Geospatial Entity Representation: A Step Towards City Foundation Models

Gao Cong

Nanyang Technological University

第二届中新数字经济与人工智能高峰论坛
THE 2nd SINO-SINGAPORE SUMMIT FOR DIGITAL ECONOMY AND ARTIFICIAL INTELLIGENCE

中国人工智能学会

丛　高 Gao CONG

南洋理工大学计算机科学与工程学院教授、新加坡电信-南洋理工人工智能联合实验室项目负责人、实验室主任

Professor, school of Computer Science and Engineering at Nanyang Technological University, a co-director for Singtel Cognitive and Artificial Intelligence Lab for Enterprises@NTU

Research Overview

Geo-spatial data mining and applications

Point
- POI recommendation & prediction
- Knowledge graph (e.g., entity resolution)

Trajectory
- Trajectory representation, similarity, simplification
- Spatial temporal prediction: Speed, travel time, route prediction, anomaly

Region
- Region search, (e.g., burst region)
- Region representation (e.g., topic)

City Foundation Models for Representation

Enriched spatial data systems

Distributed streaming systems
- Snapshot queries (OLTP, OLAP)
- Continuous queries
- Load balance & Workload adjustment
- Index & query processing

Spatial +X systems
- OLTP, OLAP, Data exploration
- Index &Query processing
- Storage

Machine learning for spatial DB

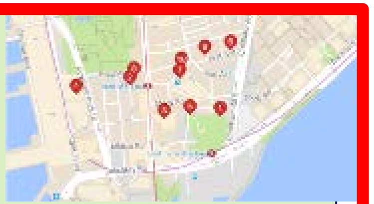
Geo-spatial + X (e.g., text, temporal) data

Outline

- Target on a specific problem on **point spatial entity**
  - Spatial relationship extraction (SIGMOD’23)
  - Geospatial IR or Spatial Keyword Search (VLDB’09--SIGMOD’23)
  - POI recommendations (SIGIR’13 --)
- Self-supervised learning for geospatial entity representation
  - Road Network Representation for Road Network Applications (CIKM’21)
  - Region Representation for Region-Level Applications (KDD’23)
  - Application of Foundation Models for Geospatial Applications
  - Efforts toward City Foundation Models.

Our Research on Point Spatial Entity

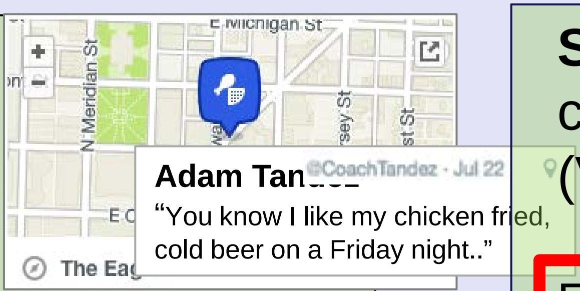
Geospatial IR
(Spatial keyword query)
(VLDB'09 --)



POI / User recommendation
(SIGIR'13 --)

POI data exploration
(SIGMOD'18)

Integrating POI with tweets & photos
(CIKM'17)



POI Annotation

Semantic relationship extraction: competitiveness/complementary (VLDB'22)

Entity resolution/ **geospatial relationship extraction** (WWW'22, SIGMOD'23)

Constructing knowledge graphs

Extracting locations and entity linkage (WWW'18)

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

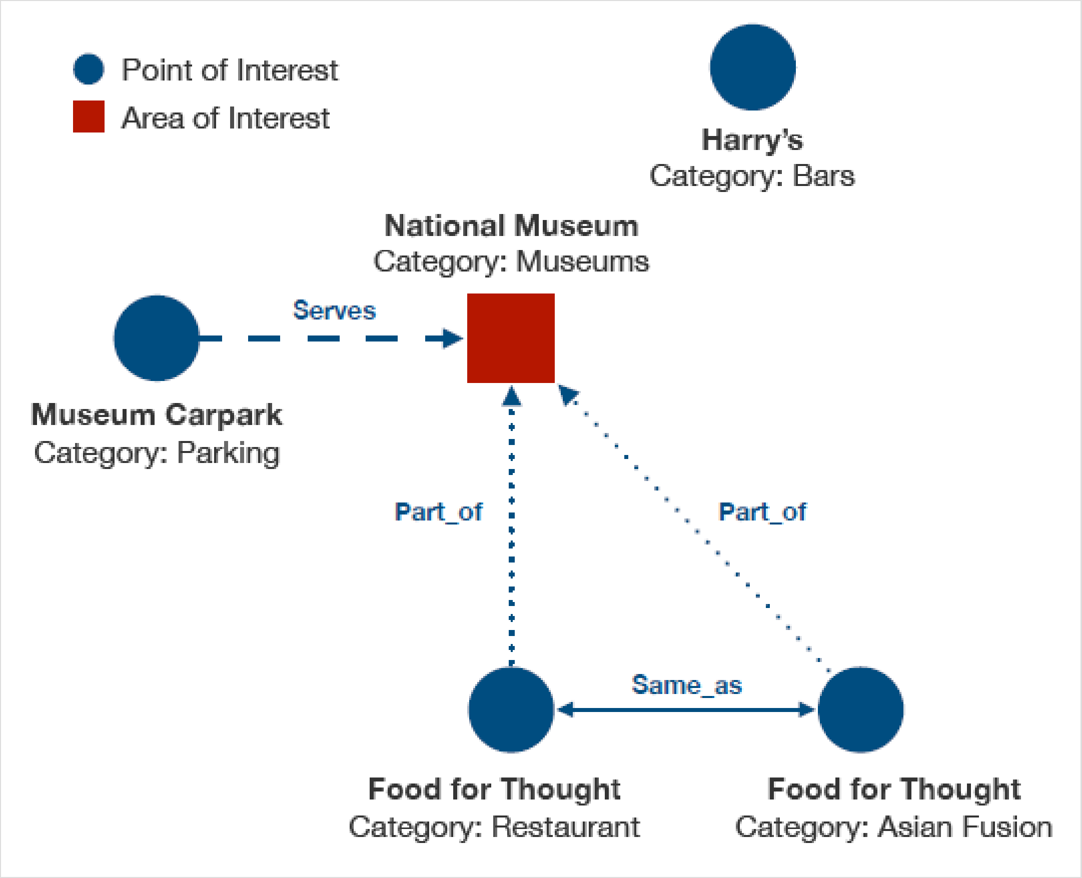
Although convenient, this representation hinders the exploration of **geospatial relationships** between the entities

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

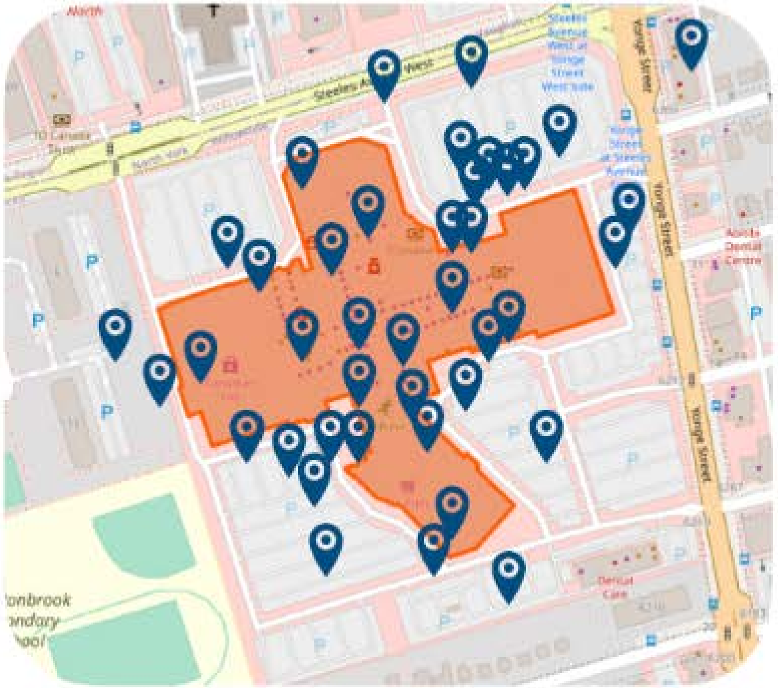
DBPedia

Challenges

- Scarcity of complex geometries (i.e. **polygons**)
- **Inaccuracy** of the geo-positional systems

In reality, all the POIs are located inside the Mall

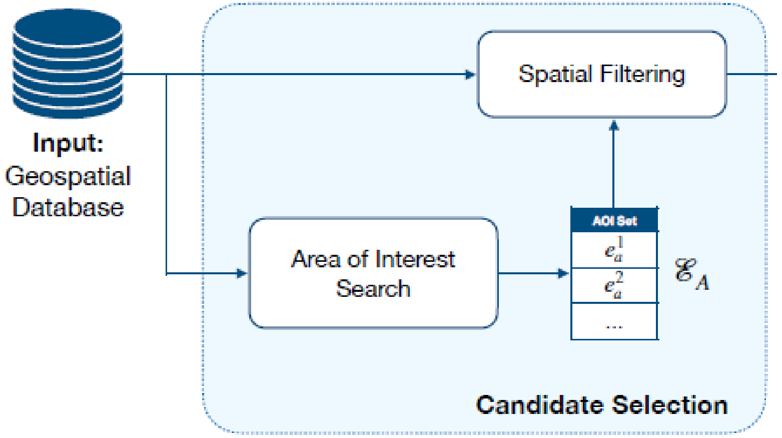
Sources: YELP, OSM

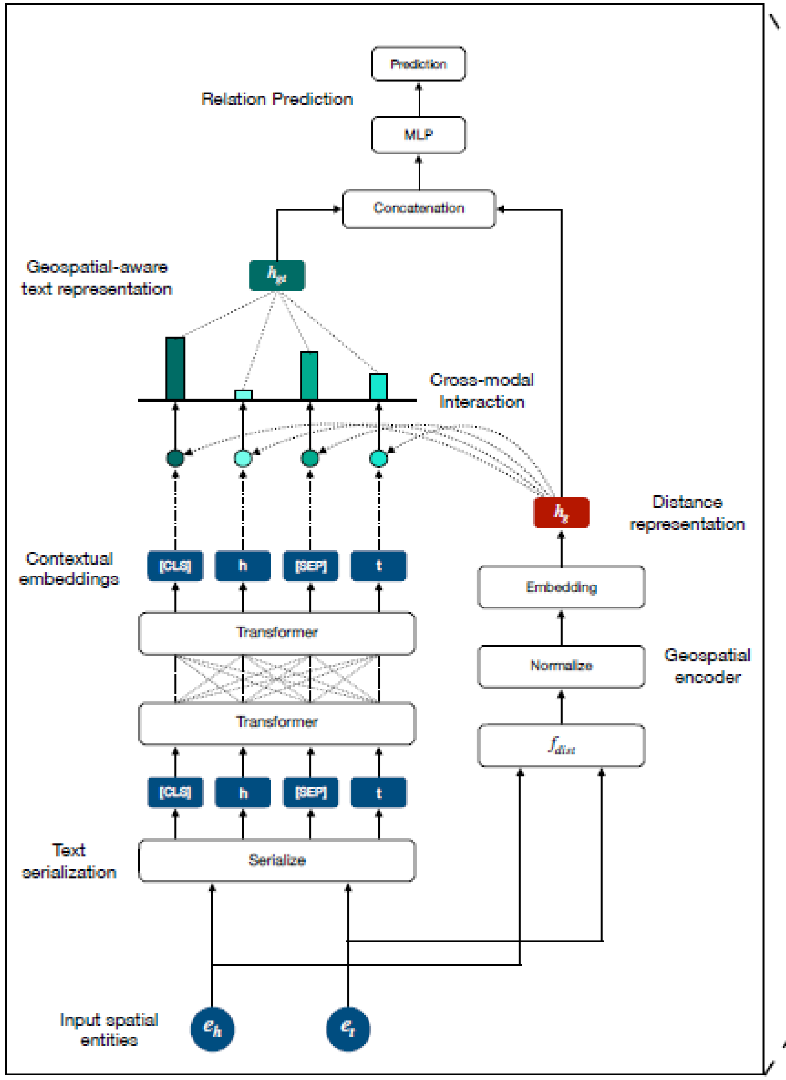


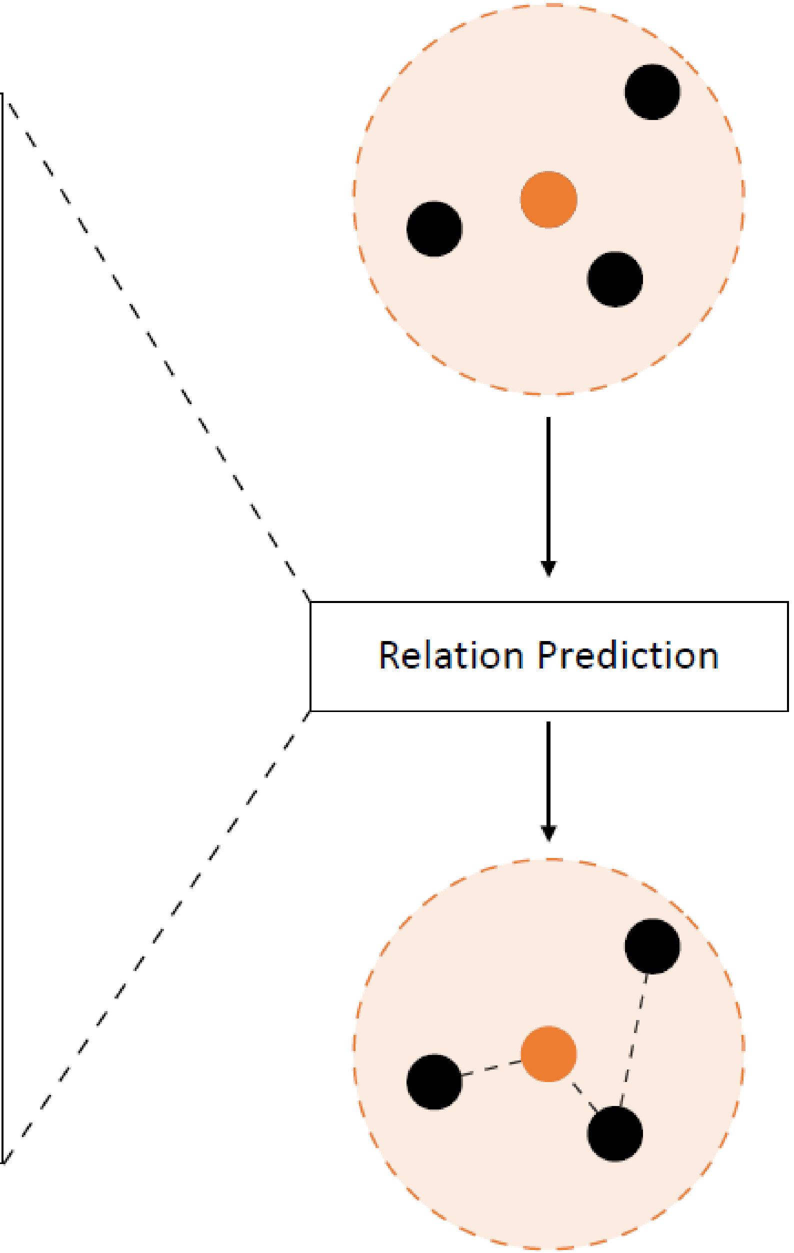
Centerpoint Mall, Toronto

Existing algorithms for KGC are not designed to take into account the **spatial position** of the nodes

Proposed solution





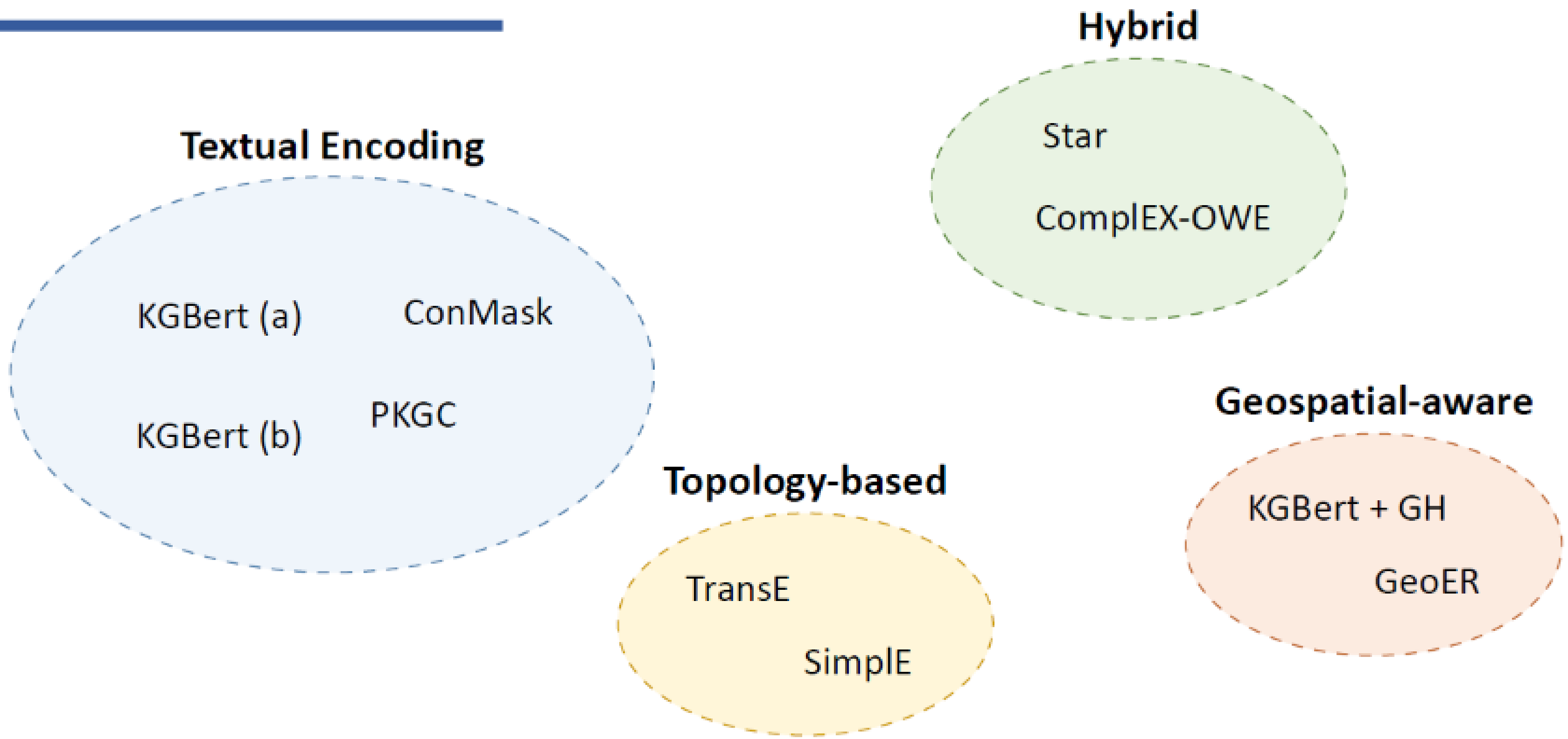


- **Candidate Selection Step**: Aim
  relationships
- **Relation Prediction:** Aim at ide
- **The KG refinement**: Aim to ext
  correctness

Experiments

				# of Triples			Relations				
City	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{E}_A\rvert$	$\lvert\mathcal{R}\rvert$	Train	Valid	Test	*part_of*	*same_as*	*serves*	Category	Address
Singapore	17092	370	4	13076	5229	7852	8526	1547	2656	99.79%	67.21%
Toronto	18911	179	4	8488	3390	5101	5744	1262	1188	99.92%	62.87%
Seattle	10504	500	4	7906	3162	4747	4257	1138	1215	99.85%	68.06%
Melbourne	13473	190	4	3058	1220	1839	2675	610	432	99.94%	62.45%

Baselines



Experimental results

LLM-based Textual encoding approaches perform well

Model	Singapore			Toronto			Seattle			Melbourne		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
TransE [6]	4.71	16.3	7.29 (± 0.41)	5.28	12.82	7.47 (± 0.28)	4.51	11.07	6.4 (± 0.79)	4.25	9.56	5.88 (± 0.55)
SimplE [30]	6.64	19.27	9.87 (± 0.96)	4.98	18.3	7.82 (± 0.71)	7.75	9.84	8.66 (± 1.05)	6.9	13.33	9.09 (± 0.93)
ComplEx-OWE [47]	60.18	44.29	51.02 (± 0.39)	60.5	41.13	48.97 (± 0.88)	40.21	29.77	34.2 (± 0.87)	66.54	41.9	51.42 (± 1.02)
ConMask [50]	63.28	50.1	55.93 (± 0.46)	67.86	41.66	51.58 (± 0.78)	58.81	40.45	47.98 (± 0.32)	72.89	44.44	55.21 (± 0.6)
KG-BERT (a) [65]	85.38	86.2	80.64 (± 1.31)	77.55	75.46	76.33 (± 0.99)	74.81	70.27	72.39 (± 1.44)	78.1	76.46	77.25 (± 1.71)
KG-BERT (b) [65]	85.80	78.11	81.77 (± 0.7)	82.58	77.21	79.78 (± 1.25)	77.61	69.11	73.02 (± 1.09)	76.44	72.24	74.52 (± 1.95)
PKGC [38]	80.55	73.38	76.79 (± 1.09)	84.13	67.87	75.13 (± 0.91)	78.44	62.58	69.61 (± 0.9)	77.7	73.96	75.78 (± 2.26)
StAR [58]	65.15	72.66	68.7 (± 0.72)	76.48	80.1	78.24 (± 1.51)	60.96	58.24	59.56 (± 0.47)	81.92	83.97	82.93 (± 0.86)
KG-BERT (+GH)	82.99	86.66	84.78 (± 1.11)	86.26	78.01	81.92 (± 1.28)	73.8	78.95	76.28 (± 1.67)	84.11	77.28	80.55 (± 1.23)
Geo-ER [4]	88.27	84.7	86.44 (± 0.88)	87.25	81.74	84.4 (± 1.16)	78.58	78.91	78.74 (± 1.25)	82.6	88.21	85.31 (± 1.47)
GTMiner	90.07	88.15	89.1* (± 1.04)	86.91	88.4	87.64* (± 1.49)	80.56	80.95	80.75* (± 1.21)	87.87	87.86	87.87* (± 1.31)
GTMiner (+Ex)	90.17	89.25	89.65 (± 1.13)	87.0	89.29	88.13 (± 1.39)	80.8	82.37	81.57 (± 1.29)	88.1	88.78	88.24 (± 1.22)
GTMiner (+Ex +Re)	91.33	89.25	**90.27** (± 1.09)	88.08	89.23	**88.66** (± 1.33)	81.27	82.37	**81.81** (± 1.28)	88.27	88.69	**88.47** (± 1.2)
Δ_{F1}			+3.82%			+4.26%			+3.07%			+3.16%

Experimental results

Model	Singapore P	Singapore R	Singapore F1	Toronto P	Toronto R	Toronto F1	Seattle P	Seattle R	Seattle F1	Melbourne P	Melbourne R	Melbourne F1
TransE [6]	4.71	16.3	7.29 (± 0.41)	5.28	12.82	7.47 (± 0.28)	4.51	11.07	6.4 (± 0.79)	4.25	9.56	5.88 (± 0.55)
SimplE [30]	6.64	19.27	9.87 (± 0.96)	4.98	18.3	7.82 (± 0.71)	7.75	9.84	8.66 (± 1.05)	6.9	13.33	9.09 (± 0.93)
ComplEx-OWE [47]	60.18	44.29	51.02 (± 0.39)	60.5	41.13	48.97 (± 0.88)	40.21	29.77	34.2 (± 0.87)	66.54	41.9	51.42 (± 1.02)
ConMask [50]	63.28	50.1	55.93 (± 0.46)	67.86	41.66	51.58 (± 0.78)	58.81	40.45	47.98 (± 0.32)	72.89	44.44	55.21 (± 0.6)
KG-BERT (a) [65]	85.38	86.2	80.64 (± 1.31)	77.55	75.46	76.33 (± 0.99)	74.81	70.27	72.39 (± 1.44)	78.1	76.46	77.25 (± 1.71)
KG-BERT (b) [65]	85.80	78.11	81.77 (± 0.7)	82.58	77.21	79.78 (± 1.25)	77.61	69.11	73.02 (± 1.09)	76.44	72.24	74.52 (± 1.95)
PKGC [38]	80.55	73.38	76.79 (± 1.09)	84.13	67.87	75.13 (± 0.91)	78.44	62.58	69.61 (± 0.9)	77.7	73.96	75.78 (± 2.26)
StAR [58]	65.15	72.66	68.7 (± 0.72)	76.48	80.1	78.24 (± 1.51)	60.96	58.24	59.56 (± 0.47)	81.92	83.97	82.93 (± 0.86)
KG-BERT (+GH)	82.99	86.66	84.78 (± 1.11)	86.26	78.01	81.92 (± 1.28)	73.8	78.95	76.28 (± 1.67)	84.11	77.28	80.55 (± 1.23)
Geo-ER [4]	88.27	84.7	86.44 (± 0.88)	87.25	81.74	84.4 (± 1.16)	78.58	78.91	78.74 (± 1.25)	82.6	88.21	85.31 (± 1.47)
GTMiner	90.07	88.15	89.1* (± 1.04)	86.91	88.4	87.64* (± 1.49)	80.56	80.95	80.75* (± 1.21)	87.87	87.86	87.87* (± 1.31)
GTMiner (+Ex)	90.17	89.25	89.65 (± 1.13)	87.0	89.29	88.13 (± 1.39)	80.8	82.37	81.57 (± 1.29)	88.1	88.78	88.24 (± 1.22)
GTMiner (+Ex +Re)	91.33	89.25	**90.27** (± 1.09)	88.08	89.23	**88.66** (± 1.33)	81.27	82.37	**81.81** (± 1.28)	88.27	88.69	**88.47** (± 1.2)
Δ_{F1}			+3.82%			+4.26%			+3.07%			+3.16%

Including geospatial information further improves the results

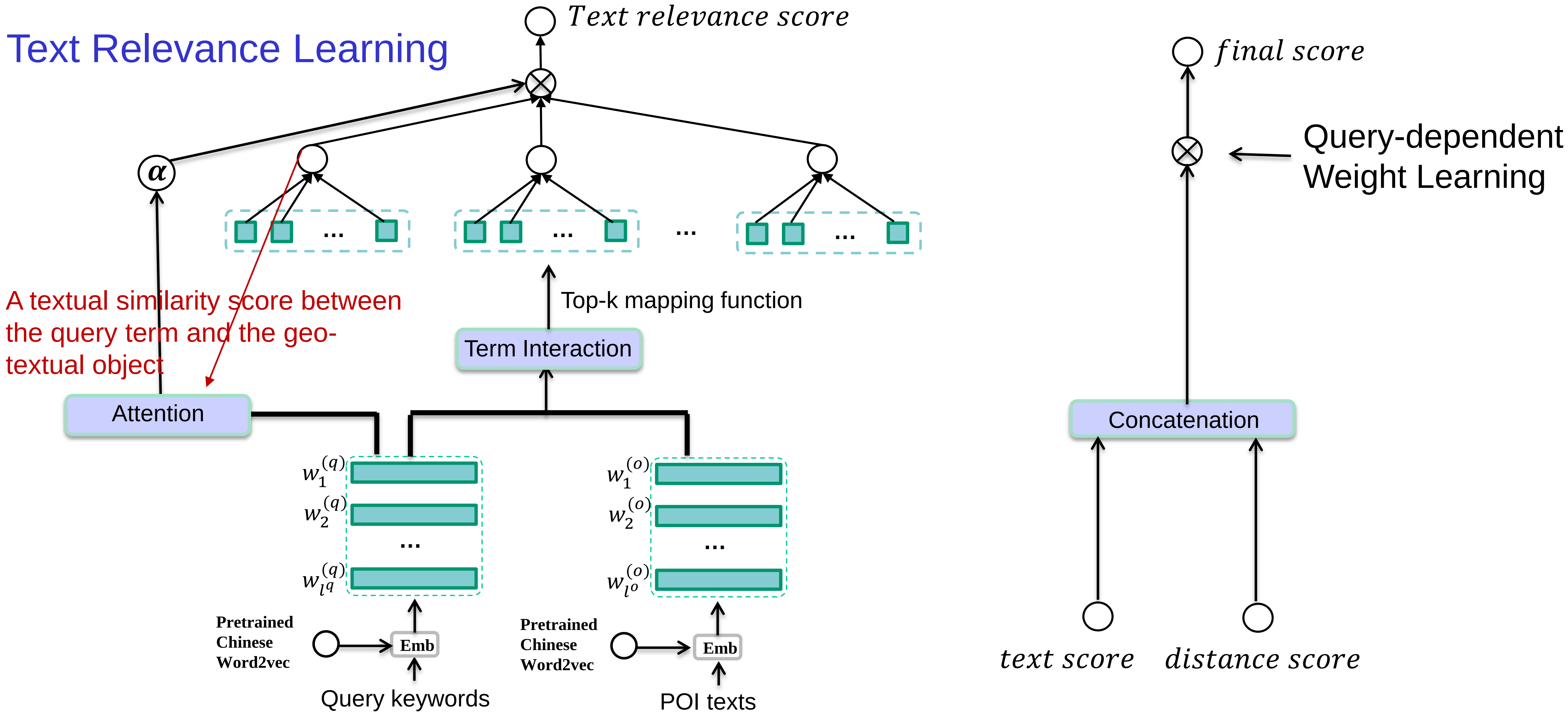
Spatial Keyword Query (Geographic IR)

- 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

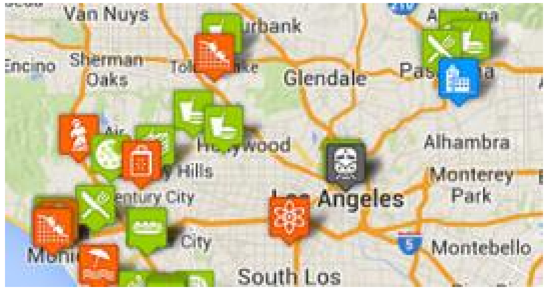


Shang Liu, Gao Cong, Kaiyu Feng, Wanli Gu, Fuzheng Zhang: Effectiveness Perspectives and a Deep Relevance Model for Spatial Keyword Queries. SIGMOD 2023

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

Other Types of POI recommendation

Context-aware POI recommendation

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

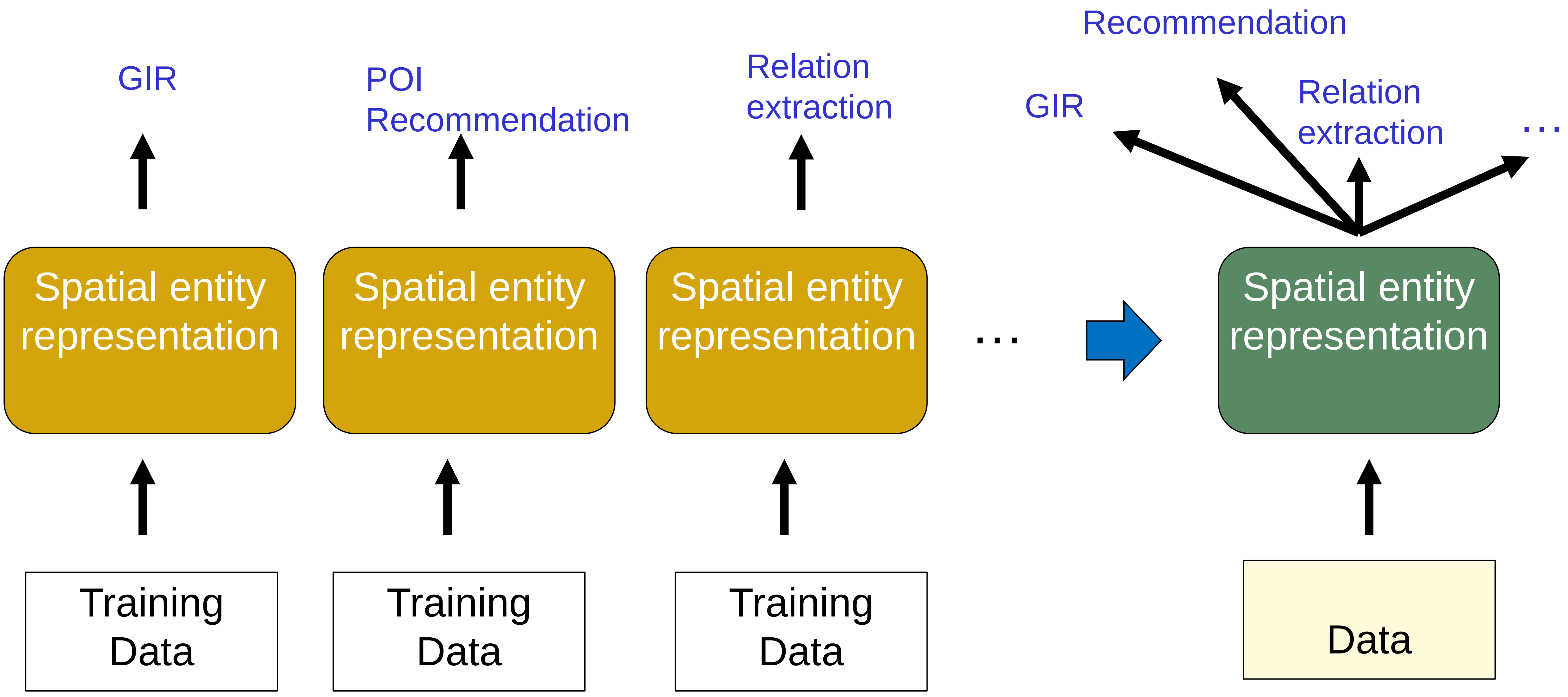
Requirement-aware POI recommendation (w/o Group)

- 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

A Summary

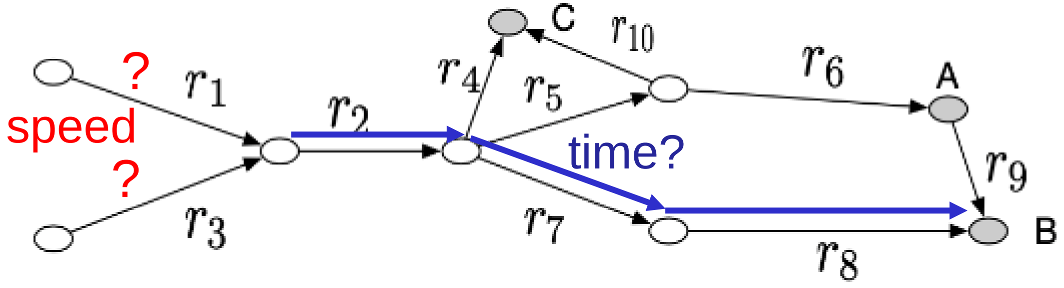

GIR
POI Recommendation
Relation extraction
Recommendation
GIR
Relation extraction
…
Spatial entity representation
Spatial entity representation
Spatial entity representation
…
Spatial entity representation
Training Data
Training Data
Training Data
Data

Outline

- Target on a specific problem on **point spatial entity**
  - Geospatial IR or Spatial Keyword Search (VLDB’09--SIGMOD’23)
  - POI recommendations (SIGIR’13 --)
  - Spatial relationship extraction (SIGMOD’23)
- Self-supervised learning for geospatial entity representation
  - Road Network Representation for Road Network Applications (CIKM’21)
  - Region Representation for Region-Level Applications (KDD’23)
  - Application of Foundation Models for Geospatial Applications
  - Efforts toward City Foundation Models.

Representation Learning for Road Networks

- ▷ **Motivation: many** applications are built upon road networks, such as travel time estimation, traffic inference, etc.

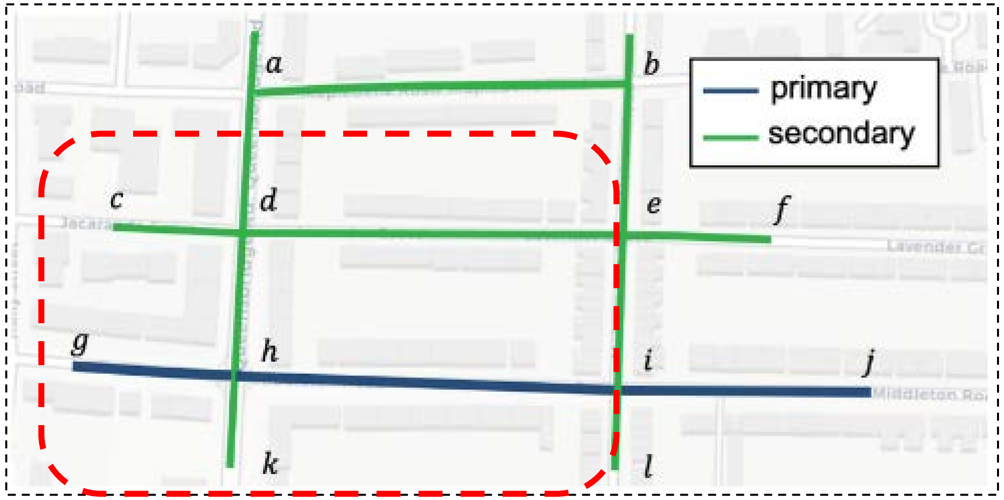


- ▷ **Target:** derive effective representations that are robust and generic for downstream applications.
  - Road segment-based & trajectory-based applications

Yile Chen, Xiucheng Li, Gao Cong, Zhifeng Bao, Cheng Long, Yiding Liu, Arun Kumar Chandran, Richard Ellison: Robust Road Network Representation Learning: When Traffic Patterns Meet Traveling Semantics. CIKM 2021

Challenges

▷ **Common assumptions in graph learning may not hold**



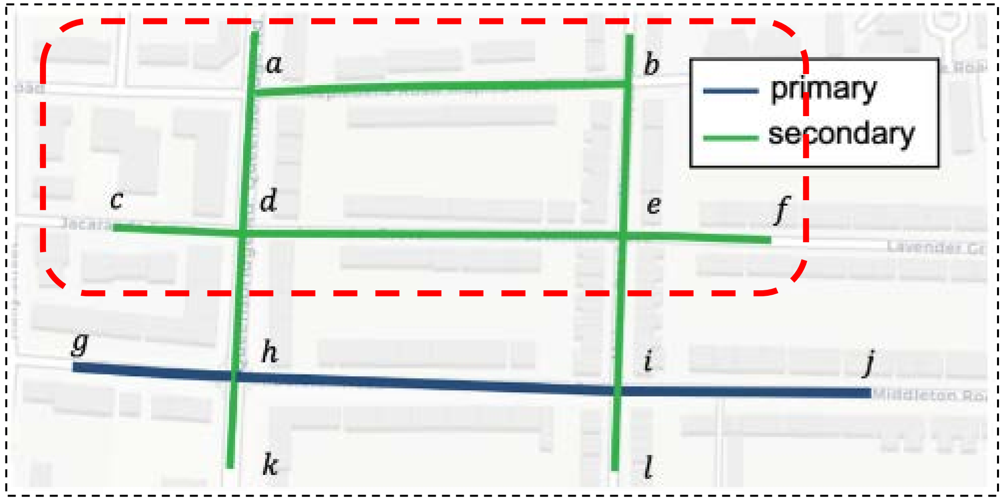
▷ According to homophily, inter-connected nodes are more similar than distant ones.

$\rightarrow$ *dh*, *gh*, *hi*, *hk* should be similar

▷ In reality, *dh*, *hk* (secondary roads) have less traffic volume than *gh*, *hi* (primary roads).

Challenges

▷ **Feature uniformity**



▷ Road networks in some regions share same features (e.g., residential area).

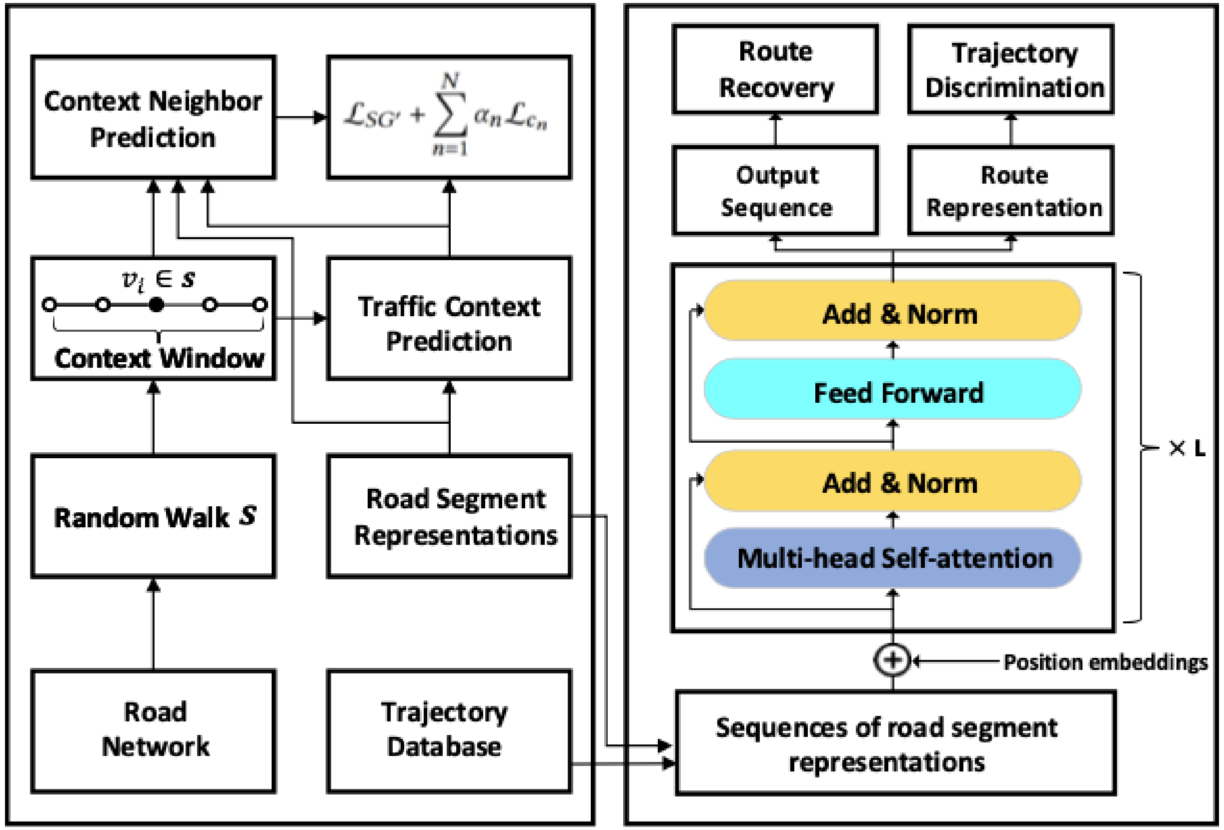
→ *GNN aggregation will end up with same representations*

▷ *de*, *ad*, *ab*, *be* will be the same.

→ undesirable: *de* should be more correlated to *cd*, *ef*.

Method - Toast

▷ Overview



Traffic context aware skip-gram module

Trajectory-enhanced Transformer module

▷ **Traffic context-aware skip-gram module:**

- Capture **traffic patterns** (e.g., volume) to distinguish the discrepancies in challenge 1.

▷ **Trajectory-enhanced Transformer module:**

- Consider **traveling semantics** (e.g., transition patterns) to avoid feature uniformity in challenge 2.

Experiments

▷ **Datasets:**

- 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

Road segment-based

- Trajectory similarity search
- Travel time estimation

Trajectory-based

Experiments

▷ **Baselines:**

❑ **Road segment representation learning**

- Representative graph embedding methods:
  - Deepwalk (KDD' 14), node2vec (KDD' 16), GAE (NIPS' 16), GraphSAGE (NIPS' 17)
- Road segment specific embedding methods:
  - RFN (SIGSPATIAL' 19), IRN2Vec (SIGSPATIAL' 19), HRNR (KDD' 20)

❑ **Trajectory representation learning**

- para2vec (ICML' 14), t2vec (ICDE' 18)

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**

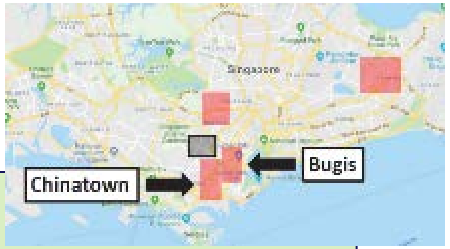
Travel time estimation

	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**

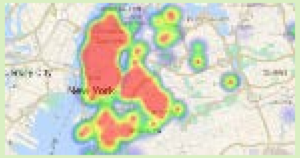
Outline

- Target on a specific problem on **point spatial entity**
  - Geospatial IR or Spatial Keyword Search (VLDB’09--SIGMOD’23)
  - POI recommendations (SIGIR’13 --)
  - Spatial relationship extraction (SIGMOD’23)
- Self-supervised learning for geospatial entity representation
  - Road Network Representation for Road Network Applications (CIKM’21)
  - Region Representation for Region-Level Applications (KDD’23)
  - Application of Foundation Models for Geospatial Applications
  - Efforts toward City Foundation Models.

Region Spatial Entity Data



Region Detection (Burst region) TKDE’22



Applications：event detection

Region Search

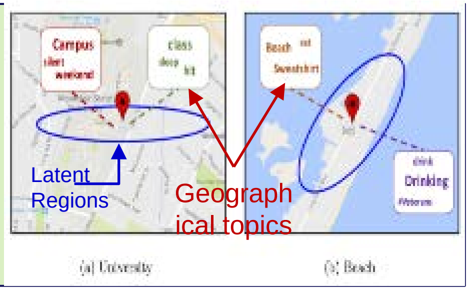
KDD’18,SIGMOD’16, PVLDB’16,ICDE’18, VLDBJ’19,VLDB’22

Applications：
Similar region search for urban planning,
Keyword based search

Region Recommendation & User Mobility Prediction

[WSDM’17, WWW’17]

Geo-topic modeling TKDE’22



Region topic exploration

SIGMOD’16， VLDBJ

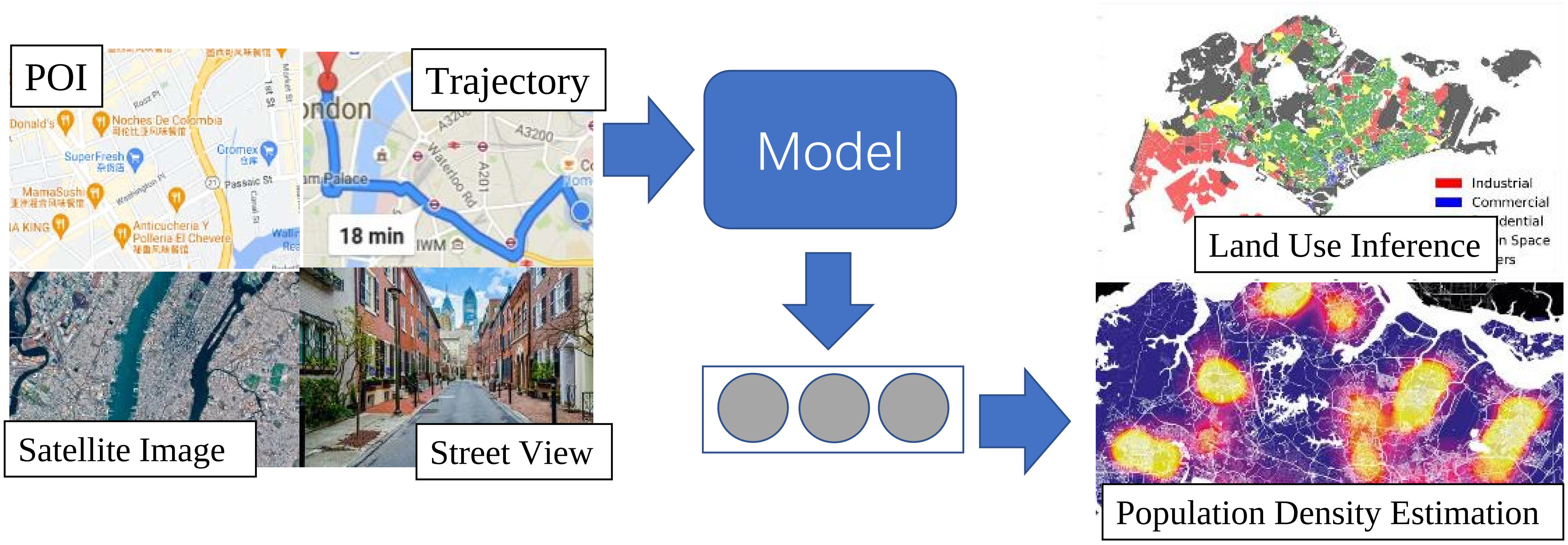
Business analytics

Deep Region Representation KDD’23

Applications：
Region analysis, such as functions, property price, crime rate, populations, etc

Problem of Urban Region Representation Learning

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



Data Representations 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

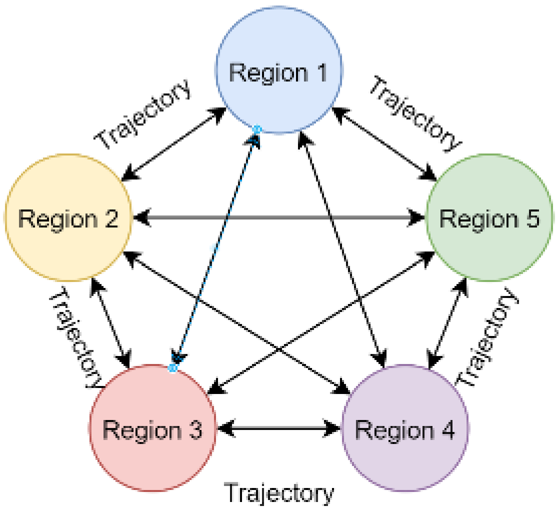
Introduction

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

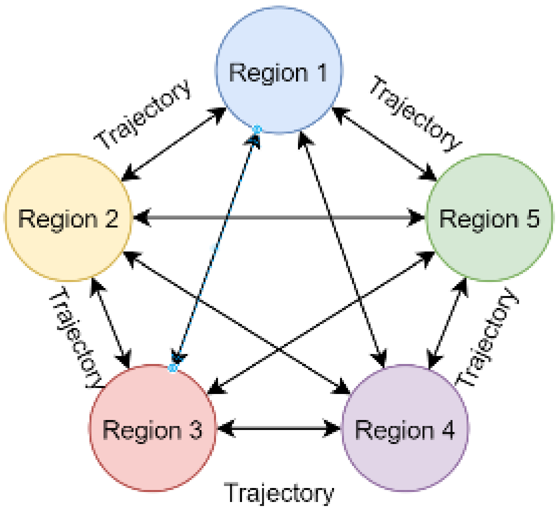
Related Work



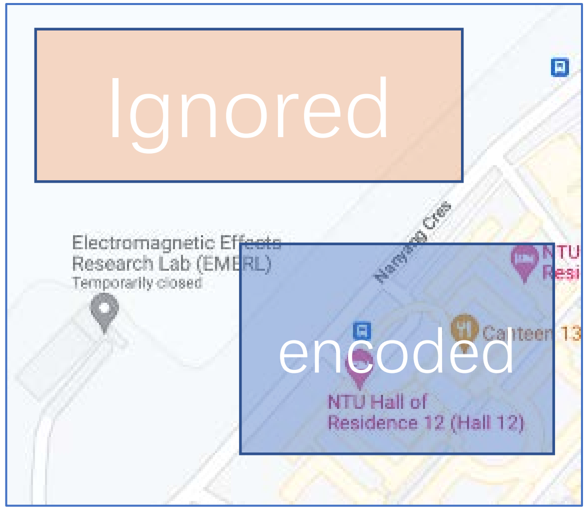
- Heavily rely on trajectory data

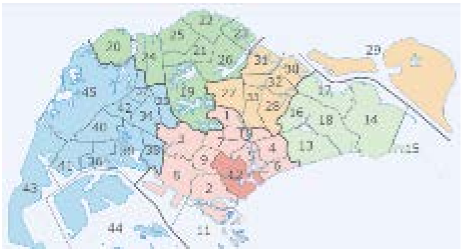
Model	Data Source
Place2vec [29]	POI
Doc2Vec [30]	POI
HDGE [8]	trajectory, POI, demographic/geographic features
ZE-Mob [9]	trajectory, check-in
Fu et al. [11]	trajectory, POI
Zhang et al. [12]	trajectory, POI
ReMVC [31]	trajectory, POI
RegionEncoder [10]	trajectory, POI, satellite image
MVURE. [13]	trajectory, POI, check-in
MGFN [14]	trajectory
RegionDCL	building footprints, POI

Related Work

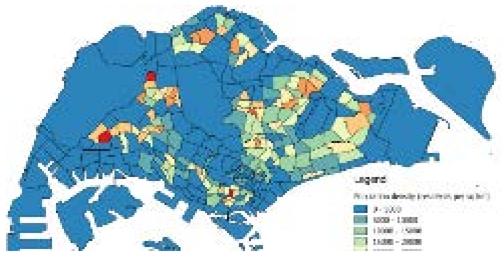


- Heavily rely on trajectory data
- Ignore data-sparse areas
- Can't adapt to multiple region partition schemes





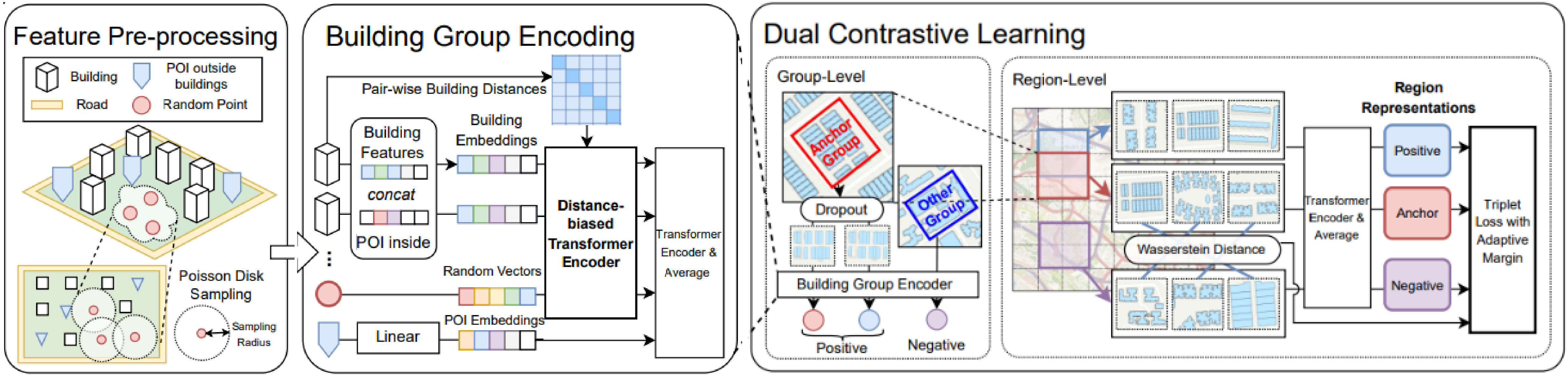
Regions in Land Use Classification



Regions in Population Density Estimation

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

- Dataset: Singapore & New York City
- Partition=Singapore Subzone & New York Census Block

Table 1: Dataset Statistics

City	Buildings	POIs	Building Patterns
Singapore	109,877	17,088	5,824
New York City	1,081,256	41,963	29,008

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**

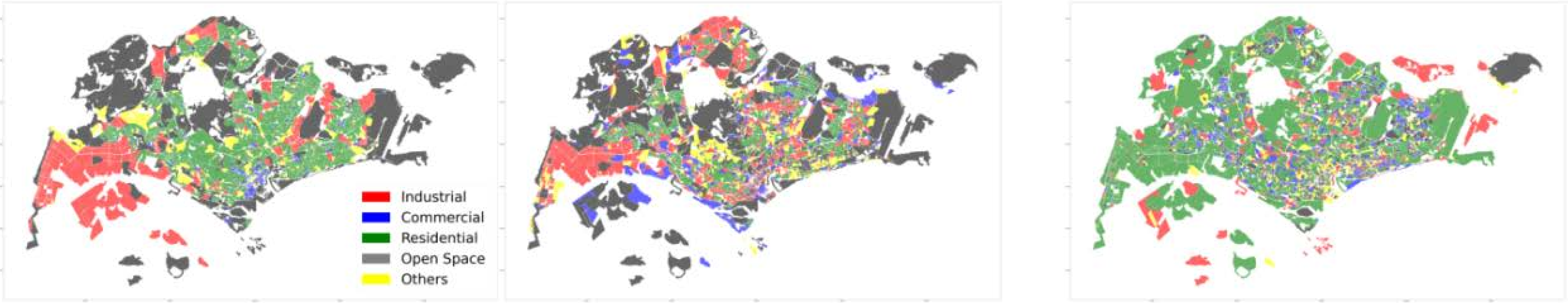
Population Density Inference

- 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.

Outline

- Target on a specific problem on **point spatial entity**
  - Geospatial IR or Spatial Keyword Search (VLDB’09--SIGMOD’23)
  - POI recommendations (SIGIR’13 --)
  - Spatial relationship extraction (SIGMOD’23)
- Self-supervised learning for geospatial entity representation
  - Road Network Representation for Road Network Applications (CIKM’21)
  - Region Representation for Region-Level Applications (KDD’23)
  - Application of Foundation Models for Geospatial Applications
  - Efforts toward City Foundation Models.

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 <u>Internet-scale</u> 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)

GPT3

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)

DALL·E 2

Zero-shot Transfer

Various CV Tasks

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

Large Language Model

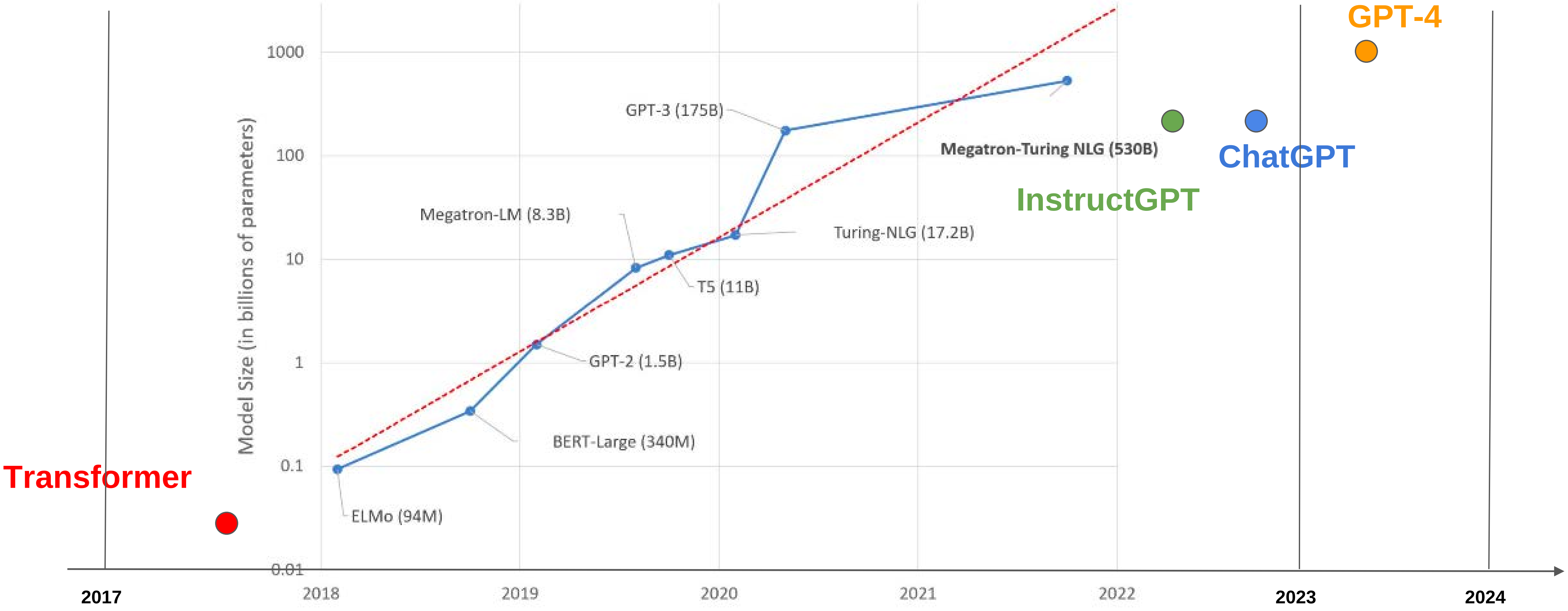
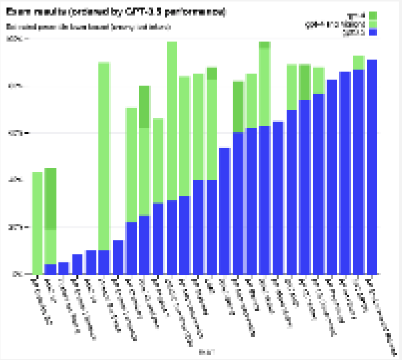


Image from Huggingface (https://huggingface.co/blog/large-language-models)

Foundation Models in Different Domains

Natural Language Processing

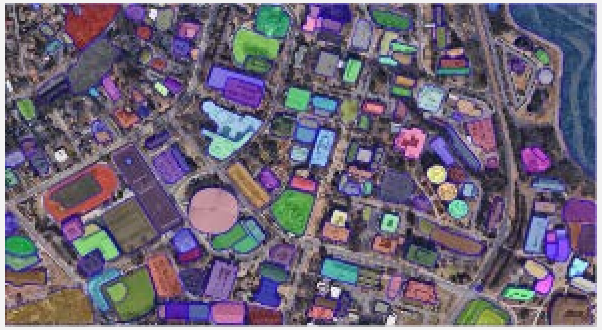
Stanford Alpaca



ChatGPT/GPT-4 (OpenAI. 2023)

Computer Vision

Imagen (Saharia et al. 2022)

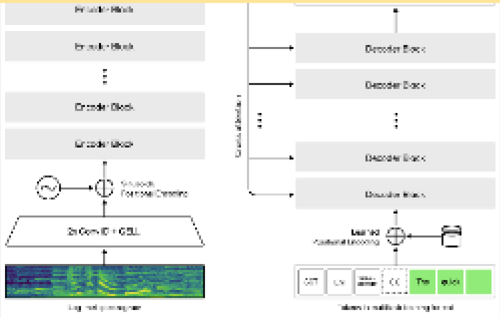


Segment Anying (Kirillov et al, 2023)

Reinforcement Learning

Gato (Reed et al. 2022)

Signal Processing



Whisper (Radford et al. 2022)

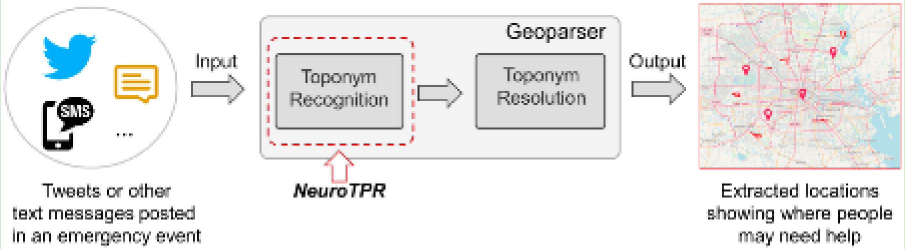
The Big Question

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.

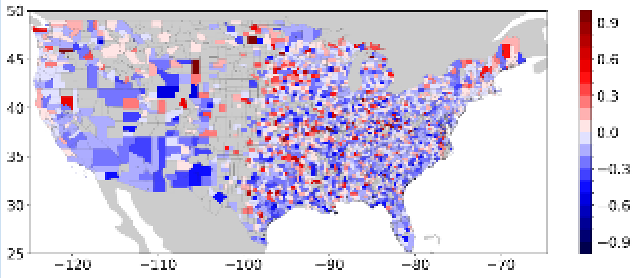


Urban Geography

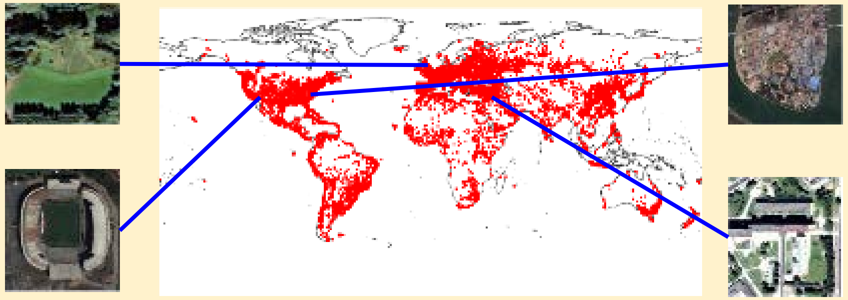
Urban Function

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
```

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
```

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

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 *SingaporeSVI579* 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

Remote Sensing

Task 7: Remote Sensing Image Scene Classification

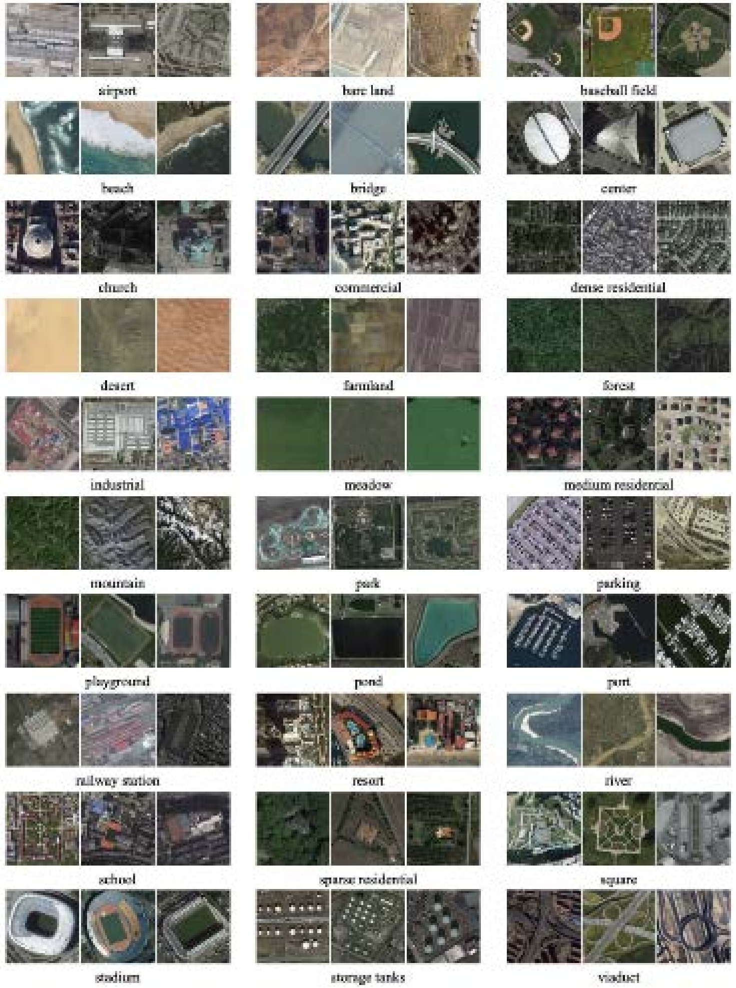


Figure 1: Samples of AID: three examples of each semantic scene class are shown. There are 10000 images within 30 classes.

Figure from Xia et al. (2016)

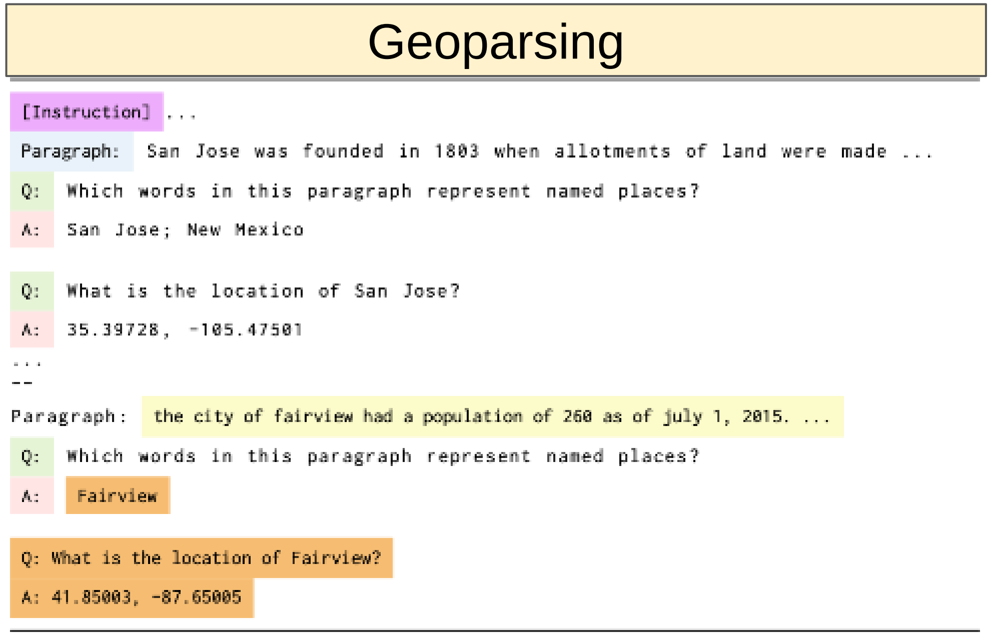
Table 7. Evaluation results of various vision-language foundation models and baselines on the remote sensing image scene classification dataset, *AID* [144]. We use the same model set as Table 6. "(*Origin*)" denotes we use the original remote sensing image scene class name from *AID* to populate the prompt while "(*Updated*)"indicates we update some class names to improve its semantic interpretation for FMs. We use accuracy and F1 score as evaluation metrics.

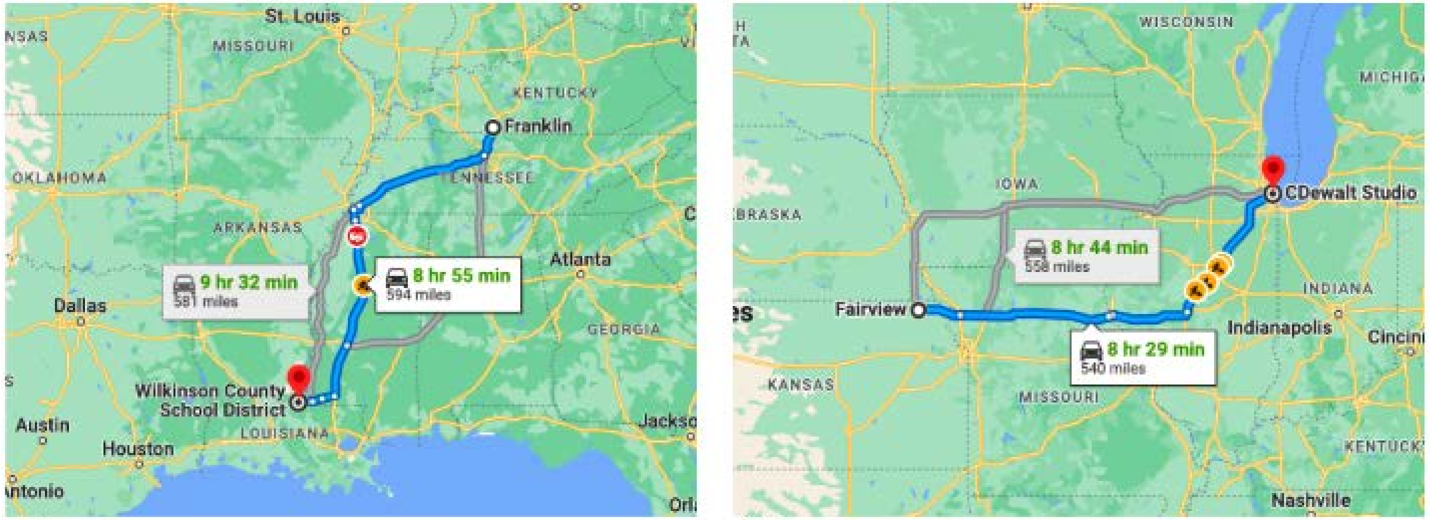
	Model	#Param	Accuracy	F1
Supervised Finetuned CNNs	AlexNet [74]	58M	**0.831**	**0.827**
	ResNet18 [37]	11M	0.752	0.730
	ResNet50 [37]	24M	0.757	0.738
	DenseNet161 [48]	27M	0.818	0.807
Zero-shot FMs	OpenCLIP-L (*Origin*) [54, 113, 127]	427M	0.708	0.688
	OpenCLIP-L (*Updated*) [54, 113, 127]	427M	**0.710**	**0.698**
	OpenCLIP-B (*Origin*) [54, 113, 127]	2.5B	0.699	0.668
	OpenCLIP-B (*Updated*) [54, 113, 127]	2.5B	0.705	0.686
	BLIP (*Origin*) [82]	2.5B	0.500	0.473
	BLIP (*Updated*) [82]	2.5B	0.520	0.494
	OpenFlamingo-9B [11]	8.3B	0.206	0.154

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

- The predicted coordinates are not accurate





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

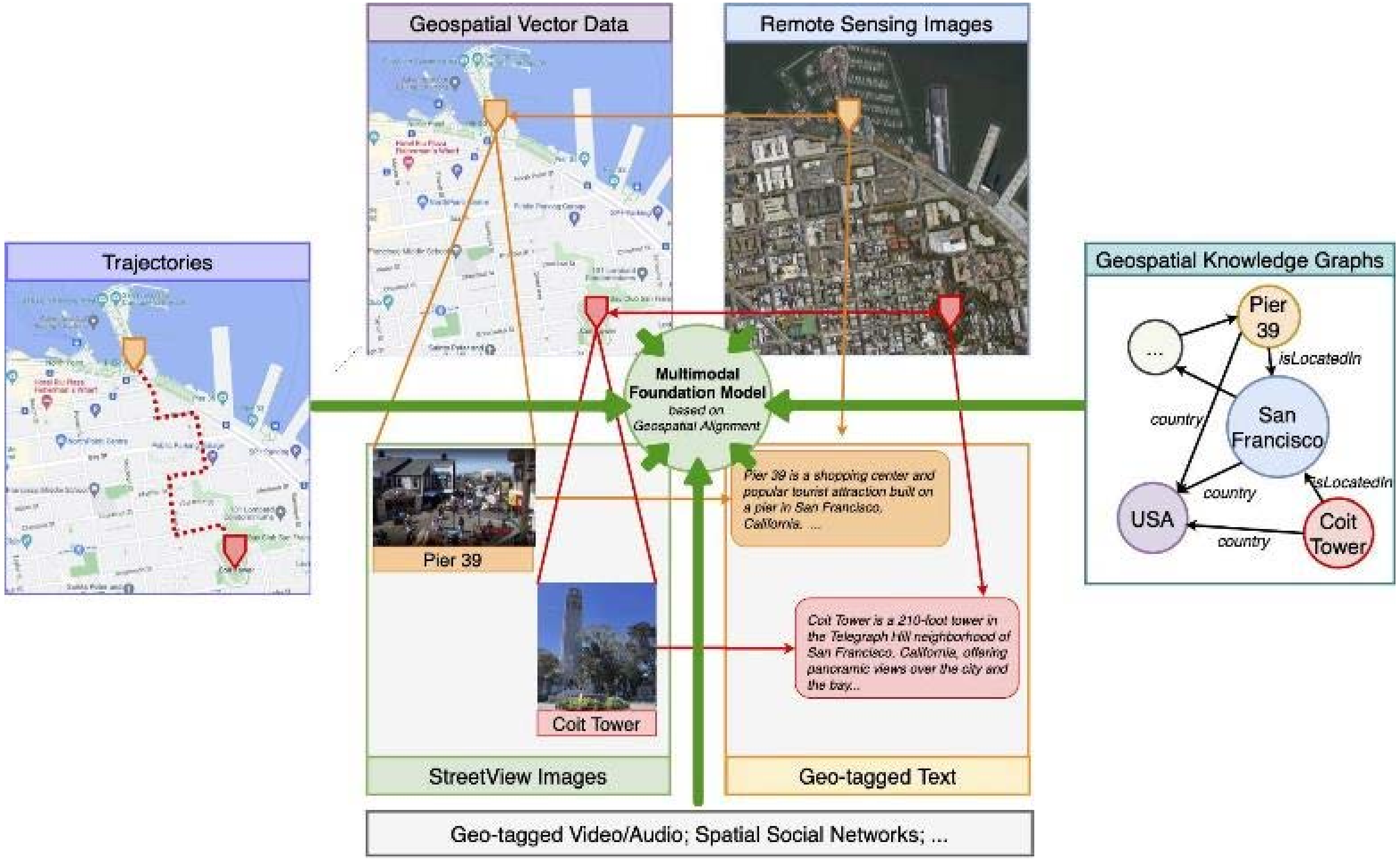
(b) [TEXT]: the city of Fairview had a population of 260 as of july 1, 2015. ...

Outline

- Target on a specific problem on **point spatial entity**
  - Geospatial IR or Spatial Keyword Search (VLDB’09--SIGMOD’23)
  - POI recommendations (SIGIR’13 --)
  - Spatial relationship extraction (SIGMOD’23)
- Self-supervised learning for geospatial entity representation
  - Road Network Representation for Road Network Applications (CIKM’21)
  - Region Representation for Region-Level Applications (KDD’23)
  - Application of Foundation Models for Geospatial Applications
  - Efforts toward City Foundation Models.

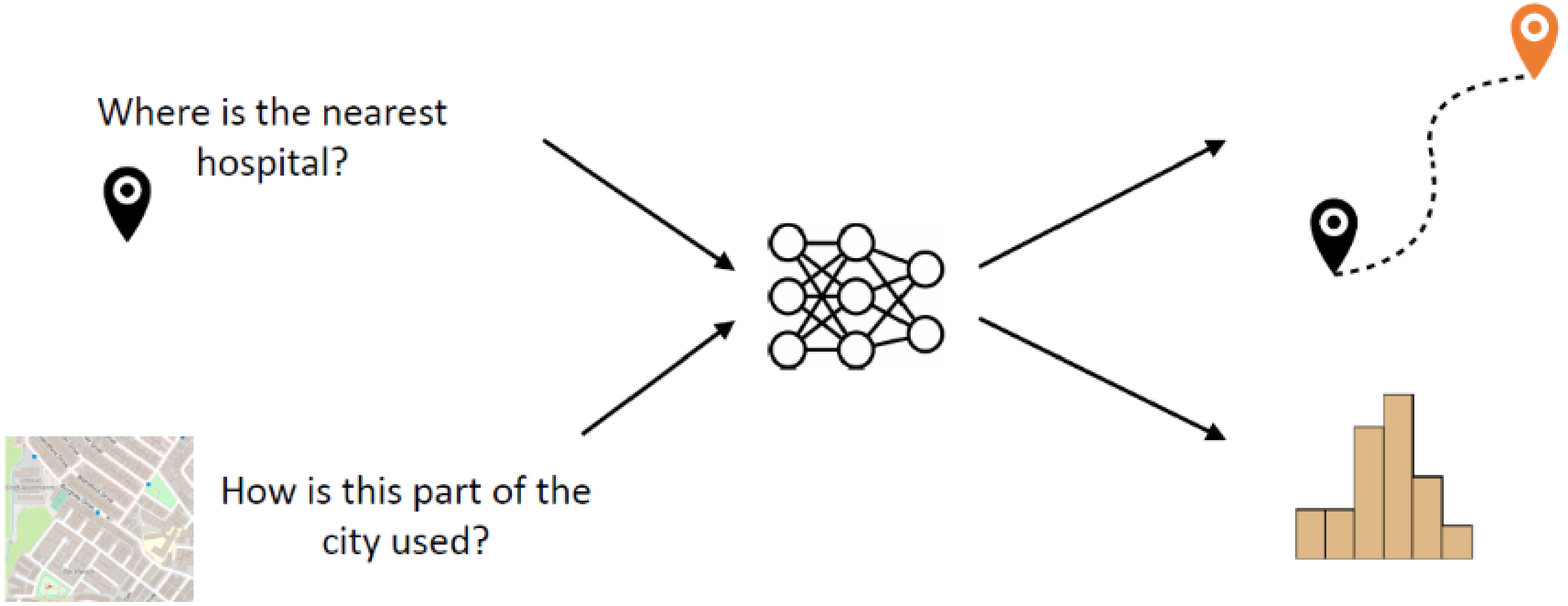
A Multimodal City FM for GeoAI

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



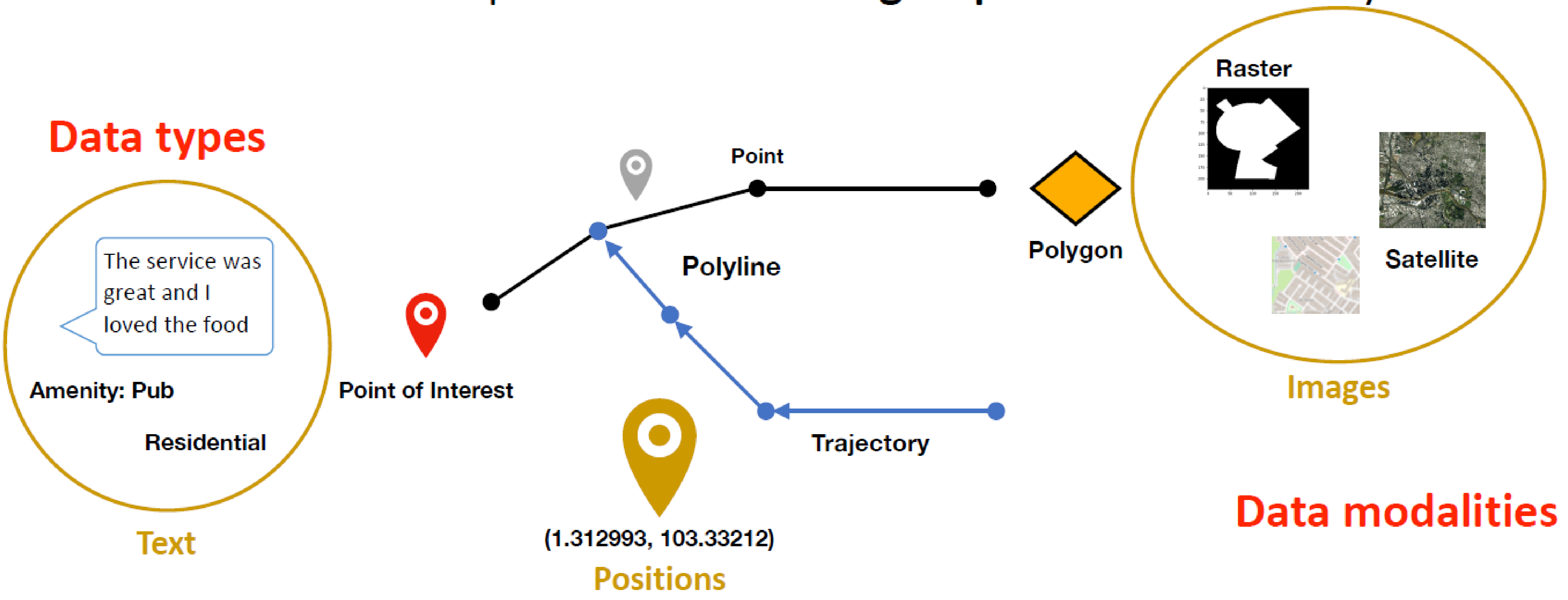
Motivations of City Foundation Models

FMs have the potential to revolutionise the way we use **geospatial data**



Challenges

A slower adoption of FMs in the **geospatial** domain… why?



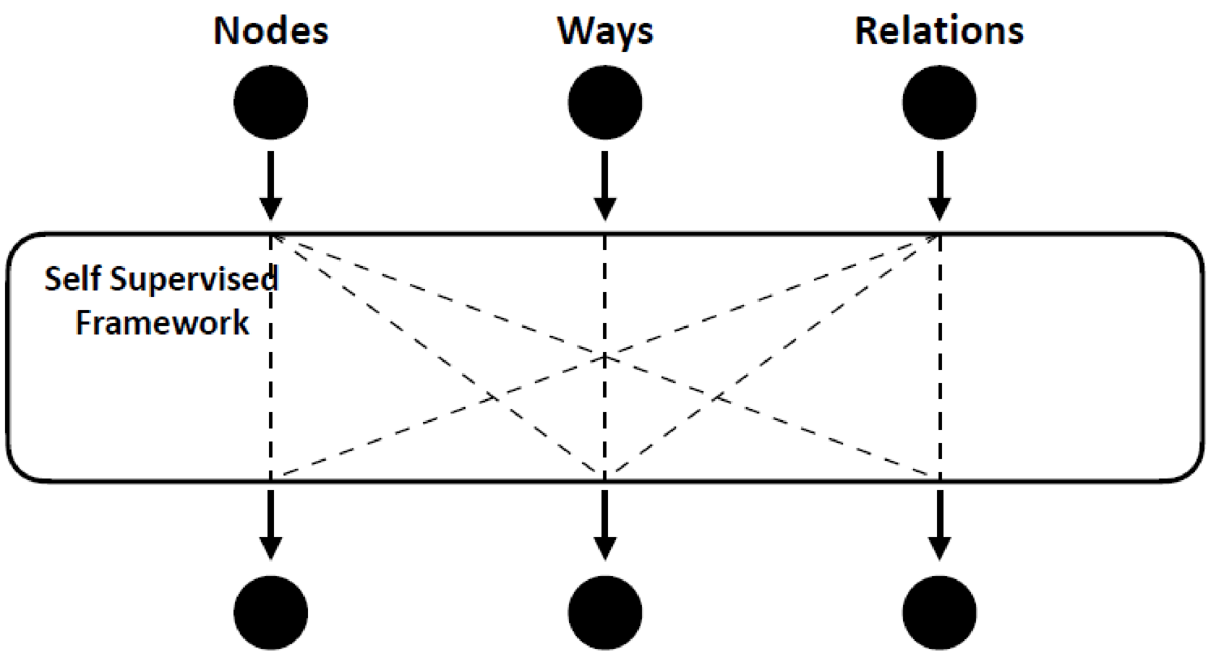
Challenges

A slower adoption of FMs in the **geospatial** domain... why?

Data sources also present a challenge, different data comes from different providers, and is available in different places!

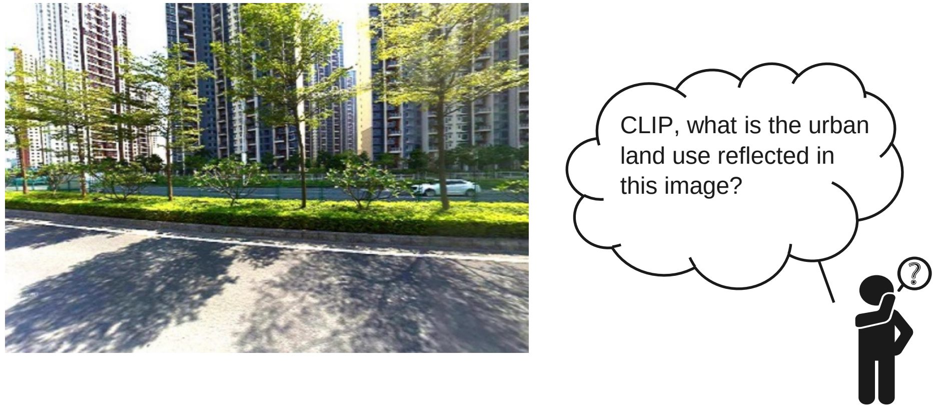
Attemp1: Use OpenStreetMap to Build a City Foundation Model

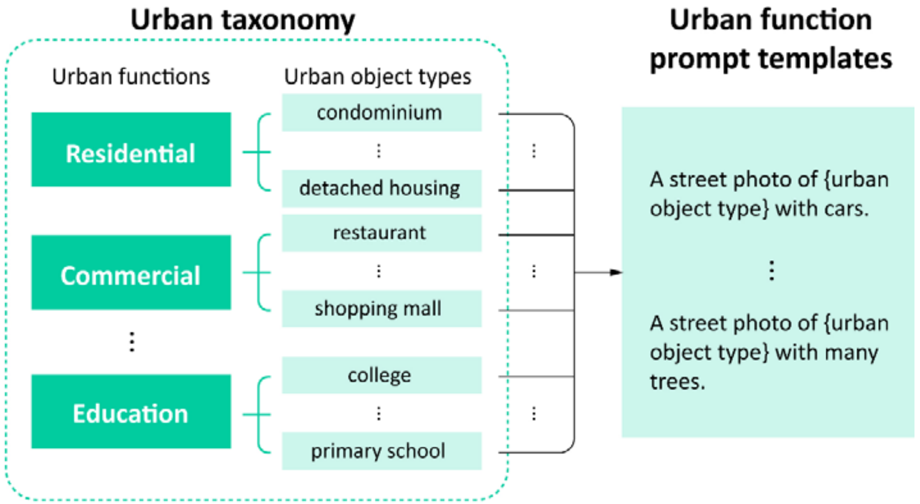
How to leverage the different data types and modalities in OSM, to **pre-train a geospatial FM**?

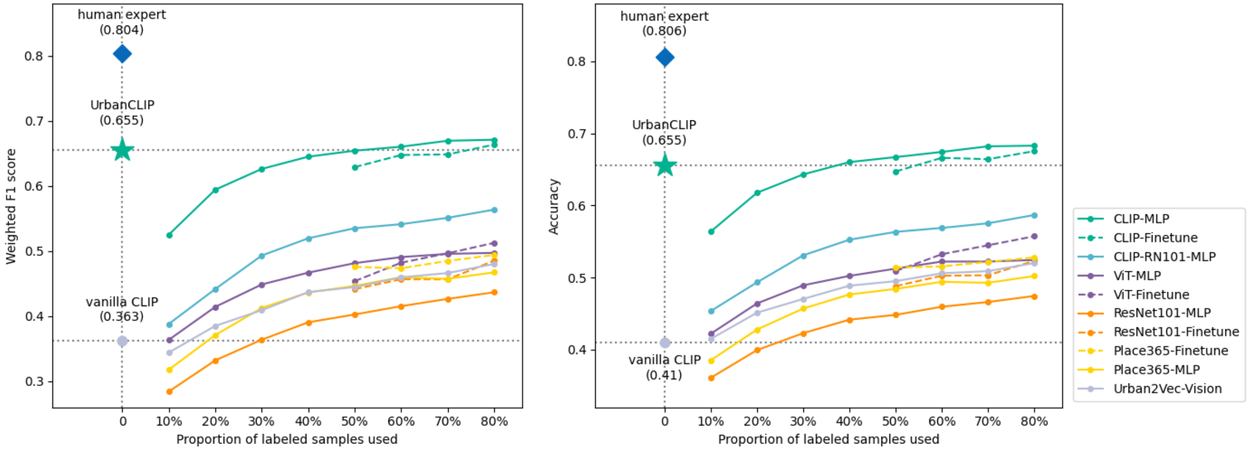


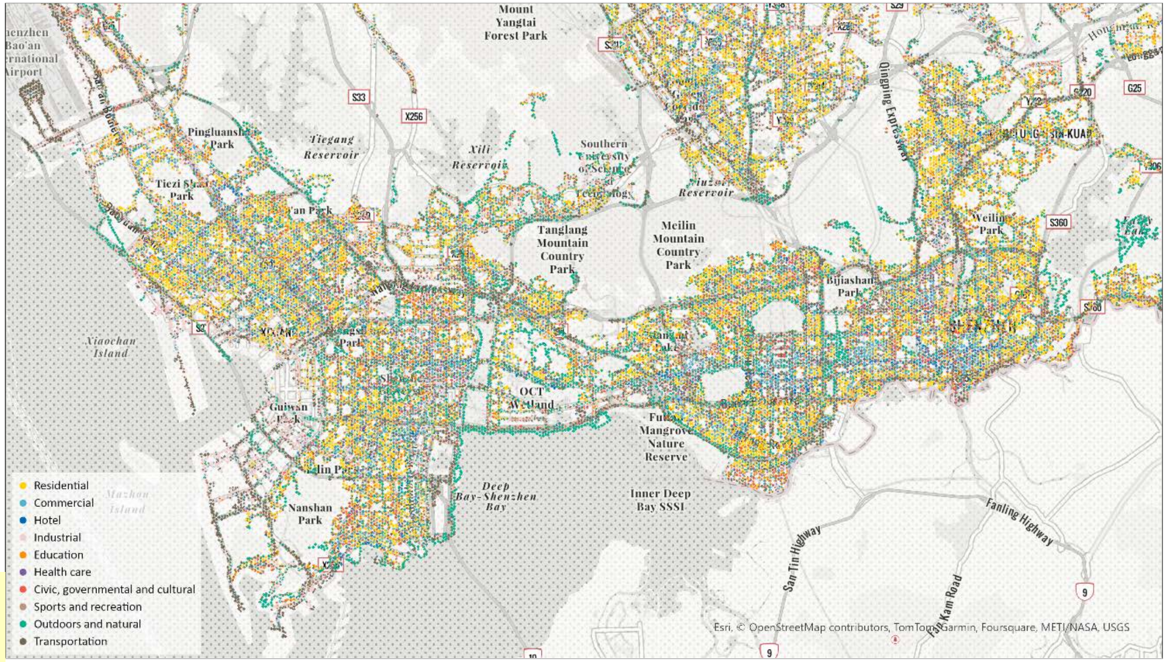
Pasquale Balsebre, et al. **Towards City Foundation Models**. On arXiv.

Attemp2: UrbanCLIP – a prompting framework for zero-shot urban land use inference



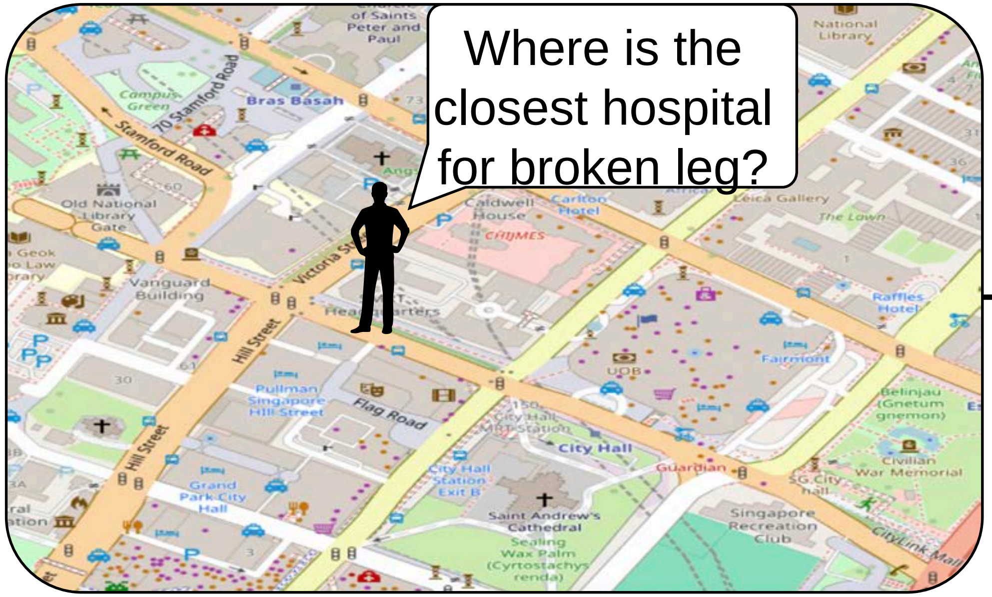






Huang, W., Wang, J., Cong, G. **Zero-shot urban function inference with street view images through prompting a vision-language model.** *International Journal of Geographical Information Science,* in press.

LLM for conversational city search



…such general models lack **city-specific knowledge**!

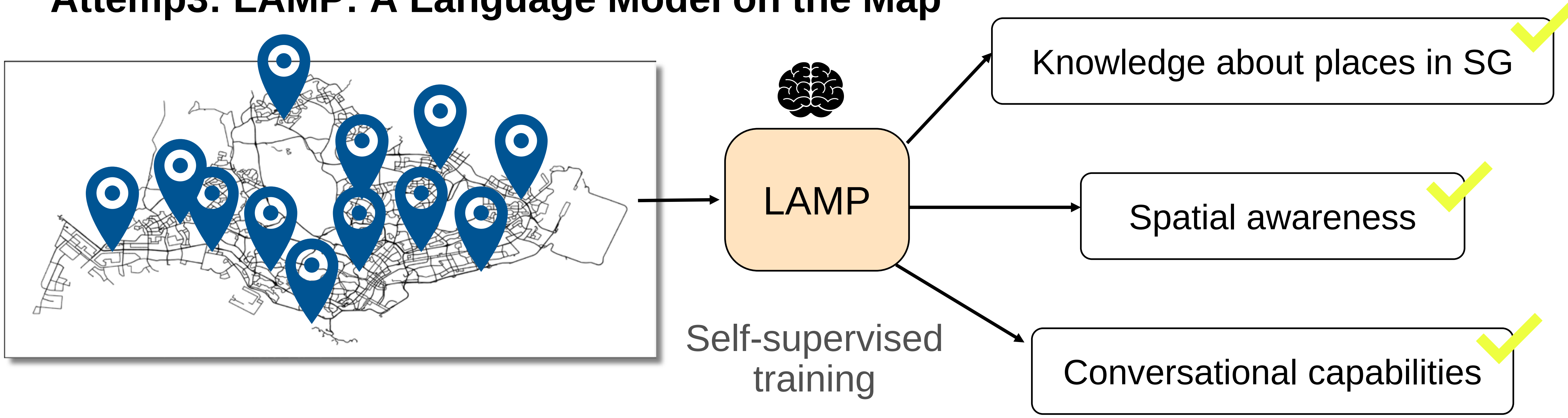
LLM for conversational city search

This type of place queries in plain language

- Can hardly be answered by traditional keyword-based map engines
- Are conversational in nature
- Require city-specific knowledge and spatial awareness, which current general purpose LLMs lack.

How about developing a chatbot on the map?

Attemp3: LAMP: A Language Model on the Map



We trained LAMP on Singapore places data, including POIs, addresses, opening hours, positions…

In this way, LAMP becomes aware of fine-grained urban places in Singapore, with a spatial sense (proximity)

Pasquale Balsebre, Weiming Huang, Gao Cong. **LAMP: A Language Model on the Map**. On arXiv.

LAMP: A Language Model on the Map

Current position: 81, Victoria Street, 188065, Singapore

Query: I'd like to have japanese food. Could you find a place nearby?

LLaMa-2-70B-Chat: Sure, I'd be happy to help! There are several Japanese restaurants near Victoria Street, Museum, Singapore. Here are a few options:

1. Sushi Tei - Located at 100 Victoria Street, #01-01, Singapore 188064.
2. Tsukemen Ginza - Located at 111 Middle Road, #01-01, Singapore 188969.
3. Ramen Nagi - Located at 100 Victoria Street, #01-02, Singapore 188064.

ChatGPT 3.5: Sure, one nearby Japanese restaurant to 81 Victoria Street, Singapore 188065 is "Shin Minori Japanese Restaurant" located at 81 Victoria Street, #01-01, Singapore 188065.

LAMP: I found two Japanese restaurants near your location:

- Shirokiya, located at 30 Victoria Street Chijmes, #01-05/06, 187996,
- Gyu Kaku Japanese BBQ, located at 30 Victoria Street, CHIJMES #01-01/03, 187996. It offers takeaway services and it is open until 11pm.

Let me know if you need any more recommendations!

When queried to find a specific POI. ChatGPT and LLaMa-2 are prone to hallucinating, and create places that do not exist, to satisfy the user's query, often replicating the address of the user or their zip code.

hallucinated information shaded in Orange, and correct information in Green.

LAMP: A Language Model on the Map

Current position: Rochor Road, Bugis, Downtown Core, Singapore, Central, 189352, Singapore

Query: Please recommend me some nearby food options

ChatGPT-4: After a quick search, I found a variety of food options near Rochor Road, Bugis.
Here are some examples:

1. 328 Katong Laksa: famous for its Laksa, a spicy noodle soup.
2. Jing Hua Xiao Chi: offers Chinese and Asian cuisine.

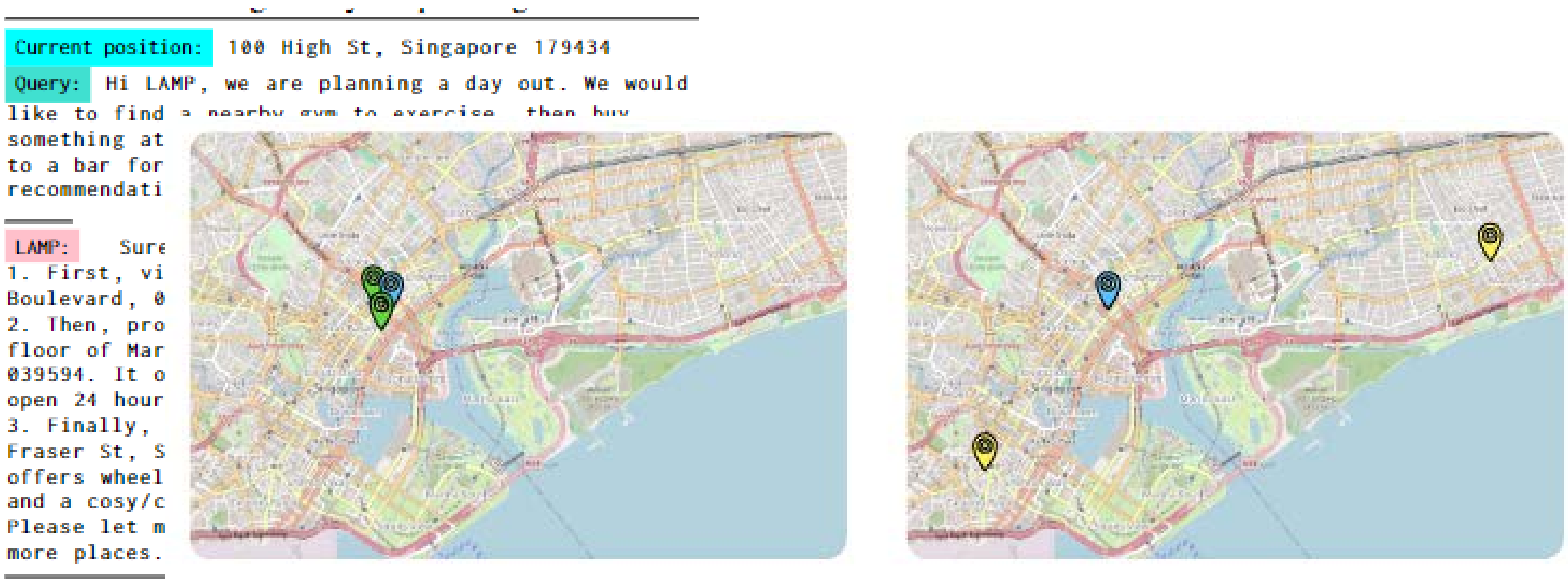
LAMP: Here are 2 possible choices:

- THE SALADBOX, 191 Rochor Rd, #B2-08, Bugis MRT Station, 188476 ,
- Ji De Chi, 8 Liang Seah St, 189029. It is highly rated and has a price range of $$!

In some cases, GPT-4 provides places that are both relevant and existing. But such places are farther apart. In contrast, LAMP continues to provide nearby relevant places.

Relevant while far away information shaded in Yellow , and correct information in Green

LAMP: A Language Model on the Map



The left places are recommended by LAMP, while the right ones are from GPT-4.

LAMP: A Language Model on the Map

Model	Truthfulness	Spatial Awareness	Semantic Relatedness
LLaMa-2-7B	0.12	0.2	0.76
LLaMa-2-70B	0.3	0.36	0.94
Claude	0.22	0.32	0.96
ChatGPT-3.5	0.68	0.6	0.98
ChatGPT-4 (Browsing)	**0.94**	0.82	**1.0**
LAMP-no-rq	0.76	0.84	**1.0**
LAMP	0.86	**0.92**	**1.0**

Table 2: Main results on conversational POI-retrieval in Singapore.

Overall, LAMP has a broad range of impacts in a smart city context:

- Can help users with every day's needs, finding places or planning a day out;
- Can increase the public services efficiency, e.g., find the closest government office for some service.
- In general, could assist urban planners and GIS experts in their job.

References and Acknowledgement

- Shang Liu, Gao Cong, Kaiyu Feng, Wanli Gu, Fuzheng Zhang. **Effectiveness Perspectives and a Deep Relevance Model for Spatial Keyword Queries.** SIGMOD 2023
- Pasquale Balsebre, Dezhong Yao, Weiming Huang, Gao Cong, Zhen Hai. **Mining Geospatial Relationships from Text.** SIGMOD 2023
- Yile Chen, Xiucheng Li, Gao Cong, Zhifeng Bao, et al.: **Robust Road Network Representation Learning: When Traffic Patterns Meet Traveling Semantics**. CIKM 2021
- Yi Li, Weiming Huang, Gao Cong, Hao Wang, and Zheng Wang. **Urban Region Representation Learning with OpenStreetMap Building Footprints**. SIGKDD 2023
- Gengchen Mai, Weiming Huang, et al. . On the Opportunities and Challenges of Foundation Models for Geospatial Artificial Intelligence. ACM TSAS 2024
- Weiming Huang, Jing Wang, Gao Cong. **Zero-shot urban function inference with street view images through prompting a pre-trained vision-language model**. IJGIS in press.
- Pasquale Balsebre, Weiming Huang, Gao Cong. **LAMP: A Language Model on the Map**. On arXiv.
- Pasquale Balsebre, Weiming Huang, Gao Cong, Yi Li **Towards City Foundation Models**. On arXiv.


Shang Liu


Yi Li


Pasquale Balsebre


Yile Chen


Di Yao


Kaiyu Feng

Xiucheng Li


Gengchen Mai


Weiming Huang