Link Representation Learning for Probabilistic Travel Time Estimation

Chen Xu, Qiang Wang, and Lijun Sun

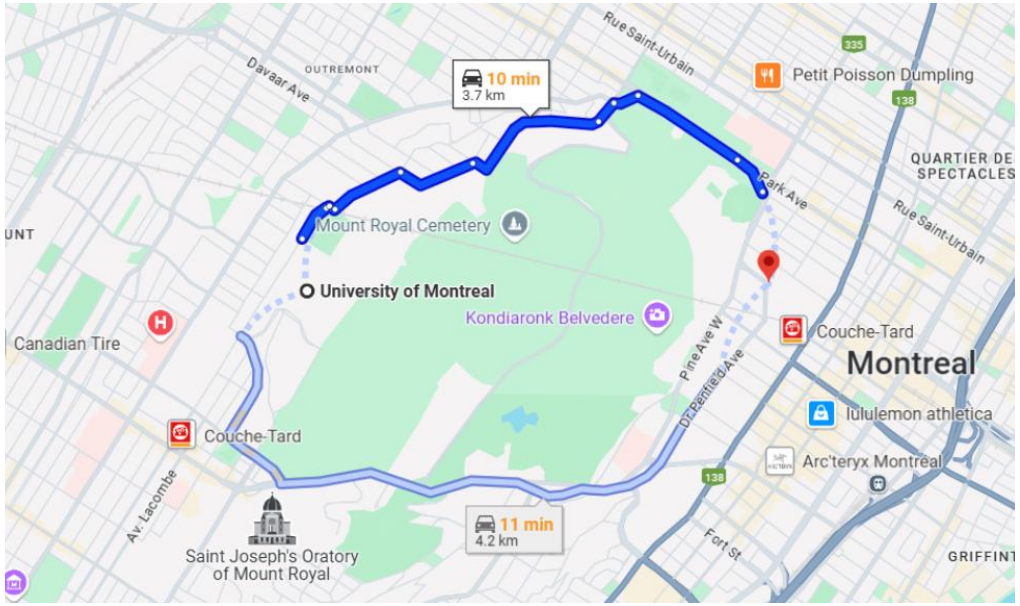
Background

When will you arrive at your destination?

- **Travel Time Estimation (TTE)** focuses on predicting the time it takes a vehicle to traverse any path connecting two locations (i.e., origin and destination).

  — influenced by external factors in both time and space.

  — exhibits variability across different drivers/routes.



➢ **Applications**

- Intelligent Transportation Systems (ITS): Navigation and route planning, traffic signal optimization.
- Logistics and Freight Optimization: Vehicle scheduling, delivery time prediction.

Background

Mainstream Methods for Travel Time Estimation

- **Origin-Destination (OD)-bsed**
  - Relies on coarse-grained OD data. Route information is not required.
  - Regression models to predict the average travel time between comparable OD pairs.
  - Simplicity and efficiency.
- **Route-based**
  - Requires detailed trajectory or routing data (GPS sequence).
  - Machine learning approaches modeling road networks, where travel time is computed by summing segment-level predictions.
  - Real-time and higher accuracy.

Background

Limitations of existing methods

- **Assumption of independent and identically distributed (i.i.d.) errors.**
  - Errors are independent of each other:

$$Cov(\varepsilon_i, \varepsilon_j) = 0 \quad for\ \ i \neq j$$

  - All errors are drawn from the same probability distribution:

$$\varepsilon_i \sim N(0, \sigma^2) \quad \forall\ i$$

  - Degenerates into a deterministic model and incapable of uncertainty quantification.

- **Anefficient utilization of GPS data.**
  - Only the origin and destination timestamps from the GPS data are utilized, while the timestamps within the trip is discarded.

Background

Our Motivation and Contribution

- **Motivation**: Modeling the multi-trip joint probabilistic distribution of travel times.
- **Contribution:**
  - — A multi-trip **joint probabilistic distribution model**, ProbETA, is introduced for estimating travel times, leveraging efficient low-rank parameterizations to capture inter-trip and intra-trip correlations across transportation network segments.
  - — A **sub-sampling data augmentation** approach is proposed to enhance sample balance and optimization efficiency, facilitating fine-grained modeling of link features.
  - — Experimental evaluations on two real-world datasets highlight the effectiveness of the proposed approach.

Background

Problem Formulation

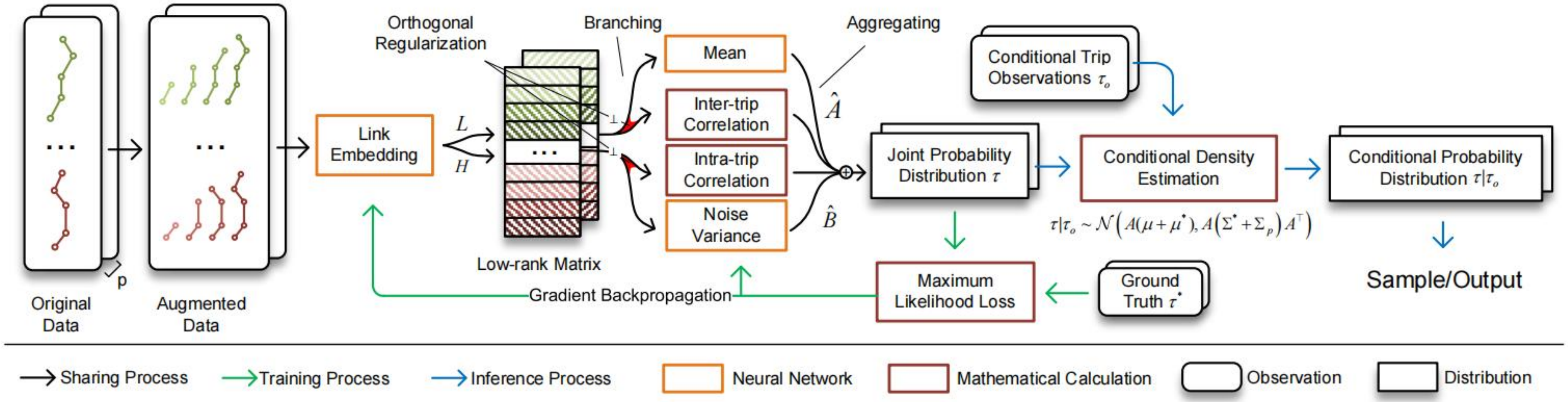
- Given a dataset $D = \{T_i | i = 1,2,...,N\}$ consisting of N historical trips in a road network, our objective is to estimate the probability distribution of the travel time τ_q for a query trip T_q that is not part of D.
- We consider the travel time τ_q to follow a Gaussian distribution:

$$\tau_q \sim N(\mu_q, \sigma_q^2),$$

where $\mu_q = f_\mu(T_q)$ and $\sigma_q^2 = f_\sigma(T_q)$, and $f_\mu(\cdot)$ and $f_\sigma(\cdot)$ are two designed models for estimating the mean and the standard deviation of travel time τ_q, respectively.

Methodology

Overview



- **Hierarchical Structure**: Incorporates both day-specific and trip-specific random effects.
- **Low-rank Representations**: Efficiently parameterizes link-level covariance matrices for travel time modeling.
- **Data Augmentation**: Uses sub-sampling to enhance link-level representation learning.
- **Conditional Estimation**: Predicts travel time distributions based on spatiotemporally related observed trips.

Methodology

Parameterizing Link Travel Time Distribution

We model the travel time $t_{l,i,q}$ for link l, day i, and trip q by a hierarchical structure to capture multi-level variations in travel time:

$$t_{l,i,q} = \mu_l + \eta_{l,i} + \epsilon_{l,i,q}$$

- **Top level** μ_l: overall mean travel time for link l (work of traditional deterministic models).
- **Middle level** $\eta_{l,i}$: day-specific deviations (e.g., weather conditions, etc.).
- **Bottom level** $\epsilon_{l,i,q}$: trip-specific deviations (e.g., driver behavior, etc.).

This hierarchical formulation separates travel time variations into global (day-specific) and local (trip-specific) factors, enabling more granular and interpretable travel time modeling.

Methodology

Day-specific random effect $\boldsymbol{\eta}$

Assume the covariance:

$$Cov(\eta_{l,i}, \eta_{l',i'}) = \delta(i, i') \times \Sigma_d(i, i'),$$

- $\delta(i, i') = 1$ when $i = i'$, and 0 otherwise, Σ_d is used to characterize inter-trip correlations.
- the joint distribution of travel times on all links on day i:

$$x_i = \mu + \eta_i \sim N(\mu, \Sigma_d)$$

Model Σ_d with a low-rank parameterization:

$$\Sigma_d = L\omega_{\Sigma_d}\omega_{\Sigma_d}^T L^T.$$

The mean value μ is parameterized by:

$$\mu = L\omega_\mu r_\mu$$

- L is reprenstation of links, ω_{Σ_d}, ω_μ are learnable parameters to update the representaion, r_μ is the regression parameter used to generate the mean value.

Methodology

Trip-specific random effect ϵ

Assume the covariance:

$$Cov(\epsilon_{l,i,q}, \epsilon_{l',i',q'}) = \delta(i, i') \times \delta(q, q') \times \Sigma_p(l, l'),$$

- Σ_d is used to characterize intra-trip correlations.
- trip-level random effects are only correlated within the same trip and are independent between different trips.

Model Σ_d with a low-rank-plus-diagonal parameterization:

$$\Sigma_p = H\omega_{\Sigma_p}\omega_{\Sigma_p}^T H^T + D.$$

$$diag(D) = log(1 + exp(H\omega_d r_d))$$

- H is another set of reprenstation of links, ω_{Σ_p}, ω_d are learnable parameters to update the representaion, r_d is the regression parameter.

Methodology

Joint Distribution for Multiple Trips

- Link travel time $t_{l,i,q}$ is not observable in the raw data.
- Transformation matrix A_i, B_i construct the relationship between link travel time t_i and trip travel time τ_i:

$$A_i = [a_{q,l}],$$

$a_{q,l}$=1 if trip q uses link l, otherwise 0.

$$B_i = blkdiag(\{A_{i,q}\}),$$

- The joint distribution of τ_i:

$$\tau_i \sim N(A_i \mu, A_i \Sigma_d A_i^T + B_i (I_{Q_i} \otimes \Sigma_p) B_i^T)$$

$$A_i = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad B_i = \begin{bmatrix} 1\ 1\ 1\ 0\ 0\ 0\ 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0\ 0\ 1\ 1\ 1\ 0\ 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0\ 0\ 0\ 1\ 1\ 1\ 1 \end{bmatrix}$$

Methodology

Data Augmentation

- Sub-sampling trips to effectively utilize the whole GPS trajectory:

$$\begin{cases} T^1 = sub(T,\rho) = \{l_1,\ldots,l_{\rho|T|}\} \\ \ldots \\ T^k = sub(T,\rho k) = \{l_1,\ldots,l_{\rho k|T|}\} \end{cases}$$

Augment one sample into k new subsamples based on the sampling rate ρ.

- The update of transformation matrix $\widehat{A}_i, \widehat{B}_i$:

$$\hat{A}_i = \begin{bmatrix} 1\ 0\ 0\ 0\ 0\ 0\ 0 \\ 1\ 1\ 0\ 0\ 0\ 0\ 0 \\ 1\ 1\ 1\ 0\ 0\ 0\ 0 \\ 0\ 0\ 1\ 0\ 0\ 0\ 0 \\ 0\ 0\ 1\ 1\ 0\ 0\ 0 \\ 0\ 0\ 1\ 1\ 1\ 0\ 0 \\ 0\ 0\ 0\ 1\ 1\ 0\ 0 \\ 0\ 0\ 0\ 1\ 1\ 1\ 0 \\ 0\ 0\ 0\ 1\ 1\ 1\ 1 \end{bmatrix} \quad \hat{B}_i = \begin{bmatrix} \begin{matrix} 1\ 0\ 0\ 0\ 0\ 0\ 0 \\ 1\ 1\ 0\ 0\ 0\ 0\ 0 \\ 1\ 1\ 1\ 0\ 0\ 0\ 0 \end{matrix} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \begin{matrix} 0\ 0\ 1\ 0\ 0\ 0\ 0 \\ 0\ 0\ 1\ 1\ 0\ 0\ 0 \\ 0\ 0\ 1\ 1\ 1\ 0\ 0 \end{matrix} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \begin{matrix} 0\ 0\ 0\ 1\ 1\ 0\ 0 \\ 0\ 0\ 0\ 1\ 1\ 1\ 0 \\ 0\ 0\ 0\ 1\ 1\ 1\ 1 \end{matrix} \end{bmatrix}$$

- The joint distribution of τ_i after augmentation:

$$\tau_i \sim N(\widehat{A}_i\mu, \widehat{A}_i\Sigma_d\widehat{A}_i^T + \widehat{B}_i(I_{Q_i} \otimes \Sigma_p)\widehat{B}_i^T)$$

Methodology

Loss Function

The negative log-likelihood loss is used to optimize all trainable parameters in the model:

$$L_{pre}(\theta;\tau^*) \propto -\frac{1}{2}[log\ det(\widetilde{\Sigma}) + (\tau^* - \tilde{\mu})^T\widetilde{\Sigma}^{-1}(\tau^* - \tilde{\mu})]$$

- where $\tilde{\mu} = A_i\mu$, $\widetilde{\Sigma}$=$A_i\Sigma_d A_i^T + B_i(I_{Q_i} \otimes \Sigma_p)B_i^T$.

The low-rank-plus-diagonal parameterization of covariance can reduce the computational cost of the loss:

- Woodbury matrix identity

$$\widetilde{\Sigma}^{-1} = \Lambda^{-1} - \Lambda^{-1}V(I + V^T\Lambda^{-1}V)^{-1}V^TV^{-1}$$

- Matrix determinant lemma:

$$det(\widetilde{\Sigma}) = det(I + V^T\Lambda^{-1}V)det(\Lambda)$$

$$\Lambda = B_i(I_{Q_i} \otimes \Sigma_p)B_i^T, \quad V = A_iL$$

Methodology

Conditional Travel Time Estimation

Leveraging the learned covariance matrix, the conditional distribution is refined based on spatiotemporally adjacent trip samples.

- The joint distribution of estimated error and observed values:

$$\begin{bmatrix} \eta \\ \tau_o \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ A_o\mu \end{bmatrix}, \begin{bmatrix} \Sigma_d & \Sigma_d A_o^T \\ A_o\Sigma_d & A_o\Sigma_d A_o^T + B_o(I \otimes \Sigma_p)B_o^T \end{bmatrix}\right)$$

- The conditional distribution of $\eta|\tau_o$:

$$\eta|\tau_o = v \sim N(\mu^*, \Sigma^*)$$

$$\mu^* = \Sigma_d A_o^T (A_o\Sigma_d A_o^T + B_o(I \otimes \Sigma_p)B_o^T)^{-1}(\mathrm{v} - A_o\mu)$$

$$\Sigma^* = \Sigma_d - \Sigma_d A_o^T (A_o\Sigma_d A_o^T + B_o(I \otimes \Sigma_p)B_o^T)^{-1} A_o\Sigma_d$$

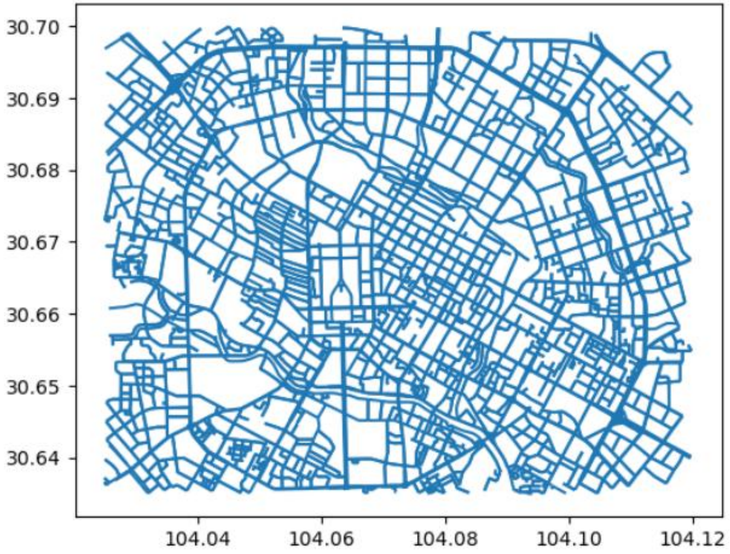
- The updated travel time distribution conditional on observed trips:

$$\tau|\tau_o = v \sim N(A(\mu + \mu^*), A(\Sigma^* + \Sigma_p)A^T)$$

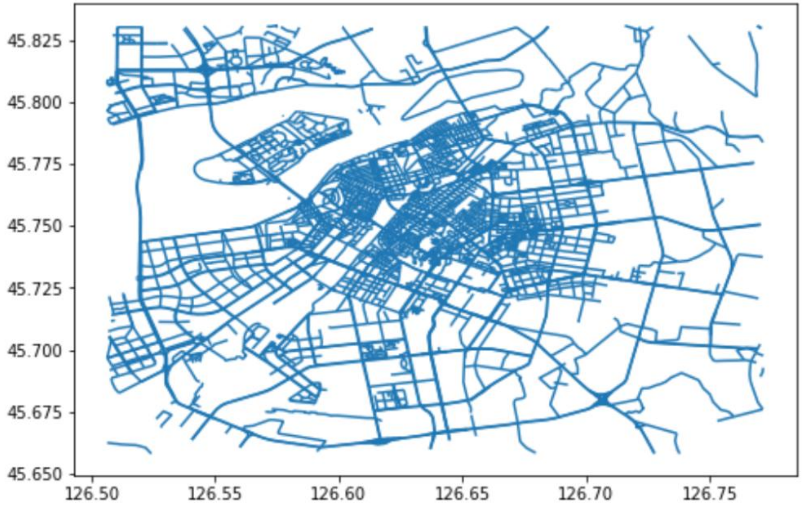
Experiment

Datasets

- **Chengdu:**
  - — 3,186 links
  - — 346,074 trip samples
  - — travel times range from 420s - 2,880s
  - — mean of 786 seconds



- **Harbin:**
  - — 8,497 links
  - — 1,268,139 trip samples
  - — travel times range from 420s - 2994s
  - — mean of 912 seconds



Experiment

Main results

MODEL PERFORMANCE COMPARISON ON CHENGDU AND HARBIN DATASETS.

Model	Chengdu				Harbin			
	RMSE	MAE	MAPE(%)	CPRS	RMSE	MAE	MAPE(%)	CPRS
DeepTTE	181.31	130.10	17.20	—	224.23	162.59	18.36	—
HierETA	155.26	111.34	14.68	—	187.93	136.45	15.62	—
MulT-TTE	<u>149.77</u>	<u>105.63</u>	<u>13.89</u>	—	178.39	129.81	14.86	—
DeepGTT	165.17	118.68	15.65	1.46	191.23	143.97	16.41	1.56
GMDNet	151.43	107.73	13.97	<u>1.31</u>	<u>176.98</u>	<u>128.11</u>	<u>14.65</u>	<u>1.41</u>
ProbETA†	134.79	95.11	12.31	1.18	156.94	112.15	12.82	1.24
ProbETA	**131.25**	**93.73**	**12.14**	**1.15**	**153.27**	**111.14**	**12.71**	**1.22**
Improvement	12.37%	11.27%	12.60%	12.21%	13.40%	13.25%	13.24%	13.48%

- In the Chengdu dataset, ProbETA outperforms the deterministic baselines by over 1.75% in MAPE and the probabilistic baselines by over 1.83% in MAPE and over 0.16 in CRPS.
- In the Harbin dataset, ProbETA outperforms the deterministic baselines by over 2.15% in MAPE and the probabilistic baselines by over 1.94% in MAPE and over 0.19 in CRPS.

Experiment

Ablation and Visualization Study

We constructed three variants of ProbETA:

- ProbETA-w/o MT: Remove multi-trip modeling.
- ProbETA-w/o DA: Remove data augmentation.
- ProbETA-w/o TD: Remove time discretization.

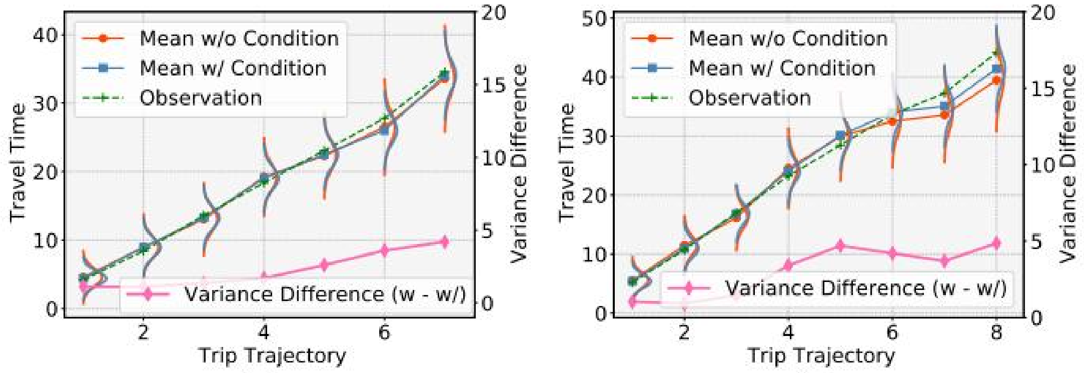
ABLATION EXPERIMENT ON CHENGDU/HARBIN DATASET.

Model	RMSE (s)	MAE (s)	MAPE (%)	CRPS
w/o MT	155.63/187.67	110.06/134.77	14.25/15.30	1.39/1.47
w/o DA	137.82/161.98	97.75/115.34	12.66/13.15	1.20/1.27
w/o TD	135.37/160.10	96.68/113.91	12.62/13.03	1.20/1.26
ProbETA	**131.25/153.27**	**93.73/111.14**	**12.14/12.71**	**1.15/1.23**

- Multi-trip joint modeling: primary source of advantage.
- Data augmentation and time discretization: a relatively small performance improvement.

Visualization of the mean and variance of travel times for a single trip in link level.

- completed trips are used as conditional information to adjust the distribution of the travel time for each link in the query trip

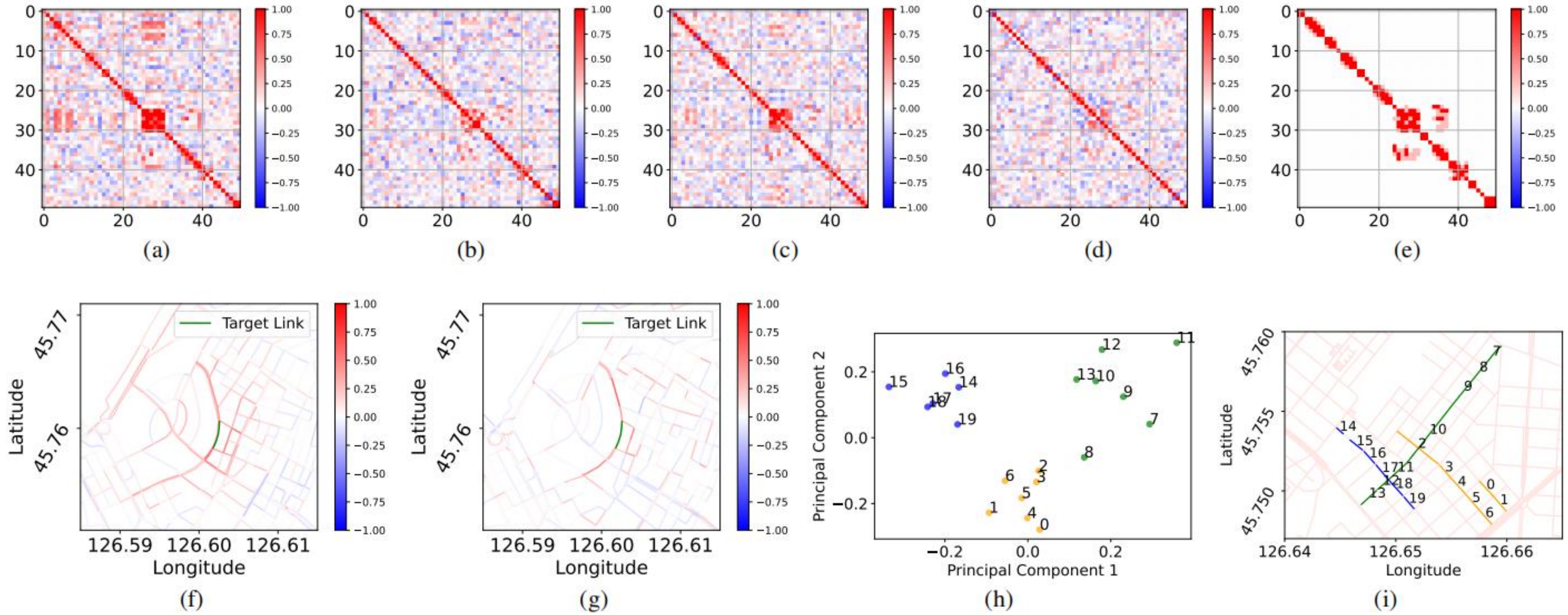


(a) (b)

Conditional information significantly reduces the prediction error and variance.

Experiment

Interpretability Analysis



- We visualized the learned link correlations Σ_d and Σ_p, then compared them with the corresponding 2-hop real adjacency matrix. The two exhibit a high degree of similarity.
- We utilized Principal Component Analysis (PCA) to project the link vectors into a two-dimensional space, and compared them with their actual map locations. The vector's clustering is similar to their actual locations.

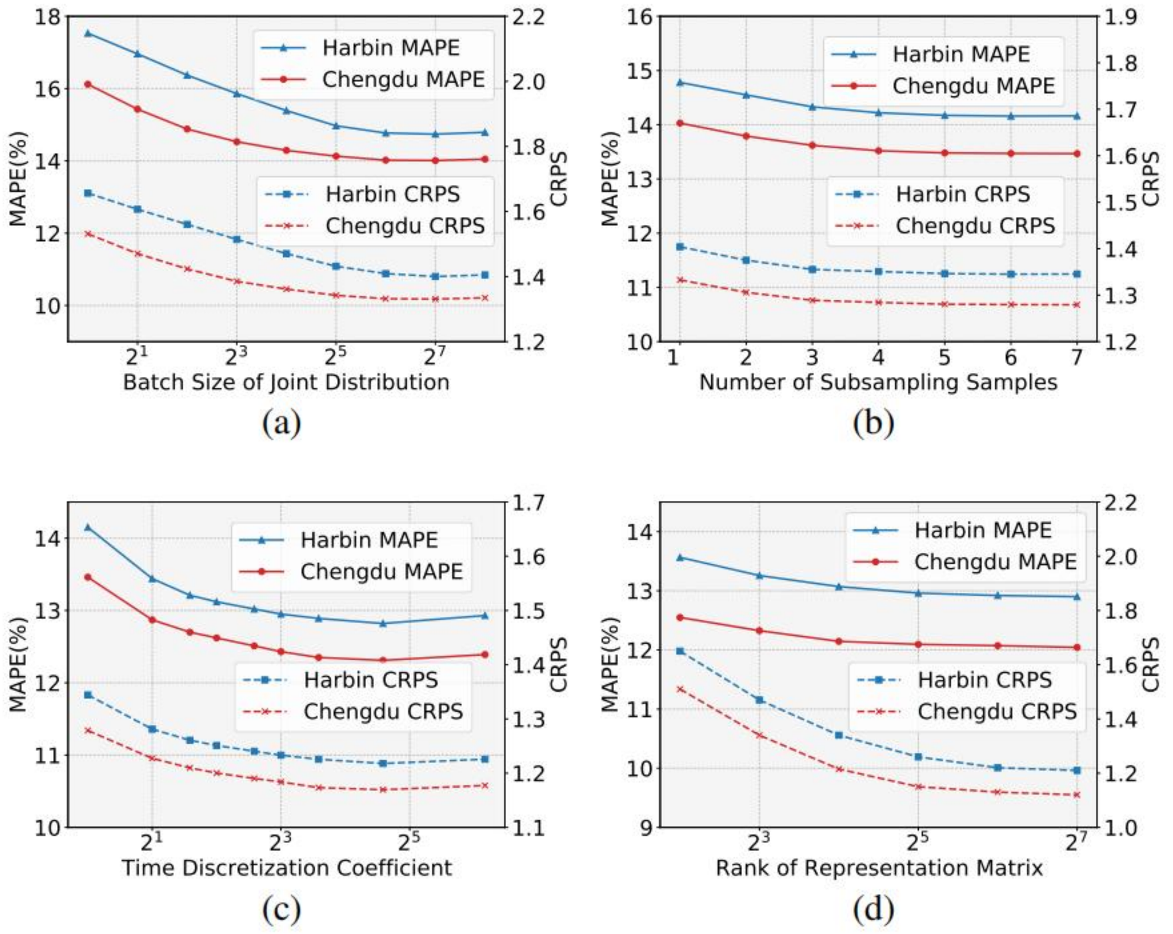
Experiment

Analysis of Model Parameters

Regarding the important parameters in the model, including

- the number of trips in the joint distribution,
- the amount of data augmentation,
- the level of temporal discretization,
- the value of rank,

we tested the model's performance under different values to determine the optimal settings for these parameters.



Conclusion

- ProbETA models the mean, day-specific random effects, and trip-specific random effects of trip travel times through a hierarchical structure. The low-rank plus diagonal covariance matrix construction enables efficient measurement of correlations between trips.
- By leveraging the learned covariance matrix and incorporating spatiotemporal proximity from historical trip samples, ProbETA can further refines the travel time distribution of the query trip.
- The data augmentation method based on subsampling data effectively improves the utilization of GPS data, which is beneficial for the individualized optimization of link representation vector learning in trips.
- Experiments have demonstrated that ProbETA improves trip time estimation performance, and the learned link correlations are interpretable.

Thank you