Multi-Modal Transportation Recommendation with Unified Route Representation Learning

Hao Liu[1], Jindong Han[1], Yanjie Fu[2], Jingbo Zhou[1], Xinjiang Lu[1], Hui Xiong[3]
[1]Business Intelligence Lab, Baidu Research, [2]University of Central Florida, [3]Rutgers University
[1]{liuhao30, v_hanjindong, zhoujingbo, luxinjiang}@baidu.com, [2]yanjie.fu@ucf.edu, [3]hxiong@rutgers.edu

ABSTRACT

Multi-modal transportation recommendation aims to provide the most appropriate travel route with various transportation modes according to certain criteria. After analyzing large-scale navigation data, we find that route representations exhibit two patterns: spatio-temporal autocorrelations within transportation networks and the semantic coherence of route sequences. However, there are few studies that consider both patterns when developing multi-modal transportation systems. To this end, in this paper, we study multi-modal transportation recommendation with unified route representation learning by exploiting both spatio-temporal dependencies in transportation networks and the semantic coherence of historical routes. Specifically, we propose to unify both dynamic graph representation learning and hierarchical multi-task learning for multi-modal transportation recommendations. Along this line, we first transform the multi-modal transportation network into time-dependent multi-view transportation graphs and propose a *spatiotemporal graph neural network* module to capture the spatial and temporal autocorrelation. Then, we introduce a *coherent-aware attentive route representation learning* module to project arbitrary-length routes into fixed-length representation vectors, with explicit modeling of route coherence from historical routes. Moreover, we develop a *hierarchical multi-task learning* module to differentiate route representations for different transport modes, and this is guided by the final recommendation feedback as well as multiple auxiliary tasks equipped in different network layers. Extensive experimental results on two large-scale real-world datasets demonstrate the performance of the proposed system outperforms eight baselines.

PVLDB Reference Format:
Hao Liu, Jindong Han, Yanjie Fu, Jingbo Zhou, Xinjiang Lu, Hui Xiong. Multi-Modal Transportation Recommendation with Unified Route Representation Learning. PVLDB, 14(3): 342-350, 2021.
doi:10.14778/3430915.3430924

1 INTRODUCTION

The increasing prevalence of various transport modes (*e.g.*, car, bus, shared-bike, ride-sharing, etc.) and the rapidly expanding transportation networks (*e.g.*, road network, bus network, pedestrian network, etc.) have provided overwhelming alternatives for travelers to reach a destination. In recent years, multi-modal transportation recommendation has become an emerging routing service in many navigation and ride-hailing applications, such as Baidu Maps [31], Here [39], and Didi Chuxing [11]. The target of multi-modal transportation recommendation is to help users find the most appropriate route from one place to another, by jointly considering one or more transport modes on a constrained transportation network. Therefore, accurate and intelligent multi-modal transportation recommendations can significantly help reduce the traveler's decision cost and ultimately improve the user experience.

Existing studies on multi-modal transportation recommendation mainly fall into two categories. (1) *Searching based multi-modal route recommendation* aims to retrieve the shortest path on the transportation network, with a predefined distance metric (*e.g.*, geographical distance, travel time, etc.). Most methods in this category [2, 10] focus on extending graph search algorithms (*e.g.*, Dijkstra, Bellman–Ford and contraction hierarchies [13]) to the multi-modal transportation network [30]. Such approaches are highly dependent on the pre-defined metric and overlook latent factors hidden in the data (*e.g.*, mode and route preferences under different situational contexts [29]). (2) *Learning based transport mode recommendation* has partially addressed the problem by inferring coarse-grained transport mode preferences based on supervised or unsupervised machine learning techniques. A common routine in such methods [29] is to explicitly extract features (*e.g.*, distance, estimated time of arrival (ETA)) from user historical data, such as GPS trajectories [41] and in-app clicks [1]. Such methods make recommendations based on empirically defined features, thus highly rely on the comprehensiveness of feature engineering. More recent studies have applied network embedding [26] and deep learning [46] for transport mode recommendation. However, such methods focus on learning coarse-grained vertex representations (*e.g.*, user and origin-destination pair) or forecasting future travel costs (*e.g.*, ETA), and are not capable of route-specific multi-modal transportation recommendation.

Indeed, the recent emergence of representation learning and multi-task learning techniques provides great potentials to overcome the above limitations. In this paper, we investigate the multi-modal transportation recommendation problem via the unified multi-task route representation learning, by exploiting both spatiotemporal dependencies from transportation networks and the semantic coherence from historical routes. However, three non-trivial challenges arise in achieving this goal. (1) **Spatiotemporal autocorrelation**.The multi-modal transportation network of various transport modes can be abstracted as a dynamic graph (*e.g.*, bus line maybe created or removed, traffic condition is time varying). The dynamic graph contains rich structural and contextual information in both vertices (*e.g.*, the degree, if it has a traffic light) and edges (*e.g.*, distance, ETA, average speed). The first challenge is how to capture the spatial and temporal autocorrelation in the dynamic transportation network. (2) **Route coherence representation**. After studying many routes

Proceedings of the VLDB Endowment, Vol. 14, No. 3 ISSN 2150-8097.
doi:10.14778/3430915.3430924

traveled by users, we identify another important dependency in route representation learning, which we call route coherence. We analogize a route with a sentence, where each hub (*e.g.*, road intersections, bus stations, etc.) and link (*e.g.*, road segments, bus lines, etc.) correspond to a word. In this way, the representation of each hub and link in the route should not only depend on the transportation network but also semantically consistent with the whole route. Besides, the route sequence is of arbitrary length, and the importance of each vertex may vary. How to learn fix-length route representations by incorporating semantic information in historical routes is another challenge. (3) **Transport mode differentiation**. A route in the multi-modal transportation network may be shared or partially shared by various transport modes. For example, given a bicycle route planned by navigation apps, it is with a high probability we can also travel by walk, and vice versa. The last challenge is how to differentiate the unified representation for various transport modes for recommendations.

To tackle the above challenges, we develop a *Hierarchical Multi-Task Route representation Learning* (HMTRL) framework for multi-modal transportation recommendations. Specifically, we first discretize the multi-modal transportation network into a set of graph snapshots over time and construct multi-view graphs, including (1) the *hub-centric graph* which regards transportation hubs as vertices, and (2) the *link-centric graph* which regards transportation links as vertices. After that, we propose the *spatiotemporal graph neural network* module which includes a graph convolution network layer that captures the non-linear spatial autocorrelation from multi-view graphs and a recurrent neural network (RNN) layer that captures the temporal autocorrelation across multiple graph snapshots. Further more, a *coherent-aware attentive route representation learning* module is introduced, including (1) a bi-directional RNN layer that integrates the relatedness of historical routes into the representation of hubs and links, and (2) a self-attentive layer that projects the route sequence into a fixed-length representation with explicit quantifying the contribution of each hub and link. Finally, we propose the *hierarchical multi-task learning* module to learn mode-specific representations, and equip multiple correlated auxiliary tasks in different network layers to guide the optimization of representations for final recommendations. By incorporating structural dependencies in multi-view transportation graphs and route coherence in historical routes under various supervision signals, the mode-specific route representation enables more accurate route-level multi-modal transportation comparison and recommendation.

Our contributions are summarized as follows. (1) We transform the multi-modal transportation network into time-dependent multi-view transportation graphs to characterize the time-evolving structural and contextual information. (2) We propose a spatiotemporal graph neural network to collaboratively capture spatial and temporal autocorrelation for both hubs and links. (3) We introduce a coherent-aware attentive representation learning module to project arbitrary length routes into a fixed-length representation, with explicit route coherence modeling. (4) We propose a hierarchical multi-task learning module to obtain mode-specific route representations and boost the recommendation performance by integrating various auxiliary supervision signals in different network levels. (5) We conduct extensive experiments on two large-scale real-world datasets from one of the world's largest navigation apps. The results demonstrate our model achieves the best performance compared with eight baselines.

2 PRELIMINARIES

2.1 Definitions and Problem Statement

Consider a set of transport modes $\mathcal{M} = \{m_1, m_2, \dots, m_k\}$, where each mode corresponds to a transportation network (*e.g.*, road network, bus line network) that supports vehicle or pedestrian movement. Generally, the transportation network of each transport mode is composed of a set of *hubs* (*e.g.*, road intersection, bus or metro station) and a set of *links* (*e.g.*, road segment, bus line). We formally define the multi-modal transportation network based on transportation networks of each transport mode.

DEFINITION 1. ***Multi-Modal Transportation Network (MMTN)***. *The multi-modal transportation network integrates multiple mode-specific transportation networks into a unified attributed directed graph $\mathcal{G} = (V, E, A^V, A^E, M)$, where V is the set of hubs, E is the set of links, M_n is a mapping function indicates the supported transport modes of each hub and link, A^V and A^E are respectively hub and link features, such as number of bus lines across the hub, spherical distance of the road segment, and ETA of the bus line.*

We use $M(v_i)$ and $M(e_{ij})$ to denote the supported transport modes of each hub $v_i \in V$ and $e_{ij} \in E$. Note each hub and link may support more than one transport modes (*e.g.*, walk and bicycle). We say two hubs are *adjacent* to each other if and only if there is a link connecting them, two links are *adjacent* if and only if a user can transfer from one to another by one hub. Without loss of generality, we constraint a user can only transfer to other links or transport modes in a hub, and a link is the smallest movement unit in the transportation network, *e.g.*, a road segment between two adjacent road intersections, a bus line between two adjacent bus stations.

DEFINITION 2. ***Route***. *A route is a triplet $r_i = \langle H, L, \phi \rangle$, where H is a sequence of adjacent hubs, L is a sequence of adjacent links, and ϕ is a mapping function that indicates the corresponding transport mode of each hub and link in the route.*

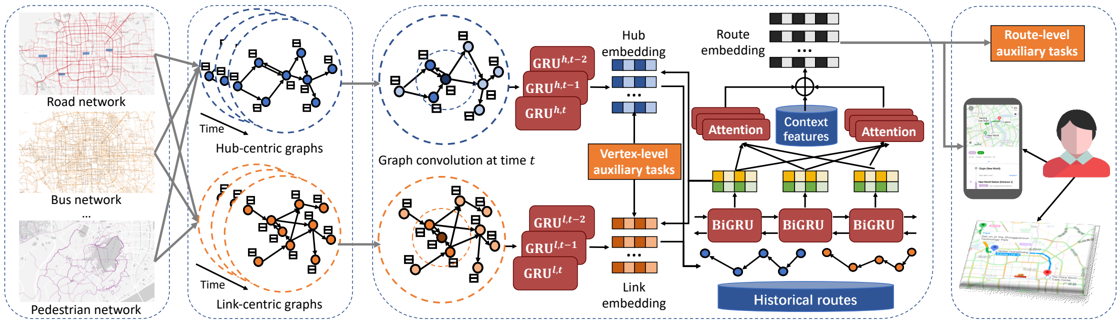
Different from the mapping function M in $\mathcal{G}$, $\phi(v_i)$ and $\phi(e_{ij})$ identify the unique transport mode in the corresponding route. In this work, we restrict a route start and terminate at a hub, and a route may consist of one or more transport modes.

DEFINITION 3. ***Routing Query***. *A routing query is defined as a triplet $q = \langle o, d, t \rangle$, where o and d are origin and destination locations represented by a pair of longitude and latitude, and t is the departure time.*

Since the origin o and the destination d are arbitrary locations, we project them to nearby hubs for recommendation. We say a route r_i is *feasible* for q if the route start from o and terminate at d.

PROBLEM 1. ***Multi-Modal Transportation Recommendation***. *Given a MMTN $\mathcal{G}$, a routing query q and a set of feasible routes Γ for q, our problem is to recommend the most appropriate route $r_i \in \Gamma$ based on the the conditional probability $\hat{y}_i \leftarrow \mathcal{F}(r_i|q, \Gamma, \mathcal{G})$, where $\mathcal{F}$ is the unified mapping function we aim to learn.*

To reduce the computational complexity, we derive the route candidate set Γ (typically less than 20 candidates) based on existing routing engines [27, 29]. To guarantee the utility of recommendations, we restrict the maximum number of mode transfer in each route candidate to three.



(a) Transportation networks (b) Time-dependent multi-view graph construction (c) Hierarchical multi-task route representation learning (d) Multi-modal transportation recommendation

Figure 1: An overview of unified route representation learning for multi-modal transportation recommendation.

2.2 Framework Overview

Figure 1 shows an overview of our approach, where the inputs are the multi-modal transportation network, historical routes, and context features such as weather condition; the output is the recommended route. Overall, there are three tasks in our approach, (1) the construction of time-dependent multi-view transportation graphs, (2) the unified route representation learning, and (3) the hierarchical multi-task learning for mode-specific representation generation and route recommendation. To be specific, in the first task, we transform the multi-modal transportation network to a set of time-dependent multi-view graphs from both the hub-centric perspective and the link-centric perspective. In the second task, the unified route representation is obtained via (1) the joint spatiotemporal autocorrelation modeling of the hub-centric graph and the link-centric graph, and (2) the coherent-aware attentive route representation learning by exploiting historical routes. In the third task, we differentiate representations for various transport modes via multiple implicit tasks and boost the recommendation performance by incorporating multiple related auxiliary tasks in different neural network layers.

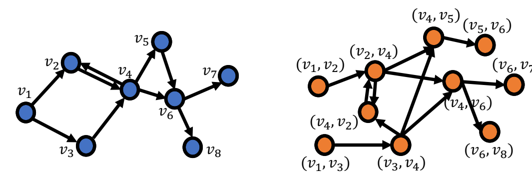
3 CONSTRUCTING TIME-DEPENDENT MULTI-VIEW TRANSPORTATION GRAPHS

We construct time-dependent multi-view transportation graphs to characterize dynamic structural and contextual information in MMTN.

The hub-centric view. The hub-centric graph is a direct mapping of MMTN, where vertices and edges are respectively transportation hubs and links. Specifically, we first discretize the time-evolving graph into a sequence of snapshots, denoted by $\mathcal{G}^h = [\mathcal{G}^{h,t_1}, \mathcal{G}^{h,t_2}, \dots, \mathcal{G}^{h,t_n}]$, where $\mathcal{G}^{h,t_i}$ is the hub-centric graph at time t_i. Figure 2(a) gives an illustrative example snapshot of time-dependent hub-centric graph. For each vertex in the graph, we attach corresponding hub features, including both time-invariant features (*e.g.*, degree, if have a traffic light) and dynamic features (*e.g.*, traffic volume). For two adjacent hubs v_i and v_j, we construct the corresponding adjacency weight based on a Gaussian kernel [25],

$$c_{ij}^h = exp(-\frac{dist(v_i, v_j)^2}{\delta^2}), \tag{1}$$

where $dist(v_i, v_j)$ denote the spherical distance [37] between v_i and v_j, and δ is the standard deviation of spherical distances. In consequence, c_{ij}^h demonstrates the geographical distance distribution among adjacent hubs.



(a) Hub-centric graph (b) Link-centric graph

Figure 2: An example of multi-view transportation graphs.

The link-centric view. The link-centric graph flips vertices and edges in MMTN to preserve structural and contextual information in transportation links. Similar to the hub-centric graph, we discretize the time-evolving link-centric graph into a sequence of snapshots, denoted by $\mathcal{G}^l = [\mathcal{G}^{l,t_1}, \mathcal{G}^{l,t_2}, \dots, \mathcal{G}^{l,t_n}]$, where $\mathcal{G}^{l,t_i}$ indicates the link-centric graph at time t_i. Figure 2(b) shows an illustrative example of the time-dependent link-centric graph at a specific time slice. For each vertex (*i.e.*, link) in the graph, we attach corresponding time-invariant features (*e.g.*, distance, road level) and dynamic features (*e.g.*, average speed, ETA). Consider two links $e_i = (v_1, v_2)$ and $e_j = (v_3, v_4)$, we set the adjacency constraint as

$$c_{ij}^l = \begin{cases} 1, & v_2 = v_3 \ and \ i \neq j \\ 0, & otherwise \end{cases}. \tag{2}$$

There is a directed edge from e_i to e_j if and only if $v_2 = v_3$ in the corresponding MMTN. In other word, an edge in the link-centric graph forms a 2-hop route in the MMTN. Different with the hub-centric graph, $c_{ij}^l \in \{0, 1\}$ preserves the connectivity information.

Mathematically, the link-centric graph is an edge-to-vertex dual of the hub-centric graph at the same time slice. We can re-construct the MMTN from either the hub-centric graph or the link-centric graph [9]. The time-dependent multi-view graphs therefore preserves the temporal dynamics and the structural integrity of MMTN for subsequent graph representation learning.

4 HIERARCHICAL MULTI-TASK ROUTE REPRESENTATION LEARNING

Based on time-dependent multi-view transportation graphs, we obtain mode-specific route representation and make multi-modal transportation recommendations with the following intuitions.

Intuition 1: Graph autocorrelation preservation. The time-dependent multi-view transportation graphs contain rich structural and contextual information that varies over time. The vertex (*i.e.*, hub and link) in each graph at different time slices is both spatially and temporally autocorrelated with other vertices, and can contribute to the overall route representation. Therefore, the model should be able to collaboratively learn the spatial and temporal autocorrelation for both hubs and links.

Intuition 2: Coherent-aware route representation learning. The routes are arbitrary-length sequences and different hubs and links are playing different important roles in different routes. The fixed-length route representation should pay different attention to hubs and links in the sequence to distill salient features of each route. Besides, each hub and link is semantically coherent with its historical routes. Therefore, the representation of hubs and links should reflect a higher relevance with historical routes it involved in.

Intuition 3: Multi-modal route representation differentiation. A transportation hub or link may be shared by various transport modes. Correlating tasks such as link ETA prediction and route preference inference offer potential auxiliary signals to help differentiate mode-specific representations. In consequence, the proposed method should be capable of integrating various auxiliary tasks in different granularity (*e.g.*, vertex level and route level) for route representation differentiation and recommendation.

4.1 Spatiotemporal Autocorrelation Modeling

We first introduce the *spatiotemporal graph neural network* module to capture spatiotemporal autocorrelations based on time-dependent multi-view transportation graphs.

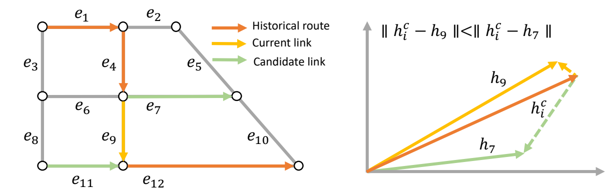
Modeling spatial autocorrelation. We employ Graph Neural Network (GNN) [20] to capture spatial autocorrelation at each time step. By iteratively aggregating and transforming neighbor representations [15], GNN obtains locally smoothed representations where spatially adjacent hubs and links tend to be close in the latent space.

Formally, consider a transportation graph $\mathcal{G}^t$ at time t, let $\mathbf{x}_i$ denotes the d-dimensional representation of vertex $v_i \in \mathcal{G}^t$, we define the graph convolution operation (*GConv*) as

$$\mathbf{x}'_i = GConv(\mathbf{x}_i) = \sigma((\sum_{j \in \mathcal{N}_i} c_{ij}\mathbf{W}_s\mathbf{x}_j) \parallel \mathbf{x}_i), \tag{3}$$

where $\mathbf{x}'_i$ is the updated vertex representation, σ is a non-linear activation function, c_{ij} is the adjacency weight, $\mathbf{W}_s \in \mathcal{R}^{d \times d}$ is learnable weighted matrix shared by all vertices in $\mathcal{G}^t$, $\parallel$ is the concatenation operation, and $\mathcal{N}_i$ is the set of neighbor vertices of v_i in $\mathcal{G}^t$. Note that we can repeat l times graph convolution operations to capture l-hop spatial dependencies. We update representations of $v_i^{p,t} \in \mathcal{G}^{p,t}$ and $v_j^{l,t} \in \mathcal{G}^{l,t}$ by $\mathbf{x}'^{p,t}_i = GConv(\mathbf{x}_i^{p,t})$ and $\mathbf{x}'^{l,t}_j = GConv(\mathbf{x}_j^{l,t})$, respectively.

Modeling temporal autocorrelation. The representations of hubs and links are not only correlated with neighboring vertices in $\mathcal{G}^p$ and $\mathcal{G}^l$, but also influenced by their status in previous time periods. We extend GNN by Gated Recurrent Unit (GRU) [8], a simple yet effective variant of RNN, for temporal autocorrelation modeling. Consider a vertex v_i and its previous T step representations $(\mathbf{x}_i^{t-T}, \mathbf{x}_i^{t-T+1}, \ldots, \mathbf{x}_i^t)$, where $\mathbf{x}_i^t$ is the output of the graph convolution operation at time t. We denote the status of v_i at time step $t-1$ and t as $\mathbf{h}_i^{t-1}$ and $\mathbf{h}_i^t$, respectively. The hidden state $\mathbf{h}_i^t$ reflects both the spatial and temporal autocorrelation of vertex v_i in



(a) Coherence on road network (b) Coherence in vector space

Figure 3: An illustrative example of route coherence modeling.

corresponding time-dependent graphs. For each hub and link in corresponding views, we respectively derive $\mathbf{h}_i^{h,t} = GRU(\mathbf{h}_i^{h,t-1}, \mathbf{x}_i^{h,t})$ and $\mathbf{h}_i^{l,t} = GRU(\mathbf{h}_i^{l,t-1}, \mathbf{x}_i^{l,t})$ for route representation learning.

4.2 Route Representation Learning

Then we present the *coherent-aware attentive route representation learning* module, where (1) the Bi-directional RNN based route coherence modeling block first incorporates route coherence constraints in historical routes to hub and link representations, and (2) the self-attentive route representation learning block further projects arbitrary-length routes into fixed-length representation vectors by automatically learning the importance of each hub and link in the corresponding route.

Bi-directional RNN based route coherence modeling. The insight of route coherence modeling is to incorporate the relatedness of prefix and suffix sub-routes into the current hub and link representations. Figure 3(a) shows an illustrative example of the prefix sub-route coherence on a road network, where the orange arrows form a historical route traveled by a user, the yellow arrow is the current link, and the green arrows are candidate links. Given a historical route $[e_1, e_4, e_9, e_{12}]$, consider e_9 as the current link, there is another candidate link e_7 for prefix sub-route $[e_1, e_4]$ and a candidate link e_{11} for suffix sub-route $[e_{12}]$. Based on the historical route, e_9 is more relevant with the prefix sub-route $[e_1, e_4]$ and the suffix sub-route $[e_{12}]$. Therefore, the representation of e_9 should reflect not only graph dynamics in the MMTN, but also the historical route dependency. We adopt the *Bi-directional GRU* (*BiGRU*) operation to integrate the route coherence dependency into both hub and link representations from both forward direction and backward direction.

Specifically, we reuse the *GRU* operation in Equation (**??**) for hub and link representation update. Formally, for vertex v_i, consider its prefix sub-route $[\cdots, v_{i-2}, v_{i-1}]$ and suffix sub-route $[v_{i+1}, v_{i+2}, \cdots]$, we obtain the forward coherent dependency and backward coherent dependency of v_i by $\overrightarrow{\mathbf{h}}_i^c = GRU(\overrightarrow{\mathbf{h}}_{i-1}, \mathbf{x}_i)$ and $\overleftarrow{\mathbf{h}}_i^c = GRU(\overleftarrow{\mathbf{h}}_{i+1}, \mathbf{x}_i)$, and define the *BiGRU* operation as

$$\mathbf{h}_i^c = BiGRU(\overrightarrow{\mathbf{h}}_i^c, \overleftarrow{\mathbf{h}}_i^c) = \mathbf{W}_c[\overrightarrow{\mathbf{h}}_i^c \parallel \overleftarrow{\mathbf{h}}_i^c], \tag{4}$$

where $\mathbf{W}_c \in \mathcal{R}^{2d \times d}$ is the learnable parameter projecting the concatenated representation to d-dimensional vector. Take Figure 3(b) for example, denote the representation of the historical route as $\mathbf{h}_i^c$, and the representations of e_7 and e_9 as $\mathbf{h}_7$, $\mathbf{h}_9$, the route coherence modeling forces $dist(\mathbf{h}_i^c, \mathbf{h}_9) < dist(\mathbf{h}_i^c, \mathbf{h}_7)$, where $dist(\cdot,\cdot)$ is a distance function in the latent vector space, *e.g.*, the Euclidean Distance. The updated vertex representation incorporates both prefix and suffix sub-route information and is more informative for multi-modal transportation recommendations.

Self-attentive route representation learning. There are still two problems to obtain unified route representation learning: (1) the length of each route may vary, and (2) the importance of each hub and link in the route may be different. Simply averaging representations of hubs and links can not capture the diversified importance of each hub and link, while the RNN in Equation (4) suffers from the gradient vanishing problem [23]. Inspired by the recent success of the attention mechanism [35] on modeling weighted dependencies of long sentences. We analogize multi-modal routes as sentences and employ a self-attention mechanism to transform arbitrary-length routes to fixed-length route representation vectors, with explicit quantifying the importance of both hubs and links in each route.

Given a hub or route sequence of n vertices, we devise K independent self-attentive operations to stabilize the learning process. Specifically, we define the k-th attentive score of v_i as

$$\alpha_{i,k} = \frac{exp(\mathbf{W}_{a,k} tanh(\mathbf{W}_{b,k}\mathbf{h}_i),)}{\sum_{j=1}^{n} exp(\mathbf{W}_{a,k} tanh(\mathbf{W}_{b,k}\mathbf{h}_j),)}, \tag{5}$$

where $\mathbf{h}_i$ and $\mathbf{h}_i$ are representations of v_i and v_j, $\mathbf{W}_{a,k}$ and $\mathbf{W}_{b,k}$ are learnable weights in the k-th attentive operation. Then, we derive the sequence representation by

$$\mathbf{h}' = \|_{k=1}^{K} (\sum_{i=1}^{n} \alpha_{i,k}\mathbf{W}_{r,k}\mathbf{h}_i), \tag{6}$$

where $\|$ is the vector concatenation operation and $\mathbf{W}_{r,k} \in \mathcal{R}^{d\times d}$ is the learnable parameter corresponding to k-th self-attentive operation. Based on Equation (6), we derive the corresponding hub sequence representation $\mathbf{h}^{r,h}$ and link sequence representation $\mathbf{h}^{r,l}$, and derive the unified route representation as

$$\mathbf{h}^r = \mathbf{h}^{r,h} \parallel \mathbf{h}^{r,l}. \tag{7}$$

4.3 Hierarchical Multi-Task Learning

Finally we introduce the *hierarchical multi-task learning* module for multi-modal route representation differentiation and recommendation optimization. By jointly learning multiple related tasks, multi-task learning shares common knowledge in each task and, therefore, improves the generality of the model [44]. Incorporating auxiliary tasks in different granularity has been proved beneficial in many tasks such as document parsing and synonym prediction [12, 16]. In HMTRL, we introduce various auxiliary tasks as complement supervision signals, where different tasks are equipped at different neural network layers.

Specifically, we employ the hard parameter sharing [4] in HMTRL, where different tasks are sharing part of the model but have individual output layers. In HMTRL, the learning tasks can be categorized into two classes, (1) the *Vertex-level MTL* that corresponds to representation learning of vertices (*i.e.*, hubs and links) in time-dependent multi-view graphs, (2) the *Route-level MTL* that corresponds to route representation optimization and recommendation.

Vertex-level MTL. Let $\{\mathcal{T}^{v,i}\}_{i=1}^{\tau_1}$ denote a set of auxiliary vertex tasks, where each task $\mathcal{T}^{v,i}$ corresponds to a set of labels $\{y_j^i\}_{j=1}^{n_i}$ if any, where $y_j^i \in \mathcal{R}$. We first introduce transport mode differentiation tasks to generate mode-specific representations for hubs and links. Specifically, for each transport mode $m_i \in \mathcal{M}$, we define a corresponding task $\mathcal{T}^{m_i}$ to obtain the mode-specific representation after the spatiotemporal graph neural network, $\mathbf{h}_j^{m_i} \leftarrow \mathcal{F}^{m_i}(\mathbf{h}_j)$, where $\mathcal{F}^{m_i}$ is a mode-specific mapping function implemented by a fully connected multi-layer neural network. Note that because not all hubs and links are feasible for all transport modes (*e.g.*, a bus link does not support walk), we mask infeasible transport modes in optimization. Different from other auxiliary tasks, transport mode differentiation tasks do not have direct supervision signals. Instead, all such tasks are optimized based on higher-level task feedbacks (*e.g.*, the link type classification, the route distance prediction and multi-modal transportation recommendation) via back-propagation.

Thereafter, we extract various vertex attributes as supervision signals and facilitate multiple auxiliary task specific layers. Concretely, we integrate regression tasks including the distance prediction and forecasting ETA of the next time step, and integrate classification tasks including hub type (road intersection, bus station, etc.) and link type (road segment, bus line, etc.) inference.

Route-level MTL. Similarly, we define a set of auxiliary route tasks $\{\mathcal{T}^{r,i}\}_{i=1}^{\tau_2}$. Specifically, we first integrate the route coherence modeling task, by leveraging the intermediate prefix sub-route and suffix-route representation derived by Equation (4). Rather than set explicit labels, we optimize the representation in a self-supervised manner, to allow the vertex representation $\mathbf{h}_i$ in the latent space to approximate the corresponding sub-route representation $\mathbf{h}_i^c$ more closely. After that, we incorporate various route related regression tasks, including route distance prediction and future ETA prediction. Besides, for each route r_j, we facilitate the transport mode prediction task by applying a multi-class classifier.

Finally, we define the main recommendation task. Consider the route representation $\mathbf{h}_i^r$, we define the output layer as

$$\hat{y}_i = \sigma(\mathbf{w}_{main}[\mathbf{h}_i^r \parallel \mathbf{x}_i^{context}] + b_{main}), \tag{8}$$

where $\hat{y}_i$ is the estimated travel likelihood of route r_i, σ is a non-linear activation function, $\mathbf{w}_{main}$ are the learnable parameters of the main task, and b_{main} is the bias. Similar to [29], we also concatenate a context vector $\mathbf{x}_i^{context}$ to incorporate the situational context, including features such as weather condition and time periods. To facilitate the readability and reproducibility, we provide detailed settings of auxiliary tasks in the supplementary material.

4.4 Optimization

In HMTRL, we optimize both the main task as well as auxiliary tasks in different layers jointly. For the main task and auxiliary classification tasks, we employ the *cross-entropy loss* for optimization. For regression tasks such as distance and ETA prediction, the objective is to minimize the *mean square error loss*. Please refer to supplementary material for detailed auxiliary losses. Additionally, we introduce the *triplet loss* for the optimization of route coherence,

$$\mathcal{L}_c = -\frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k} max\{(\|\mathbf{h}_i^c - \mathbf{h}_i\|_2 - \|\mathbf{h}_i^c - \mathbf{h}_j\|_2 + \gamma), 0\}, \tag{9}$$

where $\mathbf{h}_j$ is the representation of the negative sample, γ is a margin constant between positive pair $(\mathbf{h}_i^c, \mathbf{h}_i)$ and negative pair $(\mathbf{h}_i^c, \mathbf{h}_j)$. We draw adjacent vertices $v_j \in \mathcal{N}(v_i)$ in the corresponding time-dependent transportation graph as negative samples, and force the representation of the vertex in the route $\mathbf{h}_i$ is closer to the coherent state $\mathbf{h}_i^c$ than negative samples $\mathbf{h}_j$.

Table 1: Statistics of datasets.

Data description	BEIJING	SHANGHAI
# of routing queries & trajectories	2,804,274	2,101,028
# of road intersections	334,421	333,163
# of road segments	420,889	426,247
# of bus lines	22,364	25,652
# of bus stations	9,651	11,587

Overall, we aim to optimize the following objective,

$$\mathcal{L} = \mathcal{L}_{main} + \beta_1 \sum_{i=1}^{\tau_1} \mathcal{L}_i^v + \beta_2 \sum_{i=1}^{\tau_2} \mathcal{L}_i^r, \tag{10}$$

where $\mathcal{L}_i^v$ and $\mathcal{L}_i^r$ are auxiliary vertex and route tasks, β_1 and β_2 are hyper-parameters control the importance of auxiliary tasks. We employ Adam optimizer [19] for training with an exponential decay.

5 EXPERIMENTS

5.1 Data Description

We conduct experiments on two real-world datasets, BEIJING and SHANGHAI. Both datasets are provided by one of the world's largest navigation applications in the world. The datasets include: (1) transportation networks of car, bus, cycle and walk, (2) routing query data extracted from user in-app logs, (3) historical trajectory data collected from user navigation events, (4) context data including user demographic attributes and weather conditions. The raw data of BEIJING and SHANGHAI are 4.13 TB and 4.36 TB, respectively. Both datasets are ranged from August 1, 2019 to October 30, 2019. The Minimum Boundary Rectangle (MBR) of BEIJING and SHANGHAI are (116.21, 39.76), (116.56, 40.03) and (121.35, 31.12), (121.65, 31.38). The statistics of each dataset are summarized in Table 1. We chronologically order each data set, take the first 80% as training set, the following 10% for validation and the rest 10% for testing.

5.2 Implementation Details

We initialize all trainable parameters randomly with the uniform distribution. We apply an embedding operation to project each categorical features to 16-dimensional embedding vectors and concatenate them with continuous features. The dimension of hidden state d is fixed to 64. We stack two layers of graph convolution to capture spatial autocorrelation, and choose LeakyReLU ($\alpha = 0.2$) as the activation function in graph convolution operation. We employ a sigmoid function in the final output layer. The hyper-parameters K, β_1, β_2, T, γ are set to 8, 0.3, 0.1, 3, 0.5, respectively. We set the learning rate $lr = 0.0001$ and the batch size 256. We fix the length of the sub-route to 6 for coherence modeling. We evaluate our model as well as all baselines on a powerful Linux server with 26 Intel Xeon Gold 5117 CPUs, 8 NVIDIA Tesla P40 GPUs, 256GB memory and 10TB disk. For a fair comparison, we carefully fine-tuned the hyper-parameters for all baselines on our datasets via grid search based on settings in their original paper. Please refer to source code[1] for more details.

[1] https://github.com/hanjindong/HMTRL-Pytorch

5.3 Metrics

We employ Hit@k and Normalized Discounted Cumulative Gain (NDCG@k) [17], two widely used metrics in recommenders, to evaluate the recommendation effectiveness. Specifically, we evaluate Hit@1, Hit@3, Hit@5, and NDCG@3, NDCG@5, NDCG@10.

5.4 Baselines

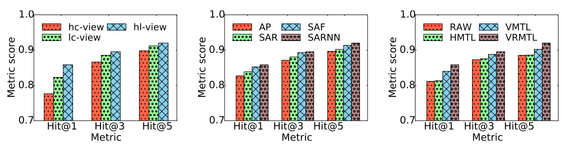
We compare HMTRL with two rule-based methods and six learning methods. **RBT** is a rule-based method that recommends the fastest route, in which we rank route candidates by ETA. **RBD** is another rule-based method that recommends the shortest route, in which we rank route candidates by road network distance. **LR** uses logistic regression [21] for recommendation. The inputs are same as raw features used in HMTRL. **GBDT** adopts the Gradient Boosting Decision Tree for recommendation, which is widely used in both academia and industry. We implement the baseline based on XGboost [7]. The input features are the same as LR. **DeepWalk** [32] is a unsupervised network embedding method that learns vertex representations of a graph. We apply random walks on the MMTN to generate vertex representations, and apply average pooling on route sequences to obtain route representation. We further apply a LR layer for recommendation. **DeepFM** [14] is the state-of-the-art recommendation model that combines the factorization machine and deep neural network to model both first-order and higher-order feature interactions. The input is the same as HMTRL. **Hydra** [29] is a state-of-the-art multi-modal transport mode recommendation method based on multi-sourced urban data. It is fed both handcrafted features as well as pre-trained latent embedding features to a gradient boosting tree-based model. We extend it by adding a regression layer to enable multi-modal route recommendation. **MURAT** [24] is a novel multi-task graph representation learning framework for travel time estimation. We also use our multi-view graphs as the input and devise the the output layer to fit our recommendation task.

5.5 Overall Performance

Table 2 shows the overall performance of our method and all the compared baselines on two datasets with respect to six evaluation metrics. Overall, HMTRL outperforms all the baselines on both datasets using all metrics, which demonstrate the advance of our model. Specifically, HMTRL achieves $(8.3\%, 3.4\%, 2.1\%)$ Hit@k and $(4.7\%, 4.6\%, 4.2\%)$ NDCG@k improvement compared with the state-of-the-art approach (MURAT) on BEIJING. Similarly, the improvement of Hit@k and NDCG@k on SHANGHAI are $(8.1\%, 4.8\%, 2.5\%)$ and $(6.5\%, 4.3\%, 5.1\%)$. Moreover, we can make the following observations. (1) The performance of RBT is much worse than RBD. This observation indicates that travel distance is a more significant indicator than ETA for user trip decision. (2) DeepWalk achieves a better performance than rule-based methods, but performs worse than other learning-based methods. The main reason is that DeepWalk can leverage the structural information but it fails to consider contextual features. Besides, due to its unsupervised property, DeepWalk neglects the user preference signal in historical data. (3) Hydra outperforms all other non-deep learning models by incorporating fine-grained handcrafted features and high-order embedding features. However, compared with deep learning-based methods, including DeepFM and MURAT, the manually extracted features limit the

Table 2: Overall performance comparison using six metrics on BEIJING and SHANGHAI.

Method	BEIJING						SHANGHAI					
	Hit@1	Hit@3	Hit@5	NDCG@3	NDCG@5	NDCG@10	Hit@1	Hit@3	Hit@5	NDCG@3	NDCG@5	NDCG@10
RBT	0.1337	0.3874	0.5794	0.3515	0.4278	0.4713	0.1596	0.4121	0.5609	0.3119	0.3730	0.4569
RBD	0.3647	0.6212	0.7339	0.4801	0.5583	0.6116	0.3178	0.4437	0.6096	0.4068	0.4337	0.5051
LR	0.7188	0.8329	0.8705	0.7864	0.8018	0.8203	0.6687	0.8010	0.8423	0.7468	0.7638	0.7827
GBDT	0.7370	0.8474	0.8851	0.8021	0.8176	0.8341	0.6814	0.8083	0.8524	0.7563	0.7745	0.7950
DeepWalk	0.5213	0.6642	0.7587	0.5955	0.6344	0.6731	0.4916	0.6591	0.7591	0.5886	0.6297	0.6687
DeepFM	0.7658	0.8538	0.8853	0.8166	0.8295	0.8452	0.7068	0.8209	0.8592	0.7743	0.7901	0.8096
Hydra	0.7604	0.8508	0.8827	0.8139	0.8270	0.8434	0.7292	0.8324	0.8713	0.7854	0.8177	0.8251
MURAT	0.7892	0.8654	0.8993	0.8345	0.8467	0.8631	0.7508	0.8415	0.8889	0.8009	0.8297	0.8334
HMTRL	**0.8545**	**0.8946**	**0.9184**	**0.8735**	**0.8856**	**0.8990**	**0.8115**	**0.8823**	**0.9111**	**0.8533**	**0.8652**	**0.8761**



(a) Effect of multi-view graphs. (b) Effect of route representation learning. (c) Effect of hierarchical multi-task learning.

Figure 4: Ablation study of HMTRL on BEIJING.

recommendation capability of the model. (4) MURAT consistently outperforms all other baselines, which demonstrate the effectiveness of multi-task graph representation learning. However, MURAT neglects the information in link-centric graphs as well as the low level supervision signals, therefore performs worse than our approach.

5.6 Ablation Study

Then we conduct ablation study on HMTRL.

Effect of multi-view graphs. We first examine the effectiveness of multi-view graphs by evaluating three variants of HMTRL, (1) *hc-view* only uses the hub-centric graph, (2) *lc-view* uses the link-centric graph only, and (3) *hl-view* uses both graphs for recommendation. As shown in Figure 4(a), the performance of *hl-view* outperforms *hc-view* and *lc-view* by $(9.6\%, 3.4\%, 2.5\%)$ and $(4.4\%, 1.2\%, 0.9\%)$ on (Hit@1, Hit@3 and Hit@5), respectively. Moreover, the *lc-view* performs better than *hc-view*, which demonstrate the structural and contextual information in transportation links plays a more important role for multi-modal transportation recommendation.

Effect of coherent-aware attentive route representation learning. We further construct and evaluate the following variants, (1) *AP* uses average pooling to aggregate hub and link representations, (2) *SAR* derives route representation by self-attention only, (3) *SAF* removes backward GRU in *BiGRU*, and (4) *SARNN* includes both self-attentive operation and the *BiGRU* to integrate route coherence. As shown in Figure 4(b), self-attentive based route aggregation achieves a better performance than *AP*. Moreover, by integrating the route coherence, *SARNN* achieves significant improvement compared with *SAR*. Additionally, compared with *SAF*, we observe *SARNN* achieves consistent improvement by incorporating backward sub-route coherence, demonstrate the effectiveness of bi-directional RNN.

Effect of hierarchical multi-task learning. We compare the following variants, (1) *RAW* directly learns route representation without auxiliary tasks, (2) *HMTL* only incorporates vertex-level hub-related tasks, (3)*VMTL* incorporates both vertex-level hub-related and link-related tasks, and (4) *VRMTL* integrates both vertex-level tasks and route-level tasks. As reported in Figure 4(c), we observe consistent improvement by respectively adding vertex level and route level auxiliary tasks, validate the effectiveness of different supervision signals for multi-modal transportation recommendation. In particular, *VMTL* achieves more significant improvement over *HMTL* than *HMTL* over *RAW*, indicating link related auxiliary tasks plays a more important role in multi-modal transportation recommendations.

5.7 Parameter Sensitivity

We further study the parameter sensitivity of HMTRL. Each time we vary a parameter, we set others to their default values.

First, we vary the dimension d from 32 to 512. The results are reported in Figure 5(a). As the dimension increases, the performance first increases and then remains stable. However, too large d will induce a higher training cost. Therefore, set the dimension to 64 is enough to capture representation information.

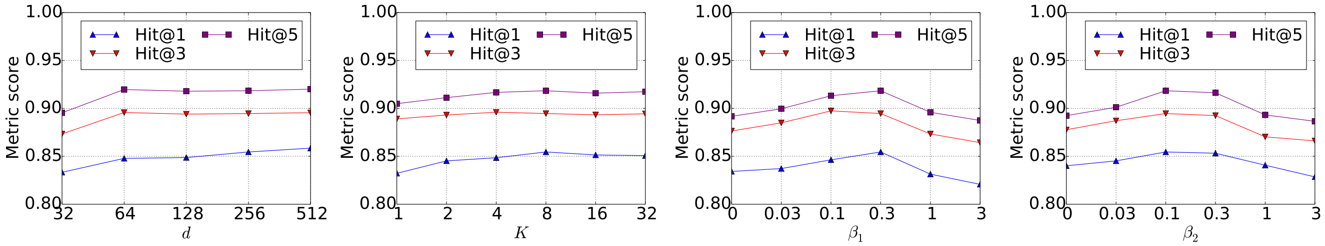
Then, we vary the number of self-attentive operations K from 1 to 32. The results are reported in Figure 5(b). We observe a performance improvement when increasing K from 1 to 8, but a slight performance degradation by further increasing K from 8 to 32. Using 8 self-attentive operations is good enough to capture diversified vertex importance for route representation learning.

After that, we vary vertex-level multi-task weight β_1 from 0 to 3. The results are reported in Figure 5(c). We observe a significant performance gain when increasing β from 0 to 0.3, and then the performance degrades when we further increase β from 0.3 to 3. Above results prove incorporating low-level supervision signals is beneficial to the main recommendation task, but may introduce more noises with too large task weight.

Finally, to test the impact of route-level auxiliary tasks weight, we vary β_2 from 0 to 3. The results are reported in Figure 5(d). HMTRL achieves the best performance when $\beta_2 = 0.1$, and we observe a performance degradation when we increase or decrease β_2. This is possibly because too small β_2 cannot fully take advantage of the common information in route level auxiliary tasks, whereas too large β_2 makes the auxiliary tasks dominate the optimization and weakens the importance of the main recommendation task.

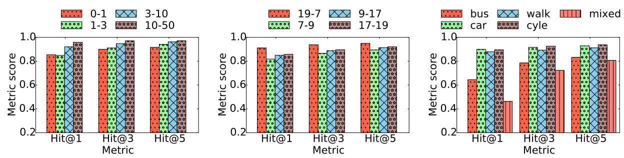
5.8 Robustness Check

A robust transportation recommendation model should perform evenly well in different routing query subgroups. We evaluate the robustness of HMTRL from the following three perspectives. First,



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

Figure 5: Parameter sensitivities on BEIJING.



(a) Group by OD distance. (b) Group by time period. (c) Group by transport mode.

Figure 6: Robustness check on BEIJING.

we group queries by OD pair distance, *i.e.*, less than 1Km, 1Km to 3Km, 3Km to 10Km, and more than 10Km. Second, we split routing queries by day using four time intervals, *i.e.*, the morning peak hour (7,9], the evening peak hour (17,19], and two off-peak intervals (19,7], (9,17]. Third, we group queries based on the selected transport mode, including bus, cycle, walk, car and mixed (*i.e.*, route with more than one transport modes). Figure 6 shows the results of HMTRL on different subgroups on BEIJING. For different OD distance intervals, we observe the performance difference is smaller than 10.9%. Besides, our model performs better on longer distance transportation recommendations, which is perhaps because routes of walk and cycle are no longer attractive for long distance trip, therefore ease the recommendation. For different time periods, we observe the difference is smaller than 10.2%. Besides, we observe a more accurate recommendation result at night, and a worse result in morning rush hour. This is possibly because the traffic condition during the morning rush hour is more complex and hard to predict. For different transport modes, we observe the performance of mixed group is notably lower than others. This may be induced by the scarcity of mixed historical routes, and the route representation with combined transport mode is more sophisticated to learn. Note that the performance of HMTRL on mixed routes still significantly higher than all baselines, please refer to supplementary material for details. The above results suggests us to pay more attention on mixed routes in the further work to obtain a better overall performance.

6 RELATED WORK

Route Recommendation has become a core component in map services (*e.g.*, Google Maps, Baidu Maps) and has gradually received more research attention [18, 40]. With the ubiquity of mobile devices and location-based services, massive historical data (*e.g.*, GPS trajectory data [41] and mobile check-in data [5, 33]) has been leveraged to improve the quality of route recommendation. For example, Chen *et al.* [6] and Wang *et al.* [36] leverage historical trajectories for better routing, but cannot be directly generalized to multi-modal recommendations. Recently, a few machine learning based multi-modal transportation recommendation techniques has been introduced. To name a few, FAVOUR [3] proposed a probabilistic model for multi-modal route recommendation based on a series of user-provided profile and survey data. Trans2vec [26] learns network embedding of users, OD pairs for transport mode recommendation. Hydra [28, 29] constructed various context features and MTRecS-DLT [1] developed a convolutional neural network based model for personalized transport mode recommendation. However, the above studies ignore rich semantic information in the transportation network and historical routes, which lead to unsatisfactory multi-modal route recommendations.

Graph Representation Learning extends the well-known convolutional neural network for capturing spatial dependencies on non-Euclidean graph structures [15, 20]. Recently, graph representation learning has been widely used in many spatiotemporal mining tasks, such as flow prediction [25, 38], region representation [43], and parking availability prediction [42]. Beyond vertex classification, a few studies investigate the classification problem of sequences on dynamic graphs [22]. However, none of the above works are dedicated to multi-modal transportation recommendations.

Multi-Task Learning is a learning paradigm that aims to improve the performance of multiple correlated tasks by sharing common information. Based on information sharing strategy, multi-task learning can be categorized into hard parameter sharing based and soft parameter sharing based [45]. Recent studies [12, 16, 34] have successfully facilitated multiple tasks in lower neural network layers to guide the overall optimization. In this paper, we employ the hierarchical multi-task learning framework by using hard parameter sharing to integrate auxiliary tasks in different network layers.

7 CONCLUSION

In this paper, we proposed HMTRL, a unified route representation learning framework for multi-modal transportation recommendation. We first constructed time-dependent multi-view transportation graphs to characterize the structural and contextual information of both hubs and links. Then, we proposed a spatiotemporal graph neural network for collaborative learning of spatial and temporal autocorrelation. After that, a coherent-aware self-attentive route representation learning module is introduced to project arbitrary-length routes into unified fixed-length route representation vectors, with explicit modeling of route coherence from historical routes. Moreover, a hierarchical multi-task learning module is proposed to derive mode-specific route representations for recommendation by integrating various supervision signals in different network layers. Finally, extensive experimental results on two real-world datasets demonstrated the performance of our approach consistently outperforms eight state-of-the-art baselines.

REFERENCES

[1] Ayat Abedalla, Ali Fadel, Ibraheem Tuffaha, Hani Al-Omari, Mohammad Omari, Malak Abdullah, and Mahmoud Al-Ayyoub. 2019. MTRecS-DLT: Multi-Modal Transport Recommender System using Deep Learning and Tree Models. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. IEEE, 274–278.

[2] Nilesh Borole, Dillip Rout, Nidhi Goel, P Vedagiri, and Tom V Mathew. 2013. Multimodal public transit trip planner with real-time transit data. *Procedia-Social and Behavioral Sciences* 104 (2013), 775–784.

[3] Paolo Campigotto, Christian Rudloff, Maximilian Leodolter, and Dietmar Bauer. 2016. Personalized and situation-aware multimodal route recommendations: the FAVOUR algorithm. *IEEE Transactions on Intelligent Transportation Systems* 18, 1 (2016), 92–102.

[4] R Caruana. 1997. Multitask learning: A knowledge-based source of inductive bias. Machine Learning. (1997).

[5] Dawei Chen, Cheng Soon Ong, and Lexing Xie. 2016. Learning points and routes to recommend trajectories. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. 2227–2232.

[6] Lisi Chen, Shuo Shang, Christian S Jensen, Bin Yao, Zhiwei Zhang, and Ling Shao. 2019. Effective and efficient reuse of past travel behavior for route recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 488–498.

[7] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 785–794.

[8] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).

[9] Dragoš Cvetkovic, Dragoš M Cvetković, Peter Rowlinson, Slobodan Simic, and Slobodan Simić. 2004. *Spectral generalizations of line graphs: On graphs with least eigenvalue-2*. Vol. 314. Cambridge University Press.

[10] Julian Dibbelt et al. 2016. Engineering Algorithms for Route Planning in Multimodal Transportation Networks. *Transportation* (2016).

[11] DiDi. 2020. Retrieved October 27, 2020 from https://en.wikipedia.org/wiki/DiDi

[12] Hongliang Fei, Shulong Tan, and Ping Li. 2019. Hierarchical Multi-Task Word Embedding Learning for Synonym Prediction. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 834–842.

[13] Robert Geisberger, Peter Sanders, Dominik Schultes, and Christian Vetter. 2012. Exact routing in large road networks using contraction hierarchies. *Transportation Science* 46, 3 (2012), 388–404.

[14] Huifeng Guo, Ruiming TANG, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. 1725–1731.

[15] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*. 1024–1034.

[16] Kazuma Hashimoto, Caiming Xiong, Yoshimasa Tsuruoka, and Richard Socher. 2017. A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 1923–1933.

[17] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)* (2002), 422–446.

[18] Evangelos Kanoulas, Yang Du, Tian Xia, and Donghui Zhang. 2006. Finding fastest paths on a road network with speed patterns. In *22nd International Conference on Data Engineering*. 10–10.

[19] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[20] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations*.

[21] David G Kleinbaum, K Dietz, M Gail, Mitchel Klein, and Mitchell Klein. 2002. *Logistic regression*. Springer.

[22] Jia Li, Zhichao Han, Hong Cheng, Jiao Su, Pengyun Wang, Jianfeng Zhang, and Lujia Pan. 2019. Predicting Path Failure In Time-Evolving Graphs. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1279–1289.

[23] Shuai Li, Wanqing Li, Chris Cook, Ce Zhu, and Yanbo Gao. 2018. Independently recurrent neural network (indrnn): Building a longer and deeper rnn. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5457–5466.

[24] Yaguang Li, Kun Fu, Zheng Wang, Cyrus Shahabi, Jieping Ye, and Yan Liu. 2018. Multi-task representation learning for travel time estimation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1695–1704.

[25] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *6th International Conference on Learning Representations*.

[26] Hao Liu, Ting Li, Renjun Hu, Yanjie Fu, Jingjing Gu, and Hui Xiong. 2019. Joint Representation Learning for Multi-Modal Transportation Recommendation. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*. 1036–1043.

[27] Hao Liu, Ying Li, Yanjie Fu, Huaibo Mei, Jingbo Zhou, Xu Ma, and Hui Xiong. 2020. Polestar: An Intelligent, Efficient and National-Wide Public Transportation Routing Engine. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2321–2329.

[28] Hao Liu, Yongxin Tong, Jindong Han, Panpan Zhang, Xinjiang Lu, and Hui Xiong. 2020. Incorporating Multi-Source Urban Data for Personalized and Context-Aware Multi-Modal Transportation Recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2020), 1–1.

[29] Hao Liu, Yongxin Tong, Panpan Zhang, Xinjiang Lu, Jianguo Duan, and Hui Xiong. 2019. Hydra: A Personalized and Context-Aware Multi-Modal Transportation Recommendation System. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2314–2324.

[30] Lu Liu. 2011. *Data model and algorithms for multimodal route planning with transportation networks*. Ph.D. Dissertation. Technische Universität München.

[31] Baidu Maps. 2020. Retrieved October 27, 2020 from https://en.wikipedia.org/wiki/Baidu_Maps

[32] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 701–710.

[33] Samia Shafique and Mohammed Eunus Ali. 2016. Recommending most popular travel path within a region of interest from historical trajectory data. In *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems*. ACM, 2–11.

[34] Anders Søgaard and Yoav Goldberg. 2016. Deep multi-task learning with low level tasks supervised at lower layers. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. 231–235.

[35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.

[36] Jingyuan Wang, Ning Wu, Wayne Xin Zhao, Fanzhang Peng, and Xin Lin. 2019. Empowering A* Search Algorithms with Neural Networks for Personalized Route Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 539–547.

[37] Pengyang Wang, Yanjie Fu, Jiawei Zhang, Pengfei Wang, Yu Zheng, and Charu Aggarwal. 2018. 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 and Data Mining*. 2457–2466.

[38] Yuandong Wang, Hongzhi Yin, Hongxu Chen, Tianyu Wo, Jie Xu, and Kai Zheng. 2019. Origin-Destination Matrix Prediction via Graph Convolution: a New Perspective of Passenger Demand Modeling. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1227–1235.

[39] Here WeGo. 2020. Retrieved October 27, 2020 from https://en.wikipedia.org/wiki/Here_WeGo

[40] Ling-Yin Wei, Yu Zheng, and Wen-Chih Peng. 2012. Constructing popular routes from uncertain trajectories. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 195–203.

[41] Jing Yuan, Yu Zheng, Chengyang Zhang, Wenlei Xie, Xing Xie, Guangzhong Sun, and Yan Huang. 2010. T-drive: driving directions based on taxi trajectories. In *Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems*. 99–108.

[42] Weijia Zhang, Hao Liu, Yanchi Liu, Jingbo Zhou, and Hui Xiong. 2020. Semi-Supervised Hierarchical Recurrent Graph Neural Network for City-Wide Parking Availability Prediction. 1186–1193.

[43] Yunchao Zhang, Yanjie Fu, Pengyang Wang, Xiaolin Li, and Yu Zheng. 2019. Unifying Inter-Region Autocorrelation and Intra-Region Structures for Spatial Embedding via Collective Adversarial Learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1700–1708.

[44] Yu Zhang and Qiang Yang. 2017. A survey on multi-task learning. *arXiv preprint arXiv:1707.08114* (2017).

[45] Yu Zhang and Qiang Yang. 2017. A survey on multi-task learning. *arXiv preprint arXiv:1707.08114* (2017).

[46] Hao Zhou, Yan Zhao, Junhua Fang, Xuanhao Chen, and Kai Zeng. 2019. Hybrid route recommendation with taxi and shared bicycles. *Distributed and Parallel Databases* (2019), 1–21.