Improvement of the knowledge management based on the use of data-mining techniques, for the textile and garment industry.

Denimal Jean Jacques
Laboratoire de Probabilités et Statistique
Université des Sciences et Technologies de Lille
59655 Villeneuve d'Ascq Cedex. France
E-mail: jean-jacques.denimal@univ-lille1.fr

Boussu François
Laboratoire GEMTEX (Génie et Matériaux TEXtiles)
Ecole Nationale Supérieure des Arts et Textiles
9 rue de l'ermitage, B.P. 30329
59056 Roubaix Cedex 01. France
E-mail: francois.boussu@ensait.fr

Abstract: We consider in this communication the particular case of the textile/garment channel. Specific knowledge management issues occurring in this field are first presented and discussed. The introduction of a data mining process for the design of a new garment collection is then proposed in order to reduce the uncertainty of the choices made by the different partners acting in a textile firm. More details are then given on a data-mining tool aiming at describing and visualizing the sale behaviors of a garment collection. An application to a reduced sized database is finally proposed by way of illustration.

Key words: Data-mining tools, knowledge management, PLS regression; Interpretative aids, Textile sale data.

1 Knowledge management issues in the textile channel

The creating process of a new garment collection implies a mastered synergy between all the partners of a textile firm. Within the French context, the current organization of the firms we have met is mainly composed of five profiles of know-how: the director manager (making decisions), the marketing expert, the fashion designer, the logistic/production expert, and the financial expert (Figure 1). These last four partners build their own knowledge in their respective field with an important part of subjectivity. As an example, the role of the fashion designer is to feel the new trends of the fashion and his knowledge is mostly based on intuition. His knowledge acquisition directly comes from humans as it is generally done in the knowledge management community. The final decision comes from the director manager who tries to take into account the different advice of his interlocutors. The major difficulty comes from the relevance of the advice, which is partially based on intuitions or feelings. In view of these uncertainties, the final decision may be long and could end to financial risks.

In order to reduce them, a good knowledge of the sale behaviors concerning the last garment collections may be helpful. For each of them, this knowledge may include the determination of the factors influencing the sales and the identification of the item characteristics (color, fabric, gender, type…) playing a role in the sale explanation with the weeks where this role is noticed. Therefore, the data mining techniques may bring a more scaleable approach by extracting a precise and synthetic knowledge of the sale behaviors of the latest garment collections. Moreover, a modeling of these behaviors may be obtained and can be useful for the different partners of the textile firm in order to test, validate their intuitions and enrich their own knowledge in their respective fields.

This data mining process needs complete databases. For each garment collection, it is necessary to store not only the sold quantities of its items but also all the characteristics describing them. The coding used for their transcription must be studied so that the comparison between two different garment collections should be possible, which is unfortunately not always observed in practice.

The data mining techniques we proposed is composed of two sets of tools:

- **A description and visualization tool** extracting the knowledge from the database for a good understanding of the sale behaviors of the last collection. In view of the complexity of the information stored in these databases, the previous tool has to structure and to synthesize the extracted knowledge in order to make it quickly accessible to the different partners acting in the new collection design.

-**A modeling tool** based on the latest collection information carrying out estimates or simulations for the future garment collection. We describe below some of its abilities:

- **Simulation ability:** In view of the characteristics of a new collection, what could be their future sale behaviors? Are the latter able to be estimated from the knowledge given by the previous collection information?
- **Dynamic ability**: the knowledge of the sold quantities of the current garment collection before a given date can be introduced in the model in order to estimate again (and therefore to adjust) the sale profiles for the incoming weeks. This ability plays an important role for restocking decisions.
- **Incremental learning ability:** At the end of each new collection, the method integrates the corresponding data into the model and updates it.

To sum up, the data mining process will allow each partner of the textile channel to upgrade the knowledge of his own field (**Figure 2**). Therefore, the director manager could achieve more quickly a compromise solution from the different propositions brought to him.

In the following parts of this communication, we propose to present briefly the description tool and to comment the results given by its application on a reduced sized database.

This data-mining method and its associated software are the results of a team work since three years. The first experiment (Boussu, 1998) was based on the merging of the different sales behaviors into different optimal clusters. Then, to extract the maximum of predicting information, the final step consisted in fitting the most suitable forecasting model for each of these clusters. Unfortunately, any kind of explanation of the different sales behaviors was possible. The first introduction of descriptive parameters of items into the explanation model was done through the discriminant analysis (Denimal et al, 1999). The results was not so accurate and, to improve it, the PLS regression was introduced into our method (Boussu and Denimal, 2000). The final approach, presented in this paper, tends to combine the data-mining method with a powerful data-visualization tool.

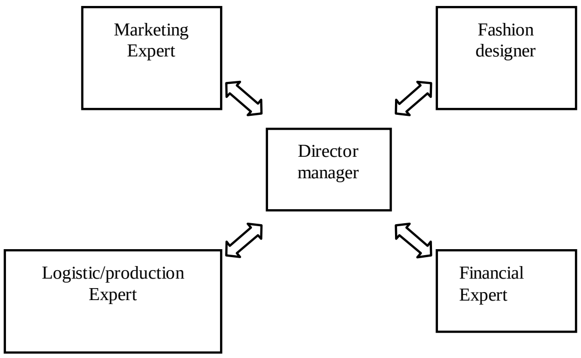


Figure 1. The main partners of the textile/garment firm

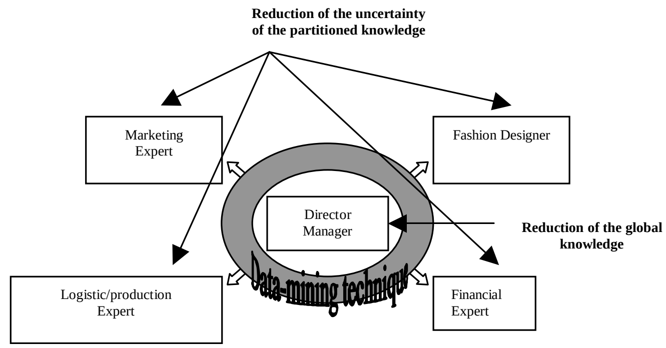


Figure 2: The impacts of the data mining process

2 Brief presentation of the description and visualization tool

2.1 Global presentation

This tool presents different steps based on particular statistical methods. Two data tables Y and X are first considered: the former is composed of the sale profiles of the items calculated on each week of the sale profile and the latter is a binary table crossing the items and their characteristics.

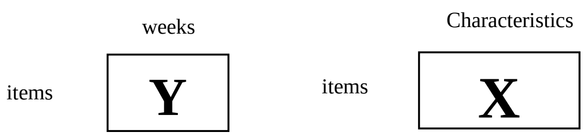


Figure 3. The data tables

First step: The characteristics (colors, fabric, type, gender...) may be very numerous, defining a large sized table X. We first submit X to the Correspondence Analysis (Benzecri et al. 1982) in order to concentrate the information of X in the reduced sized table with little loss of information.

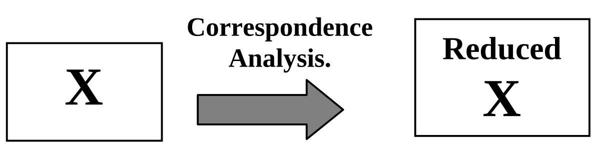


Figure 4. Creation of a reduced table X.

Second step: A Partial Least Square Regression (PLS Regression, Wold 1966-75; Esbensen et al; 194; Martens and Naes 1989; Hoskuldsson 1996) is then used for the determination of the relations between Y and Reduced X

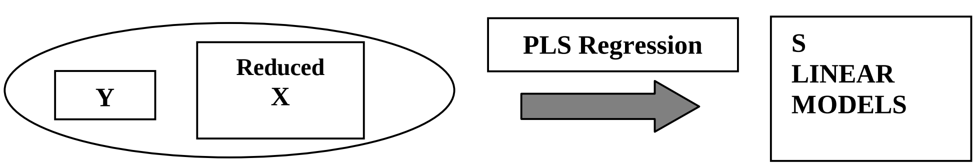


Figure 5. The use of the PLS regression.

If we denote S the number of weeks of the considered sale season, we deduce S linear models explaining the sale profiles of the items from their characteristics.
Third Step: A particular hierarchical classification (Denimal, 2001) of these S linear models is then carried out, which gives us a hierarchical classification of the sale season weeks. In this approach, the obtained classes create periods of time where the sold items have globally the same characteristics. Each retained node of this hierarchy represents a break between two periods where some items present an important modification of their sales. The main interest of this classification is to built for each node a graphic representation (which is in fact a factorial plane) showing these items and their common characteristics

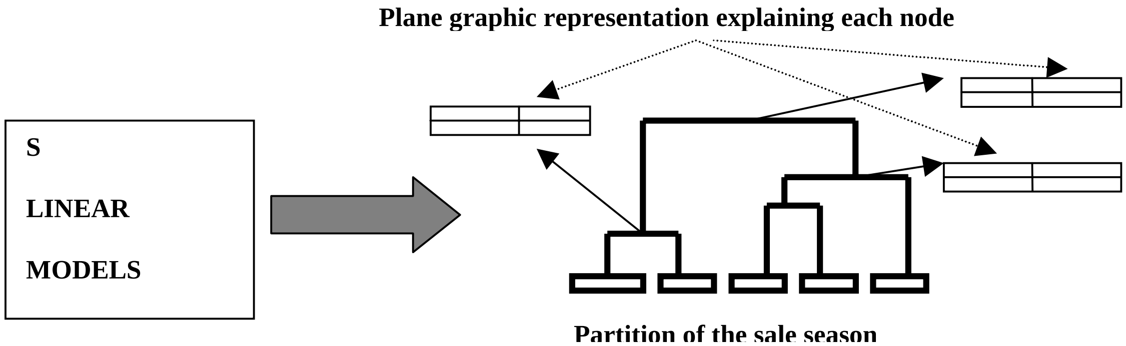


Figure 6. Hierarchical classification of sale weeks.

2.2 Mathematical formulation

Let us denote S the number of columns of Y (Number of weeks of the sale season) and X_0 the table composed of the extracted PLS factors. The proposed ascending hierarchical classification is built on the set $E=\{P_{X_0}(y^j)/1\leq j\leq S\}$ composed of the orthogonal projections of y^j (column vectors of Y) on the subspace spanned by the PLS factors (column vectors of X_0).
We will denote: $z^j = P_{X_0}(y^j), \quad 1\leq j\leq S.$
A hierarchical clustering can be defined as a succession of partitions $Q_0, Q_1, \ldots, Q_{S-1}$ where Q_0 is the partition of the singletons and Q_{S-1} is the one reduced to the single cluster E . At each stage, a partition Q_k is obtained from the previous one Q_{k-1} by aggregating two clusters of Q_{k-1}.
The research of the two clusters of Q_{k-1} which will be aggregated and will constitute the next node n_k is obtained in two successive stages:
First step: for each pair $(z^q, z^{q'}), q\in Q_{k-1}, q'\in Q_{k-1}$, the table crossing the set I of objects and the two variables z^q and $z^{q'}$ is submitted to the Principal Component Analysis (PCA).
Second step: For each of these analyses, the second eigenvalue $\lambda_2(q,q')$ equal to the projected inertia on the second factorial axis can be interpreted as a measure of dissimilarity between z^q and $z^{q'}$. As a consequence, the second step consists on the research for the pair $(z^q, z^{q'}), q\in Q_{k-1}$, $q'\in Q_{k-1}$ which minimizes the second eigenvalue $\lambda_2(q,q')$. By definition, the aggregation index $\nu(n_k)$ of the found node $n_k=(q,q')$ is: $\nu(n_k)=\lambda_2(q,q')$ and the new cluster $q\cup q'$ is represented by the variable $z^{q\cup q'}=\alpha_k.z^q+\beta_k.z^{q'}$ where (α_k,β_k) is the normed eigenvector associated to the biggest eigenvalue $\lambda_1(q,q')$.
Once the hierarchical tree is obtained, a cutting procedure is proposed in order to determine the set of upper nodes to be kept. Let us denote $X(q\cup q')=\alpha_k.X(q)+\beta_k.X(q')$ for each created node $n_k=(q,q')$ with the condition $X(q)=X_0$ for any singleton q.

For each level of the hierarchy, a partition of $E = \{P_{X_0}(y^j)/1 \le j \le S\}$ composed of s clusters $\{q_1, q_2, \ldots q_s\}$ is obtained and a linear model associated with this level is defined (Model 2). The latter can be compared with the initial model (Model 1) coupled with the singletons partition. When the model 1 is replaced by the model 2, it can be proved that the loss of explained variance is equal to the sum of the aggregation indexes of the nodes created up to the considered level. A statistical test based on the classical Fisher law can then be carried out in order to reveal a significant difference between the two models. In the treated example, the application of this test shows that the 10 highest nodes of the hierarchy can be considered as significant.

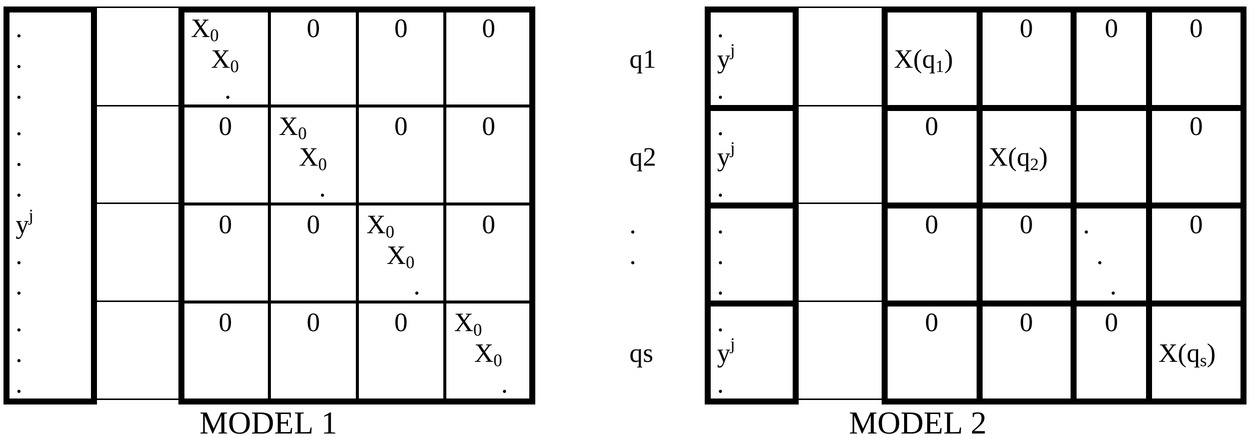


Figure 7: The two models 1 and 2.

Finally, as each node of this hierarchy is obtained after a Principal Component Analysis, graphic representations can be obtained and allow us to display the similarities and the dissimilarities between the two sub-groups merging at the considered node.

3 Application to textile data mining

3.1 Illustrating the descriptive tool.

As an example, the method is applied to a little subset of 58 sport items (Cycling, swimming, gymnastics) characterized by 6 main qualitative variables. These data are provided by a French garment maker specialized in sport items. Another table playing the role of Y gathers all the sold quantities of these items during one year (51 weeks). The items and the weeks are respectively the rows and the columns of this table.
A description of the different values of the qualitative characteristics (defining the table X), and their numbers, is given into the Table 1.

Qualitative characteristics	Categories	Number of categories
Gender	Swim or Gym	2
Type	Child, woman or man	3
Description	Swimsuit, body, etc.	11
Fabric number	1345, 1226, etc.	16
Model number	2014, 2045, etc.	23
Color	Black, White, Pink, Yellow, etc.	28
	Total	**83**

Table 1. Description of Textile items characteristics.

The textile sales present some typical particularities concerning first, the sold item and second, its environment. More precisely, the former is characterized by a short lifetime and a large number of categories and the latter by the influence of many factors (fashion trends, calendar or seasonality effects, economic indexes).

3.2 Applying the three steps of the descriptive tool.

As described in §2.2, a hierarchical classification is built on the set of the season weeks. The application of the Fisher test leads us to keep the highest 10 nodes.
The restricted hierarchy limited to them leads to 11 clusters partitioning the set of all the weeks of a year. In the representation of the dendogramm and the resulted partition (**Figure 8**), each week is labeled by its number varying between 1 and 51. For instance, the number 1 corresponds to the week, which begins the 3th of January 1994 (03/01/94) and ends the 9th of January 1994 (09/01/94). In the same way, each node is labeled by a number varying from 52 for the lowest node to 101 for the highest one.

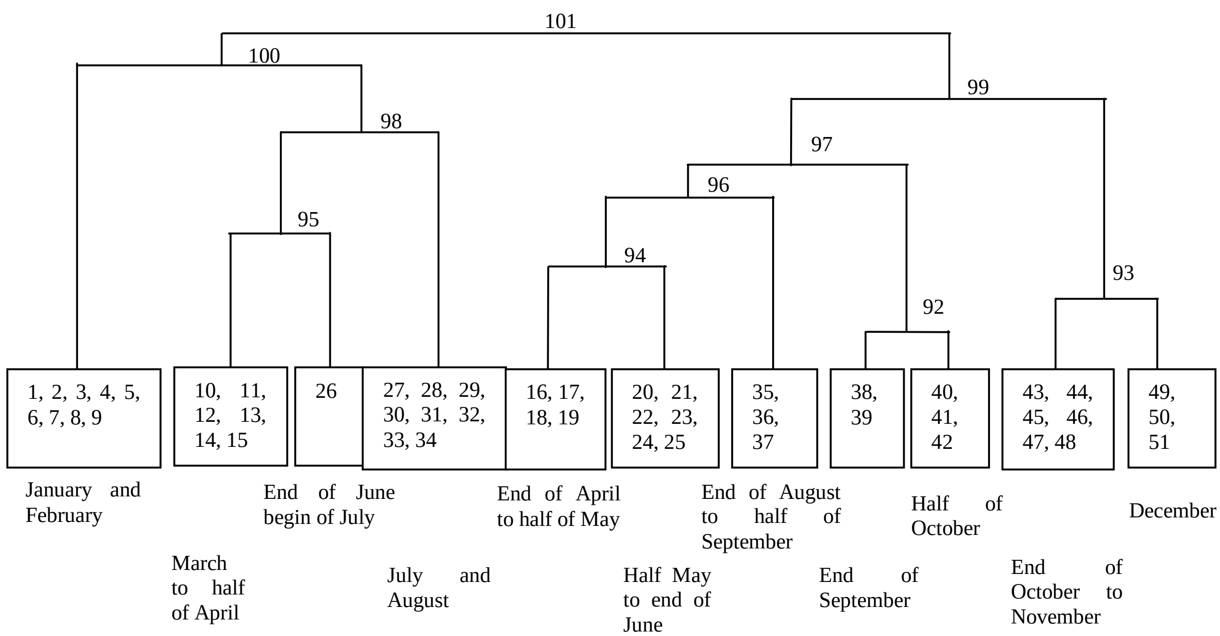


Figure 8. Representation of the restricted hierarchy

Besides, we know that this classification method can be considered as a succession of Principal Component Analyses. Since each of them is associated with a node of the hierarchy, graphic representations especially adapted for its interpretation can be obtained. In the following sections, we have chosen to give the interpretation of two nodes: the highest node 101 and the node 94. These two examples will be sufficient to show the usefulness of our classification method.

3.3 Explanation of the node 101

In Figure 9, the opposition between spring, summer, autumn and winter reveals the seasonality effect. Moreover, the breaks defined by the New year's day and Easter holiday can also be noticed. In the same way, two main transition periods such as the end of spring to the begin of summer (week 26) and the end of summer to the begin of Autumn (week 35) can be detected. In the case of the sport items, these transition periods clearly show a major psychological behavior of the customer. This global information represents a first step of the knowledge improvement of the sales behaviors.

In the following figures, the distribution of some item characteristics (color and item description) in the node 101 planes gives us a general information of the sales explanation. For instance, the summer period is described by the colors "9040 black" and "deep blue", by the swimsuits "Girl one body swimsuit", "Boy swimsuit", "Woman large size swimsuit", by the type "swim" and by the gender "woman".

It can be noticed that the united colors: 9040 black and deep blue are representative of the summer period. By the same, the united colors: bianco ao and ciliega salmon represent the winter period for the textile sport items of the firm. On the contrary, the mixed colors: v1 black/blue and v2 black/red are representative of the autumn period. It seems that the united colors are expected in the summer and winter period and the mixed colors in the autumn period. This information brings also to the marketing and designer directors a specific knowledge of the sales behaviors as regards the colors choice.

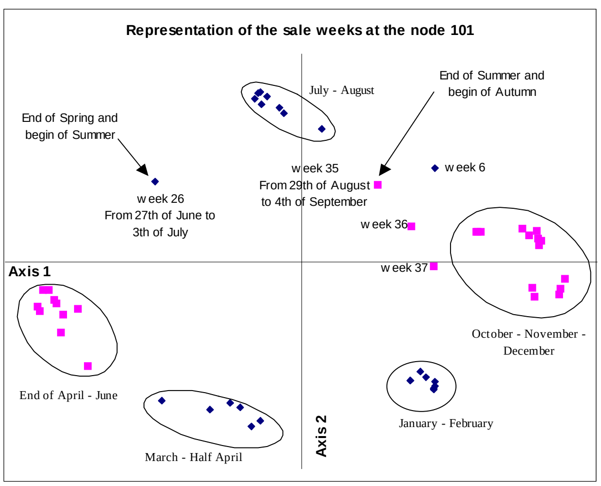


Figure 9. Representation of the weeks in the node 101 plane.

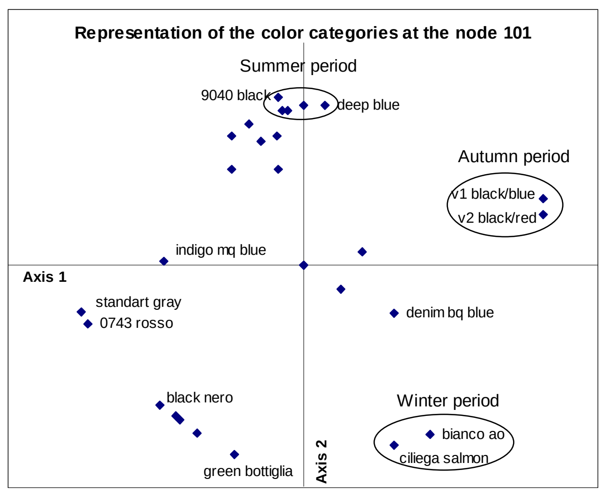


Figure 10. Representation of the color categories in the node 101 plane.

It can be noticed in the following figure the main opposition between the item man swimsuit and the group composed of girl, woman and boy swimsuits. Sale behaviors of women are recovered since they usually buy their swimsuit but also those of their children. On the contrary, men buy alone their swimsuit. These simple sales behaviors displayed in this figure and revealed by the data-mining method bring another knowledge. The designer may be helped in the building of the future collection. In the same way, the logistic and production experts may schedule with more accuracy the external material flows and adjust their production and storage capacities. Finally, the financial expert may optimize his cash needs with the best rate investment.

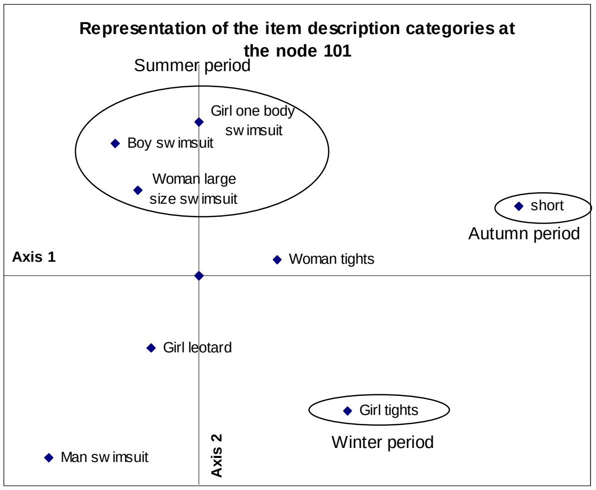


Figure 11. Representation of the item description categories in the node 101 plane.

3.4 Explanation of the node 94.

In contrast with the node 101, the information contained in the node 94 representations is totally new and couldn't be found in the different previous planes.

As a matter of fact, the node 94 opposes the periods weeks 16 to 19 (from the 18th of April to the 15th of May) and weeks 20 to 25 (from the 16th of May to the 26th of June). This opposition clearly shows the change to appear before and after the Easter holiday. In the following figures, the distribution of some item characteristics (color, model number and fabric number) in the node 94 planes gives us a general information of the sales explanation.

It can be noticed that the specific colors: 0743 rosso and standard gray are representative of the first two weeks of June contrary to the color green bottiglia more present in the begin of May. Then, two groups of colors clearly identified by the data-mining method may influence the sales behaviors and brings another information to the global knowledge of the firm.

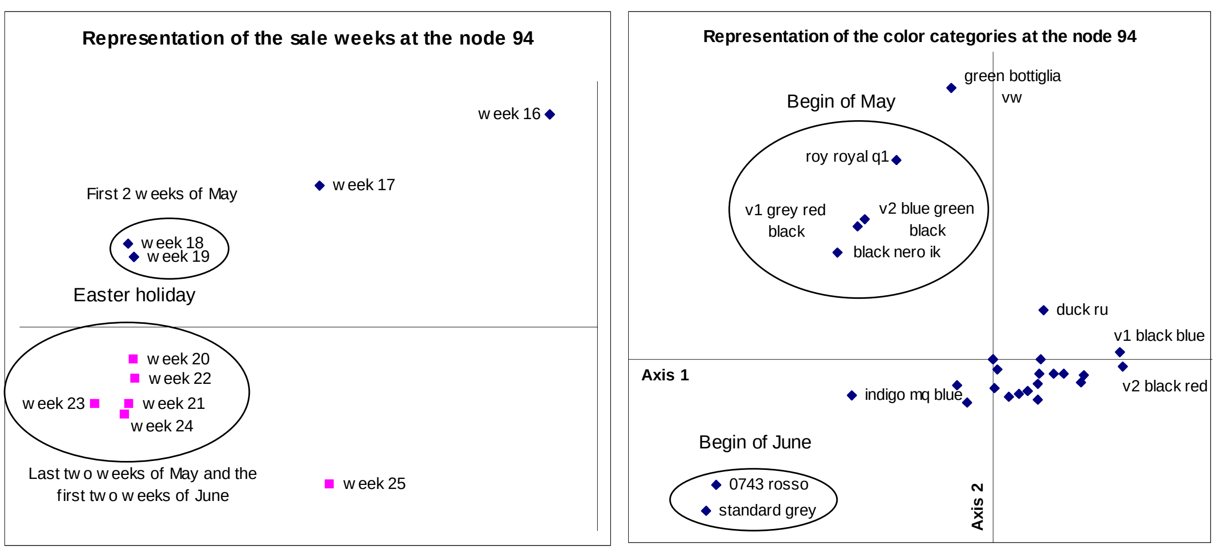
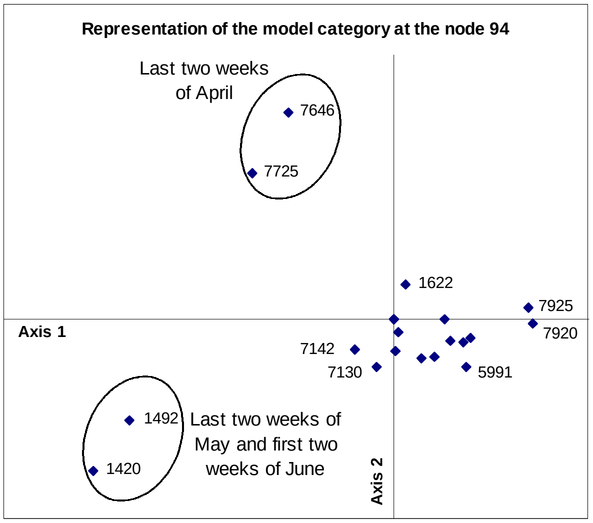


Figure 12. Representation of the weeks and the color categories in the node 94 plane.



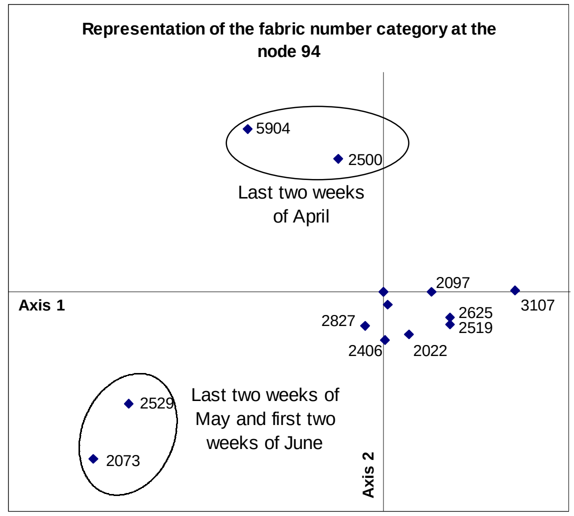


Figure 13. Representation of the model and fabric number categories in the node 94

The major information contained in the previous figures, revealed by the data-mining technique, may help the fashion designer to better understand the different sales behaviors of the model group (7646 and 7725) and the model group (1492 and 1420). Thus, it can be possible to integrate the economic aspect of a different sales behavior in the future collection for the models choice. By the same, the different group of the fabric number brings another information in the sales behaviors knowledge. The logistic and production expert may optimize its fabric needs and could better plan its orders to its weaving suppliers.

The transition defined by the Easter holidays implies for some items significant modifications of their sales profiles. For instance, the items "bottiglia green-man swimsuit" and "standard gray-woman leotard" (Figure 15) are characterized by significant changes of their sales profiles after and before Easter holidays. These two items showing opposite behaviors are represented by two opposite points along the second axis of the factorial plane associated with the node 94 (Figure 14).

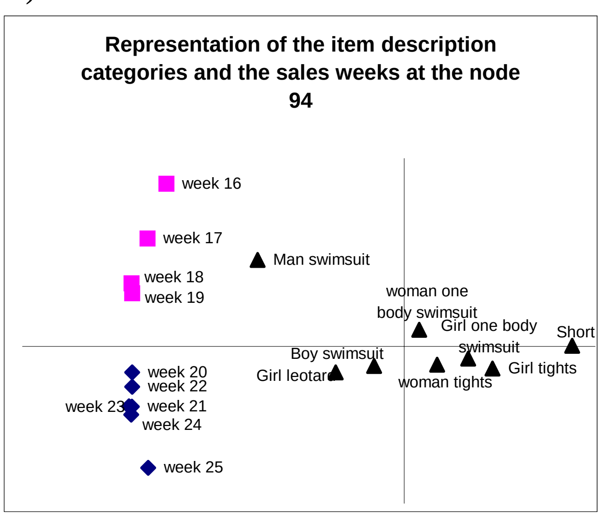


Figure 14. Representation of the item description categories and sale weeks at the node 94.

On the contrary, the first axis of this plane displays the items whose sales profiles are not modified through Easter Holidays. As a matter of fact, short the item man short shows such a sale behavior due to its very low sales in the two considered periods.(Figure 16)

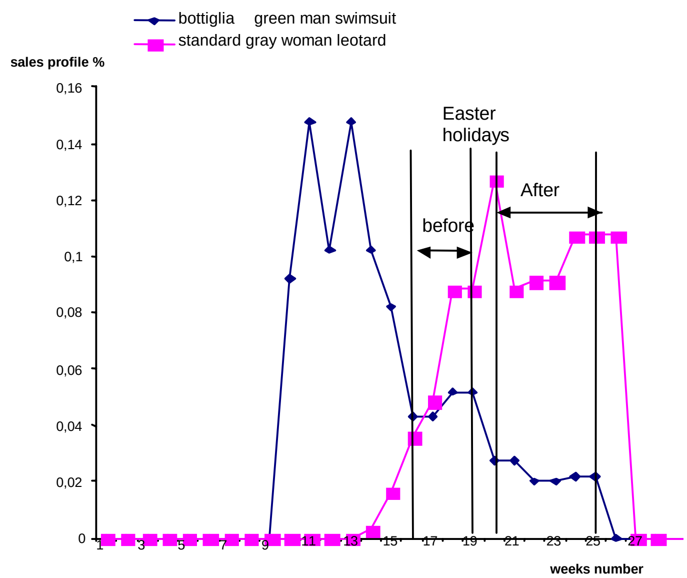


Figure 15. Profiles of the items bottiglia green man swimsuit and standard gray woman leotard.

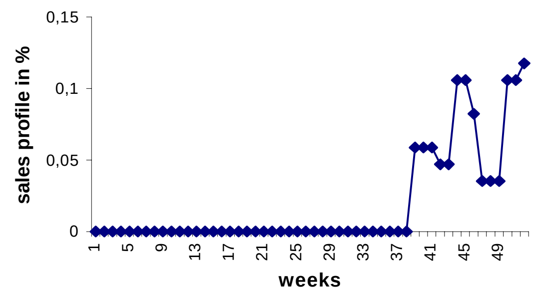


Figure 16. Profile of the item man short

4 Conclusion

The previous sale season analysis given by this approach brought a better understanding of the sales behaviors. This additional knowledge allowed the production manager to optimize the items quantities to be stocked during the next season. As a consequence, a 50% stocks reduction was reached at the end of this next season. However, the application of this analysis for each new collection remains necessary since the corresponding items present new characteristics and therefore modified sales behaviors. These modifications have to be taken into account in order to remain close to the market variations.

This data mining process could be applied to other kinds of items and more generally could be adapted to the needs of the distribution channel. A better understanding of sale behaviors observed during the latest years and good estimates of those of the current year are necessary for producers and distributors as well since they have both to minimize their buffer stock sizes as much as possible and to avoid at the same time the out of stock. As a consequence, the use of adapted data mining techniques may help the industrial managers in their choices and decisions.

5 References

BENZECRI, J.P. (1982), *L'analyse des données,* Vol. 2, Dunod, Paris.

BOUSSU F. (1998), *Simulation de la filière Textile/Habillement/Distribution : Réduction de la complexité en vue d'une meilleure prévision de ventes*, Phd Report, Université des Sciences et Techniques de Lille1.

BOUSSU F., DENIMAL J.J. (2000), *The PLS regression: a logistic tool to the textile sales estimation*, Workshop RIRL 2000, Acts on CD-ROM, Trois-Rivières, Canada.

BOUSSU F., DENIMAL J.J. (2000), Identification of Textile Sales Behaviors for Better Estimates. *Journal of Textile Institute*, 91 Part. 2, n°1.

BOUSSU F., DENIMAL J.J. (2000), *The PLS Regression : A Response to the Textile sales estimation*, WAC / ISIAC 2000, Fourth biannual World Automation Congress, Acts on CD-ROM, Hawaï, USA.

DENIMAL, J.J., DIALLO O.W., BOUSSU F. (1999), *Estimation of textile sales behaviour*, IMACS/IEEE CSCC'99 International Multi conference, Athens (Greece), paper number 454 (CD ROM proceeding acts).

DENIMAL, J.J. (2001), *Hierarchical Factorial Analysis*. 10th Symposium of Applied Stochastic Models and Data Analysis, Volume 1, Compiègne, France.

ESBENSEN K., SCHONKOPF S., MIDTGAARD T. (1994), *Multivariate Analysis in Practice*, CAMO, Olav Tryggvasons gt. 24, N-7011 Trondheim, Norway.

HOSKULDSSON A (1996), *Prediction Methods in Science and Technology*, Vol. 1. Basic Theory, Thor Publishing, Arnegaards Alle 7,Holte, Denmark.

MARTENS H., NAES T. (1989), *Multivariate Calibration*, John Wiley and Sons, New York.

WOLD H. (1966), Estimation of Principal Components and Related Models by Iterative Least Squares, in *Multivariate Analysis*, Krishnnaiah P.R.(ed), Academic Press, New York, pp 391-420.

WOLD H. (1975), *Modeling in Complex Situations with Soft Information.* Third World Congress of Econometric Society, August 21-26, Toronto, Canada.