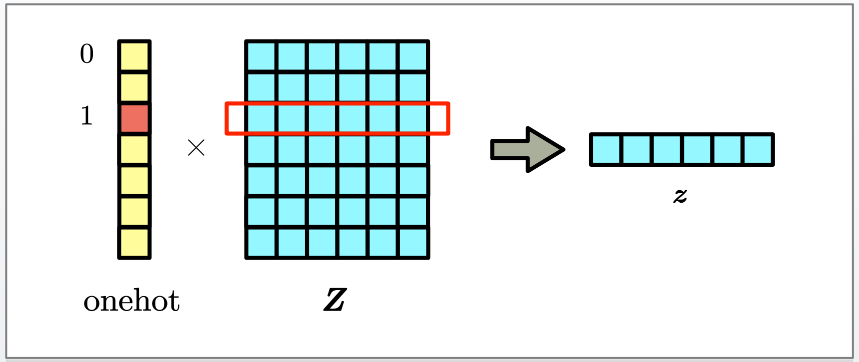
Pre-training Context and Time Aware Location Embeddings from Spatial-Temporal Trajectories for User Next Location Prediction

Yan Lin, BJTU
2021.2

Pre-training Location Embeddings



How FC embedding layer works

➢ Spatial-temporal trajectory mining

- Increasing availability of LBS data led to a burst of studies on location prediction

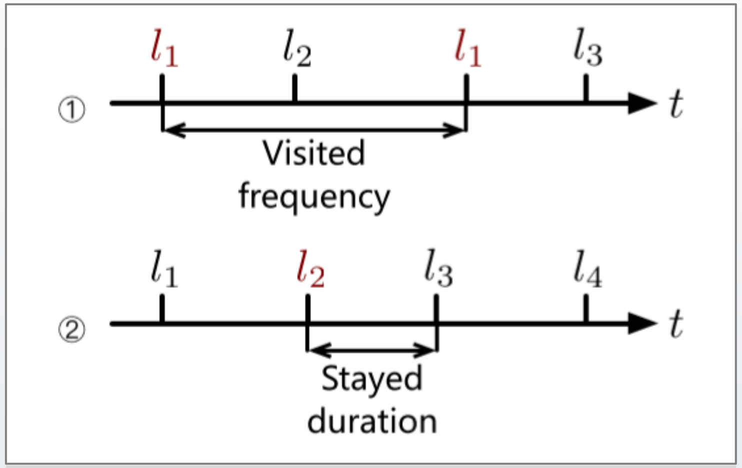
➢ Learning location embeddings

- A fundamental problem, essential for accurate prediction
- Fully-connected embedding layers
  - Hard to migrate to other models
  - Might suffer from over-fitting problems
- Pre-trained embedding vectors
  - Can be shared across various downstream models
  - Incorporate more general and comprehensive information

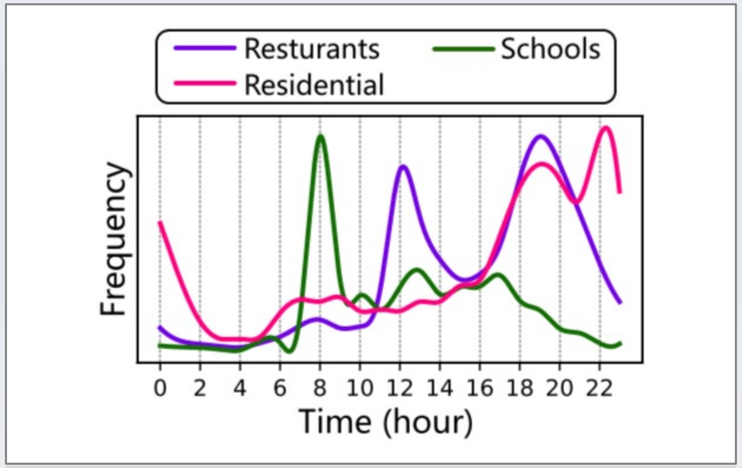
Pre-training Location Embeddings

- **Basic idea of pre-training location embeddings**
  - Borrowed from language modeling in NLP
  - Based on distributed word representations like word2vec[1]

- **Temporal information**
  - Relative visited time difference between locations: reflect visited frequency or stayed duration
  - Absolute visited time: reflect locations' functionalities
  - Incorporated by multiple location embedding methods[2,3], yet rarely simultaneously considered



Relative visited time difference

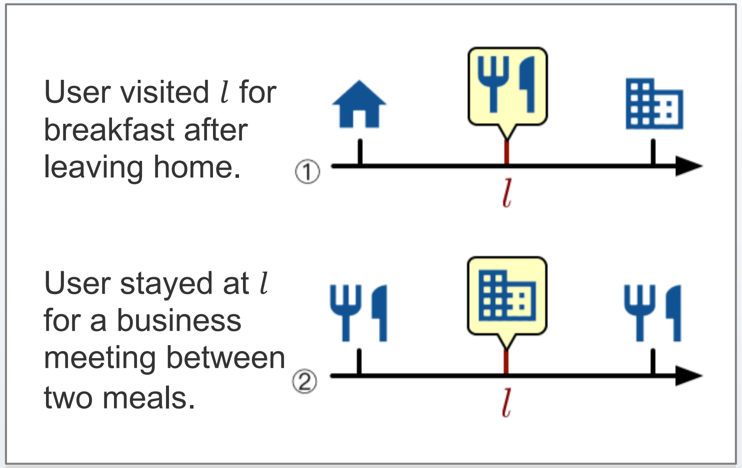


Absolute visited time distributions

Pre-training Location Embeddings

➢ Multi-functionality of locations

- People may visit the same location for different purposes in dissimilar contextual neighbors
- Incorporate the specific context of a location into embedding methods can yield higher quality representations
- Distributed representation-based methods mix a location's various functionalities into one latent vector

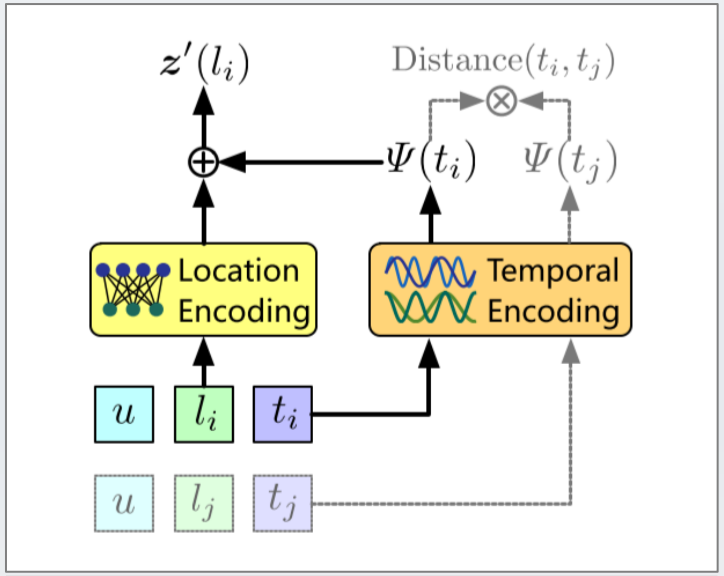


Different functionalities of a location in dissimilar contextual neighbors

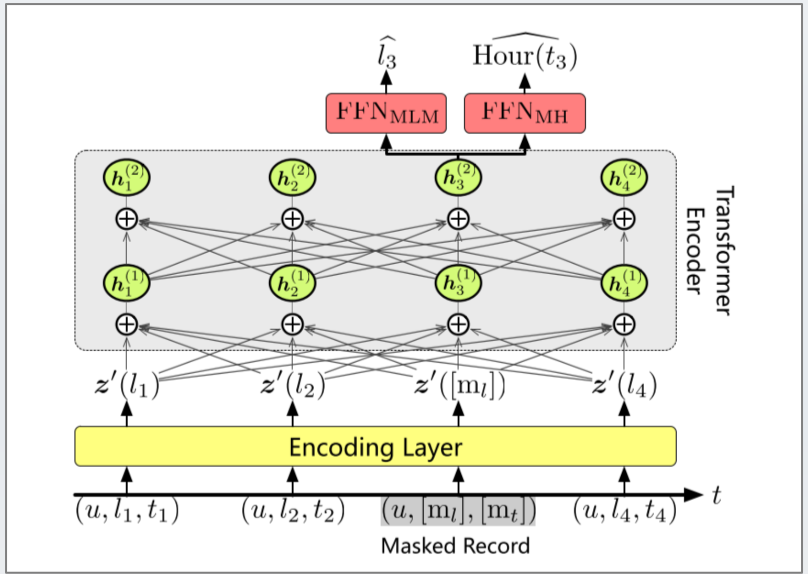
Overview of the Proposed Method

➢ Context and Time Aware Location Embedding (CTLE)

- Calculate target location’s embedding by a mapping function of its contextual neighbors
  - Incorporate context-specific functionalities
- Incorporate two aspects of temporal information
  - Further improve embedding quality



Encoding layer



Overall architecture

Embed Locations into Latent Vectors

➢ Necessity of location embedding

- Feature-based models require locations begin represented by latent vectors

➢ Fully-connected embedding layers

- Randomly initialize one latent vector for each location
- Trained with task-specific objectives
- Hard to migrate to other models
- Suffers from over-fitting when dealing with small-scale data

Pre-trained Location Embeddings

➢ Pre-trained latent representation

- Trained with un-supervised or self-supervised objectives
- Common practice in Language Process and Computer Vision

➢ Pre-training location embeddings from trajectories

- Akin to pre-training word embeddings from sentences
- Model the co-occurrence of targets and contexts to extract functionality information

Sentence

The proposed method can incorporate temporal information ...

The proposed method can fuse temporal information ...

context **target** **context**

Trajectory

Beijing Jiaotong University→ Xi Zhi Men → Peking University

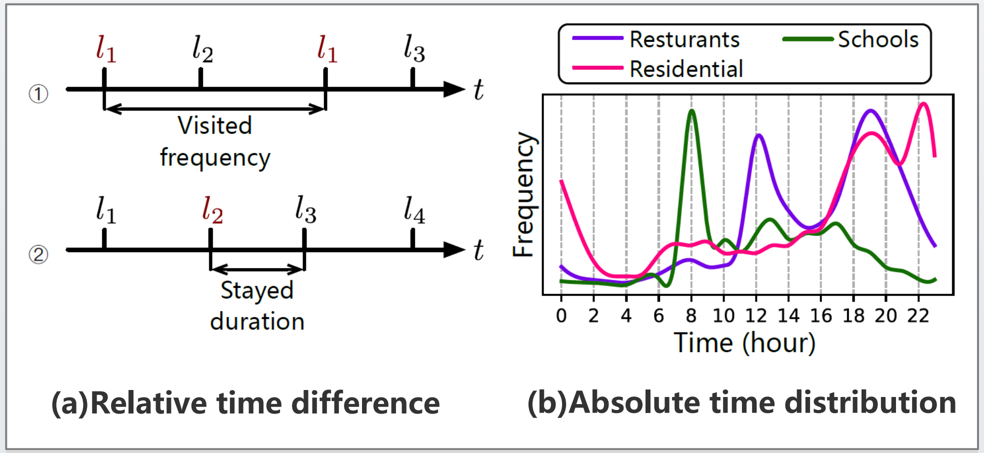
Beijing Jiaotong University→Tian An Men→ Peking University

Similarity between sentences and trajectories

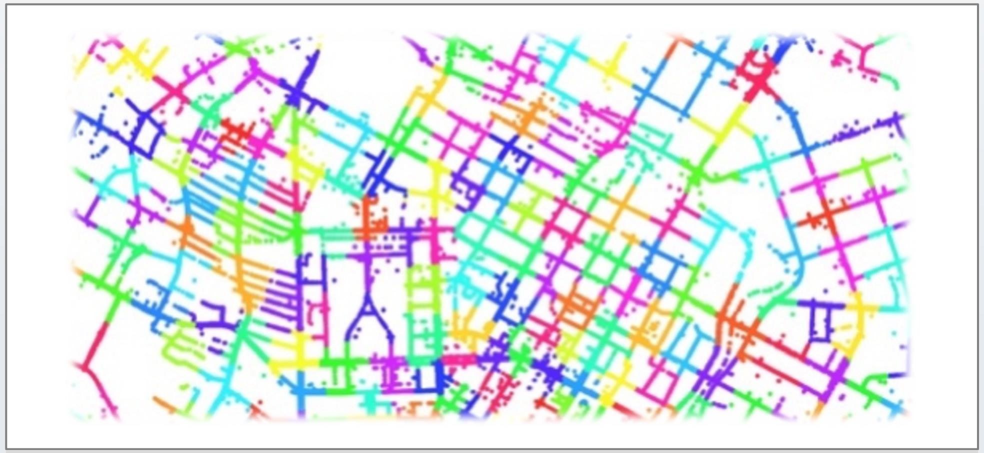
Uniqueness of Trajectories

➢ **Temporal correlation**

- Extract more accurate characteristic information



Temporal information in trajectories



Spatial information in trajectories

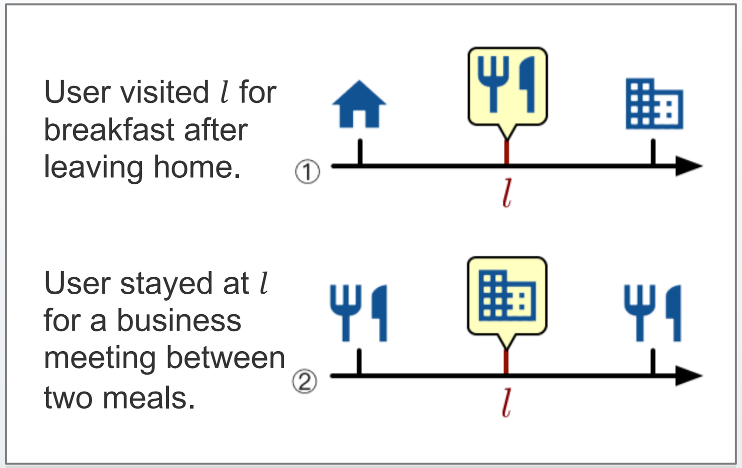
Multi-functionality of Locations

➢ Multi-functional locations

- Common in real-world
- People' s one visit to a location is still single-purposed
- Indicate the specific functionality through contexts

➢ Inspiration from NLP

- Tackle the problem ambiguity of words: contextual embedding
- Migrate to location embedding: calculate representation vectors based on targets' specific contextual neighbors



Different functionalities of a location in dissimilar contextual neighbors

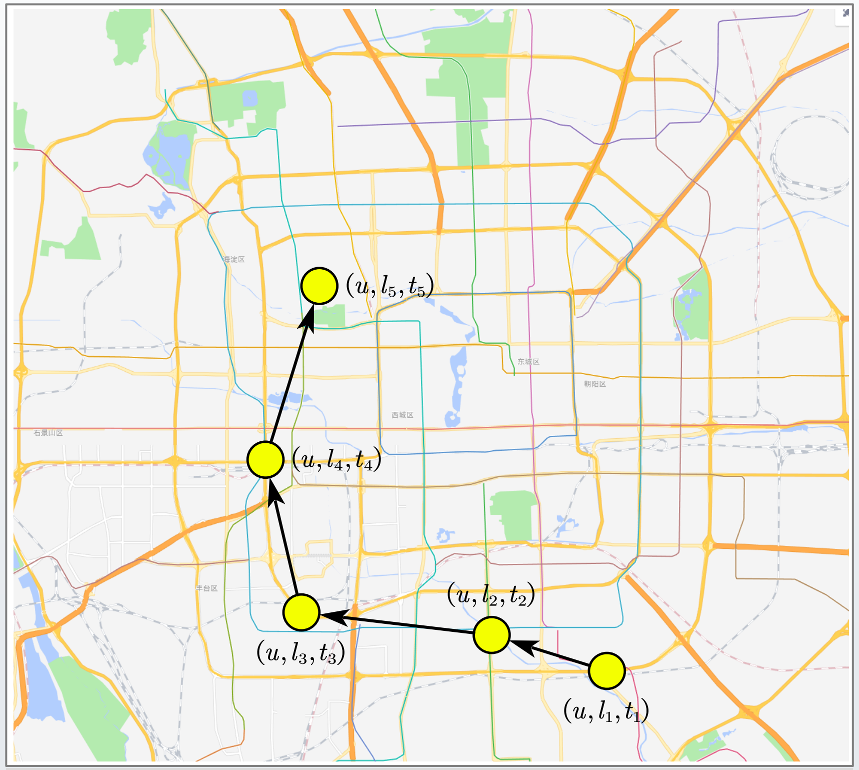
Preliminary Definitions

➢ Spatial-Temporal trajectory

- A user’s movements during a certain period, represented by a sequence of visiting records s
- A visiting record (u, l, t): user u visited location l at time t

➢ Problem definition

- Pre-train a mapping function f
- Object: generate a contextual embedding vector $\boldsymbol{z}(l)$ for target location l
- Given: the location’s context $C(l)$



A trajectory generated by user

Context-aware Location Embedding

➢ Construct mapping function f

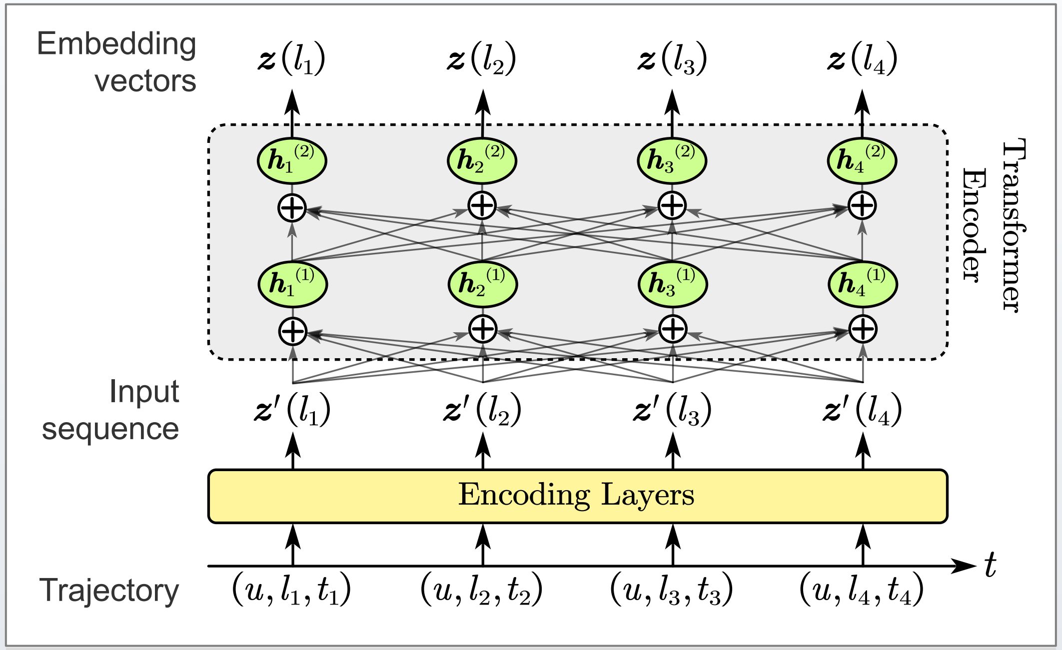
- Calculate embedding vector $\boldsymbol{z}(l)$ for location l via:

  $\boldsymbol{z}(l) = f\big(l, C(l)\big)$①
- Implement function f utilizing a bidirectional Transformer[4] Encoder

➢ Calculate embedding vectors

- Given: trajectory $s = \{(u, l_1, t_1), (u, l_2, t_2), \dots, (u, l_n, t_n)\}$
- Define all locations in s except l_i as its context $C(l_i)$
- Fetch an input vector $\boldsymbol{z}'(l)$ for each location l: $\boldsymbol{z}'(l) = \Omega(l)$

  Forming input sequence $\{\boldsymbol{z}'(l_1), \boldsymbol{z}'(l_2), \dots, \boldsymbol{z}'(l_n)\}$
- Feed the Transformer encoder with input sequence
  - Regard i-th item of output memory sequence $\boldsymbol{h}_i^{(N)}$ as $\boldsymbol{z}(l_i)$
- Corresponds to the abstract form ①

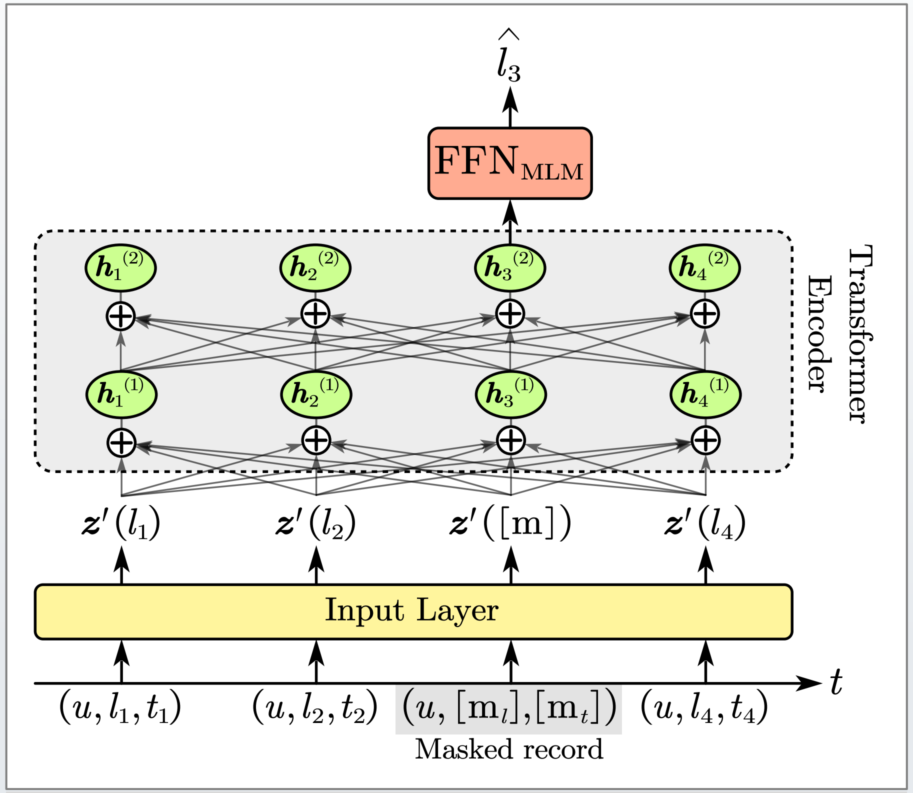


Calculate embedding vectors

Context-aware Location Embedding

➢ Pre-training mapping function f

- Make f understand correlation between targets and contexts
- Mask a portion of the trajectory, then predict the masked locations
  - Replace the masked visiting records with $(u, [m_l], [m_t])$
  - Predict the original locations using a fully-connected network
- An implementation of the Masked Language Model[5]

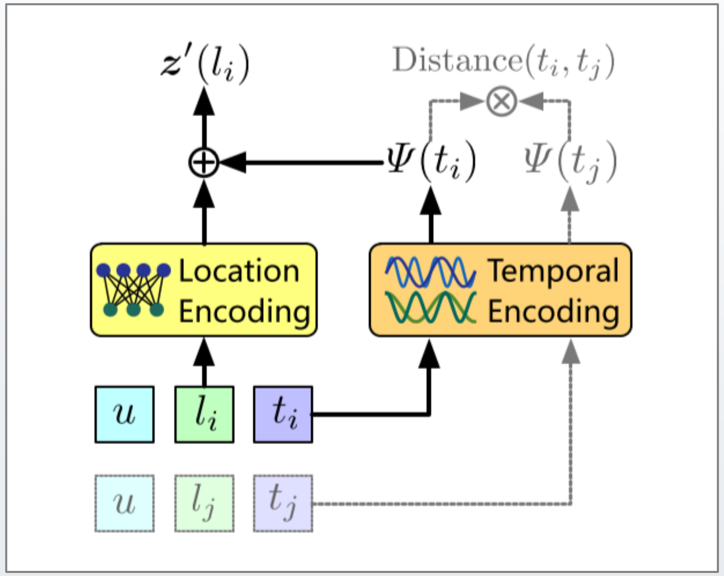


Pre-train the model using MLM

Incorporating Temporal Information

➢ Temporal Encoding

- Incorporate relative visited time difference
- A small modification of positional encoding
- $\Psi(t) = [\cos(\omega_1 t), \sin(\omega_1 t), \ldots, \cos(\omega_d t), \sin(\omega_d t)]$
  - $\Psi(t) \cdot \Psi(t+\delta) = \cos(\omega_1 \delta) + \cdots + \cos(\omega_d \delta)$, means the distance between two encoding vectors is learnable, and is independent of t
- Modify the calculation of input vector to $\boldsymbol{z}'(l) = \Omega(l) + \Psi(t)$

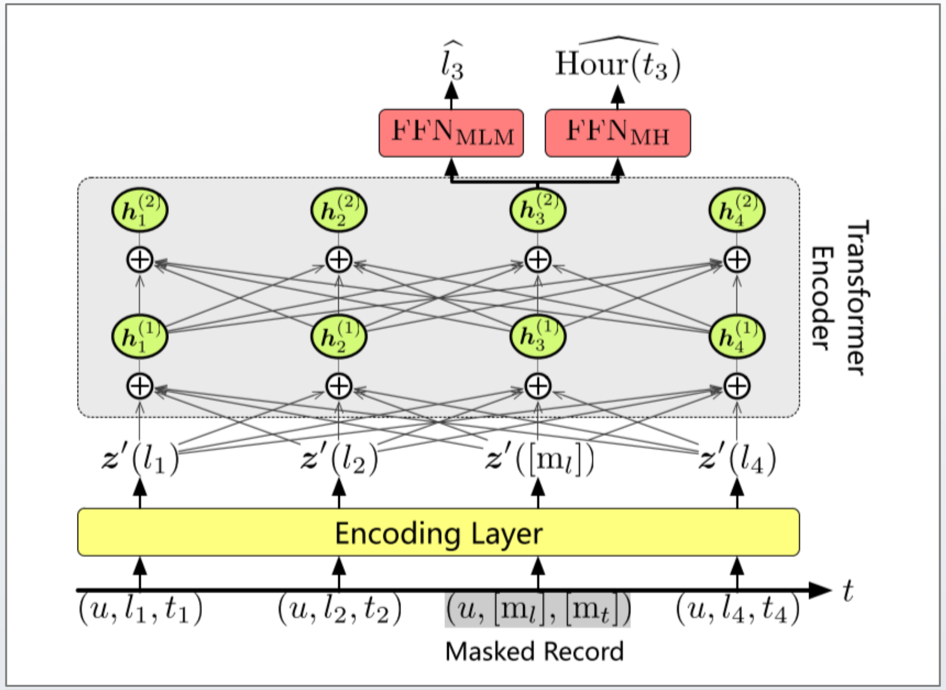


Encoding layer with temporal encoding

Incorporating Temporal Information

➢ Masked Hour Pre-training Objective

- Incorporate absolute visited time
- Predict the original visited hour index for a masked visiting record
- Forms a multi-task learning objective with the prediction of location index



Pre-train the model with MLM and MH

Dataset and Pre-process

➢ Dataset

- Two real-world mobile signaling data from Beijing and Shenyang, denoted as mobile-PEK and mobile-SHE
- Record users' movements through switching events between telecommunication base stations
  - Base stations → Locations
  - Users' entering to service area → Visiting records

Dataset	#Users	#Locations	#Records	Time span
Mobile-PEK	12,691	7,279	1,383,422	5 days
Mobile-SHE	10,564	7,201	607,581	11 days

Statistics of datasets

➢ Pre-process

- Filter out passing points
  - Remove visiting records with short duration
- One trajectory: a user's record sequence in one day

Baselines

➢ Baseline Location Embedding Methods

- Two classic distributed embedding models and some state-of-the-art location embedding models

Baselines	Pre-trained	Extra information
FC Layer	No	N/A
Skip-Gram	Yes	N/A
POI2Vec	Yes	Spatial
Geo-Teaser	Yes	Temporal + Spatial
TALE	Yes	Temporal
HIER	Yes	Temporal

➢ Downstream Prediction Models

- User next location prediction models
- Utilized for evaluating the effectiveness of embedding methods

Models	Based-on	Extra information
ST-RNN	RNN	Spatial and temporal intervals
ERPP	LSTM	Visited timestamp
ST-LSTM	LSTM	Spatial and temporal intervals

Results and Analysis

Prediction Model		ST-RNN			ERPP			ST-LSTM		
Dataset	Metric / Embedding Method	Accuracy (%)	macro-Recall (%)	macro-F1 (%)	Accuracy (%)	macro-Recall (%)	macro-F1 (%)	Accuracy (%)	macro-Recall (%)	macro-F1 (%)
Mobile-PEK	FC Layer*	3.744±0.10	1.739±0.06	1.449±0.19	4.373±0.14	2.017±0.04	1.595±0.05	4.542±0.15	2.092±0.09	1.689±0.07
	Skip-gram	3.671±0.11	1.777±0.11	1.423±0.05	4.611±0.01	2.368±0.07	1.779±0.04	4.877±0.05	2.586±0.06	1.947±0.04
	POI2Vec	3.992±0.08	2.281±0.08	1.838±0.06	5.024±0.08	2.595±0.07	2.035±0.06	5.163±0.10	2.682±0.08	2.077±0.10
	Geo-Teaser	3.998±0.13	2.166±0.07	1.796±0.07	5.159±0.05	2.671±0.08	2.039±0.03	5.305±0.05	2.739±0.04	2.084±0.02
	TALE	4.199±0.05	2.240±0.07	1.815±0.06	5.457±0.03	3.237±0.07	2.587±0.03	5.511±0.05	3.152±0.10	2.486±0.13
	HIER	4.339±0.04	2.440±0.07	1.862±0.08	5.607±0.09	2.870±0.08	2.176±0.04	5.589±0.15	2.839±0.11	2.165±0.02
	CTLE (ours)	**5.068±0.05**	**2.890±0.11**	**2.312±0.02**	**6.481±0.05**	**4.002±0.04**	**3.066±0.06**	**6.473±0.09**	**4.072±0.13**	**3.097±0.13**
Mobile-SHE	FC Layer*	3.674±0.07	2.408±0.07	1.946±0.05	4.343±0.18	2.454±0.10	2.037±0.09	4.416±0.20	2.450±0.14	2.005±0.11
	Skip-gram	3.646±0.05	2.278±0.08	1.809±0.05	4.405±0.06	2.459±0.06	1.974±0.05	4.508±0.05	2.507±0.07	1.998±0.07
	POI2Vec	3.936±0.04	2.605±0.04	2.084±0.03	4.923±0.06	2.992±0.02	2.408±0.02	4.930±0.07	2.890±0.10	2.305±0.08
	Geo-Teaser	4.006±0.05	2.455±0.03	1.897±0.02	4.932±0.12	2.895±0.02	2.410±0.07	5.130±0.15	2.754±0.07	2.245±0.06
	TALE	4.689±0.10	**3.444±0.09**	**2.761±0.08**	5.179±0.09	3.446±0.06	2.883±0.04	5.204±0.06	3.399±0.11	2.787±0.11
	HIER	4.539±0.22	3.117±0.15	2.521±0.09	5.624±0.16	3.273±0.17	2.708±0.18	5.672±0.09	3.252±0.07	2.680±0.05
	CTLE (ours)	**5.124±0.20**	3.392±0.11	2.720±0.07	**6.311±0.04**	**3.984±0.05**	**3.340±0.07**	**6.325±0.08**	**3.950±0.11**	**3.291±0.06**

Prediction performance comparison of different approaches

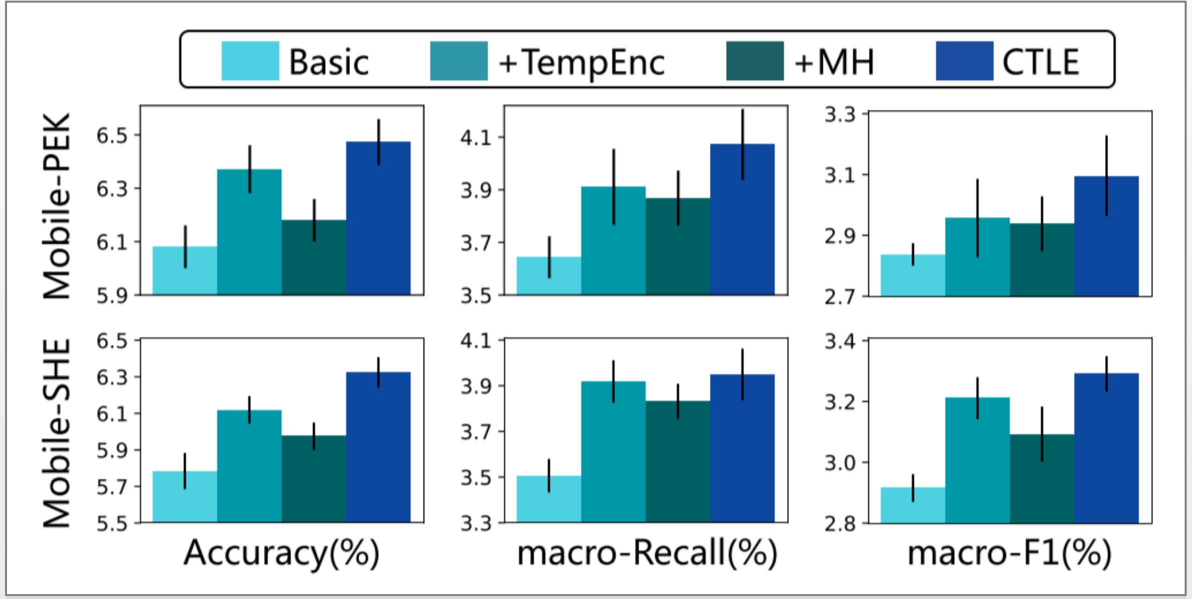
- Incorporating temporal information can improve representational quality
- Context-aware embedding can provide more accurate representation for multi-functional locations

Results and Analysis

➢ Component analysis

- Comparing our proposed method with three of its variants

Variants	Encoding	Pre-train objective
Basic	Positional Encoding	MLM
+TempEnc	Temporal Encoding	MLM
+MH	Positional Encoding	MLM+MH
CTLE	Temporal Encoding	MLM+MH



Component analysis results

- Combine two aspects of temporal information can be beneficial

Future Works

End

- How to more accurately model temporal information in absolute visited time distribution?
- How to further incorporate spatial correlation?

Thank you for your time!