Processing Time Series Data by Means of Forecasting Models

Zita Bošnjak

Faculty of Economics Subotica, Segedinski put 9-11, 24000 Subotica, Serbia and Montenegro, bzita@eccf.su.ac.yu

Miodrag Stojković

Public Transport Company "Suboticatrans", 24000 Subotica, Serbia and Montenegro, jumper@suonline.net

Saša Bošnjak

Faculty of Economics Subotica, Segedinski put 9-11, 24000 Subotica, Serbia and Montenegro, bsale@eccf.su.ac.yu

Abstract: The goal of our research was to find an appropriate model to predict the number of passengers in public transport in order to transform the vast amount of raw data into information adequate for supporting the decision maker in resources planning and allocation of buses to existing bus lines. For this purpose, we applied and compared two different approaches - one linear and one non-linear. As a linear model we used the multiple linear regression model, while the feed forward back-propagation neural network served as a non-linear model. The rational behind selecting these two models, the development process as well as the forecasting results are described in this paper.

Keywords: time-series analysis, neural networks, data mining

1 Introduction

There are many reasons for the tremendous rise of artificial intelligence applications in contemporary business. One is the fact that the amount of collected data is doubled every two years resulting in less than 5% of them being ever processed, and intelligent systems can help to reduce this data overload to manageable amount. Furthermore, over 60% of the time and efforts spent on data processing is directed to discover valuable relations and patterns, less than 20% of time is used up for the interpretation of discovered patterns, and even less than 10% of time is spent for gaining competitive advantage and benefits thanks to discovered patterns[1]. Such a situation results in decrease of decision making ability and can be improved by means of intelligent systems.

Dr. Yogesh Malhotra, the founder of the @BRINT knowledge base[2], stated that humans rely on computer technology in problem solving, but are not aware of the need for synergy of human and technological resources in exploiting the available amounts of data in order to develop useful products and services. Jeremy D. Jones in [15] also pointed out the human tendency of collecting data in organizations and services, but at the same time he emphasized that data are worthless unless transformed into information and used for achievement of defined goals.

Data mining as the process of employing one or more computer learning techniques to automatically analyze and extract knowledge from data contained within a database is the right concept to turn to. Data mining is a term used to describe knowledge discovery in databases and includes tasks known as knowledge extraction, data exploration, data pattern processing, and information harvesting. All these activities are conducted automatically and allow quick discovery even by nonprogrammers.

Data mining method is an algorithm designed to analyze data, or to extract patterns from data. Most data mining methods are based on concepts from machine learning, pattern recognition and statistics. The array of different data mining algorithms can often be quite bewildering to both the experienced data analyst and the novice.

The underlying idea of our work, described in this paper, was to analyse the travelling services data by means of available data mining methods and techniques in order to reveal some new relations or patterns in the behaviour of passengers. Such findings could result in more accurate prediction of passengers' number and decrease the risk and uncertainty in the decision making process.

2 Data Mining Technique Selection

Data mining is a confluence of disciplines and consequently it uses a full spectrum of techniques. Furthermore, depending on the kinds of data to be mined or on the given data mining application, the data mining system may also integrate techniques from respective fields. Collier et al. in [7] presented a framework for evaluating data mining software tools in which the first criteria among functionality criteria is the so called Algorithmic Variety: Does the software

[1] Data are reported by the Gartner Group and can be found on Internet

[2] www.brint.com

provide an adequate variety of mining techniques and algorithms including neural networks, rule induction, decision trees, clustering, etc.? Hence, the main problem is that there are still no established criteria for deciding which data mining methods and techniques to use in which circumstances. Kwang in [8] described a prototype expert system which addresses the need to help the user select an appropriate statistical test for a research problems experienced by statistically naïve medical researchers. Roiger et al. in [9] gave some guidelines for choosing a data mining technique in a form of partial goal tree. However, selection of data mining method is not simply a matter of selecting the best method for all purposes. Instead, the practicing professional must consider methods with respect to their particular environment, i.e. features of the data collection, and analysis needs - the devised data mining goal. There are some approaches based on heuristic approximations that help potential users to distinguish data mining systems and identify those that best match their needs. In [10] one expert system based on heuristics for evaluation and selection of the most appropriate data mining method(s) for a given problem is described. We used this expert system to obtain an initial recommendation which data mining method is most adequate for the prediction of the number of sold bus tickets in public transport. The data, subject to our analysis, were collected from the Public Transport Company “Suboticatrans" in Subotica and the Hydro-meteorological Service of the Republic of Serbia and originated from the period from January 1st 2000 to December 31st 2001, resulting in 730 daily observations. The output of the expert system is shown on Figure 1. Following the above listed recommendations, we selected neural networks, as a non-linear model and a multiple linear regression model, as supervised learning strategies, for data mining purposes.

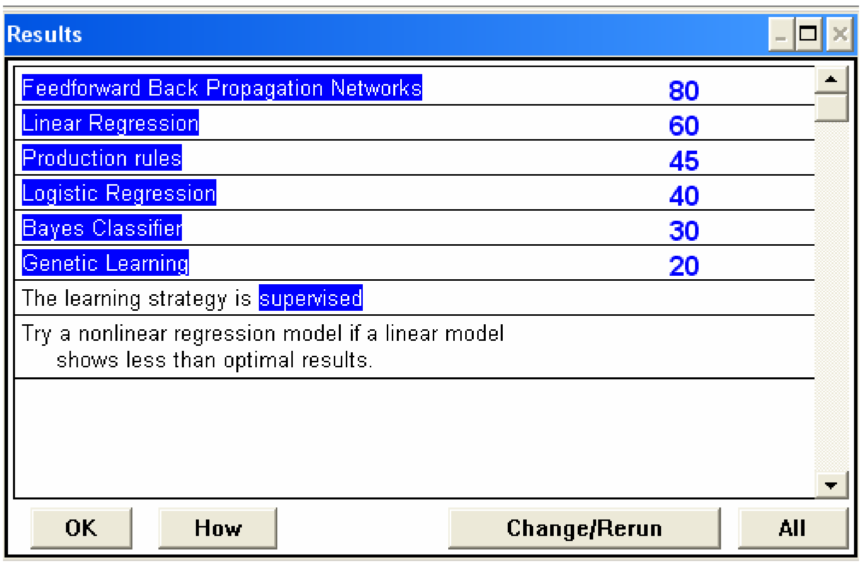

Figure 1
Results of the expert system run

The heuristics on which the knowledge base of the expert system is founded, are listed in Appendix 1.

3 Neural Networks for Time Series Analysis in Public Transportation

Many authors investigated the possibilities and advantages of neural network application in prediction and forecasting problems ([5], [11], [12], [13]). Despite the wide range of problems successfully solved by neural network models, quite few authors applied this methodology in the transportation domain. Himanen et al. [9] described one unsupervised neural network model developed with aim to explain daily traveling by personal and household attributes using data from national travel surveys. Compared to mathematical models such as regression-, factor- and cluster-analysis, the Kohonen self organizing map was significantly better modeling tool.

Our goal was to model the number of sold bus tickets in public transport by neural network trained in the supervised training process. The behavior of passengers in public transport in urban areas is characterized by the degree of their mobility, traffic demands, travelling distances, the available bus lines and the number of passengers, as an indicator of (dis)balance between provided services and demands on the market [6]. If known, these parameters may contribute to efficient and correct decision making at all levels of management in transport organizations. Oscillations in the number of passengers known in advance can easily be incorporated into resource allocation plan of a public transport company [6], thus the problem lays in unpredictable behaviour of people that use the transport services quite often, but make a decision about travelling by bus *ad hoc* and this way they contribute to the increased demand for transportation. Consequently, the task is to predict the number of “irregular” users, i.e. to analyze the number of singular bus tickets sold. We divided the number of sold bus tickets into two main categories: tickets sold by full price and tickets sold by reduced price to 50% (available for elderly people, pensioners, and children from 6 to 10 years).

3.1 Data Preparation for the Analysis

The first step in data analysis was the data pre-processing in a sense of detecting and correcting or eliminating erroneous data. Since 3.5% of meteorological data was missing due to the war conditions, the data were gathered from the other closest meteorological stations in the area, and new, interpolated data on whether condition in Subotica in the time period under investigation were calculated instead of the missing ones.

The next activity was the cleansing of data from redundancies, i.e. detecting data with strong correlation to each other. After the correlation analysis, some of the input variables were discarded from further analysis, while some other variables were grouped together in order to get a new, cumulative variable. Consequently, the final input variables were: *Type of a day, Price of gasoline/Price of tickets*, *Earnings/Expenditures*, *Air pressure*, *The lowest night temperature*, *Air moisture*, *Direction of the wind*, *Speed of the wind, Cloudiness, General weather condition, Rainfalls, Condition of the ground*, and *Height of snow.*

Furthermore, records collected on some special occasions, such as the May 1st, November 1st, etc. containing outliers, were removed from the data set, as they always show extremely high values due to specific habits and behaviour of local inhabitants, which could "confuse" the neural network during the training phase.

3.2 Neural Network Architecture

For the initial architecture of the neural network we selected the following physical configuration developed in DataEngine [14] neural network software tool:

(1) The number of nodes in the output layer is two as the neural network should model the behaviour of two output variables: the number of sold tickets by full price and by the reduced price;

(2) The number of nodes in the input layer is determined by the number of input variables and equals 13. The number of hidden layers was set to one;

(3) The number of neurons in the hidden layer was decreased from the initial 25, with a step of 5, to 5 neurons in the hidden layer, resulting in five different physical architectures;

(4) As a discriminator we selected the standard propagation function:

$$I_i = \sum_j w_{ij} x_j$$

that calculates the weighted sum I of all signals which are active at the input connections. The w_{ij} components of this equation represent the connecting weights between neuron j and neuron i, and the x_j components represent the signals at the connection concerned;

(5) At the beginning of training we set the transfer function of the input and the output layer to linear function, while the activation function of neurons in the hidden layer was set to hyperbolic tangens and afterwards it was changed to parabolic and sigmoid functions;

(6) The standard error back propagation algorithm and the modified Delta rule were used in the model. The amount of data in the training set compared to

the amount of data in the test set was 80:20. We chose the random presentation of samples from the training and the test data sets to the neural network. The desired accuracy we were looking for was set to 90%. This value was calculated by RMS error and the maximal error during training, by formula:

$$\varepsilon = 0{,}9\varepsilon_k + 0{,}1\varepsilon_m \tag{1}$$

where ε is the calculated error that shouldn't exceed 10%, ε_k stands for RMS error and ε_m for maximal error;

(7) The values of learning parameters μ (momentum) $\mu \in [0.0, 0.9]$ and α (learning rate) $\alpha \in [0.1, 0.9]$, were varied during the training phase with a step of 0.1.

3.3 Modelling Results

Five neural network architectures that achieved the best test results under the evaluation criteria are shown in Table 1. As can be seen, the best predictive performance was achieved by the network that had 5 neurons in the hidden layer, and the transfer functions of neurons in the hidden and the output layer were set to sigmoid functions. Further improvement and the final neural network architecture were gained after removing individual connections that were under the relevance threshold, set to 0.02, in the I ranked architecture. As a result of this change, the maximal training error of the network was reduced to 0.4065, the training RMS error was 0.0836, while the maximal testing error of the network was 0.4070 and the test RMS error was 0.1097.

Layers of neurons		Learning parameters		RMS	Epoch	Rank
Hidden	Output					
No. of neurons		Learning Rate	Momentum			
5	2	0.4	0.7	0.1132	780	I
10	2	0.8	0.4	0.1137	530	II
10	2	0.6	0.5	0.1148	620	III
10	2	0.1	0.0	0.1152	4660	IV
20	2	0.1	0.2	0.1155	2580	V

Table 1

Structure, learning parameters, and RMS error for five "best" neural network architectures

4 The Multiple Regression Model

As an alternative to the nonlinear neural network model for predicting the number of passengers, we applied, on the same data set, the multiple linear regression methodology. Firstly we tried to predict the number of passengers who pay the full price of a bus ticket, and secondly we repeated the process for forecasting the number of sold tickets by reduced price to 50% of its value. After the estimation of regression parameters on the training sample and the application of the obtained regression function to the test sample, we computed the RMS error for both predicted variables. The results are summarised in table 2.

	Number of sold bus tickets	
	by full price (100%)	by reduced price (50%)
RMS error	0,222630421	0,131068527
Significance	$6{,}28 > F_{0{,}05}(13{,}129) = 1{,}74918$ $6{,}28 > F_{0{,}01}(13{,}129) = 2{,}18846$	$1{,}91 > F_{0{,}05}(13{,}129) = 1{,}74918$ $1{,}91 < F_{0{,}01}(13{,}129) = 2{,}18846$

Table 2

Prediction accuracy of multiple linear regression models

Conclusions

In [11] the error rate of 1% is mentioned as the desirable prediction accuracy for neural networks, while in [12] it is stated that in some financial applications even error rates up to 30% might result in profit. Our experience gained during the case study implies that the ideal error of 1% is hard to achieve. The 10% error, that we managed to achieve, is in most practical applications quite acceptable, having it in mind when obtained results are utilised. When compared to linear regression modelling, the final neural network model was still more successful, having smaller prediction error. Since the RMS error of the multiply linear regression is worse than the RMS obtained with nonlinear MLP (RMS=0.1132), for both the prediction of the number of sold bus tickets by full and by reduced price, we may conclude that linear regression is not the best methodology for the problem in hand and suggest neural network method as more adequate for forecasting the number of passengers in public transport.

References

[1] Plummer E. A.: Time Series Forecasting with Feed Forward Neural Networks: Guidelines and Limitations, thesis submitted for the degree of MASTER OF SCIENCE in COMPUTER SCIENCE, Laramie, Wyoming, 2000

[2] Iwamya H.; Kermanashahi B.: Sensitivity Analysis and Artificial Neural Network used for Long-term Forecasting of 9 Japanese Power Utilities,

Department of Electronics & Information Engineering Tokyo University of Agriculture and Technology, Tokyo, 2000

[3] Kutsurelis J. E.: Forecasting Financial Markets Using Neural Network: An Analysis of Methods and Accuracy, Naval Postgraduate School, Monterey, California, 1998

[4] Jiawei H.; Kamber M.: Data Mining Concepts and Techniques, Morgan Kaufman Publishers, San Francisco, 2001

[5] Hill T.; Marquez L.; O'Connor M.; Remus W.: Artificial Neural Network Models for Forecasting and Decision Making, Int. Journal of Forecasting, Vol. 10, Issue 1, 1994, pp. 5-15

[6] Himanen V.; Nijkamp P.; Reggiani A.; Raitio J.: Neural Networks in Transport Application, Ashgate Publishing LimitedGower House, Hampshire, England, 1998

[7] Collier K., Carey B., Sautter D., Marjaniemi K. (2004) A Methodology for Evaluating and Selecting Data Mining Software, Center for Data Insight, Northern Arizona University

[8] Kwang B. Lee: A Design and Development of Expert Systems for Selecting an Efficient Data Mining Method Using Visual Basic, Texas A&M University, 2004

[9] Roiger, R. J., Geatz M. W.: Data Mining – A Tutorial Based Primer, Addison Wesley, Pearson Education Inc., 2003

[10] Bošnjak, Z., Bošnjak, S.: Expert System Support in Data Mining Method Selection, 20th European Conference on Operations Research, Volume of Abstracts, 2004, pp. 202

[11] Bigus J. P. : Data Mining with Neural Networks, McGraw-Hill, 1996

[12] Podding T.: Modelling and Forecasting Integrated Financial Markets using Artificial Neural Networks, Proceedings of EUFIT'96, 1996, pp. 2177-2180

[13] Torres R., Macado M. A., Souza R.: Failure Time Series Prediction in Industrial Maintenance Using neural Networks, IFORS 2005, Book of Abstracts, 2005, pp. 139

[14] DataEngine –Tutorials and Theory MIT GmbH, Aachen, Germany, 1997

[15] Jones D. J.: Artificial Intelligence and Modern Business Practices, Bauer College of Business, University of Houston, 2001

APPENDIX 1: Data Mining Method Selection Heuristics

The heuristics the expert system for data mining method selection captures in the knowledge base are devised as follows:

[H1] If the learning strategy is supervised then decision trees, production rules, feed forward back propagation networks, linear regression, logistic regression and Bayes classifier are potentially useful data mining techniques.

[H2] If a primary goal of a data analysis is to determine future outcomes rather than current behaviour, and the data we wish to analyse doesn't contain a time dimension then the learning strategy is supervised and the problem in hand is prediction. If the data contain a time dimension, the problem in hand is time-series analysis.

[H3] Poor choice if explanations are important, are regression models and neural networks, because they are black box structures. If the ability of a model to explain the results is desirable but not obligatory, one can use, amongst other data mining techniques, feed forward back propagation networks and linear regression for modelling.

[H4] If the learning strategy is supervised and the input values are numeric (or categorical but can be transformed into numeric values) and there is a single output attribute, and the output attribute is numeric (or categorical that can be transformed into numeric) and the values of the output variable are not restricted to the [0,1] interval range, then linear regression, feed forward back propagation networks and production rules may be applied for data analysis.

[H5] If the learning strategy is supervised and there is more than one output attribute, then production rules, genetic learning and feed forward back propagation networks are appropriate, while linear regression, logistic regression, decision trees or Bayes classifier are not an appropriate data mining technique.

[H6] If the learning strategy is supervised and we assume attributes are of equal importance, methods such as Bayes classifier, feed forward back propagation networks, genetic learning and production rules are a good choice for analysis, while methods such as logistic regression, linear regression and decision trees are quite poor choice for data analysis.

[H7] If the learning strategy is supervised and we want to promote attribute significance, we should prefer genetic learning, linear regression, logistic regression and decision trees to data mining methods such as Bayes classifier and feed forward back propagation networks.

[H8] If the learning strategy is supervised and the data distribution is not normal or is unknown, linear regression, logistic regression, Bayes classifier and feed forward back propagation networks should be avoided. We should use decision trees, production rules or genetic learning instead.

[H9] If the learning strategy is supervised and the data set has some missing values or contains wealth of noisy data, feed forward back propagation networks are highly recommended for utilisation.

[H10] If the learning strategy is supervised and in data analysis process, computational time is an issue, then decision trees, production rules, Bayes classifier, linear and logistic regression can give good results, while the same performance can't be expected from genetic learning or feed forward back propagation networks.

[H11] If the learning strategy is supervised and the problem in hand is time-series analysis, best data analysis results are obtained by linear or logistic regression, but feed forward back propagation networks can also be tried, while decision trees, genetic learning, Bayes classifier and production rules show less than optimal results.