Forecasting Demand in the Clothing Industry

Eduardo Miguel Rodrigues[1], Manuel Carlos Figueiredo[2]
1Eduard.m.rodrigues@gmail.com, 2 mcf@dps.uminho.pt

1. Department of Production and Systems, University of Minho, Portugal
2. Centro Algoritmi, University of Minho, Portugal.

ABSTRACT

For many clothing companies the range of goods sold is renewed twice a year. Each new collection includes a large number of new items which have a short and well defined selling period corresponding to one selling season (20-30 weeks). The absence of past sales data, changes in fashion and product design causes difficulties in forecasting demand accurately. Thus, the predominant factors in this environment are the difficulty in obtaining accurate demand forecasts and the short selling season for goods. An inventory management system designed to operate in this context is therefore constrained by the fact that demand for many goods will not generally continue into the future. Using data for one particular company, the accuracy of demand forecasts obtained by traditional profile methods is analysed and a new approach to demand forecasting using artificial neural networks is presented. Some of the main questions concerning the implementation of neural network models are discussed.

Keywords: Forecasting Demand, Clothing Industry, Artificial Neural Networks

1. INTRODUCTION

In a typical clothing business, a large proportion of the goods sold have a short and well-defined selling period. Fashion goods are typically sold for only one season (6 months or less), and more standard products are rarely included in more than two or three consecutive seasons. An inventory management system designed to operate in this context is therefore constrained by the fact that demand for goods will not generally continue into the future, (Thomassey, S. [2010]).

Thus, the predominant factors of inventory management in this environment are the difficulty in obtaining accurate demand forecasts and the short selling season for goods. The difficulty in sales forecasting arises from business and economic changes from year to year, changes in fashion and product design, and from the absence of sales data for new products. In addition, the selling season length ranges usually from only twenty to thirty weeks, with most items being sold just for one or two seasons.

Another important feature is the length of suppliers' delivery lead times. For items having short lead times it may be possible to make several "small" production orders during the season, and therefore the effect of initial and early season forecast errors may be corrected or at least minimised as the season progresses and demand forecasts become more accurate. For items with long lead times there are usually just one or two stock-replenishment opportunities during the season. A large quantity must therefore be made prior to the start of the season and any serious forecast errors made at this stage will result in overstocking or shortages difficult to correct.

One main purpose of this paper is to present a different approach to demand forecasting using neural networks. The main questions concerning the implementation of neural network models are discussed in section 2. The results of the neural network models developed for forecasting are presented in section 3. Conclusions are presented in section 4.

Although this paper is based on data obtained from a particular Portuguese company, it is anticipated that this work will be relevant to many other situations concerning the stock control of products with short life cycles, particularly those where fashion plays an important role.

2. FORECASTING DEMAND

2.1 Introduction

Forecasting demand for products is not easy but is essential. It is not easy because most lines are sold for only one selling season therefore there is no demand history for most products. It is essential because there are significant costs related to stockouts and end-of-season stock surpluses, (Aksoy, et al. [2012]).

An important issue is the knowledge we have of the demand process at different stages of the operation. The decision to sell a new product is mainly based on subjective estimates of its sales potential which are naturally subject to a large amount of error. Initially, there is a great uncertainty concerning the sales potential for new products. A company will therefore need to improve these initial estimates to avoid situations of early season stockouts or end-of-season stock surpluses.

Once the season has started and actual demand data become available, it is possible to obtain much better estimates using profile methods (Chambers and Eglese [1998]). These methods are based on the assumption that the cumulative demand for a line will build up throughout the selling season as for similar lines in the past. All the lines are divided in groups with similar characteristics (for example, men's long sleeve shirts, jeans, etc.) and a prior estimate is made for the percentage of the final demand expected for the item by each week of the selling season.

A forecast, F_{iw}, of total demand for product i, calculated at end of week w of the selling season, is given by:

$$F_{iw} = \frac{\sum_{t=1}^{w} d_{it}}{P_{iw}}$$

d_{it} = demand received for product i during week t of the season.
P_{iw} = percentage of the final demand expected by end of week w, according to the profile assigned to product i.

The errors in forecasting total demand for a product i, may be caused by random fluctuations in the observed demand d_{it}, or by the inaccuracy in estimating the weekly percentages P_{iw}.

Thus for this approach to be effective, the groups have to be chosen so that the products in each group have similar sales patterns. However if the definition of a group is too specific it will be difficult to allocate new products to appropriate groups. Therefore errors in estimating the profile for a group will arise because the true product profile may be different from that of the group as a whole. Moreover, the method is based on the assumption that group profiles will not change from one year to the next. In practice however some variation will occur due to changes in weather, business environment, fashion trends, etc. Thus, although these demand forecasts are much more accurate than the pre-season demand estimates, they may still be rather inaccurate for the first weeks of the selling season.

A method used in practice to improve these forecasts consists of applying appropriate scaling factors for each of the different departments into which merchandise is typically divided (footwear, knitwear, etc.).

An estimate of the revenue expected for a particular product may be obtained by multiplying the sales price by its demand forecast. Summing these estimates over all the products for a particular department gives the expected income for that department. Managers can then use their experience and knowledge of the market to decide whether the expected income for each department is reasonable and eventually they can decide to revise forecasts. Individual forecasts are then multiplied by an appropriate scaling factor to get a new forecast. In general, the accuracy of early season forecasts can benefit from this process of departmental scaling.

2.2. Artificial neural networks

An artificial neural network (ANN) is a computational structure composed of simple, interconnected processing elements. These processing elements, called neurons, process and exchange information as the biological neurons and synapses within the brain. Although the motivation for the development of artificial neural networks came from the idea of imitating the operation of the human brain, biological neural networks are much more complicated than the mathematical models used to represent artificial neural networks.

Neural networks are organised in layers. These layers are made up of a number of interconnected neurons. Patterns are presented to the network through the input layer, which is connected to the hidden layers where the data processing is done through a system of weighted connections. The hidden layers then link to an output layer where the result is presented. When a network has been structured for a particular application, this network is ready to start learning.

$i = [i_1, i_2, i_3] =$ input vector
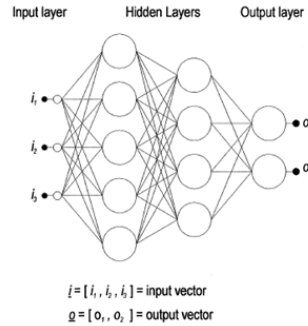$o = [o_1, o_2] =$ output vector

Figure 1: Artificial Neural Network

A neuron can receive many inputs. Each input has its own weight which gives the input influence on the neuron's summation function. These weights perform the same type of function as do the varying synaptic strengths of biological neurons. In both cases, some inputs are made more important than others so that they have a greater effect on the neuron as they are combined to produce the neural response. Thus, weights are the coefficients within the network that determine the intensity of the input signal sent to an artificial neuron. They are a measure of the strength of a connection. These weights are modified as the network is trained in response to the training data and according to the particular learning algorithm used.

The first step in the neural network operation is for each of the inputs to be multiplied by their respective weighting factor. Then these modified inputs are fed into the summing function, which usually just sums all the terms (although a different operation may be selected). These values are then propagated forward and the output of the summing function is sent to a transfer function. The application of this function produces the output of a neuron. In some transfer functions the input is compared with a threshold value to determine the neural output. If the result is greater than the threshold value the neuron generates a signal, otherwise no signal is generated. The most popular choice for the transfer function is an S-shaped curve, such as the logistic and the hyperbolic tangent. These functions have the advantage of being continuous and also having continuous derivatives. Sometimes, after the transfer function, the result can be limited. Limits are used to insure that the result does not exceed an upper or lower bound. Each neuron then sends its output to the following neurons it is connected to. This output is, in general, the transfer function result.

The brain basically learns from experience. Neural networks are sometimes called machine learning algorithms, because changing its connection weights (training the network) makes the network learn to solve a particular problem. The strength of the connections between neurons is stored and the system "learns" by adjusting these connection weights. Thus, neural networks "learn" from examples and have some capacity to generalise. Learning is achieved by adjusting the weights between the connections of the network, in order to minimise the performance error over a set of example inputs and outputs. This characteristic of learning from examples makes neural network techniques different from traditional methods.

Most artificial neural networks have been trained with supervision. In this mode, the actual output of a neural network is compared to the desired output. Weights, which are usually randomly set at the beginning, are then adjusted by the network so that the next iteration will produce a smaller difference between the desired and the actual output. The most widely used supervised training method is called backpropagation. This term derives from backward propagation of the error, the method used for computing the error gradient for a feedforward network, an algorithm reported by Rumelhart, et al. [1986a], [1986b]. This method is based on the idea of continuously modifying the strengths of the input connections to reduce the difference between the desired output value and the actual output. Thus, this rule changes the connection weights in the way that minimises the error of the network. The error is back propagated into previous layers one layer at a time. The process of back-propagating the network errors continues until the first layer is reached.

The way data is represented is another important issue. Artificial neural networks need numeric input data. Therefore, the raw data must often be conveniently converted. It may be necessary to scale the data according to network characteristics. Many scaling techniques, which directly apply to artificial neural network implementations, are available. It is then up to the network designer to find the best data format and network architecture for a particular application.

2.3. Architecture of the Neural Networks

Artificial neural networks are all based on the concept of neurons, connections and transfer functions, and therefore there is a similarity between the different architectures of neural networks. The differences among neural networks come from the various learning rules and how those rules modify a network.

A backpropagation neural network has generally an input layer, an output layer, and at least one hidden layer (although the number of hidden layers may be higher few applications use more than two). Each layer is fully connected to the succeeding layer.

The number of layers and the number of neurons per layer are important decisions. It seems that there is no definite theory concerning the best architecture for a particular problem. There are however some empirical rules that have emerged from successful applications.

After the architecture of the network has been decided, the next stage is the learning process of the network. The process generally used for a feedforward network is the Delta Rule or one of its variants. The first step consists of calculating the difference between the actual outputs and the desired outputs. Using this error, the connection weights are increased proportionally to the error times a scaling factor for global accuracy. Doing this for an individual neuron means that the inputs, the output, and the desired output all have to be present at the same neuron. The complex part of this learning mechanism is for the system to determine which input contributed the most to an incorrect output and how that element should be changed to correct the error.

The training data is presented to the network through the input layer and the output obtained is compared with the desired output at the output layer. The difference between the output of the final layer and the desired output is then back-propagated to the previous layer, generally modified by the derivative of the transfer function, and the connection weights are normally adjusted using the Delta Rule. This process continues until the input layer is reached.

In general, backpropagation requires a considerable amount of training data. One of the main applications for multilayer, feedforward, backpropagation neural networks is forecasting.

2.4 Issues in building an ANN for forecasting

There are many different ways to construct and implement neural networks for forecasting, although most practical applications use multi-layer perceptrons (MLP). In the MLP, all the input nodes are in one input layer, all the output nodes are in one output layer and the hidden nodes are distributed into one or more hidden layers.

There are many different approaches for determining the optimal architecture of an ANN such as the pruning algorithm, the polynomial time algorithm or the network information criterion (Murata et al. [1994]). However, these methods cannot guarantee the optimal solution for all types of forecasting problems and are usually quite complex and difficult to implement. Thus, several heuristic guidelines have been proposed to limit design complexity in ANN modelling (Crone, [2005]).

The hidden nodes are very important for the success of a neural network application since they allow neural networks to identify the patterns in the data and to perform the mapping between input and output variables. Many authors use only one hidden layer for forecasting purposes and there are theoretical works, which show that a single hidden layer is sufficient for ANNs to approximate any complex nonlinear function with any desired accuracy (Hornik et al. [1989]).

According to Zhang et al. [1998], one hidden layer may be enough for most forecasting problems, although using two hidden layers may give better results for some specific problems, especially when one hidden layer network requires too many hidden nodes to give satisfactory results.

The number of input nodes corresponds to the number of input variables used by the model. In general, it is desirable to use a small number of relevant inputs able to reveal the important features present in the data. However, too few or too many input nodes can affect either the learning or the forecasting ability of the network.

Another important decision in ANN building concerns the activation function, also called the transfer function. The activation function determines the relationship between inputs and outputs of a

node and a network introducing usually nonlinearity that is important for most ANN applications. Although in theory any differentiable function may be used as an activation function, in practice only a few activation functions are used. These include the logistic, the hyperbolic tangent and the linear functions.

The training process of a neural network is an unconstrained non-linear minimisation problem. The weights of a network are iteratively modified in order to minimise the differences between the desired and actual output values.

The most used training method is the backpropagation algorithm which is basically a gradient steepest descent method. For the gradient descent algorithm, a step size, called the learning rate, must be specified. The learning rate is important for backpropagation learning algorithm since it determines the magnitude of weight changes. The algorithm can be very sensitive to the choice of the learning rate. A small learning rate will slow the learning process while a large learning rate may cause network oscillation in the weight space.

The most important measure of performance for an ANN model is the forecasting accuracy it can achieve on previously unseen data. Although there can be other performance measures, such as the training time, forecasting accuracy will be the primary concern. Accuracy measures are usually defined in terms of the forecasting error, et , which is the difference between the actual and the predicted value. Many studies concluded that the performance of ANNs is strong when compared to various more traditional forecasting methods (Sun et al [2008]).

3. RESULTS

We used the software *NeuroSolutions* to develop the neural network models. We built multi-layer, feed-forward neural networks in either three or four layer structures with linear, logistic or hyperbolic tangent transfer functions. The data consisted of the demand received for a sample of products over the 2008 and 2009 Spring/Summer selling seasons and information on the characteristics of products and customers.

The performance of the neural networks developed was measured in terms of the average absolute error (AAE) made at forecasting the total demand for the lines on a test set formed by a sample of 120 items from the Spring/Summer 2010 selling season. The choice of the AAE as the measure of performance was motivated by its close relation to the actual cost function in practice, since the estimated cost of a stockout unit is similar to the cost of a stock unit left over by the end of the season. The results of the best neural network model, in terms of the minimum average absolute error (AAE), are presented in figure 2.
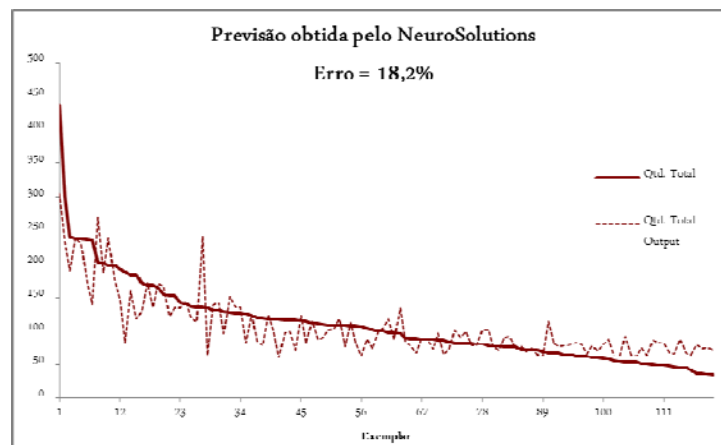

Figure 2: Artificial Neural Network Performance

The mean absolute percentage error (MAPE) for the 120 items in the sample was 18,2 %. This result was regarded as very satisfactory from the managers of the company involved in this study.

5

4. CONCLUSIONS

The life cycle of products in many business areas is becoming shorter. This trend for shorter life cycles is caused by faster changes in consumer preferences that make some products fashionable (as in the pc-games or in the toys industries). It may be also the result of a faster rate of innovation, especially in the high technology industry (as for mobile phones or personal computers). The delivery lead time or the production cycle is frequently large enough for decisions for a substantial part of the life cycle of the products to have to be made early after starting commercialisation.

Apart from this trend for shorter product life cycles, there are some industries that are traditionally fashion dependent (as are the clothing and footwear industries) where products are replaced usually every six months. In those contexts, the difficulty in forecasting accurately demand for new items is a serious obstacle to an efficient operation.

The neural network models described may be used to improve demand forecasts, allowing the revision of procurement or production decisions to avoid high obsolescence costs at the end of the life cycles of products or to avoid stockouts affecting levels of customer service.

REFERENCES

Aksoy, A., Ozturk, N. and Sucky, E.,(2012) *"A decision support system for demand forecasting in the clothing industry"*, International Journal of Clothing Science and Technology, Vol. 24 Iss: 4, pp.221 - 236

Chambers, M.L., and Eglese, R.W.,(1988), *"Forecasting demand for mail order catalogue lines during the season."*, European Journal of Operational Research, Vol. 34, pp. 131-138.

Crone,S.F.,(2005), *"Stepwise Selection of Artificial Neural Networks Models for Time Series Prediction"*, Journal of Intelligent Systems, Vol. 14, No. 2-3, pp. 99-122.

Hornik, K., Stinchcombe, M., and White, H.,(1989), *"Multilayer feedforward networks are universal approximators"* Neural Networks, 2, pp. 359-366.

Murata, N., Yoshizawa, S., and Amari, S.,(1994), *"Network information criterion - determining the number of hidden units for an artificial neural network model."*,IEEE Transactions on Neural Networks, Vol. 5, No.6, pp. 865-872.

Rumelhart, D.E., Hinton, G.E., and Williams,R.J.,(1986a), *"Learning representations by backpropagating errors"*, Nature, 323, 533-536.

Rumelhart, D.E., Hinton, G.E., and Williams,R.J.,(1986b), *"Learning internal representations by error propagation"*, in Parallel Distributed Processing : Explorations in the microstructure of Cognition, D.E. Rumelhart , J.L. McClelland (eds.), MIT Press.

Sun, L., Choi, T., Au, K. and Yu, Y. (2008), *"Sales forecasting using extreme learning machine with applications in fashion retailing"*, Decision Support Systems, Vol. 46, pp. 411-19.

Thomassey, S. (2010), *"Sales forecasts in clothing industry: the key success factor of the supply chain management"*, International Journal of Production Economics, Vol. 128, pp. 470-83.

Zhang, G., Patuwo, E., and Hu, M.,(1998), *"Forecasting with artificial neural networks: The state of the art"*, International Journal of Forecasting, 14, pp. 35-62.