2009 World Congress on Computer Science and Information Engineering

A Hybrid Subspace-Connectionist Data Mining Approach for Sales Forecasting in the Video Game Industry

Julie Marcoux, and Sid-Ahmed Selouani

Université de Moncton, Campus de Shippagan, Canada
ejm4734@umoncton.ca, selouani@umcs.ca

Abstract- *This paper addresses the issue of sales forecasting using a new approach based on connectionist and subspace decomposition methods. A tool is designed to support company management in the process of determining expected sales figures. Neural networks trained with a back-propagation algorithm are used to predict the weekly sales of a video game. For this purpose, optimal topology is found and a time-sensitive neural network is implemented. We have considered the use of many influencing indicators and parameters as inputs. In order to assess the relevance of these parameters, we perform a pre-processing based on Principal Component Analysis. The performance of the proposed system is evaluated and compared with baseline reference sales. The results are presented and discussed with regards to prediction accuracy.*

1. INTRODUCTION

Sales forecasting is an important part of business management since it provides relevant information that can be used to make strategic business decisions. Forecasting can be divided in three categories: future forecasts, environmental forecasts and industry forecasts. The first category, considered as a company forecast, aims at assessing the performance of a given company regardless of its competitors. Future forecasts can be based on achieving specific performance such as increasing sales. Environmental forecast looks at macroeconomics and provides a longer-term context. Key factors like consumer behavior or inflation are considered within this forecasting process. Industry forecast involves influencing factors such as legislation, market growth, social and cultural trends. This forecast helps produce planning assumptions for elements such as key market drivers and market size growth. In this paper, we are concerned with forecasting of the weekly sales of a video game. Many approaches have been proposed for sales prediction. In practice the methods that companies use for forecasting are mainly based on qualitative studies, focus groups, surveys and consultations, and in-store simulated testing. The sales force can be involved in forecasting since they have trading environment experience. Potential consumers can also be consulted in order to collect their opinions on the factors that could influence sales. The relevance of each approach will vary. Generally speaking, accurate methods take a long time to be effective and are the most expensive. Besides deploying such methods yields a significant financial risk that companies might not be willing to take. As an alternative to the weakness of qualitative methods, Time Series Analysis (TSA) methods have been developed. Based on analysis, coding and integration of unstructured with structured data, quantitative (data-driven) techniques have the advantage of being simple and easy to implement. They require the company to identify trends from analyzed data and to forecast based on the conclusions drawn. The principle of TSA consists of learning average sales achieved over several recent periods to predict the sales in the following period. Among TSA methods, neural networks are considered as powerful and relatively simple to implement. The use of Neural Networks (NN) has already been demonstrated in a variety of forecasting applications, such as stock market and currency exchange applications [4]. The idea consists of training a neural network by using a supervised learning algorithm in order to give the neural network the ability to generalize the mapping between the inputs (available data) and the outputs (targeted data). This mapping process allows capture of the implicit rules governing the variations of the time series and hence the prediction of its future behavior in the future. In this paper we propose a new scheme of Neural Networks, namely Autoregressive Time Delay Neural Networks (AR-TDNN). Expectation accuracy is strongly related to the quality and type of input data. In our application, relevant and non-redundant input features are obtained through the use of Principal Component Analysis (PCA). This paper is organized as follows. In section 2, we describe the basis of the PCA-based preprocessing approach, and the method we have proposed to optimize the selection of the reconstruction order. In section 3, we describe the AR-TDNN used for the prediction of video game sales. Then, we proceed in section 4 with the description of the system using the hybrid architecture PCA/AR-TDNN that we have introduced to perform the prediction. In section 5, we evaluate the hybrid technique by comparing its performance to the one obtained by baseline systems. Finally, in section 6, we conclude and discuss our results.

978-0-7695-3507-4/08 $25.00 © 2008 IEEE
DOI 10.1109/CSIE.2009.1001

2. PREPROCESSING OF KEY INPUT FEATURES

The principle of subspace decomposition techniques consists of constructing an orthonormal set of axes. These axes point in the directions of maximum variance, thus forming a representational basis that projects along the direction of maximum variability. Applied in the context of relevance determination, these axes permit us to find the optimal and relevant space of data representation. This representation optimization is done by projecting the rough data onto the subspace generated by the low-order components of PCA. The high-order components usually capture the redundancy and hence are not relevant for optimal data representation. Let $\mathbf{x} = [\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}]^{\mathbf{T}}$ be an N-dimensional observation vector, and $\mathbf{R_x}$ the covariance matrices of $\mathbf{x}$. The decomposition of $\mathbf{R_x}$ is given by $\mathbf{R_x} = \mathbf{U}\mathbf{\Lambda_x}\mathbf{U}^{\mathbf{T}}$ where $\mathbf{\Lambda_x}$=diag$(\lambda_{x1}, \lambda_{x2}, ..., \lambda_{xN})$ is the diagonal matrix of eigenvalues given in the decreasing order, and where $\mathbf{U}$ is the eigenvector matrix. Major subspace techniques assume that the most informative components are concentrated in an $r < N$ dimensional subspace, whereas the redundancy occupies the $N - r$ dimensional observation space. Reduction of the redundancies is then achieved by considering only the informative subspace in the reconstruction of the relevant key features set. This is obtained by removing high order components ($> r$). In fact r is not chosen in an arbitrary way, an automatic method for determining this optimum is used. This choice is guided by what is called the *reconstruction's quality function*, denoted by $\mathcal{Q}$. We defined $\mathcal{Q}$ as the ratio of the sum of the eigenvalues used to reconstruct the optimal input vector, to the sum of all the eigenvalues, as follows:

$$\mathcal{Q} = \frac{\sum_{i=1}^{r} \lambda_{xi}}{\sum_{i=1}^{N} \lambda_{xi}}. \tag{1}$$

The first- and second-order derivatives of $\mathcal{Q}$ are given by:

$$\Delta\mathcal{Q} = \frac{\lambda_{xr+1}}{\sum_{i=1}^{N} \lambda_{xi}} \tag{2}$$

and

$$\Delta\Delta\mathcal{Q} \quad = \quad \frac{\lambda_{xr+1} - \lambda_{xr}}{\sum_{i=1}^{N} \lambda_{xi}}. \tag{3}$$

The variations of $\Delta\Delta\mathcal{Q}$ for a certain value of r tend to zero for higher order eigenvalues. Thus, the $\mathcal{Q}$-acceleration function ($\Delta\Delta\mathcal{Q}$) allows determining the optimal component order at which we perform the reconstruction of optimal input vector.

3. AUTOREGRESSIVE TIME-DELAY NEURAL NETWORKS (AR-TDNN)

Because sales variation is a phenomenon sensitive to temporal changes, we consider Recurrent Networks (RNs) to be more adequate than feedforward networks at forecasting tasks. RNs are generally trickier to work with, but they are theoretically more powerful, having the ability to represent temporal sequences of unbounded length. Another consideration related to seasonal effects leads us to use a particular RN: the one proposed by Russel [3] and using an Autoregressive (AR) version of the backpropagation algorithm. This type of network can in principle, naturally capture relationships between long-term events. The approach we are investigating proposes integration, in addition to the AR component, of a time-delay component [5]. Through this combination, we expect that the ability of the system to discern artifacts from strong trend will be increased. The model described by Russel [3] includes an autoregressive memory which constitutes a form of self-feedback in which the output depends on the current output plus a weighted sum of previous outputs. The classical AR node equation is then:

$$y_i(t) = f\big(bias + \sum_{j=1}^{P} w_{i,j} x_j(t)\big) + \sum_{n=1}^{M} a_{i,n} y_i(t-n), \tag{4}$$

where $y_i(t)$ is the output of node *i* at time *t*, *f(x)* is the *tanh(x)* bipolar activation function, *P* is the number of input units, and *M* is the order of autoregressive prediction. Weights $w_{i,j}$, biases, and AR coefficients $a_{i,n}$ are adaptive and are optimized in order to minimize the output error. Our proposition consists of incorporating a time delay component in the input nodes of each layer and Equation 4 then becomes:

$$\begin{aligned} y_i(t) = f\big(bias + \sum_{m=0}^{L}\sum_{j=1}^{P} w_{i,j,m} x_j(t-m)\big) + \\ + \sum_{n=1}^{M} a_{i,n} y_i(t-n), \end{aligned} \tag{5}$$

where *L* is the delay order at the input. Feedforward and feedback weights were initialized from a uniform distribution in the range $[-0.8, 0.8]$. A neuron of the AR-TDNN configuration is shown in Figure 1.

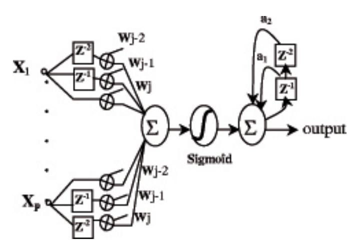

Fig. 1. AR-TDNN unit.

An autoregressive backpropagation learning algorithm performs the optimization of feedback coefficients in order to minimize the mean squared error noted *E(t)* and defined as:

$$E(t) = \frac{1}{2}\sum_{i}\big(d_i(t) - y_i(t)\big)^2, \tag{6}$$

where d_i is the desired value of the i^{th} output node. The weight and feedback coefficient changes, noted respectively $w_{j,i,m}$ and $a_{i,n}$, are accumulated within an update interval $[T_0, T_1]$. In the proposed AR-TDNN version, the update interval $[T_0, T_1]$ is fixed to correspond to the time delay of the inputs. The updated feedback coefficients are written as follows:

$$a_{i,n}^{new} = a_{i,n}^{old} + \frac{1}{T_1 - T_0} \sum_{t=T_0}^{T_1} \Delta a_{i,n}(t), \qquad (7)$$

and if T is the frame duration, the weights are as follows:

$$w_{i,j}^{new} = w_{i,j}^{old} + \frac{1}{LT} \sum_{t=T_0}^{T_1} \Delta w_{i,j}(t). \qquad (8)$$

The calculation of $\Delta a_{i,n}(t)$ variation is detailed in [3]. The optimization of weights and biases are performed as in Waibel's networks [5]. Hence, the $\Delta w_{i,j}$ variations are accumulated during the update interval after accumulating Time-Delay frames at the input.

4. HYBRID FORECASTING SYSTEM

The inputs data and the targeted sales data are combined in a nonlinear way. To approximate the required nonlinear function that performs the mapping, AR-TDNN are used. The AR-TDNN input is a vector composed of key factors. The output of the network $\hat{\mathbf{S}}$ is computed during a training phase using a convergence algorithm to update the weight vector in a manner to minimize the error between the output $\hat{\mathbf{S}}$ and the desired sales value $\mathbf{S}$. The weights of this network are calculated during a training phase with a back-propagation training algorithm using a mean square error criterion. The baseline (reference) forecasting system is composed of a Multi-Layer Perceptron (MLP) using the 18 key factors given in table 1. The proposed system is tested with PCA pre-processed input vectors and with different delay values of the AR-TDNN. The AR-TDNN is trained by using the Nguyen-Widrow initialization conditions [2]. The TDNN part of the system consists of three layers. Each unit in the hidden layer receives input weighted by the coefficients in a three-frame window of the input layer. This time delay is expected to capture the temporal (trend) component. The autoregression part is expected to help in capturing long term trends. An autoregressive order of 2 is chosen and a delay of two frames (weeks) is fixed. The preprocessing performed by the PCA leads to use of lower values of delay and order which ensures the stability of AR nodes. The overview of the hybrid system is given in Figure 2.

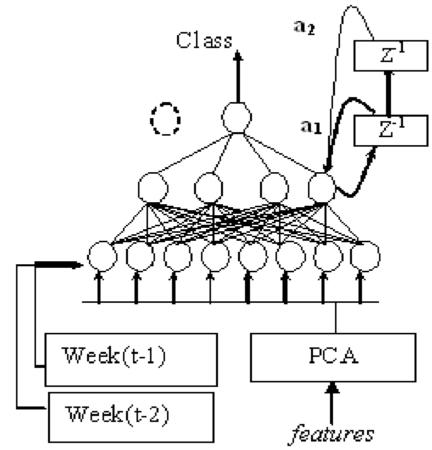

Fig. 2. Overview of the PCA/AR-TDNN system.

	Description
Gaming console	Wii, Xbox 360, etc.
Multiplayer status	single player, multiplayer or both
Third party	information with respect to publisher
Position of first week	variable coding the week of launch
Weeks elapsed	number of weeks elapsed since launch
Last week sales	number of units sold during the last week
MRSP	retail price of a video game
Multiplayer status	choices: single, multiplayer, both
Online status	choices: offline, enhanced, online
Crossover status	available on more than one console or not
Accessory status	accessory: not needed,needed
Sequel status	choices: not a sequel, sequel, re-release
Genre	type of video game as assigned by IGN
Reader reviews	number of readers who submitted evaluation
IGN rating	rating as given by IGN.com
Press average	rating as given by press organizations
Weeks elapsed since publication	weeks elapsed since launch
Units sold the first week	sold in North-America during first week

Table 1. Key input factors used for sales forecasting.

5. EXPERIMENTS AND RESULTS

The number of units sold during a week for each video game was normalized (by applying log 10) and then divided in 10 equal classes. Weekly sales numbers are not taken for the same week of the year for all video games (data entry took about one month), but the principle remains the same. Table 2 was used to train and evaluate all neural networks. Two game developers for which the most data was accumulated were selected for our evaluation. We took testing data for two neural networks with the best overall track record for predictions and analyzed the results specific to each developer.

5.1. First case study: Activision developer

Twenty-seven video games of Activision game developer were included in the testing data. For these combined twenty-seven video games, 152 029 units were sold. Video games were distributed between classes according to Table 3. The simple neural network (nn6Node33) predicts units sold for Activision with an accuracy of 85.1%. An accurate prediction is defined

Class	Normalized Unit Sold	Units Sold
1	1.5-1.9	32-79
2	1.9-2.3	79-200
3	2.3-2.7	200-501
4	2.7-3.1	501-1259
5	3.1-3.5	1259-3162
6	3.5-3.9	3162-7943
7	3.9-4.3	7943-19953
8	4.3-4.7	19953-50119
9	4.7-5.1	50119-125893
10	5.1-5.5	125893-316228

Table 2. Number of units sold during a week by class.

Class	Number of video games
1	0
2	2
3	6
4	4
5	5
6	7
7	2
8	0
9	1
10	0

Table 3. Sale distribution of Activision video games for each class.

as a prediction exactly on or only one class away from the actual class. This result is obtained by the baseline system (a simple feedforward neural network, noted NN) using 28 input units, 33 hidden units and 10 output units. A set of experiments is carried out in order to compare the baseline system (NN) with the hybrid PCA/AR-TDNN system and the PCA-NN system. We found through these experiments that using the PCA/AR-TDNN system enhances the prediction accuracy as shown in Table 4. The second derivative of the reconstruction's quality function $\Delta\Delta\mathcal{Q}$ is used to find the optimal order of the dimension reduction of the input features. For the Activision developer an optimal order of 22 was found. Consequently, the 22 principal components are used as new input features of AR-TDNNs. Note that using the AR-TDNN system without PCA leads to instability of AR nodes. A global correct rate of 89.48% is reached by the hybrid system while the baseline and the PCA-NN systems achieve 85.1 % and 84% respectively. Note that the PCA-NN configuration slightly degrades the performance but a lower number of input variables is used.

Class	NN	PCA-NN	PCA/AR-TDNN
1	1	1	1
2	3	4	1
3	2	2	5
4	6	5	5
5	4	4	4
6	9	10	8
7	1	0	1
8	1	0	1
9	0	1	1
10	0	0	0

Table 4. Prediction of Activision video game sales through the use of NN, PCA-NN and Hybrid PCA/AR-TDNN systems.

Class	Number of video games
1	0
2	0
3	2
4	11
5	14
6	6
7	8
8	4
9	2
10	0

Table 5. Sale distribution of EA video games for each class.

5.2. Second case study: EA developer

Forty-seven video games by this developer were included in the testing data. For these combined forty-seven video games, 428 103 units were sold. Video games were distributed between classes according to the distribution given in Table 5.The simple NN (nn6Node28) system predicts units sold for EA with an accuracy of 68.1%. An accurate prediction is defined as a prediction exactly on or only one class away from the actual class. This result is obtained by the baseline system (a simple feedforward neural network, noted NN) using 28 input units, 28 hidden units and 10 output units. As shown in Table 6, the use of the PCA/AR-TDNN system enhances the prediction accuracy. As it is found in the first case study, 22 is the optimal order of PCA reconstruction function. The 22 principal components are used as input features of AR-TDNNs. Similarly, the use of the AR-TDNN system without PCA leads to instability of AR nodes. A global correct rate of 75.3% is reached by the hybrid PCA/AR-TDNN system while the baseline and the PCA-NN systems achieve 68.1 % and 62.6% respectively. The PCA-NN did not improve prediction accuracy.

Class	NN	PCA-NN	PCA/AR-TDNN
1	0	2	0
2	1	1	1
3	4	3	3
4	5	8	9
5	12	11	12
6	4	6	8
7	4	6	7
8	3	6	5
9	0	2	0
10	0	2	2

Table 6. Prediction of EA video game sales through the use of NN, PCA-NN and Hybrid PCA/AR-TDNN systems.

6. CONCLUSION

In this paper a new data mining technique that aims at predicting weekly video game sales, through the use of a hybrid subspace-connectionist forecasting system (PCA/AR-TDNN), was proposed. Experiments using data of two video game developers show that the use of the hybrid PCA/AR-TDNN significantly increases the accuracy of video game sale forecasting. This method can easily be generalized to other industry sectors if categorical and numerical key factors are available. The least significant predictors could be removed by using PCA pre-processing in order to increase the effectiveness of expectations. It is important to mention that it is easier for established businesses to achieve accurate forecasting than it is for a new business, since an established business can draw on a longer history of past sales. One major advantage of our method is its adaptability to both established and new businesses thanks to the modular topology of AR-TDNN which can integrate varying amounts of a company's historic data.

7. REFERENCES

[1] P. Bazeley, "Issues in Mixing Qualitative and Quantitative Approaches to Research", Buber and Gadner, eds., pp 141-156. Hampshire, United Kingdom: Palgrave Macmillan, 2004.

[2] D. Nguyen, and B. Widrow, "Improving the Learning Speed of Two-Layer Neural Networks by Choosing Initial Values of the Adaptative Weights", Proc. of IJCNN (III), pp. 21-26, 1990.

[3] R.L. Russel, and C. Bartley "The Autoregressive Backpropagation Algorithm", Proc. of IJCNN (II), pp. 369-377, 1991.

[4] F.M. Thiesing, F.M, and O. Vornberger, "Sales forecasting using neural networks International Conference on Neural Networks, Volume 4, pp. 2125 - 2128, 1997.

[5] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, "Phoneme Recognition Using Time-Delay Neural Networks", IEEE Trans. on Audio Speech and Signal Processing, No. 37, pp. 328-339, 1989.