Traffic speed prediction method for urban networks - an ANN approach

Alfréd Csikós, Zsolt János Viharos, Krisztián Balázs Kis

Systems and Control Laboratory
Institute for Computer Sciences and Control, Hungarian
Academy of Sciences
Budapest, Hungary
{csikos, viharos.zsolt, kis.krisztian}@sztaki.mta.hu

Tamás Tettamanti, István Varga

Budapest, Hungary Dept. of Control for Transportation
and Vehicle Systems
Budapest University of Technology and Economics
Budapest, Hungary
{tettamanti, ivarga}@mail.bme.hu

Abstract—**The paper proposes a traffic speed prediction algorithm for urban road traffic networks. The motivation of the prediction is to provide short time forecast in order to support ITS (Intelligent Transport System) functionalities, such as traveler information systems, route guidance (navigation) systems, as well as adaptive traffic control systems. A potential and efficient solution to this problem is the application of a soft computing method. Namely, an artificial neural network (ANN) is used for the forecast by involving the measured speed patterns. The ANN is trained by using data produced by Vissim (a microscopic road traffic simulator) simulations. The proposed algorithm is developed and analyzed on a real-word test network (part of downtown in Budapest).**

Keywords—traffic speed; prediction; artificial neural network

I. INTRODUCTION

A. Motivation

Traffic prediction methods designed both for motorway and urban environment has been investigated by several papers in the past decades. The aim of these methods is to estimate short term traffic parameters such as traffic flow rate, travel time, volume, occupancy or the probability of congestion. These parameters are essential in the design of ITS applications. A brief overview of the latest traffic forecasting research and future challenges are given by [1].

B. Literature review of the relevant state-of-the-art methods

Traffic prediction methods can be divided into two major classes. *Classical prediction methods* are based on the traffic flow theory and use statistical methods for predicting traffic parameters [2]. Furthermore, several methods have been applied for the prediction of traffic states: Bayesian network models [3], History Average (HA) models [4], Autoregressive Integrated Moving Average (ARIMA) models [5]-[7], non-parametric regressions [8], [9] are most commonly used for prediction, as well as procedures based on Kalman Filter theory [10], [11]. These prediction methods evolve their forecast based on historical data time series analysis. Thus,

reliable prediction cannot be achieved in urban environment when the traffic conditions are changing rapidly.

For this reason, with the evolution of computational intelligence, *data-driven methods* have gained attention, offering self-learning pattern recognition methods instead of a model-based estimation. These methods are based on Artificial Neural Network (ANN) models [12]-[15], fuzzy-rule based logics [16], Support Vector Machines (SVM) [17], k-means clustering [18][19], and expectation maximization based algorithm [20]. The main advantage of the machine learning methods is that they can estimate and capture the linkage of very complex traffic flows even under rapidly changing conditions. Note, that these data driven methods also have drawbacks, for example ANN and SVM are sensitive to the training data quality. A part of these problems can be addressed by the so-called principle component analysis to handle missing input data [21][22]. Hybrid models combine statistical methods with computational intelligence to reduce the weakness of the basic method and give better prediction performance [23].

Dimitriou et al. used an adaptive hybrid fuzzy rule-based system for modeling and predicting urban traffic flow [24]. They used Genetic Algorithm for tuning the parameters of the membership functions and successfully applied the model for short-term traffic forecasts. Another application compared the multinomial logit model with artificial neural networks for detecting road traffic status using cellphone handover information [25]. Both of the models outperformed the city-wide time-of-day traffic profile which is the current standard estimation method.

By using artificial neural networks, Lint [26] has investigated the travel time estimation problem in detail. However, the results are only valid for freeway or arterial urban network. Gastaldi et al. presented a combined ANN-Fuzzy method for estimating the annual average daily traffic based on one-week traffic counts [27]. They found that this method can yield satisfactory accuracy with low computational cost compared to the current standards. Traffic volume forecasting was the goal of the authors in [28]. They used

radial basis function neural network for estimation and also considered the traffic flow of the adjacent intersections. Their results showed that using the extra information gathered from the adjacent intersections can increase the accuracy significantly.

In [29] a camera based image processing method is introduced for congestion level recognition, however, forecasting is not incorporated in the study. As another example, [23] is also mentioned which presents a hybrid model for traffic volume forecasting based on inductive loop detector data. A missing issue in this research field is the traffic speed forecast in urban area, i.e. the prediction of congestions in signalized traffic network.

Accordingly, present paper introduces a congestion prediction algorithm designed for urban road traffic network. The main novelty of the technique is a congestion forecasting pattern recognition, based on the ANN method. The input of the model is calculated from mean speed data on links of the road network during 5 minutes sampling intervals. By using real-world traffic data and Vissim microscopic traffic simulator, the proposed algorithm is trained, tested, and analyzed (applying a test network: the neighborhood of Oktogon square in Budapest, Hungary).

II. DATA GENERATION METHODOLOGY

The efficiency of a pattern recognition method can be considerably enhanced by preparing and using the appropriate data. During the construction of the dataset, three main aspects were addressed:

- realistic patterns are needed, but recurring patterns should be excluded to avoid 'over-fitting' of the ANN.
- irrelevant data need to be filtered.
- dynamic characteristics of the process need to be built into the database.

The first point is realized by creating traffic excitations as a sum of sinusoids with different frequencies. By using this scheme, occurrence of different traffic demand waves can be mimicked (e.g. the short rush before school opening during the morning rush-hour). The amplitudes of the different sinusoids are given by random variables to exclude deterministic patterns.

The second point is addressed considering topologic characteristics: spatially irrelevant information (i.e. the data of non-connected links) is excluded. When creating the database, it is reasonable to exploit the dynamic characteristics of the system. The most basic consideration is that the analysis and prediction horizon needs to be longer than the time constant of the system. Further dynamic characteristics can be involved by using statistical features, such as high-order moments, the tendencies and the highest relative variations. The additional

input features of the neural network are defined as listed in Table 1.

Notation	Description	Formula
1st moment	mean	$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$
2nd moment	variance	$\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2$
3rd moment	skewness (asymmetry in data distribution)	$\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^3$
4th moment	kurtosis (peakedness in data distribution)	$\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^4$
Max	maximum value	$\max_k x_k$
Min	minimum value	$\min_k x_k$
Max/Min	maximum/minimum	$\dfrac{\max_k x_k}{\min_k x_k}$
Max-Min	maximum-minimum	$\max_k x_k - \min_k x_k$
Tendency	tendency	$sign\,(x_n - x_1)$

Table 1: The applied statistical features

The prediction is carried out for each link separately, thus for each link a dedicated dataset was made according to the above listed rules.

III. CASE STUDY

In the case study, the objective is to predict the state of traffic around a high capacity intersection. The measured data covers only the mean speed of traffic for the network links. Based on a fix interval of 30 min measurements, state prediction is carried out for different horizon lengths (5, 15 and 30min). Using continuous input values, discrete states of traffic are forecasted.

The dataset is built using the simulated average traffic speed measurements of the network. The spatial and temporal discretization of the measurements is carried out considering the access to existing traffic data, (e.g. speed data of the online map applications). The simulations are engineered so that the obtained dataset represents the daily traffic load patterns (see Fig. 1). To avoid the 'over-fitting' of the ANN and to cover the comprehensive range of the parameter space, the generated patterns are obtained by an appropriately composed stochastic excitation of the traffic system, following the considerations of Section II.
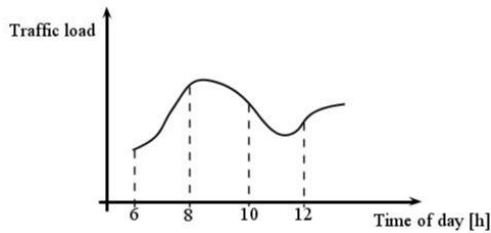
Figure 1: Modeled traffic demand

A. The applied method

The ANN-based method performs a classification to forecast a possible congestion of the system in the near future. This modeling process consists of multiple stages starting with the calculation of different statistical features from the measured average speed of the 16 segments in different time windows. These calculated features are not specific to the field of traffic prediction but a wide variety of different measures that help reveal the dynamics of the system by deriving new information from the measured values. At this point the goal is to define as many calculated features as possible to provide enough information for the next stages of the method. This stage is followed by a feature selection in order to reduce the high dimension of the problem and to select the most significant features for solving the estimation assignment, which also means that any unnecessary feature that was defined in the previous stage will be sorted out by the feature selection. This step is essential because ANN models (and other learning systems as well) are harder and more time consuming to train with high dimensional data, moreover this dimensionality reduction helps also to increase the modeling accuracy. After preprocessing the data, an ANN model can be trained to give estimation about the congestion state of the system in the future.

B. Road network scheme

The case study network models the vicinity of Oktogon square in District 6, Budapest (see Fig. 2).
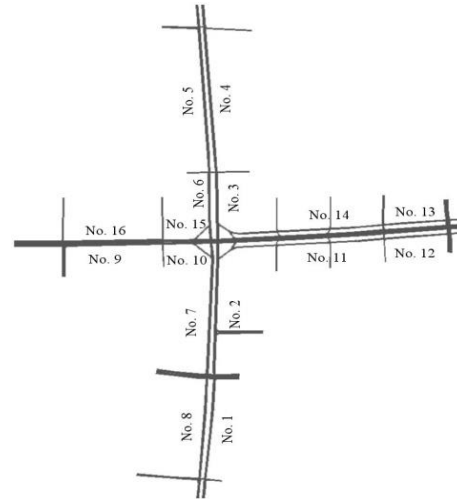

Figure 2: Scheme of the modeled road network

From the intersection, two sections are examined in all directions, respectively. The selected road sections are separated by intersections with traffic lights, thus the length of sections are different (the shortest is approximately 100 m (see No. 10 and 15 in Fig. 2), while the longest is approx. 330 m (No. 11 and 14)).

C. Database generation

During a simulation, the mean speed of traffic in each section is measured with a sampling time of 5 min. A sample is given with a row vector, containing the mean speeds of the sections. The measurements are organized in 60 min blocks. Thus, one record contains data of 12 measurements. These records are divided to two parts: the measurement data of the first 30 minutes are used as inputs (which are further modified), while the last 30 minutes serve as the logical outputs of the neural network. Applying a time-shifted framework for the measurement dataset, from a t–hour long simulation a total of $(t-1) \cdot 12$ records can be produced. In our case, 6-hour long simulations are run, resulting in 60 records. A total of 2500 records are used for training and 1000 for testing the neural network.

D. Input-output parameter selection

The prediction is carried out separately for each link, using a dedicated dataset. First, the non-relevant data are excluded: link measurements of opposite directions (e.g. for link No. 1, data of links No. 5-8 are excluded). Then, the statistical features of Table 1 for each relevant link are calculated and attached to the input vector. As a result of the input vector modifications, from a record an input vector of 266 elements are produced.

For the outputs, non-continuous discrete values are used to represent system states. The classification of continuous speed values to discrete states are performed based on the analysis of

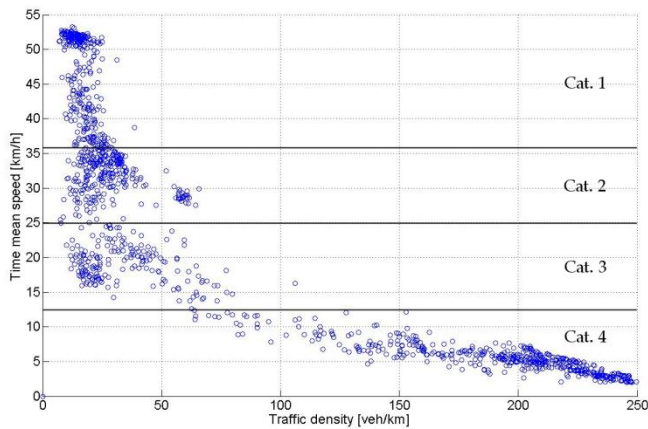the speed-density phase plot. Four discrete states are defined based on the categories shown in Fig. 3.

Figure 3: Mean speed – traffic density state pairs

The output of each pattern recognition problem is thus a three-element discrete valued vector (with the discrete state prediction of 5 mins, 15 mins and 30 mins ahead of the last input).

IV. THE APPLIED ARTIFICIAL NEURAL NETWORK MODEL

A. *ANN and the Neureca software*

Over the decades Artificial Neural Networks (ANNs) proved to be powerful computational models for solving complex estimation and classification problems as they are robust and are capable of high level generalization. The concept of the ANN was established around seven decades ago and, as the name suggests, it was inspired by the behaviour of biological neural networks inside the human brain. An ANN implements the functionality of the biological neural networks by building up a network of autonomous computational units (neurons) and connecting them via weighted links defined by the first pioneers W. S. McCulloch and W. Pitts [30].

There are several different ANN types which mostly differ in their structure and training algorithm as they were developed to fulfil different requirements of diversified applications. One of the most popular and widespread ANN model type is the MultiLayer Perceptron (MLP) [31].
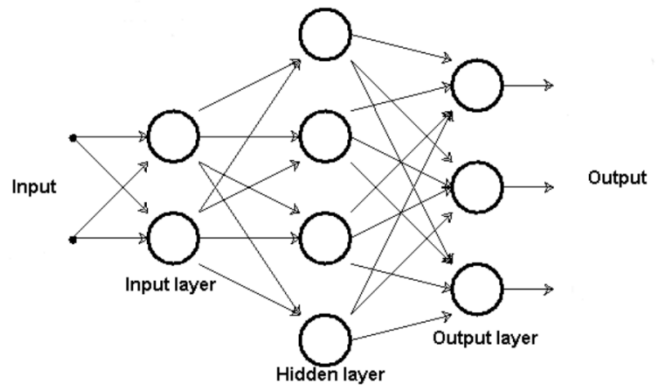

Figure 4. The MuliLayer Perceptron model

Fig. 4 shows an MLP model where the neurons are organized into layers and each layer is fully connected with the next one. Supervised training of an MLP means repeated adjustment of the weight of each link to receive more and more favourable output on specific neurons (output neurons) while stimulating other neurons (input neurons). The backpropagation algorithm achieves this by calculating the derivatives of the network's error with respect to all of its weights and adjusting the weights to a position where, based on the derivatives, the error is smaller e.g. moving the weights in the direction of the descent of the derivatives where the error is a measure of the difference between the network's output and the target values for the same input.

The MLP model was implemented in the Neureca2 software which was used in the introduced application. This software incorporates several ANN-based data mining algorithms like feature selection, adaptive learning methods, automatic recognition of input and output parameters, submodel decomposition, etc.

B. *The applied ANN method*

The applied method consists of two process stages. The first stage is a feature selection method which greatly reduces the number of parameters making it applicable for neural network training. This feature selection algorithm was originally proposed by Devijver and Kittler [32] and it assumes a pure classification task with the goal of reducing the number of inputs needed for one single output. As a generalization, continuous output parameters can be mapped onto the discrete classification scheme with an appropriate heuristics. The first step selects the output and its values encountered in the training data set are grouped into the highest possible number of clusters (i.e. intervals of equal length), so that at least one element is contained in each interval. Ranking remaining variables using suitable heuristics [22] with respect to a given "measure of distinction" delivers a list of variables and corresponding measure values. Taking more and more of them for input, the separability measure of the corresponding output deteriorates, and a further heuristic decision can determine how many of the best-ranked variables should be taken.

After the first stage selected the most significant features from the feature set the second stage can apply the ANN training on the reduced dataset. The ANN model is constructed from the selected features as inputs and the congestion/no congestion state as output. As there are three forecast settings (5, 10 and 15 minutes in the future) three different models are constructed accordingly.

V. SIMULATION RESULTS

This section discusses the classification results of the 13th link of the intersection. As a result of the simulation parameter refinement, several datasets had been created and model building had been carried out accordingly. One of the main conclusions was that it is enough to consider only one representative link because the links behave similarly to each other during the modeling process.

Three different forecast assignments had been defined to predict the four speed classes 5, 15 and 30 minutes ahead. In each case, the model building started with the feature selection in order to reduce the number of dimensions in the input space. The number of the most important features to be involved in the ANN training was determined by expert knowledge based on their importance measure. Generally, the most important features were the average speed of the last known 5 minutes of the adjacent links and the 3rd moment of the adjacent links. After the selection of input features, the ANN models had been trained for the classification task of each case.
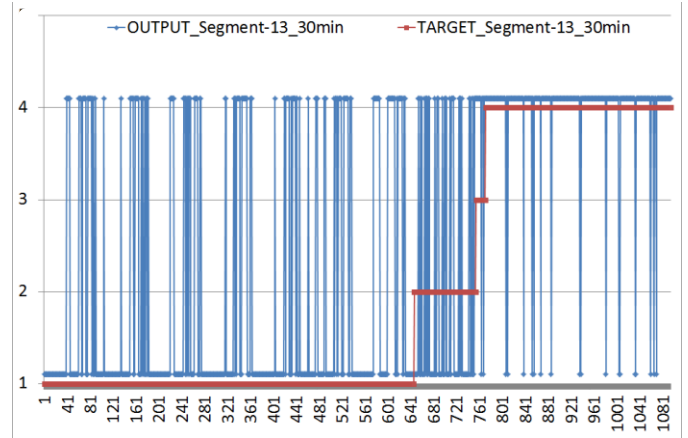

Figure 5. Classification result of 5 (top), 15 (middle) and 30 (bottom) minutes forecast

Fig. 5 shows the results of 5, 15 and 30 minutes forecast. The x axis indicates the samples ordered by the target classification and the y axis shows the classification value (the output value is shifted by 0.1 to make the diagrams more readable). On each diagram the target line (red) depicts the original classification and the output line (blue) represents the achieved classification of the model. The closer the output line is to the target line, the better the recognition rate is. The 5 minutes forecast shows that the first and fourth classes can be estimated with relatively high recognition rate (the output and target values are close to each other on the figure), but the two inner classes are highly mixed. Typically, the first and fourth classes are distinguished by the model in the other two cases too.
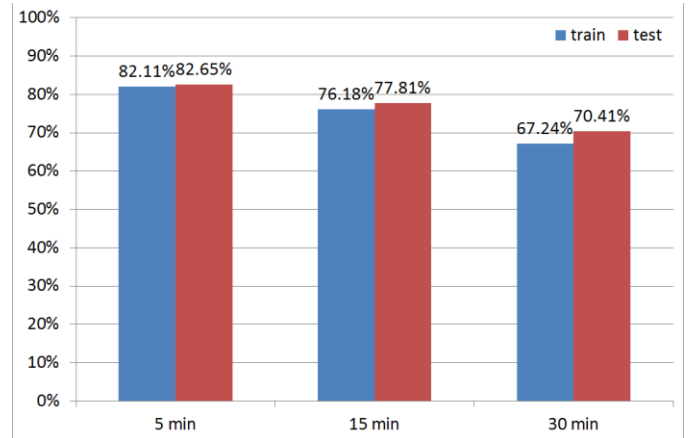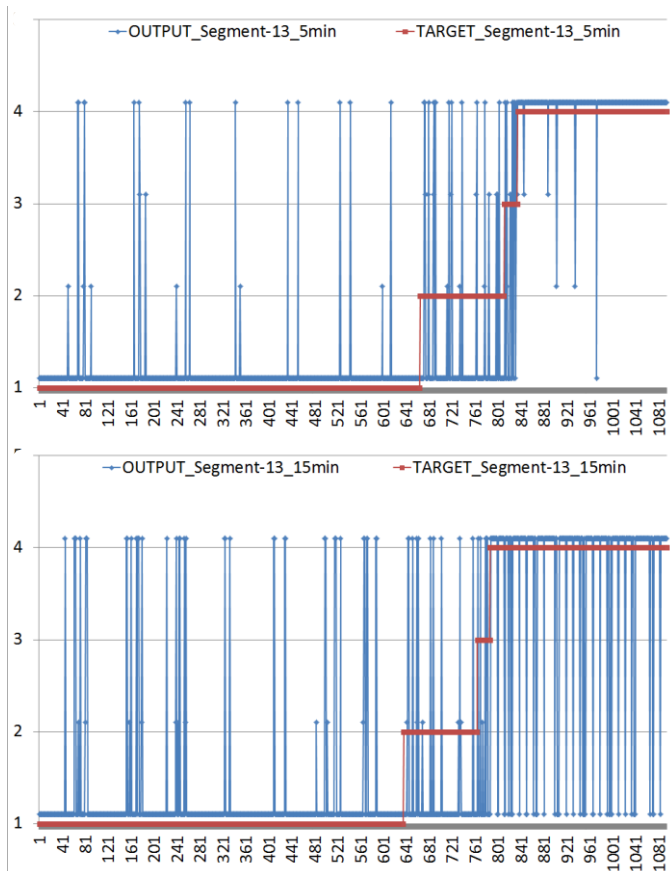

Figure 6. Decreasing recognition rate over the 5, 15 and 30 minutes forecasts

As a final evaluation of the applied method, the recognition rate has been investigated concerning the training set as well as the test set. Fig. 6 shows how the recognition rate of the model decreases as the forecast goes further in the future.

VI. CONCLUSIONS AND FUTURE WORK

A congestion prediction algorithm has been investigated specifically for urban road traffic networks. The application of ANNs for traffic state prediction has been analyzed according to different criteria. During the research, important experiences have been gained concerning the methodology for input-output parameter selection and appropriate feature selection. Numerical results attest that the generation and narrowing of input dataset plays a key role in the pattern recognition performance. It is also shown that acceptable prediction can be reached for short periods, serving as a practically applicable method for traffic control systems. The weakness of the current system lies on the user-defined classification of speed categories. As mid-speed range is present in transitive states of traffic with rapid dynamics, these classes are under-represented in the data samples compared to the extremities of free flow and congestion, which classes are confidently recognized. A higher resolution of data classes is expected to result in higher prediction performance. Accordingly, future work involves the statistical analysis of state classification. Practically, a continuous state (mean link speed) forecasting is planned to realize.

REFERENCES

[1] E. I. Vlahogianni, M. G. Karlaftis, J. C. Golias (2014): Short-term traffic forecasting: Where we are and where we're going. Transportation Research Part C - Emerging Technologies, In Press, Corrected Proof.

[2] S. Lin, Y. Xi, Y. Yang (2008): Short-Term Traffic Flow Forecasting Using Macroscopic Urban Traffic Network Model. Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems Beijing, China, pp. 134-138.

[3] X. Fei, C.-C. Lu, K. Liu (2011): A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction. Transportation Research Part C, Vol. 19, Issue 6, pp. 1306–1318.

[4] Z. Wang, Z. Huang (2003): An Analysis and Discussion on Short-Time Traffic Flow Forecasting. Systems Engineering Vol.21, No.6, pp.97-100.

[5] B.M. Williams, P.K. Durvasula, D.E. Brown (1998): Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models. Transportation Research Record: Journal of the Transportation Research Board, Vol.1644 , pp. 132–141.

[6] D. Billings, and J.-S. Yang (2006): Application of the ARIMA Models to Urban Roadway Travel Time Prediction - A Case Study. Proc. IEEE Int. Conf. Syst., Man, Cybern., Taipei, Taiwan, pp. 2529-2534.

[7] A. Guin (2006): Travel Time Prediction using a Seasonal Autoregressive Integrated Moving Average Time Series Model. Proc. IEEE Intell. Transp. Syst. Conf., Toronto, Canada, pp. 493-498.

[8] L. Wang, H. Xu, S. Zheng (2008): The Study on Forecasting for Traffic Volume in ITS. China Science and Technology Information, pp. 268-269.

[9] T. Kim, H. Kim, and D. J. Lovell (2005): Traffic Flow Forecasting: Overcoming Memoryless Property in Nearest Neighbor Non-Parametric Regression. Proc. 8th Int. IEEE Conf. Intell. Transp. Syst., Vienna, Austria, pp. 965-969.

[10] I. Okutani, Y.J. Stephanedes (1984): Dynamic prediction of traffic volume through Kalman filtering theory. Transportation Research Part B: Methodological, Vol. 18, Issue 1, 1-11.

[11] J. Guo, W. Huang, B.M. Williams (2014): Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. Transportation Research Part C - Emerging Technologies, In Press, Corrected Proof.

[12] E. I. Vlahogianni, M.G. Karlaftis, J.C. Golias (2003): Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. Transportation Research Part C, Vol. 13, pp. 211-234.

[13] M. S. Dougherty, M.R. Cobbett (1997): Short-term inter-urban traffic forecasts using neural networks. International Journal of Forecasting, Vol. 13, pp. 21-31.

[14] Y. Chen, B. Yang, Qi. Meng (2012): Small-time scale network traffic prediction based on flexible neural tree. Applied Soft Computing, Vol. 12, pp. 274-279

[15] Z. Li, D. Sun, X. Jin, D. Yu, Z. Zhang: Pattern-Based Study on Urban Transportation System State Classification and Properties. Journal of Transportation, System Engineering & IT, Vol. 8 (2008), Issue 5, pp. 83-87.

[16] L. Zhiheng, S. Dong, J. Xuexiang, Y. Di, Z. Zuo (2008): Pattern-Based Study on Urban Transportation System State Classification and Properties. J Transpn Sys Eng& IT, Vol. 8, Issue 5, pp. 83-87.

[17] R. Yu, G. Wang, J. Zheng, H. Wang (2013): Urban Road Traffic Condition Pattern Recognition Based on Support Vector Machine. J Transpn Sys Eng & IT, Vol. 13, Issue 1, pp. 130−136.

[18] M. Montazeri-Gh, A. Fotouhi (2011): Traffic condition recognition using the k-means clustering method. ScientiaIranica B Vol. 18, Issue 4, pp. 930–937.

[19] L. Lina, Y. Li, A. Sadek: A k Nearest Neighbor based Local Linear Wavelet Neural Network Model for On-line Short-term Traffic Volume Prediction. 13th COTA International Conference of Transportation Professionals (CICTP 2013), Vol. 96, pp. 2066–2077.

[20] S.-C. Lo (2013): Expectation-maximization based algorithm for pattern recognition intraffic speed

distribution. Mathematical and Computer Modelling, Vol. 58, pp. 449–456.

[21] X. Zhang, G. He: Forecasting Approach for Short-term Traffic Flow based on Principal Component Analysis and Combined Neural Network. SETP, Vol. 27 (2007), Issue 8, pp. 167–171.

[22] Zs. J. Viharos, L. Monostori, T. Vincze (2002): Training and application of artificial neural networks with incomplete data, Lecture Notes of Artificial Intelligence, LNAI 2358, The Fifteenth International Conference on Industrial & Engineering Application of Artificial Intelligence & Expert Systems, Springer Computer Science, Springer-Verlag Heidelberg; Cairns, Australia, pp. 649-659

[23] D. Srinivasan, C. Chan, P.G. Balaji: Computational intelligence-based congestion prediction for a dynamic urban street network. Neurocomputing, Vol. 72 (2009), pp. 2710-2716.

[24] L. Dimitriou, T. Tsekeris, A. Stathopoulos (2008): Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow. Transportation Reasearch Part C, Vol 16, pp. 554-573.

[25] M. G. Demissie, G. H. de Almeida Correia, C. B. (2013): Intelligent road traffic status detection system through cellular networks handover information: An exploratory study. Transportation Research Part C, Vol 32, pp. 76-88.

[26] J.W.C. van Lint (2010): Reliable travel time prediction for freeways: bridging artificial neural networks and traffic flow theory, PhD Thesis, Delft University of Technology

[27] M. Gastaldi, G. Gecchele, R. Rossi (2014): Estimation of Annual Average Daily Traffic from one-week traffic counts. A combined ANN-Fuzzy approach. Transportation Research Part C, Vol 47, pp. 86-99.

[28] J. Z. Zhu, J. X. Cao, Y. Zhu (2014): Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections. Transportation Research Part C, Vol 47, pp. 139-154.

[29] A. Lozano, G. Manfredi, L. Nieddu (2009): An algorithm for the recognition of levels of congestion in road traffic problems. Mathematics and Computers in Simulation, Vol. 79, Issue 6, pp. 1926–1934

[30] W. S. McCulloch, W. Pitts (1943): A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, Vol 5, pp. 115-133.

[31] P. J. Werbos (1974): Beyond Regression: New Tools for Prediction and Analysis in the Behaviour Sciences, Ph. D Thesis, Harward University, Cambridge.

[32] P. A. Devijver, J. Kittler, (1982): Pattern recognition, A Statistical Approach. Prentice-Hall International Inc., England, London.