Inferring the Purposes of using Ride-Hailing Services through Data Fusion of Trip Trajectories, Secondary Travel Surveys, and Land Use Data

Sanjana Hossain\(^1\) and Khandker Nurul Habib\(^1\)

Abstract
This paper presents a data fusion methodology for inferring trip purposes from GPS trajectories of ride-hailing services in Toronto. The methodology has a discrete choice model at its core that predicts the most probable purpose distributions using only basic trip-related information such as approximate pick-up and drop-off locations, trip start times, and land use characteristics around the origins and destinations. The choice model is estimated using revealed trip purpose data from a small-sample travel survey augmented by land use information from an enhanced point of interest database and the census. The methodology is applied to the trajectories of commercial ride-hailing trips made in Toronto between September 2016 and September 2018. For the core choice model, multinomial, nested, and mixed multinomial logit models are compared. Validation of the inferred trip purposes using the trip purpose proportions from another independent survey (not used in choice model estimation) reveal that the multinomial logit model can infer ride-hailing trip purpose distribution with reasonable accuracy. The inferred purpose distribution explains the nature of ride-hailing trips and provides important context of travel demand generated by the services. The results indicate that although ride-hailing services are mostly used for discretionary activities, they also play important roles in daily commuter travel. A quarter of the total weekday ride-hailing trips were made for work- and school-related activities. With increasing ridership, these services may start influencing conventional travel modes and thereby adversely affect the level of traffic congestion and transit ridership in the city.

In recent years, ride-hailing services offered by private transportation companies (e.g., Uber, Lyft, DiDi) have experienced rapid growth owing to their flexible, reliable, and cost-effective mobility options. Between September 2016 and March 2019, ride-hailing trips have grown by more than 180% in the City of Toronto (1). This increasing market penetration is making it imperative to have clear understanding of the characteristics of these trips and how they are changing the travel behavior of people. Such holistic and in-depth analysis requires large amount of data that can be extracted from the GPS trajectories of the ride-hailing trips. The trajectory data contains a wealth of travel-related information, such as when and where ride-hailing passengers move around the city in a high resolution. However, being passively collected, it does not contain semantic information about why the trips are being made. Trip purpose information directly relates to the activities for which ride-hailing is used; thus, it provides important context of travel demand generated by the services. Although questionnaire-based travel surveys can easily collect trip purposes, they suffer from limitations of high survey cost, heavy respondent burden, short time and space coverage, and trip underreporting (2). Also, these surveys usually underrepresent younger population who tend to use ride-hailing services more frequently (1, 3, 4). Thus there is a trade-off of information between ride-hailing trajectory and travel survey data, where the former provides rich spatial and temporal information but lacks trip purposes, and the latter provides detailed trip purposes but suffers from small sample size and inaccuracies. This paper proposes a data fusion methodology that leverages both of these information sources to infer ride-hailing trip purposes.

\(^1\)Department of Civil & Mineral Engineering, University of Toronto, Toronto, Ontario, Canada

**Corresponding Author:**
Khandker Nurul Habib, khandker.nurulhabib@utoronto.ca
The study uses trajectories of ride-hailing trips made in the City of Toronto between September 2016 to September 2018 to investigate the patterns of ride-hailing trip purposes at the drop-off locations. In particular, a data fusion-based methodology is proposed that uses an econometric choice model to impute the most probable purposes of individual ride-hailing trips. Application of choice theory induces important behavioral features. It captures the reality that passengers arriving at the same location can be driven by different purposes, and the purposes may also differ based on trip start times and origins. The proposed fusion method relies only on basic information such as approximate pick-up and drop-off locations of trips, start times, land use information, and revealed trip purpose data from small-sample travel surveys. The inferred trip purpose pattern extends our understanding about the increasing ride-hailing ridership, which in turn will facilitate efficient urban policy planning.

The remainder of the paper is organized as follows. The second section presents a review of methodologies used in trip purpose imputation research. The third section presents the proposed data fusion method. The fourth section presents the application of the method for the City of Toronto and discusses the results obtained. The final section summarizes the findings and limitations of the study and provides guidelines for future research.

**Previous Research on Trip Purpose Inference**

Trip purpose inference has emerged as a popular research topic over the last two decades ever since the advent of passive data collection technologies, such as GPS devices, mobile phones, transit smart cards, and so forth. The developed algorithms vary widely in their complexity, input data requirements, and performance accuracy. This section reviews the previous works based on the methods and the input/context variables used. More detailed reviews (especially for inference of purpose from GPS-based travel survey data) can be found in Gong et al. (5), Xiao et al. (6), and Ermagun et al. (7).

Earlier studies mostly relied on deterministic rule-based methods that utilize predefined heuristics to infer the trip purposes (8–11). These algorithms are criticized for lack of generalization and poor transferability. More recent studies tend to favor probabilistic methods and machine learning. For predictions, the probabilistic algorithms calculate the occurrence probability of each trip purpose, whereas the machine learning algorithms report whether a specific purpose is selected or not based on classification, and pattern recognition methods. The majority of the studies that adopt the probabilistic method utilize Bayes’ theorem of conditional probability to estimate trip purposes based on different spatial and temporal constraints (e.g., 12–14). However, a few studies (7, 10, 15, 16) also utilize the theory of choice modeling, particularly the multinomial and the nested logit models. In machine learning techniques, the most popular tools include clustering algorithms (17), decision trees (15, 18), and random forest classifiers (7, 19–21); although some studies use more complex techniques like artificial neural networks (6, 22) and continuous Markov models (23).

Researchers have compared the performance accuracy of different categories of algorithms. Ermagun et al. (7), Zhu (16) and Oliveira et al. (15) found that machine learning algorithms (random forest and decision tree) have higher predictive accuracy than probabilistic methods (regularized multinomial logit and nested logit models) while using detailed contextual information, such as people’s socio-economic attributes, previous activity type, current activity duration, and characteristics of nearby places from Google Places API. However, some of the studies made context-specific simplifying assumptions. For example, Ermagun et al. (7) and Zhu (16) did not predict home, work, and education trips, since these regular activities can be easily labeled using user-provided home and work locations from smartphone-based trajectories (7) or multiple-day continuously observed travel patterns from transit smart card data (16). This exclusion improved the prediction accuracy of the other activity types with relatively small sample sizes in the training data. On the other hand, Wu et al. (19) did not find any striking differences in performance between rule-based and random forest models when contextual attributes like time and activity data are used. These contrasting results indicate that the application context and input variables play important roles in accurate inference of trip purpose.

The most promising input variables for trip purpose inference models are reported to be the land use characteristics, points of interest information, and trip timing (trip start and end time, day of the week, etc.). Activity duration, frequency of visit to a destination, and other tour-based information also play essential roles in inference of trip purpose from GPS-based travel surveys and transit smart card data (11, 12, 23, 24). Socio-demographic characteristics of the respondents and key addresses collected in GPS-based surveys also improve the inference accuracy by providing additional contexts about activities performed (6, 20, 25). Some recent studies on inference of purpose of taxi trips have explored the potential of data repositories like Area of Interest (26) and social network check-in data (14, 17).

Many of the contextual variables, including activity duration, frequency of visit, and tour-based information are not available for inference of ride-hailing trip
purpose. Ride-hailing services record only a part of tour trajectory, and the socio-demographics of the users (if available) are not shared because of privacy concerns. Moreover, it is difficult to obtain reasonably large estimation/training samples for the inference model. All these issues pose unique challenges for ride-hailing trip purpose inference that have not been sufficiently tackled in the existing literature. Most of the studies that investigated ride-hailing travel behavior relied solely on primary or secondary data sets like household travel surveys. For example, Young and Farber (3) analyzed the purpose of ride-hailing trips using data from a large-scale household travel survey. However, trip purposes inferred from ride-hailing trajectories can extend our understanding beyond what is possible by summary of household travel surveys or small-sample ride-hailing specific surveys. They have the potential to provide important context of ride-hailing demand generation at spatial and temporal resolutions unavailable by survey results alone. Despite such immense potential, minimal research effort has been applied to inferring ride-hailing trip purposes from passively collected pick-up and drop-off locations and times. Recently, Dias et al. (4) attempted to infer purposes from publicly available anonymized ride-hailing data by fusing information from multiple sources. However, the purposes were based on locations visited and not the activities performed at those locations.

It is evident that there is a gap in the literature in relation to ride-hailing trip purpose inference. This study attempts to narrow the gap by proposing an econometric data fusion approach for inferring trip purpose from passively collected ride-hailing trajectory data. The fusion tool is based on a choice model that can infer detailed purposes of individual trips using limited context-specific variables and relatively small estimation data.

Data Fusion Methodology

Overview of the Fusion Methodology

Figure 1 presents the methodology used in this study for inferring observed destination purposes of ride-hailing trips. The core of the fusion tool is a discrete choice model that predicts the most likely trip purpose outcomes by fusing GPS attributes with travel survey and land use data. The model is first estimated using revealed trip purpose responses from the travel survey data. Land use information obtained from an enhanced point of interest data repository and census data are fused with the survey data to provide contexts about typical activities performed at the origins and the destinations. The estimated model is then applied to impute the missing purposes of the ride-hailing trips that are also augmented with land use information.

Discrete choice theory is used to develop the core model of trip purpose inference mainly to leverage its strong behavioral foundation—the random utility maximization (RUM) theory. Under the RUM assumption, the choice model calculates the probability of each trip purpose based on its systematic/observable utilities. It is anticipated that such a model will be able to accurately infer the most probable trip purposes using the (approximate) pick-up and drop-off locations, start times of the ride-hailing trips, and limited context-specific data.

Formulation of Discrete Choice Models

The study tests the performance of three discrete choice model formulations: the multinomial logit model (MNL), the nested logit model (NL), and the mixed multinomial logit model (MMNL). Among these structures, MNL is the most popular (27) because of its well-known advantage of tractability (closed-form probability function). The model assumes that the random utility components of the trip purpose alternatives are independently and identically distributed (i.i.d.) type I extreme values, which leads to the following probability of alternative \( i \) (\( i = 1, \ldots, J_n \)) for individual \( n \) (\( n = 1, \ldots, N \)):

\[
P_{im} = \frac{e^{\mu V_{in}}}{\sum_j e^{\mu V_{jn}}}
\]

where \( V_{in} = \beta X_{in} \) is the systematic utility, \( \beta \) is the parameter vector, \( X_{in} \) is a vector of observed variables relating to alternative \( i \), and \( \mu \) is the scale parameter (for estimation, \( \mu \) is unidentifiable and typically fixed at 1). However, the rigid error structure of the MNL model cannot deal with the independence from irrelevant alternatives (IIA) property, where all trip purpose alternatives are perceived to be perfect substitutes of each other (28, 29). In reality, this may not be the case. To overcome this limitation, the NL model is tested by allowing nests of alternatives, such that IIA holds within each nest, but it does not hold for alternatives in different nests (30–32). In NL formulation, the choice probability of a trip purpose \( i \) within a nest \( M \) (containing \( m \) alternatives) is:

\[
P_{im} = \frac{e^{\mu_j V_{in}}}{\sum_m e^{\mu_j V_{in}}}
\]

where \( \mu_j \) is the root scale parameter and \( \mu_M \) is the scale parameter of nest \( M \). \( \mu_j/\mu_M \) is the coefficient of expected maximum utility which should be between 0 and 1. If a particular trip purpose alternative \( l \) is non-nested, its probability is

\[
P_{ln} = \frac{e^{\mu_l V_{in}}}{e^{\mu_l V_{in}} + \sum_{j \neq l} e^{\mu_j V_{in}}}
\]
NL is still limited and cannot capture many forms of unobserved heterogeneity. Moreover, the IIA property holds for the alternatives within a nest. To test a more flexible error structure, an MMNL is estimated. Details of MMNL can be found in: Train (29), Walker (33), and Hensher and Greene (34). In this formulation, the total utility $U_{in}$ of alternative $i$ as perceived by individual $n$ is:

$$U_{in} = V_{in} + \eta_{in} + \epsilon_{in}$$

(4)

where the random utility is made up of two components: $\epsilon_{in}$ is the i.i.d. type I extreme value distributed across all alternatives $J$, and $\eta_{in}$ is the multivariate normally distributed. $\eta_{in}$ can be further decomposed as $\eta_{in} = \sigma_i \xi_{in}$, where $\xi_{in}$ is a standard normal variable and $\sigma_i$ is a positive scale factor.

While many alternative error correlation patterns can be captured by an MMNL model, in this investigation, it was found that only a heteroskedastic MMNL can be valid (based on statistical significance tests of additive error variances, $\sigma$). The covariance matrix of the model has elements in its main diagonal only, that is, there is no covariance among the model’s error terms (in other words, the covariance matrix consists of only differently-scaled variances and no off-diagonal covariances). Here, the probability of trip purpose alternative $i$ for individual $n$ can be obtained as:

$$P_{in} = \frac{1}{D} \sum_{d=1}^{D} e^{\mu \left(b_n x_n + \sigma_i \xi_{nid}^d \right)} \sum_{i} e^{\mu \left(b_i x_i + \sigma_i \xi_{iid}^d \right)}$$

(5)

where $\xi_{nid}^d$ denotes draw $d$ from the distribution of $\xi_{nid}$.

The MNL and NL models are estimated using the classical maximum likelihood estimation technique. The MMNL model is estimated by maximum simulated likelihood estimation technique and by using Halton draws for error simulation. All of these are done through programs written in GAUSS (35).

**Empirical Analysis for the City of Toronto**

**Description of Data Sets Used**

**Ride-Hailing Trip Trajectory Records.** The main data set used in this research was the trajectories of all ride-hailing trips made in Toronto between September 2016 and September 2018. The data set was obtained from the City of Toronto as part of a research collaboration. It contains trip-level information, including spatial coordinates and timestamps of pick-up and drop-off locations. However, to protect users’ privacy, the coordinates are mapped to the nearest intersections, and there are no user IDs. Thus, each record in the data set must be treated as an independent trip with no means of tracking multiple rides by the same user. A summary of the ride-hailing service usage pattern in the city based on the analysis of the data has been published by the City of Toronto Big Data Innovation Team (1).

This study uses the ride-hailing trips made between September and December 2016, which amounts to about 6.95 million trips after necessary data cleaning. Figure 2 shows the spatial distribution of these trips on weekdays and weekends. On both types of days, the destinations of the trips are highly concentrated around Downtown.
Toronto and other major activity hubs. However, almost 1.5 times more trips are generated on a weekend than on a weekday. To keep the computation burden associated with the inference process reasonable, a 20% random sample was generated that corresponds to a total of 1,390,527 ride-hailing records. As the primary objective of this research, the missing destination purposes of this sample of ride-hailing trips were inferred.

**Person Trip Survey Data.** As part of an R&D project (TTS2.0) by the authors’ research group, a web-based travel survey was conducted in August and November 2017 among the residents of Greater Toronto and Hamilton Area. The survey collected travel diaries along with home and work locations, and household and individual level socio-economic information. The diaries contain trip start times, origin and destination locations, travel modes, and reported trip purposes, classified into 13 categories (shown in Table 1). A subset of this survey data comprising trips originating and terminating within the City of Toronto was taken. After cleaning for missing information, a final data set of 5,065 trip records covering both summer and fall seasons was obtained. This data set was used for the estimation/training of the trip purpose classification models in this study.

It should noted that the estimation sample contains trips of all modes, and not exclusively of the ride-hailing mode. As such, obtaining a representative ride-hailing survey sample that is large enough to allow model estimation can be difficult. However, the performance of the fusion method is validated by matching the predicted market shares of the trajectory data with observed trip purpose market shares obtained exclusively for ride-hailing trips from an independent travel survey.

**Enhanced Points of Interest Data.** Enhanced Points of Interest (EPOI) is a national database of Canadian business and recreational points of interest maintained by DMTI Spatial Inc. (36). It contains the geocoded locations of the points of interest along with their North America Standard Industry Classification System (NAICS) codes. These codes place each EPOI in categories like healthcare facilities, shopping centers, educational services, accommodation and food services, and so forth. This open-source data was processed to generate the counts of establishments of different EPOI categories within each dissemination area. As shown in Figure 1, the EPOI data is fused with both the estimation and the inference data sets to provide detailed land use characteristics of trip origins and destinations. This, in turn, provides important context about the typical activities performed at these locations.

**2016 Canadian Census Data.** Data from the 2016 Canadian Census (37) is used to derive the number of private dwellings within each dissemination area. A dissemination area (DA) is the smallest geographic unit for which census information is publicly available in Canada. Accordingly, the location coordinates of the

![Figure 2. Spatial distribution of ride-hailing drop-off locations on an average weekday (left) and on an average weekend (right) at the dissemination area level.](image-url)
| Table 1. Descriptive Statistics of the Datasets |
|---------------------------------------------|
| Ride-hailing trajectory data \( (N = 1,390,527) \) (inference data) | Mean | SD | Minimum | Maximum |
| Count of total dropoffs at the DA level | 372.20 | 1456.68 | 0 | 55.265 |
| Weekday distribution of trips | | | | |
| Number of weekday records (Monday–Friday) | 910,475 | (65.5) | | |
| Number of weekend records (Saturday–Sunday) | 480,052 | (34.5) | | |
| Trip start time distribution | Count | Proportion (%) | | |
| Morning (6:01 a.m.–10:00 a.m.) | 172,939 | (12.44) | | |
| Midday (10:01 a.m.–3:00 p.m.) | 260,337 | (18.72) | | |
| Afternoon (3:01 p.m.–8:00 p.m.) | 384,268 | (27.63) | | |
| Evening (8:01 p.m.–12:00 a.m.) | 350,211 | (25.19) | | |
| Overnight (12:01 a.m.–6:00 a.m.) | 222,772 | (16.02%) | | |
| Person trip survey data \( (N = 5,065) \) (estimation data) | Mean | SD |
| Trip distance (km) | 8.119 | 8.821 |
| Start time distribution | Count | Proportion (%) | | |
| Morning (6:01 a.m.–10:00 a.m.) | 1,319 | (26.04) | | |
| Midday (10:01 a.m.–3:00 p.m.) | 1,116 | (22.03) | | |
| Afternoon (3:01 p.m.–8:00 p.m.) | 2,005 | (39.59) | | |
| Evening (8:01 p.m.–12:00 a.m.) | 518 | (10.23) | | |
| Overnight (12:01 a.m.–6:00 a.m.) | 107 | (2.11) | | |
| Weekday versus weekend distribution of trips | Count | Proportion (%) | | |
| Number of weekday records (Monday–Friday) | 4,470 | (88.25) | | |
| Number of weekend records (Saturday–Sunday) | 595 | (11.75) | | |
| Seasonal distribution | Count | Proportion (%) | | |
| Fall (September–December) | 3,577 | (70.62) | | |
| Summer (May–August) | 1,488 | (29.38) | | |
| Reported trip purposes | Proportion (%) | Reported trip purposes | Proportion (%) | |
| Home | 37.26 | Recreation, sports, leisure | 3.99 |
| Work | 23.71 | Arts, health and personal care | 3.30 |
| Education | 4.30 | Services | 1.15 |
| Daycare | 0.93 | Visiting friends, family | 1.78 |
| Facilitate passenger | 2.78 | Worship, religion | 0.39 |
| Shopping and errands | 11.85 | Other | 4.82 |
| Eat out | 3.75 | na | na |
| EPOI data (fused with estimation and inference data to provide land use contexts at trip ends) |
| NAICS major industry | Mean | SD | Minimum | Maximum |
| Manufacturing | 11.91 | 62.38 | 0.00 | 3387.10 |
| Retail trade | 34.53 | 105.96 | 0.00 | 4332.26 |
| Finance and insurance | 10.85 | 47.62 | 0.00 | 1241.94 |
| Professional, scientific, and technical services | 30.00 | 133.00 | 0.00 | 5132.26 |
| Educational services | 9.52 | 25.42 | 0.00 | 825.81 |
| Health care and social assistance | 25.80 | 116.81 | 0.00 | 3900.00 |
| Arts, entertainment, and recreation | 5.31 | 26.48 | 0.00 | 1174.19 |
| Accommodation and food services | 21.03 | 58.62 | 0.00 | 932.26 |
| Other services (except public administration) | 32.17 | 82.61 | 0.00 | 3219.35 |
| Public administration | 1.59 | 17.48 | 0.00 | 900.00 |

(continued)
other data sets are associated with their corresponding DA, which indirectly takes care of different accuracies of the spatial coordinates in the ride-hailing trajectories and the survey data sets. Similar to the EPOI data, the number of private dwellings within each DA is fused with the estimation and the inference data sets as a contextual variable.

2016 Transportation Tomorrow Survey Data. The Transportation Tomorrow Survey (TTS) [38] is a large-scale household travel survey conducted in the Greater Toronto and Hamilton Area once every five years. It collects travel diaries (for weekdays) and individual and household level socio-economic attributes. Its latest iteration was from September 2016 to December 2016, in which ride-hailing was included as a travel mode. This generated a sample of 1,264 ride-hailing trips in the city, with seven categories of reported trip purposes: “home,” “work,” “school,” “daycare,” “facilitate passenger,” “shopping,” and “others.” The observed trip purpose market shares from this data were used to validate the performance of the inference model.

Table 1 presents the descriptive statistics of the data sets described above. It also mentions how each of these datasets was used in the overall fusion algorithm.

Estimation Results of Discrete Choice Models
The three candidate discrete choice models described above (in the “Formulation of Discrete Choice Models” sub-section) are estimated using two categories of explanatory variables: trip attributes and land use variables. The trip-related variables include trip start time, day, season, and trip length. Land use information includes the densities of EPOI and of private dwellings within a DA. Table 2 presents the estimation results. For the NL model, different types of nesting structures were tested by grouping “similar” trip purpose alternatives based on a priori expectations. Specifically, alternatives that are expected to share common components in their random error terms were placed under the same nest. However, based on behavioral plausibility and statistical significance of the scale parameters, only one nest of mandatory trips was found to be valid. The mandatory trip nest includes “home,” “work,” “education,” and “daycare” purposes. In the MMNL model, the Cholesky factors corresponding to the variance terms of the alternatives are estimated. (Note that Table 2 report the parameters corresponding to the standard deviations [or variances by extension] of the alternatives and not of the Cholesky decomposition.) The model was tested with 50, 100, 150, and 200 draws and seemed to reach convergence after 150 draws. The model results obtained for D = 200 are presented here. Variables in the final specifications are selected based on the expected sign, and statistical significance (95% confidence interval) of corresponding parameters. Some parameters with lower than 95% confidence are still retained in the model as those are perceived to be important variables. Moreover, the same explanatory variables are kept in all models for ease of performance comparison. Among the three specifications, the MMNL model shows the best goodness-of-fit in lower log-likelihood, Akaike information criterion (AIC), and Bayesian information criterion (BIC) values against the constant-only model.

As expected, the land use variables at destinations are important predictors for most trip purposes (except “facilitate passenger,” “worship,” and “other” which are not associated with a specific land use type).
### Table 2. Empirical Model of Trip Purposes

| Variables | Trip purposes | MNL Estimates | NL Estimates | MMNL Estimates |
|-----------|---------------|---------------|--------------|----------------|
| Alternative specific constants (ASC) | | | | |
| Home | na | na | na |
| Work | 0.571 | 0.757 | 1.141 |
| Education | -2.655 | -1.349 | -2.022 |
| Daycare | -3.631 | -1.889 | -9.251 |
| Facilitate passenger | 0.048 | -0.551 | -3.232 |
| Shopping and errands | -0.077 | -0.485 | 0.350 |
| Eat out | -3.341 | -3.504 | -2.870 |
| Recreation, sports, leisure, arts | 1.151 | 1.466 | -1.405 |
| Health and personal care | -0.963 | -1.458 | -3.327 |
| Services | -2.513 | -2.996 | -4.628 |
| Visiting friends, family | -2.405 | -2.700 | -2.608 |
| Worship, religion | 1.038 | -1.649 | -1.168 |
| Other | 0.526 | -0.032 | -14.693 |
| **Trip characteristics** | | | | |
| Start time: morning | | | | |
| Work | 5.692 | 3.694 | 5.485 |
| Education | 9.244 | 6.007 | 9.009 |
| Daycare | 9.580 | 6.091 | 10.879 |
| Facilitate passenger | 6.671 | 4.492 | 6.414 |
| Shopping and errands | 3.804 | 1.308 | 3.550 |
| Eat out | 9.521 | 5.996 | 9.487 |
| Recreation, sports, leisure, arts | 9.693 | 6.862 | 9.803 |
| Health and personal care | 6.851 | 3.220 | 4.578 |
| Services | 5.010 | 2.779 | 4.697 |
| Visiting friends, family | 7.643 | 4.726 | 7.224 |
| Worship, religion | 8.262 | 6.924 | 8.154 |
| Other | 6.658 | 4.292 | 5.998 |
| Start time: midday | | | | |
| Work | -2.408 | -1.605 | -2.614 |
| Facilitate passenger | 1.103 | 1.101 | 1.143 |
| Shopping and errands | 1.473 | 1.342 | 1.594 |
| Eat out | 4.479 | 3.851 | 4.662 |
| Recreation, sports, leisure, arts | 3.838 | 3.418 | 4.279 |
| Visiting friends, family | 3.537 | 3.067 | 3.694 |
| Other | 1.418 | 1.297 | 4.438 |
| Start time: afternoon | | | | |
| Work | -3.849 | -2.597 | -4.121 |
| Daycare | 1.211 | 0.727 | 2.027 |
| Eat out | 2.683 | 2.273 | 2.781 |
| Recreation, sports, leisure, arts | 2.268 | 2.032 | 2.638 |
| Visiting friends, family | 1.831 | 1.571 | 1.893 |
| Start time: evening | | | | |
| Work | -4.683 | -3.202 | -5.066 |
| Education | -1.633 | -1.148 | -1.755 |
| Shopping and errands | -1.211 | -1.174 | -1.309 |
| Eat out | 1.724 | 1.442 | 1.747 |
| Recreation, sports, leisure, arts | 1.876 | 1.705 | 2.117 |
| Health and personal care | -1.888 | -1.784 | -2.656 |
| Other | 6.744 | 5.871 | 4.671 |
| **Start time: overnight** | | | | |
| Land use characteristics at destination DA | | | |
| Private dwelling density | Home | 0.823 | 0.684 | 1.074 |
| Manufacturing | Work | 0.551 | 0.408 | 0.667 |
| Finance and insurance | 0.366 | 0.314 | 0.491 |
| Professional | 0.239 | 0.185 | 0.289 |
| Public administration | 0.503 | 0.422 | 0.586 |
| Educational | Education | 1.409 | 1.060 | 1.641 |
| Accommodation and food | Shopping | 1.372 | 1.260 | 1.654 |
| Arts, entertainment, and recreation | Eat out | 1.565 | 1.463 | 1.843 |
| Health care | Recreation | 0.668 | 0.571 | 0.870 |
| Finance and insurance | Health | 1.286 | 1.193 | 1.839 |
| Other services | Service | 0.667 | 0.596 | 0.951 |
| Private dwelling density | Other | 0.934 | 0.883 | 1.113 |
| | Visit | 0.418 | 0.380 | 0.720 |

(continued)
For example, the coefficient for retail EPOI density is positive and significant for shopping trips, which indicates that if a trip ends in a DA with many retail establishments, it has a high probability of being a shopping trip. Land use characteristics around the trip origin provide context about the previous activity, and the associated coefficients are statistically significant for some trip purposes. For example, a trip that starts from a DA with many manufacturing or educational establishments has a higher probability of being a return home trip. Similarly, trips to daycare facilities are more likely to start from a residential area.

For trip start times, separate coefficients are estimated for each time period of the day to capture their specific effects on trip purpose. Most of the alternatives have positive parameters for the morning period with respect to return home purpose. This indicates that most of the trips in the morning are destined for some out-of-home activity location. Moreover, it is found that if a trip starts later in the day, it has a lower probability of being a work trip, and a higher probability of being a discretionary trip for facilitating passenger, eating out, recreational or social visits, and so forth.

Day of the week is an important trip purpose determinant for work and religious activities. The dummy variable representing weekdays has a positive and significant coefficient for work trips, indicating that a trip has a higher probability of being a work trip if it is made on a weekday as opposed to the weekend. On the other hand, the same variable has a negative coefficient for religious activity related trips, which generally occur over the weekend. Season of the year is also an important

| Variables                         | Trip purposes   | MNL Estimates | NL Estimates | MMNL Estimates |
|-----------------------------------|-----------------|---------------|--------------|----------------|
| Manufacturing                     | Home            | 0.181**       | 0.163**      | 0.212**        |
| Educational                       | Home            | 0.168**       | 0.118**      | 0.188**        |
| Private dwelling density          | Daycare         | 0.5697        | 0.3893       | 0.8774         |
| Trip day                          | Work            | -1.232*       | -1.370*      | -1.201*        |
|                                   | Worship         | 1.744***      | 1.150***     | 1.823***       |
| Season                            | Education       | 1.919***      | 1.262***     | 1.942***       |
| Fall                              | Shopping        | -0.251**      | -0.275***    | -0.244**       |
|                                   | Recreation      | -0.486**      | -0.519***    | -0.563***      |
|                                   | Visit           | -0.567**      | -0.543**     | -0.502**       |

Inclusive value parameter

Mandatory trips na 0.653*** na

Standard deviation

| Variables                         | MNL Estimates | NL Estimates | MMNL Estimates |
|-----------------------------------|---------------|--------------|----------------|
| Home                              | na            | na           | 0.061          |
| Work                              | na            | na           | 0.464          |
| Education                         | na            | na           | 0.043          |
| Daycare                           | na            | na           | 4.130***       |
| Facilitate passenger              | na            | na           | 3.781***       |
| Shopping and errands              | na            | na           | 0.454          |
| Eat out                           | na            | na           | 0.230          |
| Recreation, sports, leisure, arts | na            | na           | 1.437***       |
| Health and personal care          | na            | na           | 3.106***       |
| Services                          | na            | na           | 2.410***       |
| Visiting friends, family          | na            | na           | na             |
| Worship, religion                 | na            | na           | 1.713          |
| Other                             | na            | na           | 11.003***      |

Log-likelihood at convergence

|                  | MNL           | NL            | MMNL          |
|------------------|---------------|---------------|---------------|
| -9599.39         | -9599.39      | -9599.39      | -9599.39      |
| -7525.07         | -7505.42      | -7457.30      |               |

Rho-squared

|                  | MNL           | NL            | MMNL          |
|------------------|---------------|---------------|---------------|
| 0.2161           | 0.2161        | 0.2231        |               |
| 0.2093           | 0.2113        | 0.2151        |               |

Akaike information criterion (AIC)

|                  | MNL           | NL            | MMNL          |
|------------------|---------------|---------------|---------------|
| 15180            | 15143         | 15015         |               |
| 15291            | 15255         | 15147         |               |

Bayesian information criterion (BIC)

|                  | MNL           | NL            | MMNL          |
|------------------|---------------|---------------|---------------|
| 15180            | 15143         | 15015         |               |
| 15291            | 15255         | 15147         |               |

Note: MNL = multinomial logit model; NL = nested logit model; MMNL = mixed multinomial logit model; DA = “dissemination area”, a geographic unit from the Canadian Census; na = not applicable.

1All the density measures of land use characteristics are log-transformed.

***, **, * Significant at 99%, 95%, 90% levels of confidence.

For example, the coefficient for retail EPOI density is positive and significant for shopping trips, which indicates that if a trip ends in a DA with many retail establishments, it has a high probability of being a shopping trip. Land use characteristics around the trip origin provide context about the previous activity, and the associated coefficients are statistically significant for some trip purposes. For example, a trip that starts from a DA with many manufacturing or educational establishments has a higher probability of being a return home trip. Similarly, trips to daycare facilities are more likely to start from a residential area.

For trip start times, separate coefficients are estimated for each time period of the day to capture their specific effects on trip purpose. Most of the alternatives have positive parameters for the morning period with respect to return home purpose. This indicates that most of the trips in the morning are destined for some out-of-home activity location. Moreover, it is found that if a trip starts later in the day, it has a lower probability of being a work trip, and a higher probability of being a discretionary trip for facilitating passenger, eating out, recreational or social visits, and so forth.

Day of the week is an important trip purpose determinant for work and religious activities. The dummy variable representing weekdays has a positive and significant coefficient for work trips, indicating that a trip has a higher probability of being a work trip if it is made on a weekday as opposed to the weekend. On the other hand, the same variable has a negative coefficient for religious activity related trips, which generally occur over the weekend. Season of the year is also an important...
attribute for some trip purposes. For example, the fall season dummy variable has a positive coefficient for educational trips, as schools remain closed for most of the summer and reopen in fall. In general, all the variables of the models provide satisfactory and intuitive behavioral explanations for the associated trip purposes.

Inferring Ride-Hailing Trip Purposes

First, the ride-hailing trajectory data set (as described above) is fused with land use information from the EPOI and the census databases to provide contexts about potential activities at trip origins and destinations. The choice models estimated in the previous section are then applied to each trip of the augmented data set. This generates the probabilities of occurrence of the 13 purposes for each ride-hailing trip. Finally, the probabilities of all the records for each trip purpose are averaged to obtain the percentage of travel for that purpose category. Table 3 shows the aggregated prediction results.

Validation of Inferred Trip Purposes

To evaluate the effectiveness of the data fusion-based inference algorithm, the reported ride-hailing trip purposes from the 2016 TTS are considered as the ground truth. This limits the validation process to the weekday ride-hailing records, since the TTS collects weekday travel diaries only. Moreover, as mentioned above, the TTS has only seven categories of reported trip purposes, as opposed to the 13 categories inferred by the fusion algorithm. As such, to make the trip purpose categories compatible for the validation exercise, the inferred categories of “shopping and errands,” “eat out,” “recreation, sports, leisure,” “arts, health and personal care,” “services,” “visiting friends, family,” “worship, religion,” and “other” are grouped into a more general class that corresponds to the combined category of “shopping” and “others” in the TTS. The inferred and observed trip purpose distributions are shown in Figure 3.

Overall, the choice models capture the trend of trip purpose distribution well. From Figure 2 it is found that the inferred and the observed trip purpose percentages are reasonably close for “home,” “education,” and “shopping and others.” The prediction errors for these purpose categories are less than 10%, 2%, and 9% respectively. However, the models tend to overpredict the share of “work” trips made by ride-hailing services. The prediction error for these trips is roughly 28%. The inference accuracy is quite low for smaller share alternatives such as “daycare” and “facilitate passenger.” Nonetheless, the validation results are quite encouraging, especially given that the trips in the estimation data are not exclusively ride-hailing trips and as such have somewhat different spatial and temporal characteristics than the ride-hailing trip records of the inference data (see Figure 4). In spite of this, the predicted trip purpose shares match the observed shares quite well and it can be anticipated that with better estimation data, the predictions will be even more accurate. Moreover, the proposed method performed well despite the existence of mixed-use land parcels in the study area (especially the downtown core of Toronto where most of the ride-hailing trips are made) which has always been as a major challenge for trip purpose imputation (5).

The inference accuracies of the three model specifications are comparable with each other. The more complex NL and MMNL do not provide consistently improved inference. This is not surprising given that the nesting and the variance structures of these models are

| Trip purpose categories | MNL (%) | NL (%) | MMNL (%) |
|-------------------------|---------|--------|-----------|
| Home                    | 45.39   | 45.14  | 44.68     |
| Work                    | 16.72   | 16.82  | 16.83     |
| Education               | 4.68    | 4.75   | 4.84      |
| Daycare                 | 0.71    | 0.70   | 0.79      |
| Facilitate passenger    | 2.47    | 2.49   | 2.60      |
| Shopping and errands    | 12.17   | 12.11  | 12.39     |
| Eat out                 | 4.60    | 4.61   | 4.78      |
| Recreation, sports, leisure | 3.28  | 3.28   | 3.28      |
| Arts, health and personal care | 3.14  | 3.17   | 3.23      |
| Services                | 1.53    | 1.48   | 1.53      |
| Visiting friends, family | 0.29    | 0.35   | 0.12      |
| Worship, religion       | 0.47    | 0.47   | 0.44      |
| Other                   | 4.55    | 4.63   | 4.51      |

Note: MNL = multinomial logit model; NL = nested logit model; MMNL = mixed multinomial logit model.
superimposed on the estimation data, which might not hold for the ride-hailing trajectory data.

**Characteristics of Inferred Ride-Hailing Trip Purposes**

The inferred purposes help reveal the nature of ride-hailing trips in the City of Toronto. From Table 3, it is evident that ride-hailing is mostly used for discretionary activities like shopping and errands, eating out, recreation and social visits, and so forth, and for returning home. The increasing popularity of the services as a commuting mode is also noticeable. About a quarter of all weekday ride-hailing trips made within the city are for work- and education-related destinations. Moreover, a portion of the “return home” trips might originate from work or school locations, which means that this category also contributes to commuting travel on top of the “work” and “education” categories. This increasing popularity of ride-hailing as a commuting mode might have a long-term effect on the overall transportation system as the modal share of the services becomes more prominent in the future. More commuting ride-hailing trips would result in increased peak-period congestion on the city’s street networks.

The inferred trip purposes also extend our understanding of ride-hailing demand generation throughout the City of Toronto at a spatial resolution that is unavailable from survey results alone. Figure 5 shows the
spatial distributions of some of the major trip purposes predicted by the MNL model at the DA level. Such disaggregate distributions reveal important patterns of ride-hailing use in the city. For example, it is found that the downtown core attracts a wide variety of ride-hailing trips, especially “shopping and errands,” “eat out,” “recreation, sports, leisure,” “services,” and some “work” trips. However, it attracts comparatively less “education” and “visiting friends, family” trips. A large share of “work” trips are also destined for the outer employment hubs of the city. The most evident pattern is observed for “eat out” trip destinations, which are mostly concentrated around the downtown area. The opposite trend is observed for “visit friends, family” trips, which are mostly destined for the mid- or uptown locations.

A comparison between the inferred weekday and weekend trip purposes (Figure 6) reveals that more “return home” and “shopping and others” trips are made by ride-hailing over the weekends. This indicates that a higher proportion of people use these services over the weekends for trips to entertainment, bars, and other activities, which in turn may help in reducing drinking and driving.

Finally, a comparative analysis between the inferred weekday trip purpose distribution of ride-hailing obtained from the MNL model and the revealed trip purposes of other travel modes from the TTS is shown in Figure 7. As expected, the proportions of trip purposes are somewhat similar for ride-hailing and taxi, indicating the strong modal competition between these services. However, ride-hailing is used more for work-related trips than taxi. This might be attributed to the higher degree of reliability and flexibility associated with ride-hailing services compared with taxi services, making them a preferred mode for work trips. Interestingly, more educational trips are made via ride-hailing than taxis, which indicates that ride-hailing is more popular among the
Figure 5. Spatial distribution of the inferred ride-hailing trip purposes at the dissemination area level: (a) “home” trips, (b) “work” trips, (c) “education” trips, (d) “shopping and errands” trips, (e) “eat out” trips, (f) “recreation, sports, leisure” trips, (g) “arts, health and personal care” trips, (h) “services” trips, (i) “visiting friends, family” trips.
student population than taxi service, perhaps owing to its lower cost.

In general, the proportion of mandatory trips (home, work, education, daycare) among the total ride-hailing trips is found to be comparable to that of other travel modes, especially driving private car (auto driver), passenger in private car (auto passenger), and taxi. While this may be inconsequential as long as the modal share of ride-hailing is small, with its increasing popularity, ride-hailing may start to influence the ridership levels of the more substantial modes. In particular, the greater degree of convenience and reliability (at the expense of relatively cheaper cost compared with traditional taxi services) associated with ride-hailing might encourage people to switch from transit to ride-hailing for their mandatory trips. Thus, proper regulating policies need to be enforced by the municipal authorities that support the benefits of ride-hailing services by providing improved mobility options for people, but not at the expense of increased congestion and reduced transit ridership.

Conclusion and Future Work

The paper presents a data fusion methodology for inferring trip purpose from passively collected ride-hailing trajectory data. The inferred trip purposes extend our understanding about the use of this relatively new mobility option. The proposed algorithm is based on a choice model that makes use of only basic trip-related information (such as approximate pick-up and drop-off locations and trip start times), and land use characteristics around the origins and destinations to predict the probability of different trip purposes. The choice model is estimated using data from a small-sample travel survey, an EPOI database, and a census database. The estimated model is then applied to the ride-hailing trip records to obtain the most probable distribution of ride-hailing trip purposes. Nonetheless, this research has the potential to be applied for broader use beyond ride-hailing travel. The data fusion method presented here is generic enough to be applicable for trip purpose inference from other types of passive data sources, including transit smart card transactions, GPS-based travel surveys, mobile phone call detail records, taxi trajectories, and so forth. It thus demonstrates the benefit of fusing big data sources (such as trip trajectories) with small data sources (such as traditional travel surveys) to extend our understanding of urban travel demand and activity dynamics.

The proposed algorithm is empirically tested using ride-hailing trip data from the City of Toronto. Three different formulations of the choice model are estimated: an MNL structure, an NL structure with one nest of mandatory trips, and a heteroskedastic MMNL structure. The estimation results indicate that basic land use data and time of day provide good contexts about typical activities performed at trip origins and destinations. The use of a probabilistic choice model captures the reality that passengers arriving at the same location can be driven by different purposes, and the purposes may also differ based on trip start times and origins. As such, they are behaviorally more plausible and generic than deterministic rule-based approaches (e.g., 8–11). The prediction performances of the models are evaluated by comparing the inferred distribution with the revealed trip purpose proportions obtained from TTS. It is found that the models can predict trip purposes reasonably well when the training data is relatively small, and limited context-specific attributes are available. Given that the City of Toronto has mixed-use land parcels, and the estimation data has somewhat different spatial and temporal characteristics compared with the ride-hailing data, the performance of the choice models is considered satisfactory.

The inferred purpose distribution of ride-hailing trips indicate that the services are mostly used for discretionary activities (e.g., shopping and errands, eating out, recreation, social visits, etc.) and for returning home.
The services also play an important role in daily commuter travel. About a quarter of all weekday ride-hailing trips are made to work- and education-related destinations. This, along with the portion of “return home” trips that originate from work or school locations, indicates increasing popularity of the services as commuting mode. During the weekends, these services are used more for going to and returning from entertainment- and recreation-related activities, which in turn may help in reducing drinking and driving. With regard to modal competition, it is evident that taxi and ride-hailing are strong competitors that cater to similar purposes. Moreover, it is anticipated that with growing ridership, ride-hailing will start to influence other travel modes, including transit. As such, efficient policies should be mandated to reduce the adverse impacts of ride-hailing on the level of traffic congestion, the environment, and the equity of mobility services.

The proposed data fusion method includes a basic assumption that the missing ride-hailing trip purposes have the same conditional probability as the trips of all modes in the estimation data. Although the validation proves that even with this hypothesis the inference algorithm performs quite satisfactorily, this is a very strong assumption and future works might be geared to relax it by using a representative and large enough sample exclusively for ride-hailing trips. An alternative approach to relaxing the assumption might be recalibrating the alternative specific constants of the estimated choice model to reflect data on observed ride-hailing trip purpose market share, given that the information is available to the analyst from an independent ride-hailing specific survey. Such adjustment of constants would help to minimize the systematic bias associated with estimating the model using a sample that is not exclusively of ride-hailing trips. Also, more advanced context data repositories such as social network check-in data, Google Places API, and hours of operation of points of interest can be used to improve the prediction accuracy. In future work, the authors plan to compare the trip purpose predictive performance of the choice models with that of appropriate machine learning tools. Specifically, they want to investigate the suitability of unsupervised learning techniques so that the scarcity of representative training data for ride-hailing trips can be overcome.

Acknowledgments
Data were made available for the research by the Big Data Innovation Team of the City of Toronto and the Data Management Group of the University of Toronto.

Author Contributions
The authors confirm contribution to the paper as follows: study conception and design: K. N. Habib, S. Hossain; data collection: S. Hossain; analysis and interpretation of results: S. Hossain, K. N. Habib; draft manuscript preparation: S. Hossain, K. N. Habib. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was funded by a Trillium Scholarship and an NSERC Discovery Grant.

References
1. Big Data Innovation Team - City of Toronto. The Transportation Impacts of Vehicle-for-Hire in the City of Toronto. University of Toronto Transportation Research Institute, Toronto, Canada, 2019.
2. Miller, E. J., S. Srikkenthran, and B. Chung. Workshop Synthesis: Household Travel Surveys in an Era of Evolving Data Needs for Passenger Travel Demand. Transportation Research Procedia, Vol. 32, 2018, pp. 374–382.
3. Young, M., and S. Farber. The Who, Why, and When of Uber and other Ride-Hailing Trips: An Examination of a Large Sample Household Travel Survey. Transportation Research Part A: Policy and Practice, Vol. 119, 2019, pp. 383–392.
4. Dias, F. F., P. S. Lavieri, T. Kim, C. R. Bhat, and R. M. Pendyala. Fusing Multiple Sources of Data to Understand Ride-Hailing Use. Transportation Research Record: Journal of the Transportation Research Board, 2019, 2673: 214–224.
5. Gong, L., T. Morikawa, T. Yamamoto, and H. Sato. Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies. Procedia-Social and Behavioral Sciences, Vol. 138, 2014, pp. 557–565.
6. Xiao, G., Z. Juan, and C. Zhang. Detecting Trip Purposes from Smartphone-Based Travel Surveys with Artificial Neural Networks and Particle Swarm Optimization. Transportation Research Part C: Emerging Technologies, Vol. 71, 2016, pp. 447–463.
7. Ernagun, A., Y. Fan, J. Wolfson, G. Adomavicius, and K. Das. Real-Time Trip Purpose Prediction Using Online Location-Based Search and Discovery Services. Transportation Research Part C: Emerging Technologies, Vol. 77, 2017, pp. 96–112.
8. Wolf, J., R. Guensler, and W. Bachman. Elimination of the Travel Diary: Experiment to Derive Trip Purpose from Global Positioning System Travel Data. Transportation Research Record: Journal of the Transportation Research Board, 2001, 1768: 125–134.
9. Bohte, W., and K. Maat. Deriving and Validating Trip Purposes and Travel Modes for Multi-Day GPS-Based Travel Surveys: A Large-Scale Application in the
10. Chen, C., H. Gong, C. Lawson, and E. Bialostozky. Evaluating the Feasibility of a Passive Travel Survey Collection in a Complex Urban Environment: Lessons Learned from the New York City Case Study. *Transportation Research Part A: Policy and Practice*, Vol. 44, No. 10, 2010, pp. 830–840.

11. Shen, L., and P. R. Stopher. A Process for Trip Purpose Imputation from Global Positioning System Data. *Transportation Research Part C: Emerging Technologies*, Vol. 36, 2013, pp. 261–267.

12. Kusakabe, T., and Y. Asakura. Behavioural Data Mining of Transit Smart Card Data: A Data Fusion Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 46, 2014, pp. 179–191.

13. Gong, L., X. Liu, L. Wu, and Y. Liu. Inferring Trip Purposes and Uncovering Travel Patterns from Taxi Trajectory Data. *Cartography and Geographic Information Science*, Vol. 43, No. 2, 2016, pp. 103–114.

14. Chen, C., S. Jiao, S. Zhang, W. Liu, L. Feng, and Y. Wang. TripImputor: Real-Time Imputing Taxi Trip Purpose Leveraging Multi-Sourced Urban Data. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 19, No. 10, 2018, pp. 3292–3304.

15. Oliveira, M. G. S., P. Vovsha, J. Wolf, and M. Mitchell. Evaluation of Two Methods for Identifying Trip Purpose in GPS-Based Household Travel Surveys. *Transportation Research Record: Journal of the Transportation Research Board*, 2014, 2405: 33–41.

16. Zhu, Y. Estimating the Activity Types of Transit Travelers Using Smart Card Transaction Data: A Case Study of Singapore. *Transportation*, Vol. 47, 2018, pp. 1–28.

17. Chen, C., C. Liao, X. Xie, Y. Wang, and J. Zhao. Trip2Vec: A Deep Embedding Approach for Clustering and Profiling Taxi Trip Purposes. *Personal and Ubiquitous Computing*, Vol. 23, No. 1, 2019, pp. 53–66.

18. Lu, Y., and L. Zhang. Imputing Trip Purposes for Long-Distance Travel. *Transportation*, Vol. 42, No. 4, 2015, pp. 581–595.

19. Wu, J., C. Jiang, D. Houston, D. Baker, and R. Delfino. Automated Time Activity Classification Based on Global Positioning System (GPS) Tracking Data. *Environmental Health*, Vol. 10, No. 1, 2011, p. 101.

20. Montini, L., N. Rieser-Schüssler, A. Horni, and K. W. Axhausen. Trip Purpose Identification from GPS Tracks. *Transportation Research Record: Journal of the Transportation Research Board*, 2014, 2405: 16–23.

21. Yazdizadeh, A., Z. Patterson, and B. Farooq. An Automated Approach from GPS Traces to Complete Trip Information. *International Journal of Transportation Science and Technology*, Vol. 8, No. 1, 2019, pp. 82–100.

22. Martin, H., D. Bucher, E. Suel, P. Zhao, F. Perez-Cruz, and M. Raubal. Graph Convolutional Neural Networks for Human Activity Purpose Imputation. *Proc., NIPS Spatiotemporal Workshop at the 32nd Annual Conference on Neural Information Processing Systems (NIPS 2018)*, Montreal, Canada, 2018.

23. Han, G., and K. Sohn. Activity Imputation for Trip-Chains Elicited from Smart-Card Data Using a Continuous Hidden Markov Model. *Transportation Research Part B: Methodological*, Vol. 83, 2016, pp. 121–135.

24. Alsger, A., A. Tavassoli, M. Mesbah, L. Ferreira, and M. Hickman. Public Transport Trip Purpose Inference Using Smart Card Fare Data. *Transportation Research Part C: Emerging Technologies*, Vol. 87, 2018, pp. 123–137.

25. Sari Aslam, N., T. Cheng, J. Cheshire, and Y. Zhang. Trip Purpose Identification Using Pairwise Constraints Based Semi-Supervised Clustering. *Proc., 27th Conference on GIS Research UK (GISRUK)*, Geographical Information Science Research UK (GISRUK), Newcastle, UK, 2019.

26. Li, A., and K.W. Axhausen. Trip Purpose Imputation for Taxi Data. *Proc., 18th Swiss Transport Research Conference, Monte Verità Ascona*, 2018.

27. McFadden, D. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics* (P. Zarembka, ed.), Academic Press, New York, NY, 1973, pp. 105–142.

28. Wen, C. H., and F. S. Koppelman. The Generalized Nested Logit Model. *Transportation Research Part B: Methodological*, Vol. 35, No. 7, 2001, pp. 627–641.

29. Train, K. E. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, NY, 2009.

30. Daly, A., and S. Zachary. Improved Multiple Choice Models. In *Determinants of Travel Choice* (D. Hensher, and M. Dalvi, eds.), Saxon House, Sussex, UK, 1978.

31. McFadden, D. Modeling the Choice of Residential Location. *Transportation Research Record: Journal of the Transportation Research Board*, 1978, 673: 72–77.

32. Williams, H. C. On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning A*, Vol. 9, No. 3, 1977, pp. 285–344.

33. Walker, J. L. *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*. Doctoral dissertation. Massachusetts Institute of Technology, Cambridge, MA, 2001.

34. Hensher, D. A., and W. H. Greene. The Mixed Logit Model: The State of Practice. *Transportation*, Vol. 30, No. 2, 2003, pp. 133–176.

35. Aptech Inc. GAUSS Programming Language. 2019. https://www.aptech.com/. Accessed July 2019.

36. DMTI Spatial Inc. CanMap Content Suite data dictionary. 2016. http://canue.ca/wp-content/uploads/2018/03/Data_Dictionary_CanMap_Content_Suite_v2016_3.pdf. Accessed June 2019.

37. Statistics Canada. Census Profile, 2016 Census. 2017. https://www12.statcan.gc.ca/census-recensement/index-eng.cfm. Accessed June 2019.

38. Data Management Group (DMG). 2016 Transportation Tomorrow Survey. 2019. http://dmg.utoronto.ca/pdf/tts/2016/2016TTS_Conduct.pdf. Accessed July 2019.

The opinions expressed in this paper are those of the authors and not those of NSERC. The authors claim the sole responsibility for all results, comments, and interpretations made in the paper.