NONLINEAR STRATEGY TO CLASSIFY TIME SERIES OF THE SEMICONDUCTOR MARKET TREND

M. Bucolo[1], F. Caizzone[2] , L. Fortuna[1], L. Galvagno[1] ,G. Tomarchio[1,2]

[1] *DIEES Universita' degli studi di Catania, e-mail mbucolo@diees.unict.it*
[2] *STMicroelectronics Catania Site, e-mail giusy.tomarchio@st.com*

Abstract: Semiconductor firm sales are a complex function of what direct customers require and how efficiently the company is able to satisfy them. The task of solving the management of production planning with a wide products portofolio is not an easy one.
The paper deals with the adoption of an unsupervised clustering strategy to classify products, not just in relation with the business parameters, but also by considering the historical evolution of sales and customer demand. Different nonlinear techniques have been considered to face the problem both from the mathematical point of view than from the economic one. Particularly, this strategy has been implemented to classify the products of Discrete and Standard Ics Group of the worldwide semiconductor firm, STMicroelectronics in relation to their sales versus the different customer type.
Two different approaches have been used: the hierarchical clustering with an optimization procedure and GHSOM structures. The results show the validity of the cluster strategy, in fact the time series carrying on the same qualitative information are grouped, and the flexibility of the GHSOM. Moreover, the GHSOM, as a clustering technique, requires a low supervision level and brings a reduction of the problem complexity and speeds up the process.*Copyrigh©2005IFAC*

Keywords: Time series, Decision Supporting System, Products, Neural-Networks, Unsupervised Clustering.

1. INTRODUCTION

Semiconductor firms, as well as any other indirect supplier, are situated in a supply chain that starts from the final customer and ends with the raw material supplier. The company is at the end of a supply chain starting from the final consumer, therefore the sales are obviously the result of what the final user requires, propagated along the whole supply chain (Bucolo, et al., 2002).
In the supply chain between the final customer and the semiconductor firm, several other manufacturing firms act with different functions like distributors, direct assembler, and good producers. These firms represent the direct customers of the semiconductor firm.

Their demand versus the semiconductor firm is modulated both in amplitude and phase as a result of their different production strategies and production parameters like time cycle, stock policies and production capacity; nevertheless, an important role is played by the capability of these customers to forecast their own final demand (Bucolo, et al., 2003).
Semiconductor firm sales are a complex function of what direct customers require and how much and how fast the company is able to satisfy them.
In this work the products of the Discrete and Standard Group (DSG), of STMicroelectronics, constitute the case study. This group is organized in 15 Product Families (P&L) that have different business models in terms of sales distribution along worldwide regions, class of customers, market and applications; moreover

the product families' sales are spread over a very wide product portfolio.

The aim of this work is two fold: to develop a classification strategy to identify the different behavioural classes of products with the same qualitative evolution in sales to be used in a long term perspective and to model each class in order to build a forecasting system of sales at product level (Bucolo, et al., 2003).

As described in Section 2, the case study under consideration is the sales of one of the DSG product family, and by highlighting how these sales span over in the different sale channels.

The classification strategy that has been developed consists of two phases. The first phase, reported in Section 2, represents more than simple preprocessing phase because it gives the possibility to make an initial screening among the products and to focus on the final target of the classification.

In the last phase the Growing Hierarchical Self Organizing Maps (GHSOM), described in Section 3, have been implemented as final step of this strategy (Dittenbach, et al., 2002).

The GHSOM, as a clustering technique, requires a low supervision level; to perform the classification it is only necessary to set two parameters that are strictly in relation with the granularity of the clustering and the nature of the time series.

The obtained results, shown in Section 4, highlight a big reduction of the problem complexity. This strategy can represent an important step for other specific types of analysis like sale forecast of products, but carried on by class and no longer by time series.

2. THE CASE STUDY

The case study that has been analysed is related to the sale channels of the Discrete and Standard Group (DSG) of STMicroelectronics, one of the world's biggest semiconductor firm. This group is organized in 15 Product Families that manage very standard and commodity products as well as very innovative and application specific ones [3]. Particularly, the DSG sales have, mainly, two different sales channels, distribution (here indicated as dist) and manufacturing customers (here indicated as no dist).

In this work the three variables that have been considered are in relation with the two different sale channels for a specific product family of the DSG, the Voltage Regulators.

1. *billing_dist*: the billing for the distribution channel;
2. *billing_no_dist*: the billing for the no distribution channel;
3. *resale*: the resale of the Distributors.

Moreover, the resale variable has been introduced to characterize the sales of distributor customers. The samples collected are over a monthly time series in the period between June 2002 and March 2004, therefore there is an availability of 11 points for each variable. The initial analysis has been carried out in three different steps.

In the first phase all the time series have been filtered to avoid the presence of false values and all invalid data has been replaced with zero values. The second step, the data segmentation, is the most significant one in this pre-analysis. During this phase a first screening of the product has been performed in relation to an economical parameter: the selected products represent 90% of the business volume. The segmentation phase has been carried out by using the "pareto rule". The average business volume ( ) has been computed for every product and the products have been sorted out starting from the highest value to the lowest one. The cumulative percentage upon all average values has been computed for each level of this list. Specifically, the cumulative percentage at j level, per_cum(j), upon all products is evaluated through the equation reported in eq. 1:

$$per\_cum(j) = \frac{\sum_{k=1}^{j} \mu_k}{\sum_{i=1}^{n} \mu_i} \qquad (1)$$

where μ_k is the average value of the product of the list over the j level, meanwhile n is the total number of products. The products whose cumulative percentage is more than the predefined threshold have been chosen.

The last step of this data analysis consist in a pre-processing phase and has been performed by using two different possibilities in relation to the financial target of the classification.

If the adopted classification is by business volume then a scaling processing of the data has to be applied. Since the original proportions between time series have to be maintained, a suitable scale factor has been considered, in a way that the dynamics of all the time series are reported between 0 and 1.

Instead if it is necessary to focus the attention on the trend of the time series, the preprocessing procedure consists on data normalization: a linear function is applied to each time series so that its minimum and its maximum are, respectively, 0 and 1.

For each variable two different views have been considered, therefore the billing_dist has been analysed by business volume and by trend, just like the billing_no_dist and the resale.

3. GROWING HIERARCHICAL SELF ORGANIZING MAPS

The GHSOM algorithm is based upon a structure composed of Self-Organizing Maps (SOM) that grow, by increasing the number of nonlinear units, until a quality criterion of the classification has been satisfied (Dittenbach, et al., 2000). The growing process builds a hierarchical structure composed by SOM in which lower layers classify data with more granularity, as shown in Fig. 1.

The first layer of the hierarchy is constituted by only one SOM meanwhile in the next layer a new SOM can be connected to any nonlinear units of the first map. This link is repeated with each map of the each layer versus their previous layers, until a performance target is reached.
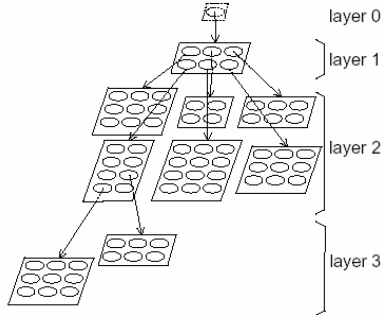

Fig. 1- Hierarchical structure of a GHSOM.

3.1 Self Organizing Map

The fundamental element of a GHSOM structure is the Self-Organizing Map (SOM) that is a structure composed of nonlinear mathematical units, called neurons that are organized in a grid and connected to data inputs. The number of neurons may vary from few dozens to thousands of elements. Each nonlinear unit is represented by a weight vector, or prototype vector, , where n is the length of the input vectors. All these units are connected to their neighbours through neighbourhood relations that characterize the topological structure of the map and can be rectangular or hexagonal.

By using the SOM two phases can be distinguished, a learning phase in which the SOM is adapted to the input vectors, and a recognizing phase.

At the end of the learning phase the SOM is a regular neuron grid, where neighbour units have similar weight vectors. In the recognizing phase, by introducing an input vector in the map, its best-adapted region, that means all the neurons that have the minimum distance between the prototype vector and the input vector, has been activated.

The SOM learning phase is based on a sequential learning algorithm, in which at each step, a vector x from the input data set is chosen and all distances between x and all weight vectors m are computed using a predefined metric. The neuron c, whose weight vector has a minimum distance with x, is called the Best-Matching Unit (BMU), and has the following properties:

$$\|x - m_c\| = \min_i \left\{ \|x - m_i\| \right\} \quad (2)$$

where   is the chosen metric to compute the distance, generally the Euclidean distance is used. Therefore, the equation (2) can be rewritten as follows:

$$\|x - m_c\| = \min_i \left\{ \sum_{k=1}^{n} (x_k - m_k) \right\} \quad (3)$$

where n is the length of the input vectors.

In the next step, all weight vectors of the map are refreshed in relation to the distance from the BMU neuron. This procedure moves the BMU neuron and its neighbours towards the input vector, as shown in Fig. 2. The refresh rule for the weight vectors of BMU neighbours is:

$$m_i(t+1) = m_i(t) + \alpha(t)h_{ci}(t)[x(t) - m_i(t)] \quad (4),$$

where t is the learning step, $h_{ci}(t)$ is a kernel function relative to the BMU neighbours, usually the gaussian function, and $\alpha(t)$ is the learning speed. The kernel function $h_{ci}(t)$ is related to the distance between the i-th unit from the BMU. This function defines the region of influence of the input vector on the map.
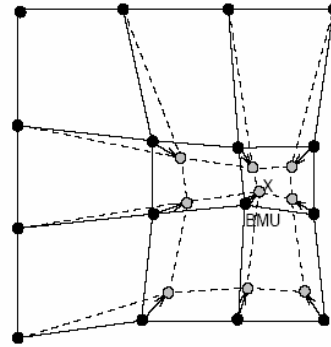

Fig. 2 –The BMU neuron and its neighbors in the map.

3.2 GHSOM Algorithm

The learning algorithm of a GHOSM structure starts with a layer 0 constituted by a SOM with only one neuron. The weight vector of this neuron is initialised with the mean of all vectors of the classifying data set. The learning process goes on with a SOM map in the layer 1 of 2x2 neurons that organize itself with the SOM learning algorithm, by associating the input vectors with the different map regions.

This training process is repeated for a number n of iterations that represents the number of the input vectors. After n iterations the neuron with the highest deviation between its weight vector and the input vectors related to this region, is chosen as the error unit. Between the error unit and its most dissimilar neighbour a new row or a new column of neurons is inserted, as shown in Fig. 3. The weight vectors of these new units are assumed to be equal to the neighbours mean.
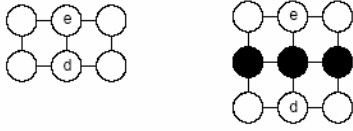
Fig. 3- The insertion of a new row between error unit
(e) and its most dissimilar neighbour (d).

The quantization error qi guides the learning process, which is the sum of the distances among the weight vectors of the i-th unit and the input vectors mapped on this unit:

$$q_i = \sum_{j=1}^{p} \left\| w_i - I_j \right\| \qquad (5),$$

where m_i is the weight vector of the *i-th* unit and I_j is one of the *p* input vectors.

The q_i is necessary to evaluate the Mean Quantization Error (MQE) of the neurons in the map.

$$MQE = \frac{1}{m} \sum_{j=1}^{m} q_j \qquad (6),$$

where N_m is the number of neurons of the map.
A map in the "*layer j*" grows by the introduction of a new row (or column) until its MQE is reduced by a fraction τ_1 of q_i of the *i-th* unit, to which it is connected in the previous "*layer j-1*" of the hierarchy:

$$MQE \leq \tau_1 \, q_i \qquad (7).$$

The map in the *j* layer now represent the input data associated to the *i-th* neuron in the *j-1* layer with more details.
The initial GHSOM architecture is formed by only one SOM, this architecture grows by another layer if some dissimilar input data is mapped on the same neurons. These neurons are characterized by a high quantization error that is higher than the threshold. This threshold basically indicates the depth level of the hierarchical structure for the data representation, and is a fraction of the quantization error in the layer 0.
A new map is added to the hierarchy when the quantization error of unit i at layer j is higher than a fraction τ2 of the quantization error in layer *0:*

$$q_i \geq \tau_2 \, q_0 \qquad (8).$$

Therefore the input data mapped to neuron "*i*" are mapped in their respective map "*layer j+1*" that grows until its MQE is reduced of a fraction τ_1 of the quantization q_i of the *layer j*.

4. RESULTS: PRODUCTS CLASSIFICATION

The product classification in relation to the different sales channel has been carried out by following all the steps of this strategy. The first significant results have been obtained through the segmentation of the data set performed by using the pareto rule. By considering all the products that represent 90% of the business volume for each sales channel a relevant reduction of the number of products has been considered. Specifically, as reported in Table 2, only the 12% of the products has been selected for the distribution channel, while 17% for the no-distribution one. The reduction increases and just the 7% of the initial set of products are useful for the resale variable.
This strong reduction of the initial data set is fundamental to speed up the following phases of the analysis. As reported in section II the same variable has been studied under two different points of view. In the first case a classification based on the business volume has been performed by scaling the variable; while in the second case a normalization pre-processing phase has rendered possible a classification based on the historical trend. A detailed description of the results obtained for each channel, is reported following.

4.1 Results: Resale

Before the actual classification, an important study has been carried out focus the attention on how the results of clustering can change in relation to a different value of parameters τ_1 and τ_2.
The results of this analysis are reported in Table 1, in which the obtained numbers of classes versus different τ_1 / τ_2 are compared. The variable considered is the resale by trend.
By changing the τ_1 / τ_2 threshold it is possible to obtain a very deep hierarchy with little maps or a less deep hierarchy with big maps.
The result of the learning process proves that a balanced structure is not necessarily the best solution. Nevertheless, a structure in which all branches have the same depth is a good one, because the hierarchy adapts itself to the input data. There are also areas, in the input space, that require more units for their representation and create branches, which are deeper than others.
A good classification has been obtained by using values τ_1=0.5 and τ_2=0.1. With these values the map has the possibility to grow in both directions, with a few preferences for a horizontal expansion rather than a vertical one.

Table1- GHSOM topologies by varying τ1 and τ2

τ_1 / τ_2	1	0.2	0.1
1	4 clas. /1 map	4clas. /3 map	4 clas. /4 map
0.8	4/1	4/3	4/4
0.6	14/1	14/3	14/4
0.4	23/1	23/1	23/1
0.2	31/1	31/1	31/1

In this first classification the resale variable by business volume and by trend has been analyzed. Amongst 727 products that represent this variable only 56 represent 90% of the business volume.

It is possible to represent the resale by business volume data set with 4 classes, while it is necessary to consider 21 classes by trend, as reported in Table 2.

The need to have more classes in a classification by trend is coherent with the economic meaning of this analysis.

In the first case the obtained classes have 4 ranges of mean business volume, so monthly pieces, that can be clearly distinguished:

- ❑ Class 1: average range $[0\text{-}2.5] *10^5$;
- ❑ Class 2: average range $[2.5 \text{-}4] *10^5$;
- ❑ Class 3: average range $[10 \text{--}15] *10^5$;
- ❑ Class 4: average range $[20 \text{-}40] *10^5$.

However, this classification requires more detail in the trend analysis. In Fig. 4 the two classes that have been obtained, have been plotted respectively in the top subplot class 2 and in the bottom subplot class 15.

The blue bold line in each subplot is the centroid of the class; class 2 has two products, while class 15 three products. The time series in both cases are really close.
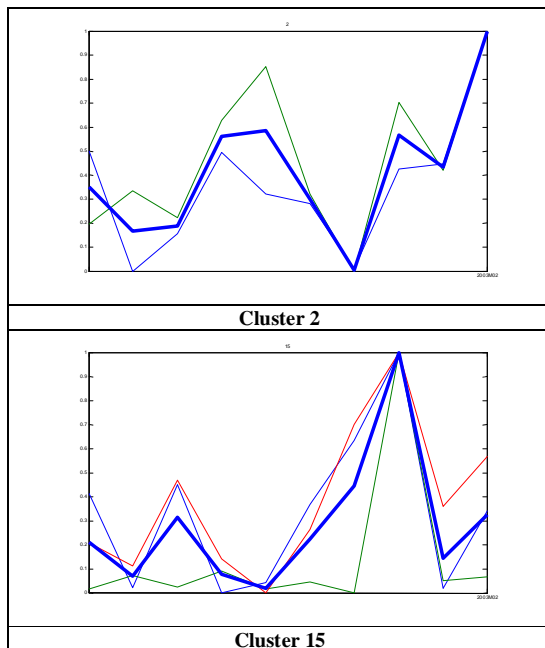

Cluster 2

Cluster 15

Fig. 4- Resale by trend: two classes.

4.2 Results: Billing_dist

Among 830 products that represent this variable only 95 are included in 90% of the business volume.

It is possible to represent the data set by business volume with 6 classes, while it is necessary to consider 33 classes for the billing distribution by trend, see Table 2.

The obtained classes have 6 ranges of mean business volume in terms of monthly pieces, as follows:

- ❑ Class 1: average range $[0.1\text{-}2.3] *10^4$;
- ❑ Class 2: average range $[1 \text{--}5] *10^4$;
- ❑ Class 3: average range $[5\text{-}15] *10^4$;
- ❑ Class 4: average range $[1.5\text{-}2.4] *10^5$;
- ❑ Class 5: average range $[4\text{-}15] *10^5$;
- ❑ Class 6: average range $[20 \text{--}35] *10^5$.

In Fig. 5 the two classes that have been obtained, are plotted, respectively in the top subplot it is reported class 18 with three products and in the bottom subplot class 2, also with three products. The blue marked line in each subplot is the centroid.
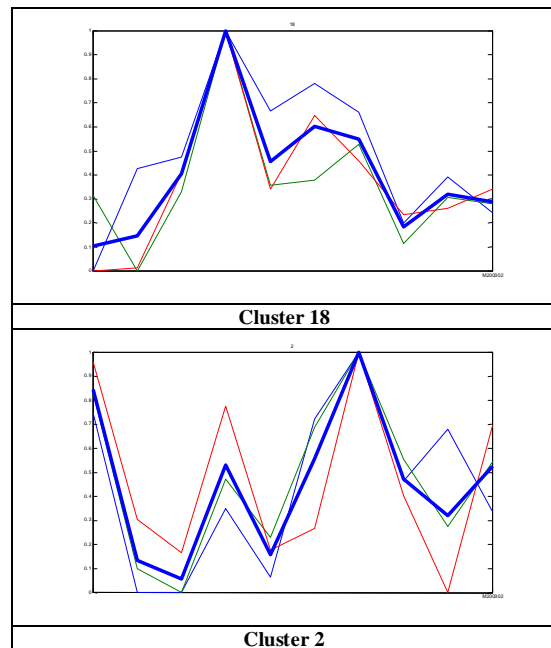

Cluster 18

Cluster 2

Fig. 5- Billing_dist by trend: two classes.

4.3 Results: Billing_no_dist

Amongst the 617 products that represent this variable only 108 are included in 90% of the business volume.

Also in this case it is possible to represent the data set by business volume with 6 classes, 33 classes for the billing_no_dist by trend, see Table 2.

The obtained classes have 6 ranges of mean business volume in terms of monthly pieces, as follows:

- ❑ Class 1: average range $[1\text{-}2] *10^5$;
- ❑ Class 2: average range $[2 \text{--}3] *10^5$;
- ❑ Class 3: average range $[3\text{-}5] *10^5$;
- ❑ Class 4: average range $[1.5\text{-}2.4] *10^5$;
- ❑ Class 5: average range $[20\text{-}35] *10^5$;
- ❑ Class 6: average range $[40 \text{--}50] *10^5$.

In Fig. 6 the two classes that have been obtained, are plotted, respectively in the top subplot it is reported class 1 with five products and in the bottom subplot

class 2, with three products. The blue marked line in each subplot is the centroid.
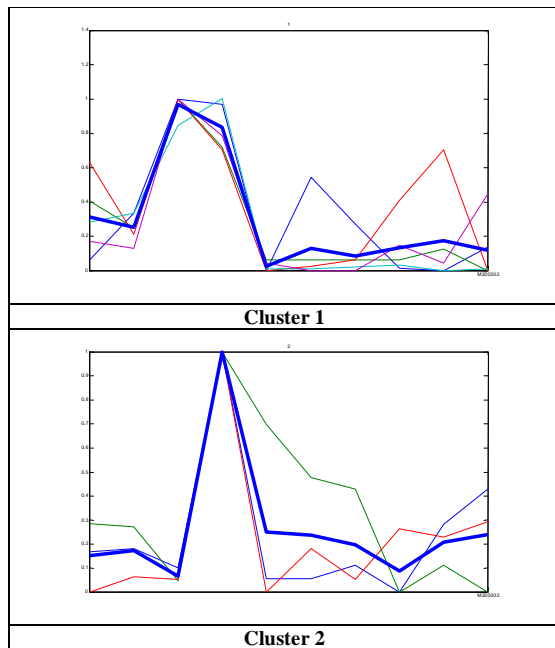
Cluster 1

Cluster 2

Fig. 6- Billing_no_dist by trend: two classes.

5. CONCLUSIONS

The aim of this work is to develop a strategy to classify the different behavioural clusters of products with the same qualitative sales evolution.

Particularly an unsupervised clustering strategy to classify products by considering the historical semiconductor firm, for the two main sales channels: distribution and direct manufacturing customers.

The applied clustering technique is the GHSOM that requires low supervision level and therefore they speeds up the process.

The results show that it is possible to find homogeneous classes in terms of qualitative information resulting in a reduction of the problem complexity.

Finally this approach can be used in a long-term perspective for the modelling of each class in order to build sales a forecast system at product level.

evolution of the sales has been described. Different nonlinear techniques have been considered to face the problem both from the mathematical point of view and the economic one. This strategy has been performed to classify the products of Discrete and Standard Ics Group of the STMicroelectronics, a worldwide

REFERENCES

M. Bucolo, F. Caizzone, L. Fortuna, G. Tomarchio, 'Nonlinear Models for Semiconductor Market Forecast', *Nonlinear Phenomena in Complex Systems*, Vol. 6, No. 2, pp. 669-677, Education and Upbringing Publishers, National Accademy of Science Belarus and Belarusian State University, 2003.

M. Bucolo, F. Conti, L. Fortuna, G. Tomarchio, 'Nonlinear Clustering Strategy to Characterize Financial time Series', *11th International IEEE Workshop on Nonlinear Dynamics of Electronic Systems (NDES 2003)*, Scuol, May 2003.

M. Bucolo, F.Caizzone, L. Fortuna, G. Tomarchio, 'Application of Advanced Techniques for Financial Data Analysis to support Medium and Long Term Planning Activity', Activity Report, 2002.

M. Berthold, D. J. Hand, 'Intelligent Data Analysis - An Introduction', Springer, 1999.

Michael Dittenbach, Dieter Merkl, Andreas Rauber, 'The Growing Hierarchical Self-Organizing Map', '*Proc. of the International Joint Conference on Neural Networks'*,pp. 15 - 19, Lug. 2000.

Michael Dittenbach, Dieter Merkl, Andreas Rauber, 'Organizing and Exploring High-Dimensional Data with the Growing Hierarchical Self-Organizing Map', '*Proc. of the 1st International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2002)'*, pp. 626-630, Nov. 2002.