THE ARCHIVES OF TRANSPORT

VOL. XXIV NO 4 2012

10.2478/v10174-012-0032-2

Classification and Prediction of Traffic Flow Based on Real Data Using Neural Networks

Teresa Pamuła*

Received September 2012

Abstract

This paper presents a method of classification of time series of traffic flow, on the section of the main road leading into the city of Gliwice. Video detectors recorded traffic volume data was used, covering the period of one year in 5-minute intervals – from June 2011 to May 2012. In order to classify the data a statistical analysis was performed, which resulted in the proposition of splitting the daily time series into four classes. The series were smoothed to obtain hourly flow rates. The classification was performed using neural networks with different structures and using a variable number of input data. The purpose of classification is the prediction of traffic flow rates in the afternoon basing on the morning traffic and the assessment of daily traffic volumes for a particular day of the week. The results can be utilized by intelligent urban traffic management systems.

1. Introduction

The rise in the number of cars on the roads and thus the growth of traffic flow rates has coerced the development of multiple solutions for controlling traffic flow. Among these solutions are systems for collecting traffic data, which significantly contribute to the improvement of the efficiency of road traffic control.

The knowledge of the traffic flow parameters is necessary to determine [4]:

- the distributions of traffic in the road network,
- the variability of traffic flows in particular measurement intervals,
- the vehicle type structure of flows,
- the directional structure of flows,

* Faculty of Transport, Silesian University of Technology, Krasińskiego 8, 40-019 Katowice, Poland, e-mail: teresa.pamula@polsl.pl

- the traffic volume forecasts on routes, and the level of utilization of routes capacity,
- the loads on the roadway.

The length of the basic interval for collecting traffic data is one hour. In the case of estimating roads, capacity intervals of 15 and 5 min are used.

Various models of the traffic prediction algorithms have been developed in recent two decades. Majority of the papers present approaches, based on the analysis of travel times in the road network. The Neural Network (NN) models have also been used in the field of various traffic parameter forecasting [8]. These provide the road users with e.g. routing information for efficient travelling through the network and enable the traffic supervisors to adapt control strategies to the traffic demand [7].

In the paper [3] sets of daily traffic data in Duisburg are organized into four basic classes and a matching process is proposed that attributes the current traffic data into these classes. Furthermore, two models for short-term forecast are examined: the constant and the linear model. These are compared with a prediction based on heuristics. The traffic description data plays a crucial role in the functioning of Intelligent Transportation Systems (ITS) applications.

The paper [5] presents three prediction models based on historical data. The performance of the models is examined using usual traffic patterns and traffic data with incidents. The results indicate that historical patterns provide a poor basis for predicting flows of traffic upset by incidents.

The paper [6] discusses a novel approach for traffic flow prediction, which uses a self-adaptive neural network. Authors emphasize the low computational complexity of this approach.

The authors in [16] successfully used a neural network to predict the values of traffic flow. Data for training the network were obtained from a road-tolling database. The prediction results were used for network planning and development, for traffic control and for the improvement of road traffic safety.

2. Traffic Intensity Characteristics

An important element of the traffic flow analysis is the inspection of the traffic fluctuations, related to the following factors:

- time (season, day of week, time of day, the time interval in hours),
- traffic type (urban, suburban, extra-urban, passenger and freight transport),
- structure of the generic traffic,
- the type and course of the road,
- capacity utilisation rate.

Significant fluctuations of traffic, occurring on particular days, strongly depend on the type and nature of road traffic.

On weekdays, the traffic intensity in urban and suburban roads generally remains at similar, high levels. During weekends traffic intensity decreases – with the exception of weekend return hours and the most popular weekend routes.

2.1. The analysed area

The urban roads are characterised by a certain backlog of traffic on Fridays, before Saturdays and Sundays. The intensity in the rush hours usually reaches 8-12% of daily traffic intensity.

The urban traffic is characterised by very high flow rates in rush hours. Morning peak is generally steeper than the afternoon peak, which is spread in time. Other factor, which strongly influences the distribution of fluctuations of traffic intensity, is the course of artery and its function.

The street under analysis is shown in Fig. 1 – the street in the town.

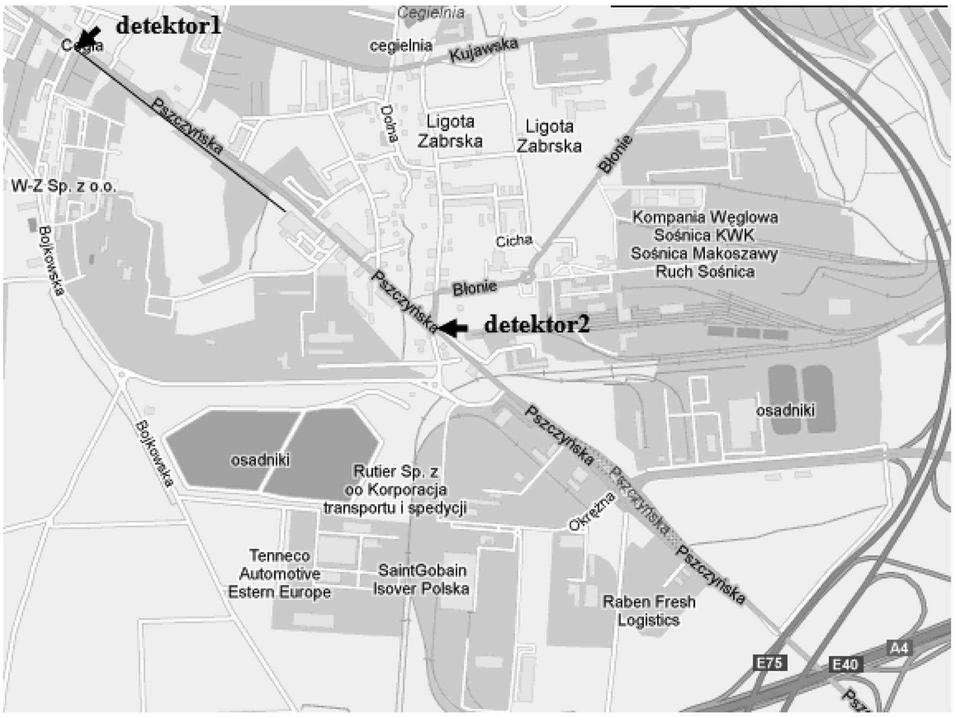

Fig. 1. The placement of the detectors on the Pszczyńska street in Gliwice (source Google maps)

The purpose of the presented study is to work out an automatic recognition method of the traffic flow class based on the time series of flow measurements. Knowledge of the flow class characteristics will allow the prediction of traffic flow in consecutive time intervals for instance for traffic control or planning such activities as road works [13].

The data used in this study come from the detector 1, registering vehicles travelling towards the city centre of Gliwice, and from the detector number 2 recording traffic flow values for vehicles leaving the city in the direction of the A4 motorway. Both detectors record the traffic flow data in five-minute intervals. Records cover the whole year from 2011-06-01 until 2012-05-31. There are over 200 thousand traffic data records.

The current study uses hourly traffic flow values sampled every hour. This proves to be sufficient for analysis of daily traffic. Originally, recorded data will be used for on-line predictions in future research.

2.2. Daily characteristics

The traffic data collected during the year were used to determine average daily traffic – the number of cars as a function of time. The Figure 2 shows the averaged data – Q_{AVd} from the detector 1, which counts the number of vehicles driving towards the centre. The time series shown in Figure 2 represent the average rate per day for working days from Monday to Friday, the average for Saturday, Sunday and holidays in the period from June 2011 to May 2012.

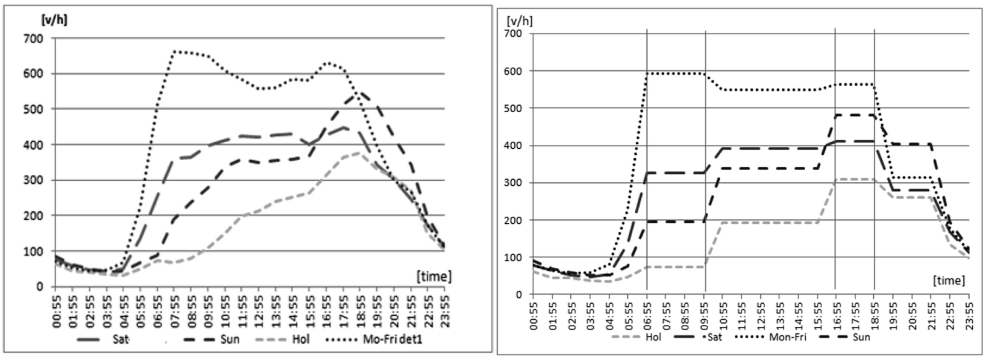

Fig. 2. The average traffic time series for all weekdays for vehicles going to the city centre

The traffic flows on days before holidays, treated in [3,13] as a separate class, in this study cannot be combined together because the absolute differences of flow values were too large. These traffic flows were attributed to classes corresponding to the days of the week when the flow was measured.

$$Q_{AVd}(t) = \frac{1}{n_d} \sum_{i=1}^{nd} q_{Day}(t) \tag{1}$$

where:

$q_{Day}(t)$ – is the average traffic volume in vehicles/hour for hour period, calculated for a given class,

n_d – numbers of days in analysed year depends from selected class,

t – time, hourly intervals, covering the whole day.

Four distinct classes of traffic flow graphs were defined: Monday-Friday, Saturday, Sunday and holidays.

A similar analysis was performed for the outbound traffic. It can be observed that the traffic flow rate distributions at the inspected road segment depend on the day of the week. The rates of the traffic flow measured in the working days from Monday to Friday are comparable. The largest differences appear in the days before holidays, because for many road users these appear to be working days. Therefore, days before holidays were allocated to other classes.

Figure 3 shows the averaged data from the detector 2, which counts the number of vehicles driving from the centre of Gliwice in the direction of A4 motorway.

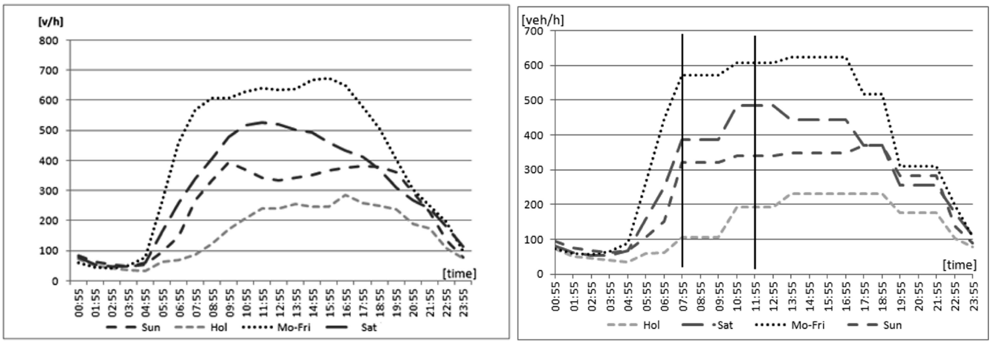

Fig. 3. The average traffic time series for all weekdays for vehicles going from the city centre

It is possible to distinguish clearly the flow classes of daily time series based on the analysis of average monthly values of traffic flow and using elementary statistic parameters. Four classes of the traffic flow graphs were selected for use with NN applications [1moja]. These are Mondays-Fridays, Saturdays, Sundays and the fourth class – holidays.

The hourly flow rates values measured every hour in the periods 6.55-9.55 and 16.55-18.55 hours were selected for performing classifications of flows using NN, for vehicles going to the centre. For vehicles going from the centre of Gliwice to the A4 motorway, the values from the interval 7.55-11.55 hours were selected for performing classification and prediction.

3. Neural Network

Finally, the following distinct traffic flow classes are defined:

- traffic on working days from Monday till Friday, except holidays (Mo–Fri),
- traffic on Saturdays, except holidays (Sat),
- traffic on Sundays, except holidays (Sun),

• holiday traffic (Hol).

In the paper [13], an attempt is presented to attribute flows to classes based on time series recorded during whole days. In the interval, 18.55 to 7.55 differences between the flow values for particular days of the week do not exceed on average 5%, which leads to ambiguous classifications of some flows by the neural network.

This article describes the results of two studies for vehicles going to the centre and one test for vehicles going in the direction of the A4 motorway. The first classification study used four consecutive samples of hourly flow rates in the morning at 6.55, 7.55, 8.55, and 9.55. Attribution to a specific flow class allows the prediction of traffic flow changes in the following hours (midday and afternoon) based on the morning flow.

3.1. The structure of the neural network

For performing the first classification study a back-propagation neural network, with four inputs and four outputs and one hidden layer was designed. The number of neurons in the hidden layer was chosen experimentally. The proposed network has the structure of 4-8-4. The inputs are fed with the values of hourly traffic flow rates. The network has four outputs. Network outputs correspond to the four classes of traffic flow.

The structure of the Neural Network is presented on Fig. 4.

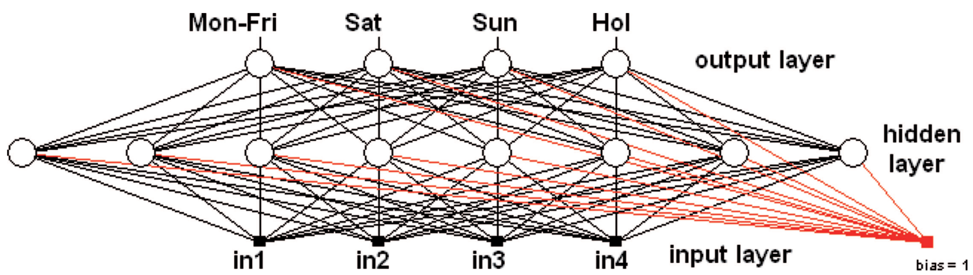

Fig. 4. Structure of the proposed NN for classifying traffic flows

3.2. The learning sequences and parameters of the learning process

The learning sequence consisted of 60 input vectors of four elements, 15 vectors for each class. The corresponding output vectors, that are outputs of the NN, consisted of four binary values; the value '1' indicated to which class the input vector belonged. The learning process was carried out until the root mean square error (RMSE) fell below the value 0,01. RMSE error is defined as follows [9]:

$$RMSE = \sqrt{\frac{1}{m}\sum_{j=0}^{m-1}(T_j - O_j)^2} \tag{2}$$

where

m – value of number of learning patterns

T_j – expected network output value,

O_j – measured network output value.

Neurons use the following transition function:

$$f(x) = \frac{1}{1 - e^{-z}} \tag{3}$$

where:

$z = \sum_i x_i w_i$, i – number of inputs of the neuron.

The neural network learning rate λ =0,9 and momentum α =0,7.

4. Classification Results

After learning, the network was tested. The test set consisted of vectors of hourly traffic flow values measured in the mornings at: 6.55, 7.55, 8.55, and 9.55, on each day. The effectiveness of the classification was examined using all available measurements.

Attribution to a specific class of traffic flow allows the prediction of the course of traffic flow in the next hours based on the flow values in the morning. For example, given traffic data in intervals 6.55-9.55 of the day, one can predict what will be the volume of traffic in the next hours on this day, based on averaged time series data for all recognized groups.

Classification results for the NN 4-8-4 structure are shown in Table 1.

The data in the table shows, that Saturday and Sunday traffic was the most poorly recognized. Saturday traffic was classified in four cases as Sunday traffic and in one case as holiday traffic. The reason for the Saturday false classifications may be ascribed to low values of traffic flow much lower than on usual Saturdays. Traffic on Saturday 5th May, 2012 was classified as holiday traffic, because earlier the 3rd May 2012 was a public holiday and many people went for a long weekend. The class Sunday traffic was erroneously identified 12 times; eight instances were identified as being Saturday traffic, and four as holiday traffic. Such results can be attributed to a high similarity of traffic flow on these days in the morning hours 6.55 – 9.55. The largest differences in traffic between the two classes occur in the afternoon. The graph on Figure 3 illustrates this case indicating larger average traffic volumes in the afternoon on Sundays than on Saturdays. Therefore, the network attributed

Table 1

Classification results for the days of the week for the 4-8-4 network

Groups	Day of week	Number of days	Recognized	Unrecognized	Unrecognized [%]
Mon-Fri	Monday	49	47	2 (Sat)	4%
	Tuesday	50	48	2 (Sat)	4%
	Wednesday	52	51	1 (Sat)	2%
	Thursday	51	50	1 (Sat)	2%
	Friday	50	43	7 (Sat)	14%
Sat	Saturday	52	47	5 (4 Sun, 1 Hol)	10%
Sun	Sunday	49	37	12 (8 Sat, 4 Hol)	24%
Hol	Holiday	13	13	0	0%

these flows to incorrect classes. To decrease the recognition error of Sunday traffic flows additional samples of the daily traffic flow were selected to be fed to the network. These are afternoon hourly traffic flow values at 16.55, 17.55, and 18.55. The number of data affecting the classification has thus increased to seven.

The next classification study used a back-propagation neural network, with seven inputs, four outputs, and one hidden layer. The number of neurons in the hidden layer was chosen experimentally. The proposed network has the structure of 7-15-4. The inputs are the values of hourly traffic flow measured every hour in the periods 6.55-9.55 and 16.55-18.55. The network has four outputs. Network outputs correspond to the four classes of the traffic flow.

Classification results based on seven traffic flow values from a day series are shown in Table 2.

Table 2

Classification results for the days of the week for the 7-15-4 network

Groups	Day of week	Number of days	Recognized	Unrecognized	Unrecognized [%]
Mon-Fri	Monday	49	48	1 (Sat)	2%
	Tuesday	50	50	0	0%
	Wednesday	52	52	0	0%
	Thursday	51	50	1 (Sat)	2%
	Friday	50	46	4 (Sat)	8%
Sat	Saturday	52	47	5 (Sun)	10%
Sun	Sunday	49	49	0	0%
Hol	Holiday	13	7	6 (Sun)	46%

Results show that increasing the number of inputs of the classification NN highly improves the average recognition rate of weekday traffic but significantly spoils the results for holiday traffic recognition. This arises from the fact that in Poland, some holidays are treated as Sundays, for example, the traffic during the second day of Christmas, or during the second day of Easter is much higher than on the first days of these holidays – Figure 5, so the use of additional afternoon samples highly changes the classification results.

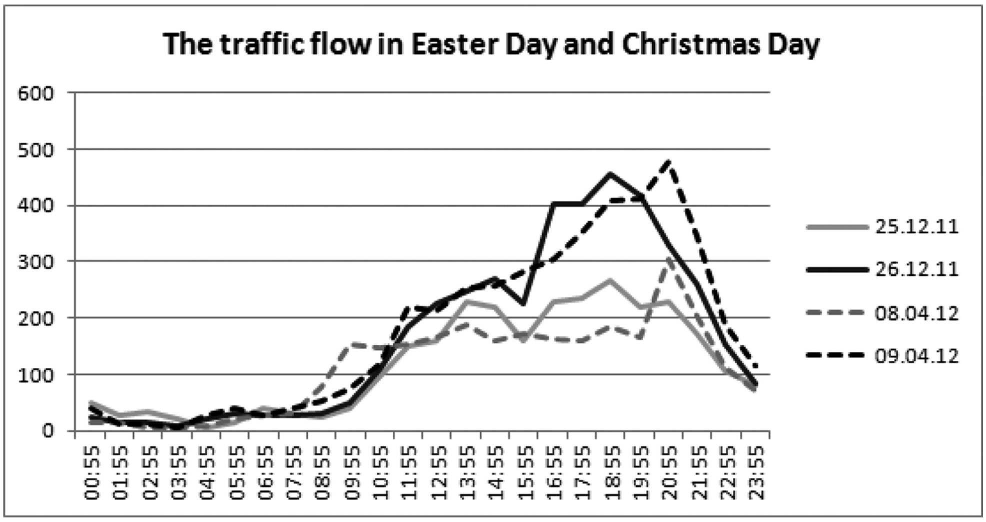

Fig. 5. The traffic flow on Easter Day and Christmas Day for vehicles going to the city centre

Traffic flow of vehicles going from Gliwice in the direction of A4 motorway, was processed in a similar way. The classification results are presented in Table 3.

Table 3

Classification results for the days of the week for vehicles going from the centre

Groups	Day of week	Number of days	Unrecognized NN structure **5-10-4**	Unrecognized NN structure **5-10-4** [%]	Unrecognized NN structure **5-8-4**	Unrecognized NN structure **5-8-4** [%]
Mon-Fri	Monday	49	5 (Sat)	10%	4 Sat	8%
	Tuesday	50	7 (Sat)	14%	3 Sat	6%
	Wednesday	52	8(Sat)	15%	3 Sat	6%
	Thursday	51	8 (Sat)	16%	4 Sat	8%
	Friday	50	6 (Sat) 2 (Sun)	16%	4 Sat 1 Sun	10%
Sat	Saturday	52	8 (5 Sun, 3 Working days)	15%	5 Work days 5 Sun	20%
Sun	Sunday	49	6 (6 Hol)	12%	14 Sun (5 Sat, 9 Hol)	29%
Hol	Holiday	13	2 Sat	15%	1 Sat	8%

For vehicles, travelling from Gliwice the greatest mean differences of hourly traffic flow rates during the day occurred at 7.55, 8.55, 9.55, 10.55, and 11.55. These samples were selected as inputs for the classification network.

The operation of two networks with different structures: 5-10-4 and 5-8-4 was examined. The networks differ in the number of neurons in the hidden layer. The results of classification are shown in the table. The network with eight neurons in the hidden layer classifies daily traffic better than the network with 10 neurons in the hidden layer. Weak classification results were obtained for the traffic classes Saturday and Sunday. These results can be ascribed to the fact that on these days some mass events generated traffic corresponding to the values of traffic on workdays.

Generally, classification results can be considered satisfactory, because even if the traffic is ascribed to a wrong day class during a week, the afternoon traffic flow values may be estimated based on average annual rates for this day.

5. Summary and Conclusions

The results of the performed tests prove that the proposed neural networks are effective for classifying traffic flow required by traffic prediction tasks. The average classification error was 8%. False flow classifications arose when traffic flow was significantly different than usually, such were for instance traffic flows on workdays between holidays on the 01.05 and 03.05 or before holidays. Although misclassified as holiday traffic their rate values correspond to holiday traffic and therefore resulting classifications give correct predictions for afternoon traffic.

Data recorded on days with abnormal traffic flow due to road incidents or extreme weather conditions, which were not documented, were excluded from the examination. In a future study, these instances of traffic flow can be selected to establish an additional class of flows indicating traffic incidents. The set of these abnormal flows is very small, in the time scale of one year, and it would require a supplement of artificially generated incident flows in order to perform network training.

The relatively good results indicate that it is possible to use neural networks, in traffic control systems. Determination of the class of the morning traffic allows planning of traffic control sequences adjusted to appropriate afternoon flows. The simple design of the NN and a very limited set of input variables required to function, facilitates the implementation of this solution in control systems.

Acknowledgments

The author wishes to thank ZIR-SSR Bytom for providing video detector data from the Gliwice site.

References

1. Awad W.H.: "Estimating traffic capacity for weaving segments using neural networks technique", Applied Soft Computing 4 (2004), pp. 395-404.

2. Cai C., Wong C.K., Heydecker B.G.: Adaptive traffic signal control using approximate dynamic programming, Transportation Research Part C, 17(5), pp. 456-474, 2009.
3. Chrobok R., Kaumann O., Wahle J., Schreckenberg M.: Different methods of traffic forecast based on real data. European Journal of Operational Research 155 (3), pp. 558-568, 2004.
4. Gaca S., Tracz M., Suchorzewski W.: „Inżynieria ruchu drogowego", WKiŁ Warszawa 2008.
5. Guo F., Polak J., Krishnan R.: "Comparison of Modelling Approaches for Short Term Traffic Prediction under Normal and Abnormal Conditions" Annual Conference on ITS Madeira Island, Portugal, September 19-22, 2010, pp. 1209-1214.
6. Haixiang D., Jingjing T.: "Prediction of Traffic Flow at Intersection Based on Self-Adaptive Neural Network", Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on Vol: 8, 2010, pp. 95-98.
7. Karlaftis M.G., Vlahogianni E.I.: Statistical methods versus neural networks in transportation research: differences, similarities and some insights. Transportation Research, Part C. Emerging Technologies 19 (3), pp. 387-399, 2011.
8. Ledoux C.: An urban traffic flow model integrating neural networks, Transportation Research C, vol. 5, No 5, pp. 287-300, 1997.
9. Osowski S: „Sieci neuronowe w ujęciu algorytmicznym", Wydawnictwo Naukowo-Techniczne, Warszawa 1996.
10. Pamuła T., Król A.: „Model systemu zarządzania ruchem pojazdów w obszarze miejskim z wykorzystaniem sieci neuronowych", Zeszyty Naukowe P.Ś., Gliwice 2010, Seria: TRANSPORT z. 67, pp. 91-96.
11. Pamuła T.: Road traffic parameters prediction in urban traffic management systems using neutal networks, Transport Problems, Vol. 6, Issue 3, Wyd. Pol. Śląskiej, pp. 123-129, 2011.
12. Pamuła T.: The neural network implementation for traffic junctions volume prediction, Transactions on Transport Systems, Telematics&Safety (Monograph), pp. 57-66, Gliwice 2011.
13. Pamuła T.: Traffic flow analysis based on the real data using neural networks, Telematics in the transport environment. Ed. Jerzy Mikulski. s. 364-371, bibliogr. 13 poz. (Communications in Computer and Information Science; vol. 329), Berlin: Springer, 2012.
14. Satsangi P.S., Mishra D.S., Gaur S.K., Singh B.K.: "Systems dynamics modelling, simulation and optimization of integrated urban systems", A soft computing approach, The Emerald Research Journal, Vol. 32 No. 5/6, 2003, pp. 808-817.
15. Srinivasan M.C., Choy R.L. Cheu: "Neural networks for real-time traffic signal control," IEEE Trans. Intelligent Transportation Systems, vol. 7, no. 3, pp. 261-271, Sep. 2006.
16. Xiaoying Li: "Prediction of Traffic Flow Base on Neural Network", Intelligent Computation Technology and Automation, 2009. ICICTA '09, 2009, pp. 374-377.