Structural Optimization of the Travel Time Prediction Model Based on Hierarchical Regression

Anton Agafonov[1,2] and Vladislav Myasnikov[1,2]

[1] Samara State Aerospace University (SSAU), Samara, Russia,
[2] Image Processing Systems Institute of the Russian Academy of Sciences
(IPSI RAS), Samara, Russia
ant.agafonov@gmail.com, vmyas@geosamara.ru

Abstract. In this paper we consider a problem of public transport arrival time prediction for a large city in real time. We investigate the algorithm based on a model of an adaptive combination of elementary prediction algorithms. Adaptability means that parameters of the constructed combination depend on a number of control parameters of the model. We compare our model with the nonlinear artificial neural network model. Comparison is performed on data of the public transport network in Samara, Russia.

Keywords: travel time prediction, combination of algorithms, neural network, hierarchical regression

1 Introduction

Geoinformation systems ans intelligent transportation systems (ITS) based on automated vehicle location data are widely used by public transportation companies around the world. ITS can be used for the improvement of public transport reliability using different operational planning and control strategies. Such strategies can be used to improve the public transport network design, evaluate the schedule reliability and improve schedule planning [1].

Travel time prediction problem is one of the most common transportation problems. Travel time prediction can be used for solving many other problems in different contexts, including management, navigation, planning, monitoring and control problem, mentioned early.

There are much research focused on the travel time prediction problem. All approaches can be divided into the four following categories [1].

1. Historical data based models. These approaches predict travel time in a certain period by the averaged over historical values during the same period in the previous days. The simplicity is the main drawback of these models, because models ignore both current road situation (such as congestions or accidents) and complex relationship between travel times and other variables in urban public transport networks.

2. Traffic theory based models [2]. These models perform travel time prediction based on traffic flow / traffic density estimations. The requirement of an actual traffic flow model is the main drawback of these models.
3. Time-series models [3]. These approaches assume that the future travel time depends only on the most recent data samples. Accuracy of models decreases for a longer forecasting horizon.
4. Machine learning and regression methods. These methods construct a regression function depending on a set of independent variables, which may include real-time and historical data of the road links travel times, road situation, passenger, weather situation, delays at stops, etc [4]. Over the last two decades, regression models have been the state of the art for this kind of applications. Artificial neural networks (ANN) are widely used for the travel time prediction [5, 6], can represent complex nonlinear relationships between the target travel time value and the set of independent variables. Support vector regression (SVR) is used for travel time prediction in [7, 8]. Nonlinear regressions methods (such as ANN or SVR) has several drawbacks compared with other regression methods, including high training computational complexity and possibility for overfitting.

In this paper we investigate the adaptive model proposed in [9]. This model is based on adaptive combination of the elementary prediction algorithms. The model uses hierarchical regression for representing complex nonlinear relationships between the target travel time value and the set of independent variables. Main purposes of this paper are the structural optimisation of the hierarchical regression based model and the comparison of the proposed model with an artificial neural network model.

The structure of this paper is organized as follows: the second section contains a problem statement. The third section contains a description of the base model. The next section describes modified structure of the base model, used elementary predictors and aggregated function. In the fifth section we present an experimental study of the algorithm. Lastly we present our conclusions and recommendations for further work.

2   Problem statement

We use the following notations:

- w is the road link in a transport network;
- j is the index of the specified vehicle;
- s is the transport type (bus, tram, trolley), S is the set of transport types;
- $\Delta_j \in [0, 1]$ is the relative position of the vehicle j on the road link;
- t_c is the current time instant (at which predictions are calculated).

The position of any vehicle in the transport network we define by the pair (w, Δ).

Let the vehicle j at the time t_c be in the position (w_0, Δ_0). We should estimate the arrival time, after which the vehicle be at the stop with position (w_K, Δ_K).

Let the route between two positions (w_0, Δ_0) and (w_K, Δ_K) includes the following edges $(w_0, \Delta_0) \rightarrow w_1 \rightarrow w_2 \rightarrow \cdots \rightarrow w_{K-1} \rightarrow (w_K, \Delta_K)$. Let $T_j^w(t_c, t)$ be the travel time along the edge w starting at the time t, assuming that the prediction is calculated at the time t_c.

Assuming, that the vehicle j moves monotonically on each road link, the predictive arrival time to the stop (w_K, Δ_K), calculated at the time t_c, can be written as follows:

$$
\begin{aligned}
T_j^{(w_0, \Delta_0) \rightarrow (w_K, \Delta_K)}(t_c) = {} & (1 - \Delta_0) T_j^{w_0}(t_c, t_c) \\
& + \sum_{k=1}^{K-1} T_j^{w_k}\left(t_c, t_c + T_j^{(w_0, \Delta_0) \rightarrow (w_k, 0)}(t_c)\right) \\
& + \Delta_K T_j^{w_K}\left(t_c, t_c + T_j^{(w_0, \Delta_0) \rightarrow (w_K, 0)}(t_c)\right) \ .
\end{aligned}
\tag{1}
$$

As it can be seen from the above expression, its main components are terms $T_j^{w_k}(\cdots)$ that characterize the travel time of the road links.

For simplification, further we consider the travel time prediction problem at the specified edge w_k and the specified time t:

$$
T^w(t_c, t) \ .
\tag{2}
$$

In next section we briefly describe our prediction model proposed in [9].

3 The base model

Calculation of prediction (2) can be performed using different information usually established in urban public transport networks. Firstly we describe the different travel time estimations and control parameters that affect the predictive travel time value (2):

1. $\left\{T_{\Sigma(s)}^w(t)\right\}_s$ is the averaged historical travel time of the edge w by the vehicles of the transport type s;
2. $T_{last}^w(t_c)$ is the most recent travel time of the link w;
3. $T_0^w(t)$ is the normative travel time by the planning schedule;
4. $T_{v(j)}^w(t_c)$ is the estimated travel time by the mean speed of the specified vehicle;
5. $\{(t - t_j, T_j(t_c))\}_j^w$ is the set of precedents containing time instants and travel times of the edge w before the time t_c;
6. τ^w is the time elapsed since the last vehicle has passed the link w;
7. $t - t_c$ is the required prediction horizon;
8. $\gamma(t) \in [0, 1]$ is the parameter defining the driving difficulty at the time t. This parameter takes into account the light and the weather conditions integrally;
9. $\rho^w(t) \in [0, 1]$ is the parameter defining the traffic density;
10. $\eta^w(t)$ is the parameter defining dynamic of the traffic flow on the edge w.

These values have different meaningful information that affects the way it is used. Parameters 1-5 present individual or aggregated statistical "precedents" of the travel times. These values can be used (together or separately, directly or indirectly) in the construction of the required estimate $T^w(t_c, t)$ as some realizations of this assessment. Parameters 6-10 also directly affect this estimate. However, these parameters cannot be considered as the precedents of the travel times. We use these parameters as the control parameters of the model (further we use notation p_i for them).

To predict travel time $T^w(t_c, t)$ we proposed to use a model based on the combination of the elementary prediction algorithms [9]. Fig. 1 provides a scheme of the proposed model.
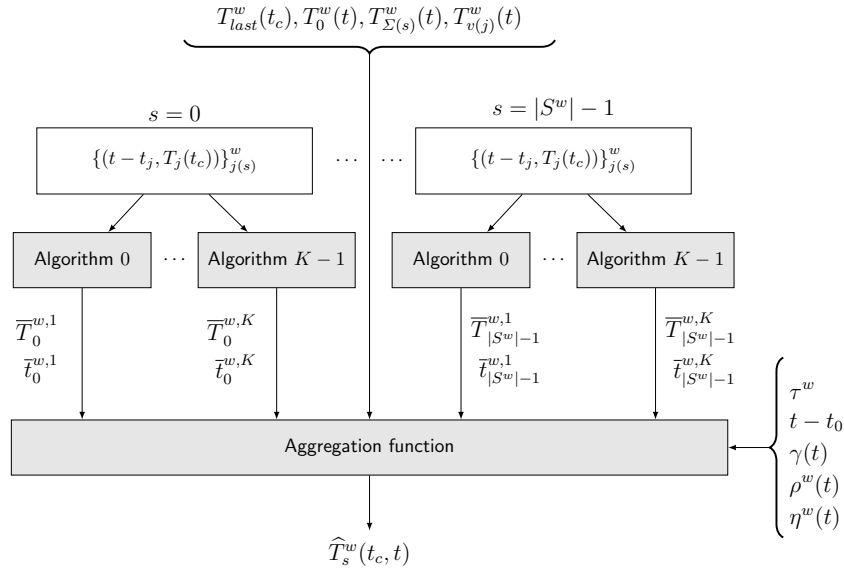

Fig. 1. The model scheme.

The elementary prediction algorithms take as input a set of precedents $\{(t - t_j, T_j(t_c))\}_j^w$ and return predictions $\overline{T}^{w,k}(t_c, t)$ For this mapping we used the following kernel functions:

$$
\begin{aligned}
\overline{T}_s^{w,k}(t_c, t) &= \frac{\sum_j \omega_k(t - t_j) T_j(t_c)}{\sum_j \omega_k(t - t_j)}, \\
\overline{t}_s^{w,k}(t_c, t) &= \frac{\sum_j \omega_k(t - t_j) t_j(t_c)}{\sum_j \omega_k(t - t_j)} \ .
\end{aligned}
\tag{3}
$$

We used four kernels: rectangular kernel ($k = 0$), triangular kernel ($k = 1$), exponential kernel ($k = 2$) and rational kernel ($k = 3$).

As the aggregation function we used the following function:

$$\widehat{T}_s^w(t_c,t) = \sum_{k=0}^{K-1} \sum_{r \in S} a_{rs}^k \overline{T}_r^{w,k}(t_c,t) +$$
$$+ a_1 T_{last}^w(t_c) + a_2 T_0^w(t) + a_3 T_{\Sigma(s)}^w(t) + a_4 T_{v(j)}^w(t_c) \ . \tag{4}$$

Since our algorithm should be adaptive to the control parameters, we proposed to use hierarchical regression similar to regression tree. Parameters of the aggregation function (4) are defined for each subdomain of definition of the control parameters. Subdomains are calculated automatically using the hierarchical regression with the following approach:

1) domain of definition for each control parameter p_i is determined: $[p_i^{min}, p_i^{max}]$;
2) interval $[p_i^{min}, p_i^{max}]$ is splitted into two intervals by an average value.

Fig. 2 shows an example of the hierarchical regression construction for two control parameters.
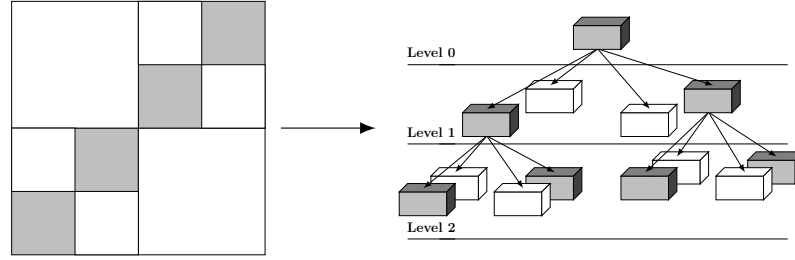

Fig. 2. Example of the hierarchical regression construction.

Proposed hierarchical regression is used for representing complex nonlinear relationships between the target travel time value and the set of independent variables.

In the next section we describe main modifications of the model.

4 Modifications of the model

To increase accuracy and decrease computation complexity of the proposed model, we modify used elementary prediction algorithms and the aggregation function. Experiments also show that the information, affected the travel time and described in the previous section, has redundancy. We decide to use only significant parameters in our modified model.

First of all, we added a new transport type car to an input data of the modified model. We decided to use the following most significant information that affect the predictive travel time value (2):

1. $\left\{ T^w_{\Sigma(s)}(t) \right\}_s$ is the averaged historical travel time of the edge w by the vehicles of the transport type s;

2. $\{(t - t_j, T_j(t_c))\}^w_j$ is the set of precedents containing time instants and travel times of the edge w before the time t_c;

3. $t - t_c$ is the required prediction horizon;

4. $\gamma(t) \in [0, 1]$ is the parameter defines the driving difficulty at the time t. This parameter takes into account the light and the weather conditions integrally;

5. $\eta^w(t)$ is the parameter defining dynamics of the traffic flow on the edge w;

6. $\rho^w(t) \in [0, 1]$ is the parameter defining the traffic density .

The elementary prediction algorithms (3) based on kernel functions have the similar accuracy and, moreover, have high correlation. To avoid these disadvantages we use the best one algorithm with the exponential kernel.

As an aggregation function we use ridge regression (regression with L_2 regularization) of the elementary prediction algorithms. Ridge regression penalizes the euclidian norm of regression coefficients.

We also modify the method for defining the intervals $[p^{min}_i, p^{max}_i]$ of the control parameters for the each node in the hierarchical regression:

1) for each interval $[p^{min}_i, p^{max}_i]$ we calculate the median value p^{median}_i using histogram of the control parameter p_i distribution;

2) divide each interval $[p^{min}_i, p^{max}_i]$ into two using the median value:
$[p^{min}_i, p^{max}_i] \rightarrow [p^{min}_i, p^{median}_i], [p^{median}_i, p^{max}_i]$.

5   Test results and analysis

Main purpose of experiments is to compare proposed hierarchical regression based model with other complex nonlinear methods, since in [9] we compare our model only with elementary prediction algorithms. As such nonlinear method we choose an artificial neural network, because this method is widely used for the travel time prediction and require not so much time for training procedure as support vector regression. Input dataset of the ANN contains both travel time estimations and control parameter values. As the ANN we use feed forward network with one hidden layer.

Models has been tested with the data of all the bus routes in Samara, Russia. The road network consists of 3387 edges. The number of the vehicles equipped with GPS sensors is more than 1500; coordinates of the vehicles positions are obtained every 30 seconds.

Tests were performed on data of all the public vehicles positions in Samara for one week. A validation set size is about 140 million records; a control set size is about 20 million records.

We compare the modified adaptive model with an elementary prediction algorithm based on the exponential kernel, the averaged historical travel time (statistics on the figures) and the neural network model. Number of hidden units were selected by the criterion of the mean absolute error minimization of

the ANN on the training set. We used ANN with 3 hidden units in a hidden layer. The exponential kernel width was also selected experimentally.

The prediction accuracy is evaluated by computing the mean absolute error (MAE) and the mean absolute percentage error (MAPE) depending on the prediction horizon. Fig. 3 shows the experimental results on the training set, fig. 4 shows the experimental results on the control set.
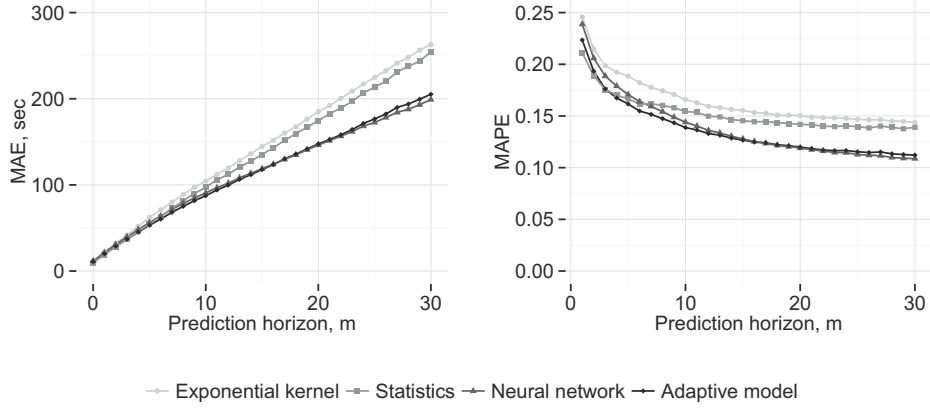

Fig. 3. MAE and MAPE depending on the prediction horizon for the training set.

Fig. 4. MAE and MAPE depending on the prediction horizon for the control set.

In terms of considered accuracy measure the proposed adaptive model based on the hierarchical regression and the neural network model show similar results.

Since the prediction model should be run on-line, the computation time of the model is a critical issue. Next experiment shows the computation time of the

adaptive and the neural network models for a different simultaneously processing number of the vehicles, using laptop computer (Intel Core i5-3740 3.20 GHz, 8 GB RAM) (fig. 5).
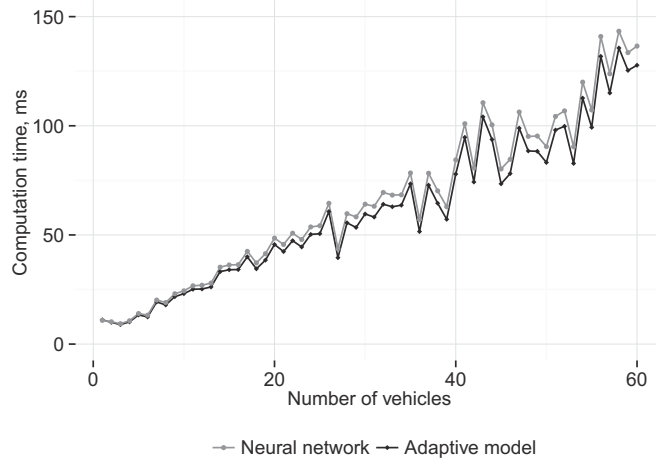

Fig. 5. Computation time of the models.

The processing time for one vehicle takes about 7 ms on average, mainly depends on the required number of travel time predictions and is determined by the route and the position of the vehicle. Computation time of the neural network model is a little bit bigger than computation time of the adaptive model. Nonetheless both models can be used to predict the arrival time of all public vehicles in a large city in real time.

6   Conclusion and future work

In this paper we modify travel time prediction model based on adaptive combination of the elementary prediction algorithms. Our modifications concern the set of elementary prediction algorithms and the aggregation function. The modified model has the same advantages as the base model:

- allows to include prediction algorithms of any type in the adaptive combination;
- has adaptability to the traffic situation changes;
- has adaptability to the factors that has direct or indirect impact on the transport movement and / or prediction result.

In experiments performed on data of all bus routes in Samara (Russia) the modified adaptive model provides almost the same results as the neural network.

The proposed model doesn't require so much time and memory resources for the training as the neural network. This is the main advantage of our model.

Comparison between the models of different classes, including time-series models and nonlinear regression models, will be conducted as our future work.

References

1. Moreira-Matias, L., Mendes-Moreira, J., De Sousa, J.F., Gama, J.: Improving Mass Transit Operations by Using AVL-Based Systems: A Survey. In: IEEE Transactions on Intelligent Transportation Systems, vol. 16 (4), pp. 1636-1653 (2015)
2. Hoogendoorn, S.P.: State-of-the-art of vehicular traffic flow modeling. In: Proceedings of the Institution of Mechanical Engineers. Part I: Journal of Systems and Control Engineering, vol. 215(4), pp. 283-303 (2001)
3. Cryer, J., Chan, K.: Time Series Analysis With Applications in R. Springer-Verlag, New York (2008)
4. Sun, H., Liu, H.X., Xiao, H., He, R.R., Ran, B.: Use of Local Linear Regression Model for Short-term Traffic Forecasting. In: Transportation Research Record, vol. 1836, pp. 143150 (2003)
5. Chen, M., Liu, X., Xia, J., Chien, S.I.: A dynamic bus-arrival time prediction model based on APC data. In: Computer-Aided Civil and Infrastructure Engineering, vol. 19(5), pp.364-376 (2004)
6. Jeong, R., Rilett, L.R.: Bus arrival time prediction using artificial neural network model. In: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, pp. 988-983 (2004)
7. Bin, Y., Zhongzhen, Y., Baozhen, Y.: Bus arrival time prediction using support vector machines. In: Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, vol. 10 (4), pp. 151-158 (2007)
8. Wu, C.-H., Ho, J.-M., Lee, D.T.: Travel-time prediction with support vector regression. In: IEEE Transactions on Intelligent Transportation Systems, vol. 5(4), pp. 276-281 (2004)
9. Agafonov, A., Myasnikov, V.: An Adaptive Algorithm for Public Transport Arrival Time Prediction Based on Hierarhical Regression. In: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, pp. 2776-2781 (2015)