A Novel Road Segment Representation Method for Travel Time Estimation

Wei Liu[1], Jiayu He[2], Haiming Wang[1,2], Huaijie Zhu[1,2(✉)], and Jian Yin[1,2]

[1] School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou, China
{liuw259,zhuhuaijie,issjyin}@mail.sysu.edu.cn,
{hejy47,wanghm39}@mail2.sysu.edu.cn
[2] Laboratory of Big Data Analysis and Processing, Guangzhou 510006, China

Abstract. Road segment representation is important for evaluating travel time, route recovery and traffic anomaly detection. Recent works mainly consider topology information of road network based on graph neural network, while dynamic character of topology relationship is usually ignored. Especially, the relationship between road segments is evolving with time elapsing. To obtain road segment representation based on dynamic spatial information, we propose a model named temporal and spatial deep graph infomax network (ST-DGI). It not only captures road topology relationship, but also denotes road segment representation under different time intervals. Meanwhile, the global traffic status/flow will also affect local road segments' traffic situation. Our model would learn the mutual relationship between them, with maximizing mutual information between road segment (local) representation and traffic status/flow (global) representation. Furthermore, it would make road segment representation more distinguishable by this kind of unsupervised learning, and be helpful for downstream application. Extensive experiments are conducted on two important traffic datasets. Compared with the state-of-the-arts models, the experiment results demonstrate the superior effectiveness of our model.

Keywords: Road segment representation · Travel time estimation

1 Introduction

Travel time estimation of a path is an important task in recent years, which is helpful for route planning, ride-sharing, navigation and so on [10,21,23]. Nowadays, almost all the travel service applications have this function, including Google Map, Baidu Map, Uber and Didi. Based on accuracy travel time estimating service, user could obtain road's status, plan personalized trip and avoid wasting time on congested roads. Meanwhile, many researchers have devoted themselves on study of high quality travel time estimation [9,13,22]. However, as travel time estimation is a complex task and many factors need to consider, it is still a challenge to provide accuracy estimation.

C. S. Jensen et al. (Eds.): DASFAA 2021 Workshops, LNCS 12680, pp. 398–413, 2021.
https://doi.org/10.1007/978-3-030-73216-5_27

To achieve accuracy travel time estimation, an effective road segment representation is necessary. Recent works mostly utilized graph-based neural network to learning representation of road segment, for instance graph auto-encoder [9], DeepWalk [13], which would make road segment representation be similar to neighborhoods. There are three drawbacks in these methods:

1) These methods lead to adjacent road segments undistinguishable and will not be beneficial for downstream tasks. Such as in Fig. 1, the adjacent road segments i, j's status would be similar. While, in fact, road segment j would be more special, since it is apt to block up with the flows from neighbor road segments.
2) They only focuses on local feature in graph, and could not learn mutual influence between global traffic condition and each road segment status. Global traffic condition has influence on individual road segment, and some critical road segments' status may also have a great impact on global traffic condition. As in Fig. 1, the traffic flows which come from office areas A, B to residential area C would affect related road segments, making road segment j congested. Meanwhile, road segment j's congestion will also influence other road segments, not just the adjacent road segments. Sometimes, parts of road segments representation at some time intervals are absent, while global (or high-level) traffic condition is rather easily obtained. If we could infer road segment presentation from global traffic condition, it would be beneficial for overcoming data sparsity.
3) They couldn't consider road segment's dynamic status. Under different time interval, road segments will have unique status, which includes not only their dynamic status, but also the special relationship with corresponding global traffic condition. For instance, road segment j would be unblocked in the morning, and the adjacent road segments are also unimpeded. But in the afternoon, j might be congested, and affect the adjacent road segments. Therefore, it requires a model could consider local-global relationship and temporal factors in road network, simultaneously.

Since compared with large volumes of road segments, trajectories are too sparse to denote the distinguishable features of each road segment. Especially, when considering temporal factors, the sparsity problem is much more severe. Meanwhile, traffic system is integrated and complex, the mutual relations between global and local is difficult to model. To solve problems above, we propose a model based on deep graph infomax, which would maximize mutual information between road segment (local) representation and road network (global) representation. It is beneficial for denoising unrelated information and make road segment representation be consistent with entire road network's condition, which could make road segment representation unique and capture similar structures in the whole network. To denote road network's condition, we not only make use of representations from whole graph itself, but also take advantage of geographical traffic status and flows condition. Besides, since road segments' status is dynamic, we would character road segment's representation under different time interval.

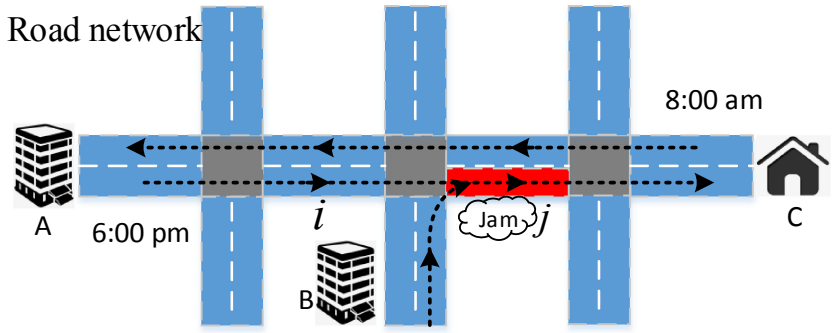


Fig. 1. An example.

In summary, our contribution could be concluded as below:

- A spatial and temporal deep graph infomax network is proposed to learn road segment representation, which could reflect local and global feature of road network simultaneously, especially capture similar structures from whole network view.
- Traffic condition and flows status are both introduced in our model, which make road representation could perceive global condition, section-level global condition with CNN. And dynamic road segment representation is designed to character the temporal traffic condition.
- Extensive experiments are conducted on three public traffic datasets. The results on downstream task (travel time estimation) consistently outperforms the state-of-the-art methods, which proves our road segment representation model is effective and performs more excellent in the sparse traffic setting.

2 Related Work

Traffic forecasting is a popular research topic in recent years. Firstly, we would introduce several main tasks in traffic forecasting and the related works. Especially, we focus the related works in travel time estimation task. Then we would analyze the difference between our work and the most related works.

Since plenty of data are from sensors and GPS, more and more researchers focus on the traffic forecasting to provide convenient service for individuals and traffic management [15,24,26]. In traffic forecasting, there are mainly three kinds of tasks:

1) Travel Time Estimation. The corresponding methods are categorized to two kinds: route-based methods and neighbor- based methods. Route-based methods would map trajectory onto road segments, estimate each segment's travel time and aggregate these time as final estimation. [23] treated travel time estimation as a regression problem and proposed wide-deep-recurrent network

to capture spatial feature of each road segment and sequential feature for estimating the travel time. Beside of GPS traces and road network, [5] utilized smartphone inertial data for customized travel time estimation (CTTE). [20] utilized convolutional neural network to extract spatial feature of road segment and employs recurrent neural network to learn the sequential feature of trajectories. [13] utilized a deep generative model to learn travel time distribution, considering spatial feature of road segments with DeepWalk in road network and temporal feature with the real-time traffic. Different with route-based method, neighbor-based methods utilize neighboring trips with a nearby origin and destination to estimate travel time. [21] found similar trip with the target trip and utilized the travel time of those similar trips to estimate the travel time of target trip.

2) Travel Speed Estimation. Although travel speed estimation is usually related with travel time estimation, they are still different tasks in traffic forecasting. Because travel time estimation contains more factors, such as traffic lights and making left/right turns, than travel speed estimation. In this kind of task, graph neural network and recurrent neural network are generally utilized for extracting topology feature and sequential feature of road network. [28] utilized graph convolutional network to extract spatial topological structure, and gated recurrent unit to capture dynamic variation of road segments' speed distribution. It fuses these spatio-temporal features together to predict traffic speed, it is also similar to [4,14,24,27] which introduced diffusion convolutional recurrent neural network, a deep learning framework for traffic forecasting that incorporates both spatial and temporal dependency in the traffic flow. [3] used history road status (speed) to predict next time's road speed by considering multi-hop adjacent matrix and LSTM. [9] utilized graph convolutional weight completion to learn each road's speed distribution. [11] proposed graph convolutional generative autoencoder to fully address the real-time traffic speed estimation problem.

3) Traffic Flow Prediction. In addition to travel speed, traffic flow is another important sign of traffic condition. To represent the high-level feature of traffic flow, convolutional neural network are usually used. [17,25,26] transformed traffic flow data into a tensor, and utilized a convolutional neural network and residual neural network to extract spatio-temporal feature for urban traffic prediction. [16] employed a sequence-to-sequence architecture to make urban traffic predictions step by step for both of traffic flow and speed prediction. Besides, there are also some other tasks. [15] predicted the readings of a geo-sensor over several future hours by considering multiple sensors' readings, meteorological data, and spatial data. [22] developed a Peer and Temporal-Aware Representation Learning based framework (PTARL) for driving behavior analysis with GPS trajectory data.

In traffic forecasting, road segment representation effect greatly on prediction performance. This paper would focus on high-quality representation of road segment. The most related works with ours is GTT [13], GCWC [9], ST-MetaNet [16]. The differences with them are mainly two points as below: 1) Our model first propose road segment representation based on local and global traffic conditions

simultaneously. Since road network is a complex system, each road segment is not only linked with local adjacent neighborhoods, but also global traffic dynamics. GCWC, ST-MetaNet only considers local features, ignoring global traffic condition. Though GTT considers real-time traffic condition, it learns static road segmentation and global traffic condition independently, fusing them linearly with concatenation operation, which loses sight of the mutual impacts between them. 2) Our model considers the dynamic relationship between road network. In GTT and GCWC, the relationship between road segments is just the topology relationship in road network and changeless. However, in fact it would evolve with time elapsing. Different with ST-MetaNet learning sequential temporal features with gated recurrent unit which needs much denser data sets and more computation costs, we pay attention on periodic and non-uniform temporal features. In summary, our model would not only consider local and global features in traffic simultaneously and mutually, but also model the dynamic status of each road segment. Based on these factors, our model could get a more comprehensive and adaptive road segment representation, which is beneficial for travel time estimation.

3 Proposed Model

Problem Definition: Given a road network $G(V, E)$, historical trajectory dataset $\mathcal{H} = \{T_{(k)}\}_{k=1}^{K}$, our objective is to learn the representation $\mathbf{h}_i^{(t)}$ for each road segment r_i under different time interval t.

We hope the road segment representation $\mathbf{r}_i^{(t)}$ not only could capture the topology relationship between road segment r_i and joint roads under different time interval, but also could denote potential relationship between r_i and global traffic condition. Road segment representation could be used for travel time estimation [13], traffic speed prediction [16] (seen in experiments), route planning and so on.

3.1 Model Overview

Our model would be introduced by four parts. Firstly, we introduce a basic graph constructed by road network, and propose a static road segment representation method based on maximization mutual information (Static Version). Secondly, the temporal factors are considered and fused into the model to obtain dynamic road segment representation (Temporal Version). Thirdly, the traffic status and flows from global view are modeled and mixed into our model (Dynamic Version). Finally, it is the optimization method for our model.

3.2 Static Road Segment Representation

Road segment has complex topology and spatial relations with other road segments. To capture the relations, graph convolutional network (GCN) is usually

adopted. Especially, graph is constructed by adjacent topology relations, spatial information could not be captured by GCN. Since GCN excessively emphasizes proximity information between neighbors, the in-dependence of road segment is easily ignored. In this part, we first introduce a basic version of road segment representation. Then to consider spatial information simultaneously, a grid-based CNN is utilized to represent section which road segment belongs to. Based on the two steps, an optimized version for maximizing mutual information will be introduced.

Feature Initialization. The road network could be represented by a topology *graph*, in which *node* denotes road segment, *edge* between nodes means the link relationship between road segments. Accordingly, we assume a generic graph-based unsupervised machine learning setup: we are provided with a set of road segment features, $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N\}$, where N is the number of road segments in the graph and $\mathbf{r}_i \in \mathbb{R}^F$ represents the features of road segment r_i, where F is the dimension of road segment basic feature. In our datasets, there are totally 14 kinds of road types[1], 7 kinds of lane number (from 1 to 7), 2 kinds of way direction (one-way or bidirectional), so the F would be 23. Directly, we could use one-hot encoding to represent each road segment, while it would miss the detail characteristics of each road segment, such as number of lanes, speed limit, road shape and so on. Also it could consider neighbor Point-of-Interests as road segment's meta information [20], which could be extended in future. Shown in Fig. 2(a), we would use multi-hot encoding to initialize each road segment feature which denotes road segment's special attribute, and reduce dimension with fully-connect layer, similar to the amortization technique [13].

Feature Propogation Based on GCN. Based on road network, we could obtain relational information between these road segments in the form of an adjacency matrix, $\mathbf{A} \in \mathbb{R}^{N \times N}$. In all our experiments we will assume the graphs to be unweighted, i.e. $A_{ij} = 1$ if there exists a connection $r_i \rightarrow r_j$ in the road network and $A_{ij} = 0$ otherwise. Here, the adjacency matrix could be obtained by the road network. However, it is static and could not reflect the road real-time relationship under special time interval. In fact, road relationship will be dynamic with time elapsing. In Sect. 3.3, we would introduce a temporal adjacency matrix $A(t)$ based on historical trajectories.

To learn road segment representation, we would build an encoder, $\mathcal{E} : \mathbb{R}^{N \times F} \times \mathbb{R}^{N \times N} \rightarrow \mathbb{R}^{N \times F'}$, such that $\mathcal{E}(\mathbf{R}, \mathbf{A}) = \mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_N\}$ represents high-level representation $\mathbf{h}_i \in \mathbb{R}^{F'}$ for road segment r_i . These representations may then be retrieved and used for travel time estimation task, speed prediction and so on. The definition of function $\mathcal{E}$ could be seen as following:

$$\mathcal{E}(\mathbf{R}, \mathbf{A}) = \sigma(\hat{\mathbf{D}}^{-\frac{1}{2}}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-\frac{1}{2}}\mathbf{R}\mathbf{W}) \tag{1}$$

[1] i.e., living street, motorway, motorway link, primary, primary link, residential, secondary, secondary link, service, tertiary, tertiary link, trunk, trunk link, unclassifie.

where $\mathbf{A} = \mathbf{A} + \mathbf{I}_N$ is the adjacent matrix with added self-connections, $\mathbf{I}_N$ is the identity matrix, $\hat{\mathbf{D}}$ is the degree matrix, $\hat{\mathbf{D}} = \sum_j \hat{\mathbf{A}}_{ij}$. σ is a nonlinear activation function. $\mathbf{W} \in \mathbb{R}^{F \times F'}$ is a learnable linear transformation applied to every node.

Here we will focus on graph convolutional encoders – a flexible class of node embedding architectures, which generate road segment representations by repeated aggregation over local road neighborhoods [6]. A key consequence is that the produced road embeddings, $\mathbf{h}_i$, summarize a local patch of the graph centered around road segment r_i rather than just the road segment itself. In what follows, we will often refer to $\mathbf{h}_i$ as local representation to emphasize this point.

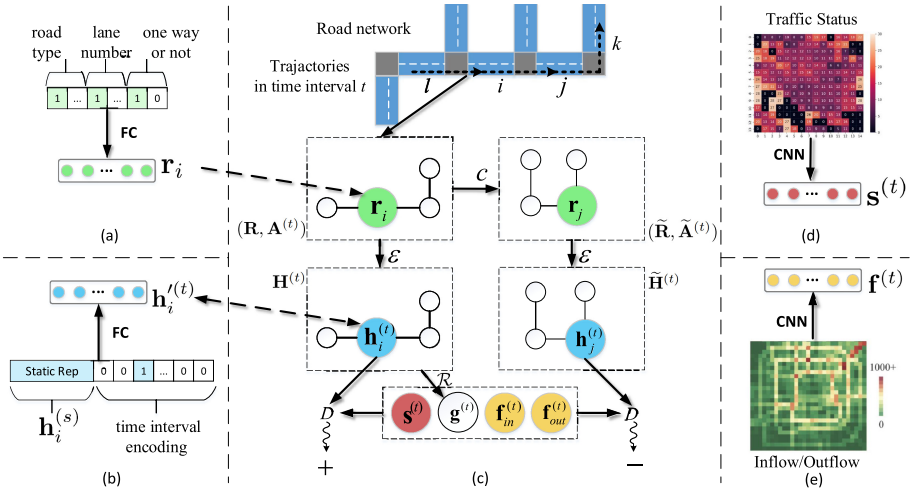


Fig. 2. Framework of our model. Road network (and trajectories) could be transformed to node representation. Based on GCN, each node obtains a local representation. A corrupted graph is also generated. Based on global feature, a discriminator is utilized to distinguish the real or fake local feature from different graph.

Spatial Region Feature Construction. Since graph can only capture topology structures, the spatial information is easily lost. We utilize grid-index of road segments to construct spatial relations between them. Each grid is denoted by a multi-hot embedding, $\mathbf{p} \in \mathbb{R}^N$. If a road segment crosses the grid, the corresponding element of embedding is set 1, and 0 otherwise. Multiplying with road segment representation matrix $\mathbf{R}$, grid representation is achieved $\mathbf{q} = \mathbf{pR}$, which would subsume representations of inner road segments.

Since road segment may cross several grids, we also consider adjacent regions to include the spatial related road segments with target road segment as much as possible. Figure 3 shows region-level embedding for target road segment.

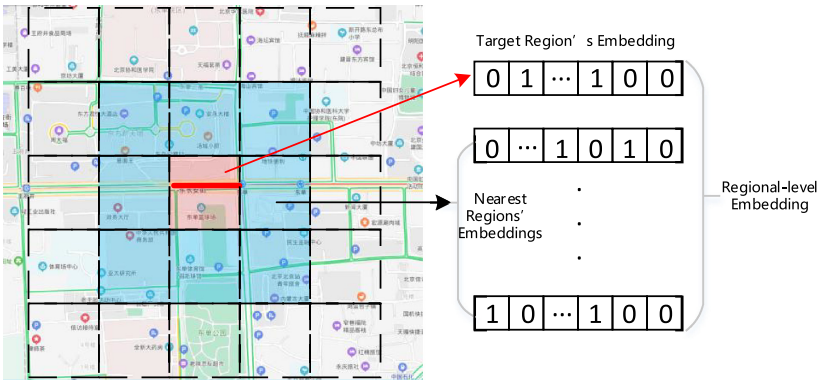


Fig. 3. Capturing region-level representation

Feature Optimization with Mutual Information. Since road network is an entire systems, a road segment is not only related to adjacent neighborhoods, but also the whole traffic system. The models based on only GCN are insufficient for road segment representation. Therefore, we need to obtain road segment (i.e., local) representations that capture the global information of the entire road network. As the general graph auto encoder (GAE) [2] could not realize this object, which directly optimizes the Euclidean Distance (or discrepancy) between input and output, we adopt maximizing local-global mutual information [8,19] between road segment (local) representation and entire road network (global) representation, which could make road segment representation not only unique but also containing global feature. To achieve this objective, there are 4 questions to solve:

1) How to represent the feature of road segment?
2) How to represent the feature of entire road network information?
3) How to compute the local-global mutual information?
4) How to maximize the local-global mutual information?

For **question 1)**, we would follow the previous method to represent road segment as $\mathbf{h}_i$. Next, for **question 2)**, in order to obtain the global-level road network features, $\mathbf{g}$, we leverage a readout function, $\mathcal{R} : \mathbb{R}^{N \times F} \rightarrow \mathbb{R}^F$, and use it to gather the obtained local (road segment) representations into a global-level representation; i.e., $\mathbf{g} = \mathcal{R}(\mathcal{E}(\mathbf{R}, \mathbf{A}))$. A simple but efficient choice of $\mathcal{R}$ is average function, $\mathcal{R}(\mathbf{H}) = \sigma(\frac{1}{N}\sum_{i=1}^{N} \mathbf{h}_i)$, where σ is the activation function.

$$I(H;G) = H(H) - H(H|G), \tag{2}$$

$$= \int_{\mathcal{H}\times\mathcal{G}} log \frac{d\mathbb{P}_{HG}}{d\mathbb{P}_H \otimes \mathbb{P}_G} d\mathbb{P}_{HG}, \tag{3}$$

$$= D_{KL}(\mathbb{P}_{HG} || \mathbb{P}_H \otimes \mathbb{P}_G) \tag{4}$$

For **question 3)**, the computation of local-global mutual information is shown in Eq. 4, where $\mathbb{P}_{HG}$ is the joint distribution of two variables, $\mathbb{P}_H \otimes \mathbb{P}_G$ is the product of margins, D_{KL} is the KL divergency between two distributions.

$$I(H;G) \leq \mathbb{E}_{\mathbb{P}_{HG}}[\mathcal{D}_\theta(\mathbf{h},\mathbf{g})] - log(\mathbb{E}_{\mathbb{P}_H} \otimes [e^{\mathcal{D}_\theta(\mathbf{h},\mathbf{g})}]) \tag{5}$$

For **question 4)**, to maximize the local-global mutual information, the mutual information (ML) in KL divergency form could be transformed to Donsker-Varadhan (DV) representation as dual representations [1] in Eq. 5. A discriminator is employed to make DV representation to approximate MI, $\mathcal{D} : \mathbb{R}^F \times \mathbb{R}^F \to \mathbb{R}$, such that $\mathcal{D}(\mathbf{h},\mathbf{g})$ represents the probability scores assigned to this local-global pair. The training of discriminator and maximization MI would be processed simultaneously. Here contrastive method is adopt [7,12,18], which is to train the discriminator $\mathcal{D}$ to score contrastively between local representations (positive examples) that contain features of the whole and those undesirable local representations (negative examples). The discriminator is defined as following.

$$\mathcal{D}(\mathbf{h},\mathbf{g}) = \sigma(\mathbf{h}^T\mathbf{W}_2\mathbf{g}) \tag{6}$$

where $\mathbf{W}_2 \in \mathbb{R}^{F' \times F'}$ is a learnable linear transformation applied to every node. Therefore, based on approximately monotonic relationship between Jensen-Shannon divergence and mutual information, a noise-contrastive type objective [8] is formularized with a standard binary cross-entropy (BCE) loss between the samples from the joint (positive examples) and the product of marginals (negative examples) for maximizing mutual information, as following equation:

$$\begin{aligned}\mathcal{L} = \frac{1}{N+M}(&\sum_{i=1}^{N} \mathbb{E}_{(\mathbf{R},\mathbf{A})})[log\mathcal{D}(\mathbf{h}_i,\mathbf{g})] + \\ &\sum_{j=1}^{M} \mathbb{E}_{(\tilde{\mathbf{R}},\tilde{\mathbf{A}})}[log(1-\mathcal{D}(\tilde{\mathbf{h}}_j,\mathbf{g}))])\end{aligned} \tag{7}$$

This approach effectively maximizes mutual information between $\mathbf{h}_i$ and $\mathbf{g}$, based on the Jensen-Shannon divergence between the joint and the product of marginals.

Negative samples for $\mathcal{D}$ are provided by pairing the global $\mathbf{g}$ from $(\mathbf{R},\mathbf{A})$ with local representation $\tilde{h}_j$ of an alternative graph, $(\tilde{\mathbf{R}},\tilde{\mathbf{A}})$, where $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{A}}$ are corruption versions of original data, respectively. For the road network graph, an explicit (stochastic) corruption function, $\mathcal{C} : \mathbb{R}^{N\times F} \times \mathbb{R}^{N\times N} \to \mathbb{R}^{M\times F} \times \mathbb{R}^{M\times M}$ (M is the node number of corruption graph) is required to obtain a negative example from the original graph, i.e., $(\tilde{\mathbf{R}},\tilde{\mathbf{A}}) = \mathcal{C}(R,A)$. Corruption function $\mathcal{C}$ could be feature-based by row-wise shuffling feature matrices or adjacent-matrix-based by changing part of adjacent matrix elements, which would be discussed in experiments. The choice of the negative sampling procedure will govern the specific kinds of structural information that is desirable to be captured as a byproduct of this maximization.

As all of the derived local representations are driven to preserve mutual information with the global road network representation, this allows for discovering and preserving dependency on the local-level – for example, distant roads with related structural roles. For road network and traffic condition, our aim is for the road segment to establish link to related road segment across the road network, rather than enforcing the global representation to contain all of these correlations.

3.3 Temporal Adjacent Matrix

As mentioned before, road segment's status is not static. At different time intervals, road segment's status will be different. The same to relationship between adjacent road segments. To model the dynamic road segment representation, we propose to construct temporal adjacent matrix $A^{(t)}$, which is based on trajectories $T^{(t)} = \{tr_1, tr_2, ...tr_{|T_t|}\}$ in corresponding time interval t. For instance, $t = [t_s, t_e)$, where t_s means the start of the time interval, t_e means the end of the time interval. Based on the temporal adjacent matrix $A^{(t)}$ and encoder $\mathcal{E}$, dynamic road segment representation $\mathbf{h}_i^{(t)}$ could be obtained.

However, if each road segment has an independent representation at each time interval, it would cause overfit and need large storage memory. Moreover, not all road segments have trajectories at each time interval, a.k.a. data sparsity problem. Here, we utilize *amortization* technique, which makes road segment r_i's static representation $\mathbf{h}_i^{(s)}$ and one-hot encoding of time interval $\mathbf{e}_t$ mapping into a low-dimensional representation by fully-connected layer to denote dynamic road segment representation $\mathbf{h}_i'^{(t)}$, shown in Eq. 8. To learn the parameters in the fully-connected layer, with $\mathbf{h}_i^{(t)}$ obtained by temporal adjacent matrix and the encoder $\mathcal{E}$ denoted in Eq. 9, we utilize L_2 to minimize the error between $\mathbf{h}_i'^{(t)}$ and $\mathbf{h}^{(t)}$, which would optimize parameters in the fully-connected layer. Based on the fully-connected layer, we could get road segment's dynamic representation.

$$\mathbf{h}_i'^{(t)} = \sigma(w(\mathbf{h}_i^{(s)} \oplus t) + b) \tag{8}$$

$$\mathbf{H}^{(t)} = \mathcal{E}(\mathbf{R}, \mathbf{A}^{(t)}) \tag{9}$$

$$\mathcal{L} = min||\mathbf{h}_i^{(t)}, \mathbf{h}_i'^{(t)}||_2^2 \tag{10}$$

3.4 Global Traffic Condition and Flows

To optimize the mutual information between local and global features in traffic graph, we utilize GCN to extract each road segment representation as the local feature. And the global feature is denoted by the average of all the road segment representation. As in Sect. 3.3, the local feature could be dynamic. Directly, we could also update the global feature with the temporal local feature. However, the global feature is insufficient by aggregating local features, which ignores

important traffic information, especially traffic real-time status $s^{(t)}$ and flows $f^{(t)}$ under time interval t. Here, we utilize two vectors $\mathbf{s}^{(t)} \in \mathbb{R}^F$, $\mathbf{f}_{in/out}^{(t)} \in \mathbb{R}^F$ to denote them, respectively. Therefore, we could update the global feature $\mathbf{g}^{(t)}$ by unifying these traffic feature with $\mathbf{g}'^{(t)}$. There would be several methods to mix these features, the detail would be discussed in the experiment part.

$$\mathbf{s}^{(t)} = CNN(\mathbf{S}^{(t)}) \tag{11}$$

To calculate traffic real-time condition, we could split geographical space into grids. And we calculate the average speed of trajectories in each grid under time interval t to denote the traffic condition $\mathbf{S}^{(t)} \in \mathbb{R}^{P \times Q}$, where P, Q are the girds number in rows and columns, respectively. Then convolutional neural network would be utilized to extract the high-level traffic condition as shown in Fig. 2(d). To represent traffic flow, we could count each grids's flow-in and flow-out times, which would construct matrix $\mathbf{F}_{in}^{(t)}, \mathbf{F}_{out}^{(t)} \in \mathbb{R}^{P \times Q}$ to denote the traffic flow separately, as depicted in Fig. 2(e). The process of extracting high-level feature of traffic flow is similar to traffic real-time condition's. To avoid utilizing the future data in prediction, we could not directly utilize the current traffic data. Here, we adopt traffic condition $\mathbf{s}^{(t-1)}$ under time interval $t-1$ as the global traffic condition. In future, we could utilize LSTM (Long Short Term Memory) to predict current traffic condition $\mathbf{s}^{(t)}$ as global traffic condition by the history traffic condition $\{\mathbf{s}^{(t-k)}, ...\mathbf{s}^{(t-1)}\}$.

Optimization. Assuming the single-graph setup (i.e., $(\mathbf{R}, \mathbf{A}^{(t)})$) provided as input), we will now summarize the steps of the model optimization procedure:

1. Sample a negative example by using the corruption function: $(\tilde{\mathbf{R}}, \tilde{\mathbf{A}}^{(t)}) \sim \mathcal{C}(\mathbf{R}, \mathbf{A}^{(t)})$.
2. Obtain local representations, $\mathbf{h}_i^{(t)}$ for the input graph by passing it through the encoder: $\mathbf{H}^{(t)} = \mathcal{E}(\mathbf{R}, \mathbf{A}^{(t)}) = \{\mathbf{h}_1^{(t)}, \mathbf{h}_2^{(t)}, ..., \mathbf{h}_N^{(t)}\}$.
3. Obtain local representations, $\mathbf{h}_j^{(t)}$ for the negative example by passing it through the encoder: $\tilde{\mathbf{H}}^{(t)} = \mathcal{E}(\tilde{\mathbf{R}}, \tilde{\mathbf{A}}^{(t)}) = \{\tilde{\mathbf{h}}_1^{(t)}, \tilde{\mathbf{h}}_2^{(t)}, ..., \tilde{\mathbf{h}}_M^{(t)}\}$.
4. Summarize the input graph by passing its local representation through the readout function: $\mathbf{g}^{(t)} = \mathcal{R}(\mathbf{H}^{(t)})$, and form $\mathbf{g}'^{(t)}$ by unifying global traffic condition with $\mathbf{g}^{(t)}$.
5. Update parameters of $\mathcal{E}$, $\mathcal{R}$ and $\mathcal{D}$ by applying gradient descent to maximize Eq. 7.

This algorithm is fully depicted in Fig. 2(c).

4 Experiments

In this section, we will conduct extensive experiments to evaluate the effectiveness of our model on travel time estimation task. Firstly, we described three large

scale datasets. Secondly, the evaluation metrics are introduced. Thirdly, the compared state-of-the-art models would be presented. Next, the results compared with baselines would be analyzed. Then, we study the effectiveness of hyperparameters and each part in our model.

Task Description: Travel Time Estimation, based on each road segment's representation $\mathbf{h}_i^{(t)}$ in trajectory tr, we could initialize road segment representation of travel time estimation model, e.g., DeepGTT. After training DeepGTT, we could infer each road segment's travel speed $v_i^{(t)}$ from road segment's representation $\mathbf{h}_i^{(t)}$. Considering each road segment's distance l_i as a weight, the average travel speed is obtained as following:

$$v_r = \sum_{i=1}^{n} w_i v_i, w_i = \frac{l_i}{\sum_{k=1}^{n} l_k} \tag{12}$$

then the travel time could be achieved as below:

$$t = \frac{\sum_i^n l_i}{v_r} \tag{13}$$

Datasets. We conduct experiments on two public datasets from Didi[2] and one dataset from Harbin [13].

- Chengdu Dataset: It consists of 8,048,835 trajectories (2.07 billion GPS records) of 1,240,496 Didi Express cars in Oct 2018 in Chengdu, China. The shortest trajectory contains only 31 GPS records (0.56 km), the longest trajectory contains 18,479 GPS records (96.85 km), the average of GPS records (distance) in trajectory is 257 (4.56 km).
- Xi'an Dataset: It consists of 4,607,981 trajectories (1.4 billion GPS records) of 728,862 Didi Express cars in Oct 2018 in Xi'an, China. The shortest trajectory contains only 31 GPS records (0.27 km), the longest trajectory contains 16,326 GPS records (166.12 km), the average of GPS records (distance) in trajectory is 316 (4.96 km).
- Harbin Dataset: It consists of 517,857 trajectories (the sampling time interval between two consecutive points is around 30 s) of 13,000 taxis cars during 5 days in Harbin, China. The shortest trajectory contains only 15 GPS records (1.2 km), the longest trajectory contains 125 GPS records (60.0 km), the average of GPS records (distance) in trajectory is 43 (11.4 km).

In all datasets each trajectory is associated with the timestamp and driverID. For the first two dataset, we set trajectories in the first 18 days as the training set, trajectories in last 7 days as the testing set and the rest of trajectories as the validation set. For Harbin dataset, we set trajectories in the first 3 days as the training set, trajectories in last day as the testing set and the rest of trajectories as the validation set. We adopt Adam optimization algorithm to train

[2] http://bit.ly/366rlXf.

the parameters. The learning rate of Adam is 0.001 and the batch size during training is 128. Our model is implemented with PyTorch 1.0. We train/evaluate our model on the server with one NVIDIA RTX2080 GPU and 32 CPU (2620v4) cores.

Baselines. To prove the effectiveness of our model on travel time estimation task, we first compare our model with several baseline methods, including:

- DeepTTE: It treats road segments as tokens and compress a sequence of tokens (a route) to predict travel time by using RNN [20].
- ST-MetaNet: It employs a sequence-to-sequence architecture to make prediction step by step, in which it contains a recurrent graph attention network to capture diverse spatial correlations and temporal correlations [16].
- GCWC: It proposed a graph convolutional weight completion to fill each road's time-varying speed distribution, which would be helpful for travel time representation [9].
- DeepGTT: It utilizes a deep generative model to learn the travel time distribution for any route by conditioning on the real-time traffic [13].

Evaluation Metrics. To estimate the performance of different models on the prediction task, we adopt RMSE and MAE,

$$RMSE(t,\hat{t}) = \sqrt{\frac{1}{|t|}||t-\hat{t}||_2^2}, MAE(t,\hat{t}) = \frac{1}{|t|}||t-\hat{t}||_1 \tag{14}$$

where t, $\hat{t}$ denote ground truth and estimated value, respectively.

Performance. Compared with the state-of-the-art models, our model outperforms other models at least 5.0%, for two reasons: 1) Our model makes the road representation consider local spatial relationship and global traffic status, simultaneously, rather than considering only local spatial relationship as in DeepTTE, ST-MetaNet and DeepGTT. 2) Our model learns road segment's temporal representation under different time interval, which would denote the dynamic spatial relationship under a certain time interval, which is also ignored by other models (Table 1).

Hyperparameters and Ablation. To demonstrate the performance of our model, we tested our model under different hyperparamters on Harbin dataset, including the dimension of features $F' \in \{40, 80, 120, 160, 200\}$, the time interval $|t| \in \{10, 15, 20, 30, 60\}$ min, as shown in Fig. 4.

Dimension: From the values of RMSE and MAE of our model on Harbin dataset, we could find as dimension increase, the model's error become less and accuracy become increasing. Especially, when dimension equals 200, the RMSE/MAE

Table 1. Performance results on travel time estimation

	Chengdu		Xi'an		Harbin	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
DeepTTE	356.74	248.04	344.69	250.15	330.65	239.67
ST-MetaNet	274.90	182.89	266.72	193.78	254.96	184.97
GCWC	280.72	195.43	270.15	200.12	264.92	193.93
DeepGTT	236.44	165.36	228.46	166.77	219.10	159.11
ST-DGI	**222.92**	**154.45**	**216.33**	**156.64**	**208.43**	**150.53**
ST-DGI/S	229.27	160.56	222.82	162.15	213.44	154.61
ST-DGI/T	231.34	161.59	224.53	163.85	215.37	158.02
ST-DGI/G	228.73	159.32	221.12	161.93	212.28	153.70

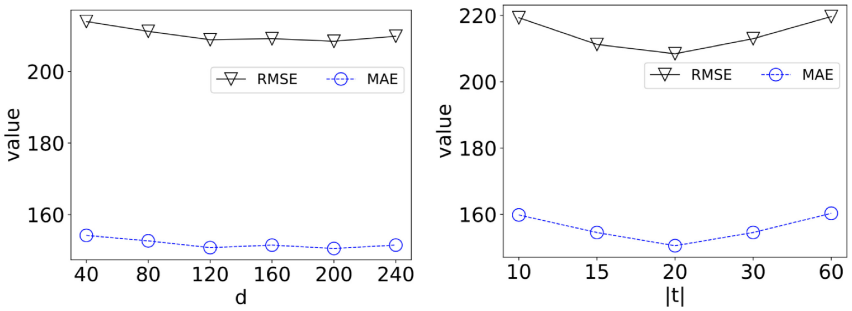
both is least, our model achieves best performance. Therefore, in the experiments we set the dimension of features in our model as 200.

Time Interval: We also test the value of time interval's effect to our model. Compared with dimension, we could find time interval's change has bigger influence to model's performance. When time interval equals 20 min, our model has the best performance. When time interval become smaller, the performance decreases. It may be caused by the data sparsity problem, in which there are too few trajectories to learn temporal road segment representation.

To prove the effectiveness of each part in our model, we estimate our model with its variants:

- ST-DGI/S: It is a variant of our model without the statistical information of each road segments. Here, we adopt stochastic initialization of road segment representation.
- ST-DGI/T: It is a variant of our model without considering temporal factor. Here, we adopt the same representation for a road segment at different time intervals.
- ST-DGI/G: It is a variant of our model without considering global traffic status and traffic flows. It just utilizes summarized representation of nodes in road network graph.

From Table 1, we could find each component is beneficial for our model. Because without considering any part of our model, the RMSE/MAE of the variants both become larger. Meanwhile, we could find the variant without temporal factor, our model's performance decreases most. It denotes the temporal factor is very important for road segment representation and travel time estimation. Besides, as without considering global traffic status and the statistical information of road segment, it proves the two factors also play an important role in travel time estimation.



(a) vary dimension (b) vary time interval

Fig. 4. Effect of two hyper parameters: dimension and time interval.

5 Conclusion

Both road's statistical information and global traffic factors and road segment's dynamic status with time changing. A mutual information loss function is utilized to learn road segment representation, which could make the road segment representation characterize global traffic status at special time interval. The experiments's performance on travel time estimation demonstrates the effectiveness of our road representation method, compared with the state-of-the-arts. In future, we would extend the application of our work on other traffic prediction tasks, such as traffic speed prediction, route planning and so on.

Acknowledgment. This work is supported by the National Natural Sci- ence Foundation of China (61902438, 61902439, U1811264, U19112031), Natu- ral Science Foundation of Guangdong Province under Grant (2019A1515011704, 2019A1515011159), National Science Foundation for Post-Doctoral Scientists of China under Grant (2018M643307, 2019M663237), and Young Teacher Training Project of Sun Yat-sen University under Grant (19lgpy214,19lgpy223).

References

1. Belghazi, M.I., et al.: Mutual information neural estimation. In: ICML, p. 530–539 (2018)
2. Berg, R.V.D., Kipf, T.N., Welling, M.: Graph convolutional matrix completion. CoRR abs 1706.02263 (2017)
3. Cui, Z., Henrickson, K., Ke, R., Wang, Y.: Traffic graph convolutional recurrent neural network: a deep learning framework for network-scale traffic learning and forecasting. IEEE TITS **21**, 4883–4894 (2019)
4. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: NeurIPS, pp. 3844–3852 (2016)
5. Gao, R., et al.: Aggressive driving saves more time? Multi-task learning for customized travel time estimation. In: IJCAI, pp. 1689–1696 (2019)
6. Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahla, G.E.: Neural message passing for quantum chemistry. arXiv:1704.01212 (2017)

7. Grover, A., Leskovec, J.: node2vec: scalable feature learning for networks. In: KDD, p. 855–864 (2016)
8. Hjelm, R.D., et al.: Learning deep representations by mutual information estimation and maximizationg. In: ICLR (2019)
9. Hu, J., Guo, C., Yang, B., Jensen, C.S.: Stochastic weight completion for road networks using graph convolutional networks. In: ICDE, pp. 1274–1285 (2019)
10. Ide, T., Sugiyama, M.: Trajectory regression on road networks. In: AAAI (2011)
11. Jian, J., Yu, Q., Gu, J.: Real-time traffic speed estimation with graph convolutional generative autoencoder. IEEE TITS **20**, 3940–3951 (2019)
12. Kipf, T.N., Welling, M.: Variational graph auto-encoders. CoRR abs/1611.07308 (2016)
13. Li, X., Cong, G., Sun, A., Cheng, Y.: Learning travel time distributions with deep generative model. In: WWW, pp. 1017–1027 (2019)
14. Li, Y., Yu, R., Shahabi, C., Liu, Y.: Diffusion convolutional recurrent neural network: data-driven traffic forecastings. In: ICLR (Poster) (2018)
15. Liang, Y., Ke, S., Zhang, J., Yi, X., Zheng, Y.: GeoMAN: multi-level attention networks for geo-sensory time series prediction. In: IJCAI, pp. 3428–3434 (2018)
16. Pan, Z., Liang, Y., Wang, W., Yu, Y., Zheng, Y., Zhang, J.: Urban traffic prediction from spatio-temporal data using deep meta learning. In: KDD, pp. 1720–1730 (2019)
17. Pan, Z., Wang, Z., Wang, W., Yu, Y., Zhang, J., Zheng, Y.: Matrix factorization for spatio-temporal neural networks with applications to urban flow prediction. In: CIKM, pp. 2683–2691 (2019)
18. Perozzi, B., Al-Rfou, R., Skiena, S.: DeepWalk: online learning of social representations. In: KDD, pp. 701–710 (2014)
19. Velickovic, P., Fedus, W., Hamilton, W.L., Lio, P., Bengio, Y., Hjelm, R.D.: Deep graph infomax. In: ICLR (Poster) (2019)
20. Wang, D., Zhang, J., Cao, W., Li, J., Zheng, Y.: When will you arrive? Estimating travel time based on deep neural networks. In: AAAI, pp. 2500–2507 (2018)
21. Wang, H., Tang, X., Kuo, Y.H., Kifer, D., Li, Z.: A simple baseline for travel time estimation using large-scale trip data. ACM Trans. Intell. Syst. Technol. **10**(2), 19:1-19:22 (2019)
22. Wang, P., Fu, Y., Zhang, J., Wang, P., Zheng, Y., Aggarwal, C.C.: You are how you drive: peer and temporal-aware representation learning for driving behavior analysis. In: KDD, pp. 2457–2466 (2018)
23. Wang, Z., Fu, K., Ye, J.: Learning to estimate the travel time. In: KDD, pp. 858–866 (2018)
24. Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In: IJCAI, pp. 3634–3640 (2018)
25. Zhang, J., Zheng, Y., Qi, D., Li, R., Yi, X., Li, T.: Predicting citywide crowd flows using deep spatio-temporal residual networks. Artif. Intell. **259**, 147–166 (2018)
26. Zhang, J., Zheng, Y., Qi, D.: Deep spatio-temporal residual networks for citywide crowd flows prediction. In: AAAI, pp. 1655–1661 (2017)
27. Zhang, Z., Li, M., Lin, X., Wang, Y., He, F.: Multistep speed prediction on traffic networks: a deep learning approach considering spatio-temporal dependencies. Transp. Res. Part C: Emerg. Technol. **105**, 297–322 (2019)
28. Zhao, L., et al.: T-GCN: a temporal graph convolutional network for traffic prediction. IEEE TITS **21**, 3848–3858 (2019)