

A visual tool for mining macroeconomics data

D. Giordano & F. Maiorana

*Dipartimento di Ingegneria Informatica e Telecomunicazioni,
Università di Catania, Italy*

Abstract

Data mining environments need tools capable of aiding results comprehension, in particular by resorting to the power of visual perception. This paper presents a tool designed and implemented in MatLab™ to facilitate visual mining of a large set of macroeconomics data on the world import and export activity of seven countries, in order to extend an analysis already performed with the aim of building models of national specialization and identifying possible market outlets. A model of the “market” is built, by using a Multilayer Perceptron; then it is fed to a graphical interface which supports queries such as: can country X expand towards country Y in one particular sector? Answer messages point out, for the selected countries, sectors in which the market could be expanded, and for what specific products. As an aid to making sense of a given suggestion, a detailed market analysis can be carried out by resorting to a set of tabular and graphical views that are able to map the productive structure of each country and also the dynamics of its evolution through the years. The advantages of MatLab™ as a development environment are pointed out. It is argued that the additional effort for developing ad-hoc task oriented interfaces is especially justified in the case of frequent and interactive mining to be performed by decision makers.

Keywords: visual data mining, neural networks, MatLab, macroeconomics.

1 Introduction

One of the issues of interest in developing usable data mining environments is to provide the analysts with tools that aid model comprehension, quick analysis, and results interpretation. Flexible tools for data exploration and results visualization are one of the most appropriate solutions towards this aim. In Kopanakis and Theodoulidis [1] visual data mining is defined as involving the



invention of visual representations that could be applied in all three data mining life cycle stages, i.e., data preparation, model derivation and validation, and results interpretation, with the objective of enhancing information and knowledge flow throughout each stage. In particular, in data preparation visual data mining attempts to carry out visually some of the pre-processing; model specification by visual means should imply guidance in the generation of the model and a gain in model understanding; whereas in the last stage, visualization should aid in the interpretation of results. Various techniques to visualize mining outputs such as association rules, relevance analysis, and classification have been proposed (e.g., [1], [2], [3]); Aggarwal [4] has also proposed, for the clustering and nearest neighbour problems, techniques to interact with the intermediate results of the mining process. This avoids the constraints imposed by entirely automated computation, and takes into account more effectively the user needs.

Interactive analysis can also be enhanced by integrating visual data mining with powerful querying systems that can assist domain experts in the generation of what-if scenarios while interacting with visual displays (Ferreira de Oliveira and Levkowitz [2]). Visualization techniques can vary in complexity, ranging from 2D to 3D; research typically addresses both general and domain specific techniques. In any case, the general underlying principle is to resort to the human perceptual channel to offload some of the complexity inherent to the task.

This paper presents a tool that has been developed with the MatLab™ environment to facilitate visual mining of a large set of macroeconomics data regarding the import/export activity of seven highly economically developed countries. This set has been analyzed previously in order to identify possible market outlets (Bordoni *et al.* [5]). The paper is organized as follows. Section 2 describes the problem and the objectives of the current analysis. Section 3 illustrates the tool's functionalities and visual interface. Section 4 summarizes the results. Section 5 offers some concluding remarks.

2 Mining macroeconomics data

The use of data mining in the macroeconomics field has just recently been taken into consideration. Economic data are primarily of observational nature, and macroeconomics data, as opposed to microeconomics data, result from aggregating over dimensions (e.g., individual, firms, regions) at the local or national level. According to Feelders [6], two considerations make a special place for data mining in macroeconomics. First, this field usually recognizes the importance of exploratory analysis. Second, in spite of the criticism that model building with data mining is “atheoretical” with respect to economic theory, this approach to model specification does not make unrealistic assumptions about the relations between the economic quantities under study (which is a problem typical of structural modelling) and often leads to models that have a better predictive performance. If prediction is a valuable aim for macroeconomics, then data mining has a role to play, to be added to hypothesis generation to develop new theories, and quite different from generating models for theory testing.



A mining analysis of macroeconomics data was performed in Bordonì *et al.* [5], by mining a database which contained over 400,000 records regarding the imports and exports of seven highly economically developed countries (France, Germany, Italy, Japan, Netherlands, United Kingdom, United States of America) in the period from 1990 to 1998. The analysis was carried out with IBM Intelligent Miner, by unsupervised neural clustering techniques. As a result, national product specialisation sectors were identified, the evolution of the economic structures of the considered countries was traced and the similarities among the countries were highlighted. Furthermore, it was possible to characterize the countries which absorb the largest proportions of Italian products of the pharmaceutical and industrial automation sectors and identify other possible market outlets for these products. The obtained results were validated by the experience of ENEA Casaccia Research Centre's economists.

The analysis presented in this paper has been conducted on the same set of data, by developing with MatLab™ the tool DEC-LAB (Data mining for EConomics). Unlike the work in [5], the model of the market built is with supervised neural networks. Aim of the tool is to provide a user-friendly interface, specific for the domain and the task, that supports the whole analysis process, by resorting to simple and flexible visualization means. The tool was designed to be consistent with the basic process advocated by visualization theorists, i.e., "Overview first", "Zoom and filter", and then "Details on demand" (Keim [3]).

3 DEC-LAB: tool description

The tool can be used in two ways:

- 1) to build a model for the market and explore if it is possible to expand a nation's export volume towards other countries in a given sector;
- 2) to perform a detailed market analysis to map the productive structure of each country and the dynamics of its evolution through the years.

3.1 Model building and classification

Starting data are patterns containing economic information. A single pattern includes the following information: exporting country, year, exporting sector, product, importing country, region, economic trade amount. Products were grouped into fifty-one categories according to the SITC Rev.3 classification [7]. The data relative to the exports of each country were aggregated so as to pass from fifty-one categories to twelve product sectors which cover the whole manufacturing division. A first pre-processing of the data set was codification to obtain numeric input patterns; subsequently these were categorized as exporting and importing patterns.

After codification we have computed both the world export volume and world import for each country c in each commercial sector s :

$$V(\exp/imp)_{c,s} \quad (1)$$

These data are used during the neural network learning and testing phase.



In order to assign to each country the sectors in which it is more competitive, the following indexes have been computed, both for the export and for the import patterns, to take into account the weight of a sector in the internal and in the world economy:

$$V(\text{exp}/\text{imp})_{c,t} \quad (2)$$

For each country, (2) is the internal export/import index, computed as the sum of all trade volume over all sectors (c as country index, t as total). This index represents the total trade amount of each country.

$$V(\text{exp}/\text{imp})_{w,s} \quad (3)$$

For each sector, (3) is the world export/import index, computed as the sum of all trade volume over all countries (w as world, s as sector index). This index represents the world trade amount in each sector.

$$I_{c,s} = \frac{V(\text{exp}/\text{imp})_{c,s}}{V(\text{exp}/\text{imp})_{c,t}} \quad (4)$$

For each country and each sector, (4) is the Internal index, computed as a sector export/import percentage over the total export/import volume of the country.

$$W_{c,s} = \frac{V(\text{exp}/\text{imp})_{c,s}}{V(\text{exp}/\text{imp})_{w,s}} \quad (5)$$

For each country and each sector, (5) is the World index, computed as percentage of export/import volume for the given country in a given sector over the world export/import volume for the given sector.

It is assumed that if the internal or world exp/imp index for a certain sector is under a threshold (0.15) then the country isn't specialized in that sector.

For every nation
For every sector

If nation i internal index value in the sector j $I_{i,j}$ is equal to nation maximum internal index

for every sector, **or if** $I_{i,s} \geq \text{threshold}$ **then** assign to target for the nation i and sector j , the considered sector, and consider as export volume $I_{i,j}$;

else if $I_{i,j}$ is lower then threshold then

if nation i world index value in the sector j $W_{i,j}$ is equal to every nation maximum world index in the sector j , or if $W_{i,j} \geq \text{threshold}$ then assign to target for the nation i and sector j , the considered sector, and consider as export $W_{i,j}$;
else assign to target for the nation i and sector j zero and consider as export volume $I_{i,j}$.

Figure 1: Algorithm to create the target.



With these data a model of the “market” is built, by using a Multilayer Perceptron (MLP) trained and tested on the historical data concerning import and export volumes of each country, towards 15 geographic areas or economic subjects. The target (used to train the neural net and then for classification) was created with the algorithm illustrated in Figure 1.

After pre-processing, the learning and testing patterns will contain the following information: exporting/importing country, exporting sector, internal or world index (as assigned by the target algorithm), target.

The tool uses a simple user interface (Figure 2). After loading the learning and testing patterns from the menu Patterns, the neural net simulation and test can be performed. We created a network with an hidden layer and an output layer, three input neurons, five neurons in the hidden layer, one neuron in the output layer, by using the log-sigmoid transfer function between the layers, the Levenberg-Marquardt as the Back-propagation network training function, the Gradient descent with momentum weight and bias as Backpropagation weight/bias learning function and Mean squared error performance function.

	Aerosp.	Medical eq.	Autom.	Chemistry	Elect. cir.	Elect. c.	Pharm.	Off.	Plast.	P. I.	Opt.	L.T.
ITA	3	0	0	0	0	0	0	5	0	0	0	0
FRA	5	0	0	5	0	0	1	3	3	0	0	0
JAP	0	3	1	3	0	2	0	0	0	0	0	0
GER	3	5	5	4	5	5	5	5	5	5	5	0
USA	5	5	3	5	5	5	5	5	5	5	5	0
UK	0	0	0	3	2	2	5	2	0	0	0	0
HOL	0	0	0	0	0	0	0	5	5	0	0	0

Figure 2: DEC-LAB Tool's interface.

Aim of network simulation is economic patterns classification. We trained the network in five years from 1990 to 1994, and used the MatLab Neural Toolbox to generate the error diagrams to evaluate the network learning performance. The classification's results for the learning phase are visualized in a table in the lower part of the tool's interface (Figure 2). In the columns there are the 12 sectors in which data were aggregated, i.e., aerospace, electro-medical equipment; industrial automation; fine chemicals; electronic components; consumer electronics; pharmaceuticals; office machines; plastics; precision instruments; optical material and instruments; low tech. Positive values in the cells point out

how many times, in the time span considered, each country was classified as specialized in the corresponding sector. An analogous table is generated for import and it can be visualized by pressing the Imports radio button. In this case positive values indicate that the country is “specialized” in importing in the corresponding sector. When the model is applied to novel patterns these tables are substituted by those ones generated from the new classification.

	Aerosp.	Medical eq.	Autom.	Chemistry	Elect. cir.	Elect. c.	Pharm.	Off.	Plast.	P. I.	Opt.	L.T.
ITA	0	0	0	0	0	0	1	0	0	0	0	0
FRA	1	0	0	1	0	1	1	1	0	0	0	0
JAP	0	0	0	0	1	1	0	0	0	0	0	0
GER	1	1	1	1	0	1	1	0	1	1	1	0
USA	1	1	1	1	1	1	1	1	1	1	1	0
UK	0	0	0	0	0	1	1	0	0	0	0	0
HOL	0	0	0	0	0	0	0	1	0	0	0	0

Figure 3: Market structure for export in 1998.

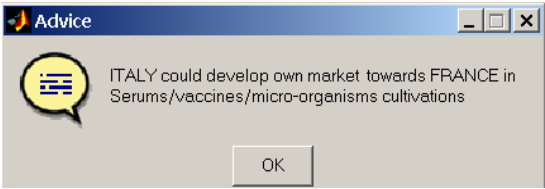


Figure 4: Advice for specific products.

The testing phase was performed with 1995 economic data. From the analysis of error as difference between real and simulated output, the error histogram and residue autocorrelation we can argue that error is quite small, with few exceptions. After testing we used the model to classify data from 1995 to 1998. Figure 3 shows the resulting classification table for export for the year 1998.

Query formulation is assisted by the table showed in Figure 3 and by its equivalent for import. These tables facilitate, respectively, identification of the couples of exporting and importing countries, which then can be specified in the combo-boxes in Figure 2. By clicking the “Advice” button, the tool generates messages that point out, for the selected countries, the possible sectors in which the market could be expanded, and for what products. Messages are generated by checking the data set to match products of the importing country for which import is high with products of the exporting country for which export is high. For example, our analysis showed that in 1998 Italy could expand its exports in the pharmaceutical sector. The message box in Figure 4 specifies the product in which Italy could expand towards France. The interface thus supports a dialogue-oriented analysis.

3.2 Market analysis

In the second part of the work, we used a simple user interface, based on a table, to visualize the market situation and perform a detailed market analysis that aids

the user in making sense of the suggestions generated from the first analysis. Starting from the same data set used for the first analysis we have computed, for each country, the total export amount for each sub-sector towards the following economic and geographic zones: Europe, ex Urss, East Europe, Africa, Asia, Oceania, Latin America, Efta, Nafta, Japan, Turkey, Cyprus, Israel:

$$V(\exp/imp)_{c,p} \quad (6)$$

The screenshot shows a software window titled 'Markets' with a menu bar (File, Reset, Exit). It has two dropdown menus: 'Select country' (set to ITALY) and 'Select sector' (set to Electronic component). There are three input fields for 'Vexp low', 'Vexp med.', and 'Vexp High'. An 'OK' button and a 'Shows Volumes' button are present. Below is a table titled 'Exp. / Imp. volumes (thousand €)'.

	Elect. circ.	Activ. comp.	Cath. tubes						
Ex URSS	397	800	1743						
E. Europe	6112	6972	14516						
U.E.	198651	986302	278245						
Asia	10444	452932	140169						
USA									
Africa	4143	3702	6189						
Latin A.	3405	2026	6455						
EFTA	12232	11390	417						
NAFTA	9825	110257	12265						
Oceania	706	3060	2531						
Italy									
Japan	124	46342	517						
Turkey	3148	5107	151306						
Cyprus	0	57	67						
Israel									

Figure 5: Interface for market analysis.

These data are shown in a tabular view that indicates the economic trade amount in thousands, for each economic and geographic zone (rows), and for each product (columns), for the country and the sector selected by user (Figure 5). This table can be used to find new markets as follows. By visual exploration of the table, one can find minimum, medium and maximum trade amount. By flexibly choosing the thresholds which determine if the export values for a given product are to be considered low, medium, or high, regions can be dynamically classified accordingly in three clusters. By clicking the Show Volumes option, the numerical information of Figure 5 can be substituted by the results of this latter classification. At this point, the user can gain a better understanding of the viability of a suggestion that has been proposed by the tool in the message boxes (e.g., Figure 4). One can check, for example, who are the competitors, for a given product, in the geographic or economic area of the country that is being analyzed as a possible new market outlet. Also, one can gain an understanding of a country's positioning in the global market, by examining if it has high volumes of export of a product mostly in less developed areas, or in technologically advanced areas. This latter situation could indicate production of competitive, state of the art products, whereas the former one indicates the contrary. If so, an

attempt to expand towards a country already importing from country renowned for the quality of its products is not likely to be successful.

To complement this analysis, the tool constructs a set of graphical views that map the productive structure of each country by showing the dynamics of its evolution through the years, allowing to focus and further reflect on specific products of a sector or exports towards a specific country. As an example of such views, Figure 6 reports for a selected country (Italy) and sector (Electronic components), export evolution of each sub-sector towards Netherlands from 1995 to 1998. Similar views were generated to show export evolution of a selected country towards each other country for the years 1995-1998.

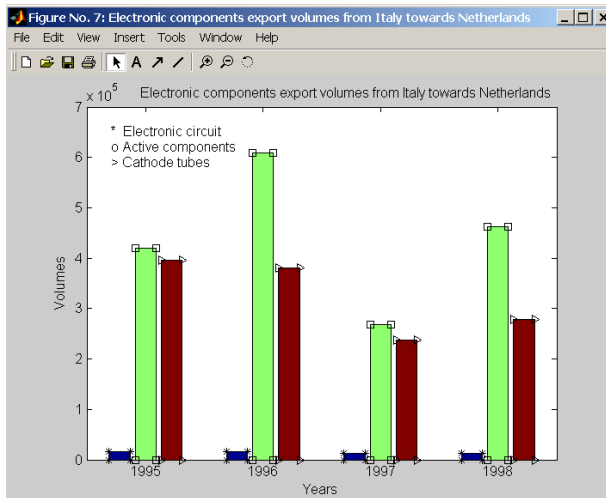


Figure 6: Italian trade evolution in electronics from 1995 to 1998 towards Netherlands.

4 Comparison with former results

The comparison with the analysis performed in Bordonì *et al.* [5] can be carried out at two levels, one concerning the qualitative suggestions obtained, and one related to process and ease of use. The results regarding national specializations, obtained by applying the multilayer perceptron, were generally compatible with those obtained in [5], where unsupervised neural cluster analysis was used. In [5] it was found that European countries, particularly France and the United Kingdom, but also the Netherlands and Germany, are strongly specialised in the advanced chemical and pharmaceutical sectors. This result was confirmed, also obtaining the further discrimination that advanced chemical is more a prerogative of Germany and France, and the additional information that also UK and Italy are to be considered specialized in the pharmaceutical sector. In the former analysis, Italy was not considered specialized in pharmaceuticals, but the sector was one of the two best scoring ones and thus was identified as one on which to

focus for market expansion. Another result in [5] was that Japan and in part Netherlands and UK aim very much at consumer electronics, whereas the United States and UK are the only countries among the seven taken into consideration who are specialised in producing and exporting helicopters and aeroplanes. In DEC-LAB the major findings for specialization were retained, but some were lost, namely, UK specialization in aerospace, and Netherlands specialization in advanced chemicals and pharmaceuticals. These variations can be explained as an effect of the threshold set to 0.15 in training the model, as is shown by the classification table appearing in Figure 2, which indicates that, in the five years used for training, UK and the Netherlands were never classified as specialist in the above mentioned sectors. To mitigate this threshold effect, visualization of the classification table could be modified to include information about the degree of specialization (very high, high, medium, none) of a given country. This can be accomplished by modifying the algorithm for calculating the target, and, accordingly, considering a network architecture with two output neurons to codify four levels of the specialization degree.

Regarding the results obtained in [5] for the identification of possible market outlets, only a partial comparison can be done. In fact, there are basic differences in the approach adopted for identification, and in the information used. In [5] further information to characterize the market, such as the opening up of trade, percentage of growth of the gross domestic product (GDP), variation of the exchange rate, presence or absence of customs unions with Italy was used to build a model, by demographic clustering, of the countries for which Italy's market share in the pharmaceuticals and industrial automation fields (where Italy scored well as specialized) was above the average. The obtained model was applied to those records of 1998 which had been excluded from the first phase, in order to identify countries that mapped well on the clusters.

The current version of the DEC-LAB tool was developed to operate on data that did not include the above mentioned information. The logic used to implement the rules that generate the suggestions for new market outlets is simply based on the result of a model of specialization in export and import and on the indexes described in section 3. Still, the two approaches can be compared from the user's point of view. A shortcoming of the clustering method used in [5] was that, to visualize the similarities in the productive structures of the countries and the dynamics of their evolution, it was necessary to filter the data year by year, obtain a sequence of instances and look into the cluster structure of each year in order to understand in which direction the economies of the seven countries were evolving. DEC-LAB, regarding market analysis, affords the advantage to provide integrated views of the market evolution, instead of having the user assemble different snapshots. Another shortcoming of the clustering method was that the main features that concurred to form the cluster did not take into account specific indications related to the products. Thus the provided information was highly qualitative, referring mostly to country identification, and had to be followed up by more detailed investigations to transform it into practical indications. Instead, the analysis mechanisms implemented in DEC-LAB is able to provide, on demand, very specific suggestions concerning not



only the sector but also the products. Thus with respect to the analysis conducted in [5], significant advantages from the user's point of view are an easier analysis process and results interpretation. DEC-LAB could be improved by allowing the user to incorporate novel information regarding variables that may affect the market and by implementing more specialized rules, possibly based on fuzzy logic, to generate the suggestions.

5 Concluding remarks

Often, the visualization mechanisms offered by commercial data mining environments lack flexibility and may present problems such as label overlapping, cluttering, cryptic representations. The tool described in this paper has been developed according to an approach that emphasizes the role of appropriate visual representations. It also aims at providing an intuitive user interface specifically designed to support a mining activity whose steps have already been defined and targeted to the domain and goal of the analysis. The advantages of MatLab™ as a development environment for this kind of tools are: the availability of a complete set of neural functionalities suitable to various mining tasks that can be activated with a line of code; capability of generating visualizations much more sophisticated than those described in this paper; an easy to develop graphical user interface. It is thus fairly simple to design personalized interfaces that are consistent with good visualization principles, e.g., overview, zoom and filter, details on demand, and also to support dialogue with the user.

Although the proposed approach involves the cost of developing ad-hoc tools, it is particularly viable when mining results should be used in an interactive and repetitive fashion, for example whenever a model must be re-generated for obsolescence but the analysis activity and the ways to interact with the model aren't likely to change. It may also afford the advantage of direct involvement of the decision makers in using the model. Usually, because of the complexity of commercial mining environments, decision-makers are just consumers of the final outcomes, and may miss the more solid understanding of the results which is developed by those who carry out the whole mining process. Thus it can be argued that also available commercial mining tools should take into account the possibility of allowing the development of easy to use, task-oriented, dialogue-oriented user interfaces, for effective use of the models by decision-makers.

References

- [1] Kopanakis, I. & Theodoulidis, B., Visual data mining modeling techniques for the visualization of mining outcomes. *Journal of Visual Languages & Computing*, **14**(6), pp. 543-589, 2003.
- [2] Ferreira de Oliveira, M.C. & Levkowitz, H., From visual data exploration to visual data mining: a survey. *IEEE Transactions on Visualization and Computer Graphics*, **9**(3), pp. 378-394, 2003.



- [3] Keim, D., Information visualization and visual data mining, *IEEE Transactions on Visualization and Computer Graphics*, **8(1)**, pp. 1-8, 2002.
- [4] Aggarwal, C.C., Towards effective and interpretable data mining by visual interaction, *SIGKDD Explorations*, **3(2)**, pp. 11-22, 2002.
- [5] Bordoni, L., Giordano, D., & Spadaro, S., Il data mining: un'applicazione agli studi macroeconomici. *Atti del convegno AICA 2002* (Associazione Italiana Calcolo Automatico), pp. 561-57, 2002.
- [6] Feelders, A.J. Data mining in economic science. Available on line at <http://www.cs.uu.nl/people/ad/dmecon.pdf>.
- [7] Standard International Trade Classification: available on line at <http://unstats.un.org/unsd/class/family/historical/sitc/default.htm>.

