

Location Encoding

CSCI 587 Lecture 25













to perform a *wide range* of tasks





to perform a *wide range* of tasks









Can we utilize current Foundation Models for GeoAl tasks?

In this lecture: LLMs



Part 1: Geospatial Semantics



(a) **Toponym Recognition**

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



Answers generated by GPT-3



Part 1: Geospatial Semantics



(a) Toponym Recognition

School of Engineering

This is a set of allow more recentition mobile
The `Deregreent' is a set of text containing places
The goal is to infer which words represent nemed places.
the goal is to infer which words represent hand places
\rightarrow in this paragraph, and split the named places
\rightarrow with ;
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: Settling in the Xenia area in 1856 the year
Talagiaph. Settining in the Achina area in 1050, the year
→ alter Bourbon County was organized in 1855, were
John van Syckle, Samuel Stephenson and Charles
Anderson.
Q: Which words in this paragraph represent named places?
A: Xenia; Bourbon
Paragraph: [TEXT]
O: Which words in this paragraph represent named places?
2. Minin Morae in this paragraph represent human process
A:
Eavy Shot Adaptation
rew Shot Audplation
TICOVAL
USC Viterbi

Location Descriptors

This is a set of location description recognition
↔ problems
The `Paragraph` is a set of text containing location
↔ descriptions.
The goal is to infer which words represent location
\hookrightarrow descriptions in this paragraph, and split
← different location descriptions with `;`.
Paragraph: Papa stranded in home. Water rising above waist
→ HELP 8111 Woodlyn Rd 77028 #houstonflood
O: Which words in this paragraph represent location
Q: which words in this paragraph represent rocation
↔ descriptions?
A: 8111 Woodlyn Rd, 77028
Paragraph: Major flooding at Clay Rd & Queenston in west
← Houston. Lots of rescues going on for ppl
↔ trapped.
Q: Which words in this paragraph represent location
↔ descriptions?
A: Clay Rd & Queenston; west Houston
Paragraph: [TEXT]
Q: Which words in this paragraph represent location
↔ descriptions?

A:

(b)

Part 1: Geospatial Semantics





	#Param	Toponym Recognition		Location Description Recognition		
Model		Hu2014	Ju2016	HaveyTweet2017		2017
		Accuracy \downarrow	Accuracy \downarrow	Precision \downarrow	Recall ↓	F-Score ↓
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

Treated as token generation task by LLMs ⇒ Location might not exist in prompt

Why/How?

- Massive text data during pretraining: Named places *frequently appear* on text.
- Contextual cues: I [visited] Paris, The house is [located at] Jefferson
- Co-occurrence patterns: in/at location, from place





[Instruction]	Few Shot Adaptation
Paragraph: San Jose was founded in 1803 when allotments of → made	land were
Q: Which words in this paragraph represent named places? A: San Jose; New Mexico	Toponym Recognition
Q: What is the location of San Jose? A: 35.39728, -105.47501	Location Prediction





[Instru	uction]	Few Shot Adaptation
Paragra	ph: San Jose was founded in 1803 when allotments of	land were
	∽ made	_
Q: Wh	ich words in this paragraph represent named places?	Toponym Pocognition
A: Sa	n Jose; New Mexico	
Q: Wh	at is the location of San Jose?	Location Prediction
A: 35	.39728, -105.47501	
Paragra	aph:	
	\hookrightarrow the city of fairview had a population of 260 as of july 1	, 2015
Q: Wh	ich words in this paragraph represent named places?	
A: Fa	airview	
Q: What	is the location of Fairview?	
A: 41.8	35003, -87.65005	

















Which Georgia cities are in the north of Athens



There are several cities in the north of Athens, Georgia. Some of the cities that are located \bigcirc \bigcirc to the north of Athens include:

- 1. Winder
- 2. Monroe
- 3. Loganville
- 4. Lawrenceville
- 5. Suwanee
- 6. Buford
- 7. Gainesville
- 8. Braselton

These are just a few examples of the many cities and towns that are located to the north of Athens, Georgia.



University of Southern California

Wrong. Monroe is SW of Athens



Effectiveness of current FMs



- Current FMs struggle in tasks that require **spatial reasoning!**
 - Lack the ability to *understand*, *represent*, and *infer* spatial relationships between objects or locations in space.





Effectiveness of current FMs



- Current FMs struggle in tasks that require **spatial reasoning!**
 - Lack the ability to *understand*, *represent*, and *infer* spatial relationships between objects or locations in space.
 - FMs could generalize these patterns across tasks like map understanding, trajectory prediction and spatial querying.







We need to develop GeoAl Foundation Models to address GeoAl tasks.

... Comes with **unique challenges**



Geospatial Data are multimodal



Not straightforward to handle



(Next Lecture)







Unsupervised text encoding





Word Embeddings

Words used in *similar* meanings and contexts should be embedded *close* together.

Similarly, we want to embed locations that share **similar spatial context** close to each other in the embedding space.



Location encoding: *Direct*







- Lack of inductive bias ⇒ High-variance model
 - *Memorize coordinate patterns instead of learning generalizable structures.*
- **Poor generalization** across scales/regions *e.g. NYC vs LA*



Location encoding: Tile-based



Discretization



- Need to choose the right *cell size*.
 - Depends on the data distribution.



Location encoding: Challenges



density of other POI types

Normalized expected number of POIs of the same type within radius



Location encoding: Motivation





Women's Clothing (Clustered Distribution)



Education (Even Distribution) Joint modeling distributions with very different characteristics ?



Location encoding: *Motivation*







Idea: Multiscale Representation Learning

Women's Clothing (Clustered Distribution)

Education (Even Distribution)



Mai, Gengchen, et al. "Multi-scale representation learning for spatial feature distributions using grid cells." *International Conference on Representation Learning (ICLR)* (2020).

Problem Formulation



Distributed representation of point-features in space:

Given a set of points $\mathcal{P} = \{p_i\}$, i.e., Point Of Interests (POIs), in L-D space (L = 2,3), each point $p_i = (\mathbf{x}_i, \mathbf{v}_i)$ is associated with a location \mathbf{x}_i and attributes \mathbf{v}_i (i.e., POI feature such as type, name). We define function

$$f_{\mathcal{P},\theta}(\mathbf{x}): \mathbb{R}^L \to \mathbb{R}^d \ (L \ll d)$$

which maps any coordinate x in space to a vector representation of d dimension

L: Original dimension. *L=2 in euclidean space*.

d: Embedding dimension.



Space2Vec: Idea

Goal: Define an encoding function *f*(*x*, *y*) that captures:

- **Relative Distance**: POIs close to each other should have similar embeddings
- Periodicity: Spatial positions that repeat in structure (e.g., 180° and –180°) share the same or similar encodings
- Multi Scale: Should capture both local and global details



School of Engineering



Space2Vec: Idea



Encoding function: $f(x, y) = [cos(\omega x), cos(\omega y)]$





Space2Vec



Encoding function: $f(x, y) = \cos(\vec{a} \cdot \vec{x})$ where $\vec{x} = (x, y)$, $\vec{a} = \text{unit vector}$



Space2Vec: Location Encoder



$$PE_{s}^{theory}(\mathbf{x}) = \left[PE_{s,1}^{theory}(\mathbf{x}); PE_{s,2}^{theory}(\mathbf{x}); PE_{s,3}^{theory}(\mathbf{x})\right],$$

$$where PE_{s,j}^{theory}(\mathbf{x}) = \left[\cos\left(\frac{\langle \mathbf{x}, \mathbf{a}_{j} \rangle}{\lambda_{min}}\right); \sin\left(\frac{\langle \mathbf{x}, \mathbf{a}_{j} \rangle}{\lambda_{min} \cdot g^{s/(S-1)}}\right)\right], \forall j = 1, 2, 3.$$

$$\min scale size$$

- $\lambda \min, \lambda \min$: scale size bounds
- S : total number of scales



Space2Vec: Location Encoder



$$PE_s^{theory}(\mathbf{x}) = [PE_{s,1}^{theory}(\mathbf{x}); PE_{s,2}^{theory}(\mathbf{x}); PE_{s,3}^{theory}(\mathbf{x})],$$

where
$$PE_{s,j}^{theory}(\mathbf{x}) = \left[\cos\left(\frac{\langle \mathbf{x}, \mathbf{a}_j \rangle}{\lambda_{min}}\right); \sin\left(\frac{\langle \mathbf{x}, \mathbf{a}_j \rangle}{\lambda_{min} \cdot g^{s/(S-1)}}\right)\right], \forall j = 1, 2, 3.$$

min scale size

$$PE^{(t)}(\mathbf{x}) = [PE_0^{(t)}(\mathbf{x}); ...; PE_s^{(t)}(\mathbf{x}); ...; PE_{S-1}^{(t)}(\mathbf{x})]$$

repeat for every scale

$$Enc(\mathbf{x}) = \underbrace{\mathbf{NN}(PE(\mathbf{x}))}_{network}, \qquad can \ be \ any \ neural network$$



Space2Vec: Feature Encoder



Point feature encoder $Enc^{(v)}()$ encodes such features \mathbf{v}_i into a feature embedding $\mathbf{e}[\mathbf{v}_i] \in \mathbb{R}^{d^{(v)}}$

For example, if each point represents a POI with multiple POI types, the feature embedding can simply be the mean of each POI types' embeddings:

$$\mathbf{e}[\mathbf{v}_i] = \frac{1}{H} \sum_{h=1}^{H} \mathbf{t}_h^{(\gamma)}$$

 $\mathbf{t}_{h}^{(\gamma)}$ indicates the hth POI type embedding of a POI pi with H POI types



Space2Vec: Decoders



So far we have defined 2 types of POI embeddings

- 1. Location / Space Embedding: e[x], through location encoding
- 2. **Feature Embedding:** e[v], through feature encoding

Location Decoder $Dec_s()$: Directly reconstructs point feature embedding $e[v_i]$ given its space embedding $e[x_i]$

$$\mathbf{e}[\mathbf{v}_i]' = Dec_s(\mathbf{x}_i; \theta_{dec_s}) = \mathbf{N}\mathbf{N}_{dec}(\mathbf{e}[\mathbf{x}_i])$$

For training we use inner product to compare the reconstructed feature embedding $\mathbf{e}[\mathbf{v}_i]'$ against the real feature embeddings $\mathbf{e}[\mathbf{v}_i]$ and other negative points



Space2Vec: Decoders

Spatial Context Decoder $Dec_c()$: reconstructs the feature embedding $\mathbf{e}[\mathbf{v}_i]$ of the center point p_i based on the space and feature embeddings $\{\mathbf{e}_{i1}, ..., \mathbf{e}_{ij}, ..., \mathbf{e}_{in}\}$ of n nearby points $\{p_{i1}, ..., p_{ij}, ..., p_{in}\}$





Space2Vec: Unsupervised Training



The unsupervised learning task can simply be maximizing the log likelihood of observing the true point p_i at position x_i among all the points in P

$$\mathcal{L}_{\mathcal{P}}(\theta) = -\sum_{p_i \in \mathcal{P}} \log P(p_i | p_{i1}, ..., p_{ij}, ..., p_{in}) = -\sum_{p_i \in \mathcal{P}} \log \frac{\exp(\mathbf{e}[\mathbf{v}_i]^T \mathbf{e}[\mathbf{v}_i]')}{\sum_{p_o \in \mathcal{P}} \exp(\mathbf{e}[\mathbf{v}_o]^T \mathbf{e}[\mathbf{v}_i]')}$$

Negative Sampling:

$$\mathcal{L}_{\mathcal{P}}'(\theta) = -\sum_{p_i \in \mathcal{P}} \left(\log \sigma(\mathbf{e}[\mathbf{v}_i]^T \mathbf{e}[\mathbf{v}_i]') + \frac{1}{|\mathcal{N}_i|} \sum_{p_o \in \mathcal{N}_i} \log \sigma(-\mathbf{e}[\mathbf{v}_o]^T \mathbf{e}[\mathbf{v}_i]') \right)$$



Space2Vec: Embedding Clustering





Space2Vec: Location Aware Image Classification





(a) Arctic Fox



(c) Bat-Eared Fox

terbi

School of Engineering





(d) Bat-Eared Fox Locations



	BirdSnap [†]	NABirds [†]
No Prior (i.e. uniform)	70.07	76.08
Nearest Neighbor (num)	77.76	79.99
Nearest Neighbor (spatial)	77.98	80.79
Adaptive Kernel (Berg et al., 2014)	78.65	81.11
tile (Tang et al., 2015) (location only)	77.19	79.58
wrap (Mac Aodha et al., 2019) (location only)	78.65	81.15
$rbf(\sigma=1k)$	78.56	81.13
theory (λ_{min} =0.0001, λ_{max} =360, S = 64)	79.35	81.59



Space2Vec: Land Use Classification



[Instructi	on] There are six land use types: (1) residential, (2) commercial, (3) industrial, (4) education, health care, civic, governmental
\hookrightarrow	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: outd	oors 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,
\hookrightarrow	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?

coutdoors and natural

				Model	Accuracy 1	Precision	Recall
				GPT2 [115]	0.318	0.105	0.158
Space 2 Vec Unsupervised Training			GPT2-Medium [115]	0.025	0.102	0.040	
Spacezvec Unsupervised Training				GPT2-Large [115]	0.005	0.001	0.002
	• Accuracy: 0.540		(B) Zero-shot LLMs	GPT2-XL [115]	0.001	0.108	0.002
	 Precision: 0.512 circulture 	cimply prodicts t	prodicts the	GPT-3 [15]	0.144	0.448	0.141
		majority class	ne	ChatGPT (Raw.) [104]	0.075	0.376	0.106
٠	• <i>Recall:</i> 0.516			ChatGPT (Con.) [104]	0.051	0.232	0.046
				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
			(C) Qne-shot LLMs	GPT2-XL [115]	0.324	0.105	0.159
USC Viterbi School of Engineering			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

References



[1] Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., ... & Lao, N. (2023). On the opportunities and challenges of foundation models for geospatial artificial intelligence. TSAS 2024.

[2] Mai, G., Janowicz, K., Yan, B., Zhu, R., Cai, L., & Lao, N. (2020). Multi-scale representation learning for spatial feature distributions using grid cells. *ICLR 2024*.

