



# Route to Time and Time to Route: Travel Time Estimation from Sparse Trajectories

Zhiwen Zhang<sup>\*2</sup>, Hongjun Wang<sup>\*1</sup>, Zipei Fan <sup>1,2</sup>,  
Jiyuan Chen<sup>1</sup>, Xuan Song , and Ryosuke Shibasaki<sup>2</sup>

<sup>1</sup> Southern University of Science and Technology, Shenzhen, China

<sup>2</sup> The University of Tokyo, Tokyo, Japan

\* Equal contribution

fanzipei@iis.u-tokyo.ac.jp, songx@sustech.edu.cn

**Abstract.** Due to the rapid development of Internet of Things (IoT) technologies, many online web apps (e.g., Google Map and Uber) estimate the travel time of trajectory data collected by mobile devices. However, in reality, complex factors, such as network communication and energy constraints, make multiple trajectories collected at a low sampling rate. In this case, this paper aims to resolve the problem of travel time estimation (TTE) and route recovery in sparse scenarios, which often leads to the uncertain label of travel time and route between continuously sampled GPS points. We formulate this problem as an inexact supervision problem in which the training data has coarsely grained labels and jointly solve the tasks of TTE and route recovery. And we argue that both two tasks are complementary to each other in the model-learning procedure and hold such a relation: more precise travel time can lead to better inference for routes (*Time*  $\rightarrow$  *Route*), in turn, resulting in a more accurate time estimation (*Route*  $\rightarrow$  *Time*). Based on this assumption, we propose an EM algorithm to alternatively estimate the travel time of inferred route through weak supervision in  $E$  step and retrieve the route based on estimated travel time in  $M$  step for sparse trajectories. We conducted experiments on three real-world trajectory datasets and demonstrated the effectiveness of the proposed method.

**Keywords:** Internet of things · Weakly supervised learning · Graph convolutional network · Travel time estimation · Route recovery

## 1 Introduction

With advances in the area of the Internet of Things (IoT), GPS modules have been widely used throughout various kinds of mobile devices. These devices collected massive trajectory data and empowered many applications in the intelligent transportation system. Among these applications, travel time estimation (TTE) is an essential task for route planning, taxi dispatching, and ride-sharing. Subsequently, a large part of relevant approaches ranging from machine learning technologies, such as Bayesian inference [15], to deep learning models [22] have

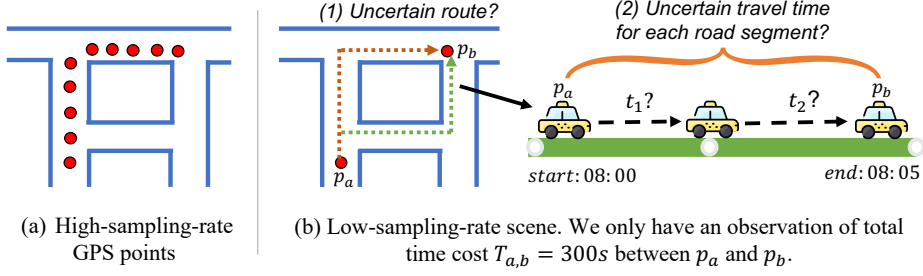


Fig. 1. Comparison between dense and sparse TTE scenarios.

been proposed to solve this task. However, due to the power and communication limitations of the mobile devices, the sampling rate of the trajectories is always low, which leads to a decrease in the accuracy for both travel time and route. Existing efforts need to label the exact travel time and route between two consecutively sampled GPS points, which is used to train the estimation model. We argue that this hypothesis sounds reasonable relying upon the scene of high-sampling-rate. In practice, we have to face a large part of trajectory data with low sampling rates [10,23].

**Motivating Scenario.** Fig. 1 gives a comparison between two TTE scenarios: dense and sparse. The exact label of travel time in each road segment and the route can be easily obtained from the dense trajectories (Fig. 1.a). However, we can not obtain the precise route and travel time label from the sparse trajectories. Fig. 1.b illustrates a case in low-sampling rate. We are hard to infer the route when given two sampled GPS points  $p_a$  and  $p_b$ , since there are multiple choices for possible route. Meanwhile, we were also challenged to acquire the exact travel time in each road segment, even though we have the ground truth route marked with a green dotted line, due to the large gap of observation  $T_{a,b} = 5$  minute. With missing supervision labels, traditional supervised learning is clumsy in giving a fine-grained prediction. This motivates us to model TTE and route recovery from sparse trajectories as a weakly supervised learning problem, more specifically a coarse labeling problem [29].

Unlike conventional supervised learning, where each sample is assigned with a label, coarse labeling annotates the label on a bag of samples. The authors in [29] summarize the task of learning from 1) the mean/sum: the arithmetic mean or the sum of  $X$ ; 2) the difference/rank: the difference  $x_i - x_j$  or the relative order  $x_i > x_j$ ; and 3) the min/max: the smallest/largest value in  $X$ . In the task of estimating travel time, the problem can be considered as learning from the mean/sum of  $X$ , since the path travel time can be equivalent to the summation of each road pass time within the path, while the exact travel time of each road segment is unknown.

As we know, travel time and route are highly correlated. In addition, the exact routes can result in a better inference of travel time ( $Route \rightarrow Time$ ), in

turn leading to a more precise route recovery ( $Time \rightarrow Route$ ). In this paper, the Expectation-Maximization (EM) algorithm [4] has been applied to alternatively estimate the travel time and route between any two consecutive GPS points. Technically, the **E step** intends to estimate the travel time of inferred route through weakly supervised learning (WSL), and the **M step** schemes to recover the route by heuristically searching for estimated travel time. Furthermore, to model the time-variant representation of road network, we generate the travel time distributions using the proposed spatio-temporal model. The Lognormal distribution is employed in this paper thanks to the excellent nature of additivity [5] and better performance in fitting real travel time.

The main contributions of this paper can be summarized as follows.

- For the first time, we integrate weakly supervised learning into the problem of TTE, which aims to infer the travel time of each road segment in a *bag* from a large gap of consecutive GPS points.
- The EM algorithm has been designed to alternatively infer the travel time distribution of each road segment and route between two consistent GPS samples ( $Route \rightarrow Time$  and  $Time \rightarrow Route$ ). In addition, we propose a spatio-temporal embedding architecture to forecast the future traffic state that integrates the spatial relational road network and temporal correlations, such as weather conditions and time-of-day.
- We conduct extensive experiments on three real-world large-scale trajectory data sets, which significantly outperform the state-of-the-art baselines for both two tasks - TTE and route recovery.

## 2 Related Work

**Weakly Supervised Learning.** Weakly supervised learning focuses on dealing with three kinds of problems [31]: 1) incomplete supervision: only part of the training data is labeled 2) inexact supervision: training data has only coarsely grained labels 3) inaccurate supervision: given labels are not always accurate. Multiple instance learning (MIL), which deals with observed data arranged in sets [7] is a branch of weakly supervised learning belonging to the category of inexact supervision. MIL has been widely applied in many fields, such as image and video classification [3], as well as document and sound classification [32]. This paper expands the concept of MIL to the application of travel time estimation in a highly sparse scenario.

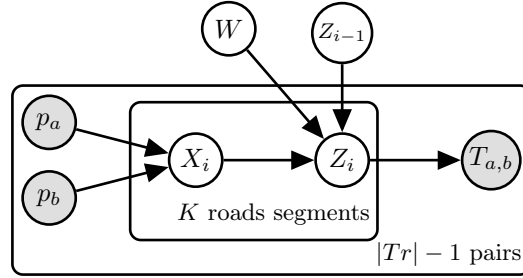
**Travel Time Estimation.** The loop detectors are firstly used in calculating the travel time by recording the individual road travel speeds and dividing it by the travel distance [13]. However, since traffic lights and left/right turns are omitted, the estimation errors are inaccurate. Therefore, road segment-based methods have been proposed, which can be approximately divided into two types: 1) nearest neighbor search [20], which sets the prediction by averaging the historical trajectory travel time; and 2) trajectory regression methods [9], which predict the travel time of road segment by road features. However, those

approaches are based on the assumption that the trajectory’s travel time is precise. Moreover, multiple trajectories with low sampling rates exist due to the network communication problem. Although some works try to conduct sparse travel time estimation[10,23], the uncertain route is also ignored. This paper aims to simultaneously resolve the problem of estimating vehicle travel time and route recovery in a highly sparse scenario.

**Route Recovery.** The route recovery problem in the low sample rate scenario is vital to reduce the uncertainty of the trajectory, and the TTE problem [17]. As we mentioned, the problems of TTE and route recovery play a role together, and this idea has been considered in previous work. For example, [24] designs a regression TTE model and applies the exact route search to obtain the potential route based on the learned travel time. [19] proposes STGAN to generate a travel time distribution in each road segment throughout the road network based on data from traffic surveillance cameras and update the possible route by posterior estimation in every iteration. Unlike the existing approach, the superior advantage of WSL-TTE is to model the sparse observation problem as weakly supervised learning, which is skilled in coarse labeling problems, and adopt it into the EM framework.

### 3 Methodology

This section first gives the problem formulation of travel time estimation based on weakly supervised learning and then introduces our proposed weakly supervised learning travel time estimation (WSL-TTE) system.



**Fig. 2.** The graphical model of the data generating process. The grey and white nodes represent the observation and hidden variable, respectively.

#### 3.1 Notation

Let  $X$  be the features of the road (e.g., road types, road lanes) and  $Z$  be the **unobserved true travel-time distribution of  $K$  road segments that we want to predict**. The goal is to learn a discriminative model  $f$  that predicts

the true target  $Z$  from the feature vector  $X$ , so as to maximize the conditional probability  $P(Z | X)$ . Here, the travel time distribution  $Z$  highly depends on the real-time traffic condition. Intuitively, we discretize one day into  $\mathcal{I}$  time steps (i.e., a specific time window  $\Delta t=30$  minutes). So given a trajectory  $Tr$  at time step  $t_s$ , which can be denoted as sequence of sampled GPS point:  $p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_{|Tr|}$ ,  $\tilde{r}$  is the most likely route (*bag*) between two continuous GPS points  $p_a$  and  $p_b$  during  $Tr$ , where  $|\tilde{r}| = K$  denotes the number of total travelled road segments in  $\tilde{r}$ . Especially,  $X_{1:K}$ <sup>3</sup> stands for the *bag* of the travelled road segments' features corresponding to  $\tilde{r}$ .  $W$  is the collection of all weight parameters in the neural network  $f$  to learn  $Z$ .  $T_{a,b}$  is the actual observation of total time cost between  $p_a$  and  $p_b$ .

Fig. 2 illustrates the data generation process with a graphical representation. Given  $p_a$  and  $p_b$ , we first infer the most likely route  $\tilde{r}$  from the road network with  $P(\tilde{r} | p_a, p_b, Z)$ , which is equal to estimate  $P(X_{1:K} | p_a, p_b, Z_{1:K})$ . Subsequently, we generate the conditional probability  $P(Z_i | \theta)$  for each travel time distribution  $Z_i$ . Specifically, we here assume that each latent variable  $Z_i$  belongs to the Lognormal distribution  $Z_i \sim \frac{1}{z\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln z - \mu)^2}{2\sigma^2}\right)$  [16]. We let the conditional probability  $P(Z_i | \theta) = P(Z_i | (\mu, \sigma) = f(X; W))$ , where  $W$  and  $X$  are the cofactors that generate the parameterized  $\mu$  and  $\sigma$  of  $Z_i$  by the deterministic function  $f$ . Consequently, total travel time  $T_{a,b}$  can be observed by an aggregate function  $Q$  with  $T_{a,b} = Q(Z_{1:K}) = \sum_{i=1}^K Z_i$ .

### 3.2 Assumptions

Here, we summarize the two basic assumptions used in our paper.

**Assumption 1** (*Aggregate observation assumption*)  $P(T_{a,b} | X_{1:K}, Z_{1:K}) = P(T_{a,b} | Z_{1:K})$

We assume that observation  $T_{a,b}$  is conditionally independent on  $X_{1:K}$  when given  $Z_{1:K}$ . This assumption is informed by existing studies [2] and conforms to the TTE problem, since given  $Z_{1:K}$ , observation  $T_{a,b}$  can be determined by aggregate function  $Q$ .

**Assumption 2** (*Markov chain assumption*)  $P(Z_{1:K} | X_{1:K}) = P(Z_1 | X_1) \prod_{i=2}^K P(Z_{i+1} | X_{i+1}; Z_i)$

We assume that  $Z_{i+1}$  are mutually independent except for  $Z_i$ . This is under the assumption of a Markov chain and is based on extensive applications in trajectory data mining [1]. Furthermore, since  $T_{a,b}$  can be determined by the function  $Q$ , the conditional probability can be defined as  $P(T_{a,b} | Z_{1:K}) = \delta_{Q(Z_{1:K})}(T_{a,b})$ , where  $\delta(\cdot)$  represents the Dirac delta function.

<sup>3</sup> The subscript for example  $X_{1:K}$  denotes an abbreviation for the set  $\{X_1, X_2, \dots, X_K\}$

### 3.3 Problem Formulation

In summary, the objective function in this paper can be written as follows:

$$\log P(T_{a,b} | p_a, p_b, W) = \log \left( \sum_Z P(T_{a,b} | p_a, p_b, Z, W) P(Z | p_a, p_b, W) \right), \quad (1)$$

which is the maximum of a posterior estimation by taking the  $Z$  as latent variables with the observation of sparse travel time  $T_{a,b}$ . Therefore, we divide the training process into expectation ( $E$  step) and maximization ( $M$  step) according to the above assumptions.

$$\begin{aligned} \mathbf{E \ step:} \quad & \mathbb{E}_Z \left[ \log P(T_{a,b}, Z | X_{1:K}, W) | X_{1:K}^{(i)}; W^{(i)} \right] \\ &= \int_{Z^K} \log P(T_{a,b} | Z_{1:K}; X_{1:K}) P(Z_{1:K} | X_{1:K}) d_{Z_{1:K}} \\ &= \int_{Z^K} \delta_{Q(z_{1:K})}(T_{a,b}) P(Z_1 | X_1) \prod_{i=2}^K P(Z_{i+1} | X_{i+1}; Z_i) d_{Z_{1:K}} \\ &\approx \mathbb{E}_{\substack{Z_i \sim p(Z_i | X_i) \\ i=1, \dots, K}} [\delta_{Q(Z_{1:K})}(T_{a,b})] \end{aligned} \quad (2)$$

$$\mathbf{M \ step:} \quad \tilde{r} = \arg \max_r \log P(T_{a,b} | \Omega_{a,b}; p_a; p_b; W) \quad (3)$$

**E step** aims to estimate the travel time of the most likely route  $\tilde{r}$  by learned travel time distribution  $Z$  through weakly supervised learning, and **M step** heuristically searches  $\tilde{r}$  from the candidate set  $\Omega_{a,b} = \{r_1, r_2, \dots, r_m\}$  to reduce the computational cost. According to the above EM procedure, we obtain the estimated route  $\tilde{r}$  for every pair of  $p_a$  and  $p_b$ , as well as travel time distribution  $Z$ . Thus, the final travel time of trajectory  $Tr$  can be obtained by summing all the estimation components  $\Theta = \{\tilde{T}_1, \tilde{T}_2, \dots, \tilde{T}_{|Tr|-1}\}$  between every continuous GPS sample with  $\tilde{T} = \sum_{i=1}^{|Tr|-1} \tilde{T}_i$ , where  $\tilde{T}_i$ , produced by  $f$ , denotes the forecast travel time of  $p_i$  and  $p_{i-1}$ .

### 3.4 System Overview

Fig. 3 shows our proposed WSL-TTE system with the EM algorithm, which consists of three main components - processor, spatio-temporal model and model training with EM procedure.

(1) **Processor** temporally partitions the low-sampling-rate historical trajectories from the datasets into each time step  $t_s$ .

(2) **Spatio-Temporal Model**  $f$  estimates the travel time distribution  $Z$  of the road segments. We firstly transform a road network into a multi-relational graph  $\mathcal{G}$ , and encode each time step  $t_s$  into a vector. Then we fuse the aforementioned spatial representation and temporal encoding as the time-variant vertex

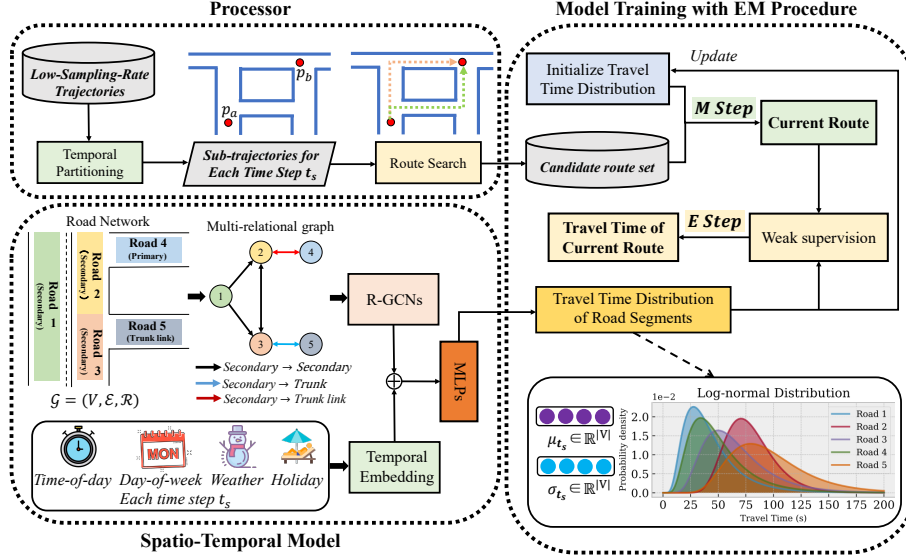


Fig. 3. The system architecture of our proposed WSL-TTE with EM procedure.

representations. In the final,  $\mu_{t_s}$  and  $\sigma_{t_s}$  of travel time distribution  $Z$  are parameterized by two MLPs (Multi-Layer Perception).

(3) **Model Training with EM Procedure** is used to optimize the learned  $Z$  in  $E$ -step using Eq. (5) and infer route  $\tilde{r}$  by Eq. (6). Travel time distribution  $Z$  is initially default as  $\mu = \frac{L_i}{S_i}$  and  $\sigma = 1$ , where  $L_i$  and  $S_i$  are the length and speed limits of  $i^{th}$  road segment, respectively. The EM algorithm will be finished when the estimated variables  $Z$  converge.

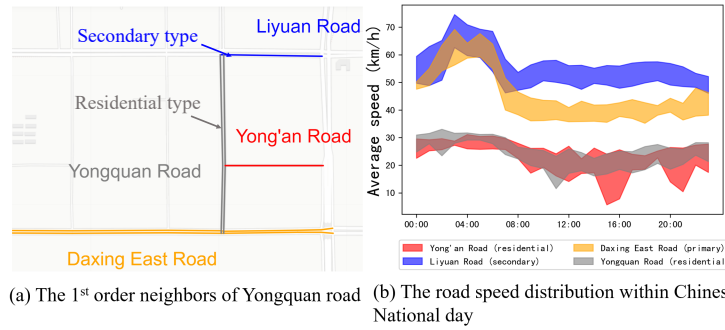


Fig. 4. Motivation example of the traffic speed in Xi'an during Chinese National Day. We used high-sampling-rate GPS trajectories to calculate the exact travel speed for these road segments, and observed that the road's speed distribution with adjoining roads is highly related with the road types.

### 3.5 Spatio-Temporal Model

The spatio-temporal model aims to learn  $f$  to estimate travel time distribution  $Z$  of road segments from road network. Fig. 4(a) shows the locations of the neighbors of Yongquan Street, and Fig. 4(b) depicts the maximum, minimum and average traffic speed during Chinese national day. We observe several phenomena: 1) road speed is highly related to the road types; and 2) the road speed is also affected by the type of connection road. For example, Yongquan road and Liyuan street are residential and secondary roads, respectively. Even though they are neighbors, the road speeds are relatively different. Thus, the road network is represented as a multi-relational graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ , where  $\mathcal{V}$  denotes the set of vertices (i.e. road segments) and  $\mathcal{E}$  denotes the set of edges. An edge  $\mathbf{e}_{ijk} = (v_i, v_j, \mathbf{r}_k) \in \mathcal{E}$  indicates that the road segment  $v_i \in \mathcal{V}$  connects to the road segment  $v_j \in \mathcal{V}$  with a relation type  $\mathbf{r}_k \in \mathcal{R}$ .

Based on this multi-relational graph, we adopt a 3-layer Relational Graph Convolution Networks (R-GCNs) [18] as the building block to learn the graph structure information.  $l^{th}$  R-GCN can be defined as

$$h_{v_i}^{(l+1)} = \sum_{\mathbf{r}_k \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^{\mathbf{r}_k}} \frac{1}{c_{i, \mathbf{r}_k}} W_{\mathbf{r}_k}^{(l)} h_j^{(l)} + W_0^{(l)} h_{v_i}^{(l)}, \quad (4)$$

where  $h_{v_i}^{(l)} \in \mathbb{R}^{d^{(l)}}$  is the hidden state of road segment  $v_i$  in the  $l^{th}$  layer of the model with dimension  $d^{(l)}$  and  $W_{\mathbf{r}_k}^{(l)}, W_0^{(l)}$  present the learnable parameters.  $\mathcal{N}_i^{\mathbf{r}_k}$  denotes the set of neighbor indices of node  $v_i$  in relation to  $\mathbf{r}_k \in \mathcal{R}$ .  $c_{i, \mathbf{r}_k}$  is the normalization constant. Note that  $h_{v_i}^{(0)}$  is the spatial feature  $X_{v_i}$  of road segment  $v_i$ . We use the embedding layer to encode the following statistical features:

- Road types: for example, primary, primary link, secondary, secondary link;
- Number of lanes: how many marked traffic lanes;
- Whether it is one way or not.

The final output of R-GCNs, represented as  $s_{v_i} \in \mathbb{R}^D$ , where  $v_i \in \mathcal{V}$ . But spatial representation  $s_{v_i}$  only provides the static representation, which could not show the temporally dynamic correlations for each road segment.

As we mentioned previously, the spatio-temporal model  $f$  is to estimate the mean value  $\mu$  and variance  $\sigma$  of Lognormal distribution for each road segment  $v_i$  at each time step  $t_s$ . Intuitively, we encode the day-of-week and time-of-day of each time step  $t_s$  into  $\mathbb{R}^7$  and  $\mathbb{R}^{\mathcal{I}}$  using one hot encoding, and concatenate them with the embedding of weather conditions and HolidayID (holiday or not). Then we use one-layer MLP to transform the above temporal embedding vector into a vector  $s_{t_s} \in \mathbb{R}^D$ , which is equal to the spatial representation  $s_{v_i}$ . To obtain the time-variant road segment representations, we fuse the above spatial representation and reconstruct temporal embedding vector: for each road segment  $s_{v_i}$  at time step  $t_s$ , the spatio-temporal representation is defined as  $F_{v_i, t_s} = s_{v_i} + s_{t_s}$ , which contains both spatial road structure and temporal information. Based on the spatio-temporal representation  $F \in \mathbb{R}^{(|\mathcal{V}| * N_{t_s}) \times D}$ , where  $N_{t_s}$  denotes the total number of time steps,  $\mu_{t_s} \in \mathbb{R}^{|\mathcal{V}|}$  and  $\sigma_{t_s} \in \mathbb{R}^{|\mathcal{V}|}$  for each time step  $t_s \in \mathbb{R}^{N_{t_s}}$  are parameterized by two-layer MLPs with shared fused representation  $F$ .



### 3.6 Model Training with EM Procedure

Next, we introduce the learning procedure of estimated travel time distribution  $Z$  through weakly supervised learning. The aforementioned *expected log-likelihood* in Equation (2) defines the aggregate expectation from  $Z$  to  $T$ . Here, we assume the distribution  $T_{a,b}$  also under the Lognormal distribution approximated by summation of all  $Z_{1:K}$  on route  $r$  [5] as

$$T \sim \text{Lognormal} \left( \mu := \sum_{i=1}^K \mu_{t_s}^{(i)}, \sigma^2 := \sum_{i=1}^K (\sigma_{t_s}^{(i)})^2 \right)$$

Thus, the term of expectation in Eq. (2) can be derived to be

$$\begin{aligned} L_{\mu, \sigma} &= \log p \left( \left\{ x_{t_s}^{(i)}, z_{t_s}^{(i)} \right\}_{i=1}^K ; W \right) \\ &= \sum_{i=1}^K \frac{\left( z_{t_s}^{(i)} - \mu_{t_s}^{(i)} \right)^2}{2\sigma_{t_s}^2} - \frac{1}{2} \log (2\pi\sigma_{t_s}^2) \\ &\approx -\frac{(Q(Z_{1:K}) - Q(\mu))^2}{2\sigma^2} - \frac{1}{2} \log (2\pi\sigma^2). \end{aligned} \quad (5)$$

In the final, we introduce how to maximize the conditional probability in Eq. (3). Given two continuous samples  $p_a, p_b$ , and the last-step parameters  $W^{(i)}$  produced from Eq. (5), our objective is to find the optimal route  $\tilde{r}$  with the travel time closest to the observation  $T_{a,b}$ . As mentioned in [8], it is natural to assume that the route  $\tilde{r}$  is very likely to be among the top  $m$ -shortest paths between  $p_a$  and  $p_b$ . Therefore, we utilize Yen's algorithm [26] to generate the candidate set  $\Omega_{a,b} = \{r_1, r_2, \dots, r_m\}$  and the optimal route  $\tilde{r}$  can be selected by

$$\tilde{r} = \arg \min_r \left| T_{a,b} - \sum_{e_i \in r_j} \mu^{(i)} \right|, \quad \forall r_j \in \Omega_{a,b}. \quad (6)$$

We perform Eq. (6) for every pair of continuous samples to update their corresponding route  $\tilde{r}$ . Furthermore, to prevent extensive overlap in  $\Omega_{a,b}$ , this paper leverages the weighted Jaccard ( $wJCD$ ) value to calculate each pair of routes  $r_i, r_j \in \Omega_{a,b}$  referring to [8]. The EM algorithm will complete when it reaches the estimated variables  $\mu$  and  $\sigma$  convergence.

## 4 Experiments

### 4.1 Experimental Settings

**Data.** We validate our proposed methods on three real-world datasets, including Xi'an, Porto and Chengdu dataset. More details can be found in Appendix A.1.

**Sampling Rate Setting.** According to [27], taxis should report their GPS positions with a low sampling rate to save communication and energy costs. We further vary the sampling ratio of the sets 3.125%, 6.25% and 12.5% to evaluate the robustness of our proposed model. Since the original trajectories are sampled every 15 seconds, the generated low-sampling-rate trajectories of 3.125%, 6.25% and 12.5% are considered to be as the average time interval of such trajectories is 8 mins, 4 mins, and 2 mins, respectively.

#### 4.2 Baseline Models & Evaluation Metrics

- **DeepTTE:** [22] is an end-to-end deep learning framework, which infers the travel time from both the entire path and each local path simultaneously.
- **DeepGTT:** [11] learns the travel time distribution through the deep generative model, which takes the real-time traffic condition into account.
- **MVSTM:** [14] is a multi-view spatial-temporal model that captures the mutual dependence of spatial-temporal relations and trajectory features.
- **MURAT:** [12] is a multi-task representation learning method by utilizing the underlying road network and the spatio-temporal prior knowledge.
- **DCRNN:** [13] exploits GCN to capture spatial dependency and then uses recurrent neural networks to model temporal dependency.
- **ConSTGAT:** [6] adopts a graph attention mechanism to explore the joint relations of spatio-temporal information.
- **T-GCN:** [30] proposes a temporal GCN model that combines the GCN and GRU to simultaneously extract the spatial and temporal dependencies.

Additionally, three state-of-the-art algorithms STRS [24], MTrajRec [17], and DeepGTT [11] are used as baseline models for route recovery. STRS and DeepGTT learn the travel time of the road network and conduct the route search for low-sampling-rate trajectories. MTrajRec recovers the route via a two-stage Seq2Seq model based on coarse grid representation.

**Evaluation Metrics.** We evaluate the task of TTE with RMSE (radial mean square error), MAE (mean absolute error) and MAPE (mean absolute percentage error). Then the route recovery performance is evaluated by route recovery accuracy, which is defined as the ratio of the length of correctly inferred road segments against the maximum value of the length of the ground truth route  $R_G$  and the inferred route  $R_I$ , that is,  $accuracy = \frac{(R_G \cap R_I).len}{\max\{R_G.len, R_I.len\}}$ .

#### 4.3 Performance Comparison

**Performance on Travel Time Estimation.** As reported in Table 1, our WSL-TTE achieves the best results among all baseline methods for three kinds of minute intervals: 2, 4, 8. We summarize the reasons for our model outperforming all baselines by a large margin: 1) The GCN production is used to learn the route’s travel time and its travel time distribution for each road segment simultaneously, helping it to yield robust and abundant features. 2) The EM iteration algorithm has been proposed to update the potential travelled path,

**Table 1.** Performance comparison for TTE under three datasets. Here, the units of both RMSE and MAE are minutes, and the unit of MAPE is percentages (%). The best performance is marked in bold font.

Data Models		2 mins			4 mins			8 mins		
		RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Xi'an	DeepTTE	2.89	1.74	14.89	3.88	2.57	15.7	3.87	2.79	18.25
	DeepGTT	4.31	3.51	32.16	5.27	4.11	39.85	7.33	6.14	45.31
	MVSTM	4.13	2.59	15.42	4.45	3.68	29.37	5.62	3.84	26.51
	MURAT	8.86	6.87	84.06	9.45	7.76	94.16	11.36	9.23	113.69
	T-GCN	3.24	2.00	14.78	4.23	3.04	15.61	4.43	3.19	18.37
	DCRNN	3.20	1.96	14.6	4.18	2.98	15.43	4.37	3.13	18.16
	ConSTGAT	3.21	1.99	14.41	4.20	3.01	15.22	4.39	3.17	17.91
	Ours	<b>1.36</b>	<b>1.07</b>	<b>9.66</b>	<b>1.53</b>	<b>1.32</b>	<b>13.35</b>	<b>1.81</b>	<b>1.92</b>	<b>15.69</b>
Porto	DeepTTE	2.14	1.46	12.59	2.90	2.10	12.36	3.38	2.48	15.01
	DeepGTT	3.39	2.83	28.58	4.07	3.03	32.05	6.23	5.08	39.34
	MVSTM	3.10	2.27	13.00	3.39	2.75	24.28	4.69	3.18	22.18
	MURAT	6.24	5.35	74.27	7.17	5.63	75.54	9.40	7.39	96.79
	T-GCN	2.04	1.31	12.13	2.85	1.83	11.92	3.16	2.23	14.26
	DCRNN	2.02	1.28	11.99	2.81	1.80	11.78	3.12	2.19	14.09
	ConSTGAT	2.03	1.29	11.83	2.82	1.82	11.62	3.13	2.22	13.90
	Ours	<b>1.11</b>	<b>0.88</b>	<b>8.25</b>	<b>1.30</b>	<b>1.14</b>	<b>10.58</b>	<b>1.72</b>	<b>1.63</b>	<b>12.77</b>
Chengdu	DeepTTE	3.13	1.97	15.11	4.09	2.99	15.96	4.28	3.14	18.79
	DeepGTT	4.97	3.83	34.31	5.75	4.32	41.39	7.89	6.44	49.24
	MVSTM	4.54	3.07	15.61	4.78	3.92	31.36	5.94	4.03	27.77
	MURAT	9.13	7.24	89.17	10.12	8.03	97.57	11.92	9.36	121.16
	T-GCN	2.99	1.77	14.56	4.02	2.61	15.39	4.01	2.83	17.85
	DCRNN	2.95	1.73	14.39	3.97	2.56	15.21	3.96	2.78	17.64
	ConSTGAT	2.97	1.75	14.20	3.98	2.59	15.01	3.97	2.81	17.40
	Ours	<b>1.89</b>	<b>1.31</b>	<b>10.15</b>	<b>2.14</b>	<b>1.93</b>	<b>13.98</b>	<b>2.55</b>	<b>2.21</b>	<b>16.28</b>

which helps our method to learn a more reasonable travel time distribution. Meanwhile, the estimation results among different sampling intervals also reflect

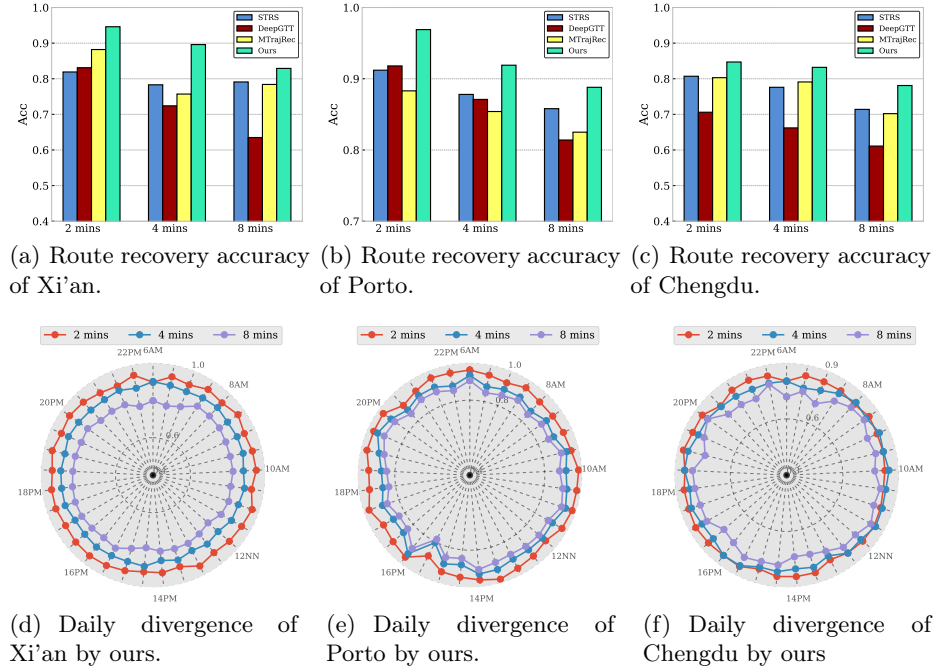
**Table 2.** Performance of our framework and ablation variants for TTE under extremely sparse scenario (8 mins).

Models	Xi'an/Porto/Chengdu		
	RMSE	MAE	MAPE
simpleGCN	1.92/2.09/2.83	2.04/1.74/2.37	15.96/13.09/16.82
GAT	1.96/2.14/2.89	2.15/1.80/2.42	15.92/13.10/16.87
GTN	1.95 /2.10/2.82	2.11/1.78/2.41	15.94/13.07/16.77
Normal Distribution	1.88/2.06/2.79	2.09/1.84/2.56	15.88/13.09/16.89
Variance=1	2.07 /2.43/3.45	2.71/2.17/2.79	16.12/13.28/17.13
Ours	<b>1.81/1.72/2.55</b>	<b>1.92/1.63/2.21</b>	<b>15.69/12.77/16.28</b>

that the uncertainty of a sparse GPS trajectory would seriously affect the model performance.

**Ablation Study.** As is shown in Table 2, in order to validate how the relational GCN modules and weak supervision can effectively capture the spatio-temporal dependencies in WSL-TTE, we first test the effects of relational GCN on modeling road network. Our WSL-TTE removes the relational GCN module and replaces it with a simple GCN, GAT[21] and graph transformer network (GTN) [28] to extract the spatial representation. The experimental results show that using our model with relational GCN can achieve better performance on two datasets. It can be explained that the complex adjacency of the road network needs to model different correlations among different road types. Next, we validate the assumption that the travel time variables belong to the Lognormal distribution. Compared to this setting, we conduct the test of normal distribution and variance=1, respectively. We find that they cannot achieve better performance than the Lognormal distribution. This validates our formulation of weak supervision regarding travel time. In sum, we can conclude that our WSL-based method is effective in travel time estimation.

**Performance on Route Recovery.** Fig. 5 reports the route recovery accuracy and daily divergence over different sampling time intervals. Fig. 5(a), Fig. 5(b) and Fig. 5(c) show that our WSL achieves better performance among

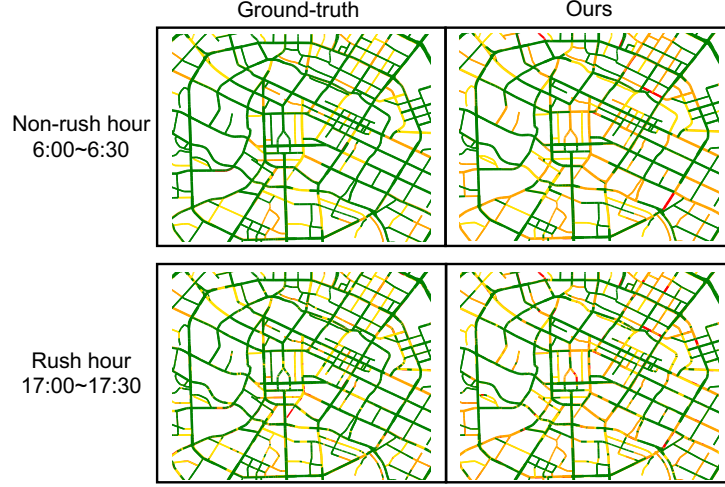


**Fig. 5.** Route Recovery Performance.

three sampling intervals, compared to DeepGTT, MTrajRec and STRS. Noted that DeepGTT achieves worse performance at 4 mins and 8 mins. This is because grid-based traffic condition tensors can not provide efficient road conditions at high sampling intervals. Fig. 5(d), Fig. 5(e) and Fig. 5(f) provide the daily divergence of the route recovery accuracy by our proposed WSL. We find that the total recovery performance stays stable from 6:00 AM to 22:00 PM for Xi'an, Porto, and Chengdu. However, the performance of Chengdu is relatively worse than both Xi'an and Porto due to the more complex road network. Furthermore, significantly as the sampling time interval increases, the accuracy of both methods drops, as expected. The reason is that a more extensive sampling time interval leads to more possible candidate routes to be inferred between two sample points. Meanwhile, the daily divergence of our WSL also shows this pattern.

#### 4.4 Case Study

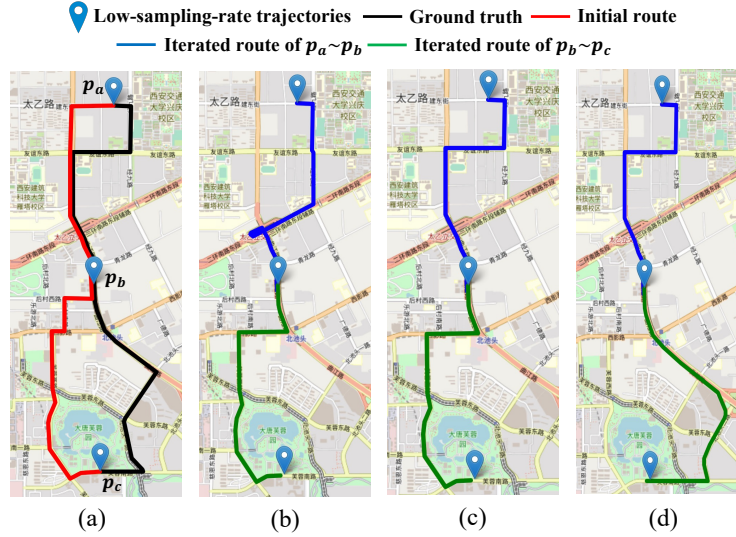
We conducted a real-world case study in Chengdu, which visualizes the learned road conditions using our proposed WSL-TTE. To acquire the road conditions of the road network, we here transform travel time  $\mu$  estimated by our WSL-TTE into the average speed by  $Speed_i = \frac{Length_i}{\mu_i}$  for each road  $v_i$ . Four kinds of colors are used to represent the different road states, which can be defined as 1) red - very congested, 2) yellow - congested, 3) orange - slow, and 4) green - unblocked. We equally divide the limiting velocity for each road type and set the speed interval for these four road states. For example, the speed limit of the primary road type is  $60kph$ , then the speed range that represents very congested



**Fig. 6.** Traffic condition comparison. We pick two-time steps, i.e., non-rush hour (6:00-6:30) and rush hour (17:00-17:30), and compute the ground truth by original dense trajectories in Chengdu, compared with the transformed speed based on the learned travel time distributions of our WSL-TTE.

is  $[0, 15)$ , congested is  $[15, 30)$ , slow is  $[30, 45)$  and unblocked is  $[45, 60)$ . We calculate the average speeds by the original dense trajectories as the ground truth. Specifically, we mark them with an unblocked state for the roads without a trajectory. The compared result is shown in Fig. 6, our model can generate approximate road conditions with ground truth for both non-rush and rush hour.

In addition, we provide a visualization example of the route update process in route recovery. As is shown in Fig. 7, our route update process with the EM algorithm can gradually find the approximate route with ground truth, owing to the precise travel time, which demonstrated our previous assumption that: the more precise travel time can lead to a better inference of routes, in turn, resulting in more accurate time estimation.



**Fig. 7.** Visualization example of the route update process. Here, Fig. (a) shows the ground truth and initial route of low-sampling-rate trajectories, and Fig. (b)~(d) shows the intermediate iteration results of route recovery.

## 5 Conclusion

This paper formulates the TTE and route recovery in a highly sparse scenario as an inexact supervision. Based on the EM algorithm, we solve the inexact travel time labeling and uncertain route choice by alternatively updating the travel time distribution through weakly supervised learning and route searching using the top  $m$ -shortest path respectively. Both two tasks are complementary to each other in the iteration process. In future work, we intend to consider more

hypotheses of travel time distribution under weakly supervised learning, such as Gamma, Weibull, as well as Burr XII distribution. Source code is available at <https://github.com/Dreamzz5/WSL-TTE>.

## Acknowledgment

This work was partially supported by National Key Research and Development Project (2021YFB1714400) of China and Guangdong Provincial Key Laboratory (2020B121201001).

## References

1. Brakatsoulas, S., Pfoser, D., Salas, R., Wenk, C.: On map-matching vehicle tracking data. In: Proceedings of the 31st international conference on Very large data bases. pp. 853–864 (2005)
2. Carboneau, M.A., Cheplygina, V., Granger, E., Gagnon, G.: Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition* **77**, 329–353 (2018)
3. Chen, Y., Bi, J., Wang, J.Z.: Miles: Multiple-instance learning via embedded instance selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(12), 1931–1947 (2006)
4. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* **39**(1), 1–22 (1977)
5. Dufresne, D.: Sums of lognormals. In: Actuarial Research Conference. pp. 1–6 (2008)
6. Fang, X., Huang, J., Wang, F., Zeng, L., Liang, H., Wang, H.: Constgat: Contextual spatial-temporal graph attention network for travel time estimation at baidu maps. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2697–2705 (2020)
7. Frénay, B., Verleysen, M.: Classification in the presence of label noise: a survey. *IEEE transactions on neural networks and learning systems* **25**(5), 845–869 (2013)
8. He, T., Bao, J., Li, R., Ruan, S., Li, Y., Song, L., He, H., Zheng, Y.: What is the human mobility in a new city: Transfer mobility knowledge across cities. In: Proceedings of The Web Conference 2020. pp. 1355–1365 (2020)
9. Idé, T., Sugiyama, M.: Trajectory regression on road networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 25 (2011)
10. Jabari, S.E., Freris, N.M., Dilip, D.M.: Sparse travel time estimation from streaming data. *Transportation Science* **54**(1), 1–20 (2020)
11. Li, X., Cong, G., Sun, A., Cheng, Y.: Learning travel time distributions with deep generative model. In: The World Wide Web Conference. pp. 1017–1027 (2019)
12. Li, Y., Fu, K., Wang, Z., Shahabi, C., Ye, J., Liu, Y.: Multi-task representation learning for travel time estimation. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1695–1704 (2018)
13. Li, Y., Yu, R., Shahabi, C., Liu, Y.: Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926* (2017)

14. Liu, Z., Wu, Z., Wang, M., Zhang, R.: Multi-view spatial-temporal model for travel time estimation. In: *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*. pp. 646–649 (2021)
15. Mil, S., Piantanakulchai, M.: Modified bayesian data fusion model for travel time estimation considering spurious data and traffic conditions. *Applied Soft Computing* **72**, 65–78 (2018)
16. Pu, W.: Analytic relationships between travel time reliability measures. *Transportation Research Record* **2254**(1), 122–130 (2011)
17. Ren, H., Ruan, S., Li, Y., Bao, J., Meng, C., Li, R., Zheng, Y.: Mtrajrec: Map-constrained trajectory recovery via seq2seq multi-task learning. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. pp. 1410–1419 (2021)
18. Schlichtkrull, M., Kipf, T.N., Bloem, P., Van Den Berg, R., Titov, I., Welling, M.: Modeling relational data with graph convolutional networks. In: *European semantic web conference*. pp. 593–607. Springer (2018)
19. Shao, K., Wang, K., Chen, L., Zhou, Z.: Estimation of urban travel time with sparse traffic surveillance data. In: *Proceedings of the 2020 4th High Performance Computing and Cluster Technologies Conference & 2020 3rd International Conference on Big Data and Artificial Intelligence*. pp. 218–223 (2020)
20. Tiesyte, D., Jensen, C.S.: Similarity-based prediction of travel times for vehicles traveling on known routes. In: *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*. pp. 1–10 (2008)
21. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017)
22. Wang, D., Zhang, J., Cao, W., Li, J., Zheng, Y.: When will you arrive? estimating travel time based on deep neural networks. In: *Thirty-Second AAAI Conference on Artificial Intelligence* (2018)
23. Wang, Y., Zheng, Y., Xue, Y.: Travel time estimation of a path using sparse trajectories. In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 25–34 (2014)
24. Wu, H., Mao, J., Sun, W., Zheng, B., Zhang, H., Chen, Z., Wang, W.: Probabilistic robust route recovery with spatio-temporal dynamics. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 1915–1924 (2016)
25. Yang, C., Gidofalvi, G.: Fast map matching, an algorithm integrating hidden markov model with precomputation. *International Journal of Geographical Information Science* **32**(3), 547–570 (2018)
26. Yen, J.Y.: Finding the k shortest loopless paths in a network. *management Science* **17**(11), 712–716 (1971)
27. Yuan, J., Zheng, Y., Zhang, C., Xie, X., Sun, G.Z.: An interactive-voting based map matching algorithm. In: *2010 Eleventh international conference on mobile data management*. pp. 43–52. IEEE (2010)
28. Yun, S., Jeong, M., Kim, R., Kang, J., Kim, H.J.: Graph transformer networks. *Advances in Neural Information Processing Systems* **32**, 11983–11993 (2019)
29. Zhang, Y., Charoenphakdee, N., Wu, Z., Sugiyama, M.: Learning from aggregate observations. *arXiv preprint arXiv:2004.06316* (2020)
30. Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., Li, H.: T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems* **21**(9), 3848–3858 (2019)



31. Zhou, Z.H.: A brief introduction to weakly supervised learning. *National science review* **5**(1), 44–53 (2018)
32. Zhou, Z.H., Sun, Y.Y., Li, Y.F.: Multi-instance learning by treating instances as non-iid samples. In: *Proceedings of the 26th annual international conference on machine learning*. pp. 1249–1256 (2009)