D maps of the economic data:

1. Normalize each statistic data range to 1.
2. Feed these normalized data into the dimensionality reduction algorithms LLE, ISOMAP and GTM.
3. Retrieve 2-D results and plot them on the plane.

2.2. Locally Linear Embedding (LLE)

According to the paper by Lawrence K. Saul and Sam T. Roweis [3], one may use LLE to apply dimensionality reduction to these economic data. The LLE algorithm is based on geometric intuition. It regards each input vector (which can be seen as a point in high dimensional space) as being rounded by K neighbors. K is a neighborhood factor that affects the result of this algorithm. Then it is possible to find a locally linear reconstruction to reconstruct a data point from its

ICONIP'2002, vol.2, pp. 787-791, Nov. 18~22, 2002, Singapore

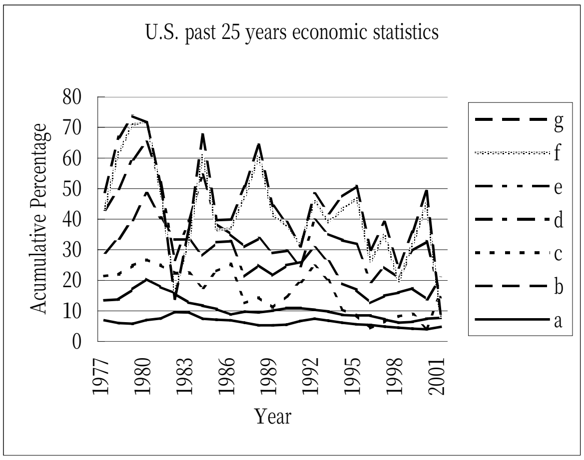

Fig. 1: U.S. economic statistics plotted in a cumulative manner. *(a)unemployment rate, (b)CPI growth rate, (c)money supply growth rate, (d)prime rate, (e)import rate, (f)export rate ,(g)economic growth rate.*

neighbors. Reconstruction errors are measured by means of the cost function:

$$\varepsilon(W) = \sum_i \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2$$

where X_i is the central point, X_j is X_i's neighbors and the weight W_{ij} summarizes the contribution of the *j*th data point to *i*th data center point in the reconstruction. By minimizing the cost function, we obtain the weights W_{ij} . According to the design, the reconstruction weights W_{ij} reflect intrinsic geometric properties of the data that are invariant to exactly such transformations. The final step of the LLE algorithm is mapping each high dimensional observation $\vec{X_i}$ into low dimensional vector $\vec{Y}_i$, which is represented by global coordinates on the manifold. This can be accomplished by choosing d-dimensional (d was set to 2 in all our simulations.) coordinates $\vec{Y}_i$ to minimize the embedding cost function:

$$\Phi(Y) = \sum_i \left| \vec{Y_i} - \sum_j W_{ij} \vec{Y}_j \right|^2$$

The only free parameter is the number of neighbors per data point, K. In our case, the statistic data had dimensionality 11. If we want to avoid the problem where the least square [3] in finding the weights does not have a unique solution, then K should be set less than 11. After setting K, we did not obtain satisfactory results. After several tries, we found that K=17 gave a better result. The result is shown in Fig.2, where each dot denotes one year.

2.3. Isometric Feature Mapping (ISOMAP)

An approach that combines the major algorithmic features of PCA and MDS [6] is called ISOMAP [4]. The ISOMAP algorithm takes the distances $d_x(i,j)$ between all pairs X_j X_i among N data points in the high-dimensional input space X, measured either in the standard Euclidean metric or in some domain-specific metric. The output vectors Y_i in a d-dimensional Euclidean space Y represent the intrinsic geometry of the data. Three steps in this algorithm are listed below:

1. Construct a neighborhood graph.
2. Compute the shortest paths.
3. Construct a d-dimensional embedding.

The only free parameter, K, which is the neighborhood factor, appears in the first step. In our study, K was set to 3, and the result is shown in Fig.3.

2.4. Generative Topographic Mapping (GTM)

The GTM algorithm, developed by Christopher M. Bishop, Markus Svensen and Christopher K.I. Williams [5], provides a principled alternative to the widely used 'self-organizing map' (SOM) algorithm [1]. The GTM consists of a constrained mixture of Gaussians, in which the modal parameters are determined based on maximum likelihood using the EM algorithm.

We generated the components of the statistical data's GTM using a 2-D latent space. The number of basis functions was set to 36, and the width of the basis functions was set to 4, a size comparable to the distance between two neighboring basis function centers. The weight regularization factor was set to off. The result is shown in Fig.4.

3. RESULTS

We summarize significant features on the maps. Comparing the results shown in Figs.2,3, and 4, we find historic events and list them in Table 1. We point out several indications on the maps. All events are more or less emphasized on the maps. When economic conditions deteriorate, the states in the figure sway outward from the cluster center to the edges or make a large jump, such as for the 1997 Asian Economic Storm indicated in Fig. 3, the 1987 New York Stock Market Crash indicated in Figs. 4 and 2, and the 2001 economic deterioration indicated in Figs. 2 and 4. We observe that the years close to the cluster center have very stable economic states, and that the years on the edges or corners have bad economic states. The years that move toward the cluster centers show how effective the economic policies were. By means of these maps, we can perceive economic states and foresee their trends.

In Fig.2, the years from 1978 to 1982, swaying near the edges, reflect the Second Oil Crisis. Tracing the dots to 1986, we see that they lead to a big jump to 1987, reflecting the October 1987 New York Stock Market Crash. In the next year (1988), in the same figure, the dots approach the edge and return in 1989. Years smoothly approach the cluster center during 1990-1993 but fall to the edge during 1994-1995. The years 1994-1995 have heavy inflation. From 1998 to 2001, the dots sway near the edge and have big jumps, indicating economic deterioration.

The LLE map reveals many significant economic features as described above. We next show how ISOMAP works. In Fig.3, the dots for the years 1977 to 1982 approach the edge corner, reflecting the Second Oil Crisis. From 1987 to 1988, the dots jump to the edge, reflecting the New York Stock Market Crash. The dots return for 1989, but fall outward again for 1990, which was the year of the Gulf War. Following on to 1997-1998, the dots make a huge jump to another edge, showing economic deterioration due to the Asian Economic Storm.

In Fig.4, the dot for 1986 at the bottom-right jumps to the upper-left in 1987, indicating a dangerous situation, the New York Stock Market crash in October 1987. We may presume that big jumps in GTM mean dangerous situations. Following the dots to 1989, another jump occurs from 1989 to 1990. The Gulf War happened in 1998. From 2000 to 2001, there is a big jump, corresponding to economic deterioration in 2001. Indeed, the Nasdaq index dropped to a low level in 2001.

These three methods (LLE, ISOMAP and GTM) produce results that display different economic features. Each method is sensitive to certain events. In Fig.3, ISOMAP is most sensitive to the Asian Economic Storm. GTM (Fig.4) is very sensitive to the stock market, and LLE (Fig.2) is sensitive to the all events except the Asian Economic Storm.

Date	Event
1978~1982	Second Oil Crisis
1987.10	New York Stock Crash
1990.8	Gulf War
1994~1995	High Inflation Exceeds 3.5%
1997.12	Asian Economic Storm
2001	Economic deterioration
2002.9	West Coast Port Closed

Table. 1: Important events in the US during the past 25 years.

Indicator	Source
Prime Rate	http://www.neatideas.com/prime.htm http://www.moea.gov.tw/~meco/stat/four/a-15.htm
Unemployment Rate	http://www.bls.gov/
Percent Changes in Consumer Price Index	http://www.moea.gov.tw/~meco/stat/four/a-14.htm
Money Supply Growth Rate	http://www.federalreserve.gov/releases/h6/hist/h6hist1.txt
Foreign Currency Reserve	http://www.moea.gov.tw/~meco/stat/four/a-19.htm
Exchange Rate Singapore/US	http://research.stlouisfed.org/fred2/data/EXSIUS.txt
Import Growth Rate	http://www.moea.gov.tw/~meco/stat/four/a-8.htm
Export Growth Rate	http://www.moea.gov.tw/~meco/stat/four/a-7.htm
Economic Growth Rate	http://www.moea.gov.tw/~meco/stat/four/a-1.htm
Gross National Product	http://www.bea.gov/bea/newsrel/gdpnewsrelease.htm http://www.marketvector.com/leading-indicator/gnp.htm
Balance of International Payments	http://www.bea.doc.gov/bea/newsrel/transnewsrelease.htm

Table. 2 : The source of economic statistics.

REFERENCES

[1] Kohonen, T.: Self-organized Formation of Topologically Correct Feature Maps. Biological Cybernetics 43, 1982, pp. 59-59

[2] Guido Deboeck and Teuvo Kohonen: Visual Explorations in Finance with Self-Organizing Maps, Springer Finance Press, 1998.

[3] Sam T. Roweis and Lawrence K. Saul: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science, Vol. 290, December 2000, pp. 2323-2325.

[4] Joshua B. Tenenbaum, Vin de Silva and John C. Langford: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science, Vol. 290, December 2000, pp. 2319-2322.

[5] Christopher M. Bishop, Markus Svensen and Christopher K.I. Williams: GTM: The Generative Topographic Mapping. NCRG/96/015, April 16, 1997.

[6] K. V. Mardia, J. T. Kent, J. M. Bibby, Multivariate Analysis, (Academic Press, London, 1979).

ICONIP'2002, vol.2, pp. 787-791, Nov. 18~22, 2002, Singapore

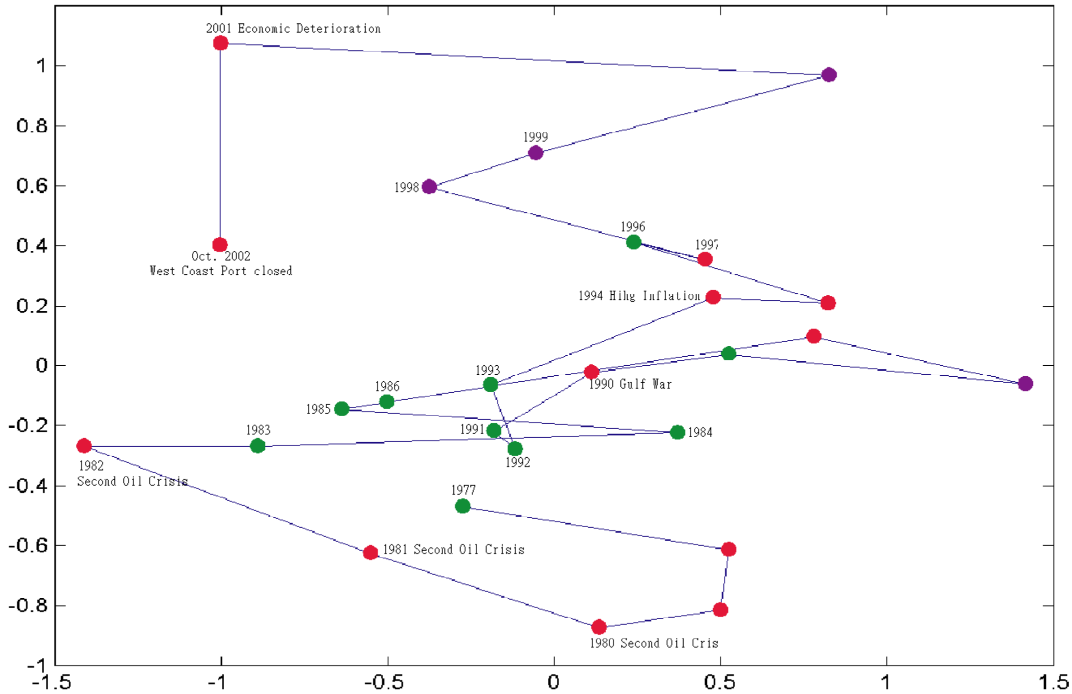

Fig. 2: U.S. economic statistic embedded from 11-D to 2-D using Isomap method. The neighborhood factor is set to 25. Red dots are years with important events. Green dots are years in the cluster center. Purple dots are years that fall to the edge.

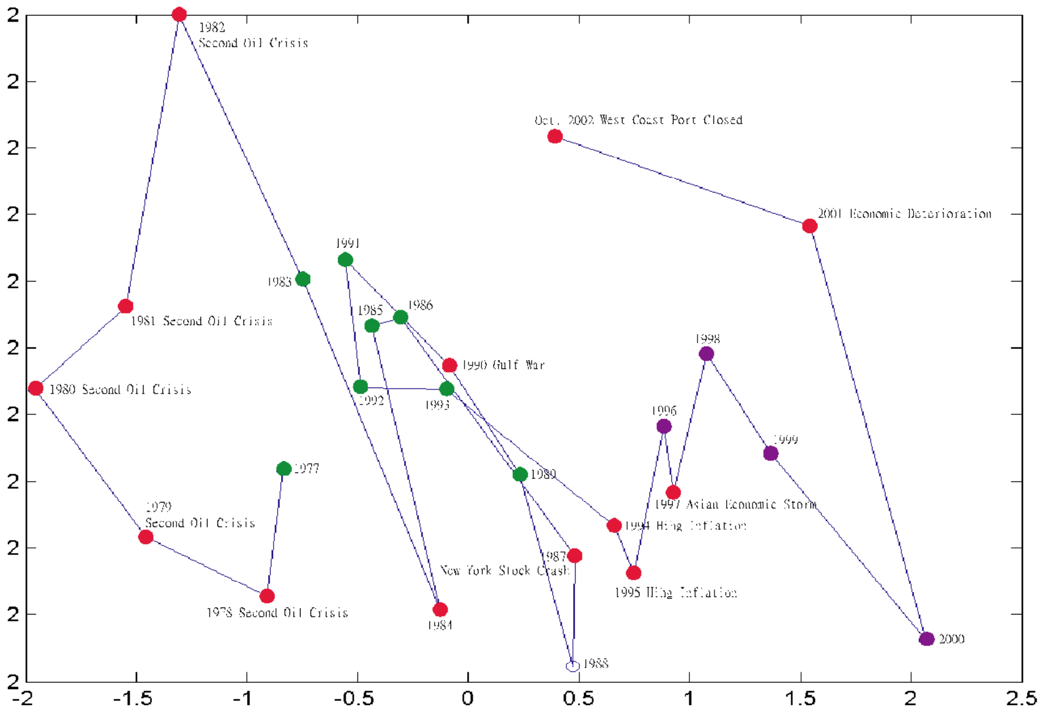

Fig. 3: U.S. Economic statistic embedded from 11-D to 2-D using LLE. The neighborhood factor is set to 10. Red dots are years with important events. Green dots are years in the cluster center. Purple dots are years that fall to the edge.

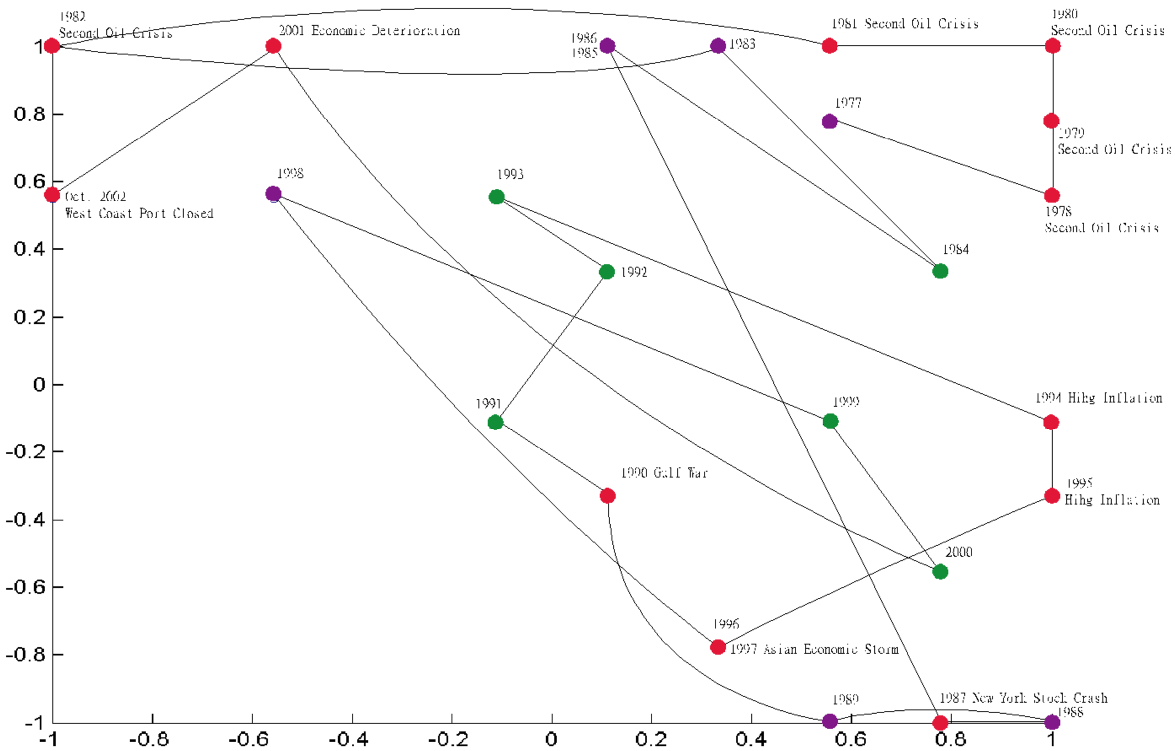

Fig. 4: U.S. Economic statistic embedded from 11-D to 2-D using GTM. The number of latent variable is set to 100. The number of basis function is set to 36 and the width of basis function is set to 4. Red dots are years with important events. Green dots are years in the cluster center. Purple dots are years that fall to the edge.