PREDICTION MODEL FOR FINANCING OF BASIC TRANSPORT SERVICES

Miloslava Kašparová
System Engineering and Informatics Institute
Faculty of Economics and Administration, University of Pardubice

Abstract

The paper is focused on comparison of actual values and predicted values of revenues to passed kilometer of bus connections. The model is based on prediction of bus carrier's revenues by force of neural networks (NN). The revenues are part of a demonstrable loss calculation.

Key words: Basic Transport Services, Financing, Prediction Model, Neural Nets

Public Services Financing in Public Passenger Transport

The basic transport services (BTS) security is a condition of access possibility to other public services.

Public passenger road transport and railway passenger transport standards are formulated universally but this universality is an advantage. It makes possible regional authorities to determine the range of BTS according to local conditions. The time and local bus line accessibility is determined in according to local requirements. The Regional Authority is a guarantor of this BTS. The Regional Authority responsibility is to grant bus carrier demonstrable loss cover. It is caused by the obligation of the public services contract fulfillment. The special-bound subsidy is appropriated from the state budget for demonstrable loss cover. This kind of the subsidy is possible to use only for a predetermined task fulfillment.

The Road Traffic Licensing Department Function

One of region functions is to determine an extent of the BTS [8]. The Regional Authority has a function of the Road Traffic Licensing Department (RTLD) in the Czech Republic. It is in subscriber role of future provided the service by bus carriers to citizens. The bus carriers are providers of the public services.

The RTLD makes contract of public service obligation to security of the BTS. There is a relation between regional budget and budget of bus carriers which is based on the demonstrable loss. This loss is compensated by the Regional Authority. This situation is described in [1] and in [9].

Existence of the public interest to the BTS security through the public service obligations is conditioned by existence of budget resources for this use reserved.

In the time of contracts conclusions (contract part is according to [9] an estimate of the future demonstrable loss that is calculated of bus carriers) RLTD will be able:

- to predicate by force of the PM revenues of bus connections that is content of contract;
- to determine extent of the financial resources that are needed for the demonstrable loss financing;
- to evaluate if the demand of bus carrier for the grant of financial resources is in accordance with reality.

Example of Subsidized Bus Connection Financing

The example describes a sample of the revenues to passed kilometer of *i* bus connection in CZK in months September and October in the years 2003 and 2004. The Fig. 1 shows forty values of the actual revenues to passed kilometer of *i* bus connection in CZK in months September and the losses to passed kilometer of *i* bus connection in CZK in September of years 2003 and 2004. The Fig. 2 shows values in October of the years 2003 and 2004. The total values are in Tab. 1 and Tab. 2.

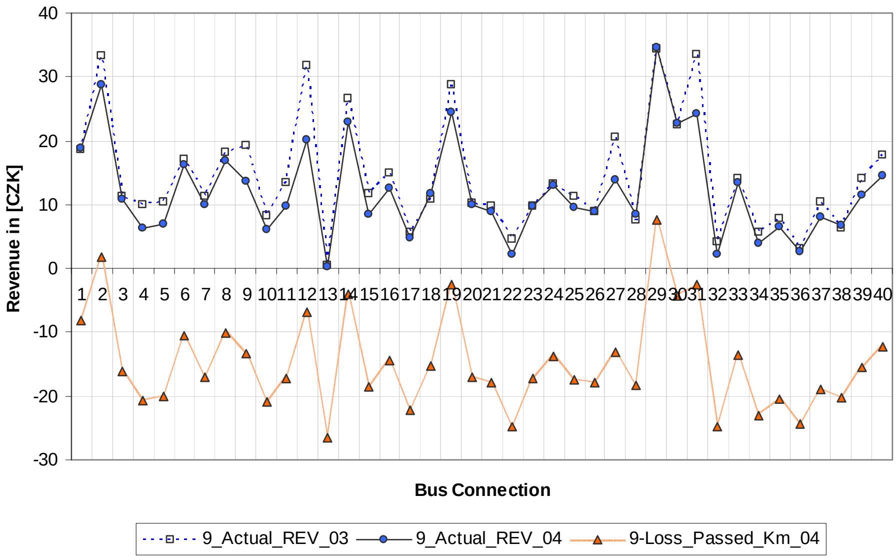

Fig. 1: Values of the revenues to passed kilometer and the losses to passed kilometer of bus connections in September in the years 2003 and 2004

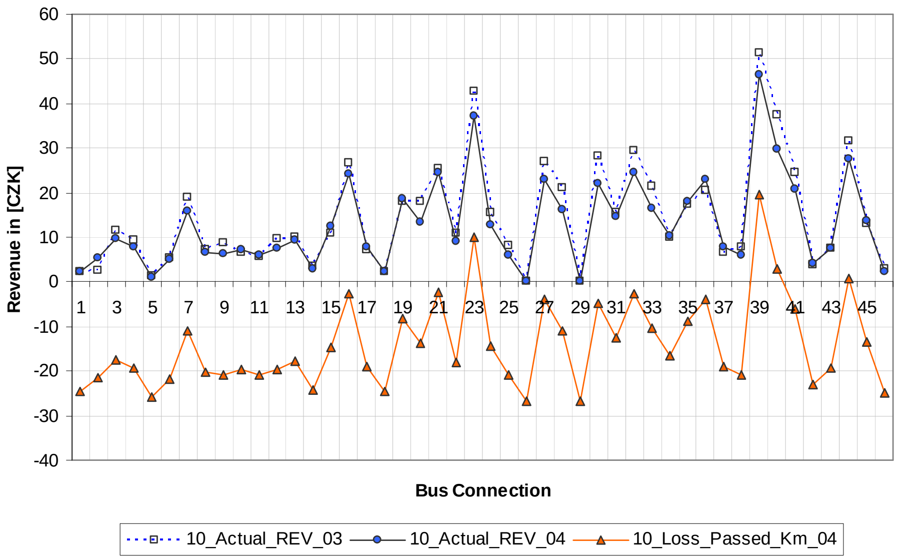

Fig. 2: Values of the revenues to the passed kilometer and the losses to the passed kilometer of bus connections in October in the years 2003 and 2004

The all bus connections are unprofitable in Fig. 1 and Fig. 2. The RLTD must to subsidy the bus connections that serve to BTS security. The subsidy size to passed kilometer is 26.00

in CZK in months September and October in the year 2003. In the year 2004 in these months it is CZK 26.95. The RLTD determines the subsidy size to passed kilometer.

Prediction Model for Financing of Basic Transport Services

Phases of the goal realization are represented in the Fig. 3. The first phase is a problem specification. The second phase is a data preprocessing and a data collection. The output of this phase is the data matrix that will be used in the phase of the PM creation. In the phase of PM creation are used feed-forward neural networks. The PM is created in software Clementine 7.0[1]. The next phase is the results evaluation and the final is results application in given public administration level.

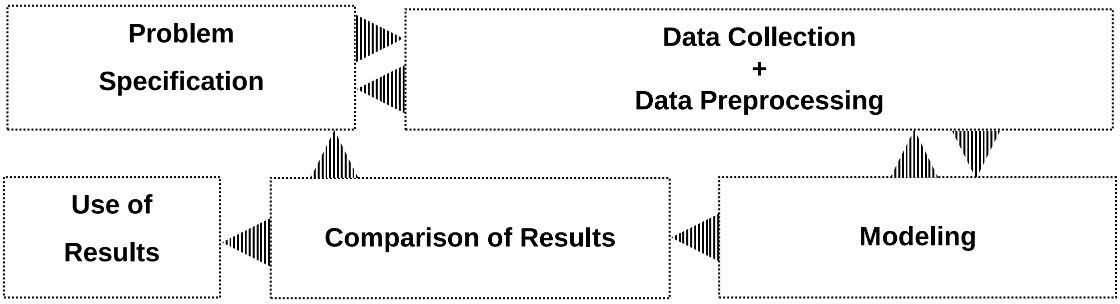

Fig. 3: Phases of prediction model creation

Feed-forward Neural Networks

The basic element of a NN is a neuron. This is a simple virtual device which accepts many inputs, sums them, applies a transfer function, and outputs the result, either as a model prediction or as input to other neurons. A NN is a structure of many such neurons, connected in a systematic way. The NNs are described for example in [2, 5, 6 ,7].

In PM in Clementine are used feed-forward NN [7]. The neurons in such networks are arranged in layers. An example of three-layer NN structure is in Fig. 4. There are typically three parts in a NN: an input layer with units representing the input fields, one or more hidden layers, and an output layer with a unit or units representing the output fields. The units are connected with varying connection strengths or weights. The network learns by examining individual records, generating a prediction for each record, and making adjustment to the weights whenever it makes an incorrect prediction. This process is repeated many times, and the NN continues to improve its predictions until one or more of the stopping criteria have been met [7].

The training of a feed-forward NN uses a method called Backpropagation of error. For each record presented to the NN during training, information (in the form of input fields) feeds forward through the NN to generate a prediction from the output layer. This prediction is compared to the recorded output value for the training record, and the difference between the predicted and actual outputs is propagated backward through the NN to adjust the connection weights to improve the prediction for similar patterns [7]. For training in Clementine are used following methods: Quick, Dynamic, Multiple, Prune and Exhaustive Prune. The methods are described in [7].

---

[1] Clementine is an enterprise data mining workbench of SPSS Inc. that enables to quickly develop predictive models using expertise and deploy them into operations to improve decision making. It supports all steps of standard methodology CRISP-DM (Cross-Industry Standard Process for Data Mining).

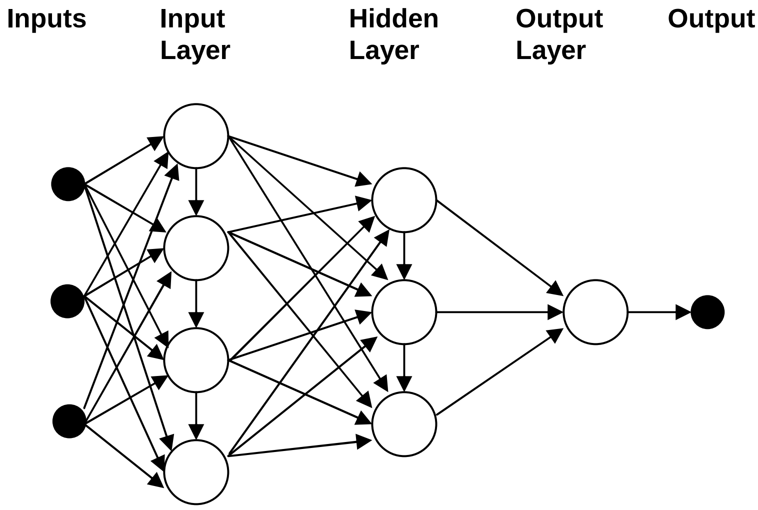

Fig. 4: The three-layer NN structure

Result of Prediction

The data matrix contents 1 dependent input variable *REV* r_{ij} (revenue to the passed kilometer of *i* bus connection in CZK in year 2003) and other 33 input variables that are independent. There are: *DEP* a_{ij} (a departure time of bus connection from starting bus stop); *MON* f_{ij} (a month when bus carrier carries on transport on bus connection and which they raised revenues); *KM* e_{ij} (a total length *i* bus connection in kilometers); *CMUnk* kn_{ij} (a municipality *n* that was assigned to cluster *k*) and *ANKMn* pn_{ij} (an average number of kilometers in municipality *n* that was assigned to cluster *k*; it follows from number of bus stops in municipality and number of bus connection kilometers passed in municipality *n*). The output of PM is predicted month revenues to passed kilometer of *i* bus connection of *j* bus line $\hat{r}_{ij}$ where $i \in <1; 67>$ and $j \in <1; 105>$. The input variables and the PM are described in [3].

The data matrix contents 5 389 objects (bus connection s_{ij} that are described with variables above). It is divided into training set and testing set. From training set it is chosen a part of date for validation. According to [6] a possible size of validation set is in range from 10 % to 50 %. The size of testing set is 1796 objects. It is one third of objects from the total size of the data matrix.

Many tests were realized in the PM. The methods above were used for the NN training. The tests differ in setting stop criterions (time 2, 5 and 10 minutes for NN training) and size of validation sets (15 % - 40 %). Model effectiveness is expressed by mean absolute error (MAE) [4, 5]. The results of all predictions are described in [3]. The MAE of the best model is 5.057 in CZK.

Comparison of Results

The comparison of actual and predicted revenues to the passed kilometer of *i* bus connection in CZK is showed on forty randomly chosen bus connections. The actual revenues in September and October of year 2004 were not used for training, validation and testing in PM. It is used only for comparison of predicted and actual revenues.

The actual revenues and the predicted revenues to the passed kilometer and losses to passed kilometer in September are in Tab. 1. In September in the year 2004 the average loss to the passed kilometer in this sample is CZK 14.83 and the total loss of all forty bus connections is CZK 270 495.78.

Tab. 1: Sample of forty bus connections – in September of the years 2003 and 2004

	Total Revenue in [CZK]
9_Actual_REV_03	571.18
9_Predicted_REV_03	489.53
9_Actual_REV_04	484.72

In Fig. 5 are showed values of actual revenues and predicted revenues to the passed kilometer of *i* bus connections in September of years 2003 and 2004.

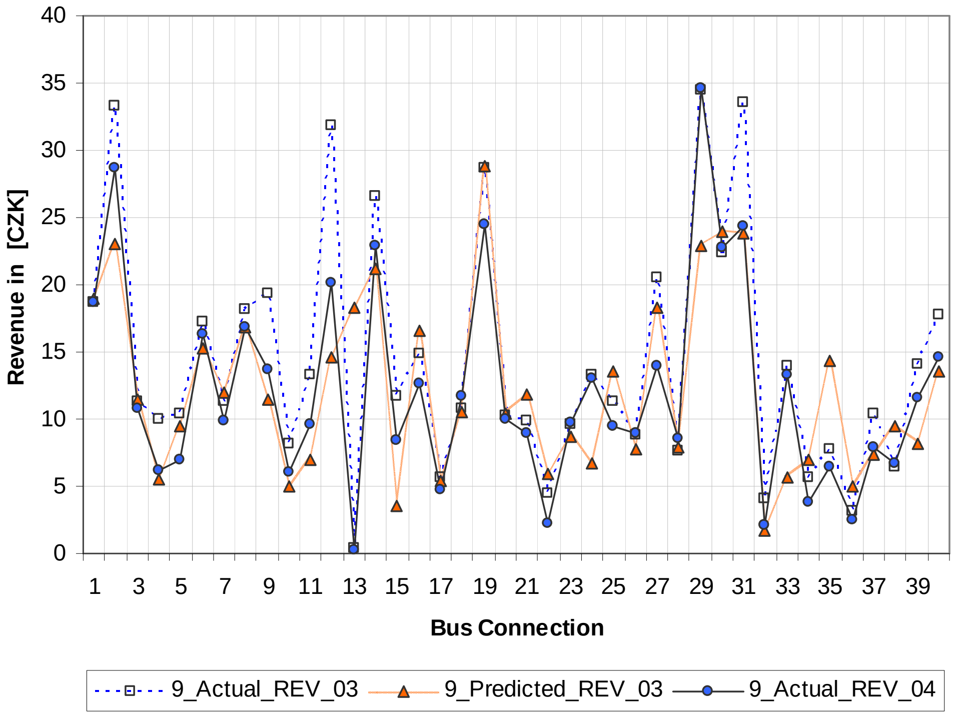

Fig. 5: The actual revenues and predicted revenues of forty bus connections in September of the years 2003 and 2004

The values of actual and predicted revenues to the passed kilometer and losses to the passed kilometer in October are in Tab. 2. In October in the year 2004 the average loss to the passed kilometer in this sample is CZK 14.01 and the total loss of all forty bus connections is CZK 198 110.07.

Tab. 2: Sample of forty bus connections – in October of the years 2003 and 2004

	Total Revenue in [CZK]
10_Actual_REV_03	665.66
10_Predicted_REV_03	601.43
10_Actual_REV_04	595.36

In Fig. 6 are showed values of actual revenues and predicted revenues to the passed kilometer of *i* bus connections in October of years 2003 and 2004.

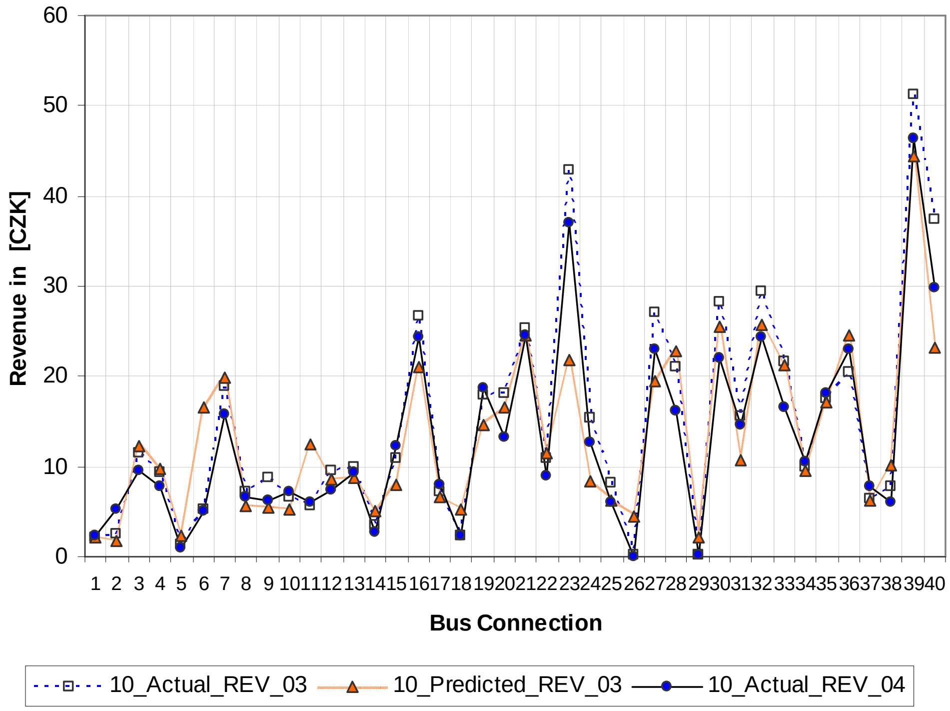

Fig. 6: The actual revenues and predicted revenues of forty bus connections in October of the years 2003 and 2004

Conclusions

By force of PM that is based on feed-forward NN it was achieved the MAE 5.057. The result comparison was made on the randomly chosen sample of forty bus connections in months September and October of the years 2003 and 2004. The total sums of predicted revenues and actual revenues are in Tab. 1. and Tab. 2. The difference between the predicted revenues and actual revenues in September of the year 2004 is CZK 4.81. In October the difference is CZK 6.07. In September of the year 2003 the difference between the predicted revenues and actual revenues is CZK 82.27. In October is CZK 64.23. In the sample of bus connections the prediction approximates to the actual revenues in September and in October of year 2004.

Literature:

[1] BARTONÍČKOVÁ, M. - BARTONÍČEK, T. Modelování tržeb ve veřejné linkové autobusové dopravě. In *Scientific Papers of University Pardubice*, No. 9, 2004, s. 6-13.

[2] BERTHOLD, M. - HAND, D. J. *Intelligent Data Analysis: An Introduction*. Berlin: Springer, 2003, ISBN 3-540-43060-1.

[3] KAŠPAROVÁ, M. Principles of Effectiveness and Transparency in Public Administration by Financing of Basic Transport Services. In: *Proceedings of the 5th WSEAS International Conference on Environment, Ecosystems and Development* [CD-ROM]. Venezia, 2. - 4. 11. 2005 (in press).

[4] MAŘÍK, V. et al.*Umělá inteligence (4)*. Praha: Academia Praha, 2003.

[5] NOVÁK, M. et al. *Umělé neuronové sítě, teorie a aplikace*. Praha: C. H. Beck, Praha, 1998, 382 s. ISBN 80 -7179-132-6.

[6] RUSELL, S. J. - NORVIG, P. *Artificial Intelligence: A Modern Approach*. New Jersey: Prentice Hall New Jersey, 2003. ISBN 0-13-080302-2.

[7] SPSS Inc. *Clementine® 7.0 User's Guide*. 2002, 741s., ISBN 1-56827-295-2.

[8] *Zákon č. 129/2000 Sb., o krajích (krajské zřízení).*

[9] *Zákon č. 111/1994 Sb., o silniční dopravě.*

Contact Address:

Ing. Miloslava Kašparová, Ph.D.
System Engineering and Informatics Institute
Faculty of Economics and Administration
University of Pardubice
Studentská 82, 532 10 Pardubice
e-mail: miloslava.kasparova@upce.cz
phone: +420 46 603 6245