The Traffic Flow Prediction Using Bayesian and Neural Networks

Teresa Pamuła and Aleksander Król

Abstract The article presents two short-term forecasting models for determining the traffic flow volumes. The road traffic characteristics are essential for identification the trends in the distribution of the road traffic in the network, determination the capacity of the roads and the traffic variability over the time. The presented model is based on the historical, detailed data concerning the road traffic. The aim of the study was to compare the short-term forecasting models based on Bayesian networks (BN) and artificial neural networks (NN), which can be used in traffic control systems especially incorporated into modules of Intelligent Transportation Systems (ITS). Additionally the comparison with forecasts provided by the Bayesian Dynamic Linear Model (DLM) was performed. The results of the research shows that artificial intelligence methods can be successfully used in traffic management systems.

1 Introduction

The knowledge of the characteristics of the traffic flow is an obligate condition for identification of the trends of traffic distribution in the network, determination of the roads capacity and variability of the traffic over the time. The description of the road traffic data plays a key role in the operation of Intelligent Transport Systems. These systems provide the helpful navigation information to the road

---

T. Pamuła (✉) · A. Król
Faculty of Transport, Silesian University of Technology,
Krasińskiego 8, 40-019 Katowice, Poland
e-mail: teresa.pamula@polsl.pl

A. Król
e-mail: aleksander.krol@polsl.pl

A. Sładkowski and W. Pamuła (eds.), *Intelligent Transportation Systems – Problems and Perspectives*, Studies in Systems, Decision and Control 32, DOI 10.1007/978-3-319-19150-8_4

users and allow the traffic supervisory authority to adjust the strategy of the traffic management to the actual traffic conditions [1].

Neural Networks (NN) are an efficient tool for classification and identification of the road traffic parameters [2, 3]. Due to their properties, first off all the ability of mapping of the multiparameter relationships as a result of the learning process. Neural Networks are applied for prediction of the traffic flow volume at the different time horizons [4]. The short term prediction of the traffic flow volumes are particularly useful for the control decision making at the crossroads [3, 5]. Modeling of the time course of the traffic flow with the use of NN is an element of the adaptive algorithms for traffic control [6, 7].

The Bayesian networks describe the set of dependences between some random variables. A Bayesian network allows to calculate the probabilities of some states with possessing the knowledge about some other states, thus it is possible to apply such the ability for the prediction [8].

The paper presents the comparison of the prediction abilities of the mentioned above neural prediction model with the prediction model based on the Bayesian networks. Both models were developed basing on historical data about the traffic volume variability. Described approaches are compared with Bayesian dynamic linear model (DLM) predictions, which is one of willingly used methods for time series analysis. The real data measured in road network of Gliwice city are used here for learning and validation. The study includes three classes of time series, which were specified on the basis of the statistical analysis [9] of the road traffic on different days of the week. The results of the prediction process will be applied for selection of the signaling modes within the crossroads and to support of the area-traffic management.

The results of comparison are gathered in the tables in the last chapter and then the conclusion summarizes the research outcomes and proposes further development of presented methods, which is aimed mainly on the improvement of the prediction accuracy.

The database of traffic flow volumes was acquired by the use of the vehicle detectors located on the main driveway to Gliwice. Gliwice, which is placed on the periphery of Upper Silesia agglomeration and has a population of almost 200,000 inhabitants is a representative research training ground for the medium-sized city. The achieved measurements include wide range of traffic flow volumes and allow to examine the different variability schemes.

The data from two video detectors located at the ends of a transit road in Gliwice town are used. Data were recorded 24 h/day for a period of one year from 1 July 2011 to 31 March 2012. The data were initially divided into groups according to the classes of the time series, such as: working days, Saturdays and Sundays [2]. Such distinction was necessary due to quite different time structure of the road traffic. Part of the data has been used to prepare training sequences and test sequences.

The observed rate of change in the measurement of the volume of traffic shows, that 5 min, coinciding with the period of measurement, prediction horizon takes into account the intensity of the fastest changes in traffic flow. Adopted to predict

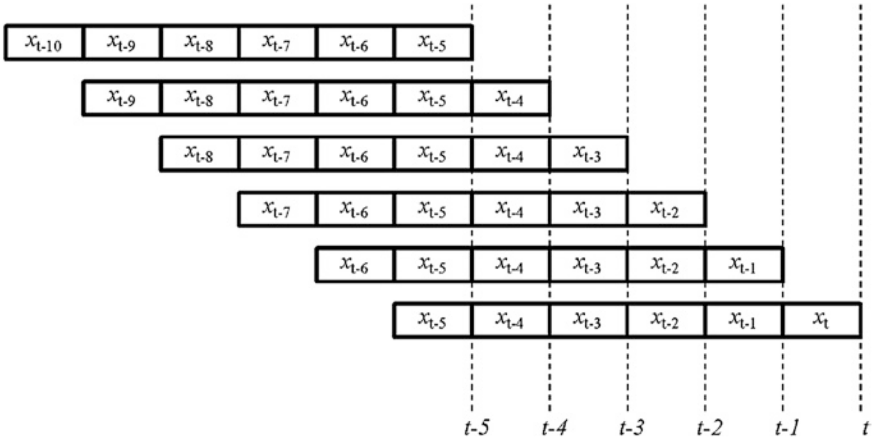

Fig. 1 The principle of the vehicle counting

the time window with a length of 6 measurement periods. This means that the model predictions, defines the data the last six traffic flow values within 30 min. The result of the model is the expected value of the traffic after the next 5 min.

2 Traffic Flow Measurement

The data was recorded automatically using the video-detection system. The length of the measurement window was 30 min, and a new count was started every 5 min, thus every 5 min the new input containing mean traffic flow in the past 30 min was provided. In this way there were six input values during the period of 30 min (Fig. 1).

The data incoming every 5 min can be treated as random variables. These random variables are strongly dependent on each other because consecutive measurements cover almost the same time interval. In fact, such data allows to calculate the number of vehicles passing in 5 min interval.

3 Traffic Flow Statistics

3.1 Traffic Flow Distribution

The number of vehicles detected in a given time interval should obey a Poisson distribution. It is important to assume the single vehicle detections are independent on each other. It is true excluding the very rare cases of passing convoys. If n

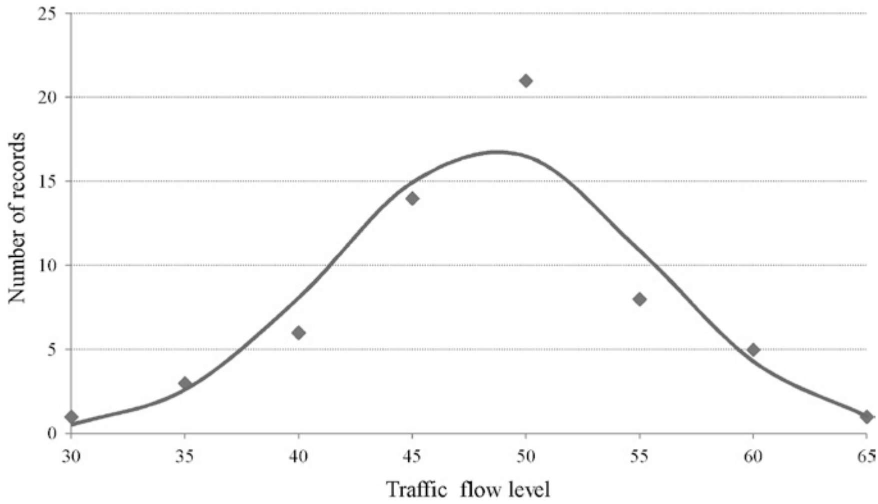

Fig. 2 Traffic flow distribution for Wednesdays in August between 1 p.m. and 2 p.m. compared with the Poisson distribution with λ = 48.5

denotes the constant average number of vehicles per the time unit, the probability of the k vehicles being detected at time interval t can be expressed as:

$$\mathrm{P}(X = k) = \frac{(nt)^k \mathrm{e}^{-nt}}{k!} \tag{1}$$

The product of nt is sometimes replaced by the average number of events detected in the time interval and denoted by λ. The important feature of Poisson distribution is that the expected value and the variance are equal and equal to λ:

$$\mathrm{E}(X) = \mathrm{V}(X) = \lambda \tag{2}$$

For very rare events ($\lambda < 1$) only the probabilities of a few events occurrence are significantly greater than 0. If $\lambda \gg 1$ Eq. (1) becomes awkward to use, but in this case the original distribution can be approximated by the normal distribution ($\mu = \sigma^2 = \lambda$):

$$\mathrm{P(X} = k) = \frac{1}{\sqrt{2\pi\lambda}} \mathrm{e}^{-\frac{(k-\lambda)^2}{2\lambda}} \tag{3}$$

The real data obtained in the video-detection system match well the theoretical distribution, but it concerns only the values calculated for 5 min intervals. As an example the traffic flow distribution for Wednesdays in August between 1 p.m. and 2 p.m. is shown in Fig. 2. The number of detected vehicles varies from 30 to 65 and the whole range was divided into intervals with a width of 5. The mean value is equal to 48.5 and the variance is equal to 47.7.

The compliance of the empirical and the theoretical distributions was confirmed also by the use of Chi-Square test (Table 1). The p-value was here about 0.96,

Table 1 Comparison of empirical and theoretical distributions

Traffic flow level	Real number of records	Theoretical distribution (Eq. 3)
30	1	0.500810
35	3	2.600963
40	6	8.064806
45	14	14.929730
50	21	16.500920
55	8	10.888370
60	5	4.289591
65	1	1.008944

$p = 0.96$

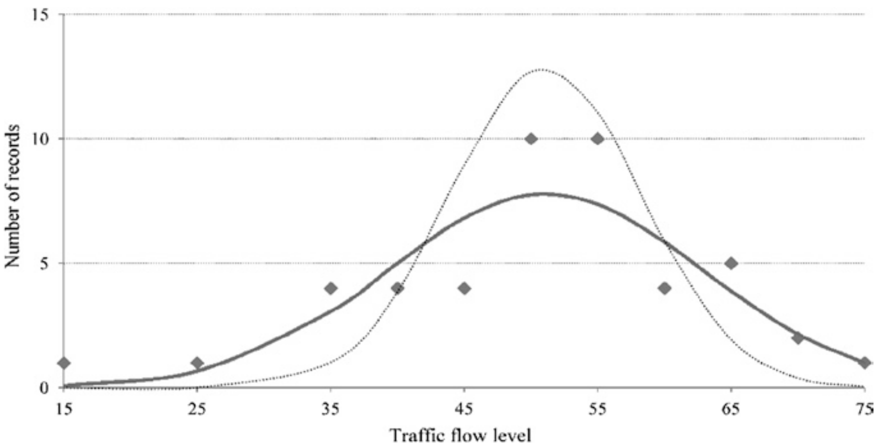

Fig. 3 Traffic flow distribution for Thursdays in August between 4 p.m. and 5 p.m. compared with the normal distribution ($\mu = 51.2$ and $\sigma 2 = 141.9$) and Poisson distribution with $\lambda = 48.5$

thus there is no presumption against the hypothesis of the compliance of both distributions.

Such good compliance concerns the low and the moderate traffic flow volumes. When high traffic flow volumes were considered (at the rush hours) the measured distributions diverged significantly from the Poisson distribution. It can be explained by the assumption that at the rush hours the traffic is flowing under the congestion conditions and the successive car detections are not independent from each other. As the example the traffic flow distribution for Thursdays in August between 4 p.m. and 5 p.m. is shown (Fig. 3). The mean value is equal to 51.2 and the variance is much greater and equal to 141.9. The distribution remains near normal, but much more diffuse.

Table 2 Descriptive terms of the time of the day

Term	Hours
MORNING	4, 5, 6, 7, 8, 9
DAY	10, 11, 12, 13, 14, 15
AFTERNOON	16, 17, 18, 19, 20, 21
NIGHT	22, 23, 0, 1, 2, 3

If distributions including more data e.g. different week days, different months were examined this observation remained true—the distribution was normal, but the variance was much greater (three and more times) than the mean value. It happened independently on the choice of the time of a day. It is due the fact that some other factors disturb the pure Poisson process and make the distribution more blurry.

3.2 Data Classification

When analyzing the distribution of road traffic over the time of day it can be noticed two peaks: one in the morning and the second in the afternoon. It is rather calmly in the night and the traffic is moderate during the day, between peaks. This obvious and rough pattern applies to all week days. But detailed study finds out that three different patterns should be distinguished. The time structure of road traffic appeared to be different for working days, Saturdays and Sundays. To confirm these two groups of tests were performed:

for each week day and for each time of day the average number of vehicle counts per one record was calculated,
for each week day and for selected traffic volume levels the average number of records per day was calculated.
For such collections of values the multi-resolution arrays were created and then with the use of Chi-Square tests the probability of mutual independence was determined. The contractual descriptive terms of the time of the day are gathered in Table 2.

Both mentioned above tests were applied for all week days (from Monday to Sunday) and for working days only. Results are shown in Table 3. The cells described by p contain the calculated p-value, which should be at least equal to the typically assumed significance level value of 0.05.

The value of p over 0.05 means the distribution due one criterion is likely to be independent on the second criterion. Thus the data presented in Table 2 clearly demonstrate that the structure of the traffic is similar for all working days

Table 3 The results of the independence tests

Average number of vehicle counts per one record—working days only						$p = 1$
Time of day	MON	TUE	WED	THU	FRI	
MORNING	202	205	202	203	199	
DAY	298	300	297	305	311	
AFTERNOON	226	232	231	236	250	
NIGHT	39	43	42	44	51	

Average number of vehicle counts per one record—the whole week							$p = 0$
Time of day	MON	TUE	WED	THU	FRI	SAT	SUN
MORNING	202	205	202	203	199	170	201
DAY	298	300	297	305	311	217	174
AFTERNOON	226	232	231	236	250	117	87
NIGHT	39	43	42	44	51	50	43

Average number of detection per day—working days only						$p = 0.77$
Volume level	MON	TUE	WED	THU	FRI	
0–150	239	229	228	226	217	
150–300	183	195	200	185	186	
Above 300	154	152	148	165	173	

Average number of detection per day—the whole week							$p = 0$
Volume level	MON	TUE	WED	THU	FRI	SAT	SUN
0–150	239	229	228	226	217	310	323
150–300	183	195	200	185	186	254	230
Above 300	154	152	148	165	173	12	15

(other, not shown here tests prove the same), but other patterns must be used for Saturdays and Sundays.

4 Description of Methods Used for Prediction

4.1 Bayesian Networks

A Bayesian network presents graphically the relationships between some random variables. The name origins from the Bayes' theorem, which postulates the belief revision in the light of new facts. Let's assume that an event (B) can occur in several mutually exclusive ways (which complete all the possibilities), then the joint probability of the event B occurrence should be expressed as:

$$\mathrm{P}(B) = \sum_{i=1}^{n} \mathrm{P}(A_i)\mathrm{P}(B|A_i) \tag{4}$$

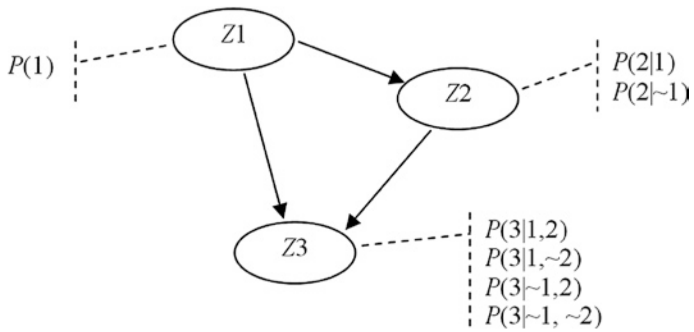

Fig. 4 A simple Bayesian network with binary (only true and false states) random variables

where

A_i the exclusive possibilities of the B event implementation
$P(B|A_i)$ conditional probability of the B at condition of A_i

Now assume that the B event has already occurred: this knowledge allows us to recalculate the prior probability of $P(A_k)$ to the new value $P(A_k|B)$:

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{P(B)} \quad (5)$$

Despite of the simplicity both formulas are powerful tools allowing to perform the bidirectional inference in any network of probabilistic relationships.

Formally a Bayesian network is an acyclic directed graph, the vertices correspond to the random variables and the edges correspond to the direct cause—effect relationships between pairs of variables. There is a table containing the conditional probabilities related to each vertex. This table describes the chances of particular states $P(X \mid P_1, P_2, \ldots)$ of given vertex X taken at different states of its direct parents P_1, P_2, … (Fig. 4). For the vertices without parents (so called root causes) the conditional probabilities become the simple probabilities [6].

The study used GeNie package developed at Decision Systems Laboratory at the University of Pittsburgh [10]. GeNie package implements the classical model of Bayesian network with additional elements supporting the use of diagnostic and decision-making processes. There is a large number of available algorithms for performing Bayesian inference. The user interface allows to enter data in the graphic form.

4.1.1 Bayesian Inference

The graph topology and the tables of conditional probabilities determine the relationships between random variables and thus allow for calculation of the probability of any state of any variable with the assumption of the knowledge concerning the remaining variables. This process is called belief updating and is performed with the use of the joint probability formula and the Bayes' formula. Because of fact that in some cases the structure of the Bayesian network can be complex and the number of possible network states can be large, the calculation may appear to be computationally hard [11].

4.1.2 Bayesian Network Construction and Learning

The knowledge of a Bayesian network is contained in the graph topology and in the tables of conditional probabilities. Thus all these properties must be determined when the network is being constructed. There are two main possibilities for the construction of a Bayesian network [12]:

- building by an expert (it may concern both the topology and the values of the probability)
- automatic generation with the use of learning data—the series of cases containing the values of random variables.

Here the automatic mode was chosen, but the topology of the Bayesian network was determined empirically. The conditional probabilities associated with the vertices are found with the application of the built-in EM algorithm.

The EM (Expectation–Maximization) algorithm is the iterative procedure to estimate the maximum likelihood in the incomplete data. It endeavors to estimate the model parameters for which the observed data are the most probably [13]. Every iteration consists of two steps:

- the E-Step (expectation): missing data are estimated given the observed data and the current estimation of the model parameters,
- the M-step (maximization): the likelihood function is maximized with the assumption the missing data are now known.

4.1.3 The Structure of the Bayesian Network for Prediction

It was assumed that the expected volume of the traffic flows depend on values observed at the current and a few previous moments. Additionally each of previous values affects all the next. It corresponds to the structure presented at Fig. 5. The vertices marked x-t correspond to values measured at the past time moments. The vertices marked xout1 and xout2 hold the forecast for respectively one and two steps ahead.

An alternative structure of the Bayesian network was examined too: there is an additional node, which represents the time of a day. This new node affects all other nodes and contains 24 states, which simply correspond to hours (Fig. 6).

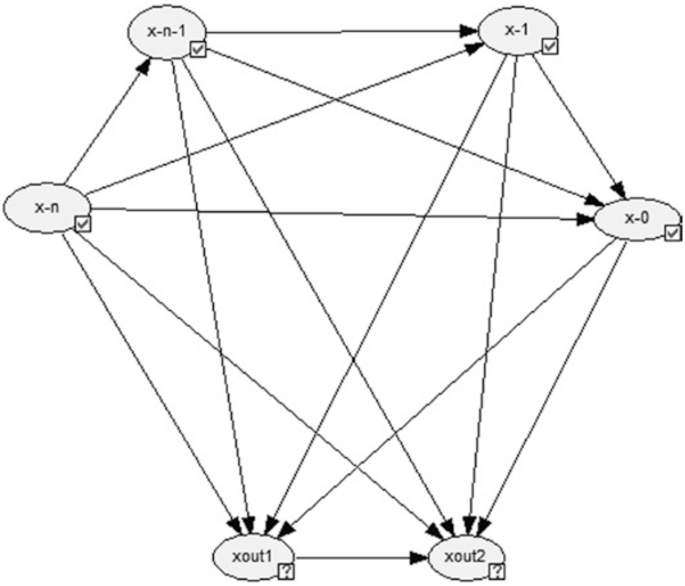

Fig. 5 The structure of the Bayesian network used for prediction

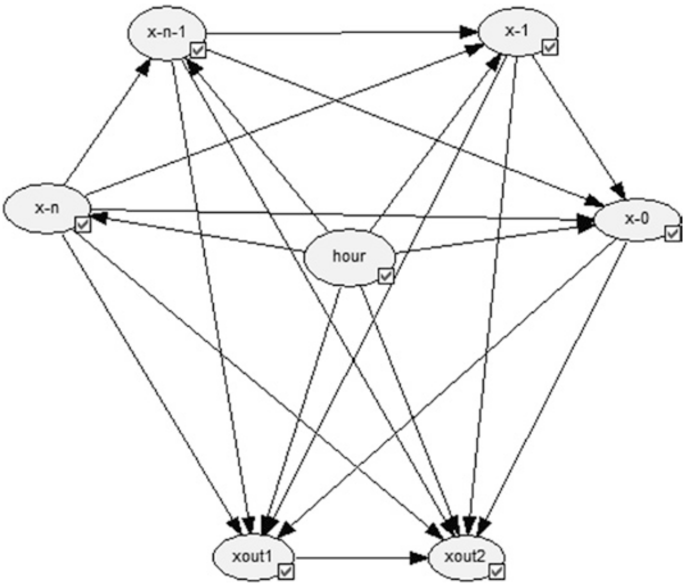

Fig. 6 The structure of the Bayesian network with 'hour' node

4.1.4 Input Data Preprocessing

The input data are given as a time series of the traffic flow volume. So a single value can be any integer number. Meanwhile the nature of a random variable requires to determine some states, which can be taken. Thus the process of discretization must be performed at the beginning. It was assumed that each of variables can take 20 states. This number is a compromise between the accuracy of the prediction and the efficiency of the Bayesian inference. As was shown above the variance of the incoming data is equal or often greater than variance of the Poisson distribution. Thus in this case the standard deviation is assumed to be somewhat greater than the square root of the number of detected vehicles. Such value is greater than the single state width. The discretization was done by the built-in hierarchical algorithm. It starts with the number of states equal to the number of input values, and then at every step the number of states is reduced by one by merging states with the nearest mean values. The discretization stops when the number of states reaches the required value.

4.1.5 Determination of the Forecast Value

As was mentioned above the input data was divided into three classes: working days, Saturdays and Sundays. These three data sets were used as learning data for Bayesian network with the structure described above, giving three different prediction models. The models were applied to the data concerning a number of selected days. For each moment, the states corresponding to the previous values of the traffic flow should be set, then after the belief update the variables xout1 and xout2 contain the forecast (Fig. 7).

The output (predicted value) is also expressed in terms of states. Instead of a single value the distribution of possible states is obtained. The required value is the weighted average of them (Eq. 6).

$$x_{out} = \sum_i p_i m_i \tag{6}$$

where

x_{out} predicted value,
p_i probability of *i*th state,
m_i the mean value corresponding to *i*th state.

4.2 *Neural Networks*

Feedforward (static) neural network is a set of interconnected artificial neurons without feedback connections. The data flow only in one direction—from inputs through hidden layers towards outputs. The simplified basic model of a single artificial neuron is shown at Fig. 8.

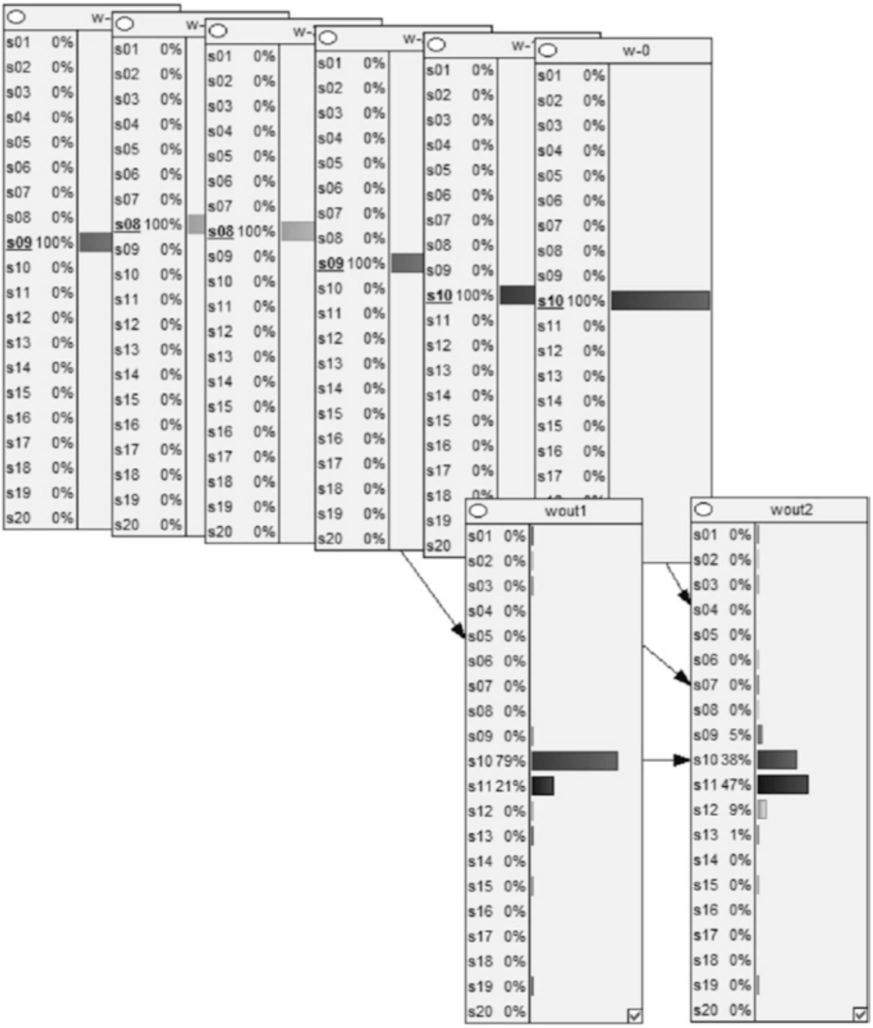

Fig. 7 The predicted value expressed as a distribution of states

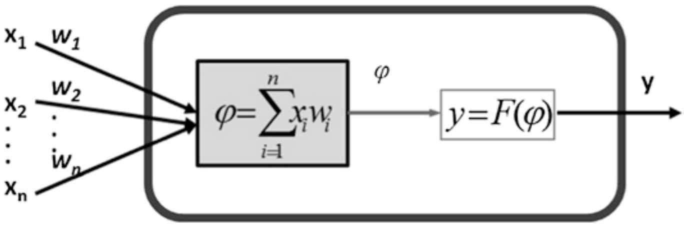

Fig. 8 The simplified basic model of a single artificial neuron

The input signals $x_1, \ldots, x_R$ reaches the neuron, then their weighted sum is calculated and the transfer function is applied to form the output signal:

$$y = f\left(\sum_{1}^{R} w_i x_i\right) \tag{7}$$

For the prediction of traffic a one-way network with sigmoidal transfer function for each neuron was proposed. A two-layer structure of the network was chosen, because the mapping of properties of waveforms requires a multi-dimensional decision-making area. The features of traffic, such as: cyclical changes of fluctuations, different speed, acceleration changes must be taken into account.

4.2.1 The Neural Networks Structures

The proposed networks had a structures of 6-22-1, 7-22-1 with one output, and 5-22-2, 6-22-2 with two outputs. Diagrams of those networks shown in Figs. 9 and 10.

The input vector consists of six or seven successive previous measurements of traffic intensity for the network with one output. The value of traffic flow was predicted for the next 5 min. A 5 min measurement period was chosen, which means

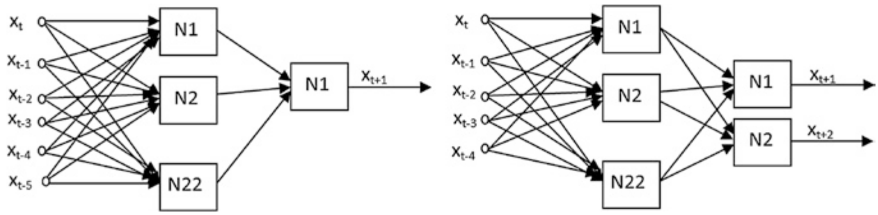

Fig. 9 The structure of the simple neural network

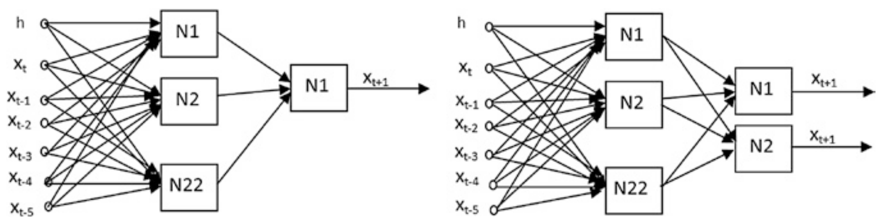

Fig. 10 The structure of the neural network with the ‘hour’ input

that the network receives as input a moving half-hour window every 5 min interval. The time window contains 5 or 6 elements of the time series of intensities traffic flow. The number of neurons in the hidden layer was determined experimentally. The prediction error is adopted as a selection criterion.

The networks have been tested with different numbers neurons in the hidden layer. The best solution was a network containing 22 neurons in the hidden layer. The networks had one or two outputs, that corresponds to the next value (values) of the value of the traffic flow in the analyzed time series.

For each traffic class a different neural network was elaborated. Weights of the neurons in the networks were evaluated by backpropagation using sets of learning sequences.

4.2.2 The Training Sequences and the Learning Process Parameters

To training set contained data collected in July, September, November, January and March. To test the network we have used data from randomly selected days of the remaining months. The length of the measurement period is 1/2 h, in which travelling vehicles are counted, a new count is started every 5 min. This means that inputs provide traffic data to the network describing the mean traffic flow in the past 1/2 h.

For each of the three classes of traffic time series (working days, Saturdays) the same structures of the neural networks were used. Depending on the group the learning sequence consisted of 3000–5000 vectors. For a group of Mon–Fri the training set was the longest, and the shortest was for the Sundays group.

The learning process terminates, when the error mean square (RMS) value is between from 0.022 and 0.034. The neural network learning rate is between 0.5 and 0.9 and momentum $\alpha = 0.4$–0.7. For all neural networks we have used a backpropagation learning method.

4.3 Bayesian Dynamic Linear Model

The approach basing mostly on historical data is sometimes criticized due to neglecting of incoming actual values. So called Bayesian forecasting is devoid of this disadvantage. The main assumptions is that the forecast is expressed in terms of probability distributions. The subjective probabilities represent the current uncertain knowledge and the beliefs of the forecaster.

The model of time series is here a dynamic model. The term dynamic means that all changes in this process happen with the passage of time. If the internal model dependencies are linear, it is called Dynamic Linear Model (DLM) [14]. Depiction as Bayesian indicates that the forecast is based on the past knowledge, but constantly inflowing new information become a part of the knowledge and can significantly influence the forecast. Let's assume D_i denotes the state of knowledge at *i*th moment and F_j denotes the forecast for *j*th moment. Then the process of Bayesian forecasting can be formally expressed as:

D_0	initial information set,
$\left(F_j \vert D_i\right)$	determining the forecast distribution for *j*th moment at *i*th moment, certainly must be j > i,
$D_i = \{I_i, D_{i-1})$	updating the knowledge with the new information acquired at *i*th moment.

The simplest and most widely used is the first-order polynomial model described by following equations (term a ~ N[b, c] means random variable *a* is normally distributed with known mean *b* and known variance c^2).

$$Y_t = \mu_t + \upsilon_t, \upsilon_t \sim N[0, V_t] \tag{7a}$$

Observational equation, here Yt is an observation at time t, it is the sum of series level μt and the observational error (noise) υt,

$$\mu_t = \mu_t - 1 + \omega_t, \quad \omega_t \sim N[0, W_t] \tag{7b}$$

Evolution equation, the series level can be treated as locally constant, but it undergoes to the random walk with evolution error (drift) ω_t. For all time moments s and t errors ν_t and ν_s are independent, ω_t and ω_s are independent, and ν_t and ω_s are independent. Variances V_t^2 and W_t^2 can be known or unknown and constant or variable at every time moment. Initial information is assumed to be normally distributed random variable with known mean and known variance. The details of the implementation of such model depend on the available knowledge concerning both variances.

Let's assume now, that both variances $\left(V_t^2, W_t^2\right)$ are known at each time moment. Then the forecast can be recursively expressed as:

$$m_t = A_t Y_t + (1 - A_t) m_{t-1} \tag{8}$$

where

m_t, m_{t-1}	the mean value of the series at moments t − 1 and t,

$$A_t = R_t / Q_t, R_t = W_{t-1} + W_t, \quad Q_t = R_t + V_t. \tag{9}$$

Here, the forecast (mt—estimated series level value) is a weighted average of the prior level estimate m_{t-1} and the observation Y_t. The value of the weight A_t lies between 0 and 1. The value rather close to 0 means $R_t < V_t$, so the prior mean value of the series is much more informative than the current observation, which is burden with the noise. If this value is close to 1 the prior distribution is diffuse and in such way it is less informative than the observation.

The model here described has a specific feature, which can be recognized as a limitation: when the forecasts more than one step ahead are considered, only the distribution is available instead one sharp value. For the k–step ahead forecast it is expressed as follows:

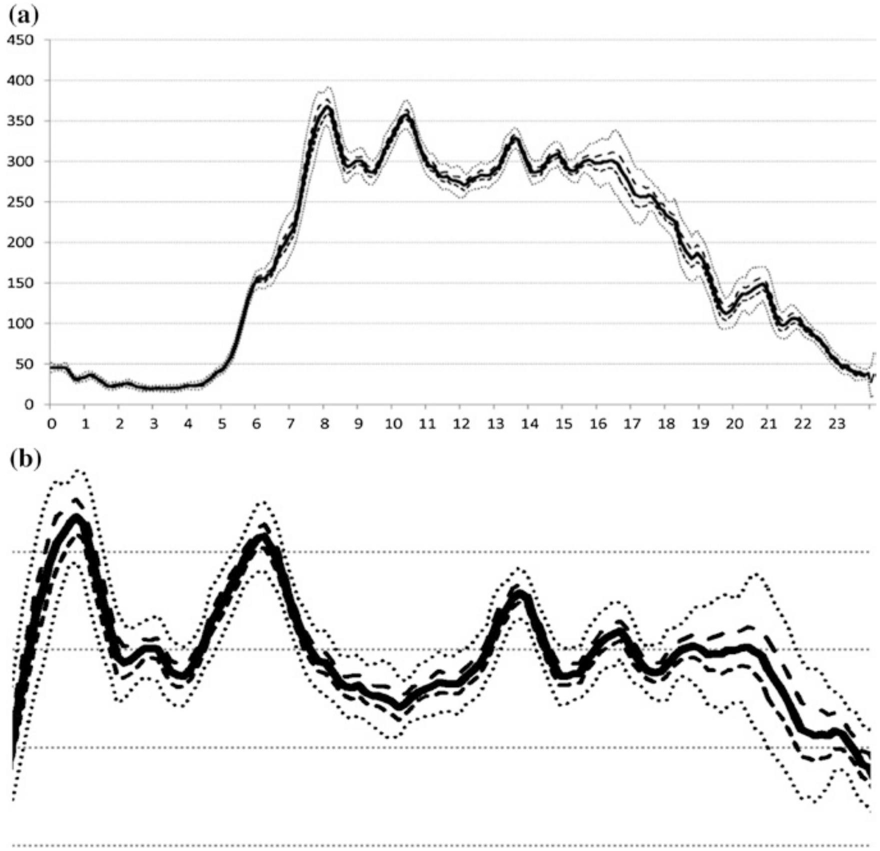

Fig. 11 Distribution of the level, noise error and level drift error: all over the day (**a**) enlarged interval between 7 a.m. and 6 p.m (**b**)

$$(Y_{t+k}|D_t) \sim N\left[m_t, Q_t(k)\right] \tag{9a}$$

$Q_t(k)$ represents the sum of all standard deviations of the level from t to t + k moments and the noise standard deviation at t + k moment, which make the distribution more diffuse:

$$Q_t(k) = W_t + \cdots + W_{t+k} + V_{t+k} \tag{9b}$$

As was mentioned above the values of V_t, W_t are required at every time moment. In this work they were calculated using the historical data.

1. For each week day category (working days, Saturday, Sunday) the calculations were performed separately.
2. For each historical series belonging to appropriate category the moving average with period of 6 was calculated (it responds to half an hour). The specificity of

road traffic variability allows to assume that so obtained averaged series well describes the level evolution.

3. For each series element the difference between average and real values was calculated (it is an equivalent of the noise).
4. For each element of averaged series the difference between the current and previous values was calculated (it is an equivalent of the level evolution).
5. The values of respectively V_t, W_t for each time moment were obtained by calculating the standard deviations of this two data sets.

Figure 11a, b shows the typical distribution of traffic flow level (solid thick line), noise error (thin dotted line) and evolution drift (thin dashed line). This data refer to all Thursdays, but the ratio of the noise error (V_t) to the evolution drift (W_t) is similar in all the cases. As can be easily found the noise error is several times greater than the evolution drift. It means that the weight At is rather close to 0, so rather the prior mean value of the series is taken as the base of the forecast.

5 Research Results

The Bayesian network (BN) and neural network (NN) models without and with 'hour' input (BNH and NNH respectively) have been evaluated. For the comparison purposes also the Bayesian Dynamic Linear Model (DLM) has been joined into tests. The research results are shown in charts and tables. Figures 12, 13, 14 and 15 show the forecast for one step ahead compared with the real series for selected days. The time window is narrowed to period from 7 a.m. to 7 p.m. due to present more details. As can be seen the predicted data fit to the real value of the traffic flow volume. The analysis shows, that the significant errors appear mainly at the rapid changes of the trend.

The best results were obtained for the class of working days. This can be explained by the greater reproducibility of similar traffic flow values, particularly during the morning and the afternoon peaks.

The forecasts quality was checked using standard error indicators: root mean squared error (RMSE), mean absolute error (MAE) and the mean absolute percentage error (MAPE) (Eq. 10a). RMSE and MAE inform about the mean of the absolute deviations of the forecast from the real value. Significant differences between RMSE and MAE indicates the occurrence of some single great deviations. MAPE evaluates the relative accuracy and additionally allows to compare the accuracy of different forecasts obtained regardless to used data range.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^p)^2} \tag{10a}$$

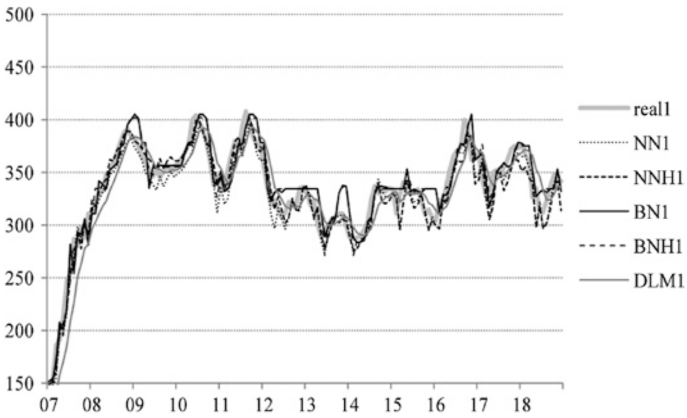

Fig. 12 One step forecast for 15.09.2011 (Thursday)

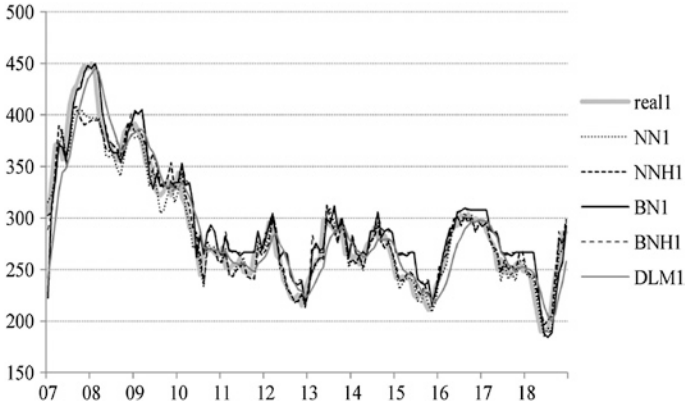

Fig. 13 One step forecast for 15.03.2012 (Thursday)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - y_i^p \right| \tag{10b}$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - y_i^p}{y_i} \right| \tag{10c}$$

where

y_i, y_{ip} real value and the forecast for *i*th moment,
n number of measurements.

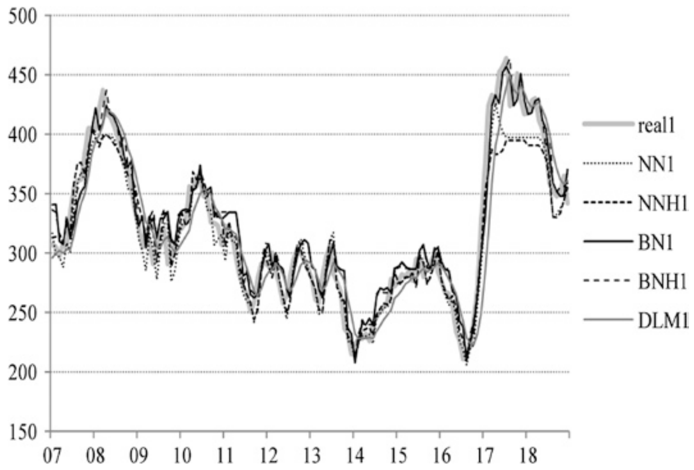

Fig. 14 One step forecast for 09.11.2011 (Wednesday)

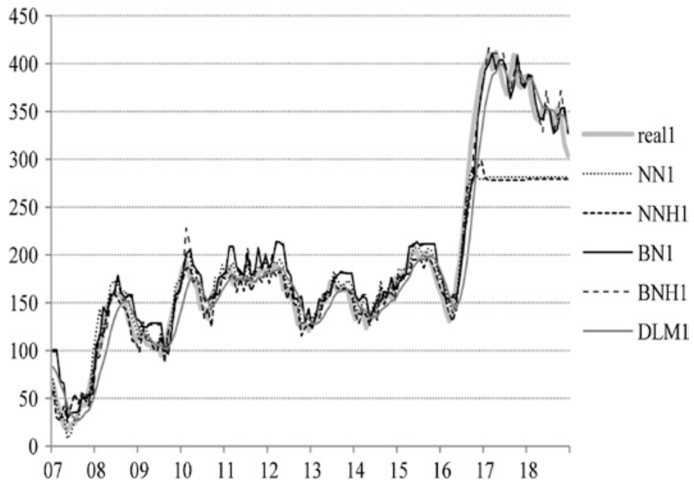

Fig. 15 One step forecast 27.11.2011 (Sunday), the scale is changed due to lower traffic volume

Additionally another error indicator was proposed: for the road traffic management purposes an important data is the knowledge about the trend. Regardless to the certain value of the forecast, the information on whether the traffic will diminish or grow can be very useful. To evaluate this Trend Tracing Indicator (TTI) was introduced:

$$\mathrm{TTI} = \frac{1}{n}\sum_{i=2}^{n}(y_i - y_{i-1})(y_i^p - y_{i-1}^p) \tag{11}$$

Table 4 List of days being tested

Week day	Dates
Wednesday	14.09.2011, 09.11.2011
Thursday	15.09.2011, 15.03.2012, 17.11.2011, 19.01.2012
Sunday	27.11.2011, 04.03.2012

Table 5 Averaged values of error indicators

	MAE1	MAPE1	RMSE1	TTI1	MAE2	MAPE2	RMSE2	TTI2
NN	14.593	0.059	20.204	72.375	21.271	0.088	28.920	63.755
NNH	14.790	0.059	21.526	73.809	20.156	0.086	27.584	62.277
BN	15.871	0.074	21.480	75.354	20.054	0.093	27.638	63.411
BNH	15.023	0.070	18.679	74.953	19.123	0.088	24.804	66.916
DLM	17.013	0.075	23.288	53.384	22.909	0.101	31.481	37.959

The TTI informs about the trend compliance between the actual series and the forecast: the great positive value confirms the good compliance, the value close to 0 or even negative indicates the large discrepancies.

All described methods were applied for some randomly chosen days, which are shown in Table 4. Thursdays and Wednesdays are commonly the subject of study when analyzing the traffic flow because of typicality, Sundays were added due to quite different traffic flow structure.

The obtained values of error indicators were too ambiguous to point out the best method. In such situation the multicriteria ranking was created. Eight criteria were taken into account: MAE, MAPE RMSE and TTI for both outputs. The error values were calculated for the forecasts in the period from 6 a.m. to 8 p.m. These values were averages for all eight days being tested. Such determined error indicators are gathered in Table 5.

All the values in Table 5 except TTI are destimulants, thus the following formula (12a) must be used to conversion the values into 0–1 interval. Value of 0 corresponds to the worse result, 1 corresponds to the best one.

$$w_{ij} = \frac{\max_i(x_{ij}) - x_{ij}}{\max_i(x_{ij}) - \min_i(x_{ij})} \tag{12a}$$

For TTI, which is stimulant conversion is performed using:

$$w_{ij} = \frac{x_{ij} - \min_i(x_{ij})}{\max_i(x_{ij}) - \min_i(x_{ij})} \tag{12b}$$

where

x_{ij} jth error indicator for ith method,
w_{ij} jth rate for ith method.

Table 6 Multicriteria ranking for all examined methods

	MAE1	MAPE1	RMSE1	TTI1	MAE2	MAPE2	RMSE2	TTI2	Final evaluation
NN	1.000	1.000	0.669	0.864	0.433	0.863	0.384	0.891	0.76
NNH	0.919	0.963	0.382	0.930	0.727	1.000	0.584	0.840	0.79
BN	0.472	0.102	0.392	1.000	0.754	0.535	0.576	0.879	0.59
BNH	0.822	0.341	1.000	0.982	1.000	0.871	1.000	1.000	0.88
DLM	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00

The ranking is shown in Table 6. The final evaluation is the average of all eight partial rates.

As can be seen Bayesian networks seem to be most promising in application for prediction. Particularly with the time of day taken into account. A bit worse results were obtained with the use of neural networks, and what is astonishing the inclusion of an hour input did not cause significant improvement of the forecast. It is a sign that some details of neural network structure and learning process should be still refined. This is especially important because the learning of a neural network is much faster than for a Bayesian network. Additionally it happened that some single results of neural network were significantly the best of all. By far the worst forecasts were provided using Bayesian DLM. In that case the lack of the sharp forecast for more steps ahead was crucial. Further study on the process of determination of the both variances (for the level evolution and the noise) can lead to better performance of this model.

6 Conclusion

The performed comparison of neural network and Bayesian network, which were applied for traffic flow prediction has proved that both approaches are here suitable. The neural network seems to be slightly more accurate while tracing the trend, but the difference is of the order of fluctuations. It can be a result of input data discretization, which is required for Bayesian networks. On the other hand Bayesian network provides not only a single predicted value, but the whole distribution of possible states. This information can be very useful in some cases.

The results of the comparison of prediction models using neural networks and Bayesian networks indicates the small differences of their behavior. The average forecast error MAPE for working day for the NN model is 6 %, and for the BN model is 10 %. It also can be due to the discretization. Both models understate the forecasts, the generated predictions are generally worse for holidays and Saturdays. The best results were obtained for the traffic on the working days.

The analysis of the input (measured) data and of the models behavior suggests that additional parameters, such as holidays, vacation, seasons and random accidents should be taken into account.

The obtained results confirm the possibility of the application of both models for selection of the signaling modes within the crossroads and to support of the area-traffic management.

References

1. Chrobok R, Kaumann O, Wahle J, Schreckenberg M (2004) Different methods of traffic forecast based on real data. Eur J Oper Res 155(3):558–568
2. Chen H, Grant-Muller S, Mussone L, Montgomery F (2001) A study of hybrid neural network approaches and the effects of missing data on traffic forecasting. Neural Comput Appl 10:277–286
3. Tan MC, Wong SC, Xu JM, Guan ZR, Zhang P (2009) An aggregation approach to short-term traffic flow prediction. IEEE Trans Intell Trans Syst 10:60–69
4. Vlahogianni EI, Karlaftis MG, Golias JC (2005) Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. Transp Res Part C 13:211–234
5. Pamuła T (2012) Traffic flow analysis based on the real data using neural networks. In: Mikulski J (ed) Telematics in the transport environment. Selected papers. Springer, Berlin, pp 364–371
6. Bolstad WM (2004) Introduction to Bayesian statistics. Wiley-Interscience, Hoboken
7. Srinivasan D, Choy MC, Cheu RL (2006) Neural networks for real-time traffic signal control. IEEE Trans Intell Trans Syst 7(3):261–271
8. Skrobisz C (2010) Bayesian prediction for non-full information on the example of electricity. Folia Pomer Univ Technol Stetin Oeconomica 280(59):99–108
9. Pamuła T (2012) Classification and prediction of traffic flow based on real data using neural networks, Arch Transp 24(4):519–530
10. GeNie package (2014). http://genie.sis.pitt.edu/
11. Cheng J, Druzdzel MJ (2000) An adaptive importance sampling algorithm for evidential reasoning in large Bayesian networks. J Artif Intell Res 13:155–188
12. Heckerman D, Geiger D, Chickering DM (1995) Learning Bayesian networks: the combination of knowledge and statistical data. Mach Learn 20(3):197–243
13. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Roy Stat Soc B 39(1):1–38
14. West M, Harrison J (1997) Bayesian forecasting and dynamic models, 2nd edn. Springer, New York, p 34