Forecasting Cash Demand in ATM Using Neural Networks and Least Square Support Vector Machine

Cristián Ramírez and Gonzalo Acuña

Universidad de Santiago de Chile, USACH, Departamento de Ingeniería Informática.
Av Libertador Bernardo OHiggins 1363, Santiago, Chile
`cristian.ramirez@gestran.cl`,
`gonzalo.acuna@usach.cl`

Abstract. In this work we forecast the daily ATM cash demand using dynamic models of type Nonlinear Autoregressive Exogeneous inputs (NARX) and Nonlinear Autoregressive Moving Average with Exogeneous Inputs (NARMAX) performed by Neural Networks (NN) and Least Square Support Vector Machine (LS-SVM) and used to predict one step (OSA) or multistep (MPO). The aim is to compare which model perform better results. We found that the Multilayer Perceptron NN presented the best index of agreement with an average of 0.87 in NARX-OSA and 0.85 in NARX-MPO. After, Radial Basis Function NN was 0.82 for both cases. Finally, LS-SVM obtained the worst results with 0.78 for NARX-OSA and 0.70 for NARX-MPO. No significant differences between NARX and NARMAX structures were found. Our contribution would have obtained the 2nd place in the NN5 competition of computational methods.

Keywords: MLP, RBF, LS-SVM, NARX, NARMAX, OSA, MPO, NN5, IA, SMAPE.

1 Introduction

Automatic teller machines (ATMs) are devices financed and managed by financial institutions that made available to customers a simple method for conducting financial transactions in a public space with almost no human intervention. According to estimates developed by ATMIA (ATM Industry Association) the number of ATMs worldwide for 2007 exceeded 1.6 million units [1].

Some banks tend to keep an excess of up to 40% more cash in their terminals (ATM) of what they really need. In this regard, many experts believe that excess of cash is near to 15% to 20%.

Costs related to keeping cash at an ATM represent from 35% to 60% of total maintenance costs [2]. Through improvements in administration and management of cash, banks can avoid falling into losses in new business opportunities

due to having high cash assets. This is why it is necessary to develop new methods and advanced ways of estimating the demand for money at an ATM, so that financial institutions can lower their operating costs.

On the other hand, banks and financial services assume that the demand for cash can be associated with certain variables that can have substantial effects on the level of demand for cash. Some of these variables that we must consider are the following [3]:

1. ATM Location
2. Seasonal factors such as weekends, holidays, etc.
3. Historical data from the ATM.

At present the tools and technological processes have become more complex, so it is necessary to develop methods and applications that succeed in improving these tasks. One way to address this problem adequately is to model the system dynamics using system identification [4].

System identification [5] has had great relevance in different areas of knowledge such as physics, chemistry, biology, economics, etc because dynamical systems -those in which the output value depends not only on the values of its inputs at the same moment, but also of its past values- abound in our environment and that is the reason why models are required for their analysis, forecasting, simulation, design and control. These models need to simulate the real behavior of the systems in cases when there is limited prior knowledge of its structure [4].

Our contribution consists in performing a comparative analysis of different model structures (Non Linear Regressive with Exogeneous Input, NARX and/or Non Linear Regressive Moving Average with Exogeneous Input NARMAX) using NN (Multilayer Perceptron, MLP and Radial Basis Function, RBF) and Least Square-Support Vector Machine (LS-SVM) for One-Step-Ahead (OSA) and Model Predictive Output (MPO) cash demand in ATM.

This document is configured as follows: after the Introduction, in Section 2 data processing is developed, followed by the methodology used (Section 3). In Section 4 results are shown and finally in Section 5 we will present the conclusions obtained in this work.

2 Data Processing

Data comes from the NN5 competition [6] and correspond to a set of 30 series of ATM's withdrawals used for training purposes. In addition, 11 series of the same characteristics that are called reduced dataset NN5, serve as a benchmark for comparison between the results here obtained and the general ranking of the competition.

NN5 time series competition includes a time series of cash withdrawals on a daily basis from ATMs located in different parts of England. All series show a strong cyclical component of 7 days, as well as some recurring seasonal periods such as summer holidays or Christmas. Almost all series contain empty values (missing values) and some of these series show long-term trends or irregularities

such as outliers or "gaps" [7]. The aim of the NN5 competition is to obtain the most accurate forecast possible with a horizon of 56 days for OSA using computational intelligence techniques. Each time series consists of 2 years of daily cash withdrawals at ATM [8].

All series include 3 types of "Gaps" or singularities.

- Observations equal to 0, indicating that no withdrawals have taken place due to "cash out" of the ATM.
- "Missing Values" indicating that on that day the client's transaction was not recorded.
- Outliers, indicating which data is above or below the normal behavior of withdrawals at the ATM.

This research addresses the 3 types of abnormalities, detecting outliers, missing values and values equal to 0.

To detect outliers the boxplot method by quartiles is used with $k = 1.5$ where the lowest quartile is $Q_1 = x_f$ with the f -th ordered observation. f is defined as:

$$f = \frac{\frac{n+1}{2} + 1}{2} \quad (1)$$

If f involves a fraction, Q_1 is the average of x_f and x_{f+1} . For obtaining Q_3 , the f observations are counted from the beginning, e.g $Q_3 = x_{n+1-f}$ [9]. Then an outlier is one that meets the following condition:

$$x < Q_1 - k(Q_3 - Q_1) \quad (2)$$

$$x > Q_3 + k(Q_3 - Q_1) \quad (3)$$

On the other hand, once each of the anomalous data is identified (a total of 870) it is replaced by cubic spline interpolation with a polynomial form $P(x) = ax^3 + bx^2 + cx + d$. Fig. 1. shows the amount of outlier identified by each ATM.

Fig. 1. Identification of outliers by the boxplot method. (*): Outlier data. In the X-axis 30 ATM are shown while in the Y-axis the amount earned by each ATM can be observed. All Outlier along the series were replaced by cubic spline interpolation.

3 Methodology

A total of 13 experiments for each series predicted separately (30 initial plus 11 reduced-series ATM) are performed with the following architectures: MLP-NARX [10]; MLP-NARMAX [11]; RBF-NARX; RBF-NARMAX [12] and LS-SVM-NARX [13]. Each structure is used for OSA and MPO predictions. 691 data coming from the set consisting of 30 ATM's are used for training while 100 data are used for testing (prediction purposes). On the other hand, for comparison reasons the NN5 reduced series, consisting of 11 ATM's will be used, under the rules of the competition, i.e. 735 data will be used for training and 56 for OSA predictions.

The quality indices used to measure the performance of each architecture are:

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{\frac{|\hat{y}_i| + |y_i|}{2}} * 100 \quad (4)$$

Symmetric mean absolute percentage error (SMAPE) used in the NN5 competition to determine the winner [14] and the Index of Agreement (IA) [15]:

$$IA = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (|y'_i| + |\hat{y}'_i|)^2} \quad (5)$$

with

$$y'_i = y_i - y_m$$

$$\hat{y}'_i = \hat{y}_i - y_m$$

To determine the amount of autoregressors the system needs a *Lipschitz* function is used [16]. The necessary amount of regressors is 4. Data were normalized in the range [0 1] using the following formula:

$$y = \frac{(y_{max} - y_{min}) * (x - x_{min})}{(x_{max} - x_{min}) + y_{min}} \quad (6)$$

Regarding the variables of the system they could be affected by the number of working days, the day of the week, the week of the month and other calendar effects such as festivals and religious events [17].

In NN5 series a strong daily component is present with a transactional peak on Wednesday, Thursday and Friday. On the other hand, a strong seasonal component is present depending on the week of the year.

Finally the day of the month (u1), day of week (u2), week (u3), month (u4) and a dummy variable (u5) to indicate special dates such as month-end, holidays and other calendar effects of interest are considered as inputs to the system. Given this the considered prediction function is as follows:

$$y_t = f(y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, u1_{t-1}, u2_{t-1}, u3_{t-1}, u4_{t-1}, u5_{t-1}) + \varrho_t \quad (7)$$

With the following associated predictor:

$$\hat{y}(t|\Theta) = \hat{y}(t|t-1, \Theta) = g(\varphi(t), \Theta) \quad (8)$$

$\varphi(t)$ is a vector containing the regressors, Θ is a vector containg the weights and g is the function realized by the neural network.

In the case of MPO NARMAX predictors:

$$\hat{y}_{MPO}(k) = f[\hat{y}(k-1), ., \hat{y}(k-4), u1(k-1), ., u5(k-1), 0, 0] \quad (9)$$

The presence of both 0 is because, for prediction, the future values of the error are not known [18].

On the other hand, to find the parameters of LS-SVM `tunelssvm` toolbox with simplex method and 20-folds cross validation is used, while for the MLP network 9 input neurons, 4 hidden neurons and 1 output neuron is used. Finally for designing the RBF network the NEWRB Matlab Toolbox is used with 9 input neurons and 1 output, the hidden layer is configured at runtime by the Toolbox.

4 Results

The results will be compared based on the best, worst and average results of the experiments using the two quality indices presented in Section 3.

Table 1. Results for all predictive structures and for a prediction horizon of 100 days

System		IA			SMAPE		
		Mean	Best	Worst	Mean	Best	Worst
MLP	NARX-OSA	0.87	0.93	0.70	21.37	15.42	32.50
	NARX-MPO	0.85	0.91	0.69	22.71	14.54	35.84
	NARMAX-OSA	0.87	0.92	0.67	21.56	14.93	28.99
	NARMAX-MPO	0.86	0.92	0.72	22.65	16.80	35.21
RBF	NARX-OSA	0.82	0.92	0.62	25.82	16.05	43.80
	NARX-MPO	0.82	0.92	0.61	25.50	16.05	43.80
	NARMAX-OSA	0.82	0.92	0.62	26.71	15.25	43.16
	NARMAX-MPO	0.83	0.92	0.63	25.91	14.26	51.66
LS-SVM	NARX-OSA	0.78	0.92	0.46	25.48	14.87	39.04
	NARX-MPO	0.70	0.93	0.28	29.70	14.54	53.53

The results show that MLP networks were those that performed better averaging IA=0.87 and IA=0.85 for OSA and MPO predictions respectively. On the other hand, RBF networks also have a good IA= 0.82 for both cases. LS-SVM had the worst results with IA=0.78 (OSA) and IA=0.70 (MPO). Regarding the

index SMAPE, MLP again has the better results with 22%, followed by RBF with 26% and LS-SVM with approximately 27%.

These results also show that the difference between NARX and NARMAX models is not significant. In many cases they get the same value. It also shows the great capacity of the MLP to perform MPO predictions (horizon=100 days) with an IA=0.86.

Figure 2 shows MLP performance NARX-OSA and NARX-MPO for ATM 22.

Fig. 2. (a) MLP-NARX-OSA and (b) NARX-MPO predictions for the ATM 22. The quality indexes for OSA are IA = 0.9348 and SMAPE 12.53%. For MPO predictions, IA = 0.9241 and SMAPE = 14.93%.

Table 2 show our results and ranking for reduced dataset NN5.

Table 2. Results obtained with the reduced set NN5

Model		Ranking		
		SMAPE	General	NN & CI methods
MLP	NARX-OSA	19.80%	3°	2°
	NARMAX-OSA	20.55%	4°	3°
RBF	NARX-OSA	25.63%	22°	14°
	NARMAX-OSA	26.03%	23°	15°
LS-SVM	NARX-OSA	24.35%	18°	12°

According to the results shown in Table 2, the MLP system, NARX-OSA would have won in the 2nd place the NN5 competition and would have obtained the 3rd place in the overall ranking of all methods.

5 Conclusions

The results show that MLP-NARX model developed here allows to make predictions with a high degree of quality over 85% for long-term predictions. On the

other hand, the NARMAX model shows no significant advantage over NARX, therefore NARX is better, aiming to simplicity considerations. Regarding the RBF networks they are also considered as a reliable alternative to tackle these problems, while LS-SVM has poorer results. In the future, new prediction methods will be studied about LS-SVM to improve the obtained results.

Acknowledgement. The authors like to acknowledge partial funding of this work by FONDECYT project 1090316.

References

1. Simutis, R., DiliJonas, D., Bastina, L.: Cash demand forecasting for ATM using neural Networks and support vector regresin algorithms. In: 20th EURO Mini Conference, Vilnius, pp. 416–421 (2008)
2. Simutis, R., DiliJonas, D., Bastina, L., Friman, J., Drobinov, P.: Optimization of Cash Management for ATM Network. Information Technology and Control 36(1A) (2007) ISSN 1392 124x
3. Justine, P.: ATMs and Cash Demand Forecasting: A Study of Two Commercial Bank. Nagoya University of Commerce and Business, 1-20-1, Nishiki, Naka-Ku, Nagoya City, JAPAN- 460 0003 (2007)
4. Velasquez, E.R.: Redes neuronales para sistemas dinmicos tipo NOE y Narmax. Tesis de magster, Universidad de Santiago de Chile. Departamento de Informtica (2007)
5. Ljung, L.: System Identification: Theory for the User. Prentice Hall, Englewood Cliffs (1987)
6. NN5, Forecasting Competition for Artificial Neural Networks & Computational Intelligence (2008), <http://www.neural-forecastingcompetition.com/NN5>
7. Jorg, D., Wichard: Forecasting the NN5 time series with hybrid models. Leibniz-Institut für Molekulare Pharmakologie (FMP), Molecular Modeling Group, Robert-Roessle-Str. 10, 13125, Berlin, Germany (2010)
8. Teddy, S.D.: Forecasting ATM Cash Demand using a local learning model of a cerebellar associative memory network. Data Mining Department, Institute for Infocomm Research Singapore 138632, Singapore (2010)
9. Tukey, J.W.: Exploratory Data Analysis. Addison Wesley (1977)
10. Bouchachia, A., Bouchachia, S.: Ensemble Learning for Time Series Prediction. In: First International Workshop on Nonlinear Dynamics Synchronization, Austria, INDS (2008)
11. Gao, Y., Joo Er, M.: NARMAX time series model prediction: feedfoward and recurrent fuzzy neural network approaches. Fuzzy Sets and Systems 150, 331–350 (2005)
12. Zemouri, R., Racocanu, D., Zerhouni, N.: Recurrent radial basis function network for time-series prediction. In: Engineering Applications of Artificial Intelligence, pp. 453–463 (2003)
13. Espinoza, M., Falck, T., Suykens, J., De Moor, B.: Time Series Prediction using LS-SVMs. Kasteelpark Arenberg 10, B-3001 Leuven, Heverlee, Belgium (2008)
14. Garcia-Pedrero, A., Gomez-Gil, P.: Time Series Forecasting using Recurrent Neural Networks and Wavelet Reconstructed Signals. In: 20th International Conference on Electronics, Communications and Computer (CONIELECOMP) (2010)

15. Rode, M., Suhr, U.: Multi-Objective Calibration of a River Water Quality Model For The Elbe River, Germany, UFZ Centre for Environmental Research Leipzig Halle, Department of Hydrological Modelling, Brckstrae 3a, D-39114 Magdeburg, Germany
16. Norgaard, M.: Neural Networks for Modelling and Control of Dynamic Systems: A Practitioner's Handbook (Advanced Textbooks in Control and Signal Processing). Springer, Heidelberg (2006)
17. Soares, P., Rodriguez, P.: Calendar Effects in Daily ATM Withdrawals. Economics and Research Department. Banco de Portugal (2010)
18. Chen, P.-H., Lee, Y.-W., Chang, T.-L.: Predicting thermal instability in a closed loop pulsating heat pipe system. Applied Thermal Engineering (2008)