Fuel consumption based on route choice

Ekaterina GILMAN *, Satu TAMMINEN **, Anja KESKINARKAUS ***, Theodoros ANAGNOSTOPOULOS ****, Susanna PIRTTIKANGAS *, and Jukka RIEKKI *

Abstract: Advances in technology equip traffic domain with instruments to gather and analyze data for safe and fuel-efficient traveling. This article elaborates on the effects from taxi drivers route selection on fuel efficiency. We fuse real driving behavior data from taxi cabs, weather, and digital map information for fuel consumption prediction. That way we compare actually driven trips and their quickest and shortest counterparts for fuel efficiency.

Key Words: fuel-efficiency, data analysis, data fusion, eco-driving.

1. Introduction

Route choice selection is an actively studied research area. Advancements in sensing technologies allow collecting data from real vehicles and analyzing these for possible factors affecting route selection. Examples of such factors are general characteristics like time, distance and even personal characteristics of drivers [4],[5].

Ericsson et al.[1] report that about 8% of fuel could be saved by selecting fuel-efficient routes. In addition, driving behavior has been shown to significantly affect fuel consumption [2]. Therefore, personalized instructions for driving and route selection are considered to be useful [3].

Taxi drivers have driving experience in the cities, hence it is beneficial to mine their route selection strategies to explore if they are fuel-efficient. Our research is based on real data retrieved from trips of taxi cabs. In related research taxi GPS traces have been used, e.g., to discover passenger searching strategies [6] or to find patterns in origin-destination flows [7]. Interested readers are referred to Castro et al.[8] for a review.

In this article, we explore whether routes selected by taxi drivers are more fuel efficient than their corresponding shortest and quickest counterparts. For this, we fuse different sources of data: driving behavior, route characteristics, and weather.

2. Data for the study

2.1 Data sources

For this study we use driving data, digital map data, and weather data.

Driving Data document driving related features with geospatial information of the route, such as speeds used and fuel spent. Driving data are retrieved by Driveco devices (http://eco.driveco.fi) collecting data from the OBDII diagnostics connectors of taxis. Details are given in [3].

Digital map data is retrieved from Digiroad (http://www.digiroad.fi), a database of Finnish road network maintained by National Land Survey of Finland, the Finnish Transport Agency and individual municipalities. This database contains information about the road network and attribute information like speed limits and traffic lights.

Weather data is fetched from Digitraffic service (http://www.infotripla.fi/digitraffic) offering information about the weather and traffic on the Finnish main roads. Road weather information is collected with road weather stations and provided via a Web service interface. One road weather station closest to the city center is selected.

2.2 Data pre-processing

Pre-processing was started by filtering, segmenting and map-matching the driving data retrieved from the Driveco devices. Trips falling outside the local region and trips having insufficient data were removed from the analysis. Data segmentation identifies single trips from Driveco device readings.

Map-matching places the driven route to the road network. We used an incremental map-matching algorithm [9] enhanced with bi-directional Dijkstra Shortest Path implementation of pgRouting package for PostgreSQL DBMS for sparse points.

For each trip we generated corresponding quickest and shortest routes having the same origin and destination as the original trip (Quickest_gen and Shortest_gen). Quickest route is the route requiring least amount of time to travel. Shortest route is the route having minimum distance to travel. Then, for each route driven and for its shortest and quickest versions, we calculated the Digiroad spatial properties and assigned weather measurements.

3. Model

For analysis we used generalized boosted regression model, GBM (gbm library in R package) to predict the fuel (l/km) from the predictor variables describing spatial and weather characteristics of the trips. The model is built with weak learners that are estimated iteratively and combined together to form a strong learner. We have experimented with other machine learning models including k-NN, C4.8, SVM and Neural Networks, but GBM proved to treat efficiently the complex relationships within our dataset.

After removing outliers, the dataset contains 2548 trips. The whole dataset was divided into 11 folds (caret R package). One

* Center for Ubiquitous Computing, University of Oulu, Finland
** Biomimetics and Intelligent Systems Group, University of Oulu, Finland
*** Center for Machine Vision and Signal Analysis, University of Oulu, Finland
**** Ordnance Survey: Britain's Mapping Agency, United Kingdom
E-mail: ekaterina.gilman@oulu.fi

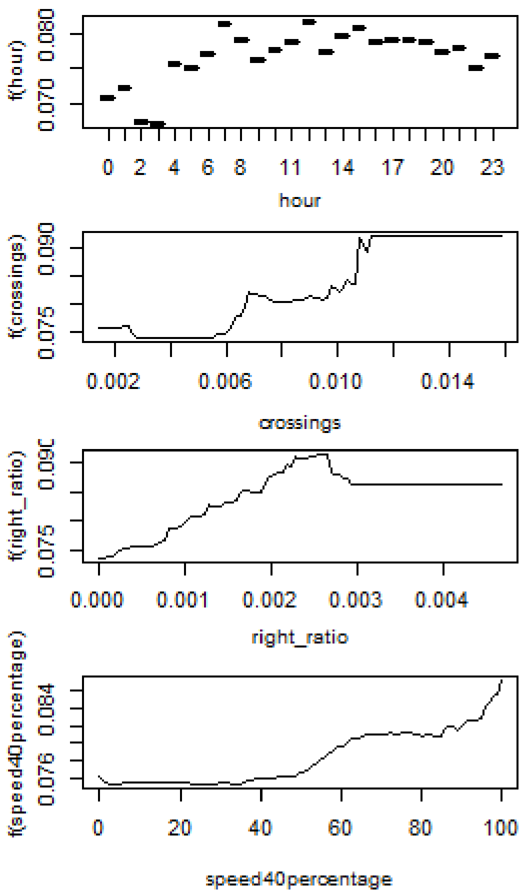

Fig. 1 Fuel prediction performance of selected model.

fold is left for validation of the model (test set), the rest 10 folds (training set) were used to select the proper set of predictors and train the model. Feature set was selected with 10-fold cross validation keeping only important predictors after each round. Based on the averaged performance of models in 10-fold cross validation, the best feature set is selected. The final model was trained with the selected feature set on the training set. Final predictors of the model are: hour of day and day of week; number of crossings and pedestrian crossings; turning right/left ratio; percent of the trip driven on the roads with speed limit 40, 50, 60, and 80 km/h; number of traffic lights; air, dew point, and road surface temperature; amount of snow; not motorway, roundabout,and slip road type; road and street type;local main, connecting, feeder street functional class of the road. Model performance: training set (RMSE=0.014, cor=0.82), test set=(RMSE=0.020, cor=0.60). This model was used to predict the fuel for Shortest_gen and Quickest_gen trips. Table 1 lists the feature set used for the model, as well as importance of predictors given by GBM.

Fig. 1 shows the effect on fuel consumption of selected predictors of the model. We can see that rush hours (hour variable), high proportion of crossings, right turns ratio (right_ratio), percent of the trip driven on the roads with speed limit of 40 km/h (speed40percentage) increase the fuel consumption.

4. Results

Created model was used to predict the possible fuel consumption (L/km) for trips which differ at least in 20% from corresponding Shortest_gen and Quickest_gen counterparts. Fig. 2 demonstrates the results obtained. Wilcoxon signed rank test

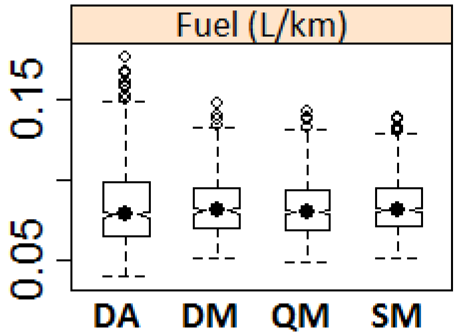

Fig. 2 Modeled fuel for the actually driven and generated routes (DA-driven actual, DM - driven modeled, QM-Quickest_gen modeled, SM-Shortest_gen modeled).

gives the following results V-value(P-value) : Driven actual less than Driven modeled = 106500 (0.3004), Driven modeled less than Shortest_gen modeled = 105300(0.2209), Driven modeled greater than Quickest_gen modeled = 116960 (0.0535), Shortest_gen modeled greater than Quickest_gen modeled = 58268(<2.2e-16). Quickest_gen trips demonstrate best fuel-efficiency per km when compared to Driven and Shortest_gen. Driven trips demonstrate less fuel consumption per km than Shortest_gen, but more fuel consumption than Quickest_gen. However, statistically significant difference (at the 0.05 significance level) is observed only between Shortest_gen modeled and Quickest_gen modeled.

References

[1] E. Ericsson, H. Larsson, K. Brundell-Freij: Optimizing route choice for lowest fuel consumption potential effects of a new driver support tool. *Transport. Res. Part C: Emerg. Technol.*, Vol. 14, No. 6, pp. 369383, 2006.

[2] M. Sivak, B. Schoettle: Eco-driving: strategic, tactical, and operational decisions of the driver that influence vehicle fuel economy. *Transport Policy*, pp. 9699, 2012.

[3] E. Gilman, A. Keskinarkaus, S. Tamminen, S. Pirttikangas, J. Rning, J. Riekki: Personalized assistance for fuel-efficient driving. *Journal of Transportation Research Part C: Emerging Technologies*, Vol. 58, Part D, pp. 681-705, 2015.

[4] M. Ben-Akiva, M.J. Bergman, A.J. Daly, R. Ramaswamy: Modeling inter-urban route choice behaviour. *Proceedings from the Ninth International Symposium on Transportation and Traffic Theory*, VNU Science Press, Utrecht, Netherlands, pp. 299330, 1984.

[5] E. Ben-Elia, R. Ishaq, Y. Shiftan: If only I had taken the other road...: regret, risk and reinforced learning in informed route-choice. *Transportation* Vol. 40, No, 2, pp. 269293, 2013.

[6] B. Li, D. Zhang, L. Sun, C. Chen, S. Li, G. Qi, Q. Yang: Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. *Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pp. 6368, 2011.

[7] W.Zhang, S. Li, G. Pan: Mining the semantics of origin-destination flows using taxi traces. *Proceedings of the Workshop of Ubiquitous Computing*, pp. 943949,2012.

[8] P.S. Castro, D. Zhang, C. Chen, S. Li, G. Pan: From taxi GPS traces to social and community dynamics: A survey. *ACM Comput. Surv.* Vol. 46, No.2, Article 17, 2013.

[9] S. Brakatsoulas, D. Pfoser, R. Salas, C. Wenk: On map-matching vehicle tracking data. *Proceedings of the 31st international conference on Very large data bases (VLDB '05)*, pp. 853-864, 2005.