European Journal of Operational Research 252 (2016) 322–333

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Large-network travel time distribution estimation for ambulances

Bradford S. Westgate*, Dawn B. Woodard, David S. Matteson, Shane G. Henderson

253 Swanson Academic Center, Alma College, 614 W. Superior St., Alma, MI 48801, United States

ARTICLE INFO

Article history:
Received 26 August 2014
Accepted 3 January 2016
Available online 11 January 2016

Keywords:
Transportation
Traffic
OR in health services
Travel time estimation
Markov chain Monte Carlo

ABSTRACT

We propose a regression approach for estimating the distribution of ambulance travel times between any two locations in a road network. Our method uses ambulance location data that can be sparse in both time and network coverage, such as Global Positioning System data. Estimates depend on the path traveled and on explanatory variables such as the time of day and day of week. By modeling at the trip level, we account for dependence between travel times on individual road segments. Our method is parsimonious and computationally tractable for large road networks. We apply our method to estimate ambulance travel time distributions in Toronto, providing improved estimates compared to a recently published method and a commercial software package. We also demonstrate our method's impact on ambulance fleet management decisions, showing substantial differences between our method and the recently published method in the predicted probability that an ambulance arrives within a time threshold.

1. Introduction

Estimates of ambulance travel times on arbitrary routes in a road network are used in ambulance dispatch decisions, base location algorithms, and real-time redeployment methods (Brotcorne, Laporte, & Semet, 2003; Dean, 2008; Goldberg, 2004; Maxwell, Restrepo, Henderson, & Topaloglu, 2010; Schmid, 2012). In many of these applications it is important to capture the uncertainty in the travel time, by predicting the entire travel time distribution rather than just the expected travel time (Ingolfsson, Budge, & Erkut, 2008; Zhen, Wang, Hu, & Chang, 2014). For instance, taking into account uncertainty in the travel time of ambulances to the scene of an emergency can substantially increase the survival rate of cardiac patients, by improving fleet management decisions and thus reducing response times (Erkut, Ingolfsson, & Erdoğan, 2008; McLay, 2010). Also, ambulance fleet performance is measured by the fraction of emergency calls for which the response time is less than a specified threshold, and forecasting this performance measure requires travel time distribution information (Mason, 2005). Travel time distributions are also used in applications for other vehicle fleets, including calculation of driving directions for private vehicles using taxi data (Yuan et al., 2010), allocation of railcars (Topaloglu, 2006), and routing and scheduling of courier vehicles (Potvin, Xu, & Benyahia, 2006).

We propose a regression approach for estimating the distribution of an ambulance travel time on an arbitrary route in a road network. The prediction depends on the route and on explanatory variables such as the time of day and day of week. Our method uses information from historical trips on the network, specifically the total travel time and estimated route for each trip. In order to predict the travel time distribution for a particular route, we do not require historical trips that take precisely the same route. Instead, our statistical approach uses information from all the historical trips by learning shared properties like the effects of time of day and types of road traversed. The model we use is intuitive and its parameters are interpretable. Our method is computationally efficient, scaling effectively to large road networks and large historical trip databases.

Two features of ambulance travel times motivate our modeling choices. First, ambulances traveling at lights-and-sirens speeds are less affected by traffic than other vehicles (Aladdini, 2010; Kolesar, Walker, & Hausner, 1975; Westgate, Woodard, Matteson, & Henderson, 2013). Therefore, historical ambulance trips are the most relevant source of information for travel times, and real-time traffic flow information from other vehicles is less useful. Second, ambulance trips are comparatively rare, implying that ambulance data is sparse in time and road network coverage. Roads that are not major thoroughfares may have only a few ambulance trips on them per year.

To estimate the route taken in the historical trips, we use Global Positioning System (GPS) measurements taken during travel. This source of data is called floating car data or automatic vehicle location data, and is increasingly available for many types

* Corresponding author. Tel.: +1 989 463 7264.
E-mail address: westgatebs@alma.edu (B.S. Westgate).

http://dx.doi.org/10.1016/j.ejor.2016.01.004
0377-2217/

of vehicles, including ambulances, taxis, and personal vehicles, via GPS-enabled smartphones or 2-way navigation devices (e.g. Garmin or TomTom) (Hofleitner, Herring, Abbeel, & Bayen, 2012a). Unlike other sources of travel time data, it does not require instrumentation on the roadway, and thus is the only source of data available to estimate travel times that has the prospect of comprehensive network coverage.

Despite the rise in availability of floating car data, there are still few methods available to utilize floating car data for travel time distribution prediction. Hofleitner, Herring, and Bayen (2012b) and Hofleitner et al. (2012a) take a traffic flow perspective, modeling at the level of the network link (a road segment between two intersections). They use a dynamic Bayesian network for the unobserved traffic conditions on links and model the link travel time distributions conditional on the traffic state. Their method is applied to a subset of the San Francisco road network with roughly 800 links, predicting travel times using taxi fleet data and validating with additional data sources.

In previous work, we introduced a Bayesian model for simultaneous travel time distribution and path estimation for a set of vehicle trips (Westgate et al., 2013). Like Hofleitner et al., we modeled travel times at the link level. Our method was applied to estimate ambulance travel times on a subregion of Toronto.

In an early paper, Erkut, Fenske, Kabanuk, Gardiner, and Davis (2001) estimate ambulance and fire truck speeds in St. Albert, AB, Canada, as part of a study to select new locations for a fire station and ambulance base. They use three road classes (freeway, main roads, and residential roads), and also account for time-of-day and season by estimating different speeds for rush hour and non-rush hour and for summer and winter. They estimate average speeds using historical data, interviews with drivers, and road tests. They do not consider the distribution of travel times.

Jenelius and Koutsopoulos (2013) propose a framework for estimating the distribution of travel times while incorporating weather, speed limit, and other explanatory factors. They point out that previous approaches such as Hofleitner et al. (2012a, 2012b); Westgate et al. (2013) assume that the link travel times are independent within a vehicle trip, perhaps conditional on the traffic state. This contrasts with empirical evidence suggesting that the link travel times are strongly correlated, even after conditioning on time of day and other explanatory factors (Bernard, Hackney, & Axhausen, 2006; Ramezani & Geroliminis, 2012). Therefore, they capture correlation using a moving average specification for the link travel times. Their framework is applied to estimate travel times for a particular route in Stockholm.

In contrast to these approaches, we model travel times at the trip level instead of the link level. This naturally incorporates dependence between link travel times. The ambulance route is taken into account in the specification of the trip travel time parameters, such as the median travel time. This trip-level approach is related to the regression approach of Budge, Ingolfsson, and Zerom (2010b), who model the travel time distribution for an ambulance trip as a function of shortest-path distance between the start and end locations. They assume that the log travel time follows a *t*-distribution, where the centering and scale parameters are either a nonparametric or parametric function of the shortest-path distance. These functional forms enable their method to be flexible but still interpretable. Like them we take a regression approach, but we also incorporate dependence on the route taken, time of day, and other explanatory factors, justifying our modeling choices empirically. This captures the fact that locations near primary roads can be reached more quickly than other locations, for example. A downside of modeling at the trip level is that travel time predictions cannot be updated to reflect changing conditions while a vehicle is enroute. However, this is not a drawback in the ambulance setting, because travel time estimates are used for ambulance dispatch decisions and base placement, rather than for route selection.

We use our method to predict ambulance travel times for the entire road network of Toronto. The size of the road network (68,272 links) is an order of magnitude larger than in previous applications of travel time distribution estimation based on floating car data (Budge et al., 2010b; Hofleitner et al., 2012a; Hofleitner et al., 2012b; Jenelius & Koutsopoulos, 2013; Westgate et al., 2013), and the number of historical vehicle trips (157,283) is also larger than these previous applications. We compare the prediction accuracy of our method to that of Budge et al. (2010b), Westgate et al. (2013), and a commercial software package for mean travel time estimation. We also consider the effect of various simplifications of our model, and investigate the accuracy of our model when the time effect on travel times is artificially inflated.

Finally, we evaluate the effect of using our method for ambulance fleet management, relative to that of Budge et al. (2010b). We select a set of representative ambulance posts in Toronto, and calculate which ambulance post is estimated to be the closest in median travel time to each intersection in Toronto. Many intersections have different estimated closest posts, according to the two methods, so the two methods would recommend that a different ambulance respond to emergencies at these locations, if the closest ambulance is dispatched. Next, we calculate the probability that an ambulance is able to respond on time (within a specified time threshold) from the closest post to each intersection of the city. We find substantial differences in these probabilities between the two methods. These appear to arise because our method captures differences in speeds between different types of roads, unlike the method of Budge et al.

Commercially available vehicle travel time estimates typically consist of estimated expected travel times rather than distribution estimates, so they cannot be used to calculate the probability an ambulance arrives within a time threshold, or for ambulance deployment algorithms where simulated travel times are needed. Also, these estimates are calculated for standard vehicle speeds, not lights-and-sirens ambulance speeds. However, they are still useful for point estimation performance comparisons, as long as they are corrected for bias. Specifically, we investigate travel time estimates from TomTom, a maker of navigation devices. Bias adjustment does not fully account for the differences between the TomTom context and ours, so our results should not be interpreted as an evaluation of the quality of TomToms estimates.

The article is organized as follows. In Section 2, we introduce the data from Toronto and highlight the exploratory data analysis that motivates our modeling choices. In Section 3, we introduce our statistical model and estimation method. In Section 4, we give results and compare to the alternative methods. We draw conclusions in Section 5. In Appendices Appendix A–Appendix C, we give details on data preprocessing, fastest path estimation, and our implementation of the method of Budge et al.

2. Toronto EMS data

We use our method for a study of ambulance travel times in Toronto, Ontario. The goal is to estimate the distribution of time required for an ambulance to drive to the scene of a high-priority emergency, in which case the ambulance uses lights and sirens, and travels at high speed. The data are provided by Toronto EMS (Emergency Medical Services), and include all such ambulance trips in Toronto during the years 2007 and 2008. We analyzed a subset of these data from the Leaside region of Toronto in previous work (Westgate et al., 2013); here we estimate travel times on the entire Toronto road network, which consists of 68,272 links.

The data associated with each trip include the approximate start and end times and locations of the trip, as well as sparse