Geospatial Entity Representation: A Step Towards City Foundation Models

Gao Cong

Nanyang Technological University

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)

Spatial 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

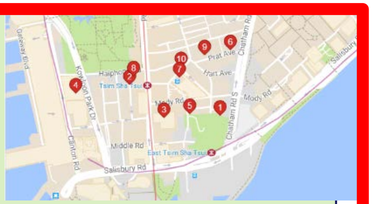
Geo-spatial + X (e.g., text, temporal) data

Outline

- Target on a specific problem on **point spatial entity**
  - 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.

Our Research on Point Spatial Entity

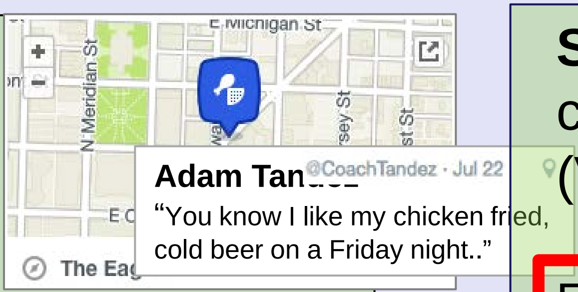
Geospatial IR
(Spatial keyword query)
(VLDB’09 --)



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)

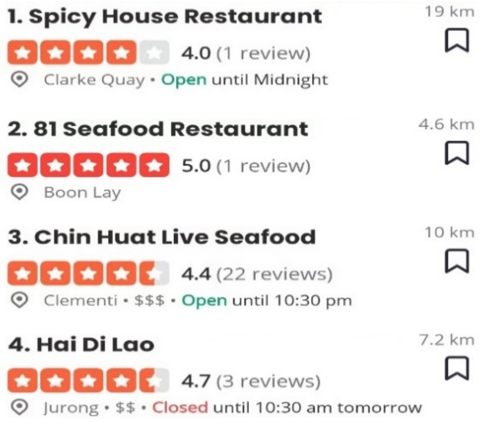
Geo-textual Data (Spatial-textual Data)

- Geo-textual data contains attributes:
  - Geo-location (*latitude, longitude*)
  - Textual content to describe the object
- Large amount of geo-textual data is being generated
  - Web pages/documents with geo-info
  - Points of Interest (POI)
  - Micro-blogging applications
  - Geo-tagged multimedia data

Spatial Keyword Query (Geographic IR)

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

- Applications of spatial keyword query:
  - o Geographic search engines
  - o location-based servic
  - o locally targeted web a



Spatial Keyword Query Example on Yelp

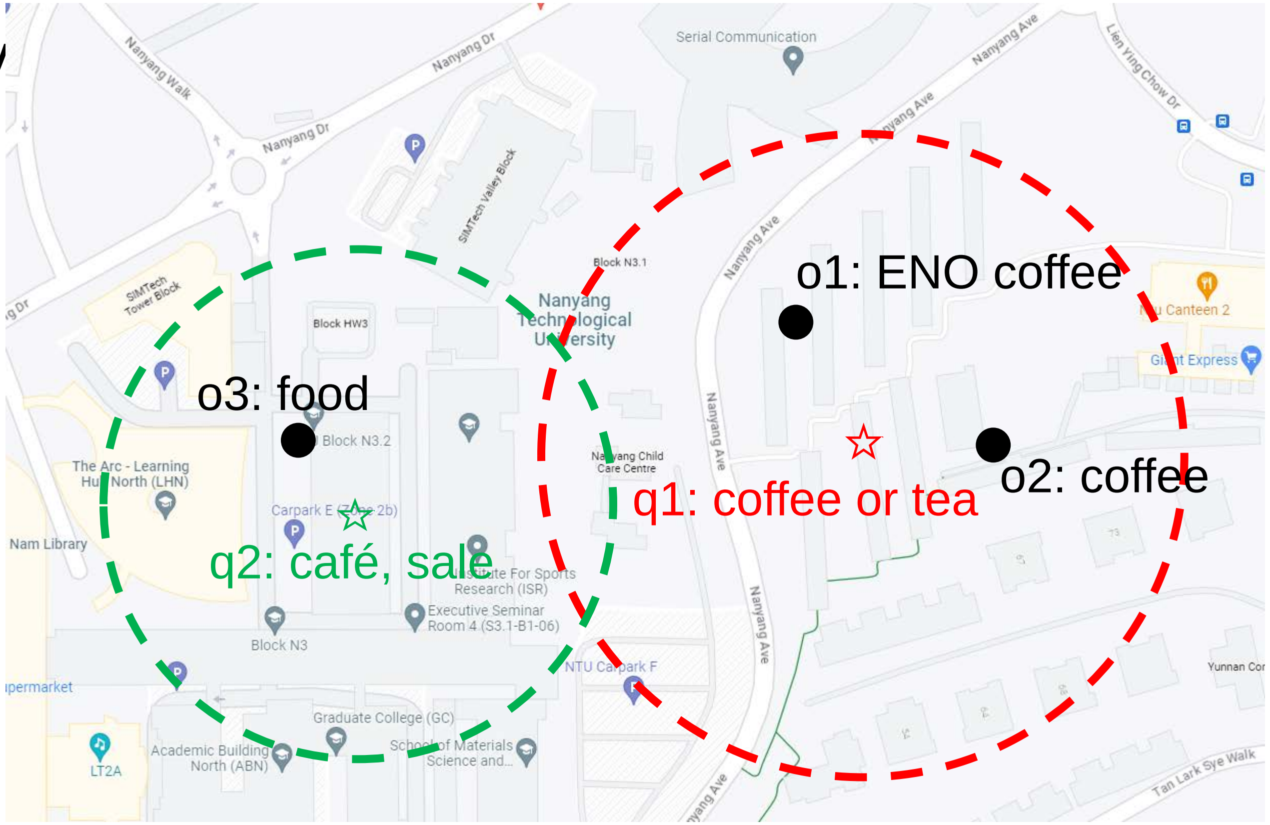
Spatial Keyword Query

Three types of basic spatial keyword queries:

- Boolean Range Query (BRQ)
- Boolean KNN Query (BkQ)
- Top-k KNN Query (TkQ)

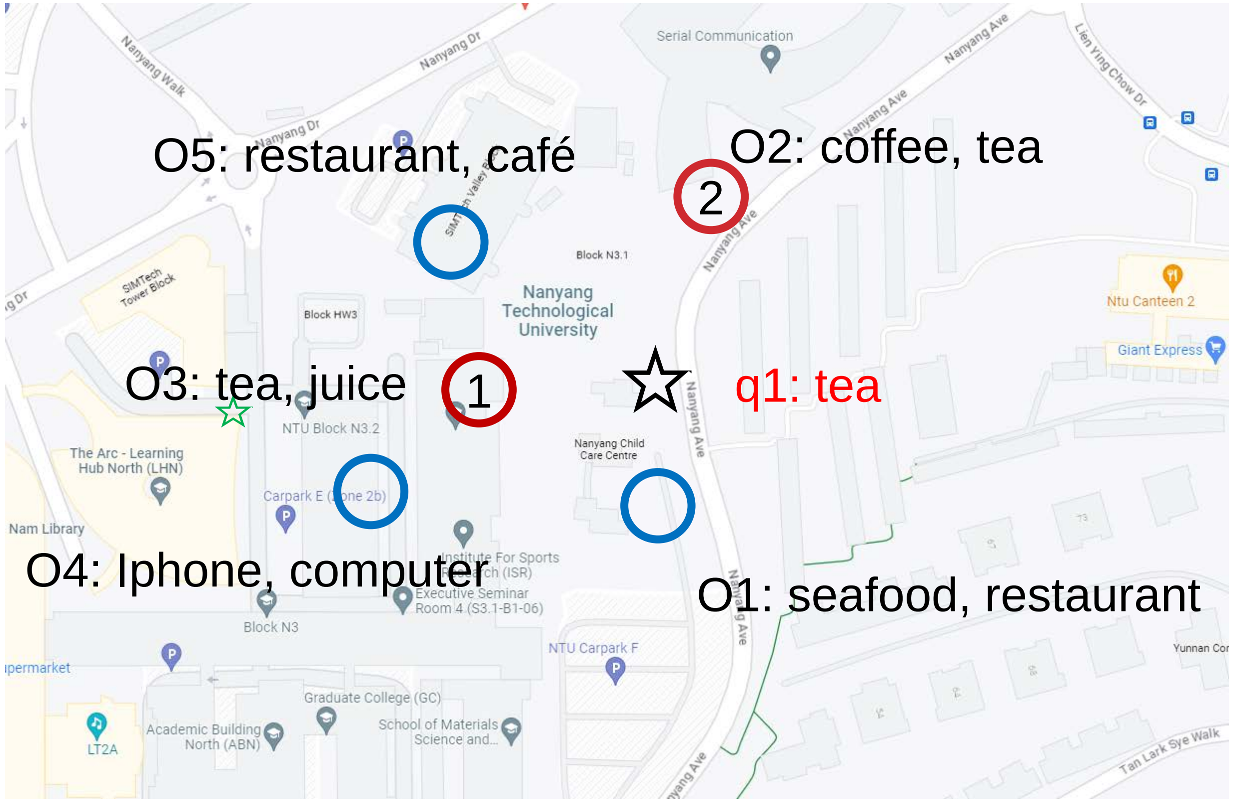
Boolean Range Query (BRQ)

- Semantics:
  - objects located inside the spatial range of a query and
  - matches the keywords of the query
- Extension: Top-k Boolean Range Query (TBRQ)



Boolean KNN Query (BkQ)

- Semantics:
  - o Identify all objects that contain query keywords
  - o Rank these objects based on the distance between objects and the location of the query



Top-k KNN Query (TkQ)

$$ST(p,q) = (1-\alpha) \times \underbrace{\boxed{(1 - SDist(p.loc, q.loc))}}_{\text{distance score}} + \alpha \times \underbrace{\boxed{TRel(p.doc, q.doc)}}_{\text{text score}},$$

- Retrieve the top-k objects ranked based on:
  - Distance between the object and the query location
  - Similarity between the keywords of the object and the query

Gao Cong, Christian S. Jensen, Dingming Wu: Efficient Retrieval of the Top-k Most Relevant Spatial Web Objects. PVLDB 2009

Evaluating Effectiveness of Spatial Keyword Queries

- Previous studies focus on **efficiency** with various indexes and query processing algorithms
- Open problem: **It is unknown about the effectiveness of the different types of queries**
- **Lack of ground truth query** makes it difficult to study the effectiveness
  - Two real-life query log datasets (Beijing and Shanghai)

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

Effectiveness Perspectives of Spatial Keyword Queries

- Datasets
  - Two datasets (Beijing and Shanghai) are from Meituan and anonymized to protect privacy.
- Evaluated queries
  - Boolean Range Query (BRQ)
  - Boolean KNN Query (BkQ)
  - Top-k KNN Query (TkQ)
  - Their variants

The basic statistics of the datasets.

	Dataset Beijing	Dataset Shanghai
Total Num of records	164,196	193,402
Total Num of Query	94,614	90,761
Total Num of POI	45,643	49,146
Avg Len of Query	5.34	5.32
Avg Len of POI text	26.76	26.04
Avg Distance (km)	5.22	8.27

Effectiveness Perspectives of Spatial Keyword Queries

The effectiveness evaluation results

	Beijing			Shanghai		
	NDCG@3	NDCG@5	MRR	NDCG@3	NDCG@5	MRR
TBRQ(1)	0.2075	0.2124	0.2077	0.1757	0.1862	0.1768
TBRQ(5)	0.2586	0.2817	0.2749	0.2401	0.2682	0.2500
TBRQ(7)	0.2777	0.2930	0.2909	0.2383	0.2550	0.2445
TBRQ(9)	0.2663	0.2868	0.2775	0.2337	0.2488	0.2458
TBRQ(13)	0.2197	0.2540	0.2453	0.2350	0.2515	0.2387
$TBRQ_v$	0.2951	0.3078	0.2971	0.2518	0.2730	0.2545
BkQ	0.3463	0.3638	0.3563	0.3372	0.3516	0.3428
$BkQ_v(0.3)$	0.3577	0.3781	0.3675	0.3643	0.3902	0.3730
$BkQ_v(0.5)$	0.3681	0.3950	0.3880	0.4069	0.4270	0.4109
$BkQ_v(0.7)$	0.3544	0.3722	0.3557	0.3692	0.3837	0.3609
$BkQ_v(0.9)$	0.2973	0.3000	0.2954	0.3010	0.3375	0.3109
TkQ(0)	0.0252	0.0299	0.0335	0.0286	0.0351	0.0373
TkQ(0.05)	0.3655	0.3812	0.3733	0.4007	0.4159	0.4021
TkQ(0.1)	**0.4099**	**0.4249**	**0.4058**	**0.4436**	**0.4608**	**0.4419**
TkQ(0.3)	0.3671	0.3928	0.3876	0.4068	0.4226	0.4099
TkQ(0.5)	0.3252	0.3478	0.3361	0.3465	0.3655	0.3660
TkQ(0.7)	0.2636	0.2840	0.2713	0.3006	0.3195	0.3024
TkQ(0.9)	0.2056	0.2256	0.2123	0.2545	0.271	0.2597
TkQ(1.0)	0.1651	0.1882	0.1768	0.2409	0.2558	0.2454
TkQ*	0.4596	0.4672	0.4474	0.5148	0.5297	0.5114

- When keywords are used as a Boolean filtering, spatial proximity ranking (BkQ) is better than text relevance ranking (TBRQ);
- Keyword information plays an indispensable role in POI search;

Effectiveness Perspectives of Spatial Keyword Queries

The effectiveness evaluation results

	Beijing			Shanghai		
	NDCG@3	NDCG@5	MRR	NDCG@3	NDCG@5	MRR
TBRQ(1)	0.2075	0.2124	0.2077	0.1757	0.1862	0.1768
TBRQ(5)	0.2586	0.2817	0.2749	0.2401	0.2682	0.2500
TBRQ(7)	0.2777	0.2930	0.2909	0.2383	0.2550	0.2445
TBRQ(9)	0.2663	0.2868	0.2775	0.2337	0.2488	0.2458
TBRQ(13)	0.2197	0.2540	0.2453	0.2350	0.2515	0.2387
$TBRQ_v$	0.2951	0.3078	0.2971	0.2518	0.2730	0.2545
BkQ	0.3463	0.3638	0.3563	0.3372	0.3516	0.3428
$BkQ_v(0.3)$	0.3577	0.3781	0.3675	0.3643	0.3902	0.3730
$BkQ_v(0.5)$	0.3681	0.3950	0.3880	0.4069	0.4270	0.4109
$BkQ_v(0.7)$	0.3544	0.3722	0.3557	0.3692	0.3837	0.3609
$BkQ_v(0.9)$	0.2973	0.3000	0.2954	0.3010	0.3375	0.3109
TkQ(0)	0.0252	0.0299	0.0335	0.0286	0.0351	0.0373
TkQ(0.05)	0.3655	0.3812	0.3733	0.4007	0.4159	0.4021
TkQ(0.1)	**0.4099**	**0.4249**	**0.4058**	**0.4436**	**0.4608**	**0.4419**
TkQ(0.3)	0.3671	0.3928	0.3876	0.4068	0.4226	0.4099
TkQ(0.5)	0.3252	0.3478	0.3361	0.3465	0.3655	0.3660
TkQ(0.7)	0.2636	0.2840	0.2713	0.3006	0.3195	0.3024
TkQ(0.9)	0.2056	0.2256	0.2123	0.2545	0.271	0.2597
TkQ(1.0)	0.1651	0.1882	0.1768	0.2409	0.2558	0.2454
TkQ*	0.4596	0.4672	0.4474	0.5148	0.5297	0.5114

- When keywords are used as a Boolean filtering, spatial proximity ranking (BkQ) is better than text relevance ranking (TBRQ);
- Keyword information plays an indispensable role in POI search;
- TkQ performs better than TBRQ and BkQ;

Effectiveness Perspectives of Spatial Keyword Queries

The effectiveness evaluation results

	Beijing			Shanghai		
	NDCG@3	NDCG@5	MRR	NDCG@3	NDCG@5	MRR
TBRQ(1)	0.2075	0.2124	0.2077	0.1757	0.1862	0.1768
TBRQ(5)	0.2586	0.2817	0.2749	0.2401	0.2682	0.2500
TBRQ(7)	0.2777	0.2930	0.2909	0.2383	0.2550	0.2445
TBRQ(9)	0.2663	0.2868	0.2775	0.2337	0.2488	0.2458
TBRQ(13)	0.2197	0.2540	0.2453	0.2350	0.2515	0.2387
$TBRQ_v$	0.2951	0.3078	0.2971	0.2518	0.2730	0.2545
BkQ	0.3463	0.3638	0.3563	0.3372	0.3516	0.3428
$BkQ_v(0.3)$	0.3577	0.3781	0.3675	0.3643	0.3902	0.3730
$BkQ_v(0.5)$	0.3681	0.3950	0.3880	0.4069	0.4270	0.4109
$BkQ_v(0.7)$	0.3544	0.3722	0.3557	0.3692	0.3837	0.3609
$BkQ_v(0.9)$	0.2973	0.3000	0.2954	0.3010	0.3375	0.3109
TkQ(0)	0.0252	0.0299	0.0335	0.0286	0.0351	0.0373
TkQ(0.05)	0.3655	0.3812	0.3733	0.4007	0.4159	0.4021
TkQ(0.1)	**0.4099**	**0.4249**	**0.4058**	**0.4436**	**0.4608**	**0.4419**
TkQ(0.3)	0.3671	0.3928	0.3876	0.4068	0.4226	0.4099
TkQ(0.5)	0.3252	0.3478	0.3361	0.3465	0.3655	0.3660
TkQ(0.7)	0.2636	0.2840	0.2713	0.3006	0.3195	0.3024
TkQ(0.9)	0.2056	0.2256	0.2123	0.2545	0.271	0.2597
TkQ(1.0)	0.1651	0.1882	0.1768	0.2409	0.2558	0.2454
TkQ*	0.4596	0.4672	0.4474	0.5148	0.5297	0.5114

- When keywords are used as a Boolean filtering, spatial proximity ranking (BkQ) is better than text relevance ranking (TBRQ);
- Keyword information plays an indispensable role in POI search;
- TkQ performs better than TBRQ and BkQ;
- The effectiveness of TBRQ increases and then drops as we increase the query radius;
- The effectiveness of TkQ is significantly affected by the parameter α;

Effectiveness Perspectives of Spatial Keyword Queries

The effectiveness evaluation results

	Beijing			Shanghai		
	NDCG@3	NDCG@5	MRR	NDCG@3	NDCG@5	MRR
TBRQ(1)	0.2075	0.2124	0.2077	0.1757	0.1862	0.1768
TBRQ(5)	0.2586	0.2817	0.2749	0.2401	0.2682	0.2500
TBRQ(7)	0.2777	0.2930	0.2909	0.2383	0.2550	0.2445
TBRQ(9)	0.2663	0.2868	0.2775	0.2337	0.2488	0.2458
TBRQ(13)	0.2197	0.2540	0.2453	0.2350	0.2515	0.2387
$TBRQ_v$	0.2951	0.3078	0.2971	0.2518	0.2730	0.2545
BkQ	0.3463	0.3638	0.3563	0.3372	0.3516	0.3428
$BkQ_v(0.3)$	0.3577	0.3781	0.3675	0.3643	0.3902	0.3730
$BkQ_v(0.5)$	0.3681	0.3950	0.3880	0.4069	0.4270	0.4109
$BkQ_v(0.7)$	0.3544	0.3722	0.3557	0.3692	0.3837	0.3609
$BkQ_v(0.9)$	0.2973	0.3000	0.2954	0.3010	0.3375	0.3109
TkQ(0)	0.0252	0.0299	0.0335	0.0286	0.0351	0.0373
TkQ(0.05)	0.3655	0.3812	0.3733	0.4007	0.4159	0.4021
TkQ(0.1)	**0.4099**	**0.4249**	**0.4058**	**0.4436**	**0.4608**	**0.4419**
TkQ(0.3)	0.3671	0.3928	0.3876	0.4068	0.4226	0.4099
TkQ(0.5)	0.3252	0.3478	0.3361	0.3465	0.3655	0.3660
TkQ(0.7)	0.2636	0.2840	0.2713	0.3006	0.3195	0.3024
TkQ(0.9)	0.2056	0.2256	0.2123	0.2545	0.271	0.2597
TkQ(1.0)	0.1651	0.1882	0.1768	0.2409	0.2558	0.2454
TkQ*	0.4596	0.4672	0.4474	0.5148	0.5297	0.5114

- When keywords are used as a Boolean filtering, spatial proximity ranking (BkQ) is better than text relevance ranking (TBRQ);
- Keyword information plays an indispensable role in POI search;
- TkQ performs better than TBRQ and BkQ;
- The effectiveness of TBRQ increases and then drops as we increase the query radius;
- The effectiveness of TkQ is significantly affected by the parameter α;
- A query dedicated parameter α can significantly improve TkQ than a uniform parameter.

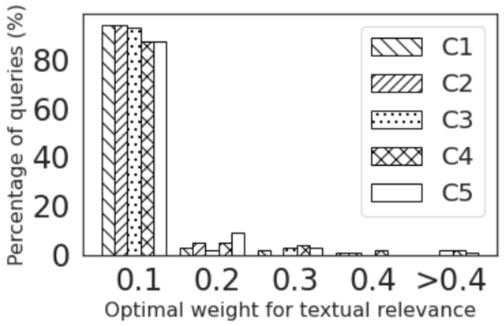
Open Problems

- Open problems
  - Designing a solution for setting a **query-dependent weight value**
  - Whether the effectiveness can be further improved by considering the **deep relevance in computing the textual similarity**.

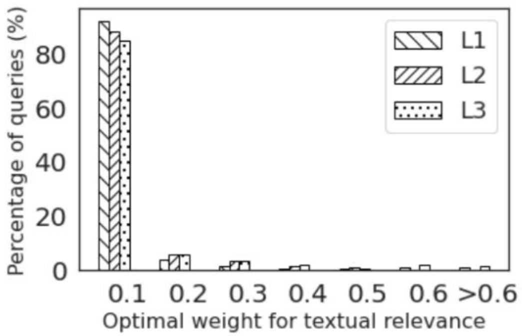
- Data analysis about query-dependent weight setting

$$ST(p,q) = (1-\alpha) \times \underbrace{\boxed{(1 - SDist(p.loc, q.loc))}}_{\text{distance score}} + \alpha \times \underbrace{\boxed{TRel(p.doc, q.doc)}}_{\text{text score}},$$

  - The optimal weight setting of α is correlated with the query keyword and location.



(a) Beijing dataset

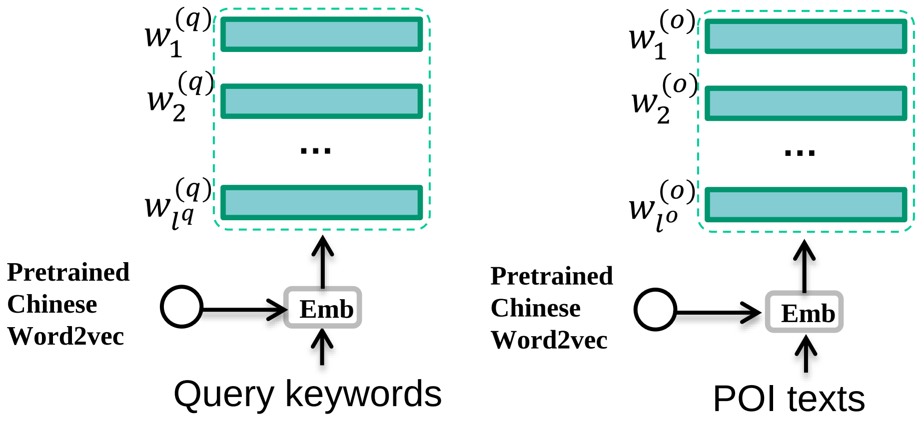


(a) Beijing dataset

- Deep relevance with Weight learning (DrW) model

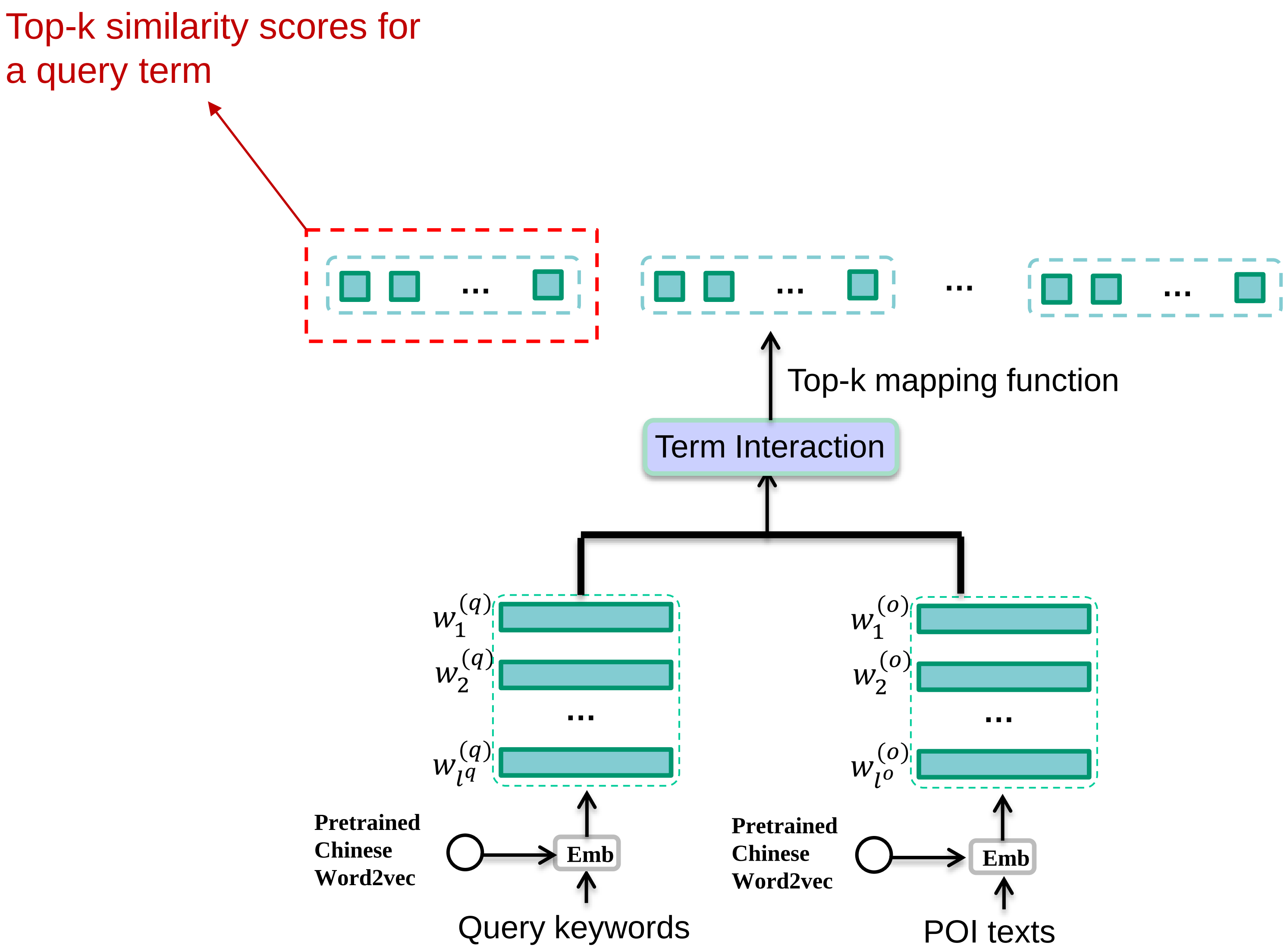
Proposed solution

- Text Relevance Learning
  - Text embedding



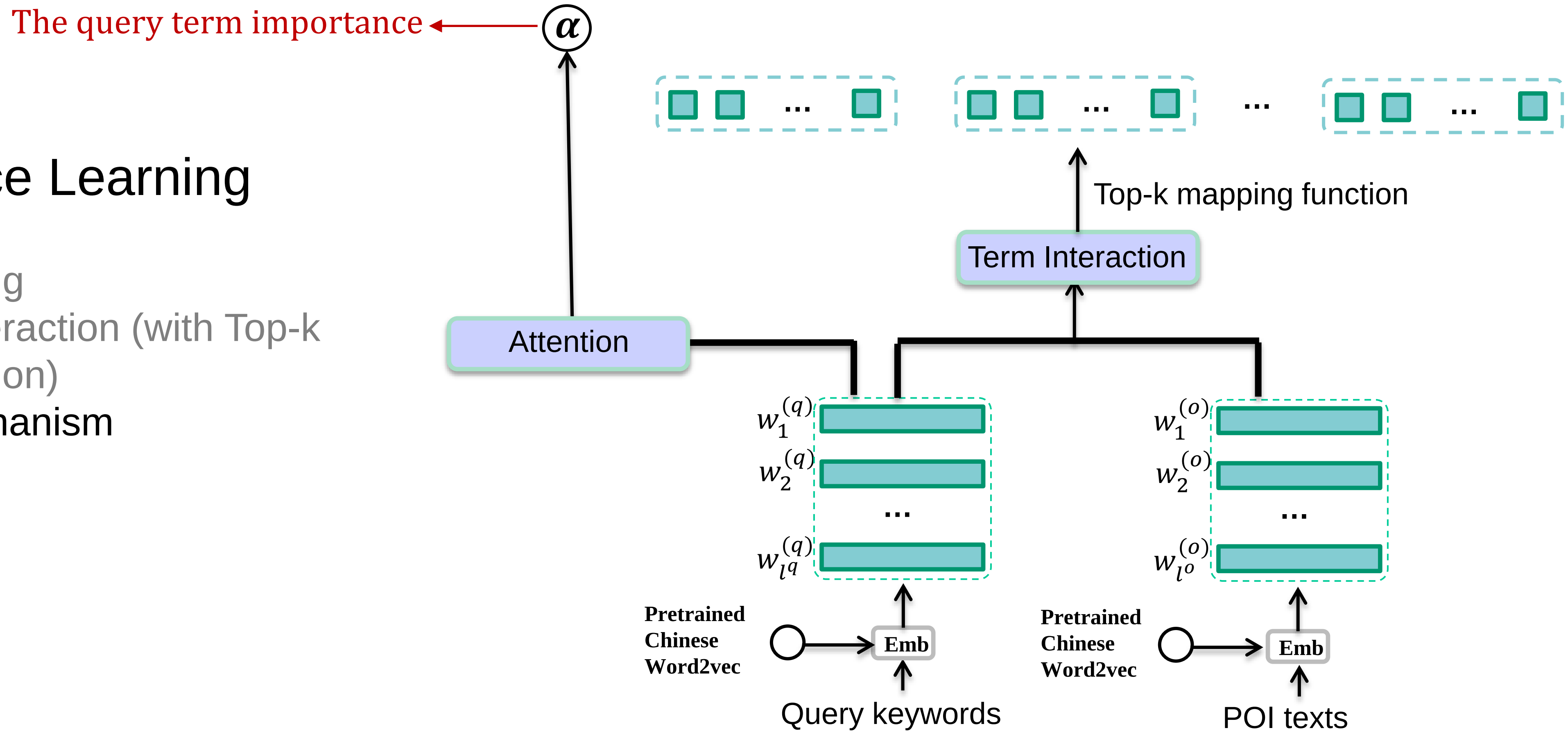
Proposed solution

- Text Relevance Learning
  - Text embedding
  - Term-level interaction (with Top-k mapping function)



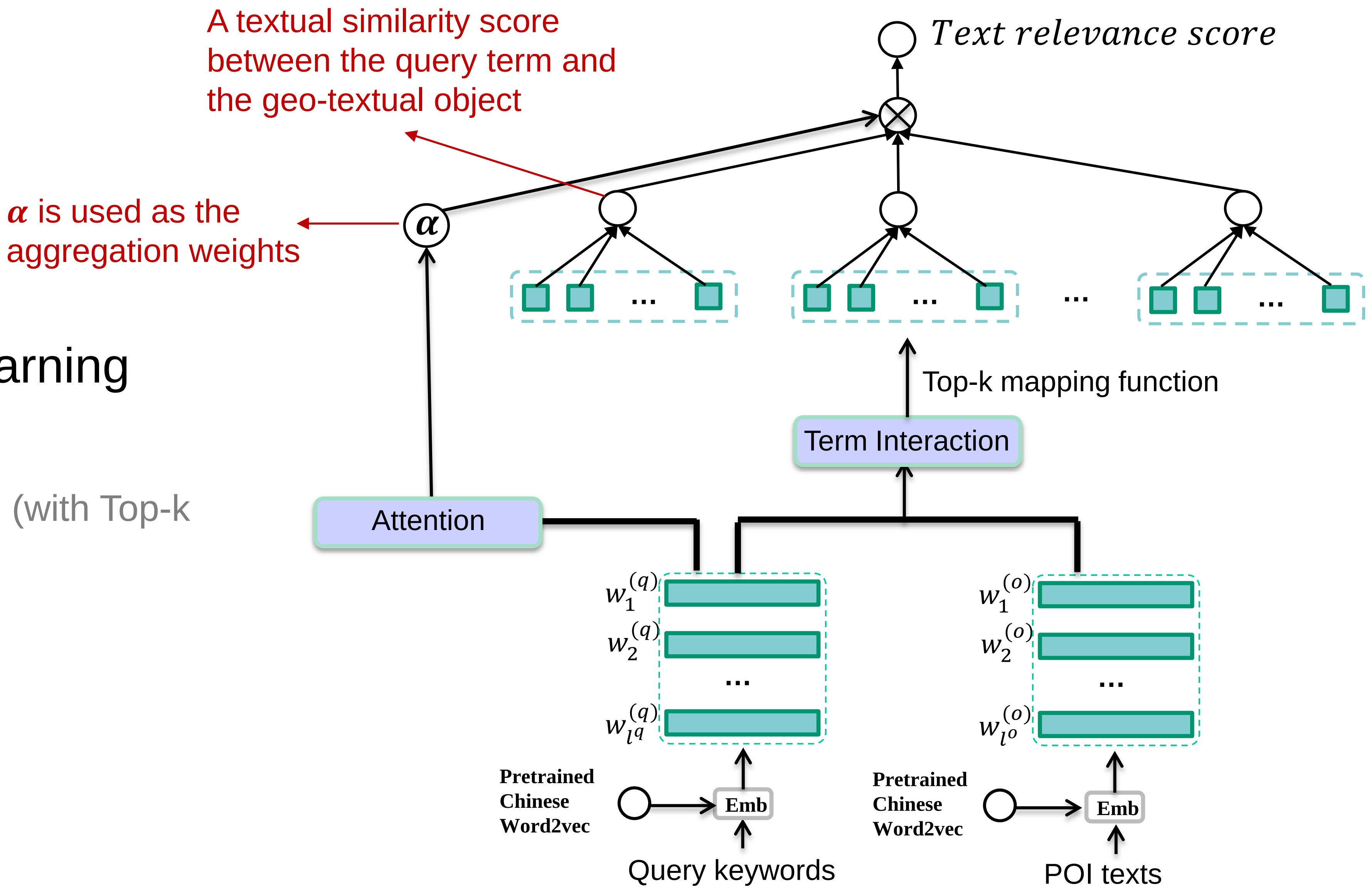
Proposed solution

- Text Relevance Learning
  - Text embedding
  - Term-level interaction (with Top-k mapping function)
  - Attention mechanism



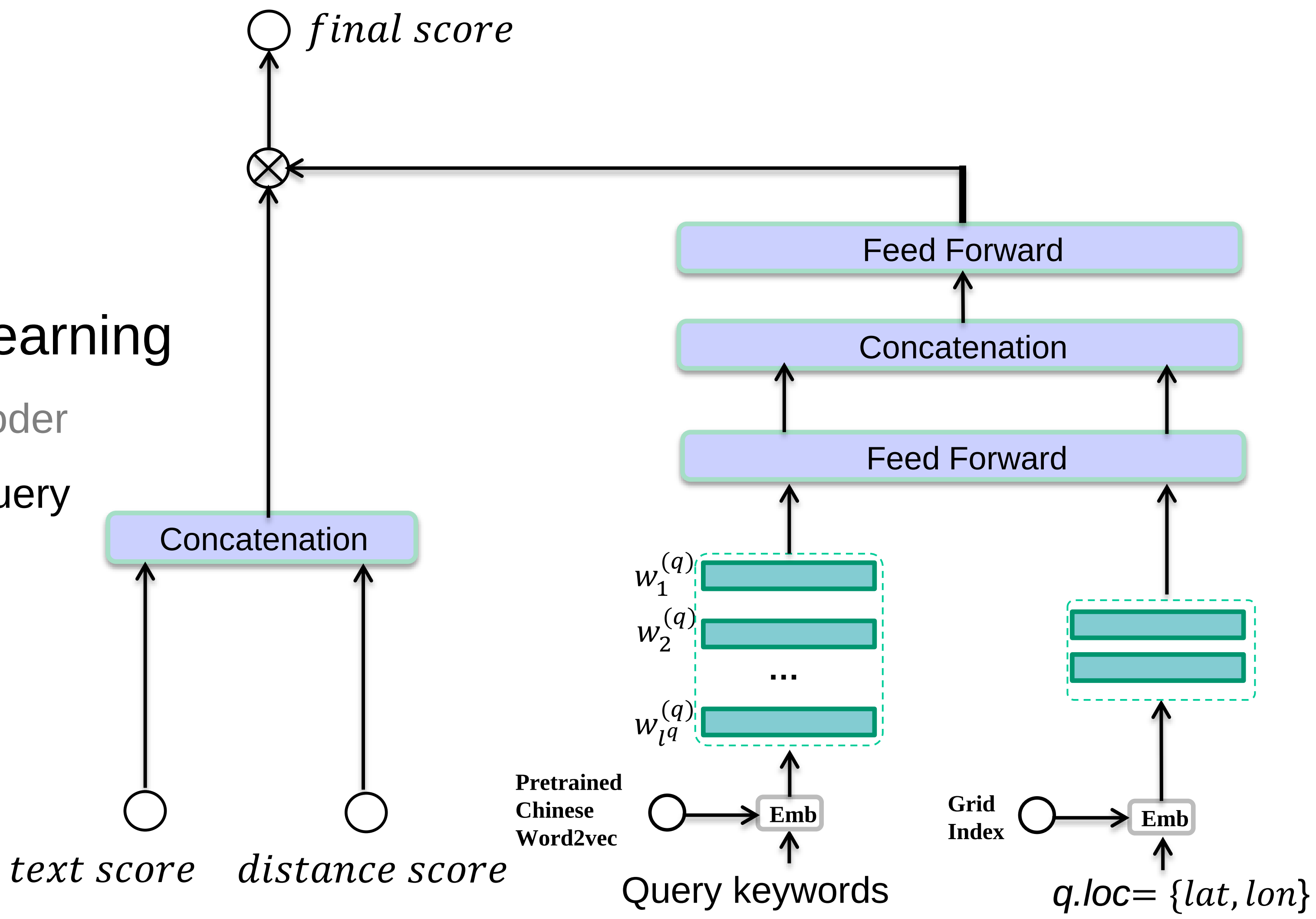
Proposed solution

- Text Relevance Learning
  - Text embedding
  - Term-level interaction (with Top-k mapping function)
  - Attention mechanism
  - Aggregation



Proposed solution

- Query-dependent Weight Learning
  - Query keywords and location encoder
  - Computing final score based on query dependent weights



Experiments

- Baselines
  - We compare DrW with text-based, text and distance-based, as well as other variant algorithms

BM25,
DRMM (CIKM2016),
ARC-I (NeurIPS2014)

TKQ,
DRMM+D,
ARC-I+D
PALM (AAAI2019)

Dr+D(1.0),
Dr+D(0.5),
DrW_{nl}

- Experimental content
  - Effectiveness
  - Study on the model parameters
  - Handle new queries and POIs
  - Efficiency study.

https://github.com/NTMC-Community/MatchZoo-py
https://github.com/Shawn-hub-hit/DrW

Experiments

- Effectiveness

Spatial keyword query results. The last row shows the improvement percentage of DrW over the best baseline

	Beijing			Shanghai		
	NDCG@3	NDCG@5	MRR	NDCG@3	NDCG@5	MRR
TkQ	0.4099*	0.4249*	0.4058*	0.4436*	0.4608*	0.4419*
BM25	0.1651	0.1882	0.1768	0.2409	0.2558	0.2454
DRMM+D	0.2219	0.2219	0.2202	0.2258	0.2365	0.2285
DRMM	0.1123	0.1223	0.1164	0.1189	0.1338	0.1271
ARC-I+D	0.3749	0.3972	0.381	0.3212	0.3397	0.3306
ARC-I	0.0931	0.1272	0.1262	0.1123	0.1333	0.1359
PALM	0.2283	0.2564	0.2518	0.2213	0.2567	0.2527
Dr+D(0.5)	0.4954	0.5083	0.5071	0.4934	0.5106	0.4947
Dr+D(1.0)	0.2131	0.2406	0.2299	0.2488	0.2679	0.2560
DrW_{nl}	0.5064	0.5363	0.5165	0.5115	0.5367	0.5176
DrW	**0.5417**	**0.5623**	**0.5397**	**0.5407**	**0.5621**	**0.5424**
Gain	32.15%	32.34%	33.00%	21.89%	21.98%	22.74%

– DrW consistently outperform other baselines on two dataset

Experiments

- Effectiveness

Spatial keyword query results. The last row shows the improvement percentage of DrW over the best baseline

	Beijing			Shanghai		
	NDCG@3	NDCG@5	MRR	NDCG@3	NDCG@5	MRR
TkQ	0.4099*	0.4249*	0.4058*	0.4436*	0.4608*	0.4419*
BM25	0.1651	0.1882	0.1768	0.2409	0.2558	0.2454
DRMM+D	0.2219	0.2219	0.2202	0.2258	0.2365	0.2285
DRMM	0.1123	0.1223	0.1164	0.1189	0.1338	0.1271
ARC-I+D	0.3749	0.3972	0.381	0.3212	0.3397	0.3306
ARC-I	0.0931	0.1272	0.1262	0.1123	0.1333	0.1359
PALM	0.2283	0.2564	0.2518	0.2213	0.2567	0.2527
Dr+D(0.5)	0.4954	0.5083	0.5071	0.4934	0.5106	0.4947
Dr+D(1.0)	0.2131	0.2406	0.2299	0.2488	0.2679	0.2560
DrW_{nl}	0.5064	0.5363	0.5165	0.5115	0.5367	0.5176
DrW	**0.5417**	**0.5623**	**0.5397**	**0.5407**	**0.5621**	**0.5424**
Gain	32.15%	32.34%	33.00%	21.89%	21.98%	22.74%

– The proposed query dependent weight learning module of DrW works

Outline

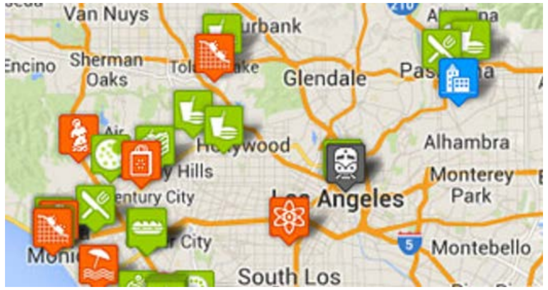
- Target on a specific problem on **point spatial entity**
  - Geospatial IR or Spatial Keyword Search (VLDB’09--SIGMOD’23)
  - POI recommendations (SIGIR’13 --)
  - Spatial relationship extraction (SIGMOD’23)
- Self-supervised learning for geospatial entity representation
  - Road Network Representation 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.

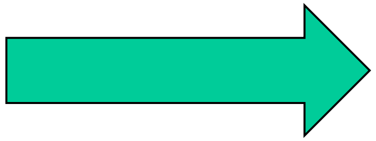
POI recommendation

- 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.

- POI recommendation helps users exploring new places and enrich their experiences.

A large number of POIs



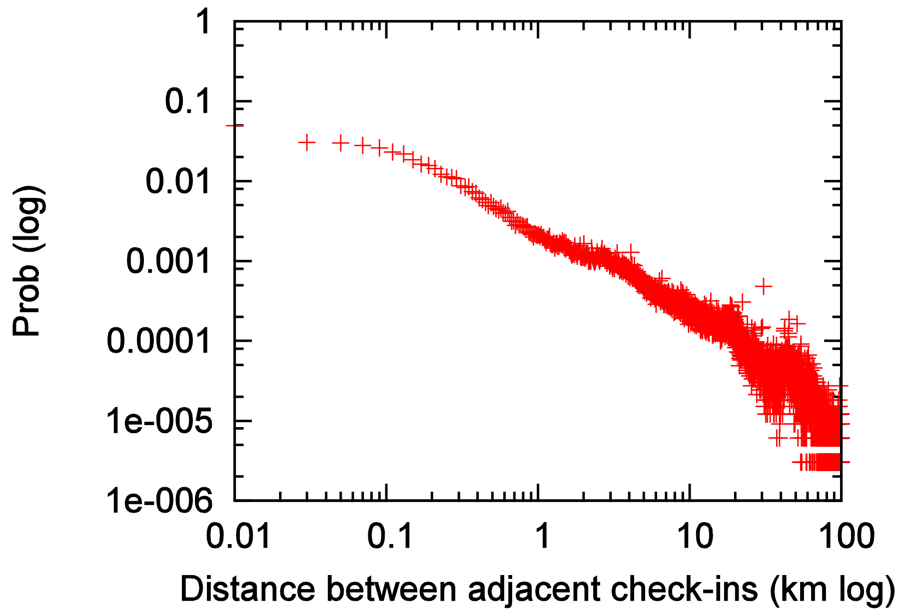


POI recommendation

Users with different interests

Challenges of POI recommendation

- **Very Sparse Data.**
  - Density for Netflix (Movies): 1.2%
  - Density for Yelp & Foursquare (POIs): 0.1%
- **Rich context.**
  - Geographical Influence
    - Human tend to visit nearby POIs
  - Temporal Influence
    - User mobility varies with time: office @ morning, pubs @ night

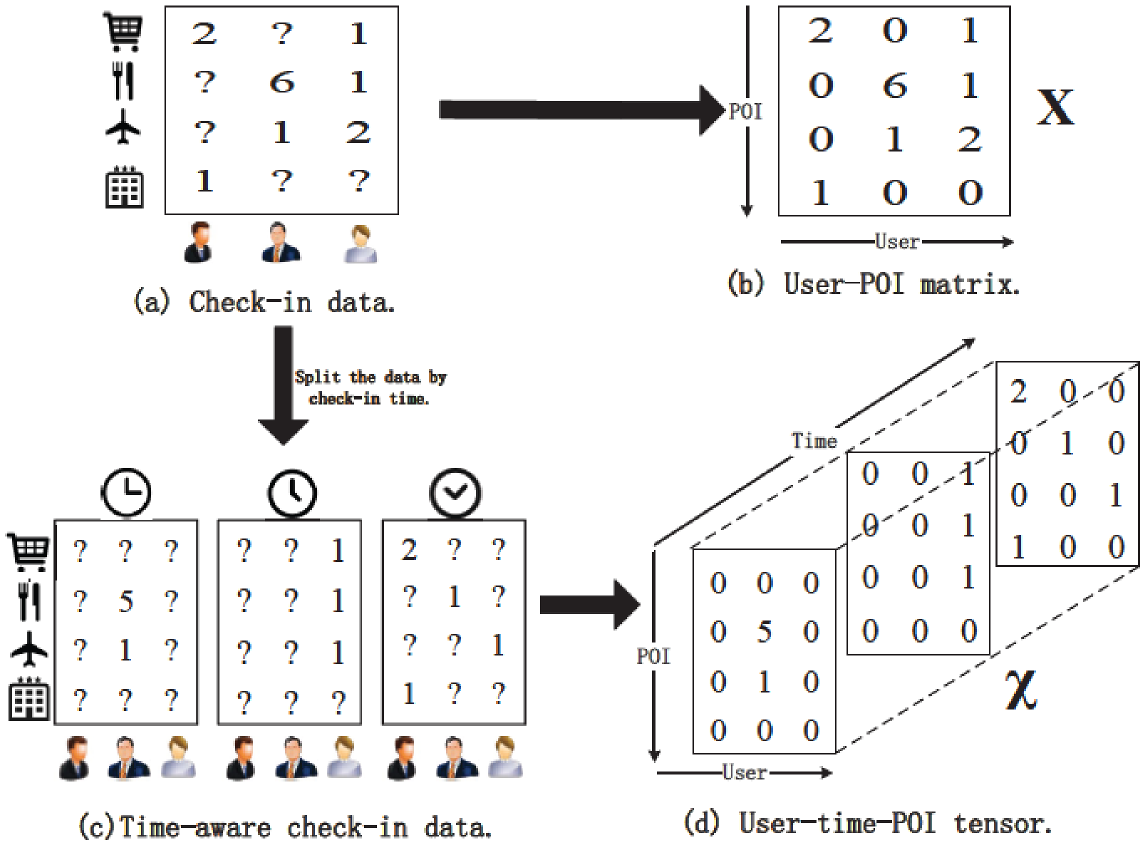


Xutao Li, Gao Cong, Xiaoli Li, Tuan-Anh Nguyen Pham, Shonali Krishnaswamy: Rank-GeoFM: A Ranking based Geographical Factorization Method for Point of Interest Recommendation. SIGIR 2015

New Types of POI recommend

Context-aware POI recommendation

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



Requirement-aware POI recommenda

- E.g., Mary wants to find a restaurant to have p
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: $$w_{u,v} = \frac{\sum_l c_{u,l} c_{v,l}}{\sqrt{\sum_l c_{u,l}^2}\sqrt{\sum_l c_{v,l}^2}}$$

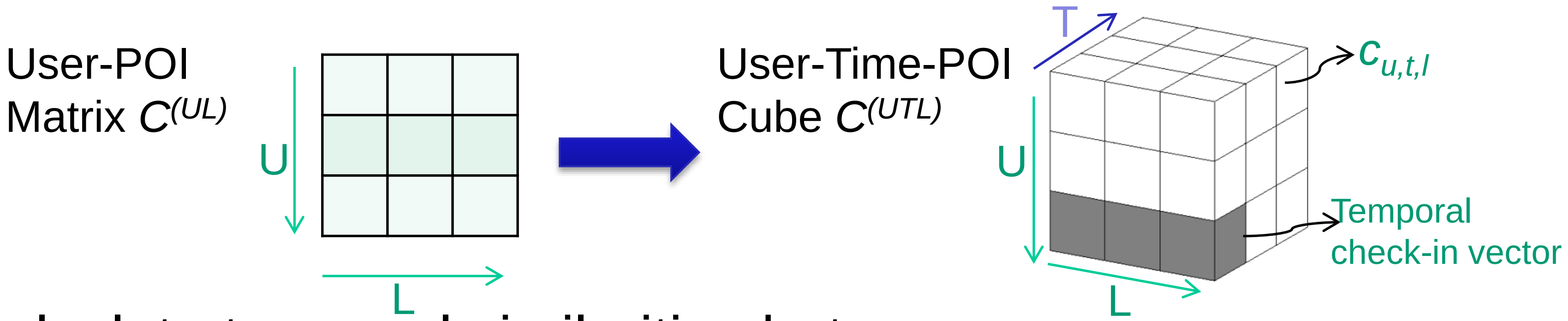
  - Produce prediction for each candidate POI *l*

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

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

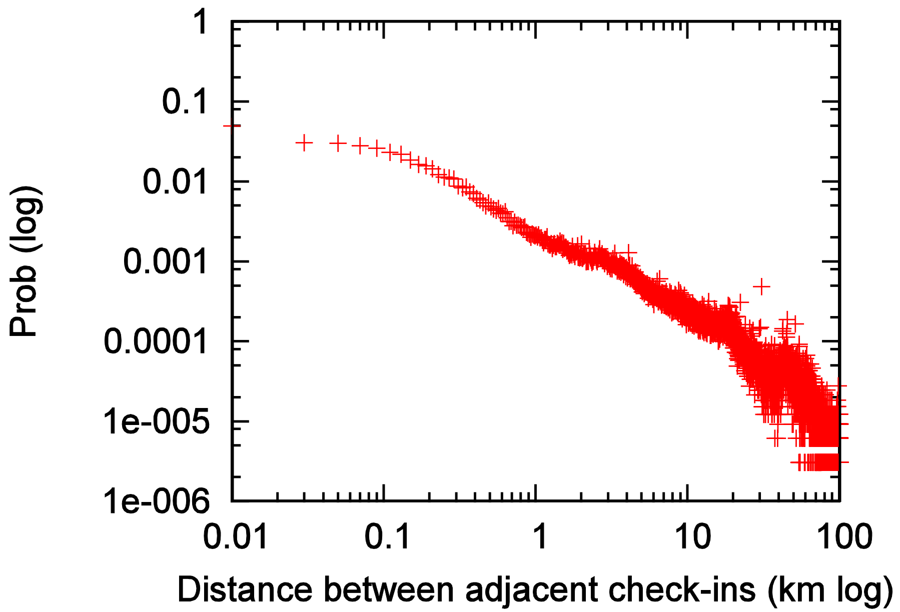
- Produce temporal predictions

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

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

Incorporating Spatial Influence

- Human tend to visit nearby POIs to their current locations.
- Calculate the distance between two POIs of every two successive check-ins, and plot the number as a function of distance.



- Power law distribution
- Users are more willing to visit nearby POIs
- The willingness of a user to visit δ *km* far away POI:

$$wi(\delta) = a \cdot \delta^k$$

Spatial Influence based Recommendation (S)

- The probability *u* at l_i will checkin l_j is proportional to the willingness:

$$P(l_j \mid l_i) = \frac{wi(\delta(l_i, l_j))}{\sum_{l_k \in L, l_k \neq l_i} wi(\delta(l_i, l_k))}$$

- Given user *u* and historical POIs L_u, we calculate $P(l|L_u)$ as the ranking score for each candidate POI *l*:

$$\hat{c}^{(s)}_{u,l} = P(l \mid L_u) \propto P(l)P(L_u \mid l) = P(l)\prod_{l' \in L_u} P(l' \mid l)$$

- *P(l)*: prior, in proportion to the number of check-ins on it.
- Users are likely to visit
  - Nearby POIs
  - Popular POIs

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

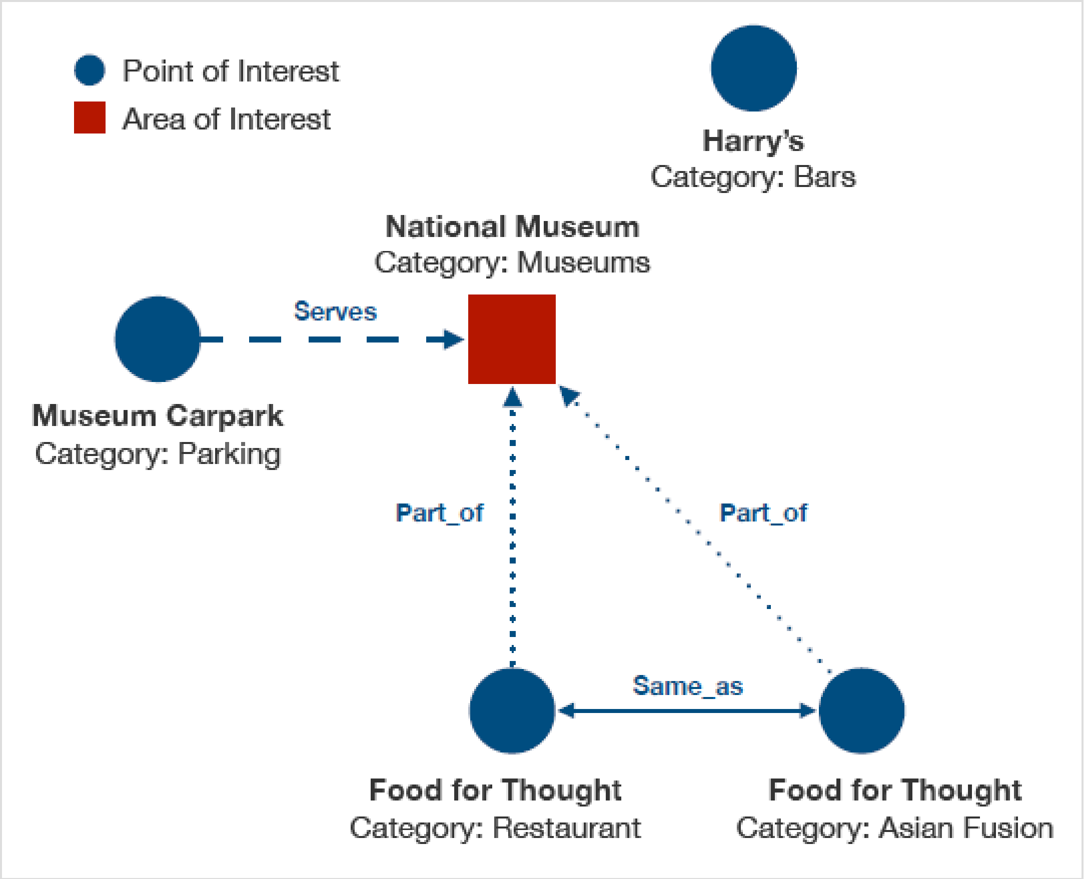
Although convenient, this representation hinders the exploration of **geospatial relationships** between the entities

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

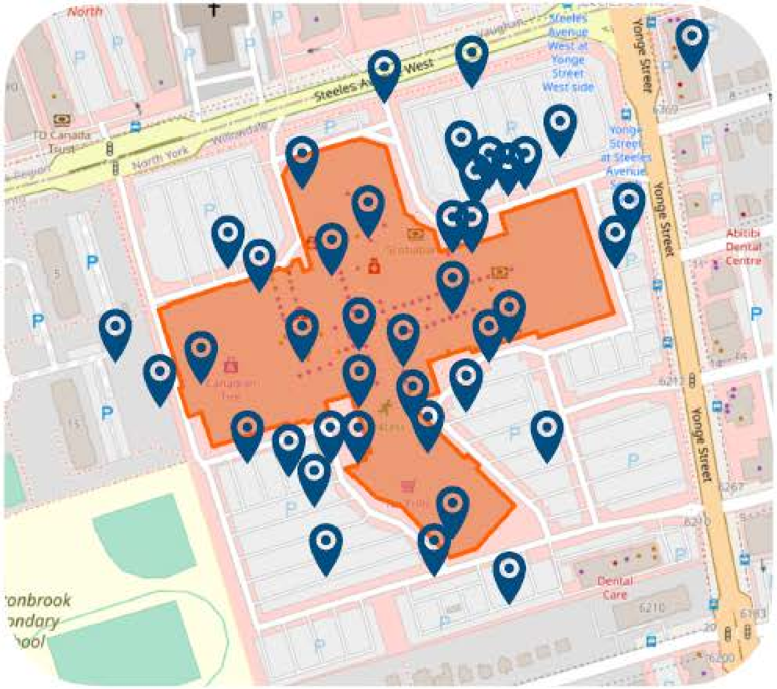
DBPedia

Challenges

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

Sources: YELP, OSM

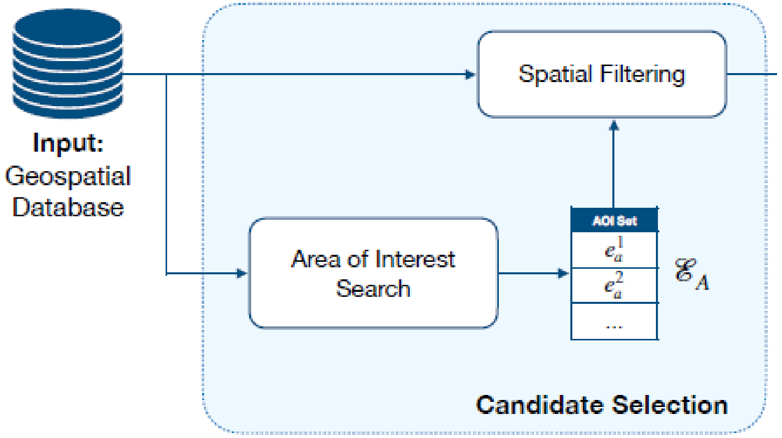
In reality, all the POIs are located inside the Mall

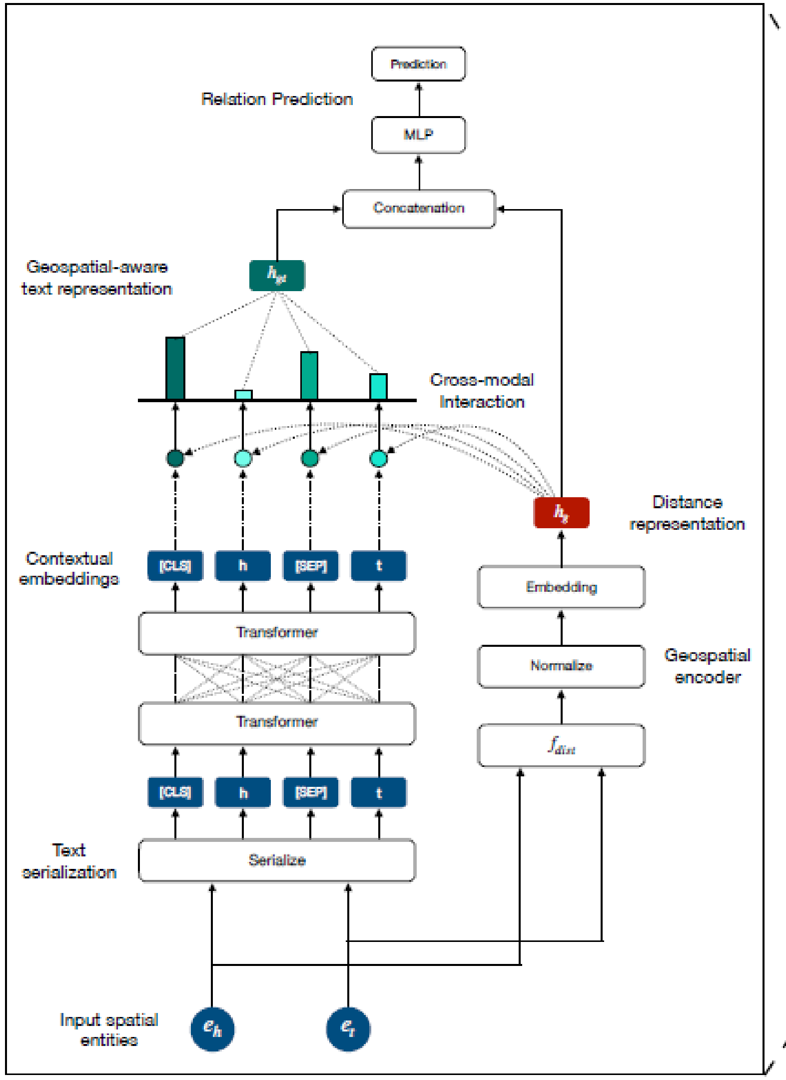


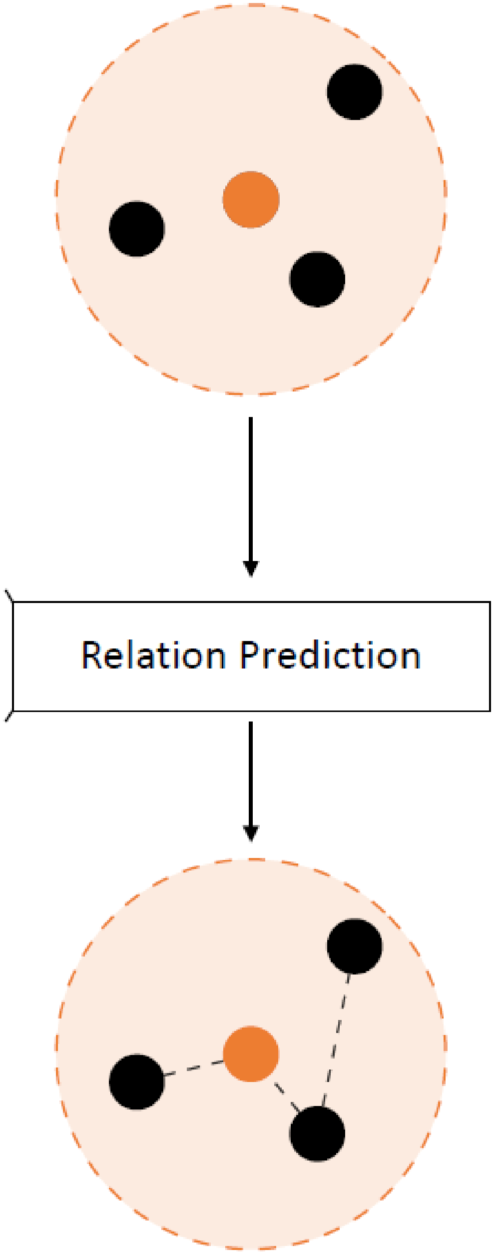
Centerpoint Mall, Toronto

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

Proposed solution





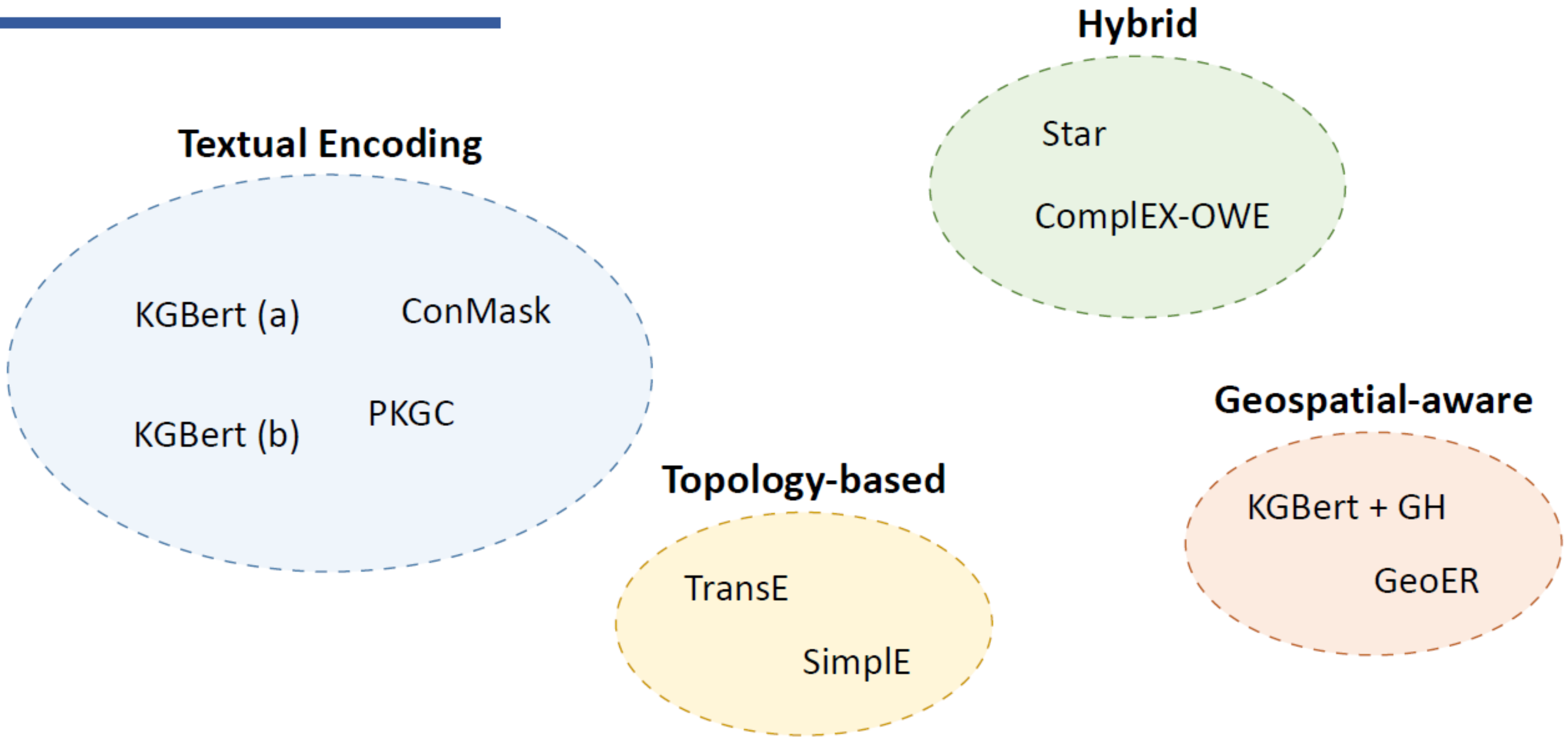


- **Candidate Selection Step**: Aim relationships
- **Relation Prediction:** Aim at ide
- **The KG refinement**: Aim to ext correctness

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

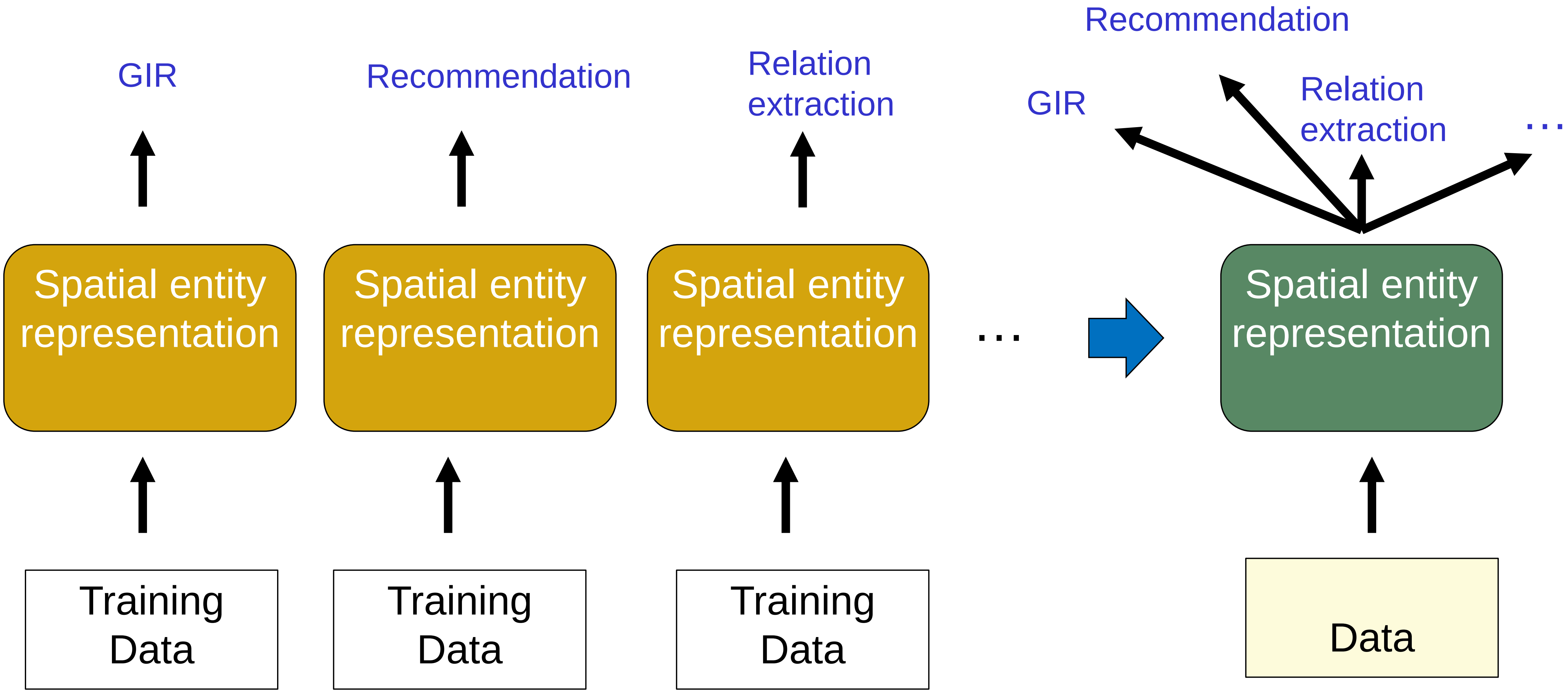
Model	Singapore			Toronto			Seattle			Melbourne		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
TransE [6]	4.71	16.3	7.29 (± 0.41)	5.28	12.82	7.47 (± 0.28)	4.51	11.07	6.4 (± 0.79)	4.25	9.56	5.88 (± 0.55)
SimplE [30]	6.64	19.27	9.87 (± 0.96)	4.98	18.3	7.82 (± 0.71)	7.75	9.84	8.66 (± 1.05)	6.9	13.33	9.09 (± 0.93)
ComplEx-OWE [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			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%

Including geospatial information further improves the results

A Summary



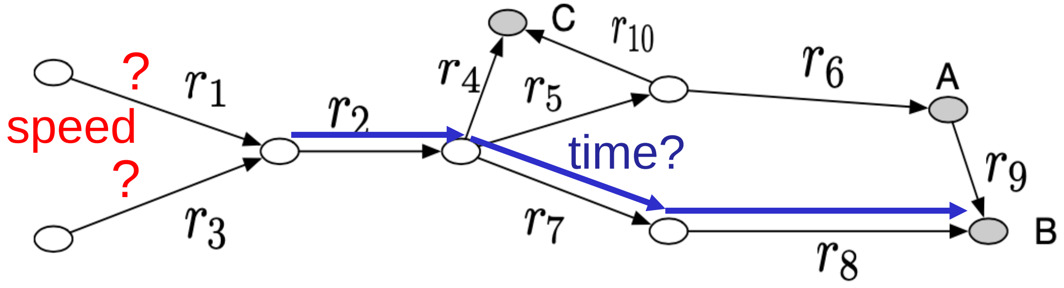
GIR
Recommendation
Relation extraction
Recommendation
GIR
Relation extraction
…
Spatial entity representation
Spatial entity representation
Spatial entity representation
…
Spatial entity representation
Training Data
Training Data
Training Data
Data

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:** numerous applications are built upon road networks, such as travel time estimation, traffic inference, etc.

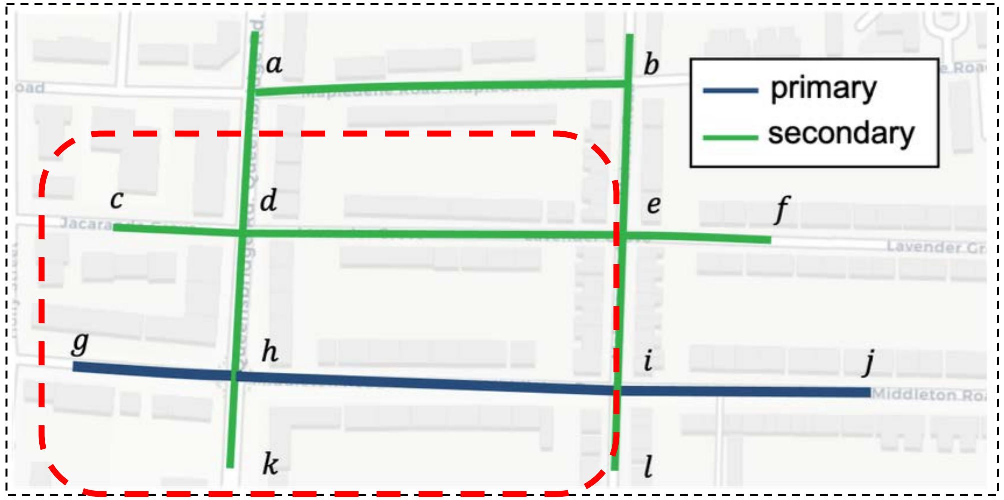


- ▷ **Target:** derive effective representations that are robust and generic for downstream applications.
  - Road segment-based & trajectory-based applications

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**

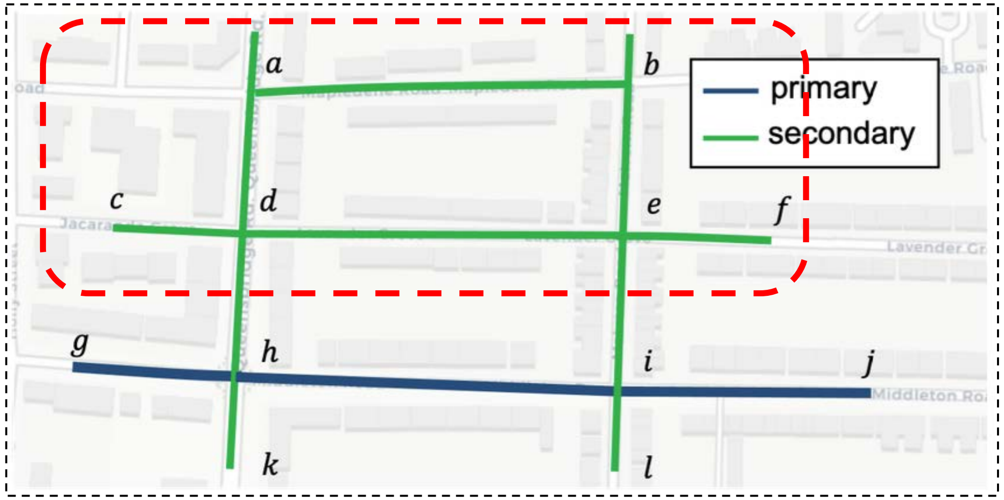


- ▷ According to homophily, inter-connected nodes are more similar than distant ones.

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

Challenges

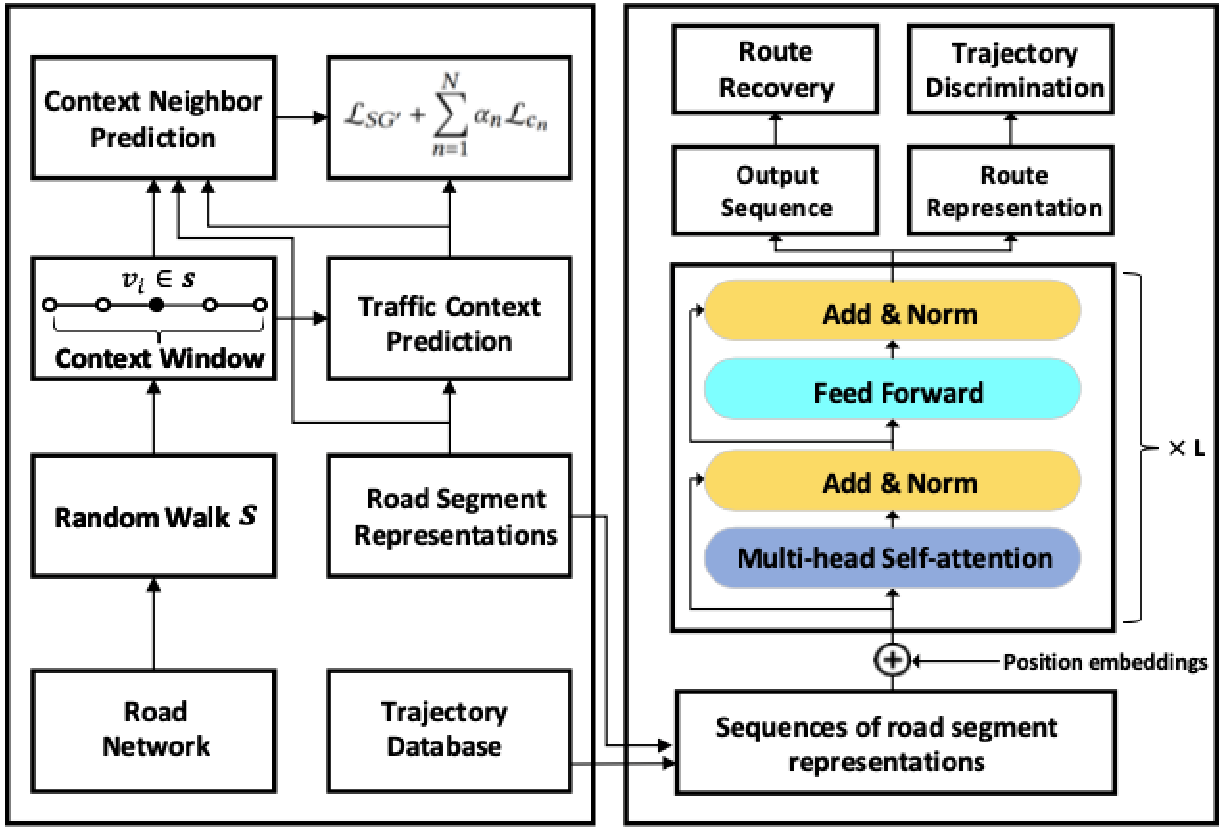
- ▷ **Feature uniformity**



- ▷ 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.

Method - Toast

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

- Basic skip-gram to encode graph structure

$$\mathcal{L}_{SG} = -\sum_{v_i \in \mathbf{s}} \sum_{v_j \in \mathcal{N}(v_i)} \log p\left(v_j | v_i\right)$$

- Auxiliary traffic context prediction:
  - predict the traffic context of target road segments

$$\mathcal{L}_{c_n} = -\sum_{v_i \in \mathbf{s}} \sum_{j=1}^{|c_n|} c_{nj}^i \log p\left(c_{nj}^i | v_i\right) + \left(1 - c_{nj}^i\right) \cdot \log\left(1 - p\left(c_{nj}^i | v_i\right)\right)$$

- Modify the basic skip-gram to be conditioned on the traffic context:

$$\mathcal{L}_{SG'} = -\sum_{v_i \in \mathbf{s}} \sum_{v_j \in \mathcal{N}(v_i)} \log p\left(v_j | v_i, \tilde{\boldsymbol{\pi}}(v_i)\right)$$

Method - Toast

- **Trajectory-enhanced Transformer module:**
  - Inspired by BERT, we propose to apply **two pre-training tasks** derived from trajectories to learn traveling semantics encoded in them.

  - **Route recovery**
    - Mask 20% consecutive road segments of a given trajectory.
      - ✓ Cannot be trivially recovered with the knowledge of road network structure
      - ✓ Can learn complex transition patterns

Method - Toast

- **Trajectory-enhanced Transformer module:**
  - Inspired by BERT, we propose to apply **two pre-training tasks** derived from trajectories to learn traveling semantics encoded in them.
  - **Trajectory discrimination**
    - Given a route on road network, the model judges whether it is a real trajectory or not.
      - ✓ Real ones are from trajectory databases, while fake ones are sampled with random walks.
      - ✓ Another way of learning transition patterns.

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
- Trajectory similarity search
- Travel time estimation

Road segment-based

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.

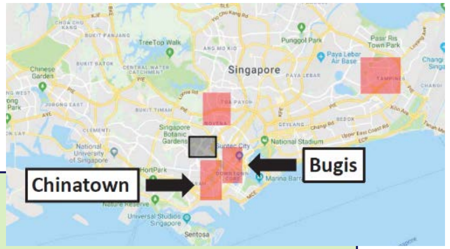
Region Spatial Entity Data

Region Detection (Burst region) TKDE'22

Applications: event detection

Region Search

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

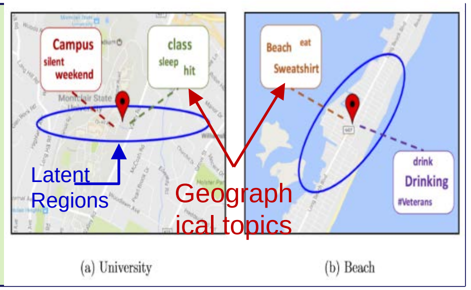


Applications: Similar region search for urban planning, Keyword based search

Region Recommendation & User Mobility Prediction

[WSDM'17, WWW'17]

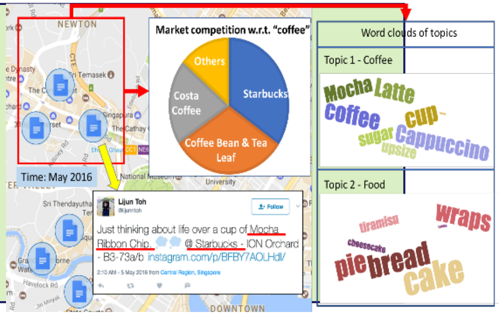
Geo-topic modeling TKDE'22



(a) University (b) Beach

Region topic exploration

SIGMOD'16, VLDBJ



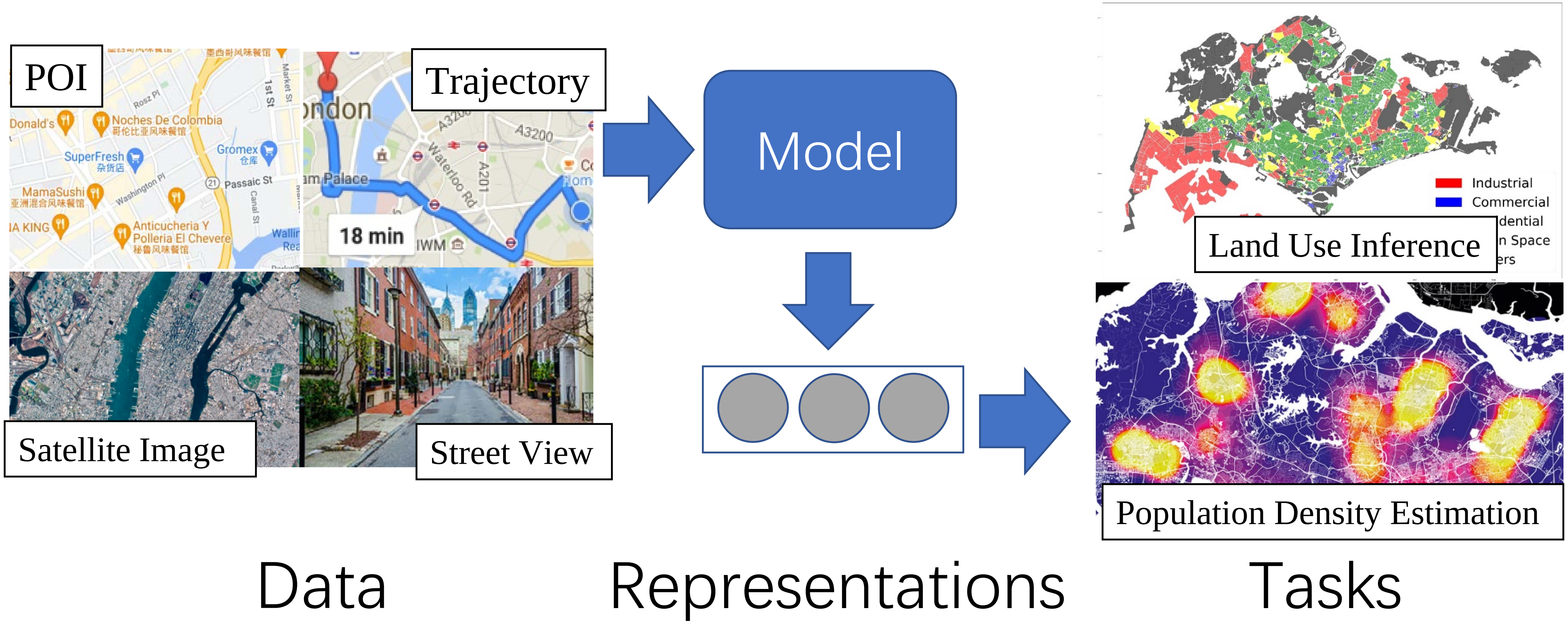
Business analytics

Deep Region Representation KDD'23

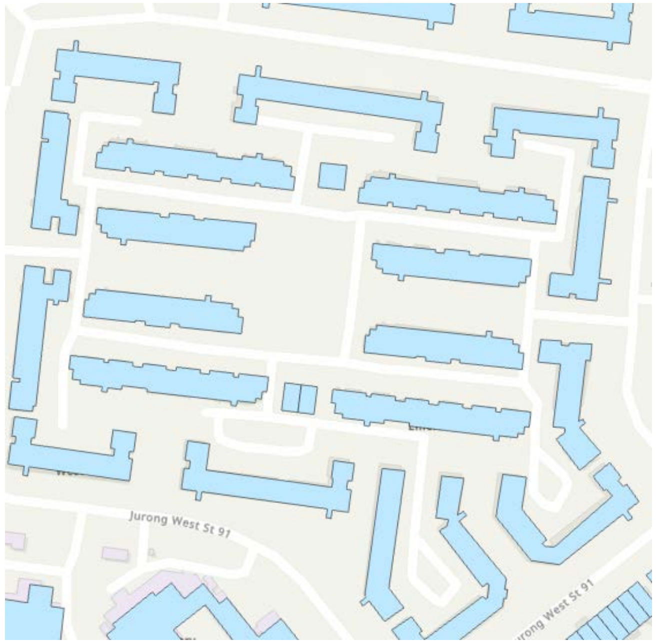
Applications: Region analysis, such as functions, property price, crime rate, populations, etc

Problem of Urban Region Representation Learning

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



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

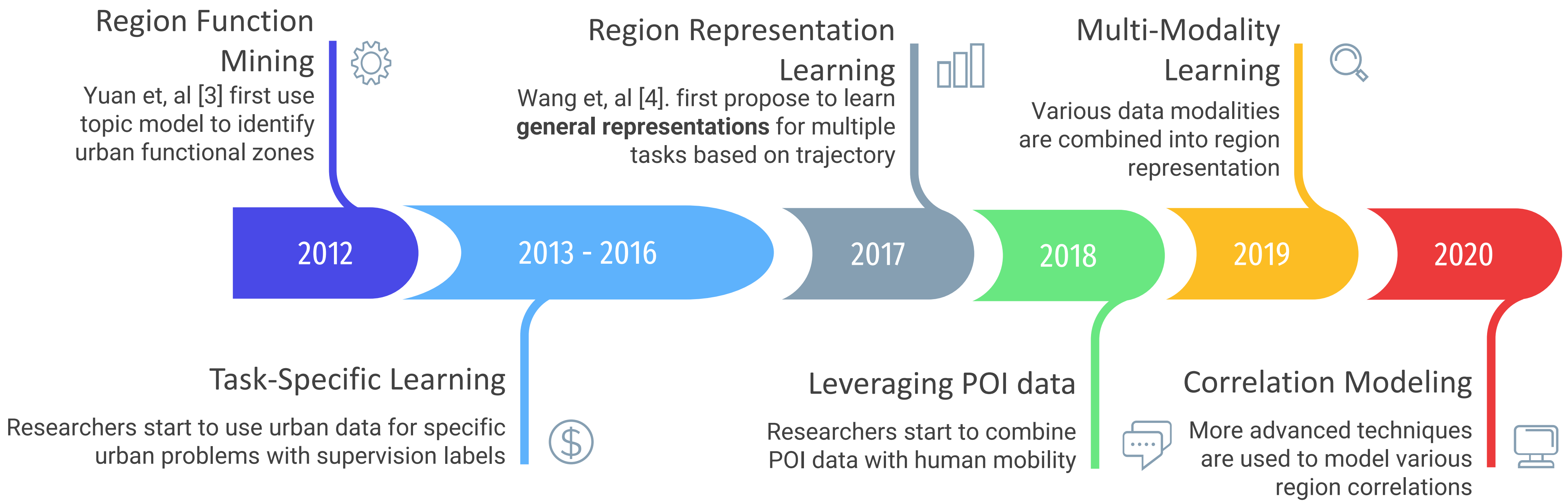
Example Building Groups with Specific Urban Functions

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

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

Related Work

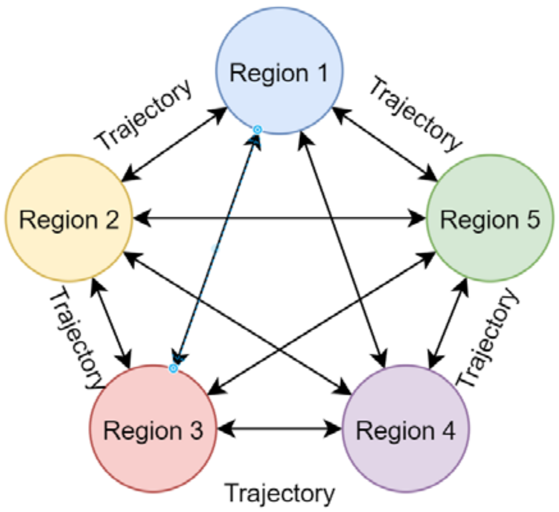
- Urban Region Representation Learning



[3] Yuan et, al. Discovering Regions of Different Functions in a City Using Human Mobility and POIs. In KDD 2012.

[4] Wang et, al. Region Representation Learning via Mobility Flow. In CIKM 2017

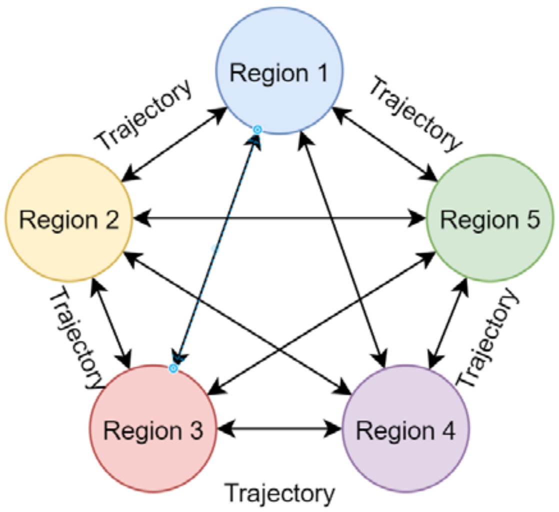
Related Work



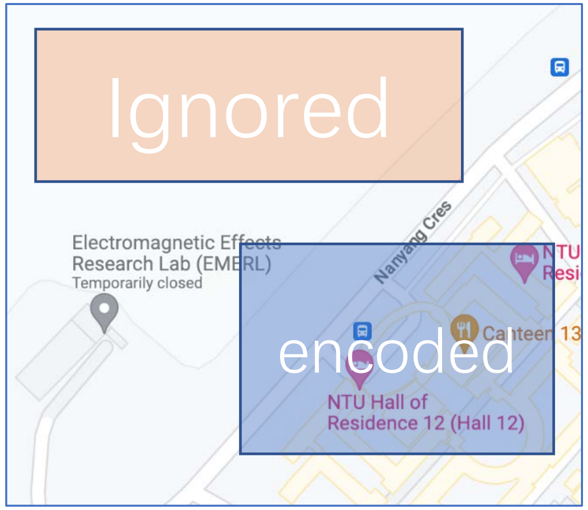
- Heavily rely on trajectory data

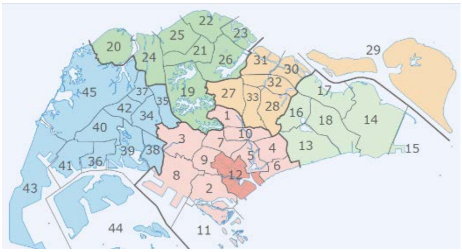
Model	Data Source
Place2vec [29]	POI
Doc2Vec [30]	POI
HDGE [8]	trajectory, POI, demographic/geographic features
ZE-Mob [9]	trajectory, check-in
Fu et al. [11]	trajectory, POI
Zhang et al. [12]	trajectory, POI
ReMVC [31]	trajectory, POI
RegionEncoder [10]	trajectory, POI, satellite image
MVURE. [13]	trajectory,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

Related Work

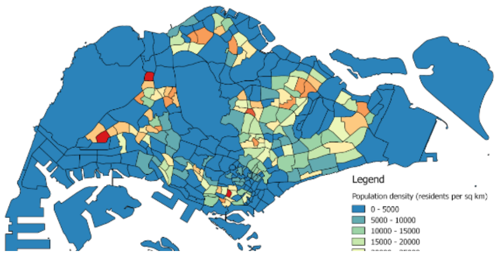


- Heavily rely on trajectory data
- Ignore data-sparse areas
- Can’t adapt to multiple region partition schemes





Regions in Land Use Classification



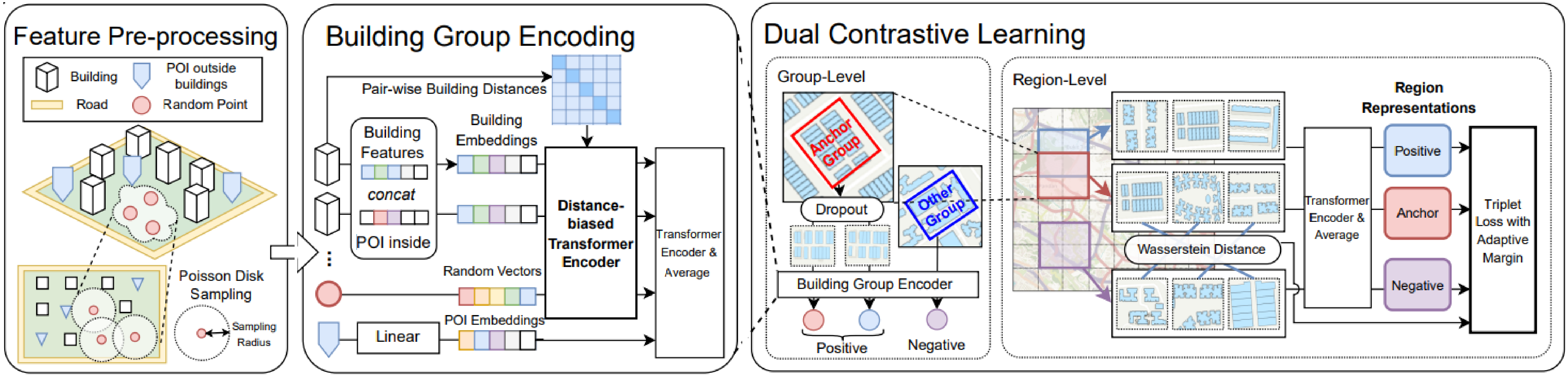
Regions in Population Density Estimation

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**

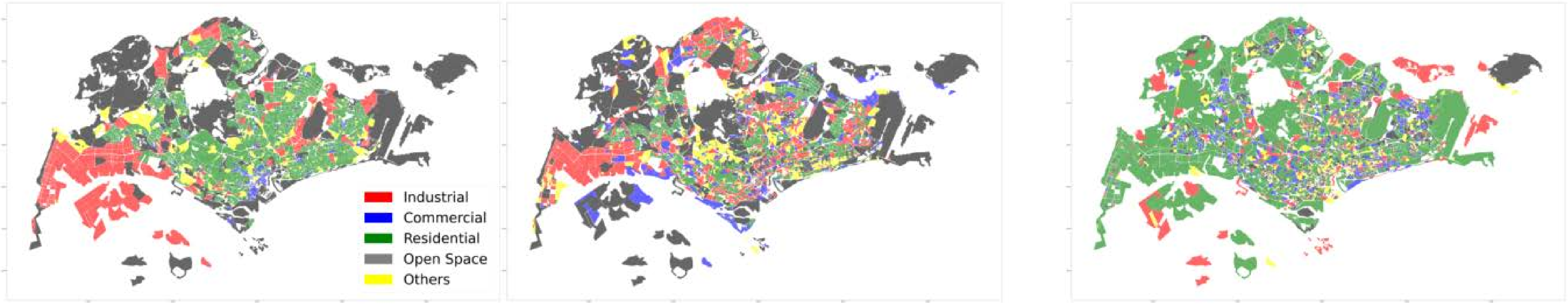
Population Density Inference

- 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 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)

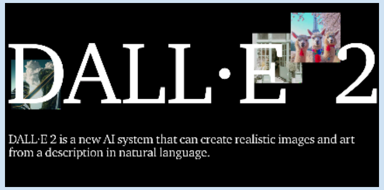
GPT3

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)



Zero-shot Transfer

Various CV Tasks

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

Large Language Model

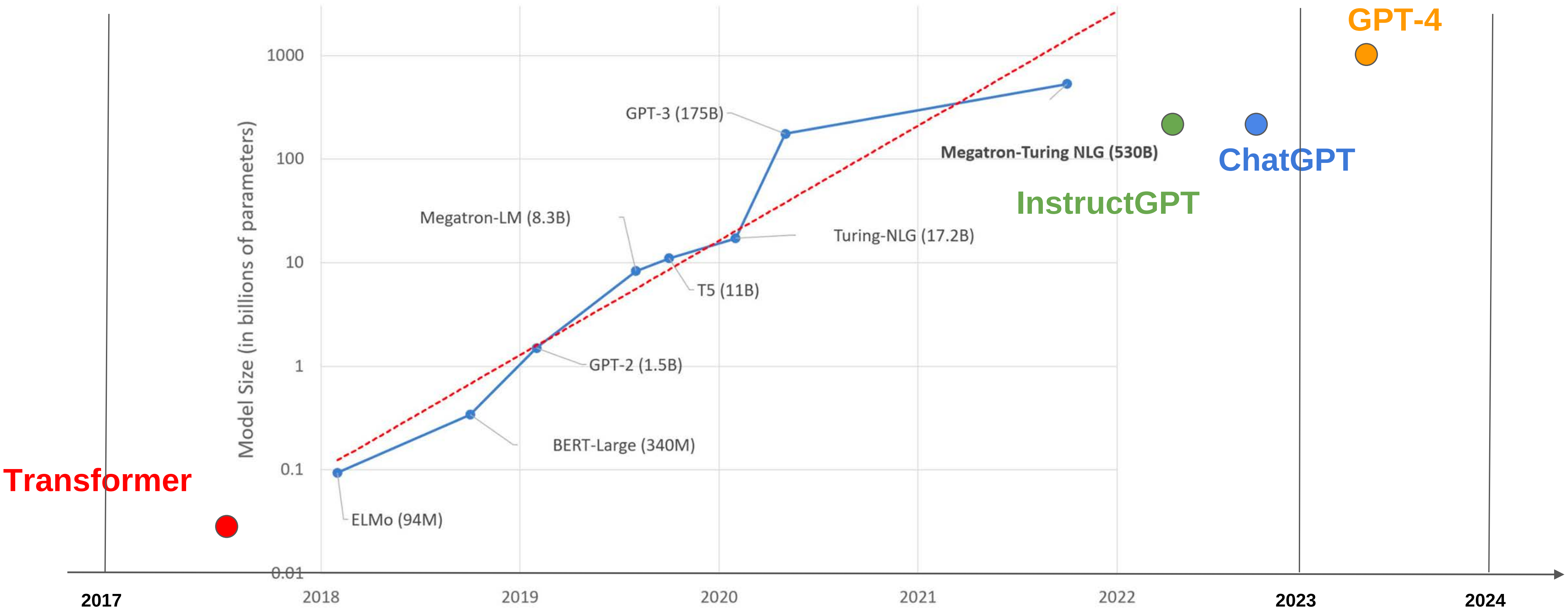
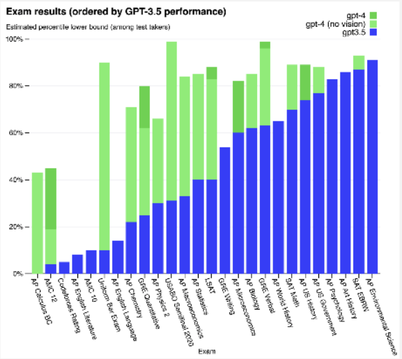


Image from Huggingface (https://huggingface.co/blog/large-language-models)

Foundation Models in Different Domains

Natural Language Processing

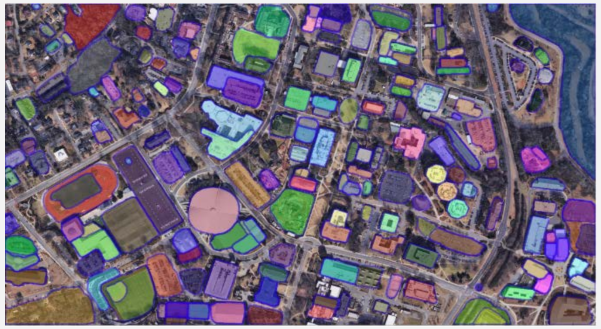
Stanford Alpaca



ChatGPT/GPT-4 (OpenAI. 2023)

Computer Vision

Imagen (Saharia et al. 2022)

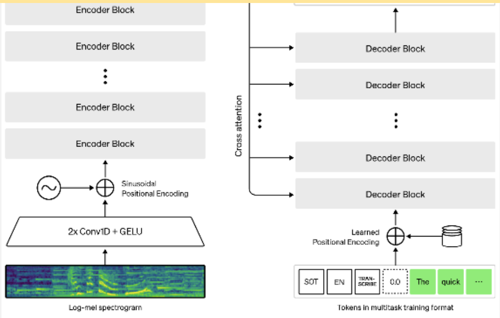


Segment Anying (Kirillov et al, 2023)

Reinforcement Learning

Gato (Reed et al. 2022)

Signal Processing



Whisper (Radford et al. 2022)

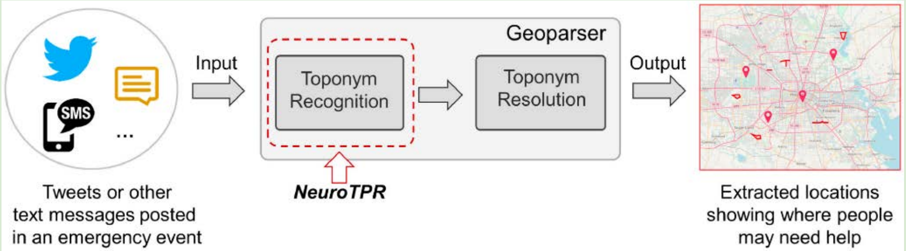
The Big Question

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.

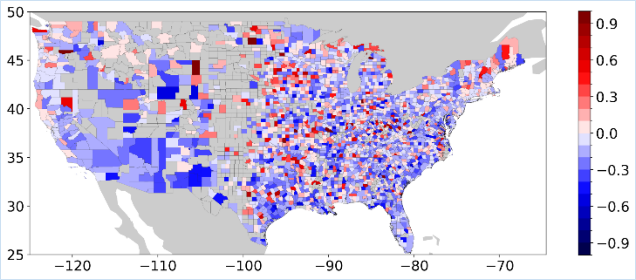


Urban Geography

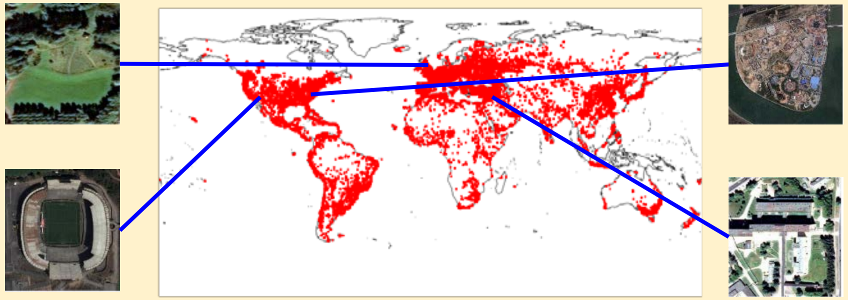
Urban Function

Urban Perception

Health Geography – Dementia Forecast



Remote Sensing – RS Image Clas.



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 5: POI-Based Urban Function Classification

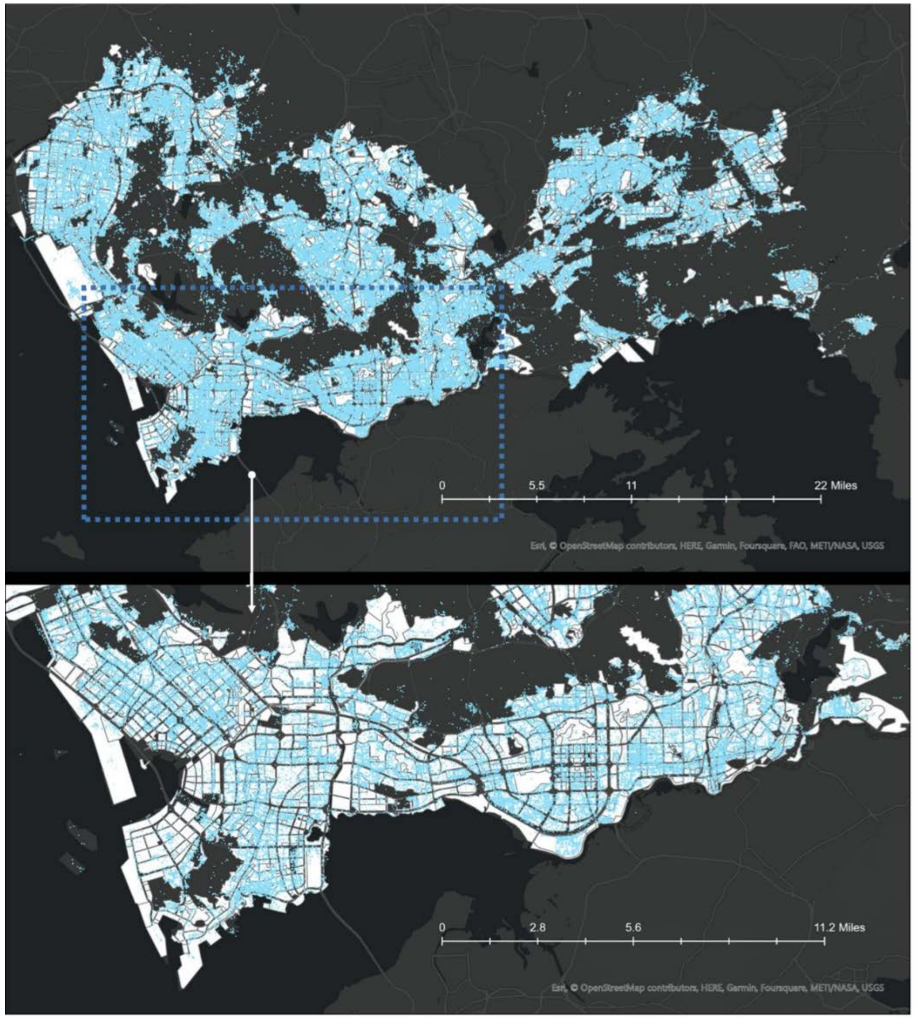


Fig. 2. The spatial distributions of POI data in the *UrbanPOI5K* dataset.

```
[Instruction] There are six land use types: (1) residential, (2) commercial, (3) industrial, (4) education, health care, civic, governmental
    ↪ and cultural, (5) transportation facilities, and (6) outdoors and natural.
Paragraph: In this urban region, there are 128 points of interest, including 2 Chinese restaurant, 1 food restaurant, 2 hotel, 2 apartment
    ↪ hotel, 1 daily life service, 1 mobile communication shop, 24 company, 1 logistics company, 1 real estate agency, 1 lottery retailer,
    ↪ 3 beauty shop, 1 manicure, 2 barber shop, 4 Internet cafe, 3 bath massage, 2 stadium, 4 training institutions, 1 pharmacy, 4
    ↪ automative sale, 6 car service, 2 car repair, 1 Car rental, 1 Automobile parts, 3 shopping, 5 shop, 5 parking lot, 5 Parking lot
    ↪ entrance, 2 transportation facility, 1 port harbor, 1 road intersection, 1 atm machine, 2 office building, 2 residential area, 7
    ↪ building, 1 real estate, 1 park, 1 factory, 7 administrative agency, 1 entrance and exit, 3 gate door, 6 convenience store, 4 home
    ↪ building materials.
Q: What is the primary land use category of this urban region?
A: outdoors and natural

Paragraph: In this urban region, there are 17 points of interest, including 1 food restaurant, 3 public toilet, 3 funeral service, 2 road station for walking and cycling,
    ↪ 1 beach, 2 parking lot, 2 road intersection, 1 corporate company enterprise, 2 administrative agency.
Q: What is the primary land use category of this urban region?
A: outdoors and natural
```

Listing 5. POI-based urban function classification with LLMs, e.g., ChatGPT (Raw.). Yellow block: the POI statistic of a new urban neighborhood to be classified. Orange box: ChatGPT (Raw.) outputs.

Table 5. Evaluation results of various GPT models and supervised baseline on the *UrbanPOI5K* dataset for the POI-based urban function classification task. We divide the models into three groups: (A) supervised learning-based neural network models; (B) Zero-shot learning with LLMs. (C) One-shot learning with LLMs. We use accuracy, weighted precision, and weighted recall as evaluation metrics. We do not include weighted F1 scores since it is the same as the accuracy score. The best model of each group is highlighted.

	Model	Accuracy	Precision	Recall
(A) Supervised NN	Place2Vec [145, 152]	0.540	0.512	0.516
	HGI [52]	**0.584**	**0.568**	**0.563**
(B) Zero-shot LLMs	GPT2 [115]	**0.318**	0.105	**0.158**
	GPT2-Medium [115]	0.025	0.102	0.040
	GPT2-Large [115]	0.005	0.001	0.002
	GPT2-XL [115]	0.001	0.108	0.002
	GPT-3 [15]	0.144	**0.448**	0.141
	ChatGPT (Raw.) [104]	0.075	0.376	0.106
	ChatGPT (Con.) [104]	0.051	0.232	0.046
(C) One-shot LLMs	GPT2 [115]	0.149	0.079	0.085
	GPT2-Medium [115]	0.317	0.104	0.156
	GPT2-Large [115]	0.057	0.083	0.021
	GPT2-XL [115]	**0.324**	0.105	0.159
	GPT-3 [15]	0.176	0.486	0.190
	ChatGPT (Raw.) [104]	0.195	**0.524**	**0.245**
	ChatGPT (Con.) [104]	0.093	0.451	0.085

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

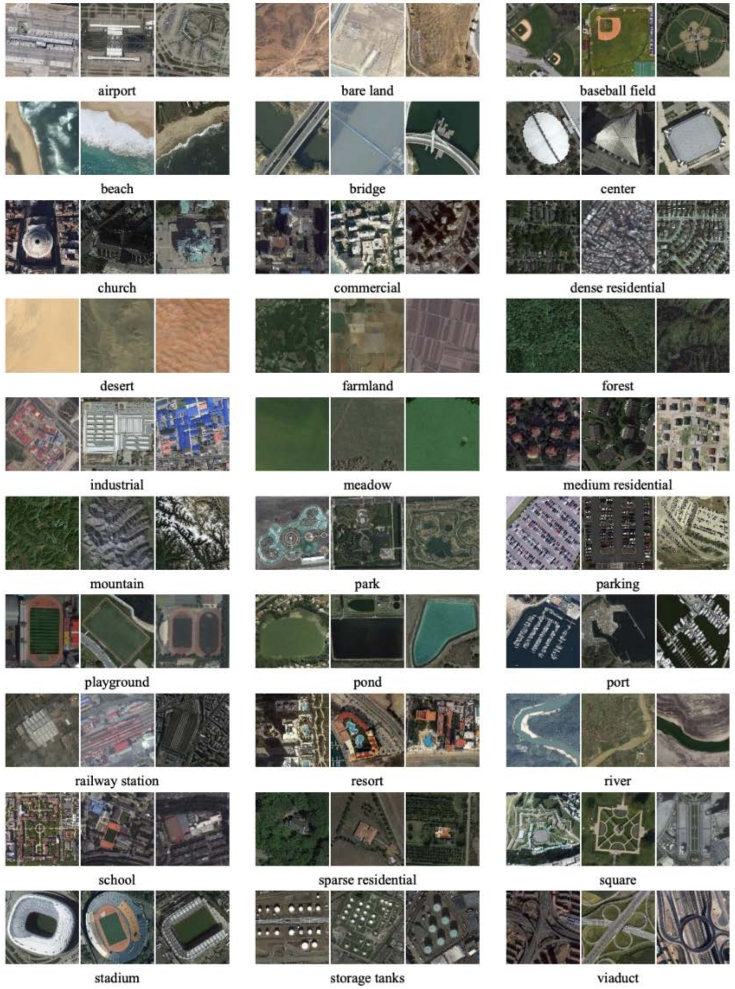


Figure 1: Samples of AID: three examples of each semantic scene class are shown. There are 10000 images within 30 classes.

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

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

...

--

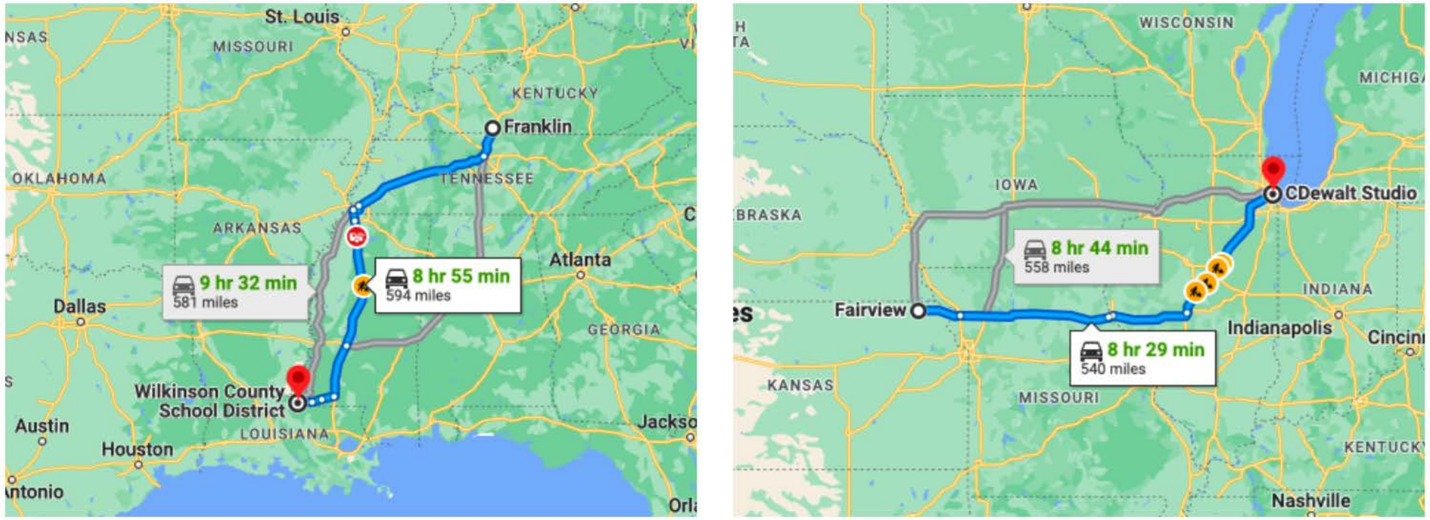
Paragraph: the city of fairview had a population of 260 as of july 1, 2015. ...

Q: Which words in this paragraph represent named places?

A: Fairview

Q: What is the location of Fairview?

A: 41.85003, -87.65005



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

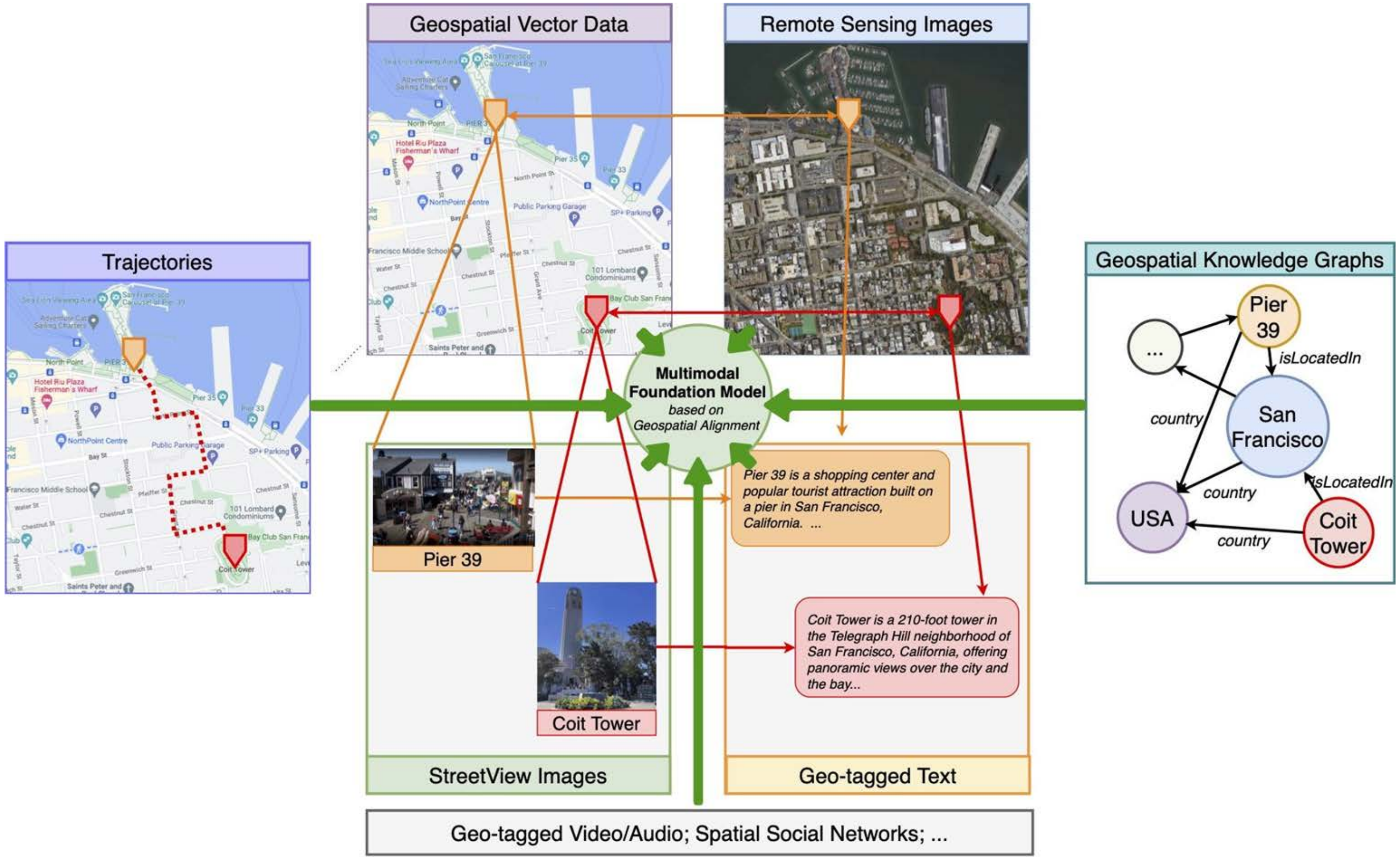
(b) [TEXT]: the city of Fairview had a population of 260 as of july 1, 2015. ...

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.

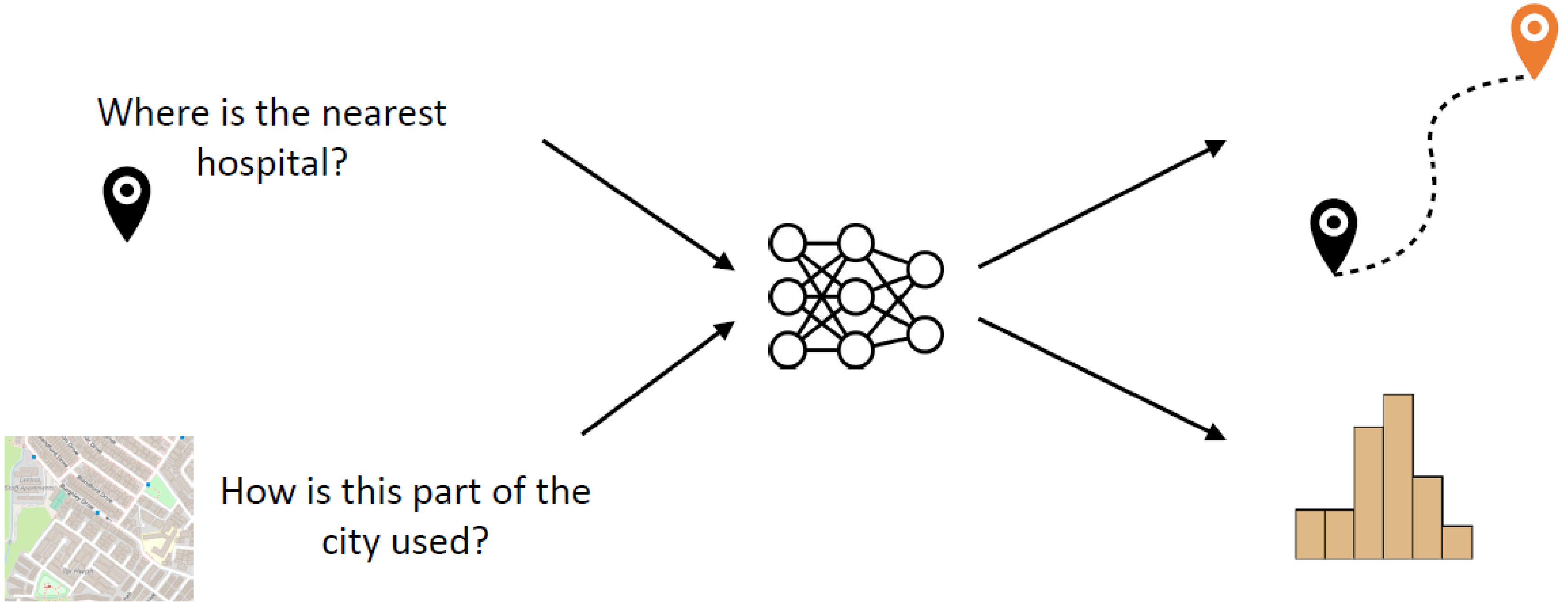
A Multimodal City FM for GeoAI

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



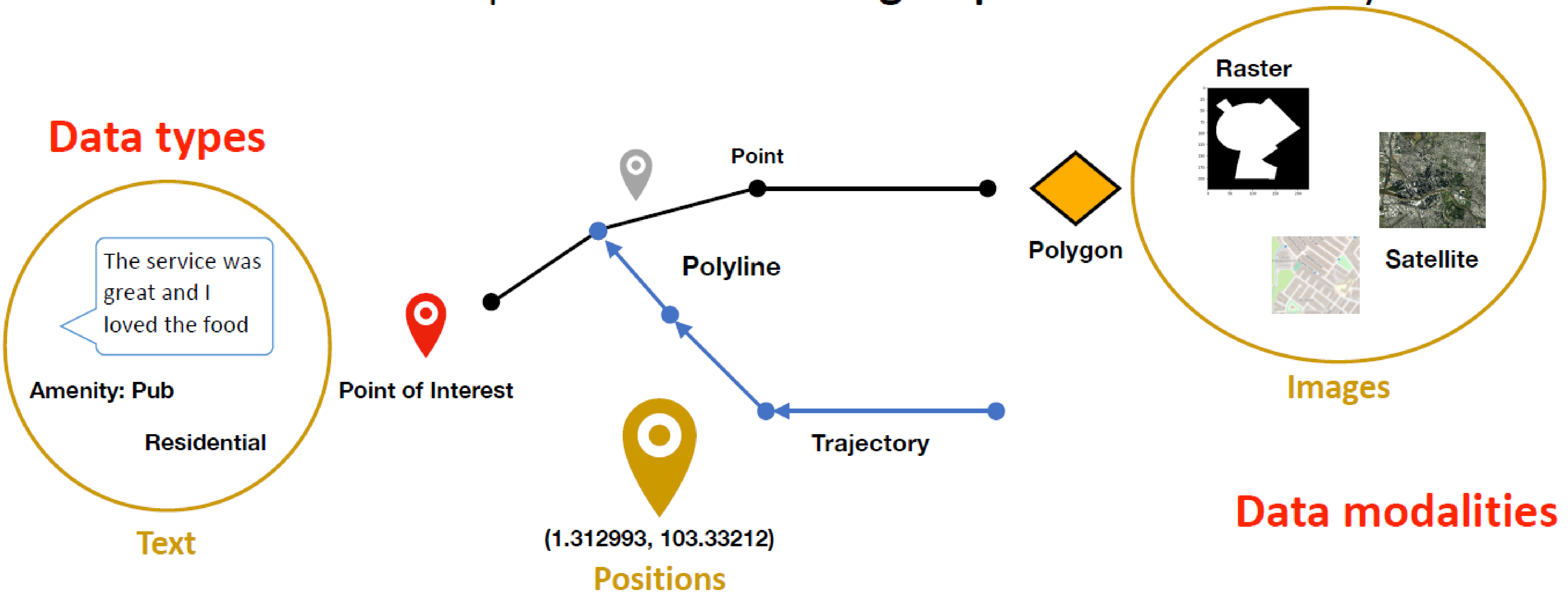
Motivations of City Foundation Models

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!

We use OpenStreetMap. Why?

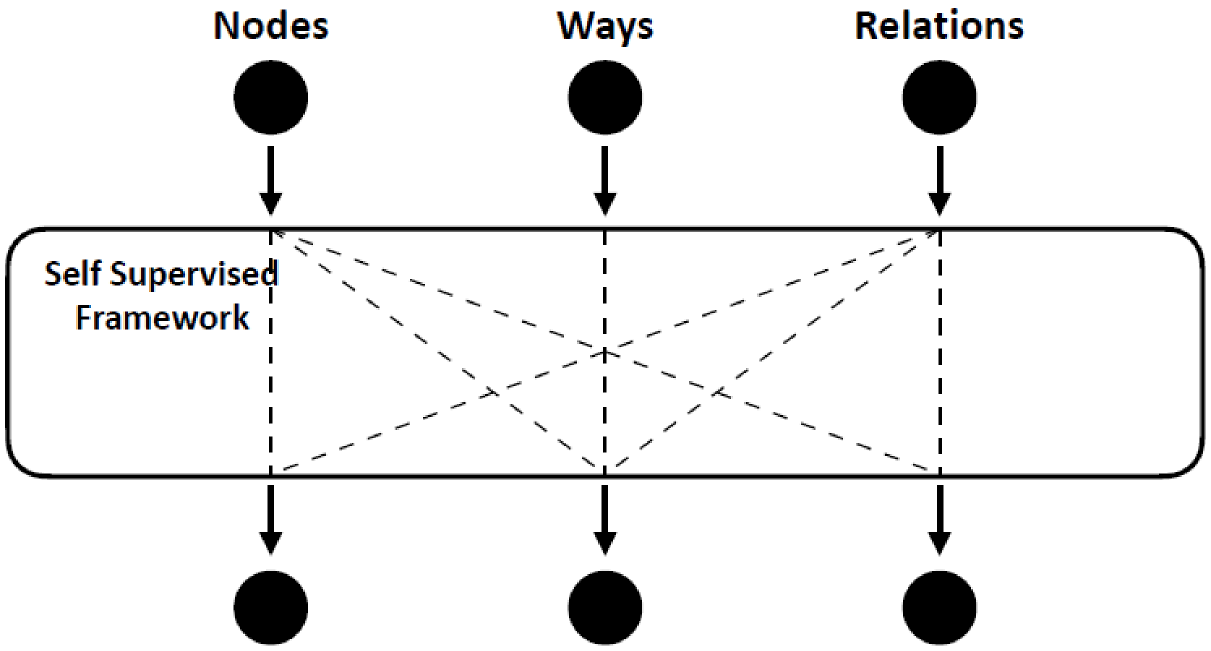
What about using free open data?

OpenStreetMap (OSM) is a collaborative project to create a free, editable map of the world:

- Available **everywhere** and free to use
- As of 2023, OSM stores **9 Billion** geospatial entities
- Several **APIs** to query, modify and expand its database

Proposed Solution

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



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. arXiv preprint arXiv: (2023)
- Pasquale Balsebre, et al. Towards City Foundation Models. Ongoing work.

Shang Liu, NTU

Meituan

Pasquale Balsebre, NTU

Yi Li, NTU

Yile Chen, NTU

Gengchen Mai, Univ of Geogia

Weiming Huang NTU