Available online at www.sciencedirect.com
ScienceDirect
Applied Soft Computing 7 (2007) 1177–1187
Applied Soft Computing
www.elsevier.com/locate/asoc

A neural clustering and classification system for sales forecasting of new apparel items

Sébastien Thomassey *, Michel Happiette

GEMTEX-ENSAIT, 9 rue de l'ermitage, 59100 Roubaix, France

Available online 28 February 2006

Abstract

The Textile-Apparel-Distribution network actors require a very accurate production and sourcing management to minimize their costs and satisfy their customers. For a such strategy, distributors rely on sales forecasting system to respond to the versatile textile market. However, the specific constraints of the textile sales (numerous and new items, short lifetime) complicate the forecasting procedure and distributors prefer to use intuitive estimation methods of the sales rather than the existing forecasting models. We propose a decision aid system, based on neural networks, which automatically performs item sales forecasting. Performances of our model are evaluated using real data from an important French textile distributor.

Keywords: Sales forecasting; Clustering; Classification; Neural networks

1. Introduction

These last decades, methods based on Supply Chain Management tools (Manufacturing Requirement Planning, Distribution Requirement Planning, Enterprise Resource Planning), have enabled an improvement of the sourcing, production and distribution of the textile items. However, due to the competitive environment, the globalization, the irreducible manufacturing lead times and the uncertainty of the customer's demand, the sales forecast is a fundamental success factor of the supply chain optimization of apparel companies [47]. The forecasting system must deal with the constraints of the textile market:

- Large number of items (about 15,000 per year).
- Items with short lifetimes (6–12 weeks).
- Substitution of most of the items for each collection (95%).
- Long lead time of textile items requires considerations of producing and planning of sourcing at a mid-term horizon (the forecasting horizon is one season or 1 year).
- Influence of many explanatory variables. These factors can be: weather data, holiday, marketing action, promotions, fashion, economic environment.

Several forecasting models, such as regression models [51,24,7], exponential smoothing and Box & Jenkins models [52,21,58], neural networks [74,77,50] or fuzzy systems [15,46,11], have been developed and provide satisfactory results in different domains [36]. However, their performances strongly depend of the field of application, the forecasting goal, the user experience, and the forecast horizon [4,13] and these methods are not easily usable in the specific textile environment. Preceding works are enabled us to carry out a global forecasting system for textile sales forecasting. This system, based on soft computing techniques, is composed of several models which automatically performs mid- [64,65] and short-terms sales forecasting [66]. However, due to the substitution of most items for each collection, the aggregation of sales by items families or by clustering procedures to obtain complete historical data of several years is required.

This paper deals with the mid-term sales forecasting for items for which we have no historical sales data. Thus, we propose to improve the system by combining clustering and classification tools. The uncertainty and the complex relationship between the sales and the descriptive criteria of items lead us to rely on neural techniques.

Section 2 describes the proposed forecasting system. Section 3 reports and analyzes the empirical results obtained with real data supplied by an important French ready-to-wear distributor.

* Corresponding author. Tel.: +33 3 20 25 64 64; fax: +33 3 20 27 25 97.
E-mail address: sebastien.thomassey@ensait.fr (S. Thomassey).

1568-4946/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2006.01.005

2. The proposed forecasting system

2.1. Time series clustering, classification and prediction

Due to the very short life span and the frequent substitution of apparel items, new items do not have historical sales data. However, the retailer is able to provide various descriptive criteria of historical and future items. Thus, for historical items, the available data, in database of the retailer, are the real sales and the following descriptive criteria: price, period of the beginning of sales and life span of items. For future items, only descriptive criteria are known.

The proposed system, called Neural Clustering and Classification (NCC) system, is composed of the three following steps (Fig. 1):

(1) A clustering procedure (Self-Organizing Map Neural Network) of the sales time series of historical items in a same family enables us to obtain the prototypes, which are characteristics of the sales behavior of items.
(2) A classification procedure (Probabilistic Neural Network) then performs the links between prototypes of time series and descriptive criteria of historical items.
(3) Finally, each new item is assigned by the classification procedure to one prototype of time series. This prototype is predicted only from the descriptive criteria of the item.

The clustering and the classification learning processes (steps 1 and 2) require the sales profiles (normalized time series) and the descriptive criteria of historical items. In order to forecast the sales profiles of future items, the predictive classification process (step 3) uses the descriptive criteria of future items.

2.1.1. Clustering procedure

The purpose is to achieve clusters of items whose the life curve are similar. The life curve of the cluster center, called prototype, is the mean life curve of all items included in the cluster. This prototype represents the behavior of the items. This step which must enable us to extract all the sales behavior of the items, is crucial for effectiveness of the downstream classification procedure.

A wide range of techniques have been developed for clustering data, which can be classified into the following categories [18]: partitioning clustering, hierarchical clustering, density-based clustering, grid-based clustering and model-based clustering. The first two categories are obviously the most frequently used methods.

The most common partitioning clustering algorithm is *k*-means algorithm, where each cluster is represented by the mean value of the objects in the cluster. One advantage of the partitioning clustering is that the procedure is dynamic, i.e., data points can move from one cluster to another. The other advantage is that some a priori knowledge, such as cluster shapes, can be incorporated in the clustering. The main drawbacks of the partitioning clustering are the following [73]:

- the difficulty to find clusters of arbitrary shapes;
- the number of clusters have to be prefixed and is sometimes hard to determine;
- the final clusters are dependant of the initialization and problems of local optima can arise.

In hierarchical clustering, the number of clusters need not be specified a priori, and there is no problem concerning initialization and local minima [68]. However, the shortcomings of the hierarchical clustering are the following [73]:

- it is not suitable for large data sets due to the high-computational complexity;
- it does not incorporate any a priori knowledge such as cluster shapes;
- it performs a static clustering, i.e., data points in a cluster at the early stage cannot move to another cluster at the latter stage.

Additionally, these two previous methods are sensitive to noise and outliers [6,70].

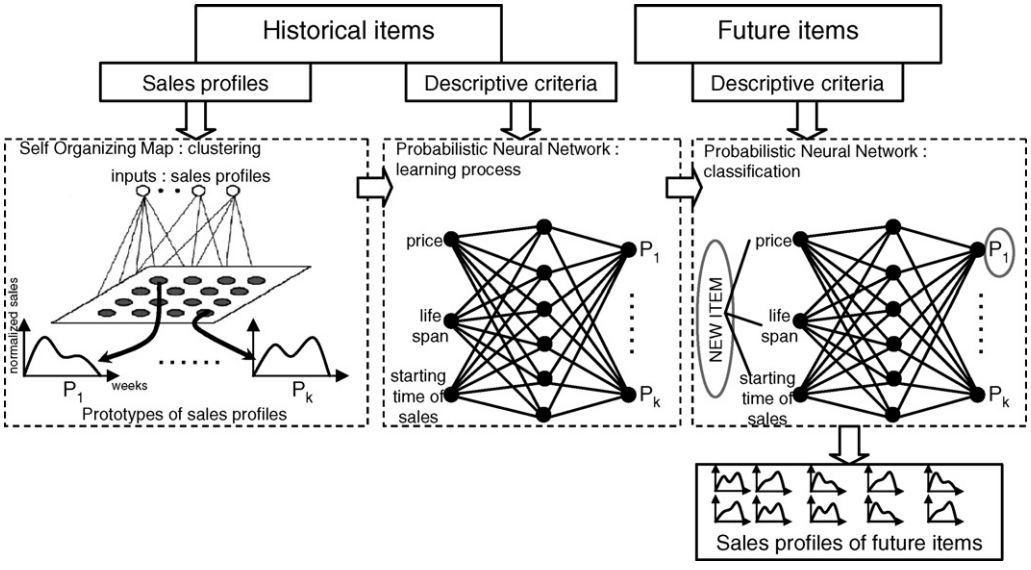

Fig. 1. Principle of the NCC forecasting model.

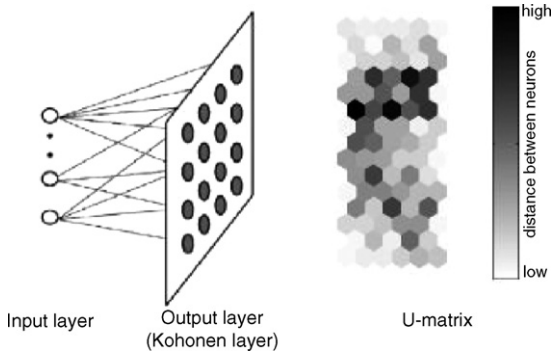

Fig. 2. Example of Self-Organizing Map network (SOM).

In order to deal with these shortcomings, more robust methods, based on neural techniques, such as Self-Organizing Map (SOM) [32] or genetic algorithms [35,44], have been developed.

SOM networks have been especially successfully applied to data clustering in many domains [68], from biology [75,33] to the travelling salesman problem [25] through ecology [16] and image classification [10].

A SOM network consists of two layers (Fig. 2): an input layer and an output layer also called Kohonen layer. The Kohonen layer is usually designed as a two-dimensional arrangement of neurons that maps N-dimensional input to two dimensions, preserving topological order [45,27]. The SOM input layer of neurons is fully connected to the Kohonen layer. The Kohonen layer computes the Euclidean distance between the weight vector for each of the Kohonen neurons and the input pattern. The Kohonen neuron that is closest (i.e., minimum distance) is the winner and its activation value is one. Unlike most other neural network approaches, the SOM network performs unsupervised training [31].

The 2D arrangement of the output layer can be easily visualized and thus gives people useful information about the input data [27]. The usual way to display the cluster structure of the data is to use a distance matrix, such as U-matrix [69]. U-matrix method usually represents the output layer as hexagonal grid using gray level according to the distance between each neighboring neurons.

The success of the SOM algorithm lies in its simplicity that makes it easy to understand, to implement and fast to compute [73,33]. The number of map units, which determines the accuracy and generalization capability, is a crucial parameter of the SOM. The bigger the map size, the lower the quantization error but the higher the topographic error. This is due to the neural network folds to reduce the quantization error. Moreover, the bigger the map size, the higher the computational cost. Therefore, there is trade-off between the increase of the topographic error and the reduction of the quantization error. A reasonable optimum solution of the compromise among the quantization error and the topographic error to determine the number of neurons of the output layer is given by the following formula [16]: $M = 5\sqrt{N}$, where M is the number of map units and N is the number of samples of the training data. The next step consists in calculating the ratio between the number of rows n_1 and the number of columns n_2. This ratio is the square root of the ratio between the two biggest eigenvalues (e_1 and e_2) of the training data such as:

$$\frac{n_1}{n_2} = \sqrt{\frac{e_1}{e_2}}$$

In order to have accurate representation of the data on the output layer, a common problem of SOM network is that [31], a larger number of nodes than the number of target groups is generally needed. Thus, an additional clustering step is required to further analyze the output map and to obtain the expected number of clusters.

The idea in this two-level procedure [70] is that the first level is composed of the SOM network which forms a 2D map of the data and the second level is composed of a second procedure which clusters the SOM outputs (Fig. 3). In the first level, a large set of prototypes which is much larger than the expected number of clusters, is formed using the SOM. These prototypes can be interpreted as ''protoclusters'', which are in the second level aggregated to form the final clusters. An interesting feature of the SOM is the notion of neighboring between the protoclusters, i.e., each protocluster is similar to the other protoclusters which are close to it, on the map (Fig. 5).

The two main benefits of the two-level approach are: the reduction of the computational cost [71] and the noise reduction. Indeed, the prototypes are local averages of the data and therefore, less sensitive to random variations than the original data. The two-level approach in clustering can only work if the prototypes reflect the properties of the data. The clusters found using the prototypes should be similar to those that would have been obtained directly from the data.

For this two-level approach, the SOM can be combined with many other clustering methods [73]. For instance, the second level can be composed of:

- A second SOM layer, which takes into account the outputs of the first SOM layer. The number of the neurons on the second map is equal to the desired number of clusters [37].
- An agglomerative contiguity-constrained clustering method [48,31].
- An agglomerative hierarchical clustering [73].
- A partitioning methods such as the k-means algorithm.

For our applications, we combine a SOM network with the k-means methods. Indeed, for this two-level approach, partitive methods seem to be better than hierarchical ones in the sense that they do not depend on previously found clusters [70]. On

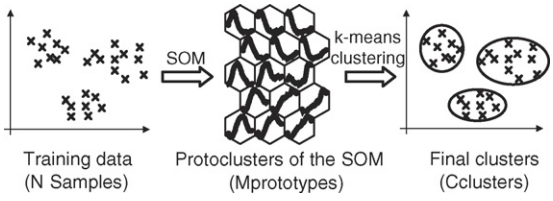

Fig. 3. The two-level clustering method.

the other hand, partitive methods make implicit assumptions on the form of clusters. For instance, *k*-means tries to find spherical clusters. The number *k* of clusters must be previously fixed and strongly influences the quality of the clustering procedure. But, in this system (Fig. 1), the number *k* also contributes to the accuracy of the classification procedure (Section 2.1.2). The optimization of the number of clusters consists in an original procedure based on a *k*-fold cross-validation which determines the best partition to perform the more accurate classification.

2.1.2. Classification procedure

In order to carry out mid-term sales forecasting for new items, it is necessary to link the sales prototypes, previously computed, with descriptive criteria. A classification model, built with historical data, enables the attribution of one prototype to each new items from their descriptive criteria. The assigned prototype is the sales forecasting of the considered item.

Many classification methods exist and they can be globally categorized into two classes: statistical methods and machine learning methods.

Statistical models, such as linear discriminant analysis [63] and logistic regression, have shown their adequacy for a great number of practical classification tasks, especially in economy [3]. They provide simple and easy-to-interpret classifiers. However, the problem with applying these methods to real life problem is that the researcher needs to impose structures (such as linearity for regression analysis) to different models [23]. These methods become also often unreliable when relationships in the input/output data set are complex and/or non-linear [1].

It is therefore recommended to employ more efficient classifiers based on machine learning techniques. Amount these methods, the most common tools are [78]: decision trees, neural networks, rule-based systems, kernel-based classifiers, baye-sian-based classifiers, ... These methods are flexible enough to perform satisfactorily in general setting. Unlike statistical methods, they also automatically extract knowledge from a data set and construct different model representations to explain the data set. Many studies investigate and compare the efficiency on these classifiers, and they generally conclude that performances are highly dependent of the context and the data [30,55,56,3]. It globally appears that:

- Decision trees such as ID3 based algorithm [8,17] and neural networks [1,5,42,62,76] achieve better results than linear discriminant analysis and logistic regression.
- Neural networks seem to be the most accurate whereas decision trees are less sensitive to reductions in sample size [39].
- Neural networks also provide best fit with numerical data, while ID3 algorithm performed best with non-numerical data [43,9].

It globally arises that neural networks [1] are considered as the most powerful classifier for their low classification error rates and robustness to noise [38]. Indeed, numerous studies have compared the classification performance of neural networks with traditional statistical techniques and provided evidence to indicate that neural networks outperform traditional techniques for some cases [57].

But, they are two shortcomings when applied to data mining problems [41,12]: neural networks require a long time to train with large databases, and lack of explanation facilities for their knowledge. Remarkably, neural network systems appear to yield acceptable results even though sometimes they are incapable of outperforming linear model [3,20,54].

However, in our context (complex relationship between sales and descriptive criteria, numerous items, numerical data, . . .), neural network should be suitable for the classification process.

Among neural network models, Learning Vector Quantification Networks [34], Counter-Propagation Networks [19] and Probabilistic Neural Networks (PNN) [59] have already been successfully used for classification in several purposes such as weather forecasting [67], medical decision [40], image recognition [49], finance [72], ...

In fact, PNN are often applied for their accuracy and generalization ability, adaptability, fast learning capability [60,26]. They are also useful for their simplicity and learning capacities, easy and instantaneous training [61]. In addition, it also has an on-line training and updating capability. A PNN, which is based on the implementation of Bayesian classication [14] and Parzen non-parametric density function estimation [53], consists of three feedforward layers (Fig. 4): input layer, pattern layer and summation layer. Its main drawbacks are the larger memory requirements due to the important network size when the training data set are numerous [26] (PNN incident traffic) and the slower execution speed for prediction of unknown patterns compared to conventional neural networks [2]. Because in our application, the accuracy and the generalization are more important than the execution time, we decide to implement PNN to classify items.

The input layer has as many neurons as there are parameters (descriptive criteria) needed or available to describe the textile items. This layer supplies features of input vector to all the neurons in the pattern layer. This one is organized such that each training input vector is represented by a neuron. When a new input is presented, the pattern layer computes distances from this input vector to the training input vectors, and produces a vector whose elements indicate how close the input is to the training inputs. Finally, the summation layer, which has as many neurons as there are classes (or prototypes) to be recognized, sums the contributions of previous vector for each class to produce a vector of membership probabilities. The output vector is composed of a 1 corresponding to class with the largest probability and 0 elsewhere.

The PNN uses a supervised training set to develop distribution function within the pattern layer. This one, as mentioned earlier, the pattern layer represents a neural implementation of a version of a Bayes classifier, where the class dependent probability density functions could be approximated using a Parzen estimator. This approach provides an optimum pattern classifier in terms of minimizing the expected risk of wrongly classifying an object.

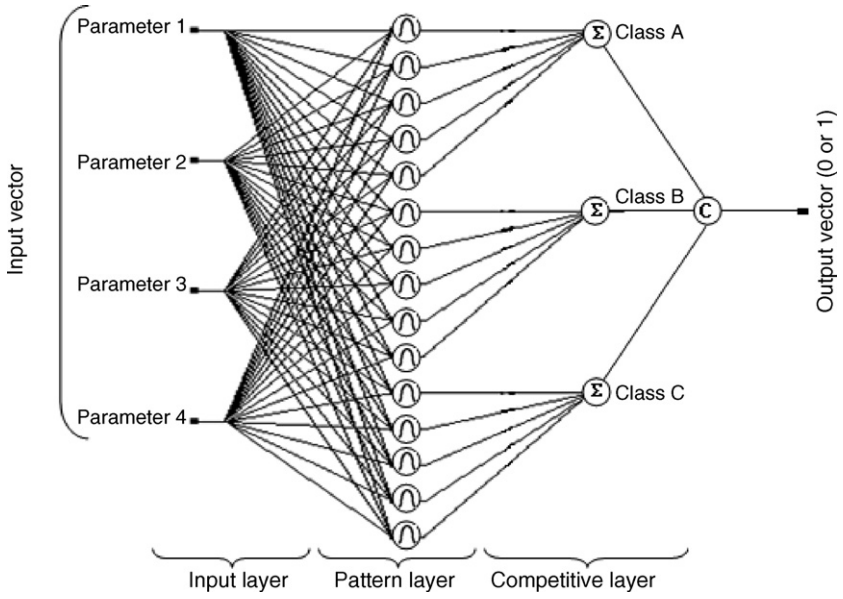

Fig. 4. Example of Probabilistic Neural Network (PNN).

3. Implementation of the NCC system on apparel sales data

3.1. Used data

Sales and descriptive data are extracted from our database relative to an important French retailer. Historical data, which are used for the learning process, are composed of 482 items (corresponding to years 1998 and 1999). Future data, which are used to evaluate the accuracy of our model, are composed of 285 items (corresponding to year 2000). In order to compare and to cluster sales profiles, life span of all items are normalized to 52 weeks.

The selected descriptive criteria of items must have a significant influence on the observed sales. However, the choice is generally imposed by the availability of these criteria in the retailer's database. In this work, the criteria are the price, the starting time of the sales and the life span of items. Additional criteria such as style or textile material are also of a great interest for forecasting purposes. However, these data are not available in our database.

3.2. Forecasting models used for comparison

It is difficult to compare our model with classical time series forecasting models since:

- the forecast horizon is too large;
- the items do not have specific sales time series (substitution of items at each collection).

Very often, retailer rely on sales forecasting based on aggregated data and mean profiles of items belonging to the same family. This mean sales profile (mean predictor) will be used later as a benchmark. In order to evaluate the accuracy of our NCC system, and especially the efficiency of the PNN, we have also tested three other classifiers:

- The most primitive algorithm, called ZeroR, simply outputs the most frequent class on the training data set. Inclusion of the ZeroR in the set of benchmark methods can help in identifying eventual overfitting problems.
- The OneR algorithm [22] generates simple rules based on the attribute that produces the smallest number of classification errors.
- The Naïve Bayesian [28] method is based on the prior probabilities of each class computed from the training data.

3.3. Clustering of historical items

The two-level clustering procedure, described in Section 2.1.1, is applied on the sales profiles of the 482 historical items. The main objective is to summarize the sales profiles with few prototypes of sales which will be the forecast of the future items.

3.3.1. Results of the SOM network

The SOM network enables a first clustering of the sales profiles. This result in ''protoclusters'' which will be clustered by the k-means method described in the next step. The main issue is the number of the ''protoclusters'', fixed by the map size of the SOM, which affects the trade off between the noise reduction and the representation of the data. The size of the used map size is determined by the formula given in Section 2.1.1. Thus, the number of map units is $M = 5\sqrt{482} \simeq 110$. In accordance to the ratio of the number of rows and columns n_1/n_2, this leads to $n_1 = 13$ and $n_2 = 8$, i.e., a map of 104 units.

The 104 protoclusters, obtained with a hexagonal topology (Fig. 5), characterize many sales behaviors. It is interesting to

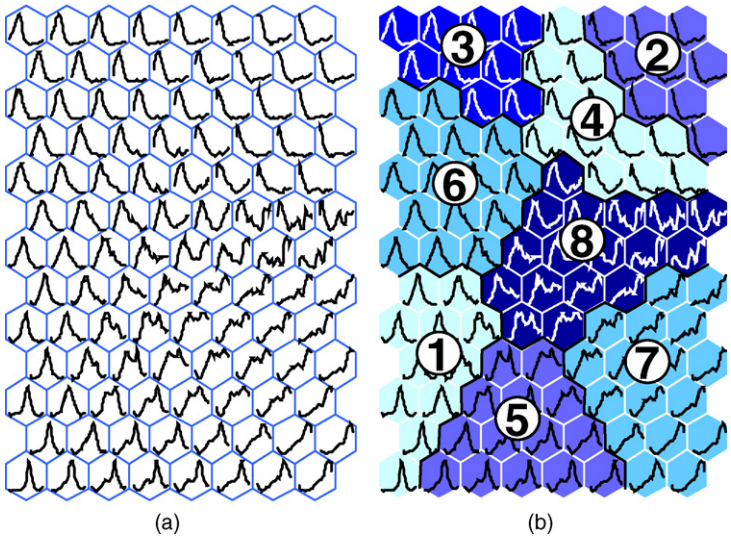

Fig. 5. The protoclusters of the SOM (a) and the clustering of the protoclusters (b).

observe the notion of neighboring between the protoclusters: the protoclusters which are close in the map are very similar whereas the protoclusters at each corner of the map are totally different. It is this notion of neighboring which makes the clustering of the protoclusters possible for the next step.

3.3.2. Results of the k-means clustering

The *k*-means clustering is performed on the 104 protoclusters. The number of final clusters n_c is a crucial parameter and strongly impacts the efficiency of the PNN for the classification procedure. This parameters is computed from historical data by a 10-fold cross-validation method. The selected n_c is the optimum number of clusters which enables the more accurate classification in term of absolute error. In this application, the optimal number of clusters, is $n_c = 8$ (Fig. 6). The eight clusters and their associated prototypes are presented in Fig. 7. Each prototype represents between 37 (for cluster 8) and 79 (for cluster 5) items. It appears that the most of the sales behaviors expected by the experts in apparel sales are included in these eight prototypes: the prototypes 2, 3 and 8 summarize

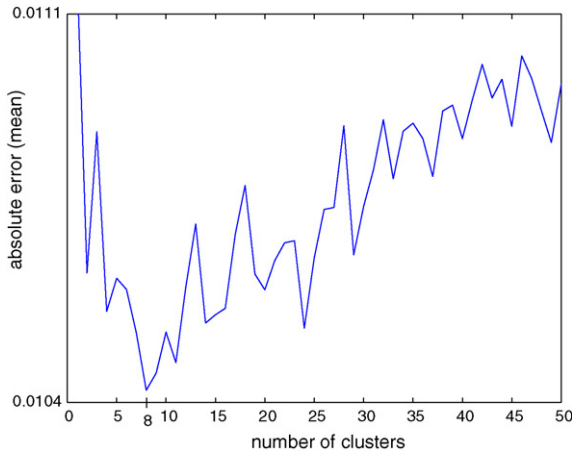

Fig. 6. Number of clusters and absolute error.

items with important sales in the beginning of the life cycle; the prototype 5 represents items which need a small period before the sales grow; items with a pick of sales during the season are characterized by prototypes 1 and 7; prototype 6 is composed by items for which the sales increase during the season; and prototype 4 proposes relatively constant sales during the season.

3.4. Classification of new items

The next stage consists in building a PNN to link the sales behavior of items (prototype) to their descriptive criteria. The learning process of the PNN is carried out with the data of the historical items: the inputs are the descriptive criteria and the target output is the cluster of the considered item previously computed.

The used PNN is composed of three layers (see Section 2.1.2): the input layer with three neurons for each of the descriptive criteria (price, starting time of the sales and life span), the pattern layer with 482 neurons for each different input vector of the training data, and the summation layer with eight neurons since the clustering procedure has built eight classes.

Fig. 8 shows the results of the NCC system. The prototypes which have been assigned by the PNN, are presented with the future items used for the simulation procedure. The sales profiles of the items classified in the clusters 1, 4, 5, 7 and 8 are relatively similar to the prototype, except for a few items. These results should be considered accurate for a sales forecasting one season ahead, without historical sales data. Results obtained with clusters 3 and 6 seem more mitigated but they stay quite acceptable. However, prototype 2 is not representative of the items assigned in the related cluster. Therefore, items are a quite similar profile but these profiles are different from the prototype. This could result from the fact that the attributes used to build the PNN are not rich enough to explain the sales behavior of these items.

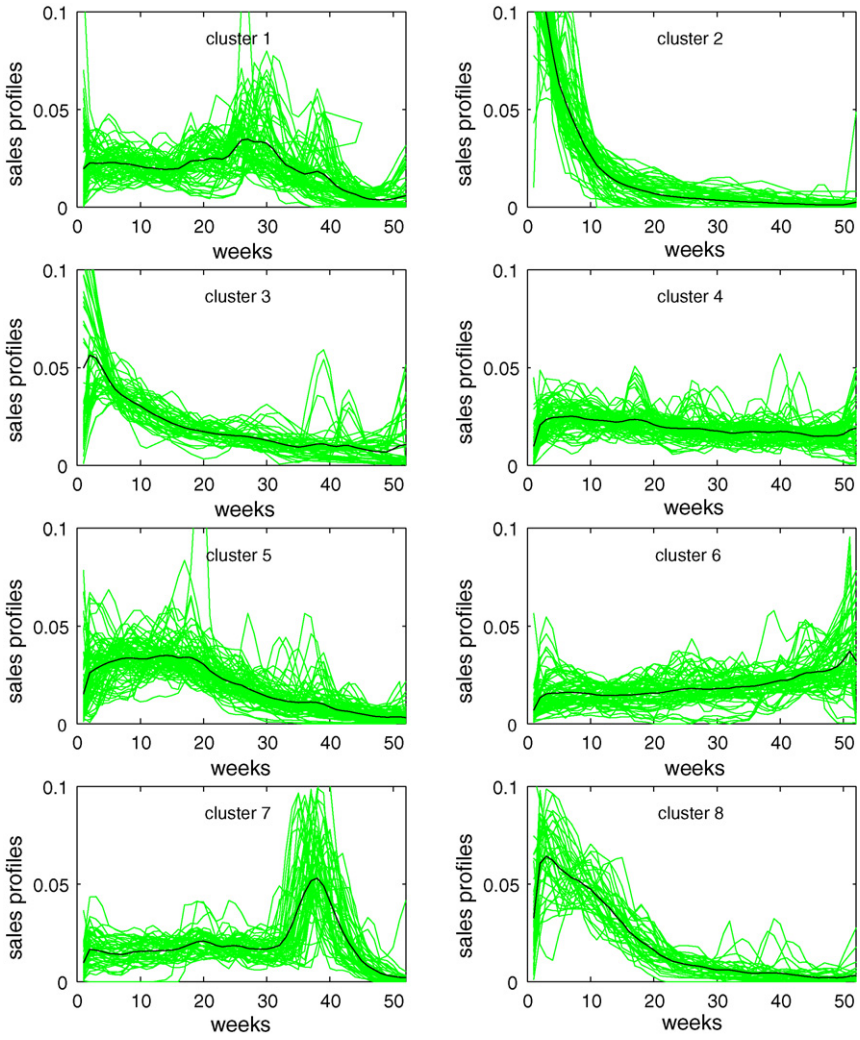

Fig. 7. The eight clusters and their associated prototypes.

The analysis of each class of new items performed by the PNN reveals that this automatic process enables a homogeneous and logical classification. The examples presented in Fig. 9 show that:

- Items associated with prototype 1 have a low starting date with a long life span or a mean starting date with a mean life span. This means that the selling period of these items is centered in the year. As a result, these items are associated with a prototype whose the sales are more important in the middle of the year.
- Items with low price and sold at the beginning of the year (low starting date) are associated with prototype 2. This prototype which presents important sales at the beginning of the year, should characterize ''one shot'' items. These items which are often fashion and low cost items, are ordered one time (often at the beginning of the season) and sold until it is out of stock without replenishment.
- The items associated with prototype 5 have a mean price, a mean life span and a wide range of starting date, i.e., they should be considered as ''mean'' items. The prototype 5 typically represents the basic life cycle of apparel items with a quick sales growth followed by a low decrease. Thus, the ''mean'' items are associated with ''mean'' life curve.
- The items with a relatively low starting date and a relatively high life span, i.e., which are sold at the end of the year, are associated with prototype 6. This prototype is characterized by an increase of sales at the end of the year.

Finally, Table 1 represents the global performances of our model and others basic methods in term of Mean Absolute Percentage Error (MAPE) and Median Absolute Percentage Error (MdAPE). The comparison with the mean predictor shows that our NCC system is globally most accurate of 25%

Table 1
Comparison of MAPE and MdAPE criteria between basic models and NCC model on 285 items

	Mean profile	ZeroR	OneR	Naive Bayesian	NCC
MAPE	203	218	194	145	146.8
MdAPE	119	136	100	91	89

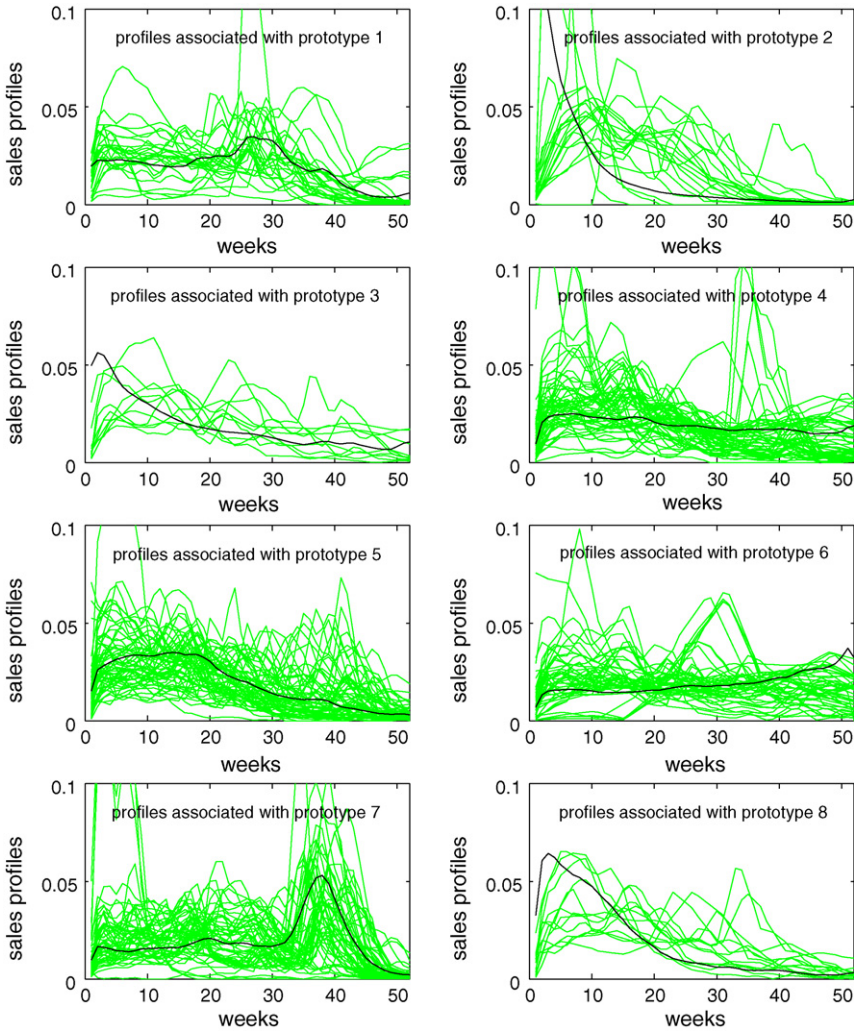

Fig. 8. Assigned prototypes and the real profiles of the 285 items used for simulation.

for the MdAPE criterion and 27% for MAPE criterion. However, the accuracy of the PNN is quite similar to the Naive Bayesian classifier which is a more basic model. Once again, this should be due to the input data which do not enable the PNN to exploit its classification capacities. Consequently, the information extracted from the data is not sufficient to perform a more accurate forecast than a simpler model such as the Naive Bayesian classifier.

4. Conclusions and perspectives

The apparel industry is very specific and requires to develop suitable tools for the sales forecasting. The proposed NCC model, based on neural technique (Self-Organizing Map and Probabilistic Neural Network) is useful to estimate sales profiles of new items for which we have no historical sales data. The clustering procedure is carried out with of a two-level method which is composed of a SOM network and a *k*-means clustering. This stage leads to the computation of prototypes which summarize the sales behavior of the historical items. The PNN, which performs the classification procedure, determines the prototype the more suitable for each future item, from its descriptive criteria. These prototypes constitute the forecasting of the sales profiles.

The results, obtained with a data base of 482 items (historical items) for the training process and 285 items (future items) for the simulation process, confirm that our model globally increases the accuracy of mid-term forecasting in comparison with the mean sales profile predictor. The 2 two-level clustering method enables us to extract eight prototypes from historical sales data. The PNN is able to perform a homogeneous and logical classification of new items from their descriptive criteria. However, a more detailed analysis shows that some mistakes arise from the PNN. In one case, the predicted prototypes are different from the real profiles of the future items. These should result from the fact that the used descriptive criteria are not rich enough to discriminate the sales behaviors of all future items. Further criteria, such as style or textile material must be considered in the hope of increasing the forecasting accuracy, but such data are not often available in the apparel industry.

Finally, it would be interesting to compare our NCC model with other powerful classifiers such as decision tree, genetic

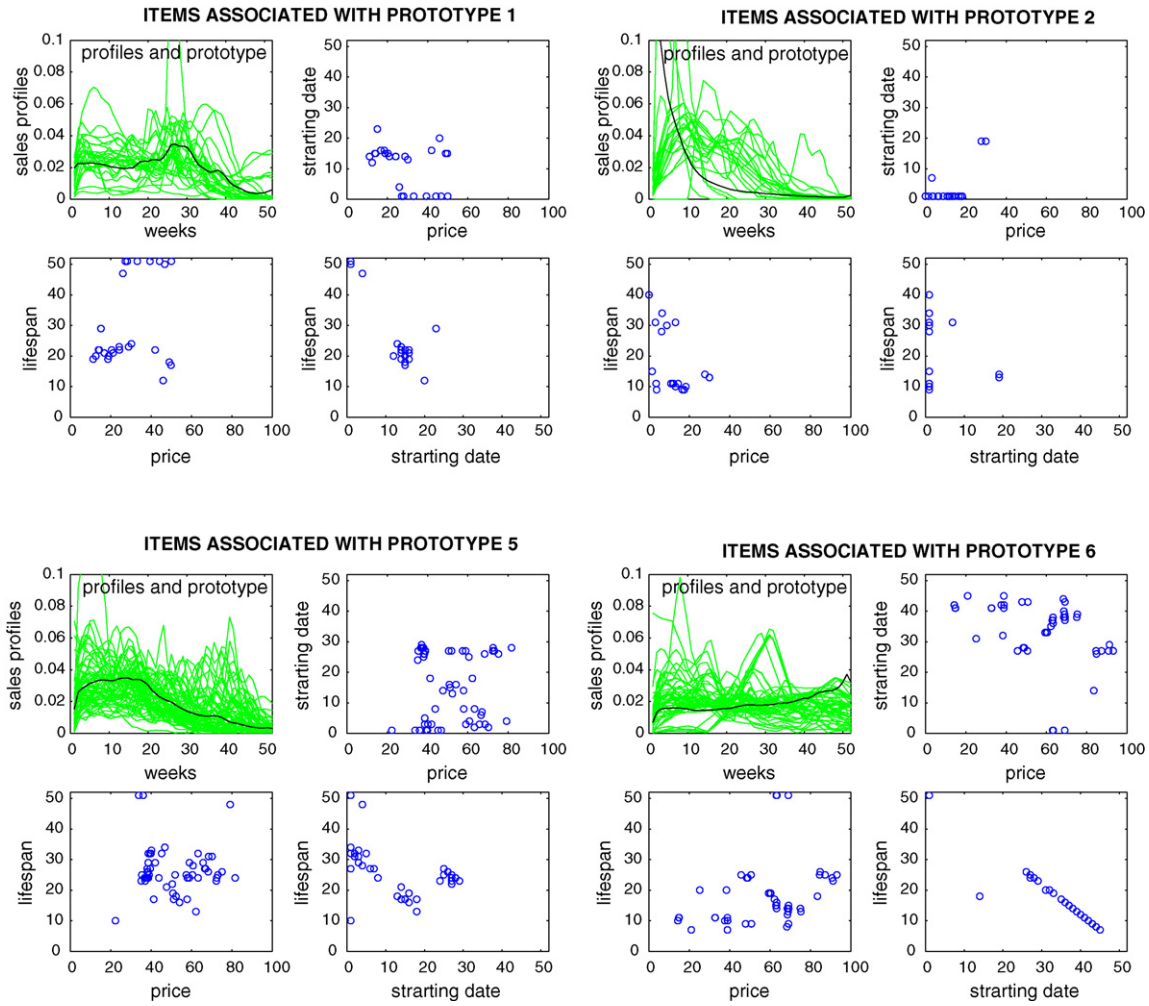

Fig. 9. Examples of descriptive criteria of new items associated with their prototype.

programming or fuzzy logic on other data sets to confirm the efficiency of our proposed system. We let these perspectives as future works.

References

[1] E.I. Altam, G. Macro, F. Varetto, Corporate distress diagnosis: comparison using linear discriminant analysis and neural networks, J. Banking Finance 18 (1994) 505–529.

[2] I. Anagnostopoulos, C. Anagnostopoulos, V. Loumos, E. Kayafas, Classifying Web pages employing a probabilistic neural network, IEE Proc. Software 151 (3) (2004) 139–150.

[3] J. Andrés, M. Landajo, P. Lorca, Forecasting business profitability by using classification techniques: a comparative analysis based on a Spanish case, Eur. J. Operat. Res. 167 (2) (2005) 518–542.

[4] J.S. Armstrong, Principles of Forecasting—A Handbook for Researchers and Practitioners, Kluwer Academic Publishers, Norwell, MA, 2001 .

[5] K. Bertels, J.M. Jacques, L. Neuberg, L. Gatot, Qualitative company performance evaluation: linear discriminant analysis and neural network models, Eur. J. Operat. Res. 115 (1999) 608–615.

[6] J.C. Bezdek, Some new indexes of cluster validity, IEEE Trans. Syst. Man Cybernetic 28 (1998) 301–315.

[7] C.S. Bos, P.H. Franses, M. Ooms, Inflation, forecast intervals and long memory regression models, Int. J. Forecast. 18 (2002) 243–264.

[8] H. Braun, J.S. Chandler, Predicting stock market behavior through rule induction: an application of the learning-from-examples approach, Decision Sci. (1987) 415–429.

[9] K.M.O. Bryson, Evaluation of decision trees: a multi-criteria approach, Comput. Operat. Res. 31 (2004) 1933–1945.

[10] S.W.K. Chan, M.W.C. Chong, Unsupervised clustering for nontextual web document classification, Decision Support Syst. 37 (2004) 377–396.

[11] T. Chen, M.J.J. Wang, Forecasting methods using fuzzy concepts, Fuzzy Sets Syst. 105 (1999) 339–352.

[12] M. Craven, J. Shavlik, Using neural networks for data mining, Future Generat. Comput. Syst. 13 (1997) 211–229.

[13] C.G. Dasgupta, G.S. Dispensa, S. Ghose, Comparative the predictive performance of a neural network model with some traditional market response models, Int. J. Forecast. 10 (1994) 235–244.

[14] D.G.T. Denison, C.C. Holmes, B.K. Mallick, A.F.M. Smith, Bayesian Methods for Nonlinear Classification and Regression, Wiley Series in Probability and Statistics, Wiley, UK, 2002.

[15] A. Fiordaliso, A nonlinear forecasts combination method based on Takagi-Sugeno fuzzy systems, Int. J. Forecast. 14 (1998) 367–379.

[16] H.L. Garcia, I.M. Gonzalez, Self-organizing map and clustering for wastewater treatment monitoring, Eng. Appl. Artif. Intelligence 17 (2004) 215–225.

[17] L.R. Garrison, R.H. Michaelsen, Symbolic concept acquisition: a new approach to determining underlying tax law constructs, J. Am. Tax Assoc. 11 (1) (Fall) (1989) 77–91.

[18] J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan-Kaufman, San Francisco, 2000.
[19] R. Hecht-Nielsen, Counter-propagation networks, in: IEEE First International Conference on Neural Networks, vol. 2, 1987.
[20] J.V. Hensen, J.B. McDonald, J.D. Stice, Artificial intelligence and generalized qualitative-response models: an empirical test on two audit decision-making domains, Decision Sci. 23 (3) (1992) 708–723.
[21] C.C. Holt, Authors retrospective on forecasting seasonals and trends by exponentially weighted moving averages, Int. J. Forecast. 20 (2004) 11–13.
[22] R.C. Holte, Very simple classification rules perform well on most commonly used datasets, Machine Learning 11 (1993) 63–91.
[23] Z. Huang, H. Chena, C.J. Hsua, W.H. Chenb, S. Wu, Credit rating analysis with support vector machines and neural networks: a market comparative study, Decision Support Syst. 37 (2004) 543–558.
[24] L. Ji, A.J. Peters, Forecasting vegetation greenness with satellite and climate data, IEEE Geosci. Remote Sens. Lett. 1 (1) (2004) 3–6.
[25] H. Jin, K.S. Leung, M.L. Wong, An integrated self-organizing map for the traveling salesman problem, in: N. Mastorakis (Ed.), Advances in Neural Networks and Applications, World Scientific and Engineering Society Press, 2001, pp. 235–240.
[26] X. Jin, D. Srinivasan, R.L. Cheu, Classification of freeway traffic patterns for incident detection using constructive probabilistic neural networks, IEEE Trans. Neural Networks 12 (5) (2001) 1173–1187.
[27] H. Jin, W.H. Shum, K.S. Leung, M.L. Wong, Expanding self-organizing map for data visualization and cluster analysis, Informat. Sci. 163 (2004) 157–173.
[28] G.H. John, P. Langley, in: Proceedings of the 11th Conference on Uncertainty in artificial Intelligence, Montreal, Canada, Morgan Kaufmann, San Francisco, Estimating continuous distributions in Bayesian classifiers (1995) 338–345.
[30] M.W. Kattan, R.B. Cooper, A simulation of factors affecting machine learning techniques: an examination of partitioning and class proportions, Omega 28 (2000) 501–512.
[31] M.Y. Kiang, Extending the Kohonen self-organizing map networks for clustering analisys, Computat. Statist. Data Anal. 38 (2001) 161–180.
[32] T. Kohonen, Self-Organizing Maps, Springer, Berlin/Heidelberg, Germany, 1995.
[33] T. Kohonen, P. Somervuo, How to make large self-organizing maps for nonvectorial data, Neural Networks 15 (2002) 945–952.
[34] T. Kohonen, Self-Organizing Maps, third extended ed., Springer Series in Information Sciences, Springer, Berlin, 2001.
[35] K. Krishna, M. Narasimha, Genetic K-means algorithm, IEEE Trans. Syst. Man Cybernetics 29 (3) (1998) 433–439.
[36] R.J. Kuo, K.C. Xue, Fuzzy neural networks with application to sales forecasting, Fuzzy Sets Syst. 108 (1999) 123–143.
[37] J. Lampinen, E. Oja, Clustering properties of hierarchical self-organizing maps, J. Math. Imag. Vis. 2 (23) (1992) 261–272.
[38] R. Li, Z. Wang, Mining classification rules using rough sets and neural networks, Eur. J. Operat. Res. 157 (2004) 439–448.
[39] T.P. Liang, J.S. Chandler, I. Han, J. Roan, An empirical investigation of some data effects on the classification accuracy of Probit, ID3 and neural networks, Contemporary Account. Res. 9 (1992) 306–328.
[40] P.J.G. Lisboa, A review of evidence of health benefit from artificial neural networks in medical intervention, Neural Networks 15 (2002) 11–39.
[41] H. Lu, R. Setiono, H. Liu, Effective data mining using neural networks, IEEE Trans. Knowledge Data Eng. 8 (6) (1996) 957–961.
[42] M.A. Mahmood, G.L. Sullivan, R.L. Tung, A new approach to evaluating business ethics: an artificial neural networks application, J. End User Comput. 11 (3) (1999) 11–19.
[43] B. Mak, T. Munakata, Rule extraction from expert heuristics: a comparative study of rough sets with neural networks and ID3, Eur. J. Operat. Res. 136 (2002) 212–229.
[44] K.F. Man, K.S. Tang, S. Kwong, Genetic Algorithms, Springer-Verlag, New York, 1999.
[45] P. Mangiameli, S.K. Chen, D. West, A constrained orthogonal least-squares method for generating TSK fuzzy models: application to short-term load forecasting, Fuzzy Sets Syst. 118 (2001) 215–233.
[46] P.A. Mastorocostas, J.B. Theocharis, V.S. Petridis, A constrained orthogonal least-squares method for generating TSK fuzzy models: application to short-term load forecasting, Fuzzy Sets Syst. 118 (2001) 215–233.
[47] H. Mattila, R. King, N. Ojala, Retail performance measures for seasonal fashion, J. Fashion Market. Manage. 6 (4) (2002) 340–351.
[48] F. Murtagh, Interpreting the Kohonen self-organizing feature map using contiguity-constrained clustering, Pattern Recognit. Lett. 16 (1995) 399–408.
[49] A. Nigrin, Neural Networks for Pattern Recognition, MIT Press, 1993.
[50] D. Olson, C. Mossman, Neural network forecasts of Canadian stock returns using accounting ratios, Int. J. Forecast. 19 (2003) 453–465.
[51] A.D. Papalexopoulos, T.C. Hesterberg, A regression-based approach to short-term system load forecasting, IEEE Trans. Power Syst. 5 (1990) 1535–1547.
[52] J.H. Park, Y.M. Park, K.Y. Lee, Composite modeling for adaptive short-term load forecasting, IEEE Transact. Power Syst. 6 (1991) 450–457.
[53] E. Parzen, On estimation of a probability density function and mode, Ann. Mathemat. Statist. 33 (1962) 1065–1076.
[54] E. Patuwo, M.Y. Hu, M.S. Hung, Two-group classification using neural networks, Decision Sci. 24 (4) (1993) 825–845.
[55] R. Pavur, A comparative study of the effect of the position of outliers on classical and nontraditional approaches to the two group classification problem, Eur. J. Operat. Res. 136 (2002) 603–615.
[56] P.C. Pendharkar, A computational study on the performance of artificial neural networks under changing structural design and data distribution, Eur. J. Operat. Res. 138 (2002) 155–177.
[57] R.S. Sexton, R.E. Dorsey, Reliable classification using neural networks: a genetic algorithm and backpropagation comparison, Decision Support Syst. 30 (2000) 11–22.
[58] R.D. Snyder, A.B. Koehler, R.J. Hyndman, J.K. Ord, Exponential smoothing models: means and variances for lead-time demand, Eur. J. Operat. Res. 158 (2004) 444–455.
[59] D.F. Specht, Probabilistic neural network, Neural Networks 3 (1) (1990) 109–118.
[60] D.F. Specht, P.D. Shapiro, Generalization accuracy of probabilistic neural networks compared with backpropogation networks, in: IJCNN91 International Joint Conference on Neural Networks, vol. 1, 1991, 887–892.
[61] D. Srinivasan, X. Jin, R.L. Cheu, Evaluation of adaptive neural network models for freeway incident detection, IEEE Trans. Intelligent Transportation Syst. 5 (1) (2004) 1–11.
[62] C.H. St. John, N. Balakrishnan, J.O. Fiet, Modeling the relationship between corporate strategy and wealth creation using neural networks, Comput. Operat. Res. 27 (11, 12) (2000) 1077–1102.
[63] B.G. Tabachnick, L.S. Fidell, Using Multivariate Statistics, fourth ed., Allyn and Bacon, Boston, 2001, Chapter 11.
[64] S. Thomassey, M. Happiette, J.M. Castelain, Textile items classification for sales forecasting, in: ESS 2002, fourth European Simulation Symposium and Exhibition Simulation in Industry, Dresde, Germany, October 23–26, 2002.
[65] S. Thomassey, M. Happiette, J.M. Castelain, X. Zeng, A Short-Term Forecasting System Adapted to Textile Distribution, IPMU 2002, Annecy, France, July 8–9, 2002.
[66] S. Thomassey, M. Happiette, J.M. Castelain, A short and mean-term automatic forecasting system—application to textile logistics, Eur. J. Operat. Res. 161 (2005) 275–284.
[67] B. Tian, M.R. Azimi-Sadjadi, T.H. Vonder Haar, D. Reinke, Temporal updating scheme for probabilistic neural network with application to satellite cloud classification, IEEE Trans. Neural Networks 11 (2000) 903–920.
[68] C.F. Tsai, C.W. Tsai, H.C. Wu, T. Yang, ACODF: a novel data clustering approach for data mining in large databases, J. Syst. Software 73 (2004) 133–145.
[69] A. Ultsch, H.P. Siemon, Kohonens self organizing feature maps for exploratory data analysis, in: Proceedings of the International Neural Network Conference, Dordrecht, Netherlands, (1990), pp. 305–308.
[70] J. Vesanto, E. Alhoniemi, Clustering of the self-organizing map, IEEE Trans. Neural Networks 11 (3) (2000) 586–600.

[71] A. Varfis, C. Versino, Clustering of socio-economic data with Kohonen maps, Neural Network World 2 (6) (1992) 813–834.
[72] R.L. Wilson, R. Sharda, Bankruptcy prediction using neural networks, decision support system 11 (5) (1994) 545–557.
[73] S. Wu, T.W.S. Chow, Clustering of the self-organizing map using a clustering validity index based on inter-cluster and intra-cluster density, Pattern Recognit. 37 (2004) 175–188.
[74] H. Yoo, R.L. Pimmel, Short-term load forecasting using a self-supervised adaptative neural network, IEEE Trans. Power Syst. 14 (2) (1999) 779–784.
[75] L.M. Zampighi, C.L. Kavanau, G.A. Zampighia, The Kohonen self-organizing map: a tool for the clustering and alignment of single particles imaged using random conical tilt, J. Struct. Biol. 146 (2004) 368–380.
[76] A. Zapranis, D. Ginoglou, Forecasting corporate failure with neural networks approach: The Greek case, J. Financial Manage. Anal. 13 (2) (2000) 1–11.
[77] G.P. Zhang, M. Qi, Neural network forecasting for seasonal and trend time series, Eur. J. Operat. Res. 160 (2) (2003) 501–514.
[78] C. Zopounidis, M. Doumpos, Multicriteria classification and sorting methods: a literature review, Eur. J. Operat. Res. 138 (2002) 229–246.