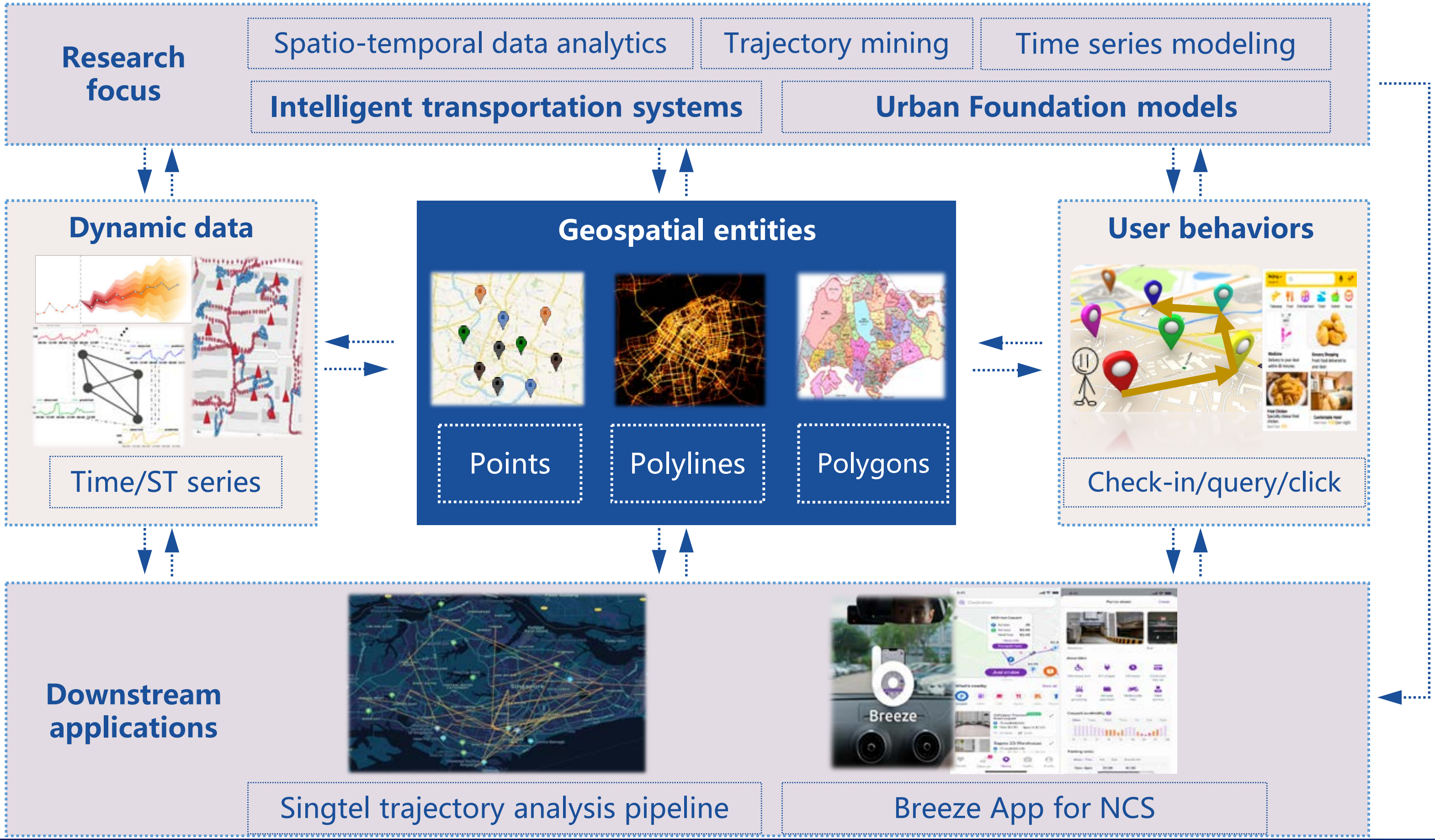
Introduction to Research on Spatial-temporal Data Mining

Research Overview

- Urban intelligence
  - Spatio-temporal data mining and analytics, smart city, user modeling



Outline

- Spatial-temporal data mining
  - Spatial relationship extraction
  - Geospatial IR or Spatial Keyword Search
  - POI recommendations
  - Road Network Representation for Road Network Applications
  - Region Representation for Region-Level Applications
- Trajectory data mining
  - Application in intelligent transportation
- Application of Foundation Models for Geospatial Applications

Geospatial database

Name	Lat	Long	Address	Category
National Museum	1.29682	103.84877	93 Stamford Rd, 178897	Museums
Food for Thought	1.2963	103.84876	93 Stamford Road #01-04, National Museum, 178897	Asian Fusion
Museum Carpark	1.296509	103.84794		Parking
Harry's	1.2976	103.84905	90 Stamford Rd, 178903	Bars
Food for Thought	1.29675	103.8486		Restaurant

Geospatial DB

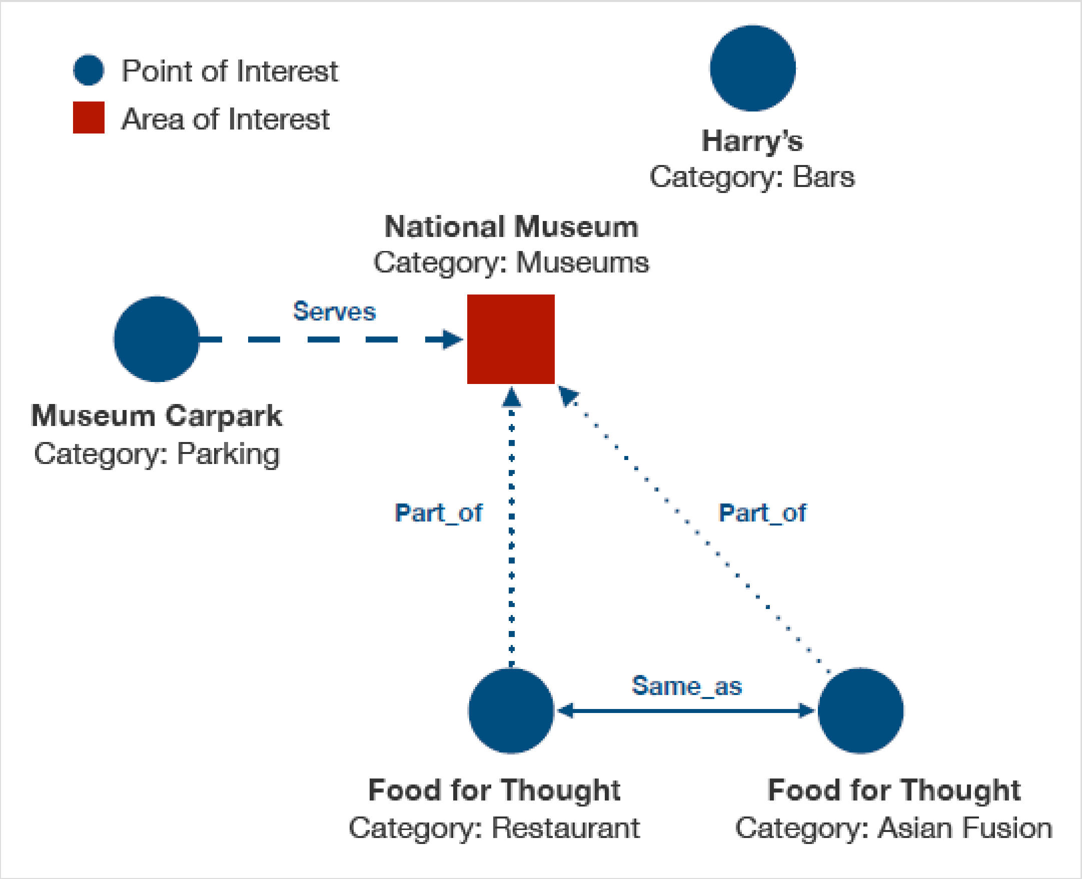
Although convenient, this representation hinders the exploration of **geospatial relationships** between the entities

Pasquale Balsebre, Dezhong Yao, Gao Cong, Weiming Huang, Zhen Hai: Mining Geospatial Relationships from Text. SIGMOD 2023

Geospatial KG

- Relationships between the entities exist and can be captured in a KG representation
- **Knowledge Graphs** are ubiquitous today and offer several advantages:
  - Machine-readable format
  - Can represent both entities and their relations
  - Widely adopted in AI applications

- Existing geoKGs represent only *coarse-grained* relationships

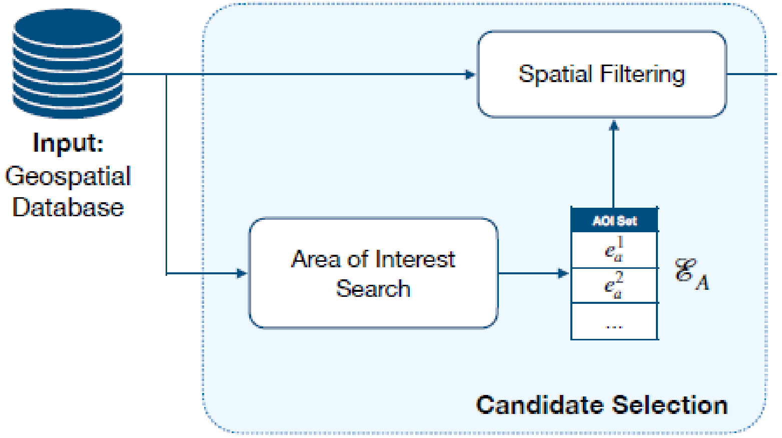


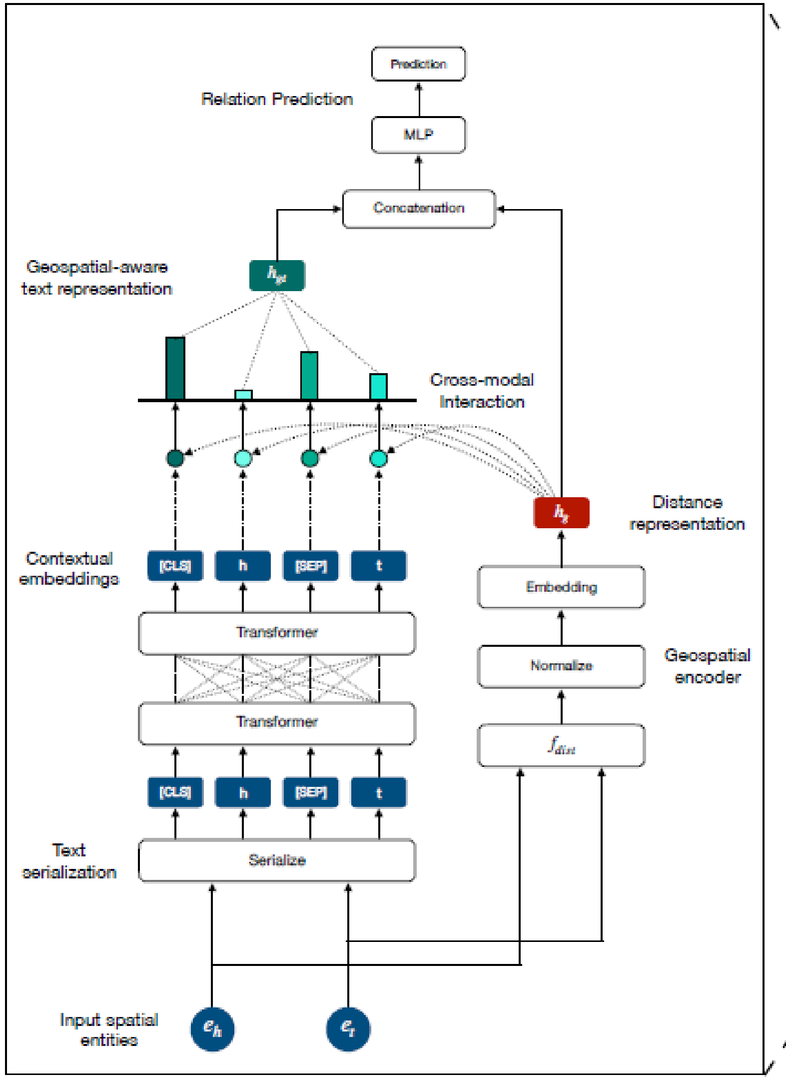
WorldKG

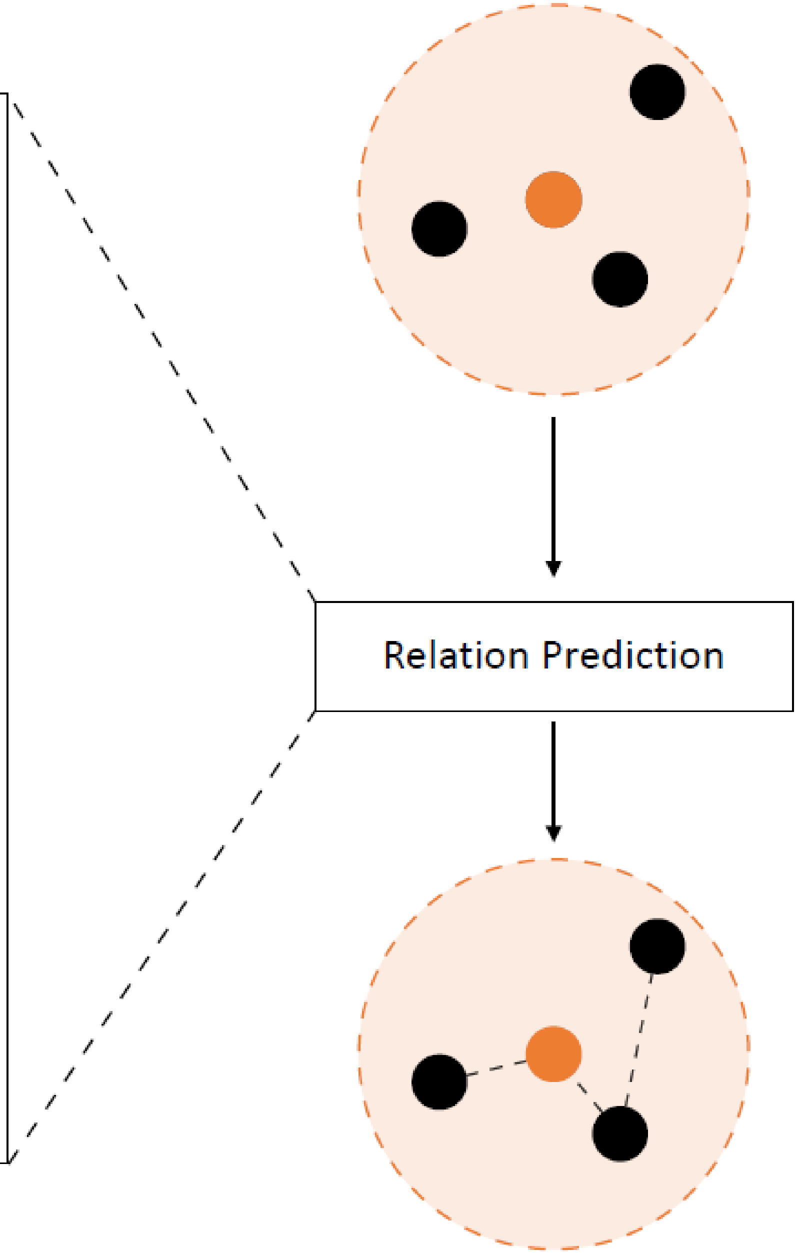
YAGO2Geo

DBPedia

Proposed solution







- **Candidate Selection Step**: Aim
  relationships
- **Relation Prediction:** Aim at ide
- **The KG refinement**: Aim to ext
  correctness

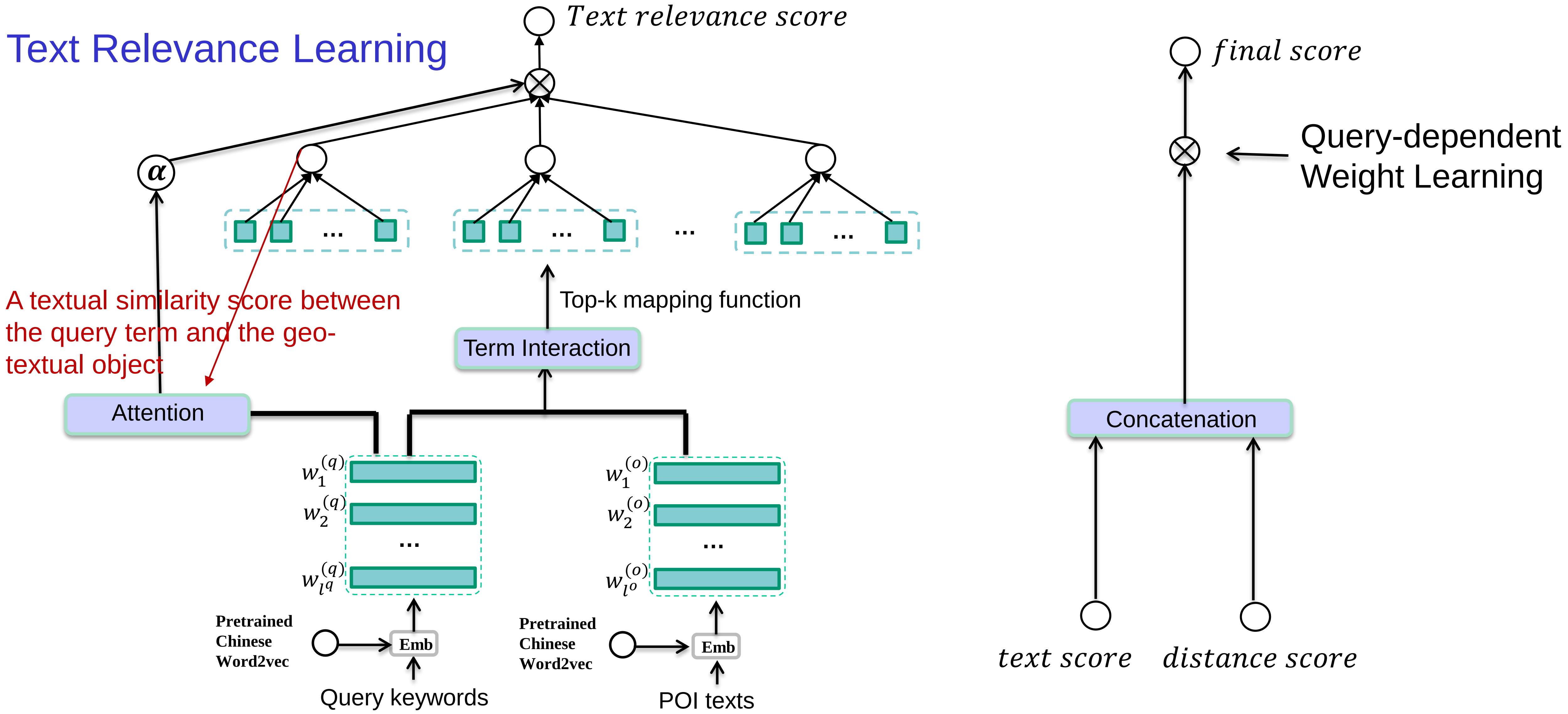
Spatial Keyword Query (Geographic IR)

- Take query keywords and location as input and output retrieved objects/documents

- Applications of spatial ke
  - o Geographic search e
  - o location-based servi
  - o locally targeted web

Spatial Keyword Query Example on Yelp (or Meituan)

Geospatial entity representation learning

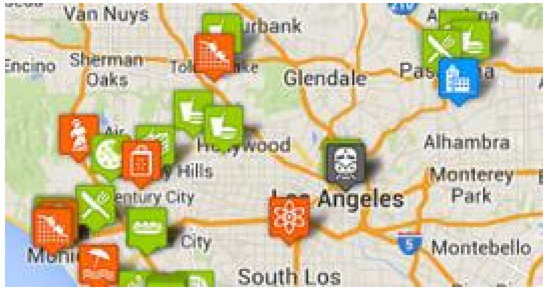


Shang Liu, Gao Cong, Kaiyu Feng, Wanli Gu, Fuzheng Zhang: Effectiveness Perspectives and a Deep Relevance Model for Spatial Keyword Queries. SIGMOD 2023

POI recommendation

- Given a set of POIs, and a set of users each associated with a set of visited POIs, POI recommendation is to recommend for each user new POIs that are likely to be visited.

A large number of POIs

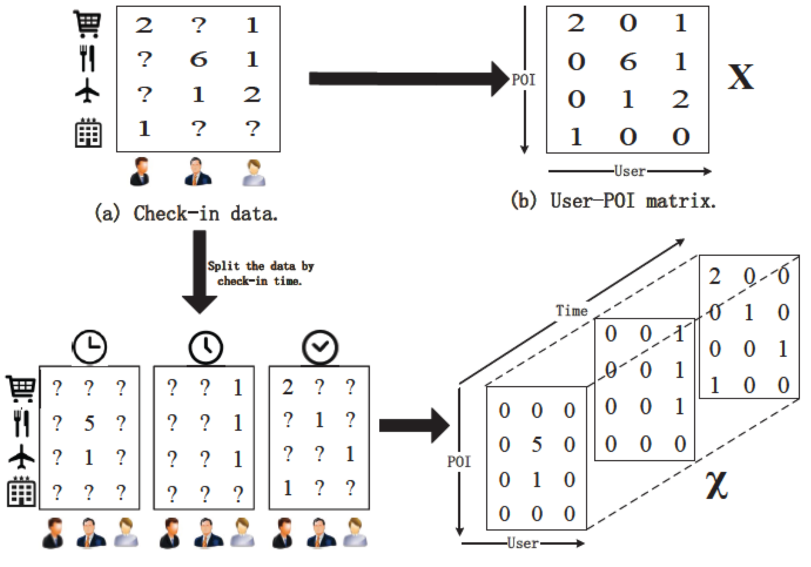


POI recommendation

Users with different interests

Quan Yuan, Gao Cong, Aixin Sun, Chin-Yew Lin, Nadia M. Thalmann. Category Hierarchy Maintenance: A Data-Driven Approach. SIGIR 2012

New Types of POI recommendation



(a) Check-in data. (b) User-POI matrix. (c) Time-aware check-in data. (d) User-time-POI tensor.

Context-aware POI recommendation

- Context: time, current location.
- E.g., Workplace + Friday Evening → Restaurant / Bar

Requirement-aware POI recommendation (w/o Group)

- E.g., Mary wants to find a restaurant to have pizza **with her friend Bob** at 7:00 PM on Friday

Predict potential visitors for a POI (for ads)

- It can help POI owners to find potential customers for marketing
- E.g., given a POI restaurant, we want to predict potential consumers who would visit this restaurant in the next several hours

User-based CF (U)

- Assumption: the interests of the target user *u* can be estimated based on the check-in histories of other users who checked-in at similar POIs with *u*.

User-POI matrix $C^{(UL)}$

$c_{u,l}$	l_1	l_2	l_3	l_4
u_1	1	1	0	0
u_2	1	1	1	0
u_3	0	1	0	1

Check-in vector of u_1

- Two steps:
  - Calculate similarities between users

similarity: $$w_{u,v} = \frac{\sum_l c_{u,l} c_{v,l}}{\sqrt{\sum_l c_{u,l}^2}\sqrt{\sum_l c_{v,l}^2}}$$

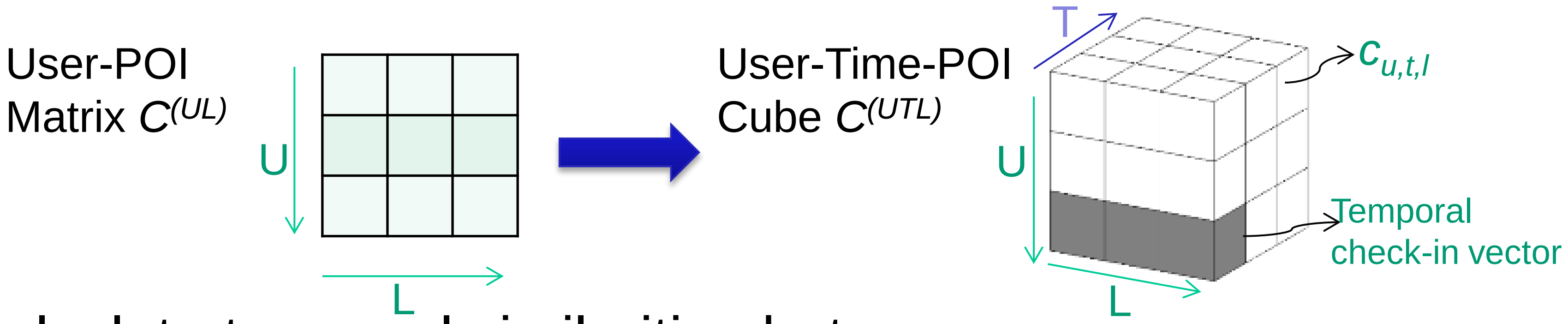
  - Produce prediction for each candidate POI *l*

score: $$\hat{c}_{u,l} = \frac{\sum_v w_{u,v} c_{v,l}}{\sum_v w_{u,v}}$$

Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, Nadia Magnenat-Thalmann: Time-aware point-of-interest recommendation. SIGIR 2013

User-based CF with Time Preference (UT)

- Introduce time dimension into the matrix:



- Calculate temporal similarities between users

$$w_{u,v} = \frac{\sum_l c_{u,l} c_{v,l}}{\sqrt{\sum_l c_{u,l}^2}\sqrt{\sum_l c_{v,l}^2}} \Rightarrow w_{u,v}^{(t)} = \frac{\boxed{\sum_{t=1}^{T}} \sum_l c_{u,t,l} c_{v,t,l}}{\sqrt{\sum_{t=1}^{T} \sum_l c_{u,t,l}^2}\sqrt{\sum_{t=1}^{T} \sum_l c_{v,t,l}^2}}$$

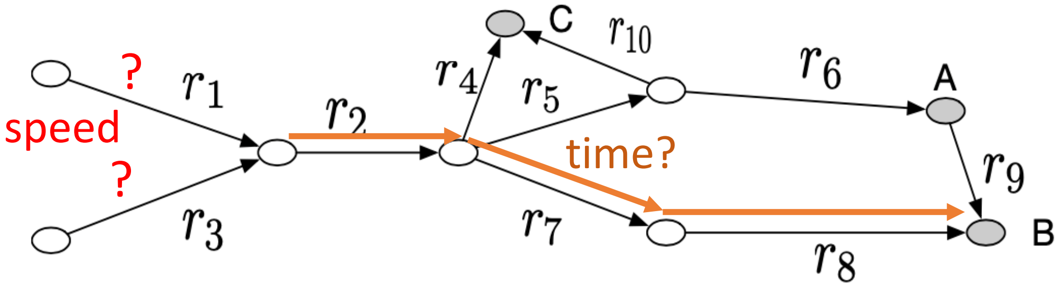
- Produce temporal predictions

$$\hat{c}_{u,l} = \frac{\sum_v w_{u,v} \boxed{c_{v,l}}}{\sum_v w_{u,v}} \Rightarrow \hat{c}_{u,t,l} = \frac{\sum_v w_{u,v}^{(t)} \boxed{c_{v,t,l}}}{\sum_v w_{u,v}^{(t)}}$$

- The recommendation score is calculated based on the check-ins at target time *t*.

Geospatial Entities

- Road networks



- Applications:
  - Generate **generic representations** for various types of road network applications.
    - ❑ Trajectory-based: travel time estimation, similarity search
    - ❑ Road segment-based: traffic inference, road attribute inference

Geospatial Entities

- Results:
  - Road networks and trajectories from two cities

Dataset	#Road Segments	#Edges	#Trajectories
Chengdu	4,885	12,446	677,492
Xi'an	5,052	13,660	373,054

  - Downstream applications:
    - Road label classification
    - Traffic inference
    - Trajectory similarity search
    - Travel time estimation

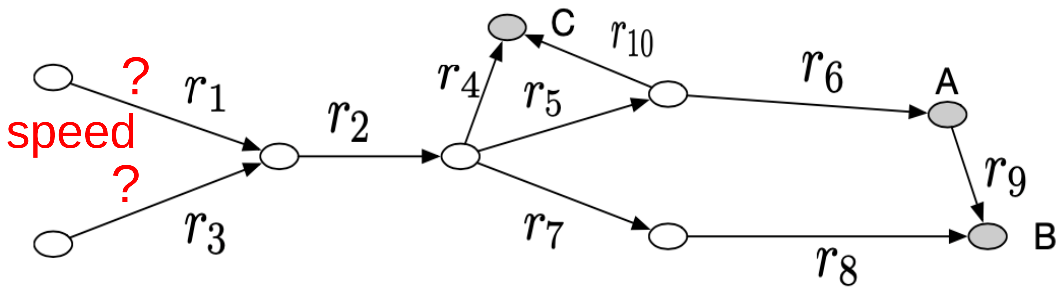
Road segment-based

Trajectory-based

Experiments

▷ **Road segment-based application result:**

Task	Road Label Classification				Traffic Inference			
	Chengdu		Xi'an		Chengdu		Xi'an	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	MAE	RMSE	MAE	RMSE
DW	0.522	0.493	0.552	0.524	7.32	9.14	6.78	8.57
node2vec	0.524	0.495	0.586	0.559	7.12	9.00	6.41	8.22
GAE	0.432	0.328	0.447	0.339	6.91	8.72	6.41	8.39
GraphSAGE	0.452	0.324	0.466	0.347	6.48	8.52	6.12	7.98
RFN	0.516	0.484	0.577	0.570	6.89	8.77	6.57	8.43
IRN2Vec	0.497	0.458	0.531	0.506	6.52	8.52	6.60	8.59
HRNR	0.541	0.527	0.631	0.609	7.03	8.82	6.52	8.45
Toast	**0.602**	**0.599**	**0.692**	**0.659**	**5.95**	**7.70**	**5.71**	**7.44**



Experiments

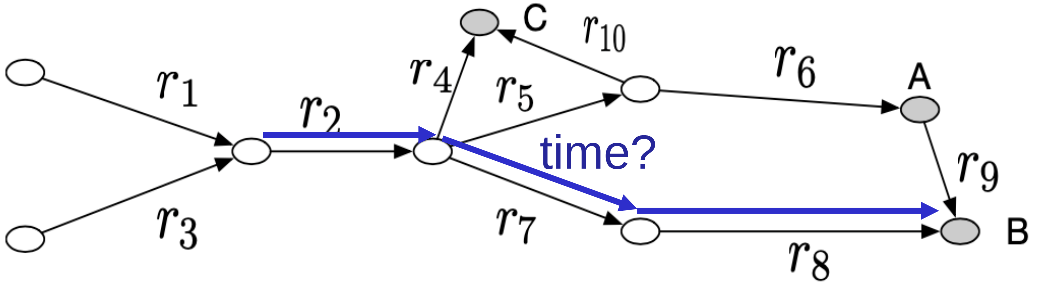
▷ **Trajectory-based application result**

Trajectory similarity search

	Chengdu		Xi'an	
	MR	HR@10	MR	HR@10
para2vec	216.92	0.251	279.38	0.205
t2vec	46.17	0.781	38.67	0.806
LCSS	67.72	0.487	83.94	0.469
EDR	458.20	0.174	529.74	0.119
Fréchet	21.17	0.847	22.79	0.894
Toast	**10.10**	**0.885**	**13.71**	**0.905**

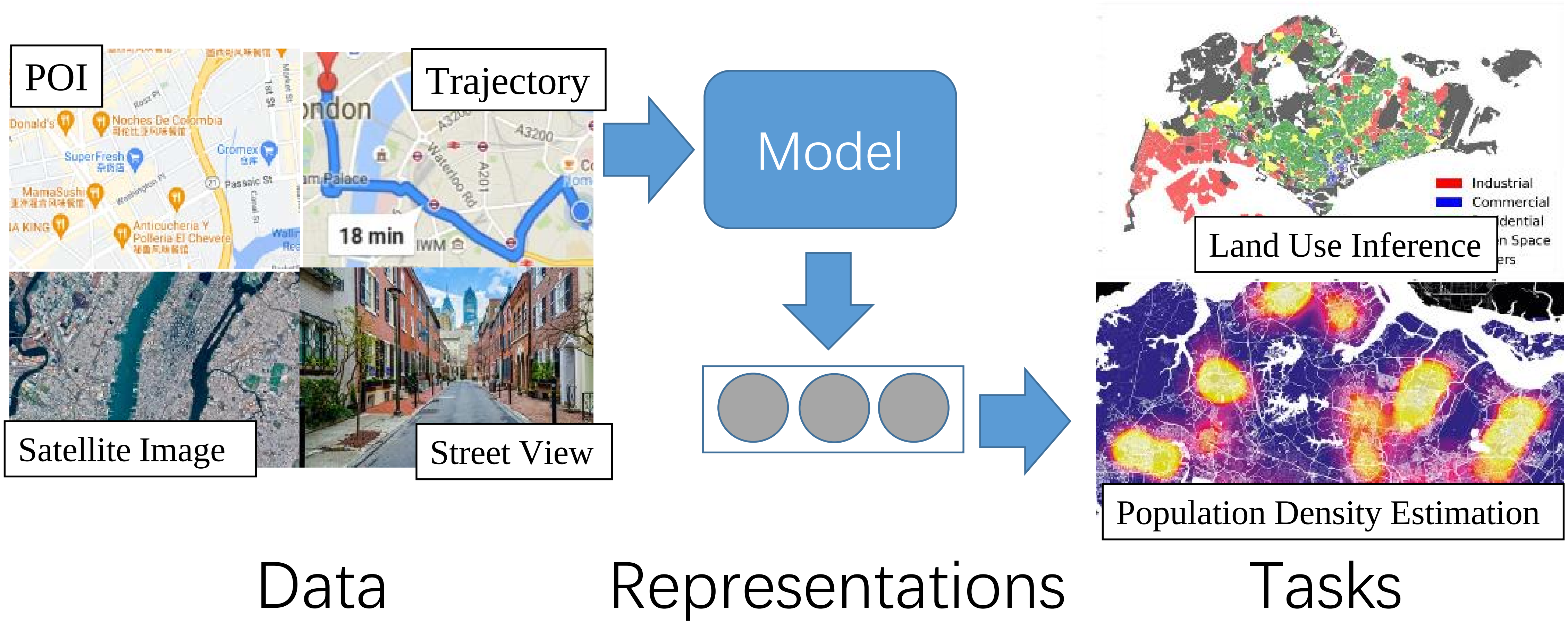
Travel time estimation

	Chengdu		Xi'an	
	MAE	RMSE	MAE	RMSE
para2vec	220.45	302.72	244.73	345.49
t2vec	165.18	240.72	207.56	311.04
Road-Pool	151.80	223.02	185.47	293.82
Toast	**127.80**	**190.86**	**175.68**	**265.09**



Problem of Urban Region Representation Learning

- Urban Region Representation Learning aims at learning effective feature vectors for urban regions to serve various downstream tasks.



Our motivations

An Example Building Group

(Singapore Public House)

We focus on **OSM buildings.**

- **Buildings**, (or formally, **building footprints),** refer to the 2-D building polygon on the map
  - size, height, type, name…
- **Building groups** refers to the **collection of buildings** in a defined spatial area.
  - We use OSM road networks to partition buildings into building groups.

Yi Li, Weiming Huang, Gao Cong, Hao Wang, and Zheng Wang. Urban Region Representation Learning with OpenStreetMap Building Footprints. SIGKDD 2023

Introduction

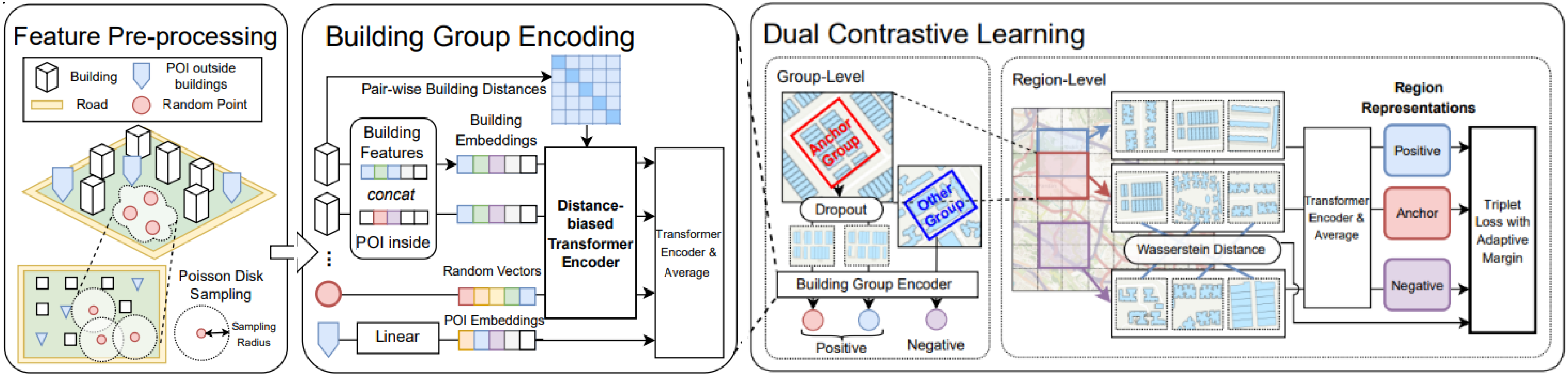
Example Building Groups with Specific Urban Functions

Comparing to other data types, building data has **advantages**:

- **Effectiveness**
  - Buildings directly carrying urban functions.
- **Availability**
  - Buildings are readily available in OSM

Method

1. **Partition** the city into building groups with road network.
2. **Encode** building groups with POIs and regions with Transformer-based encoders.
3. **Train** the encoder with Group-level and Region-level contrastive learning



Experiments: Land Use Inference

- Infer 5 types of land use (Residential, Industrial, Commercial, Open Space, Other)

Table 2: Land Use Inference in Singapore and New York City

Models	Singapore			New York City		
	L1↓	KL↓	Cosine↑	L1↓	KL↓	Cosine↑
Urban2Vec	0.657±0.033	0.467±0.043	0.804±0.017	0.473±0.018	0.295±0.015	0.890±0.007
Place2Vec	0.645±0.039	0.451±0.047	0.812±0.018	0.518±0.016	0.308±0.012	0.878±0.005
Doc2Vec	0.679±0.050	0.469±0.058	0.789±0.027	0.506±0.015	0.299±0.016	0.885±0.008
GAE	0.759±0.040	0.547±0.051	0.765±0.022	0.589±0.011	0.365±0.011	0.855±0.007
DGI	0.598±0.029	0.372±0.032	0.846±0.012	0.433±0.009	0.237±0.012	0.907±0.005
Transformer	0.556±0.046	0.357±0.070	0.850±0.026	0.436±0.020	0.251±0.018	0.903±0.008
RegionDCL-no random	0.535±0.054	0.321±0.066	0.863±0.030	0.422±0.011	0.234±0.010	0.910±0.005
RegionDCL-fixed margin	0.515±0.042	0.303±0.040	0.872±0.020	0.426±0.011	0.248±0.018	0.905±0.008
RegionDCL	**0.498±0.038**	**0.294±0.047**	**0.879±0.021**	**0.418±0.010**	**0.229±0.008**	**0.912±0.004**

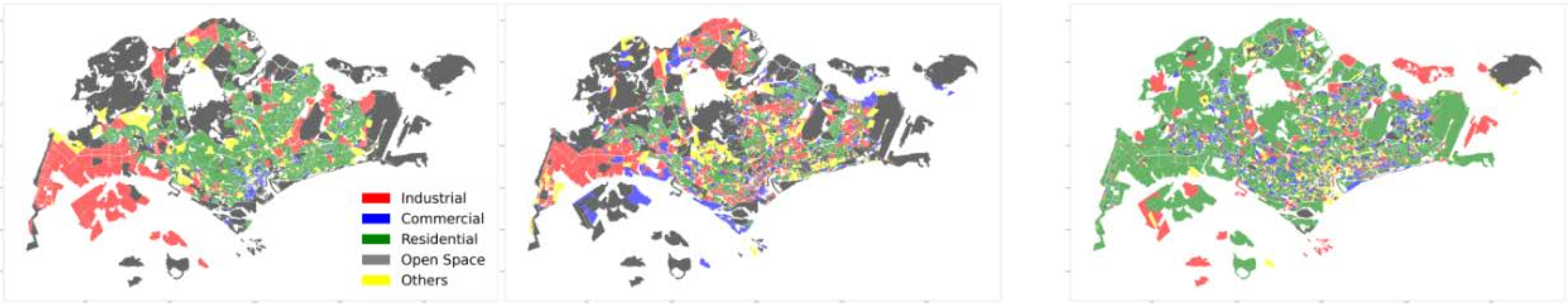
Experiments: Population Density Inference

- Similar results in inferring the population density within regions

Table 3: Population Density Inference in Singapore and New York City

Models	Singapore			New York City		
	MAE↓	RMSE↓	R^2 ↑	MAE↓	RMSE↓	R^2 ↑
Urban2Vec	6667.84±623.27	8737.27±902.41	0.303±0.119	5328.38±200.58	7410.42±261.89	0.522±0.028
Place2Vec	6952.34±713.30	9696.31±1239.65	0.171±0.121	8109.79±175.18	10228.61±261.43	0.096±0.043
Doc2Vec	6982.85±650.76	9506.81±1052.25	0.206±0.062	7734.56±247.99	9827.56±354.51	0.166±0.031
GAE	7183.24±579.82	9374.20±913.56	0.163±0.112	8010.73±290.33	10341.09±362.28	0.071±0.027
DGI	6423.44±671.25	8495.16±972.87	0.305±0.151	5330.11±261.77	7381.92±358.09	0.526±0.032
Transformer	6837.67±716.28	9042.02±1032.99	0.269±0.081	5345.17±216.30	7379.47±308.36	0.522±0.039
RegionDCL-no random	6400.50±630.35	8437.89±993.41	0.364±0.075	5228.27±210.46	7278.70±322.85	0.535±0.040
RegionDCL-fixed margin	6237.61±647.54	8387.56±948.78	0.365±0.107	5125.66±184.27	7159.65±250.12	0.551±0.033
RegionDCL	**5807.54±522.74**	**7942.74±779.44**	**0.427±0.108**	**5020.20±216.63**	**6960.51±282.35**	**0.575±0.039**

Visualization



(a) Ground truth land use (b) RegionDCL (c) Transformer

- Cluster the building group embeddings via K-Means
- Ours are visually close to the Singapore land use ground truth
- Baseline fails.

Dynamic data: ST graphs

- Input: road network and past T’ traffic speed observed at sensors
- Output: traffic speed for the next T steps

Input: Observations

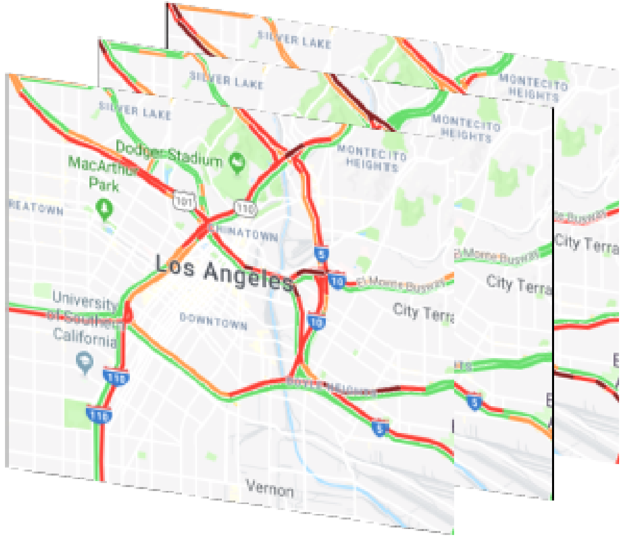
7:00 AM

...

...

8:00 AM

Output: Predictions

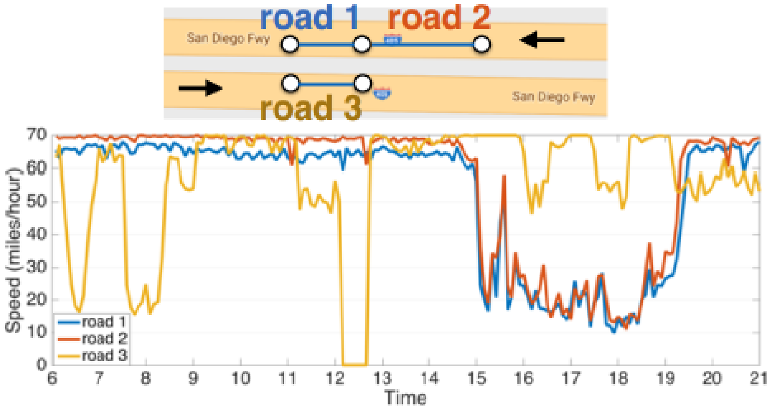


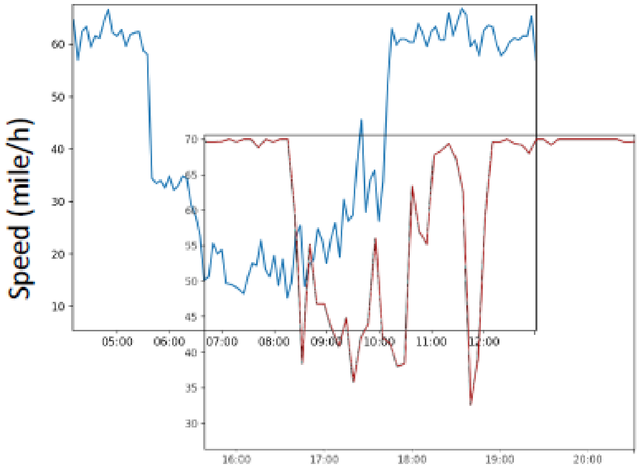
8:10AM, 8:20AM, ..., 9:00 AM

What to model?

Complex
Spatial Dependency

Non-linear, non-stationary
Temporal Dynamic





Diffusion Convolutional Recurrent Neural Network (DCRNN)

- Model spatial dependency with ***diffusion convolution of GCN***
- Sequence to sequence learning with ***encoder-decoder*** framework of RNN

Dynamic data: ST graphs

- Experiments:
  - We evaluate SAGDFN on 4 real-world datasets and we evaluate all the models on a single 32 GB Tesla V100 GPU.

STATISTICS OF OUR DATASETS

Data type	Datasets	# of sensors	Time range
Traffic speed	METR–LA	207	1 Mar - 30 June 2012
	London2000	2000	1 Jan - 31 Mar 2020
	NewYork2000	2000	1 Jan - 31 Mar 2020
Carpark lots	CARPARK1918	1918	1 May - 30 June 2021

  - We perform evaluation on multivariate time series forecasting problem with **15** baselines.

Dynamic data: ST graphs

- Comparison with baselines on Traffic Speed dataset
  - **METR-LA**: it consists of traffic speed measurements obtained from loop detectors deployed on the road network of LA County. Specifically, the dataset comprises the records from 207 sensors over four months from March to June in 2012, resulting in a total of 34,272 time slices. The sampling frequency is 5 minutes.

PERFORMANCE COMPARISON ON METR–LA DATASET.

METR–LA	Horizon 3			Horizon 6			Horizon 12		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ARIMA	3.99	8.21	9.6%	5.15	10.45	12.7%	6.9	13.23	17.4%
VAR	4.42	7.89	10.2%	5.41	9.13	12.7%	6.52	10.11	15.8%
SVR	3.99	8.45	9.3%	5.05	10.87	12.1%	6.72	13.76	16.7%
LSTM	3.44	6.30	9.6%	3.77	7.23	10.9%	4.37	8.69	13.2%
DCRNN	2.77	5.38	7.3%	3.15	6.45	8.8%	3.6	7.60	10.5%
STGCN	2.88	5.74	7.6%	3.47	7.24	9.6%	4.59	9.4	12.7%
GRAPH WaveNet	2.69	5.15	6.9%	3.07	6.22	8.4%	3.53	7.37	10.0%
GMAN	2.80	5.55	7.4%	3.12	6.49	8.7%	3.44	7.35	10.0%
AGCRN	2.87	5.58	7.7%	3.23	6.58	9.0%	3.62	7.51	10.4%
MTGNN	2.69	5.18	6.9%	3.05	6.17	8.2%	3.49	7.23	9.9%
ASTGCN	4.86	9.27	9.2%	5.43	10.61	10.1%	6.51	12.52	11.6%
STSGCN	3.31	7.62	8.1%	4.13	9.77	10.3%	5.06	11.66	12.9%
GTS	2.67	5.27	7.21%	3.04	6.25	8.4%	3.46	7.31	10.0%
STEP	2.61	4.98	**6.5%**	2.96	**5.97**	8.0%	**3.37**	**6.99**	9.6%
D2STGNN(c)	2.57	**4.93**	**6.5%**	**2.94**	**5.97**	**7.9%**	3.41	7.15	9.6%
SAGDFN	**2.56**	5.00	**6.5%**	**2.94**	6.05	**7.9%**	**3.37**	7.17	**9.5%**

Dynamic data: ST graphs

- Comparison with baselines on Carpark dataset

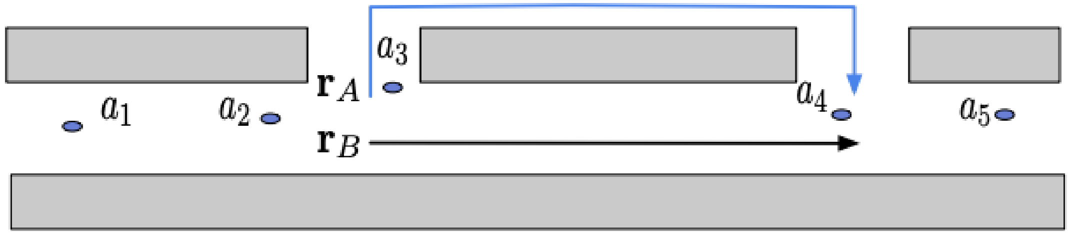
  - **CARPARK1918**: it contains parking lot availability recordings collected by the Land Transport Authority (LTA) of Singapore. Specifically, the dataset contains records from 1918 carparks over two months from May to June in 2021. The sampling frequency of the dataset is 5 minutes, resulting in a total of 17,569 time slices.

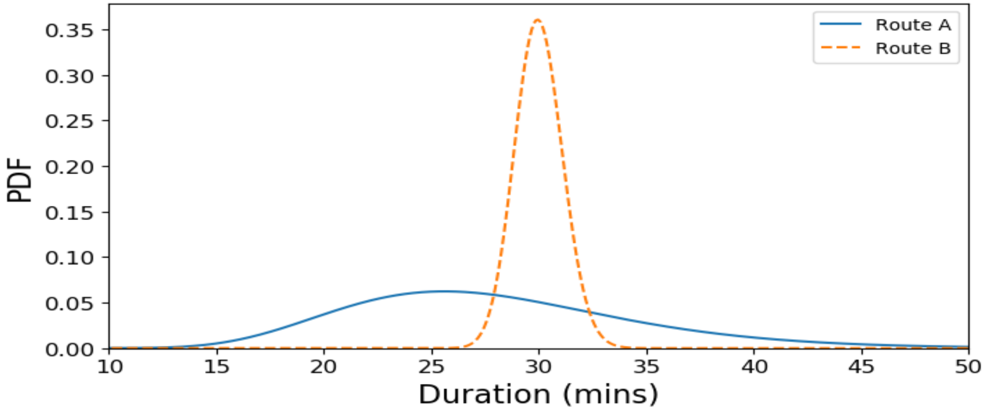
PERFORMANCE COMPARISON ON CARPARK1918 DATASET. (RESULTS MARKED '×' FOR MODEL ENCOUNTERING THE OOM ISSUE)

CARPARK1918	Horizon 3			Horizon 6			Horizon 12		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
ARIMA	3.31	7.41	7.4%	5.59	10.03	12.1%	9.15	16.68	14.1%
VAR	5.45	10.81	13.5%	7.41	14.63	19.9%	10.65	20.17	27.3%
SVR	17.71	47.42	16.11%	19.35	47.93	20.2%	21.92	49.00	22.3%
LSTM	3.15	8.90	7.0%	5.48	12.29	11.7%	9.23	18.92	16.0%
DCRNN	2.59	7.69	7.0%	4.18	10.09	10.9%	6.31	14.19	14.0%
STGCN	×	×	×	×	×	×	×	×	×
GRAPH WaveNet	7.84	17.26	15.9%	8.67	19.04	16.8%	10.49	22.75	19.8%
GMAN	×	×	×	×	×	×	×	×	×
AGCRN	×	×	×	×	×	×	×	×	×
MTGNN	3.74	9.88	8.6%	4.97	12.33	12.2%	7.57	17.02	15.8%
ASTGCN	×	×	×	×	×	×	×	×	×
STSGCN	×	×	×	×	×	×	×	×	×
GTS	×	×	×	×	×	×	×	×	×
STEP	×	×	×	×	×	×	×	×	×
D2STGNN(c)	×	×	×	×	×	×	×	×	×
SAGDFN	**2.26**	**7.74**	**6.0%**	**3.77**	**10.07**	**10.2%**	**5.36**	**13.17**	**12.7%**

Dynamic data: Trajectories

- Intelligent transportation systems: ETA



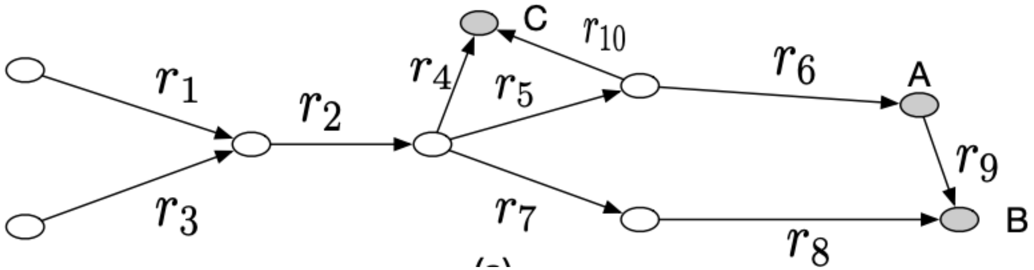


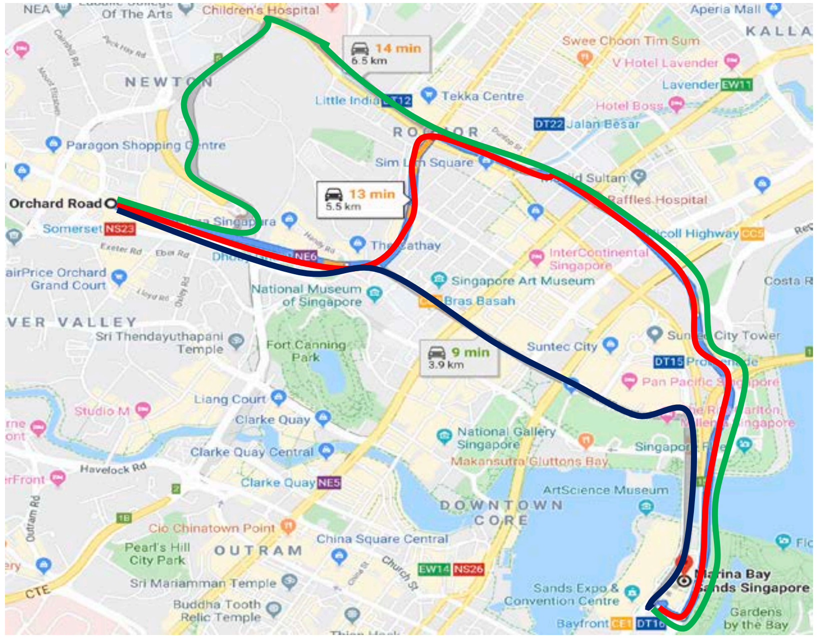
- Given a route on the road network, we aim to learn its travel time distribution (Probability Density Function) with the consideration of real-time traffic.

Xiucheng Li et al., Learning Travel Time Distributions with Deep Generative Model. WWW 2019

Dynamic data: Trajectories

- Intelligent transportation systems: route inference



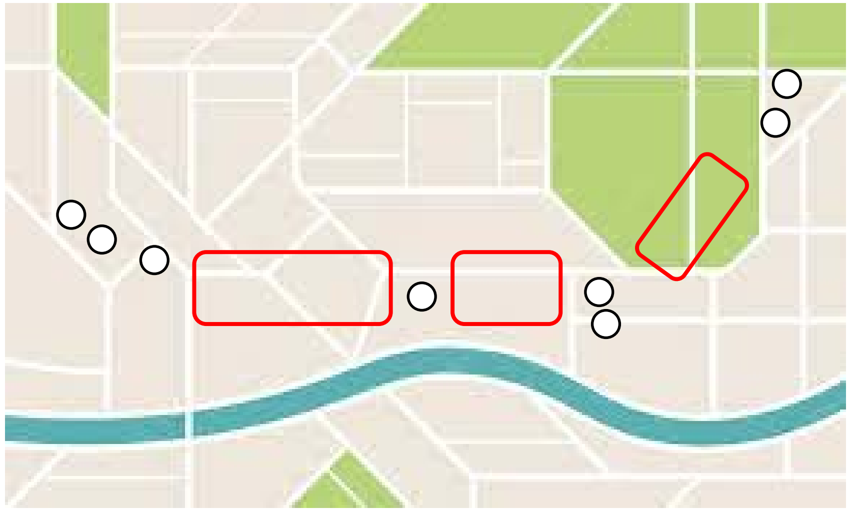


- Given the origin and destination, we aim to predict the most likely traveling route on the road network and score the likelihood of a given route.

Xiucheng Li et al., Spatial Transition Learning on Road Networks with Deep Probabilistic Models. ICDE 2021

Dynamic data: Trajectories

- Intelligent transportation systems: trajectory enrichment
  - Trajectories can be sparse and incomplete due to technical issues.



☹

  - These low-quality trajectories are not good for the analysis and model development in traffic management systems.

Yile Chen et al., TERI: An Effective Framework for Trajectory Recovery with Irregular Time Intervals. VLDB 2024

Introduction to Foundation Models

Foundations Models (FMs) represent a paradigm shift in AI

Advantages:

- Self-supervised pre-training
- Task-agnostic —> FMs develop capabilities that generalise across tasks
- Able to access Internet-scale amount of (unlabelled) data
- Easy to deploy to downstream applications (fine-tune or zero-shot)

Foundation Models

Task-Specific Models

Training **specific** models for **specific** tasks

- Question Answering Models
- Machine Translation Models
- Common Sensing Reasoning Models
- Reading Comprehension Models
- Natural Language Inference Models
- Image Classification Models
- Text-to-image Generation Models
- Image Editing Models

Paradigm shift

Foundation Models

A large task-agnostic pre-trained model which can be adapted via **fine-tuning** or **few-shot/zero-shot** learning on **a wide range of domains**. (Bommasani et al, 2021)

GPT-3 (Brown et al., 2020)

Few-shot Adaptation

Various NLP Tasks

- Closed Book Question Answering
- Machine Translation
- Common Sense Reasoning
- Reading Comprehension
- Natural Language Inference
- …

DALL·E 2 (Ramesh et al., 2022)

Zero-shot Transfer

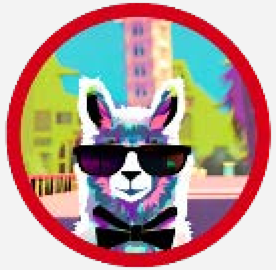
Various CV Tasks

- Text-to-image generation
- Image Completion
- Image Editing
- Style Transfer
- …

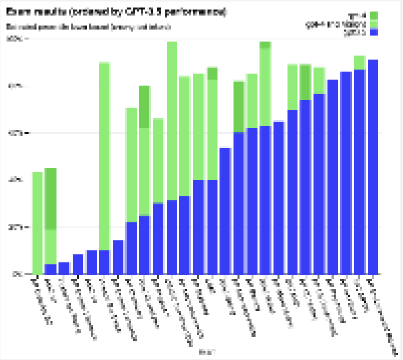
Foundation Models in Different Domains

Natural Language Processing

Stanford
Alpaca



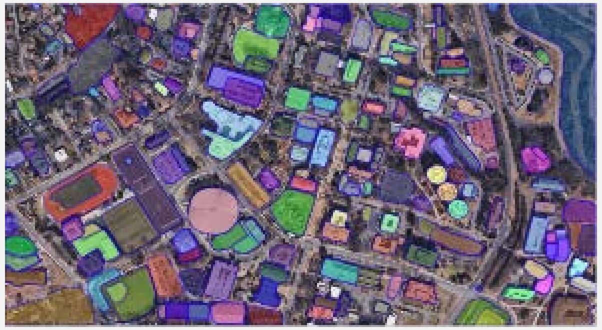
Stanford Alpaca



ChatGPT/GPT-4 (OpenAI. 2023)

Computer Vision

Imagen (Saharia et al. 2022)

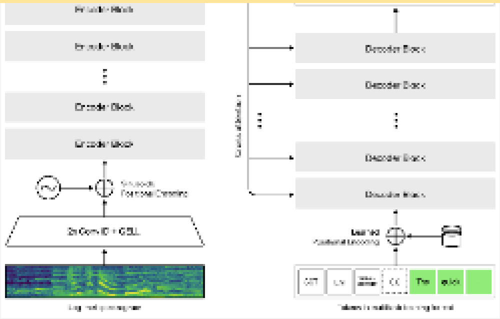


Segment Anying (Kirillov et al, 2023)

Reinforcement Learning

Gato (Reed et al. 2022)

Signal Processing



Whisper (Radford et al. 2022)

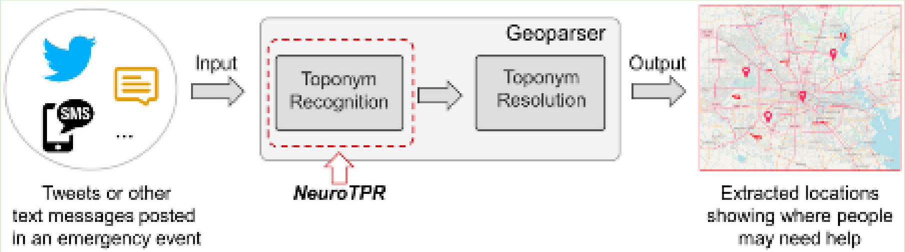
The Big Question

GPT-4 FOR

AGI on Geospatial Problems

How do the existing cutting-edge foundation models perform when compared with the state-of-the-art fully supervised task-specific models on various geospatial tasks?

Geospatial Semantics – Topo.Recg.

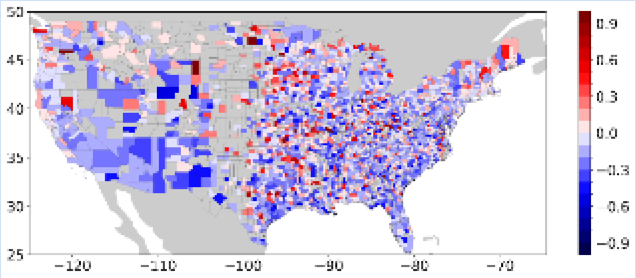


Urban Geography

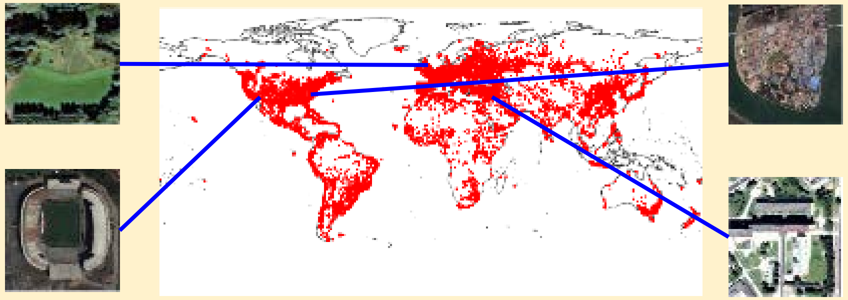
Urban Function

Urban Perception

Health Geography – Dementia Forecast



Remote Sensing – RS Image Clas.



Gengchen Mai, Weiming Huang, et al. . [On the Opportunities and Challenges of Foundation Models for Geospatial Artificial Intelligence](). ACM TSAS 2024

Geospatial Semantics

- Investigate the performance of **GPT-3** on some well established **geospatial semantic tasks**:

Typonym Recognition

```
[Instruction] ...
Paragraph: Alabama State Troopers say a Greenville man has died of his injuries
    ↪ after being hit by a pickup truck on Interstate 65 in Lowndes County.
Q: Which words in this paragraph represent named places?
A: Alabama; Greenville; Lowndes
...
--
Paragraph: The Town of Washington is to what Williamsburg is to Virginia.
Q: Which words in this paragraph represent named places?
A: Washington; Williamsburg; Virginia
```

*toponyms: proper names of places, also known as place names and geographic names.

Location Description Recognition

```
[Instruction] ...
Paragraph: Papa stranded in home. Water rising above waist. HELP 8111 Woodlyn Rd
    ↪ , 77028 #houstonflood
Q: Which words in this paragraph represent location descriptions?
A: 8111 Woodlyn Rd, 77028
...
--
Paragraph: HurricaneHarvey Help Need AT 7506 Jackrabbit Rd, Houston, TX 77095.
Q: Which words in this paragraph represent location descriptions?
A: 7506 Jackrabbit Rd, Houston, TX 77095
```

GPT-3 Fewshot Learning for Geospatial Semantic Tasks

Task 1 & 2: Toponym Recognition & Location Description Recognition

- Typonym recognition: FMs (e.g., GPT-2/3) consistently outperform the **fully-supervised** baselines with only **8 few-shot** examples

- Location Description Recognition: GPT-3 achieves the best Recall score across all methods

Typonym Recognition	Location Description Recognition

	Model	#Param	Toponym Recognition		Location Description Recognition		
			Hu2014	Ju2016	HaveyTweet2017		
			Accuracy ↓	Accuracy ↓	Precision ↓	Recall ↓	F-Score ↓
(A)	Stanford NER (nar. loc.) [30]	-	0.787	0.010	**0.828**	0.399	0.539
	Stanford NER (bro. loc.) [30]	-	-	0.012	0.729	0.44	0.548
	Retrained Stanford NER [30]	-	-	0.078	0.604	0.410	0.489
	Caseless Stanford NER (nar. loc.) [30]	-	-	0.460	0.803	0.320	0.458
	Caseless Stanford NER (bro. loc.) [30]	-	-	0.514	0.721	0.336	0.460
	spaCy NER (nar. loc.) [44]	-	0.681	0.000	0.575	0.024	0.046
	spaCy NER (bro. loc.) [44]	-	-	0.006	0.461	0.304	0.366
	DBpedia Spotlight[99]	-	0.688	0.447	-	-	-
(B)	Edinburgh [7]	-	0.656	0.000	-	-	-
	CLAVIN [134]	-	0.650	0.000	-	-	-
	TopoCluster [23]	-	0.794	0.158	-	-	-
(C)	CamCoder [33]	-	0.637	0.004	-	-	-
	Basic BiLSTM+CRF [77]	-	-	0.595	0.703	0.600	0.649
	DM NLP (top. rec.) [139]	-	-	0.723	0.729	0.680	0.703
	NeuroTPR [135]	-	$0.675^{\dagger}$	0.821	0.787	0.678	**0.728**
(D)	GPT2 [115]	117M	0.556	0.650	0.540	0.413	0.468
	GPT2-Medium [115]	345M	0.806	0.802	0.529	0.503	0.515
	GPT2-Large [115]	774M	0.813	0.779	0.598	0.458	0.518
	GPT2-XL [115]	1558M	0.869	**0.846**	0.492	0.470	0.481
	GPT-3 [15]	175B	**0.881**	0.811^{*}	0.603	**0.724**	0.658
	InstructGPT [106]	175B	0.863	0.817^{*}	0.567	0.688	0.622
	ChatGPT (Raw.) [104]	176B	0.800	0.696^{*}	0.516	0.654	0.577
	ChatGPT (Con.) [104]	176B	0.806	0.656^{*}	0.548	0.665	0.601

Health Geography

Task 4: US County-Level Dementia Time Series Forecasting

```
[Instruction] This is a set of time series forecasting problems.
The `Paragraph` is a time series of the numbers of deaths from
    ↪ alzheimer's disease for one of US counties from 1999 to 2019.
The goal is to predict the number of deaths from alzheimer's disease at
    ↪ this county in 2020. Please give a single number as the
    ↪ prediction.
--
--
Paragraph: At Santa Barbara County, CA, from 1999 to 2019, the numbers
    ↪ of deaths from alzheimer's disease are
    ↪ 126 in 1999, 114 in 2000, 124 in 2001, 127 in 2002, 156 in 2003,
    ↪ 154 in 2004, 175 in2005, 172 in 2006, 171 in 2007, 248 in 2008, 204
    ↪ in 2009, 241 in 2010, 260 in 2011, 297 in 2012, 283 in 2013, 308 in
    ↪ 2014, 358 in 2015, 365 in 2016, 334 in 2017, 363 in 2018,
    ↪ and 328 in 2019.
Q: Please forecast the number in 2020 at Santa Barbara County, CA?
A: 345
```

Listing 4. US county-level Alzimier time series forecasting with LLMs by zero-shot learning. Yellow block: the historical time series data of one US county. Orange box: the outputs of InstructGPT. Here, we use Santa Barbara County, CA as an example and the correct answer is 373.

Table 3. Evaluation results of various GPT models and baselines on the US county-level dementia time series forecasting task. We use same model set and evaluation metrics as Table 2.

	Model	#Param	MSE ↓	MAE ↓	MAPE ↓	R^2 ↑
(A) Simple	Persistence [103, 107]	-	1,648	16.9	0.189	0.979
(B) Supervised ML	ARIMA [58]	-	1,133	15.1	0.193	0.986
(C) Zero shot LLMs	GPT2 [115]	117M	77,529	92.0	0.587	-0.018
	GPT2-Medium [115]	345M	226,259	108.1	0.611	-2.824
	GPT2-Large [115]	774M	211,881	94.3	0.581	-1.706
	GPT2-XL [115]	1558M	162,778	99.8	0.627	-1.082
	GPT-3 [15]	175B	1,105	14.5	0.180	0.986
	InstructGPT [106]	175B	**831**	**13.3**	**0.179**	**0.989**
	ChatGPT (Raw.) [104]	176B	4,115	23.2	0.217	0.955
	ChatGPT (Con.) [104]	176B	3,402	20.7	0.231	0.944

Urban Geography

Task 6: Street View Image-Based Urban Noise Intensity Classification

(a) Very Quiet (0.100) (b) Quiet (0.350) (c) Noisy (0.600) (d) Very Noisy (0.965)

Fig. 6. Some street view image examples in *SingaporeSVI*579 dataset. The image caption indicates the noise intensity class this image belongs to and the numbers in parenthesis indicate the original noise intensity scores from Zhao et al. [162].

Table 6. Evaluation results of various vision-language foundation models and baselines on the urban street view image-based noise intensity classification dataset, SingaporeSVI579 [162]. We classify models into two groups: (A) Supervised finetuned convolutional neural networks (CNNs); (B) Zero-shot learning with visual-language foundation models (VLFMs). We use accuracy and weighted F1 scores as evaluation metrics. The best scores for each group are highlighted.

	Model	#Param	Accuracy	F1
(A) Supervised Finetuned CNNs	AlexNet [74]	58M	0.452	0.405
	ResNet18 [37]	11M	0.493	**0.442**
	ResNet50 [37]	24M	**0.500**	0.436
	DenseNet161 [48]	27M	0.486	0.382
(B) Zero-shot FMs	OpenCLIP-L [54, 113, 127]	427M	0.128	0.089
	OpenCLIP-B [54, 113, 127]	2.5B	0.169	0.178
	BLIP [81, 82]	3.9B	**0.452**	**0.405**
	OpenFlamingo-9B [11]	8.3B	0.262	0.127

GPT-3 Fewshot Learning for Geospatial Semantic Tasks

- **Shortcoming of text FMs**: by design they are unable to handle other data modality, e.g., geo-coordinates, toponym resolution/geoparsing

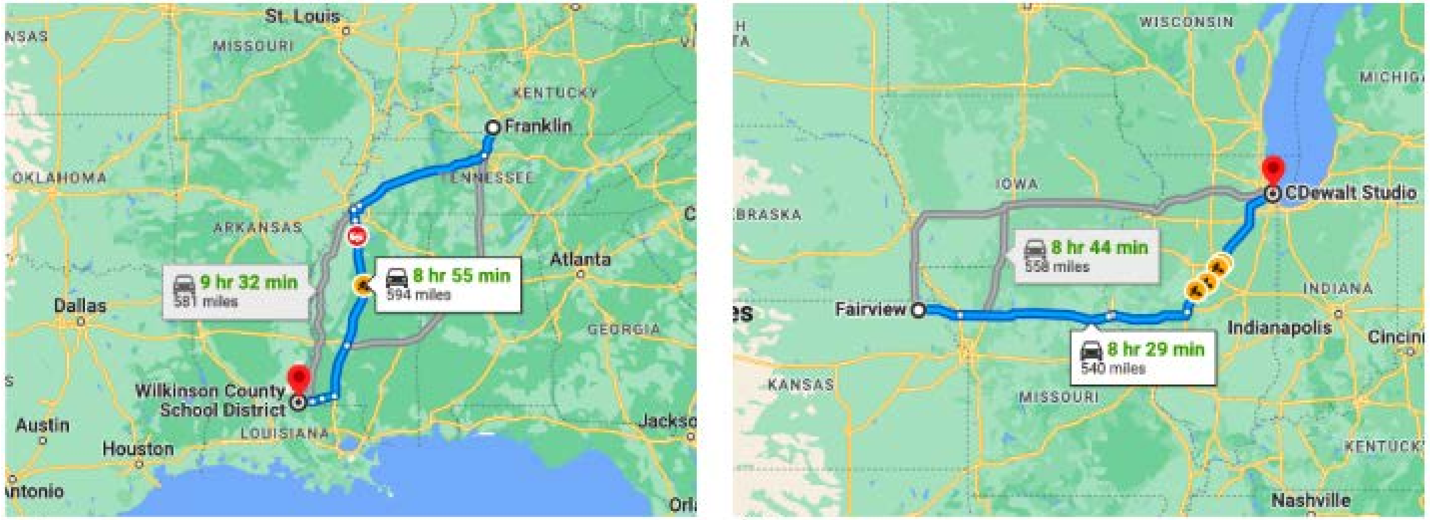
Geoparsing

```
[Instruction] ...
Paragraph: San Jose was founded in 1803 when allotments of land were made ...
Q: Which words in this paragraph represent named places?
A: San Jose; New Mexico

Q: What is the location of San Jose?
A: 35.39728, -105.47501
...
--
Paragraph: the city of fairview had a population of 260 as of july 1, 2015. ...
Q: Which words in this paragraph represent named places?
A: Fairview

Q: What is the location of Fairview?
A: 41.85003, -87.65005
```

- The predicted coordinates are not accurate

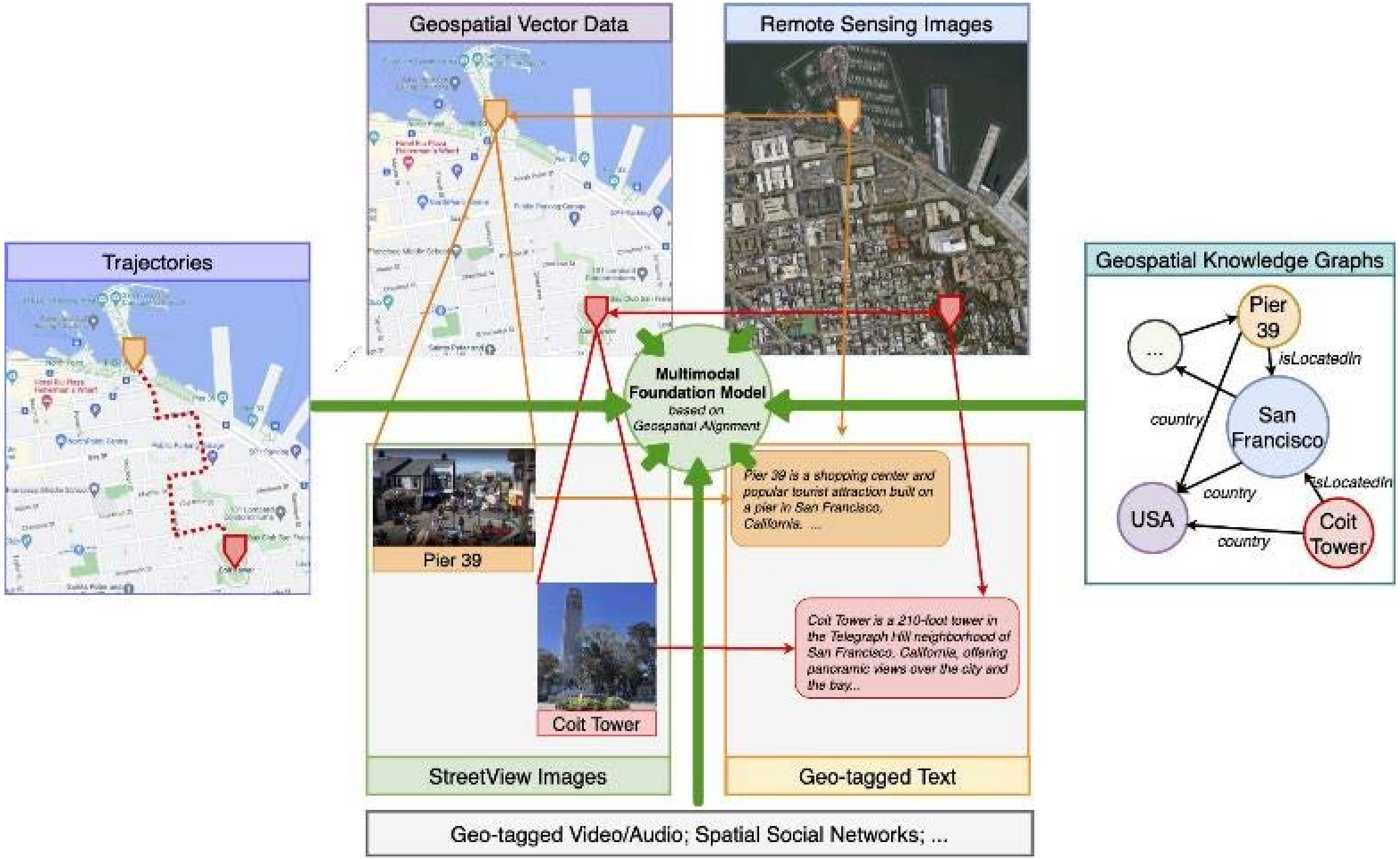


(a) [TEXT]: Franklin is a city in and the county seat of simpson county, ...

(b) [TEXT]: the city of Fairview had a population of 260 as of july 1, 2015. ...

A Multimodal City FM for GeoAI

Vision: a multimodal City FM for GeoAI that use their **geospatial relationships as alignments** among **different data modalities.**



Summary

- Spatial-temporal data mining
    - Spatial relationship extraction
    - Geospatial IR or Spatial Keyword Search
    - POI recommendations
    - Road Network Representation for Road Network Applications
    - Region Representation for Region-Level Applications
- Trajectory data mining
    - Application in intelligent transportation
- Application of Foundation Models for Geospatial Applications