On Predicting Traveling Times in Scheduled Transportation (Abstract)

Avigdor Gal AVIGAL@IE.TECHNION.AC.IL
Avishai Mandelbaum AVIM@IE.TECHNION.AC.IL
Francois Schnitzler FRANCOIS@EE.TECHNION.AC.IL
Arik Senderovich SARIKS@TX.TECHNION.AC.IL
Technion - Israel Institute of Technology, Haifa, Israel

Matthias Weidlich WEIDLIMA@INFORMATIK.HU-BERLIN.DE
Humboldt-Universität zu Berlin, Berlin, Germany

Traveling time prediction

Urban mobility impacts urban life to a great extent. People, living in cities, plan their daily schedule around anticipated traffic patterns. Some wake-up early to “beat” rush hour. Others stay at home and work during days when a convention comes to town.

To enhance urban mobility, much research was invested in traveling time prediction, see (Wu et al., 2004). That is, given an origin and destination, provide a passenger with an accurate estimation of how long a journey lasts. In particular, the ability to predict traveling time in scheduled transportation, *e.g.,* buses, was shown to be feasible (Chien et al., 2002).

In this work, we address the problem of *online travel time prediction* in the context of a bus journey. That is, a journey may be ongoing in the sense that journey events already indicated the progress of the bus on its route. For such an ongoing journey, we are interested in the current prediction of the traveling time from the current bus stop to some destination via a particular sequence of stops, which is defined by the respective journey pattern.

Prediction Approach

To address the problem of online travel time prediction, we investigate a novel use of methods from Queueing Theory and Machine Learning in the prediction process. We propose a prediction engine that, given a scheduled bus journey (route) and a ‘source/destination’ pair, provides an estimate for the traveling time, while considering both historical data and real-time streams of information that are transmitted by buses. To do so, we model buses as clients that go through a journey of segments that are interpreted as a network of queues. We propose a model that uses natural segmentation of the data according to bus stops and a set of predictors, some use learning while others are learning-free, to estimate traveling time.

The model of journey segments. As the foundation of our approach, we propose to model each bus trip by using a segmentation model as follows. A trip between two stops consists of segments, with each segment being represented by a ‘start’ stop and an ‘end’ stop, see Figure 1. Given the first stop of a trip ω_1 and the last stop of a trip ω_n, the intermediate stops are known in advance since each bus follows a predefined journey pattern. Therefore, a trip can be described by segments that are characterized by a pair of stops of the form $\langle \omega_i, \omega_{i+1} \rangle$ (Figure 1). This segmented model, in turn, allows for fine-granular grounding of the prediction of traveling time $T(\langle \omega_1, \ldots, \omega_n \rangle, t_{\omega_1})$ for a sequence of stops $\langle \omega_1, \ldots, \omega_n \rangle$ when departing at time t_{ω_1}: instead of considering only journeys that follow the same sequence of stops $\langle \omega_1, \ldots, \omega_n \rangle$, all journeys that share some segments can be used for prediction.

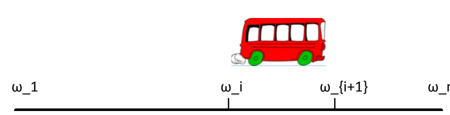

Figure 1. A segmented model of traveling times

Using information on bus stops, the prediction of the journey traveling time $T(\langle \omega_1, \ldots, \omega_n \rangle, t_{\omega_1})$ is traced back to the sum of traveling times per segment. The traveling time per segment is assumed to be independent of a specific journey pattern and, thus, also independent of a specific journey:

$$T(\langle \omega_1, \ldots, \omega_n \rangle, t_{\omega_1}) = \sum_{i=1}^{n-1} T(\langle \omega_i, \omega_{i+1} \rangle, t_{\omega_i}),$$

where $t_{\omega_{n-1}} = t_{\omega_1} + T(\langle \omega_1, \omega_{n-1} \rangle, t_{\omega_1})$.

Prediction based on the snapshot principle. A first set of predictors is grounded in heavy-traffic approximations

Proceedings of the 2^{nd} International Workshop on Mining Urban Data, Lille, France, 2015. Copyright ©2015 for this paper by its authors. Copying permitted for private and academic purposes.

in Queueing Theory. It is non-learning, in the sense that it does not generalize prediction from historical events, but rather uses recent events to predict future traveling times. Applied to our context, the main idea is that a bus that passes through a segment, will experience the same traveling time as another bus that has just passed through that segment (not necessarily of the same type, line, *etc*). Following this line, we define a single-segment snapshot predictor, called Last-Bus-to-Travel-Segment (LBTS).

To use this predictor to address the online travel time prediction problem, it needs to be lifted to a network setting. To this end, we exploit the fact that the snapshot principle holds for networks of queues, when the routing through this network is known in advance (Reiman & Simon, 1990). Clearly, in scheduled transportation, this is the case, so that we define a multi-segment (network) snapshot predictor, called Last-Bus-to-Travel-Network. It is derived by summing up the LBTS predictions for the segments of the respective journey pattern of the bus for which the prediction is made.

Prediction using Machine Learning methods. A second set of predictors comes from Machine Learning and is based on regression trees. They exploit past journey logs to learn a prediction model, and then use this model to make a prediction on new instances of the problem, in our case, traveling times as part of current journeys.

As a first step, we formalize the traveling times prediction problem as a regression problem. Features considered for the regression include the travel time of the last bus that used that segment (LBTS, as introduced above); the interval between the time the last bus left the segment and the estimated time to enter the segment; the day of the week; and the time of the day. For the resulting regression model, various generic algorithms are applied to derive an ensemble of regression trees that is then used to solve the prediction problem. Specifically, random forests, extremely randomized forests, AdaBoost, and gradient tree boosting are leveraged.

In a final step, the above methods originating from Queueing Theory and Machine Learning are combined. That is, we rely on the boosting algorithms and modify them such that the first model considered in the boosting is the snapshot predictor model.

Evaluation

To demonstrate the value of our approach, we tested the proposed predictors using bus data that comes from the bus network in the city of Dublin.[1] The data includes location of buses that is sampled in intervals of 5 to 300 seconds, depending on the current location of the bus.

Using this data, we empirically evaluated the prediction accuracy of the presented methods. The main results of our experiments have been:

- Prediction methods that combine the snapshot principle and Machine Learning techniques are superior in quality of prediction to both snapshot predictors and Machine Learning methods (that do not include the snapshot predictor).
- The prediction error increases with the number of bus stops per journey. However, when considering the relative error, it is stable for all trip lengths, i.e. the predictors do not deteriorate proportionally to the length of the journey (in stops).
- Surprisingly, the snapshot predictor does not deteriorate for longer trips, therefore contradicting the hypothesis that the snapshot predictor would be more precise for journeys with higher temporal proximity to the current journey.

Conclusion

In this work, we presented a novel approach towards predicting travel time in urban public transportation. It is grounded in a partitioning of the travel time into stop-based segments, and combines the use of Machine Learning and Queueing Theory predictors to model traveling time in each segment. Our empirical evaluations confirmed that the combination of methods indeed improves performance. Moreover, we observed that the snapshot predictor is, counter-intuitively, unaffected by the length of a journey. This leads to positive evidence in favor of applying mixed Queue and Machine Learning predictors in similar settings.

This work was supported by the EU INSIGHT project (FP7-ICT 318225).

References

Chien, Steven I-Jy, Ding, Yuqing, and Wei, Chienhung. Dynamic bus arrival time prediction with artificial neural networks. *Journal of Transportation Engineering*, 128 (5):429–438, 2002.

Reiman, Martin I and Simon, Burton. A network of priority queues in heavy traffic: One bottleneck station. *Queueing Systems*, 6(1):33–57, 1990.

Wu, Chun-Hsin, Ho, Jan-Ming, and Lee, Der-Tsai. Travel-time prediction with support vector regression. *Intelligent Transportation Systems, IEEE Transactions on*, 5(4):276–281, 2004.

[1] See also http://www.dublinked.ie/ and http://www.insight-ict.eu/