IVANA ŠEMANJSKI, Ph.D.
E-mail: ivana.semanjski@fpz.hr
University of Zagreb,
Faculty of Transport and Traffic Sciences
Vukelićeva 4, 10000 Zagreb, Croatia

Intelligent Transport Systems (ITS)
Review
Submitted: Feb. 14, 2015
Approved: Nov. 24, 2015

POTENTIAL OF BIG DATA IN FORECASTING TRAVEL TIMES

ABSTRACT

Travel time forecasting is an interesting topic for many intelligent transportation system (ITS) services. Increased availability of data collection sensors increases the availability of the predictor variables but also highlights the high processing issues related to this big data availability. The aim of this paper is to analyse the potential of big data and supervised machine learning techniques in effectively forecasting the travel times. For this purpose fused data from three data sources (Global Positioning System vehicles tracks, road network infrastructure data and meteorological data) and four machine learning techniques (k-nearest neighbours, support vector machines, boosting trees and random forest) were used. To evaluate the forecasting results they were compared in-between different road classes in the context of absolute values, measured in minutes, and the mean squared percentage error. For the road classes with high average speed and long road segments, machine learning techniques forecasted travel times with small relative error, while for the road classes with low average speeds and small segment lengths this was a more demanding task. All three data sources were proven to have high impact on the travel time forecast accuracy and the best results (taking into account all road classes) were achieved for the k-nearest neighbours and random forest techniques.

KEY WORDS

big data; support vector machines; k-nearest neighbours; boosting trees; random forest; forecasting travel times; data fusion;

1. INTRODUCTION

Travel time information is one of the key quantitative performance indicators of the transportation system as a whole. It is widely used in many intelligent transportation system (ITS) services as dynamic route guidance [1, 2, 3], traveller information system [4, 5, 6] or traffic management system [7, 8]. Another characteristic of travel time information, which distinguishes it from other data on the traffic flow, is its level of relevance and understanding among different groups of the stakeholders such as decision makers, transport system users, transportation planners, etc. [9]. In that context, it is the most widely used parameter as the relevant one when comparing different transportation modes [10]. It is also one of the highest costs of transportation, and travel time savings are often the primary justification for transportation infrastructure improvements [11].

In literature, different approaches to travel time forecasting can be found [6, 7, 12]. Yu et al. use support vector machines (SVM) to predict bus arrival times at bus stations [5], Zong et al. apply a genetic algorithm to forecast daily commute travel times in Beijing [13] and Simroth uses a nonparametric distribution-free regression model [14]. Regarding the data sources used for the travel time forecasting these are mainly global positioning system (GPS) based data [14, 15, 16], survey data [13] or data from different types of road detectors [6, 7]. Multiple data sources are rarely used [5].

When it comes to the factors affecting travel time, literature review identifies free flow travel speed, occurrence of incident situations, holidays or other uncommon events, congestion level and weather conditions [17, 18].

Nevertheless, studies have so far limited scope in the context of transferability as they use limited number of data sources (and explanatory variables) while applying travel time forecast on a very limited geographical area (just one road category) or quite a specific group of vehicles (e.g. buses or taxi service). There is very little effort given in providing a systematic overview of advanced travel time forecasting approaches which have the capability of dealing with big data collected from multiple sources.

This paper uses three data sources (road network data, GPS and weather forecast data) to analyse the potential of four supervised machine learning techniques (SVM, *k*-nearest neighbour, boosting trees and random forest) in forecasting travel times on five different road categories. The attempt is to give a systematic overview of the comparable results for travel time forecast as a good reference point for future research in this field as well as a growing number of developing travel time forecast-based services and applications.

2. DATA SET DESCRIPTION

One would think that today's availability of different data sources makes travel time forecasting easier. And indeed, it allows us to collect very diverse data sets but it also brings new challenges regarding the storing and processing of the mobility data. In the traditional trip diaries, information on the weekly trips for one person represents about 20 kilobytes of data. Today, just the GPS tracks of weekly trips for one person range between 10-20 megabytes. Other data sources just add up to this number. This brings new challenges in travel time forecasting but it also gives a higher level of details available as well as possibilities for travel time forecasting in real time.

This paper uses three data sources. Spatiotemporal data are collected via GPS vehicle tracks. The data on the road network infrastructure, on which spatiotemporal GPS data are matched, are collected from the road network database. And the meteorological data that are collected by the network of sensors are provided by the national Meteorological and hydrological service. The complete database fused from three sources, based on the location and time data, contained 39 gigabytes of data.

2.1 Spatiotemporal data

Spatiotemporal data are dynamically collected from 300 probe vehicles that used the road network in the City of Zagreb (Croatia). The installed GPS devices sent data via mobile network to the central database. These data included:

- Vehicle ID;
- Information on location (*x* and *y* coordinates);
- Vehicle speed;
- Vehicle course;
- Logging time.

In the database, these data were joined with the attributes on the day of the week, special events (holidays, scheduled traffic flow disturbances, school days, etc.), historical average of speeds recorded for the same road and standard deviation of historical speed records.

2.2 Infrastructural data

Infrastructural database contained information on the complete road network including:

- Road length;
- Road name;
- Start and end coordinates of every road segment;
- Direction (one-directional street or not);
- Traffic modes using the road;
- Number of traffic lanes.

Based on the infrastructural data (particularly the start and end coordinates of every road segment), unique vehicle IDs, logging times and the observed travel times were joined to every road segment.

2.3 Meteorological data

Meteorological data were obtained from the sensor network and provided by the national Meteorological and hydrological service. This data set included information on:

- Air temperature;
- Ground temperature;
- Pavement condition (wet or not);
- Snow (falling or not and the thickness of the new snow on the road);
- Rain;
- Humidity;
- Wind;
- Horizontal visibility.

3. METHODS

Before conducting the travel time forecast, based on the spatiotemporal and infrastructural data, a hybrid approach was used to classify the roads. This was done based on the hypothesis that by including information on actual vehicles movements along the road far better insight into similarities between different roads and way they are used can be obtained than by merely relying on infrastructure characteristics (as in traditional road classification approaches). For this purpose we used multiple regression and factor analysis to identify parameters that most influence the road classification and these were historical speed average, standard deviation of the historical speed, length of the road segment and count of the vehicles that used this road segment. *Figure 1* and *Figure 2* give an overview of five road classes based on the most distinctive variables. More detailed description of this procedure can be found in literature [18].

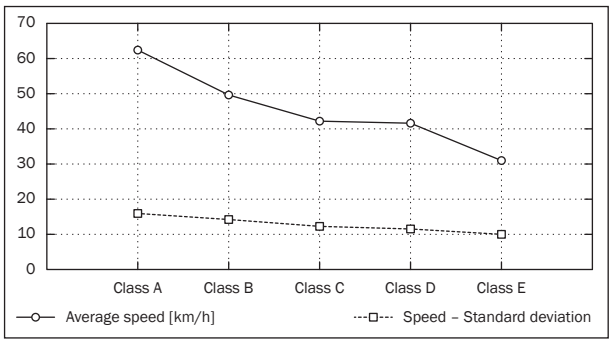

Figure 1 – Summary of road class differences based on the average speed in km/h and standard deviation of the speed

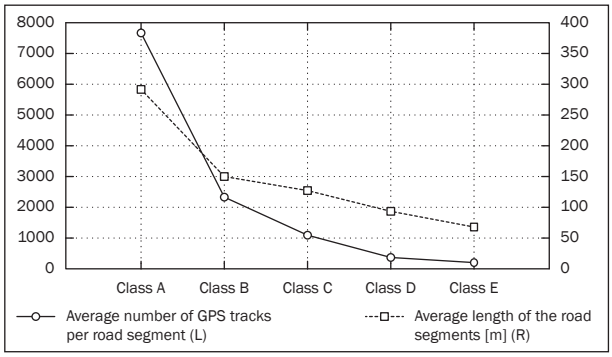

Figure 2 – Summary of road class differences based on the average length of road segments (in metres, right y-axis) and average number of records per road segment (count, left y-axis)

For the travel time forecasting, for each of the five road classes, four supervised machine learning regression techniques (*k*-nearest neighbours, support vector machines, boosting trees and random forest) were used.

3.1 The *k*-nearest neighbours

The *k*-nearest neighbours (kNN) is a non-parametric method that can be used for both regression and classification tasks [19, 20, 21, 22]. As our goal is to forecast travel time (continuous dependent variable) the focus will be primarily on the kNN regression. The kNN regression is a technique that based on the *k* closest training examples in the feature space gives the output forecast as the property value for the object (average of the values of its *k* nearest neighbours). The feature space is created based on the independent variables which can be either continuous or categorical [22]. More detailed description of the basic principles of the kNN regression can be found in literature [23, 24].

One of the main challenges when using kNN technique is the choice of *k* (neighbourhood size) as it can strongly influence the quality of forecast [25, 26]. For any given problem, a small value of *k* will lead to a large variance in predictions and a large value may lead to a large model bias. Literature suggests no exact solutions for finding the optimal size of *k* but rather to use the heuristic approach [27, 28]. For this purpose the cross-validation technique was used. Cross-validation divides the data sample into a number of *v* folds (randomly drawn, disjointed sub-samples or segments). Then, for a fixed value of k, the kNN technique is applied to make the forecast on the v-th segment (others are used as examples) and to evaluate the error. This process is then successively applied to all possible choices of *v* and various k. The value achieving the lowest error is then selected as the value for *k* [29, 30]. In travel time forecasting the original data set was divided into quarters, three of which were used for learning and one for testing. The distance between neighbours was calculated based on the squared Euclidian distances as defined by Equation (1) where *p* and *q* are the query point and a case from the examples sample, respectively.

$$D(p,q) = (p-q)^2 \tag{1}$$

The 10-fold cross validation was used to select the value of *k* for each road category separately. The search range for *k* was from 1 to 50, with the increment of one.

3.2 Support vector machine

Support vector machines (SVM) are supervised learning models with associated learning algorithms that can be used for both classification and regression analysis. For example, the SVM can be used to learn polynomial, radial basis function (RBF) and multi-layer perceptron (MLP) classifiers [31]. In principle, SVM works with the concept of decision planes that define decision boundaries (separate between a set of objects having different class memberships). In the multidimensional space these decision planes are hyperplanes that are, in a sense, equidistant from the *n* sets of objects. In the simplest case, these are linearly separable sets of objects, but in practice this is often not the case and the kernel functions are used. For a reader interested in more details about SVM technique a detailed overview is given in literature [32, 33, 34].

For the travel time forecasting, the SVM regression was used where the relationship between the independent (*y*) and dependent variables (*x*) is given by a deterministic function *f* plus the addition of some additive noise as defined by Equation 2:

$$y = f(x) + noise \tag{2}$$

The SVM model was trained to find a functional form for *f* that can correctly forecast travel times for the new cases with which the SVM had not been presented before. This was done by the sequential optimization of an error function defined by Equation 3:

$$\varepsilon = \frac{1}{2} w^T w + C\sum_{i=1}^{N} \xi_i + C\sum_{i=1}^{N} \xi_i^* \tag{3}$$

where:

- C – capacity constant;
- w – vector coefficients;
- ξ – parameters for handling non-separable data (inputs);
- N – number of training cases.

The error function was minimized in regard to:

$$w^T \phi(x_i) + b - y_i \le \varepsilon + \xi_i^* \tag{4}$$

$$y_i - w^T \phi(x_i) - b_i \le \varepsilon + \xi_i \tag{5}$$

$$\xi_i, \xi_i^* \ge 0, i = 1, \ldots, N \tag{6}$$

where:

- ϕ – kernel used to transform data from the input to the feature space;
- b – constant.

For the kernel we used radial basis function defined by Equation 7:

$$\phi = \exp\left(-\frac{|x - x_i|^2}{2\sigma^2}\right) \quad (7)$$

where:

$|x–x_i|^2$ – squared Euclidean distance between the two feature vectors;

σ – free parameter.

3.3 Boosting trees

Boosting Trees (BT) evolved from the application of boosting methods to regression trees. The main idea is to compute a sequence of simple trees, where each successive tree is built for the prediction residuals of the preceding tree and in this way, by combining many weak learners it forms a single strong one [35]. In principle, BT has a form (Equation 8) that can be trained by optimizing the scalar α_t and the weak learner $h_t(x)$ at each iteration t.

$$H(x) = \sum_t \alpha_t h_t(x) \quad (8)$$

Before the training begins, a non-negative weight (w_i) is assigned to each data sample (x_i). The weighted expansion of simple regression trees will be computed and the prediction residuals fitted on all the preceding trees. In this procedure, for an independently drawn sample of observations, each individual tree is fitted to the residuals computed thereby introducing a certain degree of randomness into the estimation procedure and avoiding overfitting [36, 37, 38].

For travel time forecasting the maximum number of nodes for each individual tree in the boosting sequence is defined to be three to create a single split or partitioning of the training sample at each step. As empirical studies from literature suggested [37], for the weighted expansion of simple regression trees, the weight (w_i) assigned before training is set to be 0.1.

3.4 Random forest

The Random Forest (RF) technique is built upon the idea that the use of different learning models increases the accuracy of forecasting. It is basically built as a large collection of decorrelated decision trees, each capable of producing a response when presented with a set of predictor values [39].

For regression problems [40, 41], as travel time forecasting, the tree responses are averaged to obtain an estimate of the dependent variable as described by Equation 9:

$$RF\ forecast = \frac{1}{k}\sum_{K=1}^{K} K^{th}\ \ tree\ response, \quad (9)$$

where:

k – index that runs over the individual trees in the forest.

As other, previously mentioned methods, RT can also be used for classification tasks. In this case the response takes the form of a class membership, which associates a set of independent predictor values with one of the categories present in the dependent variable [42, 43].

For travel time forecasting, the stopping condition for the random tree technique of 10 levels and the maximum of 100 nodes was defined. The stopping conditions were equal for every road class.

4. RESULTS

To evaluate the results of the travel time forecasting techniques, the representative roads (test sample) for every road class was selected. *Table 1* gives the names for the selected roads and *Figure 1* shows their spatial distribution in the traffic network of the city of Zagreb. When selecting the test samples, we tried to select the roads that were located in different parts of the city, but had a high frequency of GPS tracks (were often used) and had a number of linked segments classified to belong to the same road class.

4.1 The *k*-nearest neighbours results

To determine the size of the neighbourhood for every road class, the cross validation method was used. This allowed the examination of the cross validation error for every road class separately (*Figure 4*). In general, for every road class, the error was growing together with the size of the neighbourhood. For the road classes with high average speeds and long road segments (class A, B and C) the error initially grew fast but soon got stabilized, while for the road classes E and D, it grew almost linearly across a very small range of values.

When comparing the cross validation error across the road classes, the lowest overall error was for road class E. For all the rest, except for road class A, the error remained below the 0.2 margin.

For the size of the neighbourhood that yields the lowest cross validation error the travel time forecast was made. *Figure 5* shows the relationships between the observed values of the travel times, forecast values and the residuals for the kNN technique. In general, the highest accuracy of the forecast is when the raw data (the points) are aligned on the diagonal of the three-dimensional plane in the approximate height of the middle of the residual axis (z). This would mean that the observed value is equal to the forecast value, and respectively that the residual is equal to zero (middle of the z-axis). Taking into account this interpretation, one can see that the best results are achieved for road classes B, C, and D. The forecast for road class A is equally dispersed around the diagonal line and for road class E it tends to overestimate the values for the long travel times, but this overestimate is never higher than two minutes.

Table 1 – Selected representatives for the travel time forecasting for each road class

Road class	Kind of movements served	Road name
Class A	Long distance movements	Zagrebačka avenija, Slavonska avenija
Class B	Long to medium distance movements	Avenija Dubrovnik, Selska cesta, Avenija grada Vukovara
Class C	Medium distance movements	Savska cesta, Kralja Zvonimira, Maksimirska
Class D	Medium to short distance movements	Mirogojska cesta, Prisavlje
Class E	Short distance movements	Harambašićeva, Jordanovac, Hrvatskog proljeća

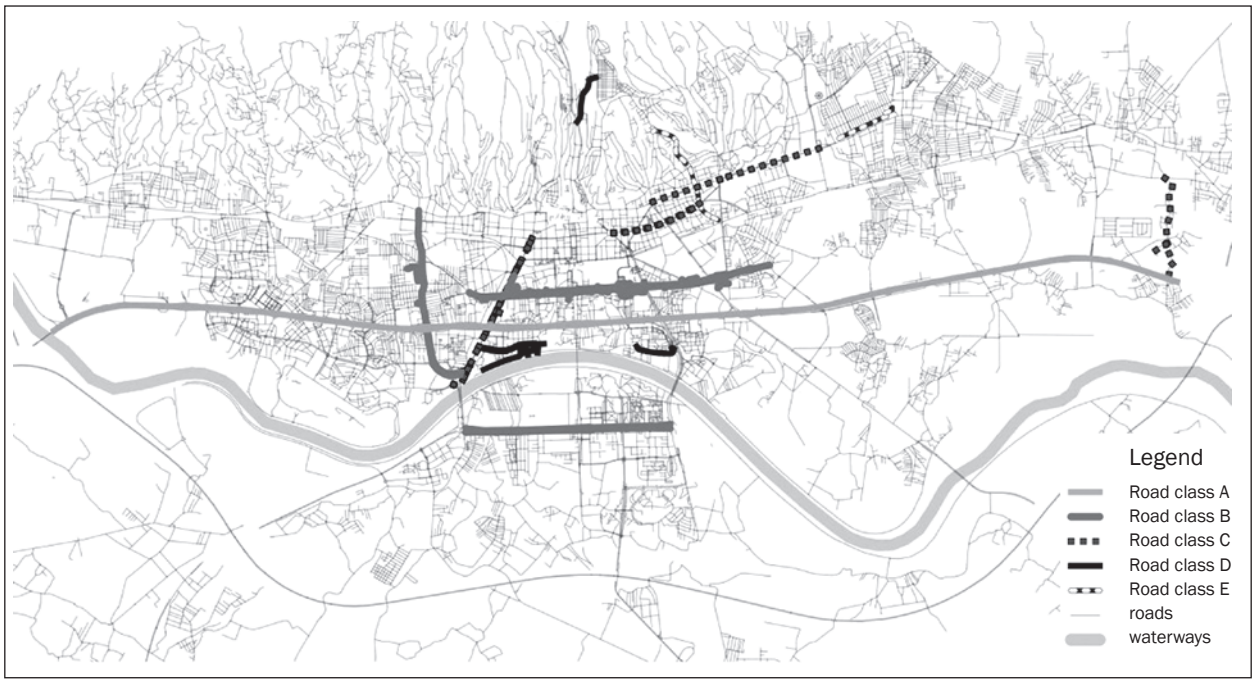

Figure 3 – Spatial distribution of the selected test samples for each road category

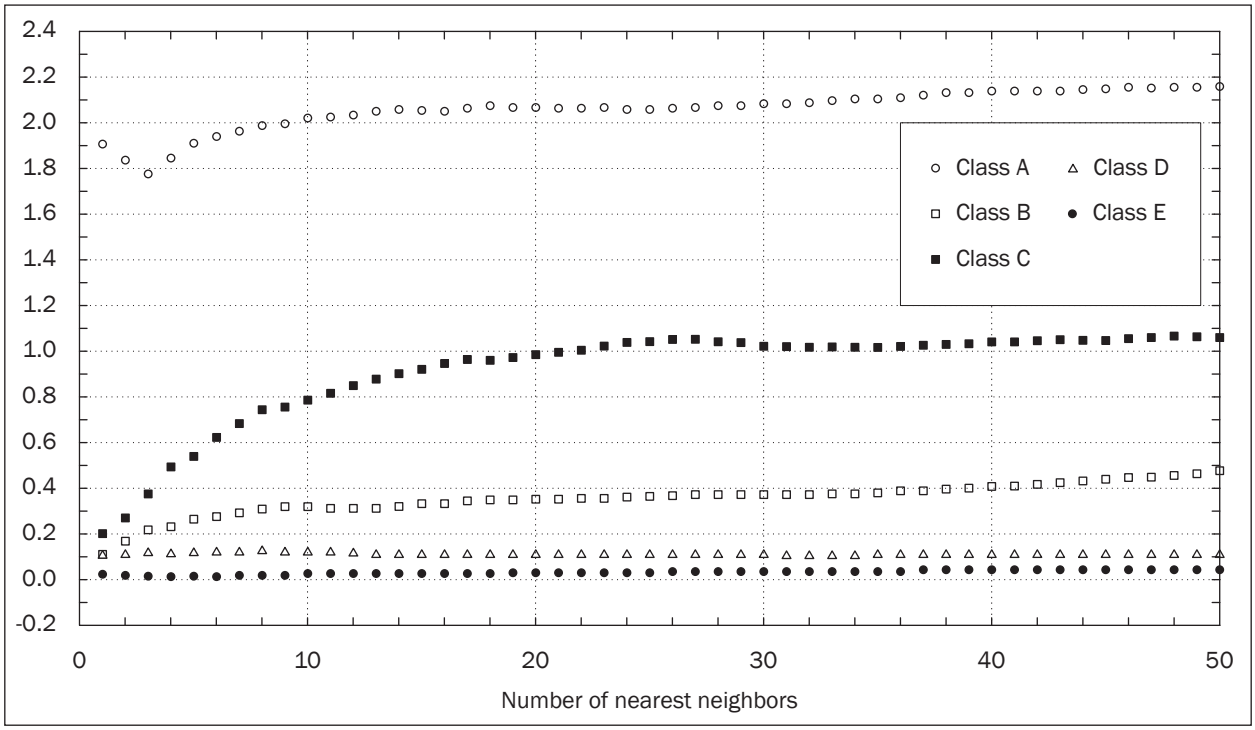

Figure 4 – Cross validation error values for all road classes

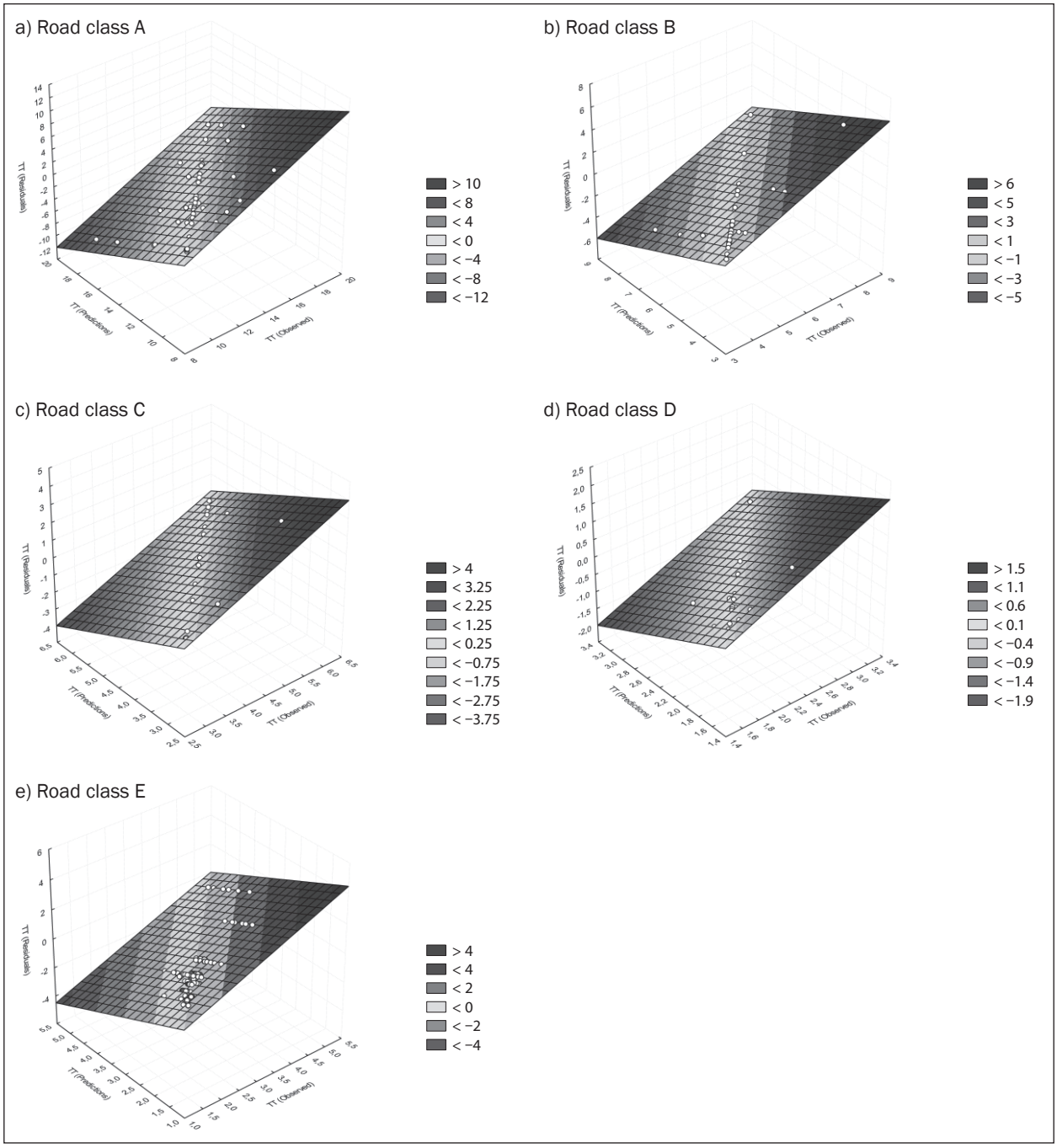

Figure 5 – Comparison of the observed and the forecast travel time values in minutes for all road classes (kNN technique)

4.2 Results of the support vector machines

For the support vector machines, the best results were achieved for road classes C, D and E *Figure 6*. For road class A, SVM actually gave a very small range of the forecast values. This resulted in the equally distributed range of residual values, but very limited sensitivity in the context of actually observed values of travel times. For road class B, the forecasts were quite good, with just a few extreme values (residual between 2 and 4 minutes).

In general, the SVM resulted in the smallest range of residual values for road classes D and E and the largest for road class A. When compared with the kNN results, SVM showed less forecasting sensitivity and obtained lower prediction accuracy for all road classes.

4.3 Boosting trees results

For the boosting tree technique, we firstly reviewed how consecutive boosting steps improved the accuracy (quality of the forecasting model) for the randomly selected training data and testing data (*Figure 7*). For every road class, the forecast actually improved with

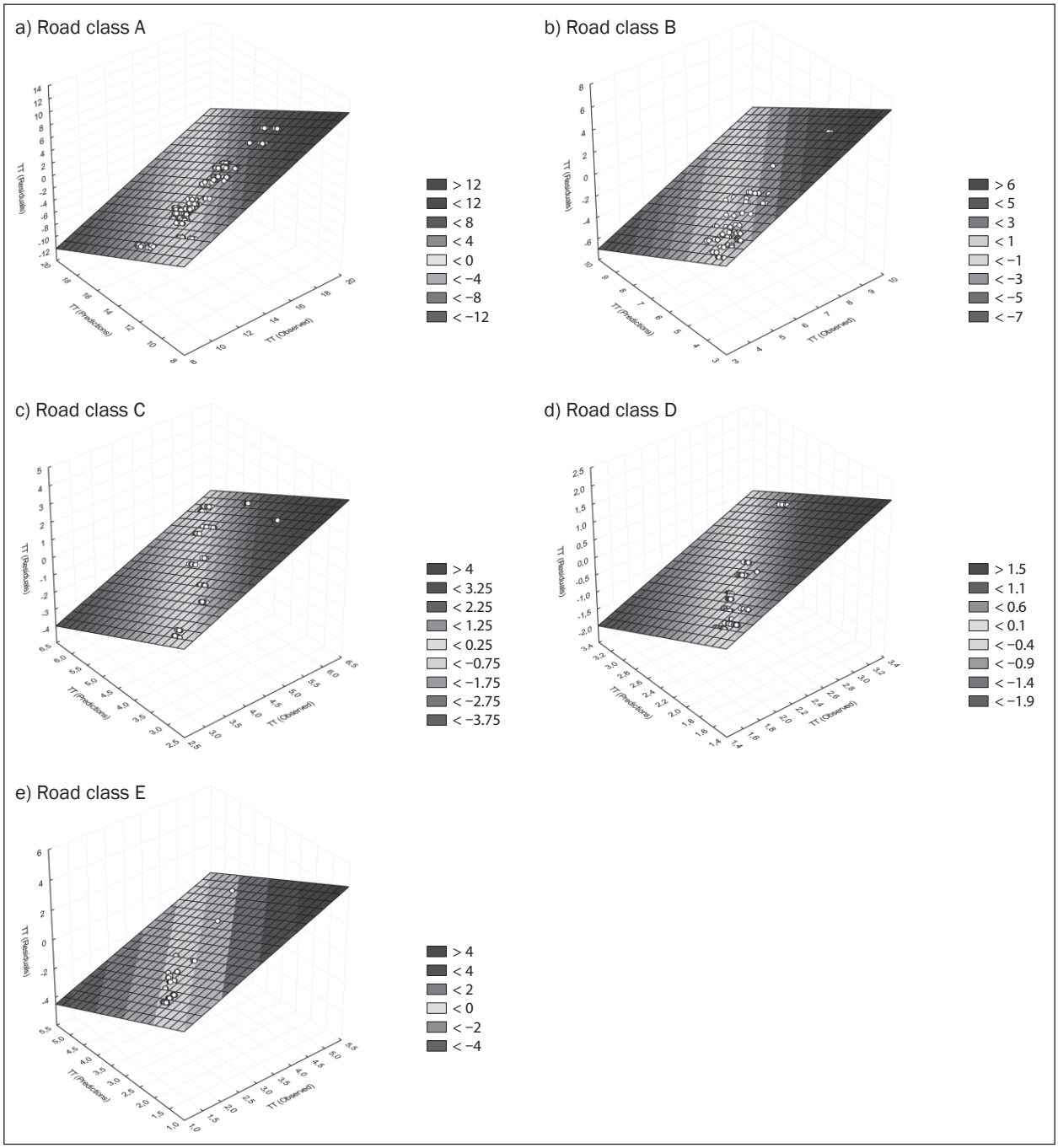

Figure 6 – Comparison of the observed and the forecast travel time values in minutes for all road classes (SVM technique)

the addition of the consecutive boosting steps and the lowest error was achieved in the final step. However, road classes A, C and E yield the lowest difference between the average squared error for the training and test data. Nevertheless, for none of the road classes the overfitting occurred.

When examining the independent variables importance for every road class, one can notice that the members of the set of just eight independent variables was among the five most important variables for the travel time forecast for every road class (*Table 2*). Equal number of times the most important predictors were the length of the road segment and the standard deviation for the speed values. It is interesting to notice that the variables from all three input data sources are among the most important predictors.

Evaluating the forecast results (*Figure 8*), one can see that the boosting trees yield good and stable forecast for all road classes. Only for road class C it happened that for the longer travel times the boosting tree technique overestimated the travel times and for the shorter travel times it underestimated the travel times, but in both cases this was never for more than +/- 2 minutes. In general, it achieved better results than SVM method.

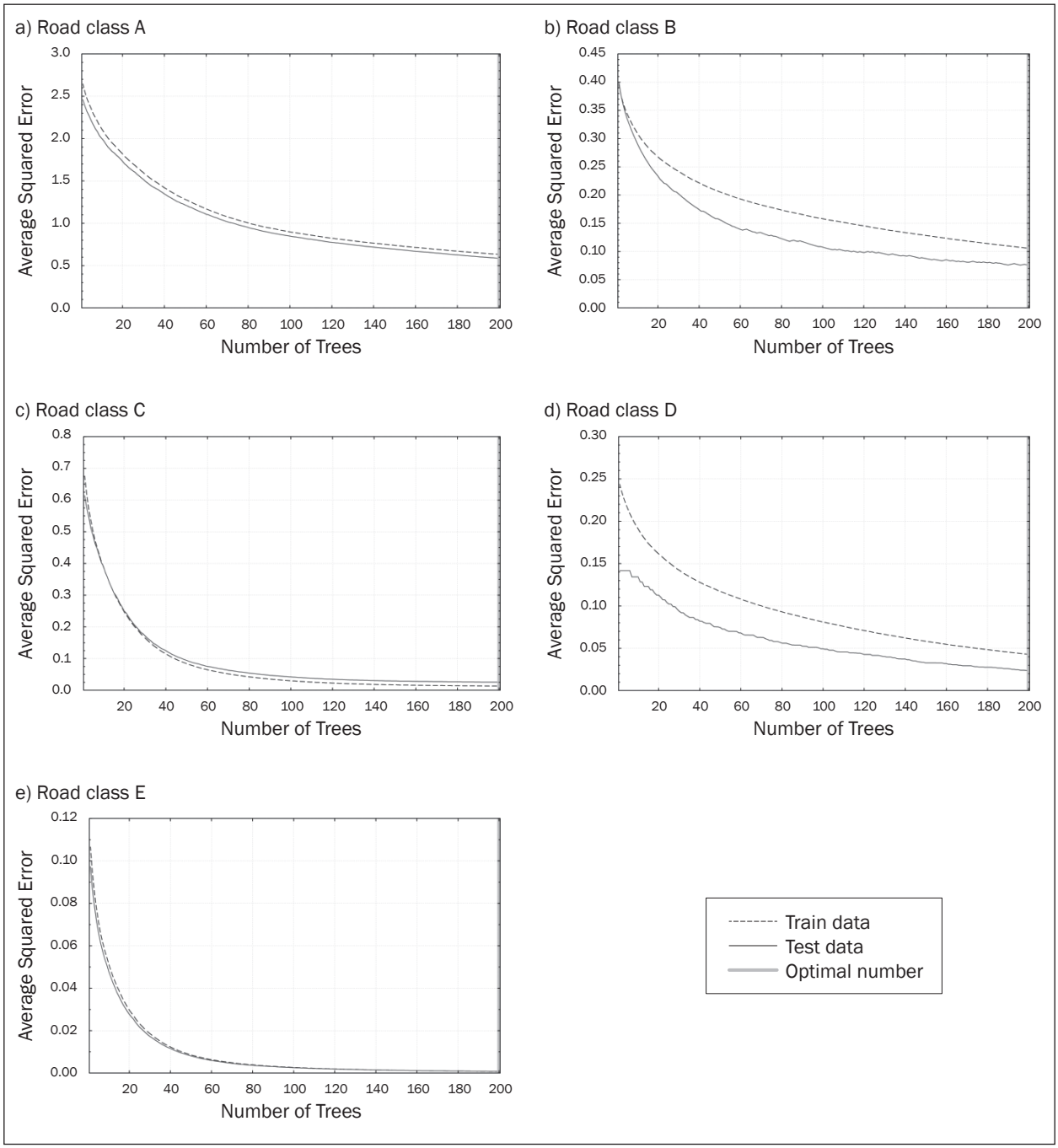

Figure 7 – Boosting trees average square error vs. number of trees

Table 2 – Rank of independent variables importance in regard to every road class (Boosting trees)

	Class A	Class B	Class C	Class D	Class E
Length	1	2	2	1	3
Count	2	–	3	5	2
Speed st_dev	3	1	4	3	1
Hour_of_day	4	4	–	2	5
Coordinate	–	3	1	4	4
Horizontal visibility	5	–	–	–	–
Precipitation	–	5	–	–	–
Snowing	–	–	5	–	–

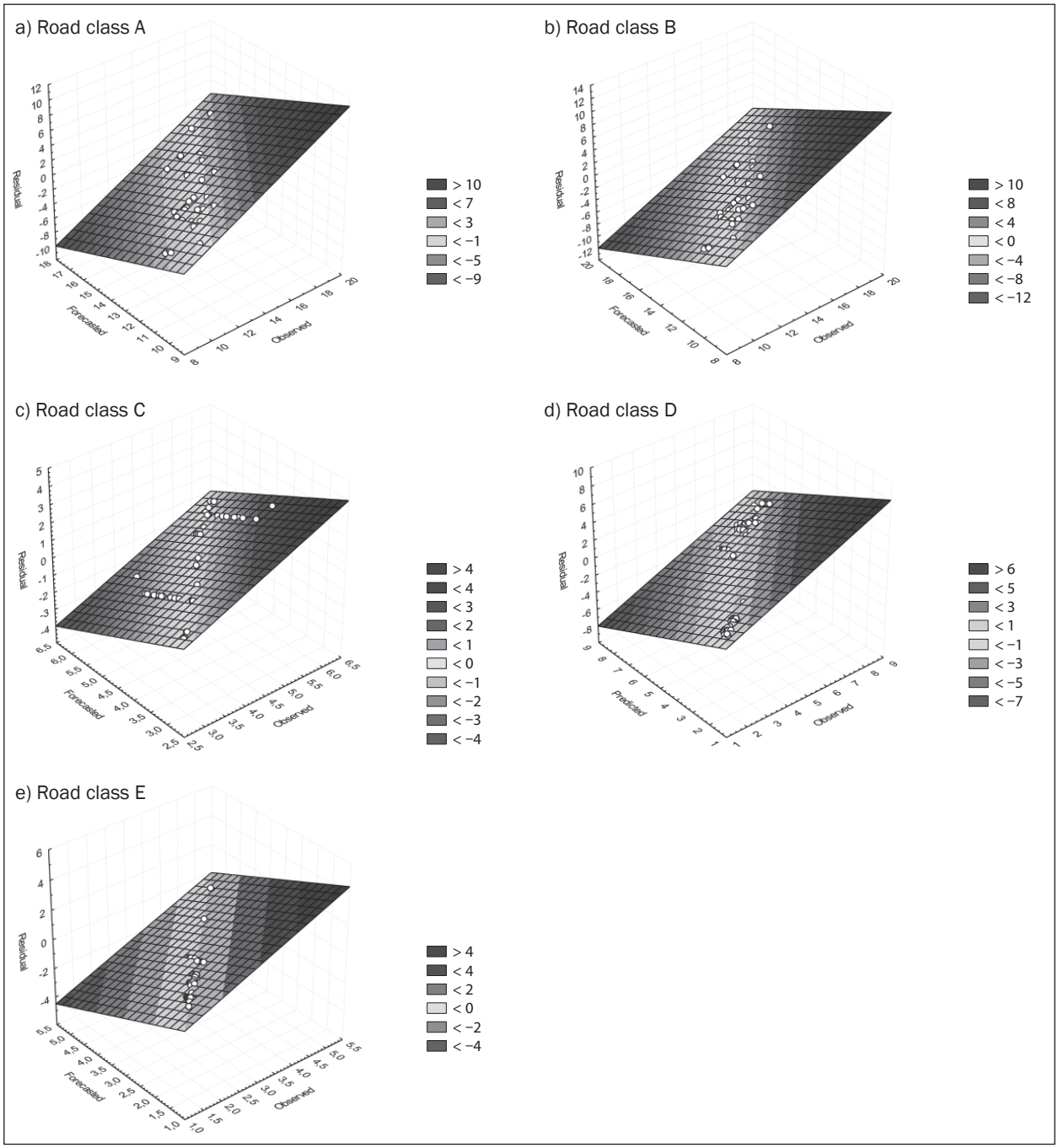

Figure 8 – Comparison of the observed and the forecast travel time values in minutes for all road classes and boosting trees technique

4.4 Random forest results

For the random forest technique, it was first reviewed how the averaged squared error changed over the entire training cycles for each road class (*Figure 9*). In general, the smaller the road segments, speed and changes of speed for the road class the lower was the error. When considering the importance of predictor variables, the set that contained five most important variables for every road class consists of the list of just seven independent variables (*Table 3*). Once again, these seven variables included information from all three sources (GPS tracks, meteorological database and road network infrastructural data).

Considering the relationship between the observed and forecast values for the random forest technique (*Figure 10*), one can notice that the highest accuracy of the travel time forecast is achieved for road classes C, D and E. For road class A the residuals are fairly evenly distributed and for road class B only extreme value of the long travel time was overestimated for 3 minutes.

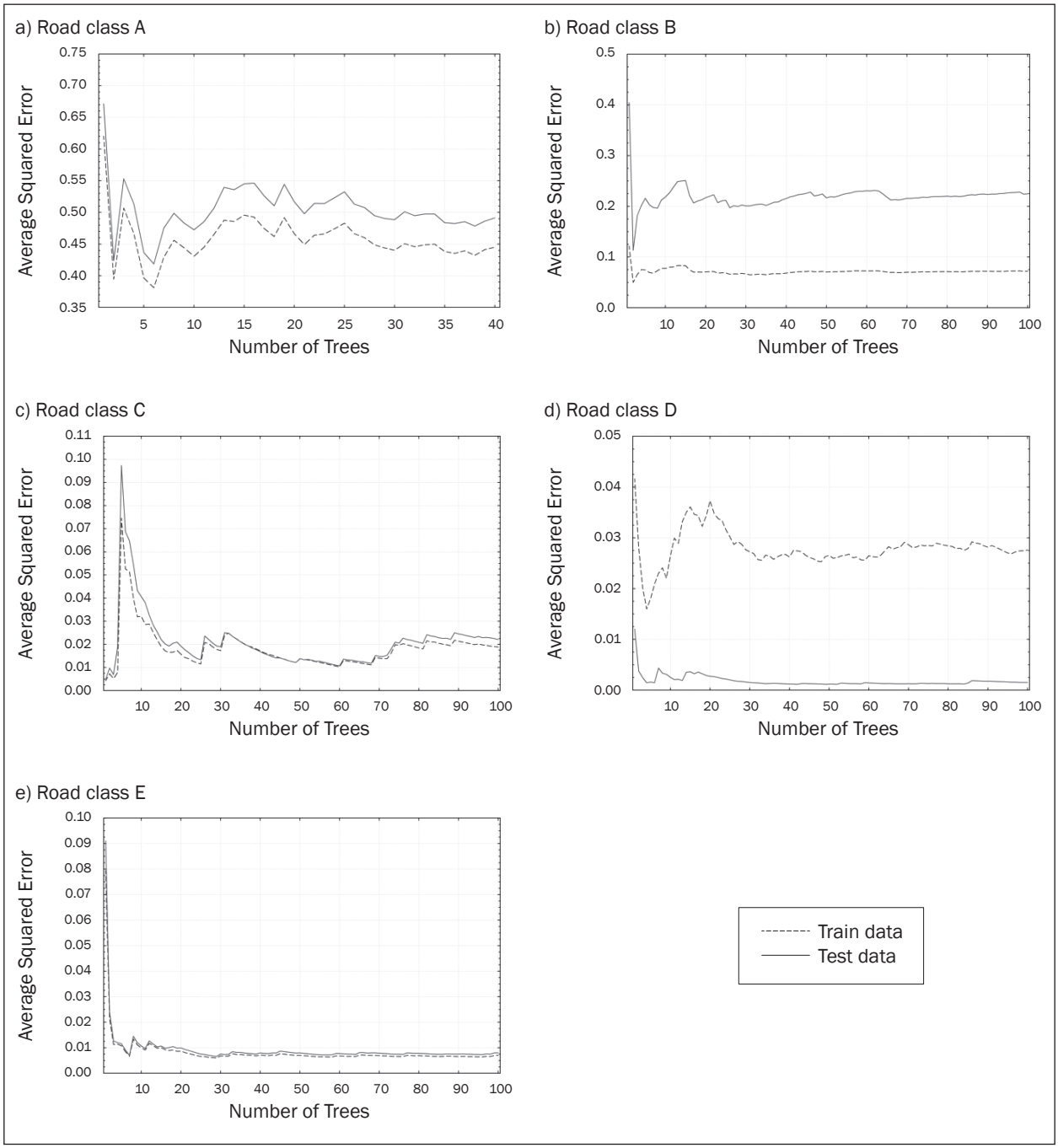

Figure 9 – Random trees average square error vs. number of trees

Table 3 – Rank of independent variable importance in regard to every road class

	Class A	Class B	Class C	Class D	Class E
Length	1	1	2	2	2
Count	2	–	3	5	1
Speed st_dev	3	2	4	3	3
Hour_of_day	4	4	5	1	5
Coordinate	–	3	1	4	4
Temperature of wet pavement	5	–	–	–	–
Air temperature	–	5	–	–	–

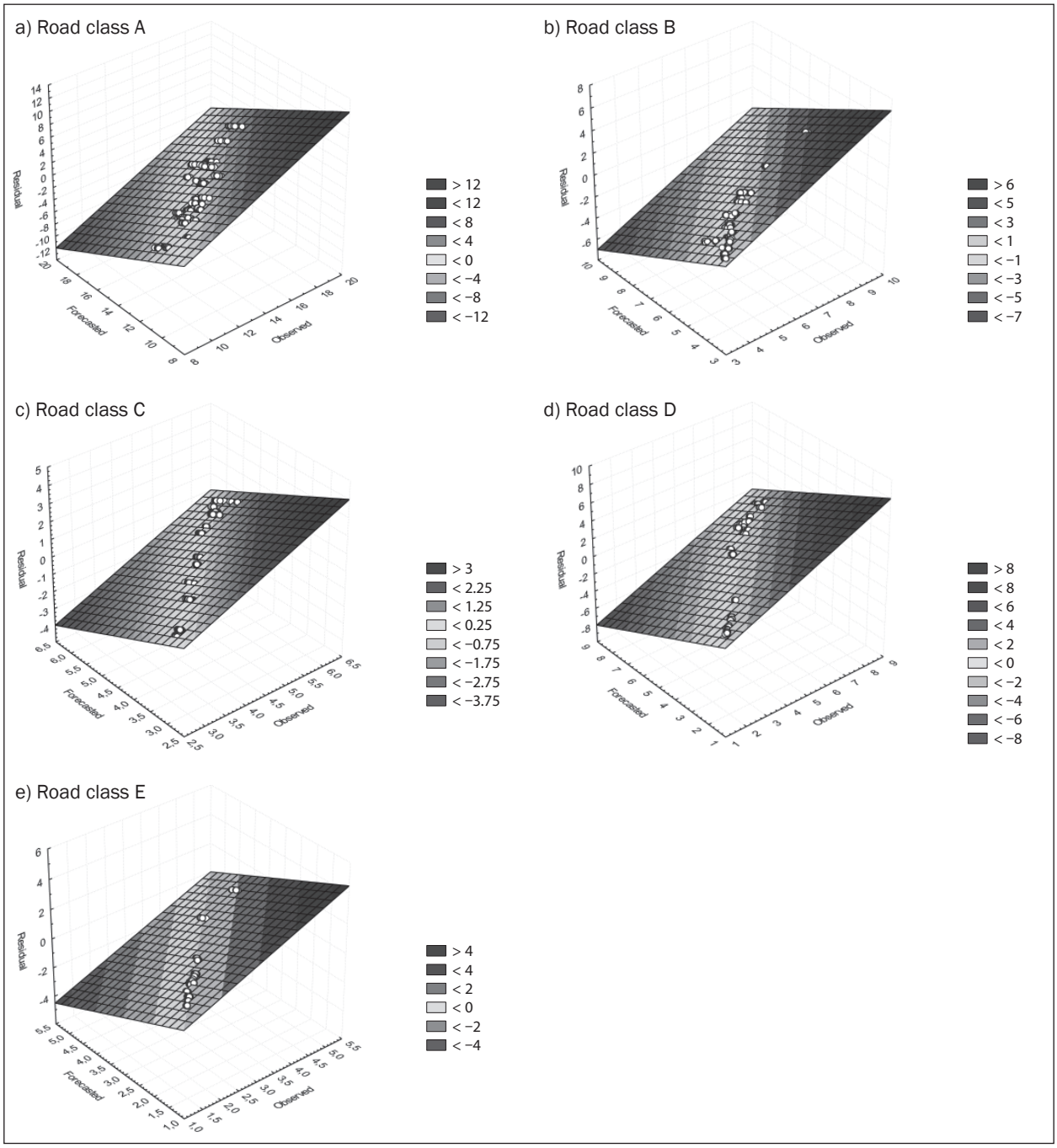

Figure 10 – Comparison of the observed and the forecast travel time values in minutes for all road classes and the random forest technique

4.5 Mean absolute percentage error of the travel time forecast

To gain a better insight in the travel time forecast among different techniques and road classes the mean absolute percentage error (MAPE) was calculated as defined by Equation 10.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \quad (10)$$

where:

A_i – observed value,

F_i – forecast value.

The largest MAPE is for road class E and the support vector machines technique (*Figure 11*). Although, when measured in minutes, the residuals were the lowest, but when taking into account the relative ratio (regarding the road length and respectively the travel time needed to travel across the whole road), this error was the highest. Overall, road class E has the highest difference in the forecast error among different techniques and road class C has the smallest.

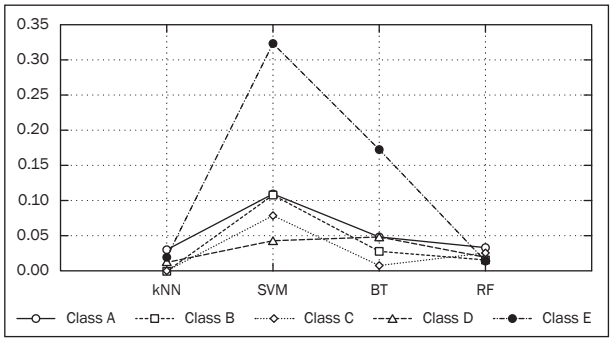

Figure 11 – MAPE values for travel time forecast for all road categories and kNN, SVM, Boosting trees and Random forest technique

5. DISCUSSION AND CONCLUSION

It was found that for the roads with high speeds and long road segments the supervised machine learning techniques analysed in this paper (*k*-nearest neighbours, support vector machines, boosting trees and the random forest) will yield the most uniform results. More demanding in this context, are the local and arterial roads. Here the impact of adequate travel time forecasting technique is crucial.

If one cannot afford a complex travel time forecasting system that would take into account different road categories then, based on our results, it would be advisable to use the *k*-nearest neighbour technique or random forest as two of these yield the lowest overall error across different road categories. Nevertheless, it should be noted that the kNN technique was more computationally expensive. One reason for this is the use of the cross validation method to estimate the optimal size of the neighbourhood, but indeed this step does not need to be repeated every time when calculating the travel time forecast (rather just the first time and then this value should be stored and recalled before every forecast calculation). Another detail that is worth noting, regarding the kNN method, is that it will gain the highest accuracy for the smallest sizes of the neighbourhood (in real time forecasting this would often mean a short forecasting period). Therefore, one could define the error margin for the forecasting quality (e.g. 5% error) and accept larger sizes of the neighbourhood, where the cross validation error is still below the defined margin.

If one would design a travel time forecasting system that would take into account different road classes, then for the high speed roads with long road segments (like highways) and streets with the average speed around 50km/h and average length of the road segment longer than 150 m we would advise the kNN technique (alternatively the random forest). It is important to notice that although the kNN gave the largest absolute error for road class A, when placed in the relative context (in regard to the overall travel time and respectively to the road length) it is actually a very small MAPE value. For the roads with the average speed close to 40 km/h the kNN and the boosting trees yield the highest accuracy. For the local and residential roads this was the case with the random forest technique.

Regarding data sources, it is worth considering multiple data sources as in our case all of them, the GPS data, the road network data and meteorological information had high importance when forecasting the travel times. Nevertheless, one should be selective when choosing among different available variables from these sensors as not all are equally important. Among the GPS data the most useful were proven to be the information on the standard deviation of the speeds for the given road together with the time of the day, the information on how often this road is used and what was the position of the vehicle along the road. Regarding the road network data the most interesting was the information on the average length of the road segments (e.g. how often the traffic flow is interrupted). For the meteorological data, information on the horizontal visibility and precipitation (is it snowing, raining and is the road surface wet and frozen). Meteorological information also had higher impact on the travel time forecasting for the roads with higher speeds as here they were influencing the travel times more severely.

Overall, the supervised machine learning techniques approach has proven to be useful when dealing with multiple data sources for the travel time forecast. All techniques were fairly sensitive to different spatio-temporal conditions among different road classes and different weather conditions. Although, for the road classes with the high average speeds and long road segments the absolute travel time error (measured in minutes) was almost always the highest, it is important to notice that simultaneously these travel times were the longest. When looking in the relative context of overall travel times, errors for these road classes are actually the smallest and the more demanding forecasting is required for road classes with small road segments and low average speeds (residential and local streets).

Doc. dr. sc. **IVANA ŠEMANJSKI**
E-mail: ivana.semanjski@fpz.hr
Sveučiliste u Zagrebu, Fakultet prometnih znanosti
Vukelićeva 4, 10000 Zagreb, Hrvatska

POTENCIJAL VELIKIH SETOVA PODATAKA U PROGNOZIRANJU VREMENA PUTOVANJA

SAŽETAK

Prognoziranje vremena putovanja je ključan element mnogih usluga u sklopu inteligentnih transportnih sustava (ITS). Povećana dostupnost raznih osjetilnih uređaja pozitivno utječe na dostupnost prediktorskih varijabli potrebnih za prognoziranje, ali i dodatno naglašava probleme vezane uz

zahtjevno procesiranje ovih velikih setova podataka. U ovom članku nastojimo analizirati potencijal metoda nadziranog strojnog učenja u savladavanju ovog problema. U tu smo svrhu koristili združene setove podataka iz tri izvora (tragove vozila prikupljene putem sustava globalnog pozicioniranja, bazu podataka o izgrađenoj cestovnoj infrastrukturi i meteorološke podatke), te četiri metode nadziranog strojnog učenja (metoda potpornih vektora, k-najbližih susjeda, metodu rastućih stabala i metodu slučajne šume). Kako bismo usporedili postignute rezultate, isti su međusobno uspoređeni među različitim klasama prometnica kao apsolutne vrijednosti prognoziranog vremena putovanja (izražene u minutama) i kao srednja kvadrirana postotna pogreška. Za one klase prometnica koje odlikuju visoke prosječne brzine kretanja vozila i dugi neprekinuti segmenti prometnica, metode nadziranog strojnog učenja su polučile rezultate s malom relativnom pogreškom, dok je za prometnice s nižim prosječnim brzinama i kratkim segmentima prognoziranje bilo puno zahtjevniji zadatak. Sva su tri korištena ulazna seta podataka opravdala svoju primjenjivost za ovu namjenu jer su imala veliki utjecaj na preciznost prognoziranja. Najbolji su rezultati (uzevši u obzir sve klase prometnica) postignuti primjenom metoda k-najbližih susjeda i slučajne šume.

KLJUČNE RIJEČI

veliki setovi podataka; metoda potpornih vektora; metoda k-najbližih susjeda; metoda rastućih stabala; metoda slučajne šume; prognoziranje vremena putovanja, fuzija podataka;

REFERENCES

[1] Xu X, Chen A, Cheng L. Assessing the effects of stochastic perception error under travel time variability. Transportation. 2013;40(3):525-548.

[2] Kim M, Miller-Hooks E, Nair R. A Geographic Information System-Based Real-Time Decision Support Framework for Routing Vehicles Carrying Hazardous Materials. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations. 2011;15(1):28-41.

[3] Yu Z, Ni M, Wang Z, Zhang Y. Dynamic Route Guidance Using Improved Genetic Algorithms, Mathematical Problems in Engineering [Internet]. 2013 [cited 2015 Feb 8]; 2013:[about 6pp.]. Available from: http://dx.doi.org/10.1155/2013/765135

[4] YinY, Lam WHK, Ieda H.Modeling ris*k*-taking behavior in queuing networks with advanced traveler information systems. Transportation and traffic theory. 2002;15:309-328.

[5] Yu B, Yang Z-Z, Chen K, Yu B. Hybrid model for prediction of bus arrival times at next station. Journal of Advanced Transportation. 2010;44(3):193-204.

[6] Sun L, Yang J, Mahmassanic H. Travel time estimation based on piecewise truncated quadratic speed trajectory. Transportation Research Part A: Policy and Practice. 2008;42(1):173-186.

[7] Zheng W, Lee D-H, Shi Q. Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach. Journal of Transportation Engineering. 2006;132(2):114-121.

[8] Mahmood M, Bashar MA, Akhter S. Traffic Management System and Travel Demand Management (TDM) Strategies: Suggestions for Urban Cities in Bangladesh. Asian Journal of Management and Humanity Sciences. 2009;4(2-3):161-178.

[9] Lyons G, Urry J. Travel time use in the information age. Transportation Research Part A: Policy and Practice. 2005;39(2-3):257-276.

[10] Brnjac N, Ćavar I. Example of Positioning Intermodal Terminals on Inland Waterways. Promet – Traffic & Transportation. 2009;21(6):433-439.

[11] Malchow M, Kanafani A, Varaiya P. The Economics of Traffic Information: A State-of-the-Art Report. University of California at Berkeley; 1996.

[12] Bhaskar A, Chung E, Dumont A-G. Analysis for the Use of Cumulative Plots for Travel Time Estimation on Signalized Network. International Journal of Intelligent Transportation Systems Research. 2010:8(3):151-163.

[13] Zong F, Lin H, Yu B, Pan X. Daily Commute Time Prediction Based on Genetic Algorithm. Mathematical Problems in Engineering, [Internet]. 2012 [cited 2015 Feb 9]; 2012:[about 20pp.]. Available from: http://dx.doi.org/10.1155/2012/321574

[14] Simroth A, Zähle Z. Travel Time Prediction Using Floating Car Data Applied to Logistics Planning. IEEE Transactions on Intelligent Transportation Systems. 2011;12(1):243-253.

[15] Huang Y, Xu L, Kuang X. Urban Road Travel Time Prediction Based on Taxi GPS Data. Improving Multimodal Transportation Systems-Information, Safety, and Integration; 2013 Jun 29-Jul 2; Wuhan, China. Reston: American Society of Civil Engineers; 2013.

[16] Anusha SP, Anand RA, Vanajakshi L. Data Fusion Based Hybrid Approach for the Estimation of Urban Arterial Travel Time. Journal of Applied Mathematics, [Internet]. 2012; [cited 2015 Jan 27]; 2012:[about 17pp.].no. Available from: http://dx.doi.org/10.1155/2012/587913

[17] Lum K, Fan H, Olszewski S. Speed-Flow Modeling of Arterial Roads in Singapore. Journal of Transportation Engineering. 1998;124(6):213-222.

[18] Ćavar I, Kavran Z, Petrović M. Hybrid Approach for Urban Roads Classification Based on GPS Tracks and Road Subsegments Data. Promet – Traffic & Transportation. 2011;23(4):289-296.

[19] Akbari M, van Overloop PJ, Af A. Clustered *k* Nearest Neighbor Algorithm for Daily Inflow Forecasting. Water Resources Management. 2011;25(5):1341-1357.

[20] Li L, Zhang Y, Zhao Y. *k*-Nearest Neighbors for automated classification of celestial objects. Science in China Series G: Physics, Mechanics and Astronomy. 2008;51(7):916-922.

[21] Valenti G, Lelli V, Cucina D. A comparative study of models for the incident duration prediction. European Transport Research Review. 2010;2(2):103-111.

[22] Poloczek J, Treiber NA, Krame O. KNN Regression as Geo-Imputation Method for Spatio-Temporal Wind Data. Advances in Intelligent Systems and Computing, 2014;299:185-193.

[23] Wang H, Düntsch I, Gediga G, Guo G. Nearest Neighbours without k. Monitoring, Security, and Rescue Techniques in Multiagent Systems. 2005;28:179-189.

[24] Battiti R, Mascia F, Brunato M. Supervised Learning. Reactive Search and Intelligent Optimization. 2009;45:1-33.
[25] Batista G, Silva DF. How *k*-Nearest Neighbor Parameters Affect its Performance.Simposio Argentino de Inteligencia Artificial, 2009 Aug 24-28; Mar del Plata, Argentina. Buenos Aires: Sociedad Argentina de Informática; 2009.
[26] Everitt BS, Landau S, Leese M, Stahl D. Miscellaneous Clustering Methods. In: Cluster Analysis. 5th ed. Chichester, UK: John Wiley & Sons, 2011; doi: 10.1002/9780470977811.ch8.
[27] Nigsch F, Bender A, van Buuren B, Tissen J, Nigsch E, Mitchell JBO. Melting Point Prediction Employing *k*-Nearest Neighbor Algorithms and Genetic Parameter Optimization. J. Chem. Inf. Model. 2006;46(6):2412-2422.
[28] Hall P, Park BU, Samwor RJ. Choice of neighbor order in nearest-neighbor classification. The Annals of Statistics. 2008;36(5):2135-2152.
[29] López-Rubio E, Ortiz-de-Lazcano JM. Automatic Model Selection by Cross-Validation for Probabilistic PCA. Neural Processing Letters. 2009;30(2):113-132.
[30] Donate JP, Cortez P, Gutierre G. Weighted Cross-Validation Evolving Artificial Neural Networks to Forecast Time Series. In: Soft Computing Models in Industrial and Environmental Applications. Salamanca, Spain; 2011.
[31] Abe S. Support Vector Machines for Pattern Classification (Advances in Computer Vision and Pattern Recognition). London, UK: Springer; 2010.
[32] Burbidge R, Buxton B. An Introduction to Support Vector Machines for Data Mining. In: Keynote papers. Nottingham, UK: University of Nottingham, Operational research society, 2001; pp. 3-16.
[33] Steinwart I, Christmann A. Support Vector Machines (Information Science and Statistics). New York, USA: Springer; 2008.
[34] Hamel LH. Knowledge Discovery with Support Vector Machines. Hoboken, Canada: Wiley-Interscience; 2009.
[35] Appel R, Fuchs T, Dollar P, Perona P. Quickly Boosting Decision Trees – Pruning Underachieving Features Early. In: 30th International Conference on Machine Learning. Atlanta, USA; 2013.
[36] Freund Y, Schapire RE. A Short Introduction to Boosting. Journal of Japanese Society for Artificial Intelligence. 1999;14(5):771-780.
[37] Friedman J, Hastie T, Tibshira R. Additive Logistic Regression: A Statistical View of Boosting. The Annals of Statistics. 2000;28(2):337-407.
[38] Chang Y-CI, Huang Y, Huang Y-P. Early stopping in L2Boosting. Computational Statistics & Data Analysis. 2010;54(10):2203-2213.
[39] Breiman L. Random Forests. Machine Learning, 2001;45(1):5-32.
[40] Biau G. Analysis of a random forests model. The Journal of Machine Learning Research. 2012;13:1063-1095.
[41] Grömping U. Variable Importance Assessment in Regression: Linear Regression versus Random Forest. The American Statistician. 2009;63(4):308-319.
[42] Biau G, Devroye L, Lugosi G. Consistency of Random Forests and Other Averaging Classifiers. Journal of Machine Learning Research. 2008;9:2015-2033.
[43] Pal M. Random forest classifier for remote sensing classification. International Journal of Remote Sensing. 2007;26(1):217-222.