Joint Representation Learning for Multi-Modal Transportation Recommendation

Hao Liu[1], **Ting Li**[2], **Renjun Hu**[3], **Yanjie Fu**[4], **Jingjing Gu**[5], **Hui Xiong**[1*]
[1]The Business Intelligence Lab, Baidu Research, National Engineering Laboratory of
Deep Learning Technology and Application, Beijing, China, liuhao30@baidu.com, xionghui@gmail.com
[2]National University of Defense Technology, Changsha, China, liting6259@gmail.com
[3]SKLSDE Lab, Beihang University, Beijing, China, hurenjun@buaa.edu.cn
[4]Missouri University of Science and Technology, Missouri, USA, fuyan@mst.edu
[5]Nanjing University of Aeronautics and Astronautics, Nanjing, China, gujingjing@nuaa.edu.cn

Abstract

Multi-modal transportation recommendation has a goal of recommending a travel plan which considers various transportation modes, such as walking, cycling, automobile, and public transit, and how to connect among these modes. The successful development of multi-modal transportation recommendation systems can help to satisfy the diversified needs of travelers and improve the efficiency of transport networks. However, existing transport recommender systems mainly focus on unimodal transport planning. To this end, in this paper, we propose a joint representation learning framework for multi-modal transportation recommendation based on a carefully-constructed multi-modal transportation graph. Specifically, we first extract a multi-modal transportation graph from large-scale map query data to describe the concurrency of users, Origin-Destination (OD) pairs, and transport modes. Then, we provide effective solutions for the optimization problem and develop an anchor embedding for transport modes to initialize the embeddings of transport modes. Moreover, we infer user relevance and OD pair relevance, and incorporate them to regularize the representation learning. Finally, we exploit the learned representations for online multi-modal transportation recommendations. Indeed, our method has been deployed into one of the largest navigation Apps to serve hundreds of millions of users, and extensive experimental results with real-world map query data demonstrate the enhanced performance of the proposed method for multi-modal transportation recommendations.

Introduction

Transport modes, such as walking, cycling, automobile, public transit, are means for traveling from an origin to a destination. Transportation mode recommendation refers to the effort of finding the most appropriate transport tools with awareness of user preferences (e.g., costs, times) and trip characteristics (e.g., purpose, distance). In this study, we investigate this problem with large-scale navigation App data.

In prior literature, a majority of work focuses on improving unimodal transport planning. For example, the studies in (Dai et al. 2015; Rogers and Langley 1998; Chen, Shen, and Zhou 2011; Luo et al. 2013) aim at recommending the most cost-effective driving routes. The study in (Liu et al. 2014) optimizes bus routing strategies based on human mobility patterns. However, in fact, a large number of transport planning requests involve the decision process of choosing an appropriate one from multiple available transport modes. According to statistics from Baidu Map[1], one of the world's largest navigation Apps, over ten million of users request the route planning information of at least two transport modes (e.g., driving and bus) for same OD pair every day. Intuitively, the transport mode preferences vary over different users and spatiotemporal contexts. For example, metros are more cost-effective than taxis for most urban commuters; economically disadvantaged people may prefer cycling and walking to others for local travel, if the transport options are inadequate. Such socio-economic personalized effects in choosing transport modes present a unique challenge, but, meanwhile, a great potential for improving transportation route planning: if we can decide an appropriate transport mode for a trip, it is easier to plan an optimized route in the following. Unfortunately, there are limited studies that intend to systematically address the multi-modal transportation recommendation problem: *given a user and an OD pair, how to effectively identify the most appropriate transport mode?*

Smart transport mode recommendations have a number of advantages, including but not limited to reducing transport times, balancing traffic flows, reducing traffic congestion, and, ultimately, promoting the development of intelligent transportation systems. However, it is a non-trivial task to infer the most appropriate transport mode. Specifically, there are three major challenges.

The first challenge comes from the transport heterogeneity. In particular, transport modes are heterogeneous in terms of the means of conveyance. For example, cycling and walking are human-powered while bus and taxi are gasoline-powered. In addition, transport modes are heterogeneous in terms of infrastructures, e.g., walking require pedestrian lanes, and cycling require bike lanes. Moreover, transport modes are heterogeneous in terms of time-efficiency, cost, and comfortableness. Consequently, users choose different transport modes for different OD pairs. Then, one research issue is: how to incorporate the transport heterogeneity for multi-modal transportation recommendation? The second

*Corresponding author.

[1]https://map.baidu.com/

challenge is that feedbacks in navigation Apps are incomplete and implicit. In particular, due to privacy issues, when a user queries a transport plan from an origin to a destination, the preference feedbacks, such as clicking and viewing a routing plan of a transport mode, are mostly implicit. Thus, for effective recommendations, it is critical to address the issue of incomplete and implicit feedbacks in data. The third challenge is geo-spatial locality. Despite the exploratory nature of human mobility (Song et al. 2010), each user usually and repeatably visits a very small number of places ($\sim$25) (Alessandretti et al. 2018). In other words, the frequently-traveled and preferred OD pairs are very limited for most users. As a result, it is challenging for the proposed method to learn robust transport mode preferences of users and OD pairs from the limited and localized feedback data.

To tackle these challenges, we propose Trans2Vec, a joint representation learning based framework for multi-modal transportation recommendation. Specifically, inspired by the recent success of word embedding (Mikolov et al. 2013) and network embedding (Cui et al. 2017), we first model users, OD pairs, and transport modes together into a multi-modal transportation graph and jointly learn the user preference and OD preference in a unified latent space. The representation of incomplete user and OD pair is curated since the preference can be learned by capturing the second order proximity from the multi-modal transportation graph. Besides, an anchor based method is proposed to model the heterogeneity of transport modes. Moreover, we incorporate the user-user and OD-OD relevance to identify the interconnections between users and between ODs to overcome the data sparsity and geo-spatial locality issues. Finally, we propose a simple yet effective method to support real-time transport mode recommendations in the online mapping service. Our major contributions are summarized as follows:

- We formally define the transport mode recommendations problem identified from a real-life scenario.
- We propose Trans2Vec, a general framework for transport mode recommendations. Trans2Vec is a joint optimization framework that combines the historical travel behaviors and user/OD relevance for learning transport mode preference. An online recommendation method is also proposed to support the real-time recommendations.
- We conduct extensive experimental evaluations based on real-world datasets. The results demonstrate the effectiveness of our proposed framework.
- Trans2Vec has been deployed in a map and navigation App and serves hundreds of millions of users every day.

Preliminaries

We first introduce some important definitions and then formalize our problem to investigate.

Assuming there is a set $\mathcal{U}$ of users, in which each user $u \in \mathcal{U}$ is associated with a vector $\mathcal{A}(u)$ of demographic attributes, such as age, gender and job type. Besides, we partition a city into a set $\mathcal{R}$ of non-overlapping regions. Each region $r \in \mathcal{R}$ is associated with a set $\mathcal{P}(r)$ of POIs, that represent the functionalities of regions (Yuan, Zheng, and Xie 2012). Considering the existence of users $\mathcal{U}$ and regions $\mathcal{R}$, we next define some essential concepts:

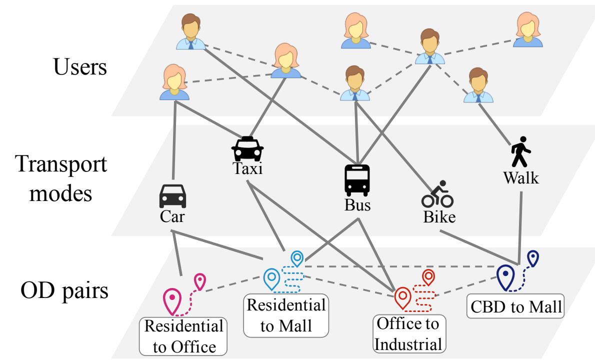


Figure 1: An illustrative example of a multi-modal transportation graph. The nodes of the graph include users, transport modes, and OD pairs. Solid lines indicate observed user-mode and OD-mode edges, whereas dashed lines indicate the user-user and OD-OD relevance edges.

Definition 1 ***Origin-Destination (OD) Pair****. An OD pair $od = (o, d)$ is a pair of regions in which there are travel demands from $o \in \mathcal{R}$ to $d \in \mathcal{R}$. Be sure to note that, given an OD pair (o, d), $o = d$ indicates an intra-region trip. We further denote the set of OD pairs by $\mathcal{OD}$.*

Definition 2 ***Transport Mode****. A transport mode $m \in \mathcal{M}$ is a mean by which passengers move from an origin to a destination. Typical transport modes are bike, bus, taxi, train, ferry, and airplane. Different transport modes yield different travel times, expenses, safety, and comfortableness. Consequently, both users and OD pair have personalized preferences and appropriateness on specific modes. In this study, we focus on intra-city transport modes: bus, taxi, train, bicycle, and walk.*

Definition 3 ***Travel Event****. A travel event $q \in \mathcal{Q}$ is a triplet $\langle u, m, od \rangle$, which represents the user $u \in \mathcal{U}$ travel between the OD pair $od \in \mathcal{OD}$ by the transport mode $m \in \mathcal{M}$.*

Definition 4 ***Multi-modal Transportation Graph(MMTG)****. A multi-modal transportation graph is a heterogeneous undirected weighted graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \mathcal{U} \cup \mathcal{OD} \cup \mathcal{M}$ is a set of heterogeneous nodes, and $\mathcal{E} = \mathcal{E}_{um} \cup \mathcal{E}_{odm} \cup \mathcal{E}_{uu} \cup \mathcal{E}_{odod}$ is a set of heterogeneous edges including user-mode edges $\mathcal{E}_{um}$, OD-mode edges $\mathcal{E}_{odm}$, user-user edges $\mathcal{E}_{uu}$ and OD-OD edges $\mathcal{E}_{odod}$. When it is clear from the context, we directly refer a node in G to as a user, an OD pair or a mode.*

Figure 1 gives an example of a multi-modal transportation graph. We obtain edges of G from both travel events and user/OD attributes. The weight of a user-mode (resp. an OD-mode) edge is the frequency of travel events that are associated with the user (resp. OD) and the transport mode. The weight of a user-user (resp. an OD-OD) edge is defined as the relevance of two corresponding nodes (details are in Section Methodology). With the aforementioned notations and concepts, we finally formalize our problem as follows:

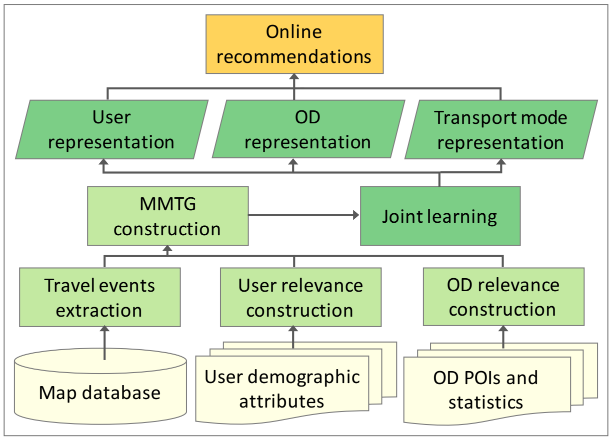


Figure 2: The Trans2Vec framework.

Definition 5 ***The Multi-modal Transportation Recommendation Problem.*** *Given a multi-modal transportation graph G, a user u, and an OD pair od, we aim to recommend the most appropriate transport mode $m \in \mathcal{M}$ for the user u to travel between the OD pair od.*

Methodology

In this section, we introduce the proposed analytic framework Trans2Vec. We first present an overview of Trans2Vec, then detail each step of the proposed framework.

Framework Overview

We study the multi-modal transportation recommendation problem via joint representation learning on multi-modal transportation graphs. Figure 2 shows an overview of the proposed analytic framework. It consists of three major tasks: (1) multi-modal transportation graph construction, (2) joint representation learning to capture the transport mode preferences of users and ODs, and (3) online transport mode recommendations.

Specifically, we construct a multi-modal transportation graph based on heterogeneous source data including a map database, user demographic attributes, and OD POIs information. The graph thus contains rich semantics, i.e., travel event occurrence and user/OD relevance. Second, we develop a joint representation learning method that, given the multi-modal transportation graph G, learns a d-dimensional ($d \ll |\mathcal{V}|$) latent representation $\mathbf{x}_v$ for each node $v \in \mathcal{V}$. Our method exploits a novel anchor embedding strategy to model and maintain the heterogeneity of different transport modes and further incorporates user and OD relevance into the base embedding model. Finally, we propose a simple yet effective online recommendation method to recommend the most proper transport mode.

Base Model

The basic idea of the proposed method is from word and network embedding (Mikolov et al. 2013; Perozzi, Al-Rfou, and Skiena 2014), which assumes that semantically-related words and correlated nodes have the similar context in sentences and network structures. We analogize travel events to sentences and random walks, in order to learn embeddings of users, OD pairs, and transport modes. If a user, an OD pair, and a transport mode co-occur in the same travel event, it implies both the user and the OD pair have certain preference for the transport mode. More specifically, a travel event $\langle u, m, od\rangle \in \mathcal{Q}$ can indicate that the embeddings of the user u and the OD pair od should be close to the embedding of the transport mode m.

We denote the corresponding embeddings of u, od and m by $\mathbf{x}_u$, $\mathbf{x}_{od}$ and $\mathbf{x}_m$, respectively. The user-mode and OD-mode relationships of all the travel events in $\mathcal{Q}$ are represented in the form of $\mathcal{E}_{um} \cup \mathcal{E}_{odm}$. We then achieve the goal by maximizing the following objective function for all of the edges in $\mathcal{E}_{um} \cup \mathcal{E}_{odm}$:

$$O_0 = \sum_{(u,m)\in\mathcal{E}_{um}} \log\sigma(\mathbf{x}_u^\top \cdot \mathbf{x}_m) + \sum_{(od,m)\in\mathcal{E}_{odm}} \log\sigma(\mathbf{x}_{od}^\top \cdot \mathbf{x}_m), \tag{1}$$

where $\sigma(x) = 1/(1+e^{-x})$ is the sigmoid function.

Note that the objective in Eq. (1) reaches a global optima when all the nodes have the same embeddings. The problem is that the captured preference is uninformative if we omit the disfavor information. To avoid converging to the meaningless trivial solution, we adopt the negative sampling strategy (Mikolov et al. 2013), such that for each $(u, m) \in \mathcal{E}_{um}$ (resp. $(od, m) \in \mathcal{E}_{odm}$), we randomly sample another mode $m' \in \mathcal{M}$, i.e., negative sample, and move the embedding $\mathbf{x}_u$ (resp. $\mathbf{x}_{od}$) away from $\mathbf{x}_{m'}$. Correspondingly, the objective function is rewritten as:

$$\begin{aligned} O_0 = &\sum_{\substack{(u,m)\in\mathcal{E}_{um}\\ m'\sim U}} \big(\log\sigma(\mathbf{x}_u^\top \cdot \mathbf{x}_m) + \log\sigma(-\mathbf{x}_u^\top \cdot \mathbf{x}_{m'})\big) \\ + &\sum_{\substack{(od,m)\in\mathcal{E}_{odm}\\ m'\sim U}} \big(\log\sigma(\mathbf{x}_{od}^\top \cdot \mathbf{x}_m) + \log\sigma(-\mathbf{x}_{od}^\top \cdot \mathbf{x}_{m'})\big). \end{aligned} \tag{2}$$

In Eq. (2), U is a uniform distribution for generating negative mode samples so that $U(m) = 1/|\mathcal{M}|$. As we can see, the trivial solution is not an optima anymore because the negative sample terms also need to be maximized.

Transport Mode Anchor Embedding

The multi-modal transportation graph has unique characteristics. In the graph, there are only several (e.g., 5 in our case) transport mode nodes whereas there are a large number of user nodes and OD nodes. A transport mode node connects with numerous users and OD pairs. Consequently, transport mode nodes can be regarded as hub nodes. In other words, almost every node in the graph is reachable to each other via two-hop paths. If we apply classic network embedding techniques (Perozzi, Al-Rfou, and Skiena 2014; Tang et al. 2015), all of the nodes will have similar embeddings due to the high second-order proximity between almost every pair of the nodes. However, this will lead to a serious issue: we are not able to learn discriminative embeddings of nodes for effective multi-modal transportation recommendation. To address such challenge, we propose an *anchor embedding* strategy for transport mode nodes. The primary objective is to ensure that each mode is assigned a discriminative embedding that reflects its inherent context information.

More specifically, we maintain each transport mode as a fixed anchor vector in the latent space. The anchor vector will not be changed after initialization. In the learning procedure, only user and OD embeddings are optimized to maximize Eq. (2). Consequently, as long as the anchor vectors are initialized properly, each user or OD pair will be allocated in an appropriate location in the hidden space that captures the distinct preference of the user or the OD pair for each transport mode.

In addition, the strategy provides great potential to incorporate prior human knowledge regarding the transport modes into the initialization step of anchor vectors. For instance, compared with the transport modes of bus, train, and airplane, the mode of walk is more similar to the one of bike because: both walk and bike are human powered while bus, train, and airplane are gasoline-powered. Such human knowledge can be reflected in the distance between the anchor vectors of the modes.

Along this line, we develop a data-driven approach to quantifying the relevance between transport modes:

$$rel(m_i, m_j) = \frac{\sum_{v_i \in \mathcal{Z}} F_{m_i}(v_i) F_{m_j}(v_i)}{\sqrt{\sum_{v_i \in \mathcal{Z}} F_{m_i}(v_i)^2}\sqrt{\sum_{v_i \in \mathcal{Z}} F_{m_j}(v_i)^2}}, \tag{3}$$

where $\mathcal{Z} = \mathcal{U} \cup \mathcal{OD}$, $F_{m_i}(\cdot)$ computes the frequency the input $v \in \mathcal{Z}$ ever co-occurred with transport mode m_i in $\mathcal{Q}$.

After measuring the mode-mode pairwise relevance, we employ a greedy strategy to initialize transport mode anchor embeddings. In particular, we first initialize the two transport modes of the highest relevance. In each of following iterations, we initialize the mode node of highest overall relevance with previously initialized mode nodes under the pairwise relevance order constraints. For example, assuming $rel(m_{bus}, m_{taxi}) > rel(m_{bus}, m_{bicycle}) > rel(m_{taxi}, m_{bicycle})$, with anchor embedding $\mathbf{x}_{bus}$ and $\mathbf{x}_{taxi}$ initialized, the distance of anchor embedding $\mathbf{x}_{bicycle}$ should satisfy $rel(\mathbf{x}_{bus}, \mathbf{x}_{taxi}) > rel(\mathbf{x}_{bus}, \mathbf{x}_{bicycle})$ and $rel(\mathbf{x}_{bus}, \mathbf{x}_{taxi}) > rel(\mathbf{x}_{taxi}, \mathbf{x}_{bicycle})$, simultaneously. Note that each entry of anchor embedding is initialized with an uniform distribution between $[-\frac{1}{2d}, \frac{1}{2d}]$. As a result, transport modes frequently co-used by same users and same ODs are closer to each other.

Modeling User and OD Relevance

In real life, the choice of transport mode is highly influenced by the characteristics of both users and OD pairs. For example, business users may prefer traveling by taxi than by bus for long-distance travel; on the other hand, the bus may be in general popular for urban commuter to go to work. That is, users (OD pairs) with similar characteristics would share similar transport mode preferences. Next, we propose methods to measure the relevance of users and OD pairs and further incorporate the relevance to refine our framework.

Measuring user relevance. We collect the demographic attributes of users in order to measure the user-user relevance. In particular, the attributes of user u is represented by a vector $\mathcal{A}(u)$ where each dimension is a categorical demographic attribute. In practice, different attributes have different impacts on the relevance of users regarding transport mode preference. We thus learn a vector $\mathbf{w}$ of weights by predicting transport modes taken by users via a linear regression model, i.e., $\arg\min_{\mathbf{w},b} \sum_{(u,m,od)\in\mathcal{Q}} (\mathbf{w}^\top \mathcal{A}(u) + b - l(m))^2$, where each transport mode m is assigned a distinct random label $l(m)$ in $\{1, \ldots, |\mathcal{M}|\}$. Given two users u and u', the user relevance $c(u, u')$ is then given by:

$$rel(u, u') = \sum_i \mathbf{w}_i I\big(\mathcal{A}(u)_i, \mathcal{A}(u')_i\big) / \sum_i \mathbf{w}_i, \tag{4}$$

where subscript i means the i-th element of a vector, $I(\cdot)$ is a 0–1 indication function which equals to 1 iff. the input categorized attributes are the same. Since mode preferences are only shared by users having high relevance, we thus collect the set $\mathcal{E}_{uu}$ of user relevance edges as $\{(u, u')|u \in \mathcal{U}, u' \in N(u)\}$ where $N(u)$ includes the top-K most relevant users of u. Note that $N(u)$ can be efficiently retrieved as the K-nearest neighbors in a KD-tree where each user u is represented by the Hadamard product $\mathbf{w} \odot \mathcal{A}(u)$. Finally, we incorporate user relevance into our framework as constraints by maximizing the objective function below:

$$O_1 = -\frac{1}{2} \sum_{(u,u')\in\mathcal{E}_{uu}} \big(\mathbf{x}_u^\top \cdot \mathbf{x}_{u'} - rel(u, u')\big)^2. \tag{5}$$

Measuring OD relevance. Considering that distance and travel purpose (e.g., home-work, home-commercial) are among the most influential factors for choosing transport modes, here we calculate the OD relevance from the above two perspectives. Given an OD pair $od = (o, d)$, the distance $dist_{od}$ is computed by Haversine formula. On the other hand, the travel purpose is highly related to the functions of regions o and d, which can further be captured by the region POI distribution (Yuan, Zheng, and Xie 2012). Specifically, a region r is modeled as a POI distribution vector $\mathbf{p}_r$ computed from the POI set $\mathcal{P}(r)$; each dimension represents the number of POIs in a certain category such as residence, entertainment and transport station. The OD pair od is represented as a concatenated vector of the above two views:

$$\mathbf{od} = d_{od} \oplus \mathbf{p}_o \oplus \mathbf{p}_d. \tag{6}$$

Since the impact of each dimension also varies, we learned another vector $\mathbf{w}$ of weights in exactly the same way as user relevance (we also reuse the notation $\mathbf{w}$ as this does not introduce confusion). The relevance between two OD pairs od and od' is computed as:

$$rel(od, od') = \exp\{-||\mathbf{w} \odot (\mathbf{od} - \mathbf{od'})||\}. \tag{7}$$

Note that Eqs. (4) and (7) differ from each other in the way computing relevance, as they are dealing with categorical and numerical features, respectively. Finally, we incorporate the OD relevance via maximizing the following objective:

$$O_2 = -\frac{1}{2} \sum_{(od,od')\in\mathcal{E}_{odod}} \big(\mathbf{x}_{od}^\top \cdot \mathbf{x}_{od'} - rel(od, od')\big)^2, \tag{8}$$

where $\mathcal{E}_{odod} = \{(od, od')|od \in \mathcal{OD}, od' \in N(od)\}$ is the set of OD relevance edges, and, $N(od)$ includes the top-K most relevant OD pairs of od retrieved by another KD-tree.

Algorithm 1: Joint learning algorithm of Trans2Vec

Input: A multi-modal transportation graph G, number d of dimensions, number K, learning rate α, parameters β_1 and β_2;

Output: $\mathbf{x}_u$/$\mathbf{x}_{od}$/$\mathbf{x}_m$ for u/od/$m \in \mathcal{U}$/$\mathcal{OD}$/$\mathcal{M}$;

Initialize entries of $\mathbf{x}_u$/$\mathbf{x}_{od}$/$\mathbf{x}_m$ with uniform$[-\frac{1}{2d}, \frac{1}{2d}]$;

Compute user and OD relevance with Eqs. (4) & (7);

$iter \leftarrow 1$;

repeat

  foreach $(u, u') \in \mathcal{E}_{uu}$ **do**

    $\mathbf{x}_u \leftarrow \mathbf{x}_u - \frac{\alpha\beta_1}{iter}\big(\mathbf{x}_u^\top \cdot \mathbf{x}_{u'} - rel(u, u')\big)\mathbf{x}_{u'}$;

    $\mathbf{x}_{u'} \leftarrow \mathbf{x}_{u'} - \frac{\alpha\beta_1}{iter}\big(\mathbf{x}_u^\top \cdot \mathbf{x}_{u'} - rel(u, u')\big)\mathbf{x}_u$;

  foreach $(od, od') \in \mathcal{E}_{odod}$ **do**

    $\mathbf{x}_{od} \leftarrow \mathbf{x}_{od} - \frac{\alpha\beta_2}{iter}\big(\mathbf{x}_{od}^\top \cdot \mathbf{x}_{od'} - rel(od, od')\big)\mathbf{x}_{od'}$;

    $\mathbf{x}_{od'} \leftarrow \mathbf{x}_{od'} - \frac{\alpha\beta_2}{iter}\big(\mathbf{x}_{od}^\top \cdot \mathbf{x}_{od'} - rel(od, od')\big)\mathbf{x}_{od}$;

  foreach $(u, m) \in \mathcal{E}_{um}$ **do**

    Sample a transport mode $m' \sim U$;

    $\mathbf{x}_u \leftarrow \mathbf{x}_u - \frac{\alpha}{iter}\big(\sigma(\mathbf{x}_u^\top \cdot \mathbf{x}_m) - 1\big)\mathbf{x}_m - \frac{\alpha}{iter}\sigma(\mathbf{x}_u^\top \cdot \mathbf{x}_{m'})\mathbf{x}_{m'}$;

  foreach $(od, m) \in \mathcal{E}_{odm}$ **do**

    Sample a transport mode $m' \sim U$;

    $\mathbf{x}_{od} \leftarrow \mathbf{x}_{od} - \frac{\alpha}{iter}\big(\sigma(\mathbf{x}_{od}^\top \cdot \mathbf{x}_m) - 1\big)\mathbf{x}_m - \frac{\alpha}{iter}\sigma(\mathbf{x}_{od}^\top \cdot \mathbf{x}_{m'})\mathbf{x}_{m'}$;

  $iter \leftarrow iter + 1$;

until *converge*;

return $\mathbf{x}_u$/$\mathbf{x}_{od}$/$\mathbf{x}_m$ for u/od/$m \in \mathcal{U}$/$\mathcal{OD}$/$\mathcal{M}$;

Model Training and Online Recommendations

The optimization objective of our method jointly considers traditional heterogeneous network embedding, negative sampling, user-user graph regularization, and OD-OD graph regularization:

$$O = O_0 + \beta_1 O_1 + \beta_2 O_2, \tag{9}$$

where β_1 and β_2 are hyperparameters that regularize the importance of user and OD relevance. We utilize stochastic gradient descent to train the embeddings. The overall learning process is presented in Algorithm 1.

After deriving the embeddings, we propose a real-time and effective method to compute the preference score of each transport mode m given a user u and an OD pair od:

$$f(u, od, m) = \gamma \mathbf{x}_u^\top \cdot \mathbf{x}_m + (1 - \gamma)\mathbf{x}_{od}^\top \cdot \mathbf{x}_m, \tag{10}$$

where γ is a hyperparameter to control the weights of user preference and OD preference. Finally, we rank the transport modes based on the computed scores and return the one with the highest score as the recommendation.

Experiments

Using large-scale real-life datasets, we present an extensive experimental study to evaluate: (1) the overall performance of Trans2Vec, (2) the parameter sensitivity, (3) the transport mode relevance and (4) the robustness of our approach.

Notation	Description	BEIJING	SHANGHAI
$\lvert\mathcal{Q}\rvert$	# of travel events	1,137,688	1,117,981
$\lvert\mathcal{U}\rvert$	# of users	318,879	316,060
$\lvert\mathcal{OD}\rvert$	# of ODs	375,165	350,904
$\lvert\mathcal{M}\rvert$	# of modes	5	5

Table 1: Statistics of datasets

Experimental Setups

Data description. We choose two datasets BEIJING and SHANGHAI to test our approach. Both of them are produced based on the map queries and user feedbacks on the Baidu Map, in the corresponding cities from April 1, 2018 till August 20, 2018. More specifically, a travel event is recorded if a user issues an OD query and clicks on a specific transport mode. We also use the POI and user profile data collected by the same navigation Apps to generate features of users and OD pairs. The statistics of our datasets are summarized in Table 1.

Evaluation metrics. We adopt the overall NDCG[2], as well as the weighted precision (PREC), recall (REC) and F1 metrics to evaluate the performance. For the last three weighted metrics, we first compute results at transport mode level and then combine the results scaled by the relative frequency of each transport mode. Also note that the NDCG metric takes the orders of all transport modes into consideration and the rest metrics only care about the top-1 recommended mode. Finally, the REC metric is equivalent to the overall accuracy in our case as we recommend a transport mode for each travel event.

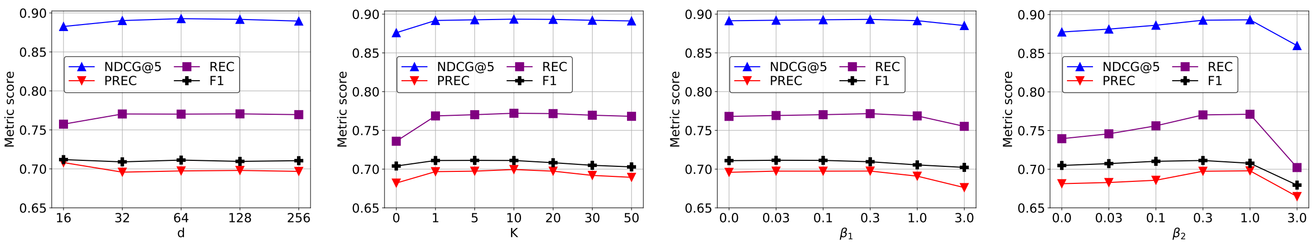
Baseline algorithms. We compare our full approach Trans2Vec with two recommendations methods, two embedding learning methods and one variant of Trans2Vec.

- LR uses logistic regression for transport mode recommendations, which, given a user u and an OD pair od, takes the user attribute vector of $\mathcal{A}(u)$, the distance d_{od} and the fraction of historical rides using each transport mode by u and between od as input features.
- LTR is a LambdaMart (Burges 2010) learning-to-rank method whose input features are the same as LR except that it only uses the fraction of historical rides using the transport mode to rank.
- PTE (Tang, Qu, and Mei 2015) first learns user, OD and transport mode embeddings by treating $\{u, od\}$ as the document and mode m as the label for each travel event $\langle u, m, od\rangle$, and utilizes logistic regression that takes the corresponding embeddings as input for recommendations.
- Metapath2Vec (Dong, Chawla, and Swami 2017) is the same to PTE except that it learns embeddings from random walks following meta-path $\mathcal{U}$–$\mathcal{M}$–$\mathcal{OD}$–$\mathcal{M}$–$\mathcal{U}$–. . . .
- BTrans2Vec is the base model of Trans2Vec which does not incorporate user and OD relevance.

[2] The overall NDCG refers to NDCG@5 in this paper.

Algorithm	Beijing				Shanghai			
	Ndcg@5	Prec	Rec	F1	Ndcg@5	Prec	Rec	F1
LR	0.804	0.704	0.589	0.633	0.848	0.682	0.657	0.658
LTR	0.824	0.667	0.662	0.664	0.830	0.671	0.666	0.668
PTE	0.770	0.493	0.518	0.499	0.807	0.564	0.610	0.585
Metapath2Vec	0.731	**0.718**	0.439	0.515	0.736	**0.728**	0.451	0.528
BTrans2Vec	0.876	0.682	0.736	0.704	0.878	0.695	0.754	**0.718**
Trans2Vec	**0.893**	0.700	**0.770**	**0.711**	**0.891**	0.708	**0.778**	**0.719**

Table 2: Overall performance



(a) Effect of d (b) Effect of K (c) Effect of β_1 (d) Effect of β_2

Figure 3: Parameter sensitivity on Beijing

Implementations. We use the data from April through July for training, i.e., learning embeddings, counting historical numbers of rides and training models, and the rest data for testing. We use the recommended parameters for all baselines, and fixed learning rate $\alpha = 0.5$ and 0.3 for Beijing and Shanghai, respectively, number d of dimensions to 64, number K of relevance neighbors to 5, and regularizing parameters β_1, β_2 and γ to 0.1, 0.3 and 0.5, respectively, by default for our BTrans2Vec and Trans2Vec (we will report the sensitivity analysis later).

Experimental Results

Overall performance. In the first set of experiments, we evaluate the overall performance of all tested approaches for transport mode recommendation. We test the performance on both Beijing and Shanghai datasets using the four metrics. The results are reported in Table 2.

As can be seen, both our basic BTrans2Vec and complete Trans2Vec approaches achieve significantly better performance compared with the rest tested baselines on both datasets using all metric except Prec. Indeed, the Ndcg@5, Rec and F1 of BTrans2Vec is already (5.1%, 5.0%, 8.9%, 14.3%)[3], (12.2%, 8.1%, 18.1%, 30.0%) and (6.6%, 4.5%, 16.9%, 19.0%) higher than those of (LR, LTR, PTE, Metapath2Vec) on average on the two datasets, respectively. Moreover, using (Ndcg@5, Rec, F1), our complete Trans2Vec further outperforms BTrans2Vec by (1.5%, 2.9%, 0.4%) on average, respectively. These results indicate that our joint representation learning technique enhanced with the anchor embedding strategy is effective for capturing transport mode preference. And the incorporated user and OD relevances do have positive impacts for transport mode recommendation.

On the other hand, our approaches are slightly worse in terms of Prec, compared with Metapath2Vec. While the Rec of BTrans2Vec and Trans2Vec are consistently much better than the one of Metapath2Vec. Indeed, BTrans2Vec and Trans2Vec strike a better balance between the Prec and Rec, as illustrated by performance evaluated by the more comprehensive metric F1.

Parameter sensitivity. Due to the space constraints, here we report the impacts of the number d of dimensions, the number K of relevance neighbors, and regularizing parameters β_1 and β_1 on Beijing data. The results on Shanghai are similar. Each time we vary a parameter, set others to their default values, and evaluate the performance of Trans2Vec using all metrics.

To test the impacts of the number d of dimensions, we vary d from 16 to 256. The results are reported in Fig. 3(a). Overall, our approach is robust to the selection of d. A small d is already sufficient for distinguishing transport modes considered in our case. Also note that when increasing d from 16 to 32, the Rec increases at the cost of obtaining worse Prec.

To test the impacts of the number K of relevance neighbors, we vary K from 0 to 50. The results are reported in Fig. 3(b), including the ones of BTrans2Vec with $K = 0$. There can be observed a remarkable improvement when varying K from 0 to 1. The performance degenerates with larger K, especially for Prec, possibly because noises are introduced. Finally, we note that it suffices to use a small K, e.g., 5 and 10.

To test the impacts of parameter β_1, we varied β_1 from 0 to 3. The results are reported in Fig. 3(c). When varying β_1, the scores of all tested metrics first slightly increase and then decrease. The best performance is achieved when

[3] % refers to absolute percentage points throughout the paper.

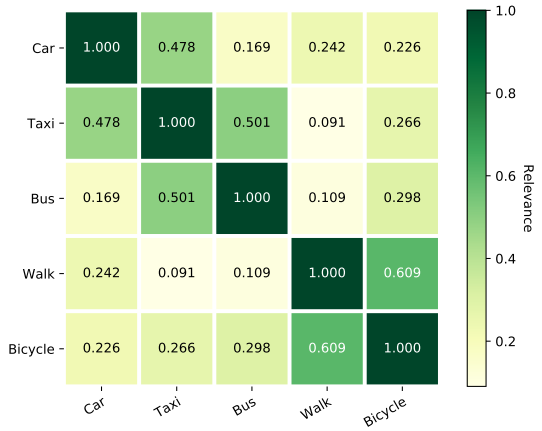
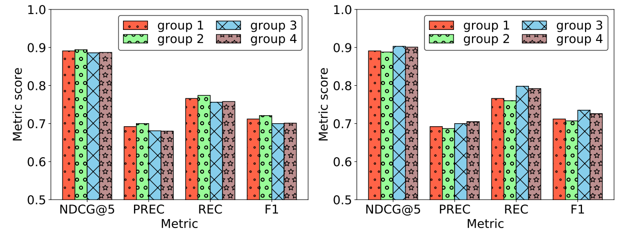


Figure 4: Pairwise transport mode relevance matrix



(a) Group by users (b) Group by OD pairs

Figure 5: Robustness check on BEIJING

$\beta_1 \in [0.1, 0.3]$. And we admit that the user relevance only has a minor improvement, possibly because the correlation between available user attributes and transport mode preference is not such strong compared the OD relevance.

To test the impacts of parameter β_2, we also vary β_2 from 0 to 3. The results are reported in Fig. 3(d). Again, when varying β_1, all tested scores first increase and then decrease. And when $\beta_2 \in [0.3, 1.0]$, incorporating OD relevance results in a remarkable improvement.

Transport mode relevance. Figure 4 reports the pairwise transport mode relevance according to Eqn. 3. We can observe that transport modes can be clustered into two groups: car, taxi, bus in one group and bus, walk, bicycle in another group. Bus, as the nexus of two groups, has relatively high relevance with taxi and bicycle. Such groups make sense since these two groups have significantly different characteristics in efficiency, cost, and environmental impact.

Robustness check. To evaluate the robustness of Trans2Vec, we first group users (resp. OD pairs) into four subgroups with K-means algorithm and then test the performance on each group of users (resp. OD pairs) based on our online recommendation method and the learned representations. Here we only report the results on BEIJING data, shown in Fig. 5, and, again, the results on SHANGHAI are similar. Observe that the performance of Trans2Vec is stable in different groups of users and OD pairs. Indeed, the standard derivative is less than 1.5% in all our tests using the four metrics. That is, our Trans2Vec framework is robust for different transport mode recommendation scenarios.

Related Work

Transportation recommendation. Unimodal transportation recommendation has been extensively studied in previous work (Chen, Shen, and Zhou 2011; Luo et al. 2013; Bao et al. 2012; Ge et al. 2011). Personalized route planning was first considered in (Rogers and Langley 1998) to improve transportation recommendation quality, and PRP (Funke and Storandt 2015) computes personalized routes on huge road networks in a real-time fashion. A few work (Liu 2011; Borole et al. 2013) discussed generating multi-modal shortest routes based on heterogeneous transportation network and real-time transit data. Although the algorithms mentioned above improve the route recommendation satisfaction, these methods cannot be utilized for multi-modal transportation recommendation.

Network embedding. Network embedding (Cui et al. 2017) has become an emerging topic since the first seminal work DeepWalk (Perozzi, Al-Rfou, and Skiena 2014). DeepWalk is inspired from word representation learning method word2vec (Mikolov et al. 2013), which uses random walks in networks to stimulate sentences in language models. And after that random walks have been a general tool for learning embeddings, such as node2vec (Grover and Leskovec 2016) and metapath2vec (Dong, Chawla, and Swami 2017), to name a few. Almost at the same time as DeepWalk, LINE (Tang et al. 2015) is proposed to learn network embedding via preserving the first and second order pairwise proximities. Finally, network embedding has also found successful applications in various geospatial tasks such as POI recommendations (Feng et al. 2017; Chang et al. 2018), anomaly detection (Hu et al. 2016) and region function profiling (Yao et al. 2018; Fu et al. 2018; Wang et al. 2018a; 2018b).

In this work, we exploit network embedding in a new scenario, i.e., multi-modal transportation recommendation, which has not been studied earlier. In addition, considering the distinct structure of our multi-modal transportation graph, we propose a novel anchor embedding strategy for maintaining transport mode embeddings.

Conclusions

In this paper, we present Trans2Vec, an analytic framework for multi-modal transportation recommendation. From a multi-modal transportation graph perspective, we construct a heterogeneous graph that encodes rich semantics from multi-source data. We then apply joint representation learning on the constructed graph to capture the transport mode preference of both users and OD pairs. In addition, a novel anchor embedding strategy and user and OD relevance were further equipped in the joint representation learning method. Extensive experiments demonstrate the effectiveness of Trans2Vec using real-life datasets.

Acknowledgments

This research is supported in part by grants from the National Natural Science Foundation of China (Grant No.91746301 and No.71531001).

References

Alessandretti, L.; Sapiezynski, P.; Sekara, V.; Lehmann, S.; and Baronchelli, A. 2018. Evidence for a conserved quantity in human mobility. *Nature Human Behaviour* 1.

Bao, T.; Cao, H.; Chen, E.; Tian, J.; and Xiong, H. 2012. An unsupervised approach to modeling personalized contexts of mobile users. *Knowledge and Information Systems* 31(2):345–370.

Borole, N.; Rout, D.; Goel, N.; Vedagiri, P.; and Mathew, T. V. 2013. Multimodal public transit trip planner with real-time transit data. *Procedia-Social and Behavioral Sciences* 104:775–784.

Burges, C. J. 2010. From ranknet to lambdarank to lambdamart: An overview. Technical report.

Chang, B.; Park, Y.; Park, D.; Kim, S.; and Kang, J. 2018. Content-aware hierarchical point-of-interest embedding model for successive poi recommendation. In *IJCAI*, 3301–3307.

Chen, Z.; Shen, H. T.; and Zhou, X. 2011. Discovering popular routes from trajectories. In *Proceedings of the 27th International Conference on Data Engineering*, ICDE, 900–911.

Cui, P.; Wang, X.; Pei, J.; and Zhu, W. 2017. A survey on network embedding. *CoRR* abs/1711.08752.

Dai, J.; Yang, B.; Guo, C.; and Ding, Z. 2015. Personalized route recommendation using big trajectory data. In *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*, 543–554.

Dong, Y.; Chawla, N. V.; and Swami, A. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 135–144.

Feng, S.; Cong, G.; An, B.; and Chee, Y. M. 2017. Poi2vec: Geographical latent representation for predicting future visitors. In *AAAI*, 102–108.

Fu, Y.; Liu, G.; Ge, Y.; Wang, P.; Zhu, H.; Li, C.; and Xiong, H. 2018. Representing urban forms: A collective learning model with heterogeneous human mobility data. *TKDE*.

Funke, S., and Storandt, S. 2015. Personalized route planning in road networks. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 45.

Ge, Y.; Liu, C.; Xiong, H.; and Chen, J. 2011. A taxi business intelligence system. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 735–738.

Grover, A., and Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Hu, R.; Aggarwal, C. C.; Ma, S.; and Huai, J. 2016. An embedding approach to anomaly detection. In *32nd IEEE International Conference on Data Engineering*.

Liu, Y.; Liu, C.; Yuan, N. J.; Duan, L.; Fu, Y.; Xiong, H.; Xu, S.; and Wu, J. 2014. Exploiting heterogeneous human mobility patterns for intelligent bus routing. In *Data Mining (ICDM), 2014 IEEE International Conference on*, 360–369.

Liu, L. 2011. *Data model and algorithms for multimodal route planning with transportation networks*. Ph.D. Dissertation, Technische Universität München.

Luo, W.; Tan, H.; Chen, L.; and Ni, L. M. 2013. Finding time period-based most frequent path in big trajectory data. In *Proceedings of the 2013 ACM SIGMOD international conference on management of data*, 713–724.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014. Deepwalk: online learning of social representations. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Rogers, S., and Langley, P. 1998. Personalized driving route recommendations. In *Proceedings of the American Association of Artificial Intelligence Workshop on Recommender Systems*, 96–100.

Song, C.; Koren, T.; Wang, P.; and Barabási, A.-L. 2010. Modelling the scaling properties of human mobility. *Nature Physics* 6(10):818.

Tang, J.; Qu, M.; Wang, M.; Zhang, M.; Yan, J.; and Mei, Q. 2015. LINE: large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*.

Tang, J.; Qu, M.; and Mei, Q. 2015. Pte: Predictive text embedding through large-scale heterogeneous text networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Wang, P.; Fu, Y.; Zhang, J.; Li, X. L.; and Lin, D. 2018a. Learning urban community structures: A collective embedding perspective with periodic spatial- temporal mobility graphs. *ACM Transactions on Intelligent Systems and Technology (TIST)*.

Wang, P.; Fu, Y.; Zhang, J.; Wang, P.; Zheng, Y.; and Aggarwal, C. 2018b. You are how you drive: Peer and temporal-aware representation learning for driving behavior analysis. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2457–2466.

Yao, Z.; Fu, Y.; Liu, B.; Hu, W.; and Xiong, H. 2018. Representing urban functions through zone embedding with human mobility patterns. In *IJCAI*, 3919–3925.

Yuan, J.; Zheng, Y.; and Xie, X. 2012. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 186–194.