Transportation
https://doi.org/10.1007/s11116-020-10120-0

Probabilistic model for destination inference and travel pattern mining from smart card data

Zhanhong Cheng[1,2] · **Martin Trépanier**[2,3] · **Lijun Sun**[1,2]

Abstract
Inferring trip destination in smart card data with only tap-in control is an important application. Most existing methods estimate trip destinations based on the continuity of trip chains, while the destinations of isolated/unlinked trips cannot be properly handled. We address this problem with a probabilistic topic model. A three-dimensional latent dirichlet allocation model is developed to extract latent topics of departure time, origin, and destination among the population; each passenger's travel behavior is characterized by a latent topic distribution defined on a three-dimensional simplex. Given the origin station and departure time, the most likely destination can be obtained by statistical inference. Furthermore, we propose to represent stations by their rank of visiting frequency, which transforms divergent spatial patterns into similar behavioral regularities. The proposed destination estimation framework is tested on Guangzhou Metro smart card data, in which the ground-truth is available. Compared with benchmark models, the topic model not only shows increased accuracy but also captures essential latent patterns in passengers' travel behavior. The proposed topic model can be used to infer the destination of unlinked trips, analyze travel patterns, and passenger clustering.

Keywords Public transit · Smart card data · Destination inference · Topic model · Passenger clustering

Introduction

Origin and destination (OD) matrix is an essential input for transit planning and operation. Most transit agencies have been relying on travel surveys to collect representative OD information. However, conducting such a survey with a reasonable scale is not only costly but also time-consuming. With the recent advances of intelligent transportation systems,

✉ Lijun Sun
lijun.sun@mcgill.ca

[1] Department of Civil Engineering, McGill University, Montreal, QC H3A 0C3, Canada

[2] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Montreal, Canada

[3] Department of Mathematics and Industrial Engineering Polytechnique Montréal, Montreal, QC H3T 1J4, Canada

Published online: 10 June 2020

researchers and practitioners have started taking advantage of the transit operation data and smart card data for better planning and operation practices (Pelletier et al. 2011).

Smart card systems are initially designed for the purpose of automatic fare collection (AFC). When the system has both tap-in and tap-out controls (e.g., using a distance-based transit fare scheme), the full itinerary (boarding time/station and alighting time/station) of each trip can be registered. However, most smart card systems across the world adopt a single fare scheme with only tap-in validation, and the alighting information (time/station) is essentially unknown. Inferring the alighting stations is a crucial problem in obtaining the OD matrix from these smart card systems.

Trip destination estimation in smart card data has always been a hot issue. Barry et al. (2002) proposed two assumptions to address this issue: (1) the alighting station of a trip is very likely to be the boarding station of the immediate next trip; (2) the last alighting station of a day is usually the first boarding station of the same day. This type of "rule-based" model soon became the workhorse algorithm for smart card destination estimation. Depending on the data, current algorithms can obtain around 60% to 85% trips' destinations; these trips are often called linked trips in the literature, and the rest un-inferred trips are referred to as unlinked trips. Without the information from consecutive trips, the destination estimation of unlinked trips is more challenging. Existing methods address this problem by seeking similar trips in the passenger's historical trips; we refer them as individual-history-based models. Such as He and Trépanier (2015) used the spatial and temporal kernel density probability of passengers' trips and get an additional 10% estimation for unlinked trips.

The prediction of unlinked trips is challenging without the help of the trip-chain continuity information. The solution lies in the regularity of human mobility. As explained by González et al. (2008), Song et al. (2010), human movement follows certain regularity and is highly predictable. However, there still lacks an appropriate framework to infer the missing destination using the mobility regularity. To address this issue, this paper attempts to build an integrated model that estimates the missing destinations drawing on the common mobility patterns among the population. We establish a probabilistic topic model for smart card data by making an analogy with the latent dirichlet allocation (LDA) model (Blei et al. 2003). We assume transit trips among the population can be summarized in a few latent topics over departure time, origin, and destination. Every passenger is characterized by a latent topic distribution and the whole population share the topic-word distributions for departure time, origin and destination. To share more information among different passengers, we represent each station by each passenger's rank of visiting frequency, as against to directly using the station ID. A case study is performed on Guangzhou Metro data, where the tap-out data is used as the ground truth to test different models. Results show our topic model has improved accuracy compared with individual-history-based models. We further demonstrate passengers' latent topic distribution is a useful feature for passenger clustering, commuter identification, and travel pattern mining.

The remainder of the paper is organized as follows. "Literature review" section briefly reviews the current research on smart card data destination inference and transit pattern mining. "Methodology: topic model for destination inference" section elaborates the topic model for transit trips, the Gibbs sampling for model inference, and the destination inference in ranked stations. The case study on Guangzhou Metro will be shown in "Case study" section, where the destination inference will be compared with individual-history-based models; the model interpretation and the passenger clustering will be demonstrated in "Case study". Finally conclusions and discussions are summarized in "Conclusions and discussion" section.

Literature review

Destination inference in smart card data

Destination inference is an important problem in smart card data. Existing methods primary take advantage of the continuity of trip chains, and infer the destinations based on assumptions or rules. In a very first study, Barry et al. (2002) proposed that the destination of a trip can be inferred by the origin of the immediate next trip, and they assumed the last destination of a day is often the first origin in the same day. Since then, many refined models have been proposed based on similar assumptions. Trépanier et al. (2007) imposed a distance constraint between consecutive trips, and they further assumed the last destination of a day can also be inferred by the first origin in the next day. Munizaga and Palma (2012) proposed to use generalized time instead of distance in destination inference. Further, Sánchez-Martínez (2017) constructed a generalized disutility minimization objective to determine the paths and transfers between the origin and destination. Research based on similar rule-based methodology has become the mainstream, and more research can be found in Zhao et al. (2007), Wang et al. (2011), Gordon et al. (2013), Alsger et al. (2016), Nunes et al. (2016). Depending on the data, the rule-based method can accomplish around 60% to 85% of the destinations; trips of which the destinations can be inferred by the rule-based model are often called linked trips.

For the O-D of unlinked trips, whose destination cannot be inferred by rule-based models, one treatment is to scale the O-D of linked trips by some methods (see e.g., Munizaga and Palma 2012; Gordon et al. 2018). This approach assumes the destination distribution of unlinked trips at each origin is the same as the linked trips, which is unverified. On the other hand, the destinations of unlinked trips can be estimated by historical similar trips (individual-history-based model), similar to supervised learning with labeled data. Such as Trépanier et al. (2007) defined a similar trip as a trip on the same route with similar departure time in the previous several days. He and Trépanier (2015) used spatial and temporal kernel density probability estimated by historical trips to infer the destination of unlinked trips. Zhang et al. (2015) conducted an interesting study, where a collaborative space alignment framework was presented to reconstruct smart card trips. Recent studies attempted to use (deep) neural networks to infer trip destinations (Jung and Sohn 2017; Assemi et al. 2020). These studies were based on smart card systems with full information and extensive features (e.g. time and location, land-use features of stations). Experiments showed promising results, while the large number of labeled destinations are essentially unavailable for a real tap-in-only system.

To summarize, existing research has developed various algorithms based on the trip continuity feature to estimate the destination of linked trips. The destination estimation of unlinked trip relies on historical similar trips. This paper provides a whole new approach to infer the destination of unlinked trips by a topic model. The proposed model is not only a prediction model but also a generative model that captures individuals' behavioral patterns.

Transit pattern mining

There has been a large body of literature on passengers' travel behavior patterns. The travel patterns are usually characterized by certain features. A series of analyses (such as commuter identification, passenger clustering, and pattern evolving) can then be conducted

using these features. Next, we briefly review related literature based on how these features are obtained.

In many studies, the features for travel patterns are designed based on domain knowledge. For example, Morency et al. (2007) defined two indicators to measure passengers' spatial and temporal variability. Then a k-mean algorithm was conducted to cluster passengers. In another research, Ma et al. (2013) designed four features based on how often did a passenger repeatedly visits the same or adjacent places on a multi-day basis; these features can be used to identify regular passengers. A similar approach is also applied in Ma et al. 2017. Mohamed (2014) established a temporal profile by passengers' travel time on a weekly basis to analyze the travel patterns. He et al. (2018) directly used time series on transit smart card activities' data as features, and used the distance between time series for passenger clustering.

On the other hand, the travel patterns can also be represented by latent features that are learned from data; the topic model developed in this paper also falls in this catalogue. For example, Goulet-Langlois et al. (2016) used principal component analysis (PCA) to extract eigen-patterns from passengers' multi-week activity sequences. Briand et al. (2016) applied a mixed Gaussian model to extract latent features to mining passengers' temporal travel patterns. Based on the same method, Briand et al. (2017) further analyzed the year-to-year pattern changes in a public transportation system. Zhao et al. (2020) applied a topic model to discover latent activity patterns from smart card data, which is very relevant to our research. Zhao et al. (2018) and this paper both extend the LDA for travel behavior mining. The main difference is that we organize the latent features in a three-dimensional manner, which captures the interaction of spatial and temporal topics.

Besides the public transportation domain, topic models have been widely applied for mobility mining. For example, Hasan and Ukkusuri (2014) classified individuals' activity patterns by applying LDA to geo-location data collected from Twitter. Sun and Axhausen (2016) applied a probabilistic tensor factorization to smart card transactions to understanding urban mobility patterns. Fan et al. (2016) applied LDA to mobile phone call data, and further developed a Hidden Markov Model for complete missing mobility data. Sun et al. (2019) developed a two-dimensional LDA on license plate recognition data, where the spatial and temporal topics are modeled separately, and their interactions are characterized in a two-dimensional simplex. We applied the same methodology as Sun et al. (2019) and extend it to smart card data with three-dimensional features (origin, destination, and time).

Methodology: topic model for destination inference

This section details the probabilistic topic model for trip destination inference in smart card data. The objective is to infer the unknown trip destination in a tap-in-only system. A large portion of the destinations of linked trips could be inferred by rule-based models (Barry et al. 2002; Trépanier et al. 2007; Munizaga and Palma 2012); some trip surveys could also provide a sample of complete trip information (Trajet 2019). We can train the proposed topic model by those trips with complete (ground truth or inferred) itineraries. Next, the destinations of unlinked trips could be inferred by the trained topic model.

Model formulation

A smart card trip could be characterized by a three-element tuple (w^t, w^o, w^d) representing the departure time, origin, and destination; where w^t is assumed to be a discrete variable in 1-hour intervals. Then, all the historical trips of a passenger u can be represented as $\mathbf{w}_u = \{(w_i^t, w_i^o, w_i^d) : i = 1, \ldots, N_u; w_i^t \in \{1, \cdots, T\}; w_i^o, w_i^d \in \{1, \cdots, S\}\}$; where N_u is the total number of trips for passenger u, T is the number of possible departure hours, and S is the number of boarding/alighting locations.

The LDA model in NLP assumes there are several topics (e.g. sport, and cooking) among the corpus, the probability for each word's occurrence varies from topic to topic (e.g. the probability for the word "basketball" occurs in the sport topic is higher than which in the cooking topic). A document is characterized by a mixture of topics, which explains the probability of each word's occurrence in the document. By making an analogy to the LDA model, we treat each trip (w^t, w^o, w^d) as a word and $\mathbf{w}_u$ as a document (a bag of words). Thus, all the trips belonging to a passenger compose a document with each trip being regarded as a word, each passengers trips are characterized by a mixture of latent topics.

The traditional LDA cannot directly model metro trips, because of the three interdependent attributes of a trip (i.e. time, origin and destination). A common solution is to combine different attributes into one dimension with the vocabulary size of $T \times S \times S$, such as in Hasan and Ukkusuri (2014), Fan et al. (2016). The main drawback of this approach is that it considerably increases the vocabulary size, while the new combined words are sparse with many unobserved/unlikely trips. Moreover, the interdependency between original attributes is lost (e.g., two trips with the same origin and destination but different times can become unrelated words). To address this problem, we use an innovative method that expands the latent topics into a three-dimensional tensor, similar to the probabilistic tensor factorization (Sun and Axhausen 2016; Sun et al. 2019). By increasing the dimension of latent topics, we have three types of topic-word distributions, which avoids the large vocabulary set and captures inter-dependencies of different types of words in the latent space.

The latent topic is organized as a three-dimensional tensor $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{J \times K \times L}$, where J, K, and L are the number of latent topics of time, origin, and destination respectively. The element $z_{j,k,l}$ of tensor $\boldsymbol{\mathcal{Z}}$ corresponds to the j^{th} temporal topic z_j^t, the k^{th} origin topic z_k^o, and the l^{th} destination topic z_l^d. Each passenger's trips are characterized by a Multinomial distribution over latent topics $\boldsymbol{\mathcal{Z}}$ (the topic distribution), parameterized by θ_u. Given a latent topic $z_{j,k,l}$, the topic-word distributions for departure time, origin, and destination are Multinomial distributions parameterized by φ_{z^t}, ψ_{z^o}, and ω_{z^d} respectively. The overall picture of the model can be clearly depicted by a graphical model shown in Fig. 1; where α, β, γ, and η are parameters for Dirichlet priors; U is the number of passengers. We describe the generative process in Fig. 1 follows:

- Draw topic distribution for each passenger $\theta_u \sim \text{Dirichlet}_{J \times K \times L}(\boldsymbol{\alpha})$.
- Draw topic-time distribution for each time topic $\varphi_j \sim \text{Dirichlet}_J(\boldsymbol{\beta})$.
- Draw origin distribution for each origin topic $\psi_k \sim \text{Dirichlet}_K(\boldsymbol{\gamma})$.
- Draw destination distribution for each destination topic $\omega_l \sim \text{Dirichlet}_L(\boldsymbol{\eta})$.
- For each passenger u, for each trip record:
  - Draw latent topic $z \sim \text{Multinomial}(\theta_u)$.
  - Obtain z^o, z^d, and z^t by z.
  - Draw $w^t \sim \text{Multinomial}(\varphi_{z^t})$.

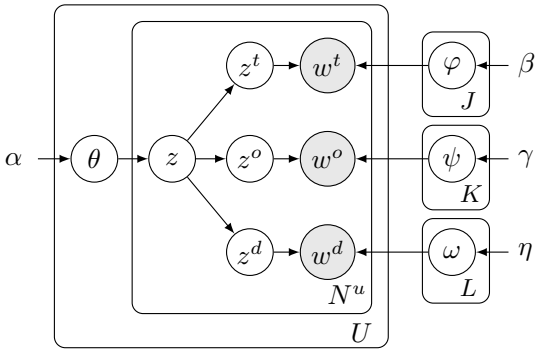

Fig. 1 Plate notation for the graphical model

- Draw $w^o \sim \text{Multinomial}(\psi_{z^o})$.
- Draw $w^d \sim \text{Multinomial}(\omega_{z^d})$.

We apply Multinomial distribution to departure time by discretizing time into 1-h intervals. This is a reasonable simplification and has been widely used in literature (Hasan and Ukkusuri 2014; Sun and Axhausen 2016; Sun et al. 2019). Continuous distributions, such as Normal and Log-Normal distributions (Zhao et al. (2018), are more refined in time representation, but they are also more computational costly and to some extent restrictive in the shape of the distribution. Considering 1-hour resolution is normally enough to distinguish different travel/activity patterns, this paper uses the discrete representation of time.

Model inference

The model inference involves estimating the parameters for latent topic distribution of each passenger and the topic-word distribution of each topic. In the generative process, each trip is generated from a latent topic z, which is unobserved. We use a collapsed Gibbs sampling algorithm Griffiths and Steyvers (2004) to iteratively sample the topic for each trip by the conditional probability shown in Eq. (1):

$$P(z_i^t = j, z_i^o = k, z_i^d = l | w_i^t = t, w_i^o = o, w_i^d = d, \mathbf{z}_{-i}^t, \mathbf{z}_{-i}^o, \mathbf{z}_{-i}^d, \mathbf{w}_{-i}^t, \mathbf{w}_{-i}^o, \mathbf{w}_{-i}^d) \propto \frac{N_{z^t=j}^{w^t=t} + \beta}{N_{z^t=j} + T\beta} \times \frac{N_{z^o=k}^{w^o=o} + \gamma}{N_{z^t=k} + S\gamma} \times \frac{N_{z^d=l}^{w^d=d} + \eta}{N_{z^t=l} + S\eta} \times \frac{N_{z^t=j,z^o=k,z^d=l}^{u} + \alpha}{N^u + JKL\alpha}. \quad (1)$$

where $\mathbf{w}_{-i}^{(\cdot)}$ and $\mathbf{z}_{-i}^{(\cdot)}$ are trip attributes and latent topics for all other trips except trip i; $N_{(\cdot)}^{(\cdot)}$ denotes the number of trips that satisfy the condition listed in the subscript and the superscript, note that the current trip i is excluded when counting N.

The sampling procedure will converge after sufficient iterations, by then we can estimate the parameters in topic distributions and topic-word distributions by Eq. 2:

$$
\begin{aligned}
\varphi_{t,j} &= \frac{N^{w^t=t}_{z^t=j} + \beta}{N_{z^t=j} + T\beta}, \\
\psi_{o,k} &= \frac{N^{w^o=o}_{z^o=k} + \gamma}{N_{z^t=k} + S\gamma}, \\
\omega_{d,l} &= \frac{N^{w^d=d}_{z^d=l} + \eta}{N_{z^t=l} + S\eta}, \\
\theta_{u,j,k,l} &= \frac{N^{u}_{z^t=j,z^o=k,z^d=l} + \alpha}{N^u + JKL\alpha}.
\end{aligned} \tag{2}
$$

Destination inference and station-to-rank transformation

Having estimated all the parameters in the model, we can infer the missing destination for a trip with only the origin and the departure time observed. According to the Bayes' theorem, the probability for passenger u alighting at a location d given the departure time t and the boarding location o takes the form:

$$
\begin{aligned}
&P(w^d = d|w^t = t, w^o = o; u) \propto P(w^t = t, w^o = o, w^d = d; u) \\
&= \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} P(w^t = t|z^t_j)P(w^o = o|z^o_k)P(w^d = d|z^d_l)P(z^t_j, z^o_k, z^d_l; u).
\end{aligned} \tag{3}
$$

Next, the most likely destination of a trip is the one that takes the highest probability in Eq. 3.

By now we have shown the complete topic model for destination inference, but there is a final impediment that prevents the model from giving a good destination estimation—the giant heterogeneity among passengers' spatial patterns. In essence, the topics of an LDA model are learned from the word co-occurrences across different documents. However, the origin-destination set is generally diverse from person to person (few word co-occurrences), which means a very large number of spatial topics are required to capture the spatial heterogeneity of the entire population. The large latent space not only fails to extract representative patterns among individuals but also increases the number of unknown parameters.

To address this problem, we do not use unique IDs for stations, instead, we label locations by the frequency-rank in each passengers' historical trips. Studies have shown the frequency of individuals' historical locations follows Zipf's law González et al. (2008), indicating most of the trips of a passenger are between several frequently visited locations. Therefore, the first several ranks can well characterize a person's travel behavior. Specifically, denote r^i_u to be the rank (by the order of visiting frequency) of station s_i in all the historical origins of passenger u. We transform each passenger's visited locations into the rank representation and store a mapping function $M_u(r^i_u) \rightarrow s_i$ to restore real stations. By doing this, the diverse spatial patterns are essentially transformed into similar behavioral regularities (e.g., travel from the most visited station to the second most visited station). The same-ranked location for different passengers' does not correspond to the same real stations, but represents a similar degree of importance of these stations to these passengers.

We build the topic model and infer the destination in the ranked reference; the estimation for real destination is then retrieved by the mapping function M_u.

Case study

We use the Guangzhou Metro smart card data—a tap-in and tap-out system—to examine the proposed topic model. As the topic model requires a portion of complete itineraries (training set) to learn passengers' travel patterns, we will investigate our model under two scenarios:

1. using a ground truth training set;
2. using an estimated training set.

In scenario 1, we randomly select 70% trips and preserve the real destinations, as a training set; the destinations inference is tested on the rest 30% data. Scenario 2 is a more realistic case where the ground truth destinations are unknown. We train the model with the destinations inferred by rule-based models (Trépanier et al. 2007) and demonstrate our model's performance under the "noisy" training set. The case study part is organized as follows, we will first introduce the data set, hyper-parameters, and baseline models, next test the destination inference accuracy, interpret the latent patterns, and finally present an application of passenger clustering.

Guangzhou Metro data

Guangzhou Metro is one of the busiest metro systems in the world. As of August 2019, Guangzhou Metro has 14 operating lines with a total length of 478 km. It is the third-largest metro system in China, after Beijing and Shanghai. The average daily ridership exceeds 8.6 million, taking over 50% of the ridership in the public transportation (Guangzhou 2019). Except line 9, 14, APM, and THZ1, our data covers other 11 lines of Guangzhou Metro with 159 stations from July 1st to September 30th, 2017. The metro operates 19 h per day from 5:00 to 24:00. Therefore, the vocabulary size for time is 19, for origin and destination is 159.

Guangzhou Metro is a tap-in and tap-out system with both origin and destination registered, we can compare the estimated destination with the real destination to test the inference accuracy. There are single pass, day pass, Yang Cheng Tong and Linnan pass (including various subclasses for students, elderly and disabled people), and digital tickets on smartphone apps. Around 1/3 trips are accomplished by single or day pass, the destinations of these short-term users are barely estimable because of the lack of information. We only focus on the passengers with a minimum of 20 observations in the 3 months; later we will discuss the effect of the number of observations to the estimation accuracy. We showcase our model in 10,000 randomly selected passengers. The total number of selected trips is 667,033, which means on average each person took 67 trips in the 3 months.

As discussed in "Destination inference and station-to-rank transformation" section, instead of station IDs, we train our model by each passenger's rank of stations. A preliminary analysis of the data shows that the first several ranks can capture most of the trips. Figure 2a shows the rank r of stations and the visiting probabilities $p(r)$ in a log-log plot. It can be found that the visiting probability drops significantly as the rank gets

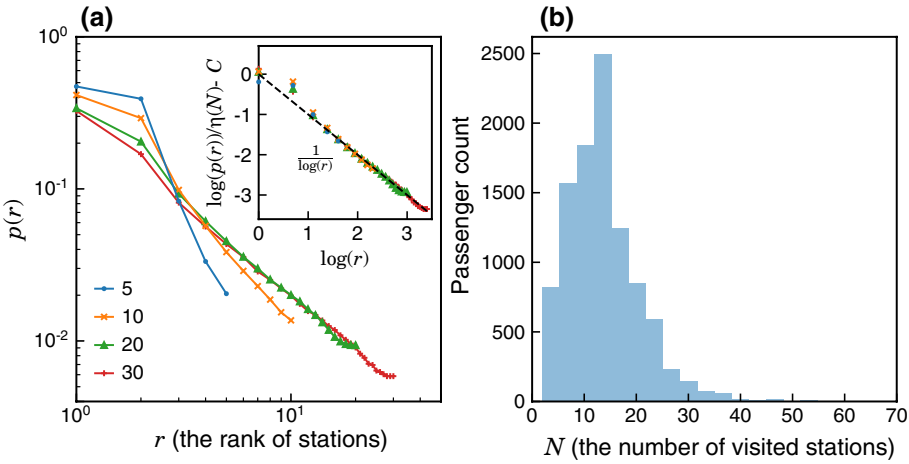

Fig. 2 The probability of visiting different stations. **a** For passengers that have been observed to visit 5 10, 20, and 30 different stations, the rank of the stations (in the order of the visit frequency) and the visiting probabilities in a log–log scale. The insert figure shows that the four groups of $p(r)$ can be well approximated by $p(r) \sim r^{-\eta}$, when apply $\eta = 3.57N^{-0.38}$. **b** The histogram for the number of different stations visited by each passenger, in the 10,000 passengers

large. Further, the relation can be approximated by the Zipf's law $p(r) \sim r^{-\eta}$ with the exponent term η relates to the number of visited stations N (Hasan et al. 2013). When applying $\eta = 3.57N^{-0.38}$, $p(r)$ and r can be approximated by a single distribution, shown in the inserted figure. Similar to Hasan et al. (2013), the rank 2 station deviates from this relation, showing higher visiting probability. This indicates that there is a bi-central mobility pattern in metro usage compared to the common Zipf's law (González et al. 2008).

Figure 2b shows the histogram for the number of different stations visited by each passenger. We can find most passengers visited between 5 and 20 different metro stations in the 3-month period; the number of people who visited more than 20 stations tails off. Therefore, we cut off the frequency-rank at 20, marking all stations ranked larger than 20 as 20. By doing this, each passenger's spatial vocabulary size is aligned at 20. Because the possibility of choosing cut stations is very low, as long as the cut-off point is not too small, the choice of cut-off point has little effect to the performance of our model. Representing stations by rank significantly decreases the number of latent topics needed on the spatial dimension.

Hyper-parameters

There are two types of hyper-parameters in our models—the number of latent topics and Dirichlet priors. In literature, the number of latent topics is often determined by perplexity, which measures the average likelihood of test data set (Blei et al. 2003; Hasan and Ukkusuri 2014). In our context, we use the destination inference accuracy in the test set to select the number of topics. We perform a grid search over $J = [3, 4, 5]$ and $K, L = [2, 3, 4, 5]$ and select the best configuration by the minimal destination inference error, and we prefer a smaller model when the errors are close. Based on the result, we choose $J = 4$ and $K = L = 4$ for scenario 1, $J = 4$ and $K = L = 3$ for scenario 2; more topics do not

contribute to the inference accuracy. Note that the number of spatial topics for scenario 2 is less than scenario 1. This is because the training set of scenario 2 is estimated from the rule-based models, the noisy training set prevents the model from learning more patterns.

There are four Dirichlet priors in our model. These hyper-parameters affect the smoothness of the Multinomial distribution; a larger value will increase the smoothness. Besides, we found hyper-parameters (within a range) have little effect to destination inference accuracy, which is more relevant to the peak rather than the smoothness of distributions. We adopt the typical value in NLP and set $\beta = \gamma = \eta = 0.1$ (Griffiths and Steyvers 2004). The hyper-parameter α affects the smoothness of individuals' topic distribution. Considering it is rare for an individual to possess a wide range of travel patterns; we apply a small value $\alpha = 5/(J \times K \times L)$ to learn a relatively sparse topic distribution that captures individual's specific character. Note a typical setting for α in NLP is 50/(number of topics) (Griffiths and Steyvers 2004).

Benchmark models

We compare our topic model with five benchmark models. For the first four benchmark models, we predict the destination by the most visited destination in a passenger's historical similar trips. The four kinds of "similar trips" are defined as follows:

- (SO) Trip with the same origin.
- (ST) Trip with the same departure time (1-hour interval).
- (SOT_O) Trip with the same origin and departure time, if no such trip, use SO.
- (SOT_T) Trip with the same origin and departure time, if no such trip, use ST.

We adopt the method proposed by He and Trépanier (2015) as the fifth benchmark model, where the destinations of unlinked trips are predicted by the multiplication of spatial and temporal kernel density at potential destinations. This method was developed for bus systems with all potential destinations on the same bus line as the origin. Because the potential destinations of metro systems could be on different lines, we extend the potential destinations with historical destinations that have the same origin as the current trip, and replace the spatial kernel density by the visiting frequency. We choose 1 h as an appropriate bandwidth for the temporal kernel density estimation after comparing different alternatives. We refer to this model as the "kernel-based" method in the following text. When any of the above benchmark models fails, the destination is predicted by the most visited destination of the corresponding passenger.

Scenario 1: using ground truth training set

In scenario 1, we randomly select 70% of the trips as the training set, where the ground truth destinations are known. Table 1 shows the destination inference accuracy of different models in both training and test sets. As the Gibbs sampling depends on the initial value, the accuracies of topic models are means of 50 runs and the standard deviations are shown in parentheses. It can be found that our rank-based topic outperforms the best benchmark models (SOT_O) around 2% in the test set. On the other hand, the no-rank topic model—directly uses station ID in the model—has the worst accuracy, even under a very large number of topics. Our station-to-rank preprocessing greatly improves the

Table 1 The destination inference accuracy of scenario 1

Method	Accuracy (test set)	Accuracy (training set)
SO	67.38%	–
ST	63.49%	–
SOT_O	67.75%	–
SOT_T	64.90%	–
Kernel-based	67.15%	–
Rank topic[a]	69.78% (0.14%)	73.24% (0.14%)
No-rank topic[b]	31.15% (0.19%)	35.30% (0.18%)

[a]The number of topics $J = 4$ and $K = L = 4$

[b]The number of topics $J = 5$ and $K = 10$, $L = 100$

inference accuracy and reduces the latent parameters. As expected, the accuracy of the training set is slightly higher than the test set.

As the topic model requires some historical trips for training, we want to evaluate the effect of an individual's number of training trips to the prediction accuracy. Besides, it is also interesting to investigate the relationship between the destination inference and individuals' travel regularity. There are many metrics to measure individual's travel regularity, such as entropy Scheiner (2014), actual entropy Song et al. (2010), and entropy rate Goulet-Langlois et al. (2017). Entropy measures the randomness of a probability distribution. In metro trips, the entropy of passenger u is defined as

$$E_u = -\sum_{i=1}^{N_u} p_u(i) \log_2 p_u(i).$$

where $p_u(i)$ is the historical probability that location i was visited, N_u is the total number of visited stations. The larger the entropy, the more random the distribution, and harder to predict. Unlike the actual entropy and the entropy rate, the order of the trips does not affect the entropy. As the LDA is a bag-of-words model regardless of the order of words, we use entropy to reflect an individual's travel regularity and evaluate its relation to the accuracy of destination inference.

Figure 3 illustrates the destination inference accuracy under different numbers of training trips and entropy levels. Overall, the number of training trips concentrates on the small end with most passengers having 10 to 25 training trips. On the contrary, the entropy distribution is centered in the middle level with a decreasing trend in the high and low levels. From the bottom right of Fig. 3, it is conspicuous that the prediction accuracy steadily increases with the decrease of the entropy, this is because more regular travelers are easier to predict. With this in mind, it is not hard to understand the relation between the number of training trips and prediction accuracy. The group with around 110 training trips has the highest prediction accuracy, this is because this group has the lowest entropy level (see bottom left of Fig. 3). It is not hard to conclude that the changes in the prediction accuracy are mainly caused by the entropy rather than the number of training trips. The most predictable people are those that have around 110 training trips (around 110/0.7=157 trips with test set) in the 3 months, this number indicates that these people are very likely to be regular commuters. The SOT_O and the rank-based topic model follow the same trend under different numbers of training trips and entropy levels, but our topic model always has higher accuracy.

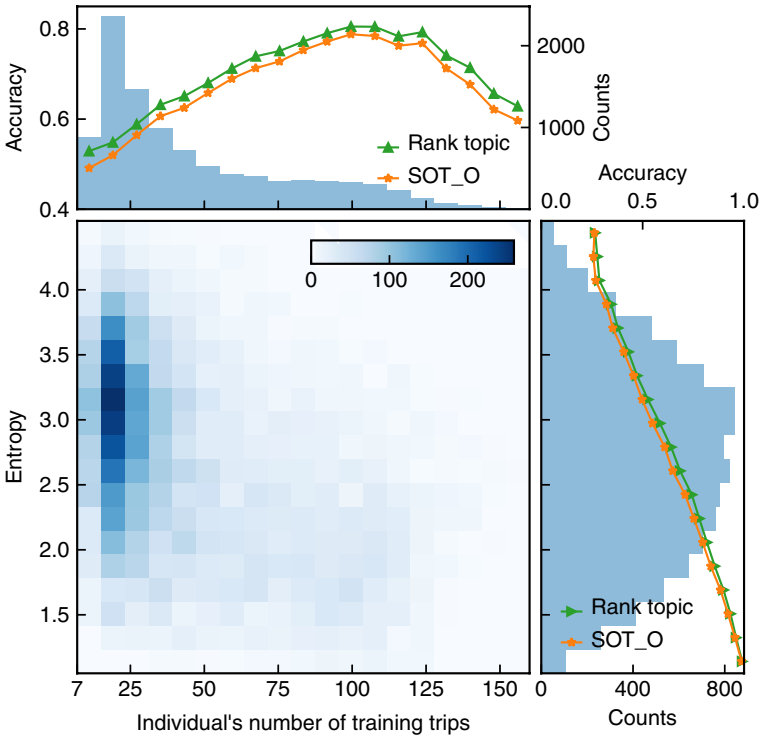

Fig. 3 Bottom left: the 2D histogram of the number of passengers, grided by the number of trips of each passenger in the training set and the entropy of their trips. Bottom right: the destination inference accuracy of two models for passengers at different entropy level and the histogram of entropy. Top: the destination inference accuracy of two models for passengers with different number of trips in the training set and the histogram of the number of trips in the training set. (Accuracies are shown by means.)

Scenario 2: using estimated training set

Scenario 2 imitates the real-world tap-in only system, where the ground truth destinations are unknown. We first use rule-based models to infer the destinations of all linked trips as a training set, and then train our topic model using the estimated training set. The rule-based model that we applied is similar to Trépanier et al. (2007):

- Rule 1: predict the destination as the origin of the next trip in the same day.
- Rule 2: predict the last destination of a day as the first origin of the same day.
- Rule 3: predict the last destination of a day as the first origin of the next day.

The next rule will be only applied when the previous rule is not applicable to a trip. Note that any two metro stations can be connected by transfers; therefore, we do not need to verify whether the origin of the next trip is in the vicinity of the first Metro line, which is different to the bus network in Trépanier et al. 2007.

The accuracy and the coverage of the three rules are shown in Table 2. The assumptions of these rules have been indirectly verified by cordon count data (Barry et al. 2002) and survey data (Barry et al. 2002; Munizaga et al. 2014), only a few study examines these

Table 2 The destination inference accuracy and coverage of scenario 2

	Coverage	Cumulative coverage	Method	Accuracy (test set)	Accuracy (training set)
Linked trips	44.44%	44.44%	Rule 1	86.33%	–
	79.93%	35.49%	Rule 2	76.80%	–
	85.26%	5.34%	Rule 3	60.50%	–
Unlinked trips	14.74%	100.00%	SO	49.63%	–
			ST	43.02%	–
			SOT_O	48.93%	–
			SOT_T	44.19%	–
			Kernel-based	50.51%	–
			Rank topic[a]	51.43% (0.14%)	66.48%(0.16%)
			No-rank topic[b]	31.14% (0.20%)	35.48%(0.18%)

[a]The number of topics $J = 4$ and $K = L = 3$

[b]The number of topics $J = 5$ and $K = 10$, $L = 100$

workhorse assumptions by ground-truth destinations (Alsger et al. 2016). We can tell from Table 2 that Rule 1 using the consecutive trips could reach 86% accuracy. Although destinations inferred by Rule 2 and Rule 3 are less reliable, they are indispensable parts for the training set, because they represent the other side of passengers' travel patterns (e.g., returning home at night). The three rules together handle 85.26% trips.

We then infer the destinations of unlinked trips by our rank-based topic model. Note that for scenario 2, we only use the origins for the ranking. Because the real destinations are unknown and the frequency of destinations is roughly the same with the origins if a passenger uses the smart card to and from. The destination inference results of the topic models and four benchmark models are shown in Table 2, the standard deviations are shown in parentheses. It can be found that the best benchmark model is the kernel-based method with 50.51% accuracy, our rank-based topic model performs slightly better than the kernel-based method with around 51.43% accuracy in the test set. It is noteworthy that the accuracy of the training set is significantly higher than the test set, despite they are both trained by the noisy data. This is because the training set and the test set are not randomly partitioned; there are some differences in the distributions of linked trips and unlinked trips. Besides the lack of ground truth, this difference further impacts the accuracy of scenario 2.

Interpreting latent topics

In the proposed topic model, each topic is characterized by a distribution over time (T), origin (O), or destination (D). By looking at these topic-word distributions, we can endow semantic meanings to latent topics. Therefore, we illustrate the topic-word distribution of scenario 1 by Fig. 4. For time topics in Fig. 4a, we can find topic T2 and T4 have very high probabilities of traveling in the morning, and could be interpreted as early and late morning peaks topics respectively. Contrarily, topic T1 indicates trips in the night and T3 takes the rest of the day. For spatial topics shown in Fig. 4b, c, it can be found that O4 and D1 take near 1 probability for the ranked 1st station, representing boarding and alighting at the most visited station respectively. Meanwhile, O2 and D3 represent boarding and alighting at the second most visited station; O1 and D2 represent boarding and alighting at the

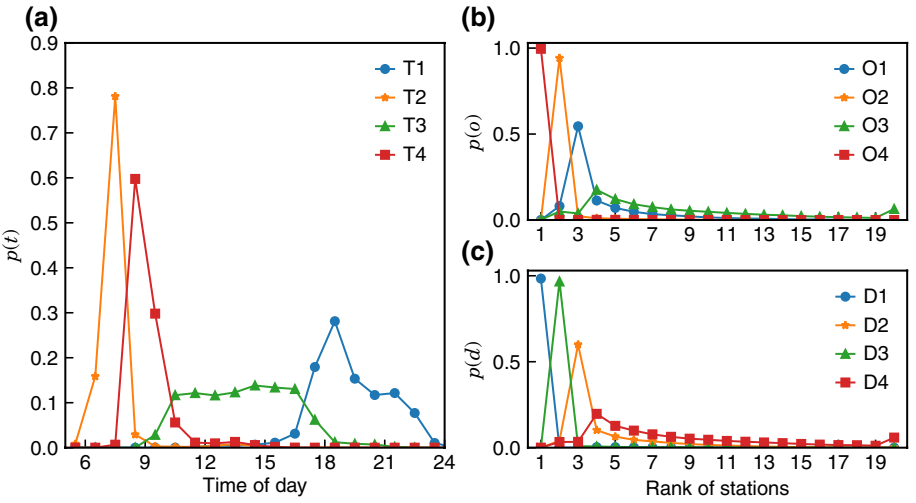

Fig. 4 Topic-word distributions. **a** The departure time distributions of the four time topics. **b** The origin distributions of the four origin topics. **c** The destination distributions of the four destination topics

third most visited station. For O3 and D4, the probabilities peak at the ranked 4th station and then gradually tail off. Moreover, we found the topic-word distribution is quite stable across different runs. Although the order of topics could switch, the shapes of the topic-word distributions maintain unchanged. This suggests the model is insensitive to initial values and the latent topics are good representations for travel patterns.

It is worth mentioning that although the probability of alighting at a station after rank 4 is not zero, it is impossible to predict the destination of a trip as a station ranked after 4 by Eq. 3. Because Eq. 3 always predicts the destination as the most likely one, which is always the most likely destination (peak) in a particular latent destination topic. This limitation causes the accuracy of ranked 4th destination being compromised by stations after rank 4; Luckily, these trips are sparse and with high randomness, the first three destinations make up the majority.

After model training, each passenger is assigned with a distribution over topics, representing to what extent the passenger belongs to each topic. This topic distribution is a high-level summary of a passenger's travel pattern. For example, Fig. 5 shows the latent topic distribution of a passenger. Each matrix represents the probabilities over origin and destination topics under a time topic. Although we don't know the exact mapping relation between the rank of a station and its real function (e.g., home/work), we can easily understand these travel patterns by common sense. It is conspicuous that there are two latent topics with significantly higher probability, indicating a possible commuting pattern. The most significant latent topic is [T2, O4, D3]; according to the semantic meaning shown in Fig. 4, [T2, O4, D3] represents this passenger frequently departure form the most visited station to the second most visited station in the early morning, indicating a possible home-work behavior. Similarly, the second significant topic [T1, O2, D1] represents traveling from the second most visited station to the most visited location in the night, which could be the work-home trip. Besides, [T1, O3, D1] also has a high probability, which could be backing home from the third most visited location (such as a shop) at night. Other noticeable topics are mostly in T3 and have relatively low probabilities, these could be recreational

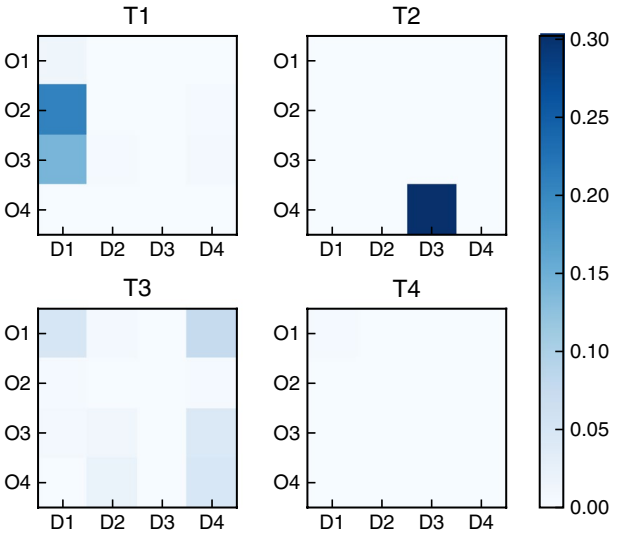

Fig. 5 The latent topic distribution of a passenger

activities. Further, we can find this passenger often use metro in the early morning (T2), night (T1), and afternoon (T3), but seldom use it between 8:00 a.m. to 10:00 a.m. (T4).

Passenger clustering

Passenger clustering is important for personalized service, improving demand models, and various applications. The latent topic distribution is an excellent feature for passenger clustering. There has been a large body of research that uses smart card data for passenger clustering and travel pattern mining. Most existing methods capture either spatial or temporal features. Ma et al. (2013) clustered passengers based on spatial and temporal features; but the two kinds of features are independently defined and then combined together. Our latent topic distribution jointly captures the spatial and temporal patterns in a compact manner, which provides a useful approach for investigating people's travel behaviors.

The feature used for clustering is passenger' latent topic distribution. Jensen-Shannon divergence (JSD) is a metric of measuring the similarity between two probability distributions; we apply the square root of the JSD as the distance between two latent topic distributions. Next, we select 500 passengers and apply hierarchical clustering to illustrate common travel behaviors among the population. Hierarchical clustering is a useful way to visualize the structure of the clustering component. It is also useful in providing the centroid and the number of clusters for faster clustering methods, such as K-means.

The hierarchical clustering of 500 passengers by their latent topic distribution is shown in Fig. 6. Noticeably, passengers on the left half of the figure (From 0 to around 275) show distinct two travel directions: one from the rank 1 station to the rank 2 station and the other from the rank 2 to the rank 1 station, indicating a commuting pattern. More specifically, the temporal topic within R1R2 and R2R1 are different in each cluster, showing these passengers regularly leave from a place at a certain time and then come back at another time, the time at which passengers leave and back distinguishes different clusters. On the other hand, passengers on the right half of Fig. 6 (around 275 to 500) do not have an as significant commuting pattern as those on the left part, and therefore correspond to non-commuters.

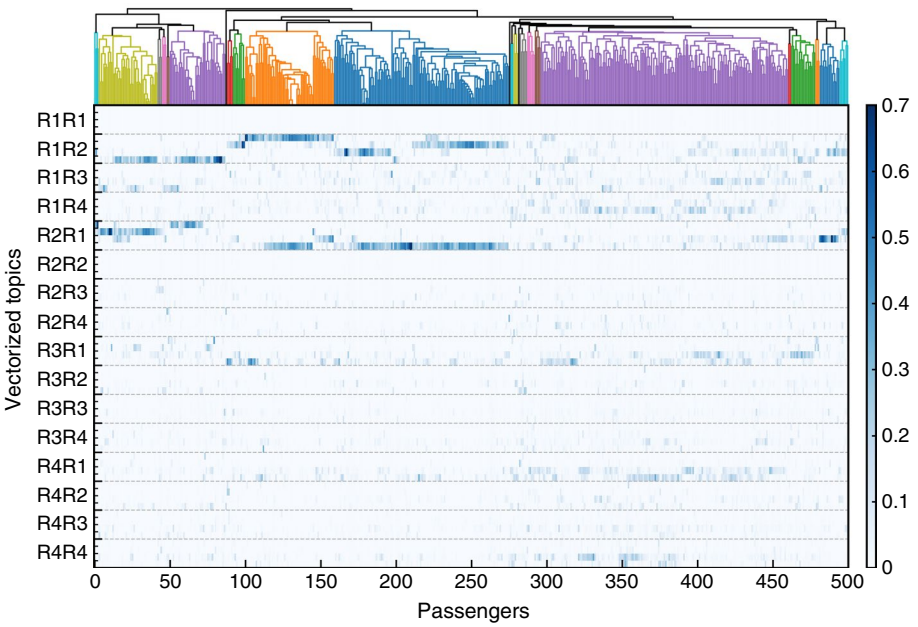

Fig. 6 Top: the dendrogram of the hierarchical clustering on 500 passengers. Bottom: the feature matrix for the clustering; each column in the matrix is a vectorized latent topic distribution of one passenger. The tick labels on y-axis represent the semantic meanings of latent topics: such as, R1R2 means the origin topic that peaks at ranked 1st station and the destinations topic that peaks at ranked 2nd station. The four temporal topics are separated by the minor ticks between every two major ticks and ordered from early morning topic to night topic.

The latent topic distributions of non-commuters show more diverse interactions between different topics, especially between rank 1 stations and others. It is also interesting to find that the early morning topic (in the top minor tick between two major ticks, corresponding to T2 in Fig. 4) mostly belongs to the commuters; most non-commuters' metro trips are in the late morning, afternoon, and night. The distinct pattern between commuters and non-commuters validates the returners and explorers dichotomy in human mobility (Pappalardo et al. 2015). By utilizing the proportion of people under different clusters, a potential application of our topic model is to generate synthetic itineraries of the population for transit simulation.

Conclusions and discussion

This paper uses a probabilistic topic model for smart card data destination estimation and travel pattern mining. We establish a three-dimensional LDA model than captures the time, origin, and destination attributes in smart card trips. Moreover, we introduce a station-to-rank preprocessing that reduces the spatial divergence among passengers to discover more compact latent topics. The case study of Guangzhou Metro shows our model outperforms individual-history-based model by around 2% more accurate, in both scenarios with ground-truth or estimated training set. As a probabilistic model, the destination estimation

accuracy is more related to an individual's travel regularity than the number of trips in the training set. Other than a prediction model, the proposed topic model is also a generative model that explains the probability of a trip by the individual's latent topics (i.e. the probability of traveling from rank o station to rank d station under time topic t) and can be used for travel pattern analysis, and passenger clustering, and trip generation.

For the spatial topics, we introduce a station-to-rank transformation that enhances word co-occurrences among passengers and greatly improves the inference accuracy. The limitation of rank representation is the loss of spatial information. As shown in Fig. 4b, c, each spatial topic actually corresponds to one rank, rather than a mixture of words. Therefore, the topic model cannot be used for spatial clustering as to the vocabulary clustering in natural language processing. Indeed, related research (e.g., Hasan and Ukkusuri (2014); Zhao et al. (2020)) primarily focused on passengers' pattern rather than the spatial similarity. How to derive spatial/region similarity from individuals' transit itineraries is an interesting direction, such as Du et al. (2019). Besides, representing stations by labels (home/work/shop) could further improve the model interpretability; incorporating geographical and land use features could be promising future research. Finally, the effect of our model in the denser bus network is also worth exploring.

There is also improvement space for temporal topics. Firstly, distinguishing weekdays and weekends could be helpful to the prediction. Secondly, how to transform the topic model to a time-varying version is an interesting direction, such as Fan et al. (2016). A problem with our smart card data is that it is a coarse sample for individuals' life trajectory. Even in our sampled 10,000 "frequent" users, around 50% of passengers use metro no more than four times a week; it is hard to utilize connections between neighboring trips under such big travel intervals. Finally, similar to Yin et al. (2017), we can include extraneous variables to improve prediction accuracy. Out of the context of smart card data, it is promising to extend our model for a more general mobility prediction, such as next trip prediction (Zhao et al. 2018).

Acknowledgements An early draft of this paper is presented at the Transitdata2019 workshop. This research is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, Mitacs Canada, the Canada Foundation for Innovation (CFI), and exo (https://exo.quebec/en).

Author contributions The authors confirm contribution to the paper as follows: all authors contributed to the research conception and design; the data analysis was performed by Lijun Sun and Zhanhong Cheng; the first draft of the manuscript was written by Zhanhong Cheng and all authors commented on previous versions of the manuscript; all authors reviewed the results and approved the final version of the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Alsger, A., Assemi, B., Mesbah, M., Ferreira, L.: Validating and improving public transport origin-destination estimation algorithm using smart card fare data. Transp. Res. Part C Emerg. Technol. **68**, 490–506 (2016)

Assemi, B., Alsger, A., Moghaddam, M., Hickman, M., Mesbah, M.: Improving alighting stop inference accuracy in the trip chaining method using neural networks. Public Transp. **12**(1), 89–121 (2020)

Barry, J.J., Newhouser, R., Rahbee, A., Sayeda, S.: Origin and destination estimation in New York city with automated fare system data. Transp. Res. Rec. **1817**(1), 183–187 (2002)

Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)

Briand, A.S., Côme, E., Mohamed, K., Oukhellou, L.: A mixture model clustering approach for temporal passenger pattern characterization in public transport. Int. J. Data Sci. Anal. **1**(1), 37–50 (2016)
Briand, A.S., Côme, E., Trépanier, M., Oukhellou, L.: Analyzing year-to-year changes in public transport passenger behaviour using smart card data. Transp. Res. Part C Emerg. Technol. **79**, 274–289 (2017)
Du, B., Zhou, W., Liu, C., Cui, Y., Xiong, H.: Transit pattern detection using tensor factorization. INFORMS J. Comput. **31**(2), 193–206 (2019)
Fan, Z., Arai, A., Song, X., Witayangkurn, A., Kanasugi, H., Shibasaki, R.: A collaborative filtering approach to citywide human mobility completion from sparse call records. In: International Joint Conference on Artificial Intelligence, pp. 2500–2506 (2016)
González, M.C., Hidalgo, C.A., Barabási, A.L.: Understanding individual human mobility patterns. Nature **453**(7196), 779 (2008)
Gordon, J.B., Koutsopoulos, H.N., Wilson, N.H., Attanucci, J.P.: Automated inference of linked transit journeys in London using fare-transaction and vehicle location data. Transp. Res. Rec. **2343**(1), 17–24 (2013)
Gordon, J.B., Koutsopoulos, H.N., Wilson, N.H.: Estimation of population origin-interchange-destination flows on multimodal transit networks. Transp. Res. Part C Emerg. Technol. **90**, 350–365 (2018)
Goulet-Langlois, G., Koutsopoulos, H.N., Zhao, J.: Inferring patterns in the multi-week activity sequences of public transport users. Transp. Res. Part C Emerg. Technol. **64**, 1–16 (2016)
Goulet-Langlois, G., Koutsopoulos, H.N., Zhao, Z., Zhao, J.: Measuring regularity of individual travel patterns. IEEE Trans. Intell. Transp. Syst. **19**(5), 1583–1592 (2017)
Griffiths, T.L., Steyvers, M.: Finding scientific topics. Proc. Natl. Acad. Sci. **101**(1), 5228–5235 (2004)
Guangzhou Metro (2019). http://www.gzmtr.com/ygwm/gsgk/gsjs/. Last accessed 02 Nov 2019
Hasan, S., Ukkusuri, S.V.: Urban activity pattern classification using topic models from online geo-location data. Transp. Res. Part C Emerg. Technol. **44**, 363–381 (2014)
Hasan, S., Schneider, C.M., Ukkusuri, S.V., González, M.C.: Spatiotemporal patterns of urban human mobility. J. Stat. Phys. **151**(1–2), 304–318 (2013)
He, L., Trépanier, M.: Estimating the destination of unlinked trips in transit smart card fare data. Transp. Res. Rec. **2535**, 97–104 (2015)
He, L., Agard, B., Trépanier, M.: A classification of public transit users with smart card data based on time series distance metrics and a hierarchical clustering method. Transp. Sci. Transp. **16**(1), 56–75 (2020)
Jung, J., Sohn, K.: Deep-learning architecture to forecast destinations of bus passengers from entry-only smart-card data. IET Intel. Transp Syst. **11**(6), 334–339 (2017)
Ma, X., Wu, Y.J., Wang, Y., Chen, F., Liu, J.: Mining smart card data for transit riders travel patterns. Transp. Res. Part C Emerg. Technol. **36**, 1–12 (2013)
Ma, X., Liu, C., Wen, H., Wang, Y., Wu, Y.J.: Understanding commuting patterns using transit smart card data. J. Transp. Geogr. **58**, 135–145 (2017)
Mohamed, K., Côme, E., Baro, J., Oukhellou, L.: Understanding passenger patterns in public transit through smart card and socioeconomic data. In: ACM SIGKDD Workshop on Urban Computing (2014)
Morency, C., Trepanier, M., Agard, B.: Measuring transit use variability with smart-card data. Transp. Policy **14**(3), 193–203 (2007)
MTL Trajet (2019). https://ville.montreal.qc.ca/mtltrajet/en/. Last accessed 08 Nov 2019
Munizaga, M., Devillaine, F., Navarrete, C., Silva, D.: Validating travel behavior estimated from smartcard data. Transp. Res. Part C Emerg. Technol. **44**, 70–79 (2014)
Munizaga, M.A., Palma, C.: Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago, Chile. Transp. Res. Part C Emerg. Technol. **24**, 9–18 (2012)
Nunes, A.A., Dias, T.G., e Cunha JF, : Passenger journey destination estimation from automated fare collection system data using spatial validation. IEEE Trans. Intell. Transp. Syst. **17**(1), 133–142 (2016)
Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., Barabási, A.L.: Returners and explorers dichotomy in human mobility. Nat. Commun. **6**, 8166 (2015)
Pelletier, M.P., Trépanier, M., Morency, C.: Smart card data use in public transit: a literature review. Transp. Res. Part C Emerg. Technol. **19**(4), 557–568 (2011)
Sánchez-Martínez, G.E.: Inference of public transportation trip destinations by using fare transaction and vehicle location data: dynamic programming approach. Transp. Res. Rec. **2652**(1), 1–7 (2017)
Scheiner, J.: The gendered complexity of daily life: effects of life-course events on changes in activity entropy and tour complexity over time. Travel Behav. Soc. **1**(3), 91–105 (2014)
Song, C., Qu, Z., Blumm, N., Barabási, A.L.: Limits of predictability in human mobility. Science **327**(5968), 1018–1021 (2010)
Sun, L., Axhausen, K.W.: Understanding urban mobility patterns with a probabilistic tensor factorization framework. Transp. Res. Part B Methodol. **91**, 511–524 (2016)

Sun, L., Chen, X., He, Z., Miranda-Moreno, L.F.: Pattern discovery and anomaly detection of individual travel behavior using license plate recognition data. In: Transportation Research Board 98th Annual Meeting (2019)
Trépanier, M., Tranchant, N., Chapleau, R.: Individual trip destination estimation in a transit smart card automated fare collection system. J. Intell.Transp. Syst. **11**(1), 1–14 (2007)
Wang, W., Attanucci, J., Wilson, N.: Bus passenger origin-destination estimation and related analyses using automated data collection systems. J. Public Transp. **14**, 131–150 (2011)
Yin, M., Sheehan, M., Feygin, S., Paiement, J.F., Pozdnoukhov, A.: A generative model of urban activities from cellular data. IEEE Trans. Intell. Transp. Syst. **19**(6), 1682–1696 (2017)
Zhang, F., Yuan, N.J., Wang, Y., Xie, X.: Reconstructing individual mobility from smart card transactions: a collaborative space alignment approach. Knowl. Inf. Syst. **44**(2), 299–323 (2015)
Zhao, J., Rahbee, A., Wilson, N.H.: Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. Comput. Aided Civ. Infrastruct. Eng. **22**(5), 376–387 (2007)
Zhao, Z., Koutsopoulos, H.N., Zhao, J.: Individual mobility prediction using transit smart card data. Transp. Res. Part C Emerg. Technol. **89**, 19–34 (2018)
Zhao, Z., Koutsopoulos, H. N., Zhao, J.: Discovering latent activity patterns from transit smart card data: A spatiotemporal topic model. Transp. Res. Part C Emerg. Technol. **116**, 102627 (2020)

Zhanhong Cheng is a Ph.D. student at McGill University. His research focuses on public transportation, spatio-temporal pattern mining, ridership forecast. He is specially interested in using machine learning for a better transportation system.

Martin Trépanier is a professor in the Department of Mathematics and Industrial Engineering at Polytechnique Montréal and the Director of the Interuniversity Center for Research on Enterprise Networks, Logistics and Transportation (CIRRELT). His research interests include information systems in logistics and production, Geographical Information Systems in Transportation (GIS-T), public transportation, vehicle routing in operational logistics, GPS, RFID, smart card data processing.

Lijun Sun is an Assistant Professor in the Department of Civil Engineering at McGill University. His research centers on the area of urban computing and smart transportation, developing innovative methodologies and applications to address efficiency, resilience, and sustainability issues in urban transportation systems. His work has been featured in popular media outlets, including Wired, Citylab, Scientific American, and MIT Technology Review.