

## DECISION TREES AS A TOOL FOR REAL-TIME TRAVEL TIME ESTIMATION ON HIGHWAYS

*This paper presents a time travel estimation model based on a decision tree. Input parameter for travel time prediction is occupancy of detectors. The proposed model was tested on the most widely used arterial in Prague and also in the Czech Republic. This road section has many unmeasured inputs and outputs and with regards to only two detectors within a section it is difficult to estimate the travel time. A temporary installation of ALPR system is used for training the decision tree model which consequently provides reliable travel time estimation.*

**Keywords:** Travel Time estimation, occupancy, Decision Tree, FCD, ALPR.

### 1. Introduction

Travel time (TT) is an important traffic system performance indicator and its great advantage is that it is easy to understand for drivers and generic users in contrast to the usual manner when the Level of Service is expressed as integer number 1 to 5 (“1” is free flow and “5” is congestion).

Travel time can be measured directly by a license plate matching system for every individual vehicle or indirectly estimated through many techniques, mostly using stationary traffic detectors as the most common type of measurement tool, ref. [1], [2]. However the automated licence plate recognition (ALPR) system is still rather expensive and the proposed method uses ALPR only for calibration of estimations of TT provided by static traffic detectors. The described method in this paper is based on a decision tree which estimates travel time using data from loop detectors which were being calibrated by ALPR. ALPR could be installed for a limited time, but it is highly recommended to share the wide range of traffic situations.

### 2. Literature review

There are many methods for travel time estimation. A trajectory method uses spot speed measures and transforms it to the estimate of the space mean speed. A first conversion formula was described as early as in 1952 by Wardrop, [3]:

$$\bar{v}_t = \bar{v}_s + \frac{\sigma_s^2}{\bar{v}_s}, \quad (1)$$

where  $\bar{v}_s$  is average space mean speed,  $\bar{v}_t$  is average time mean speed and  $\sigma_s^2$  is variance of the space-mean speed distribution. In the real traffic environment we can directly measure the speed at the spot (time-mean speed), but this formula enables to calculate it from space-mean speed, which is impossible to obtain by standard inductive loops.

Travel time is more meaningful traffic parameter than volume of traffic or occupancy of traffic detectors are. Travel time could be directly used in intelligent transport systems to inform drivers through information displays placed at highways or roads. Travel time is also measured for delay in a road network and could be used as input variable in traffic control systems. The typical example is published in paper [4]. The authors put forward new intersection control optimization method based on real-time delays estimation. The total delay is done by average delay caused by signal control, the additional delay caused by the queuing of previous vehicles and also by the stochastic delay. To obtain the total time delay, it is necessary to use analytical equations with many parameters that must be set in a complicated way. The presented method overcomes the necessity to apply analytical methods with many boundary conditions by the using of AI methods.

The neural network optimized by generic algorithm of models for the real time forecast of traffic flow is presented in [5]. Traffic flow is a better predictable parameter than travel time is. The reason is done by slowly varying curve of traffic flow which is “modulated” by pseudo stochastic randomly distributed clusters of vehicles. The average relative error of traffic flow forecast is 5.53%. However, the proposed method of estimating TT has not a larger deviation than

\* Jakub Wosyka, Pavel Pribyl

Faculty of Transportation Sciences, CVUT Prague, Czech Republic, E-mail: xwosyka@fd.cvut.cz

6%, while it is relatively less complicated than neural network oriented method.

*A. Trajectory based methods*

Trajectory methods are the simplest and the most commonly used techniques to estimate travel time, [6]. Both methods consider the measured speed is constant until the next detector (or to the half distance on both sides). In half distance method the speed measured by a detector is applicable to half the distance on both sides.

$$TT = \frac{1}{2} \left( \frac{L}{v_1} + \frac{L}{v_2} \right) \quad (2)$$

In average speed method the speed is assumed to be the average speeds measured by detectors 1 and 2.

$$TT = \frac{2L}{(v_1 + v_2)} \quad (3)$$

The main disadvantage of the constant speed trajectory methods is the decreasing performance with increasing traffic congestion.

The author in paper [1] describes a method which uses raw (not time aggregated) data from loop detectors. The spot speed of an individual vehicle is considered constant until the next measurement of another available vehicle. The trajectory of one individual vehicle in time and space is constructed from the sequence of speed measurements obtained from the same vehicle and its few successors. The author evaluated the mean error of his method result between 7 - 10 % considering constant shock wave speed at 22.4 km/h. Unfortunately, the raw input data is essential for this method, but this data is not commonly available online in the Czech Republic.

*B. Traffic flow theory based methods*

Theoretical models have also been developed for the estimation of travel time from loop detector data based on traffic flow theory.

The Adaptive Kalman Filter (AKF) is used in the paper to improve the travel time estimate from inductive loop using a small number of probe vehicles. Unfortunately probe vehicles were simulated by microscopic simulation model PARAMICS. Because of the inaccuracy of the travel time estimation from point detectors, traffic data from other sources can be incorporated to improve the estimation. Due to the recent advances in probe vehicle technologies, such as Global Positioning Systems (GPS), Automatic Vehicle Identification (AVI), cellular phone positioning, and vehicle re-identification technology, probe vehicle has shown its potential to be another valuable real-time traffic data source according to the [2].

Both data inputs (loops and probe vehicles) do not truly reflect accurate section travel time especially under recurrent or non-recurrent traffic congestion condition.

The proposed method in [2] consists of two parts. First is based on the conservation or continuity equation:

$$\frac{\partial p}{\partial t} + \frac{\partial k}{\partial x} = 0 \quad (\text{or the traffic generation rate}) \quad (4)$$

where  $q$  is flow (vehicles/hour),  $k$  is density (vehicles/km),  $x$  is location, and  $t$  is time.

For a typical urban freeway section including one on-ramp and one off-ramp, the traffic flow passing the section during time period  $(t - 1, t)$  can be estimated as:

$$q(t) = \alpha [q_u(t) + q_{on}(t)] + (1 - \alpha) [q_d(t) + q_{off}(t)] \quad (5)$$

where  $\alpha$  is a smoothing parameter that is set to 0.5.  $q_u(t)$  and  $q_d(t)$  are traffic flows of the upstream and downstream boundaries within  $(t - 1, t)$ .  $q_{on}(t)$  and  $q_{off}(t)$  are total on-ramp and off-ramp traffic flows within  $(t - 1, t)$ .

Assuming that the traffic inside of the section is homogeneous, an intuitive estimation of the section travel time is:

$$TT(t) = \frac{\Delta x}{v(t)} = \frac{\Delta x}{q(t)} k(t) \quad (6)$$

where  $\Delta x$  is length of the section between upstream and downstream detectors. Section density,  $k(t)$ , which can be represented as a time series:

$$k(t) = k(t - 1) + \frac{1}{L\Delta x} (q_u(t) + q_{in}(t) - q_d(t) - q_{out}(t)) \quad (7)$$

where  $L$  is the number of lanes on the mainline freeway.

The Kalman filter is used for correction of section density, which is used for travel time estimation. Section density is treated as a state variable and the section travel time is treated as a measurement variable. The performance of the proposed algorithm is evaluated by Mean Absolute Percentage Error (MAPE). The author also simulated constant error patterns and time-varying error pattern on the loop detectors in order to evaluate how the proposed method can handle these kinds of common errors.

The AKF Algorithm using 5% probe vehicles in the traffic stream (simulated in PARAMICS simulation model) has MAPE in range 7.6 - 9.7% compared to both single data source methods up to 16.6%, which is a significant improvement. The author also stated that with the higher probe rate available, better performance can be expected from the AKF model. When the probe rate is higher than 20% no more improvement is observed.

**3. Proposed model**

The Travel time estimation model which improves travel time based only on data from the series of loop detectors is presented in this paper (see Fig. 1) and in [7]. The Automatic licence plate recognition (ALPR) system was temporarily used for training of

decision tree model (DTM) by determining truth values of travel time on road sections of interest.

The length of the section where estimation of travel times was being tested is 3.6 kilometres and there are three profile measurements by traffic loop detector. The average distance between detectors is greater than one kilometre. The application of the analytical methods described in Chap. 2 was not successful and the difference between real measured TT and estimated time was often higher than 100 percent; depending on the actual traffic volume. The results were poorer for low or very high traffic volume. The reason is simple – the distance between detectors is too large. For instance, the recommended distance between fixed detectors on highways is about 500 m in the Netherlands.

A road without many exits is most appropriate for accurate TT estimation of TT, but it is not the case of our tested road, see Fig. 2. Jizni spojka is an arterial road in Prague that carries more than 100 000 personal cars connecting the southern and northern parts of Prague. Three significant exits and two petrol stations on this short tested section are present. Arriving and departing vehicles significantly influence accuracy of TT estimation. In order to reduce the inaccuracy to the lowest possible level the method

based on artificial intelligence is probably best to use. Well-trained decision trees have the capacity to cover all traffic scenarios at this road.

*A. Decision tree learning process*

Decision trees (DT) can be used to discover features and extract patterns in large databases that are important for predictive models. Decision trees have an established position in artificial intelligence methods. DT are used as classification and prediction tool of travel time estimation based on input parameters which occupancy of detectors is.

Detector occupancy is the ratio between the time that the vehicle is over the detector, and the scanning period expressed in percentage.

The Decision tree learning process requires input variables which are traffic parameters (occupancy) and output values as travel time. Figure 2 depicts a specific part of Jizni spojka in Prague. It is possible to see the whole section enclosed by green (start) and red (finish) flags. Two video-detectors (2 and 3) lie within the road link section, a first one is located outside, in position -364 meters from the start. The ALPR system measured TT between point red flag and green flag and it is 3644 metres long.

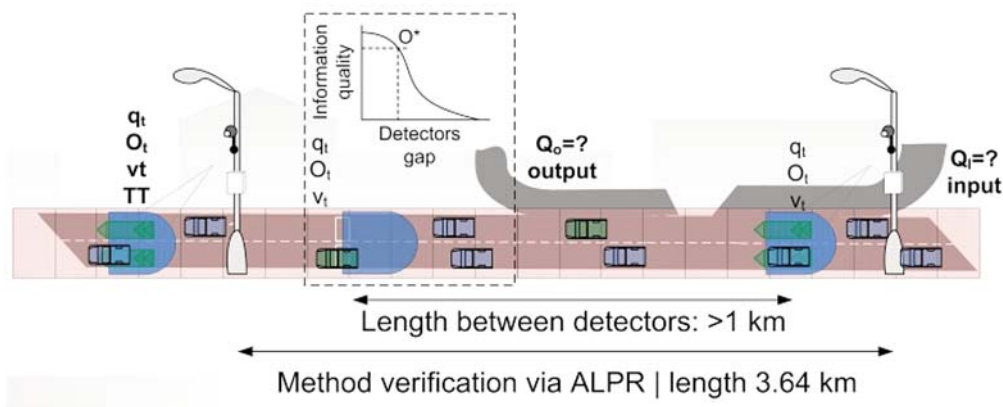


Fig. 1 Decision tree learning via automatic licence plate recognition (ALPR) system

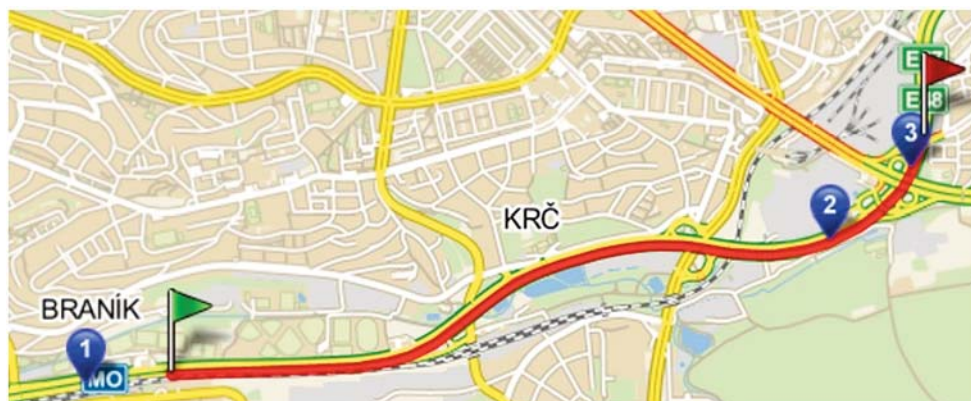


Fig. 2 Link road section in Prague (Jizni spojka)

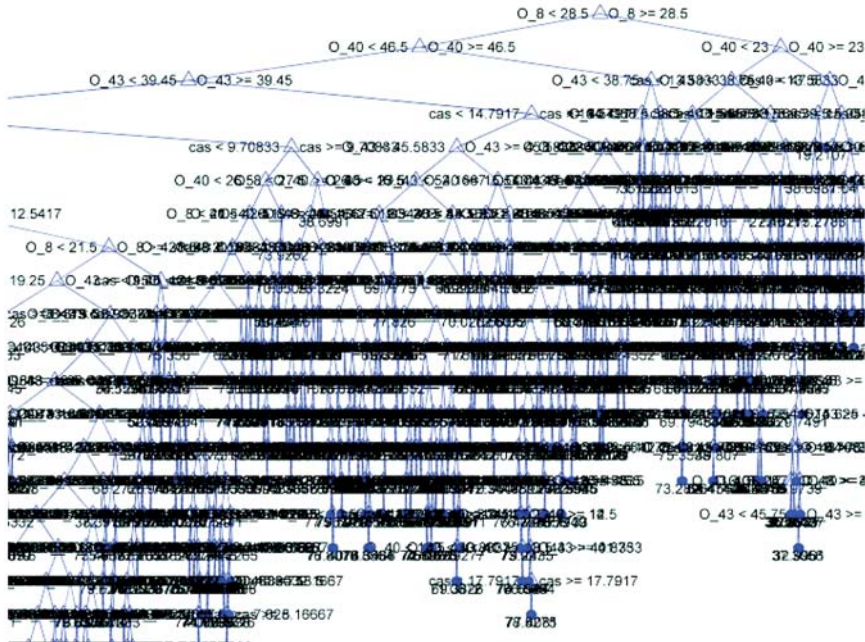


Fig. 3 Part of created decision tree for estimating travel time – Demonstration of complexity

Data recorded in one month contain both free flow and congested traffic states. 90% of data was used for learning decision tree model (training data). The rest was verification data. The preliminary DTM tests used traffic volumes as well as occupancy of detectors. Many experiments showed that the trees are too complex in this case and practically the same results can be achieved when using only occupancy.

The final model uses five minute aggregated occupancy values of all three video- detectors as the input variables. Based on these inputs the travel time estimation is provided in same time step (5 minute time step). Part of related decision tree for travel time estimation can be seen in Fig. 3, which shows how complex created DT could be.

Very fast evaluation of binary condition (if – then) is the significant advantage of decision trees for traffic engineering application. Hence the decision tree-based model can be easily implemented.

*B. Model validation and performance evaluation*

Consequently unused input data (about 10%) was used for model performance validation using the truth value of TT coming from ALPR system. Three performance indicators were used for evaluation purposes: MSE (mean square error) of average speed, MAE (mean absolute error) of average speed and SSE (sum of square error) of average speed

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad MAE = \frac{1}{n} \sum_{i=1}^n \sqrt{e_i^2} \quad SSE = \sum_{i=1}^n e_i^2 \quad (8)$$

The performance indicators corresponding to twelve different days in May and June are elaborated in Table 1. It is possible to see that the travel time error (in seconds) corresponds to the mean absolute error of the speed on the road section.

Performance evaluation

Table 1.

Date	MSE [km/h] <sup>2</sup>	MAE [km/h]	SSE [km/h] <sup>2</sup>	Mean time difference [sec.]	TT error
4. 5. 2009	33.46	4.51	6290	9	5%
5. 5. 2009	31.22	4.45	3903	9	5%
12. 5. 2009	35.09	4.66	5158	9	6%
13. 5. 2009	40.86	4.95	7722	9	6%
15. 5. 2009	41.77	4.97	7143	10	6%
18. 5. 2009	23.01	3.69	4304	7	4%
21. 5. 2009	33.27	4.55	6487	9	5%
22. 5. 2009	44.16	5.	7419	10	6%
28. 5. 2009	42.39	5.17	6316	10	6%
29. 5. 2009	37.7	4.74	4825	9	6%
5. 6. 2009	41.69	5.12	6837	10	6%
6. 6. 2009	15.75	2.97	2000	6	4%

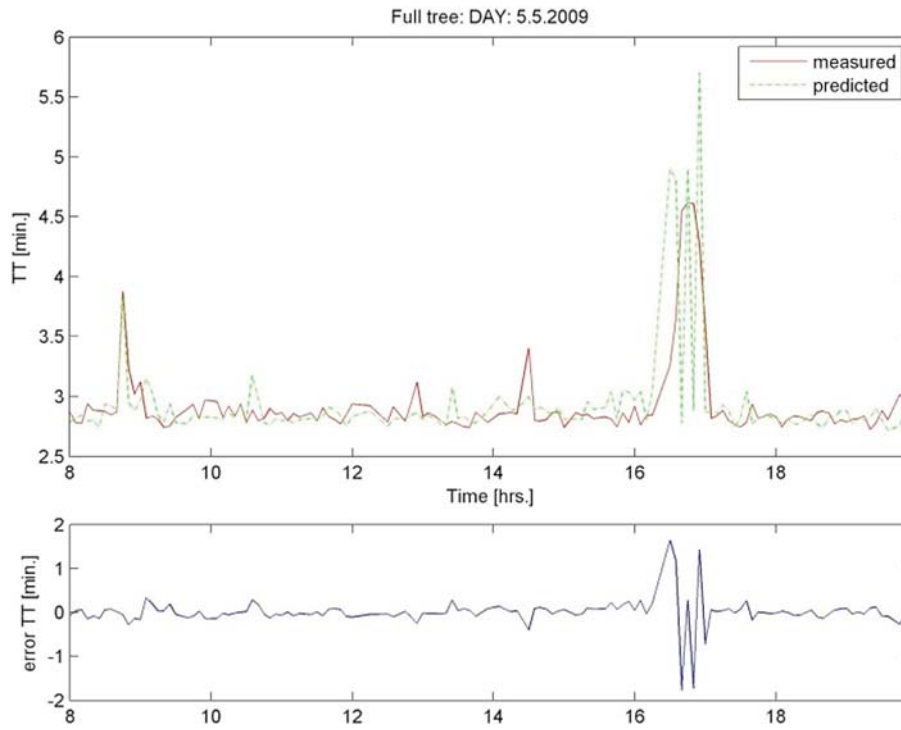


Fig. 4 Measured and predicted TT (above); residual value (bottom)

**Historický model**

čas	v [km/h]	tt [min]	I1	O1	I2	O2	I1	O2
17:57	76.1369	3.11707	113	14.5	63.3	29.125	104.5	13.9

**Výstup jednotlivých modulů**

metoda	čas	tt levý pruh [min]	tt prostřední pruh [min]	tt pravý pruh [min]	tt prumer [min]
rozhodovací stromy	17:57	2.875	3	3.4375	3.10417
neuronové site	17:57	2.55118	3.05989	5.15951	3.59019

**Časový průběh obsazenosti jednotlivých strategických detektorů**

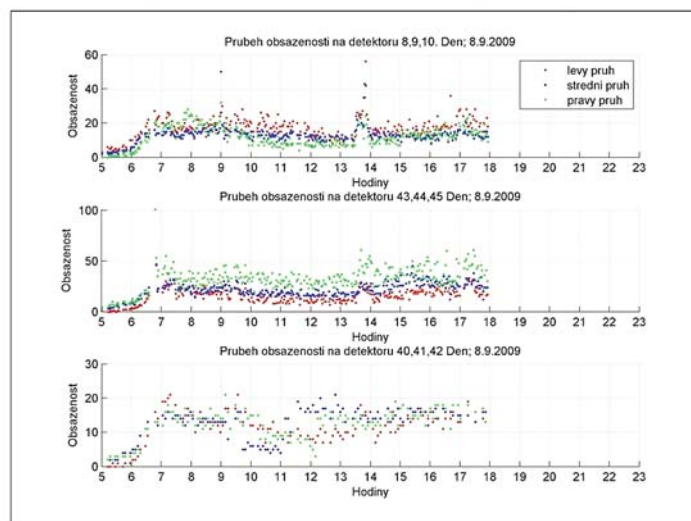


Fig. 5 Web site opened for road users [7] (in Czech)

The diagram in Fig. 4 (bellow) is graphical presentation of real measured TT by ALPR (red curve) and predicted value by DTM (green curve). X-axis expresses the time. The residual value is the curve at the bottom.

#### 4. Practical application

System for prediction of travel time based “only” on very limited set of input data was implemented as Czech web site of the City Hall of Prague to give a possibility for road users to check accuracy of the proposed system, Fig. 5.

Three different colours represent all three lanes measured on the tested link section (the legend translation is following: levý / střední / pravý pruh = left / middle / right lane). The occupancy of three detectors is depicted in the figure above and the result of predicted travel time is shown in the table. The TT was predicted by decision trees and also neural network tools. The displayed values as outputs of both AI methods were mutually compared for the research reasons.

#### 5. Conclusion and future work

The presented method is very effective because it enables us to estimate travel times relatively reliably, even when traffic detectors are far from each other. In comparison to the neural networks models, which were also tested in the frame of the project, the decision tree model is more controllable and well-arranged.

Nevertheless the temporary installation of ALPR system for TT measurement also has operational expenditures. The research continues in the direction to replace ALPR equipment by floating car data. The first measurements on the D1 highway proved a penetration of floating cars in the traffic stream volume is about 5 %. This volume of floating cars could very precisely calibrate travel times estimated by fixed detectors.

##### Acknowledgement

This project was supported by Ministry of Industry and Trade of the Czech Republic No. 2A\_1TP1/023 and it has been processed in collaboration with Faculty of Electrical Engineering, University of Žilina.

#### References

- [1] COIFMAN, B.: *Estimating Travel Times and Vehicle Trajectories on Freeways Using Dual Loop Detectors*, Ohio State University : Columbus, 2008
- [2] CHU, L., OH, J., RECKER, W.: *Adaptive Kalman Filter Based Freeway Travel time Estimation*, TRB Annual Meeting, 2005
- [3] WARDROP, J. G.: *Some Theoretical Aspects of Road Traffic Research*, Proc. of the Institute of Civil Engineers, vol. 1-2, 1952, pp. 325-378, 1952
- [4] LIU, B., ZHANG, T., LI, Q., WANG, L.: An Traffic Control Method Based on the Real Time Detector Delay, *Communications - Scientific Letter of Zilina University*, 2, 2010, pp. 76-80, ISSN 1335-4205
- [5] HAI-MIN, J., YU-LONG, P.: Neural Network Optimized by Generic Algorithm of Models for Real Time Forecast of Traffic Flow, *Communications - Scientific Letter of Zilina University*, 3, 2010, pp. 80-84, ISSN 1335-4205.
- [6] SHEN, L.: *Freeway Travel Time Estimation and Prediction Using Dynamic Neural Networks*, FIU Electronic Theses and Dissertations, Paper 17, 2008
- [7] WOSYKA, J., PRIBYL, P.: *Satelitni lokalizace vozidel v dopravní telematice - SATEL [Vehicle Satellite Localization in Traffic Telematics]*, Projekt MPO #. 2A\_1TP1/023, VZ 322-2010-1930-EEG, Praha, 2010 (in Czech).