

## Adopting Semantic Signatures to Quantify the Uncertainty of Volunteered Geographic Information

## Rui Zhu University of California, Santa Barbara







## Outline

- Motivation
- Methodology
- Experiment
- Discussion

#### **Motivation**















SpatialAccuracy2018





#### **Motivation**

Uncertainty of VGI:

- Positions: coordinates, address, postcode ...
- Geometric representations and topological relations
- Attributes: number of check-ins, number of employees
- Semantics: place types



#### **Motivation**

Semantic Uncertainty of VGI

- Restaurants in Foursquare and Google places might not be the same.
- *Mountains* in DBpedia Places are different from *Mountains* in GeoNames.
- A place should be labeled more as a restaurant or as a bar?
- Will spatial contexts help to reduce such uncertainties?



Conventional approach for uncertainty analysis

- Numeric data
  - $\circ$  Distribution of the data  $\rightarrow$ variance, entropy, Bayes theorem
- Categorical data
  - Indicator statistics
  - $\circ$  Semantic signatures  $\rightarrow$  semantic distribution of place types



# Methodology **Spatial Analysis Spatial** Temporal Semantic Signatures Spatial Signature Thematic

Janowicz, K., McKenzie, G., Hu, Y., Zhu, R., and Gao, So. (2018): <u>Using Semantic Signatures for Social Sensing in Urban Environments</u>. Mobility Patterns, Big Data and Transport Analytics.

#### SpatialAccuracy2018





Spatial Signatures

- **Spatial structure** of the data belonging to a place type is used to quantify its semantics.
- **Spatial statistics** are applied to describe such spatial structure.
- Spatial point patterns. Spatial autocorrelation analysis, spatial interaction analysis with other geographic features, place-based analysis. → 41 statistics



**SpatialAccura** 

Spatial Signature -- Spatial point patterns

- Intensity-based: local intensity, kernel density estimation
- Distance-based: nearest-neighbor distance, Ripley's K, and standard deviational analysis





Spatial Signature -- Spatial point patterns -- Examples



Figure 1: Ripley's K of Park (left) and Dam (right) from DBpedia Places.

Statistics: mean and std. of the deviation between theoretical can observed K curves



Spatial Signature -- Spatial AutoCorrelation Analysis

- Moran's I: how intensities of cells differ from their neighbors
- Semivarigram: measure the variation of cell intensities in a specific distance lag class. Dams in GeoNames



Cell size : 36 km \* 22.2 km

Cell value: number of instances falling in the cell

Rui Zhu



Spatial Signature -- Spatial point patterns -- Examples



Figure 2: Experimental semivariogram of Park (left) and Dam (right) from TGN.

Statistics: mean and std. of the semivariance at first, median and last lag distance



Spatial Signature -- Spatial Interaction with Other Geographic features

- Population
- Climate
- Road network





Spatial Signature -- Spatial Interaction with Other Geographic features -- Examples



Figure 3: Histogram of road types for Amusement Park (left) and Restaurant from Google.



Spatial Signature -- Place-based statistics

In contrast to spatial statistics, they focus more on describing the *topological* and *hierarchical relations* between places.

- The number (and entropy) of distinct states (or counties) a place type occurs in;
- The number (and entropy) of adjacent states (or counties) that also contain features of the target type;



Spatial Signature -- Place-based statistics -- Examples

• Distinguishing feature types such as Glacier (which occur in eight US-states according to DBpedia) and River (which occur in all states).





Spati	al Point Pattern	Spatial Autocorrelations	ocorrelations Spatial Interaction with Other Geographic Features			Place-based statistics	
	Intensity			min	Number of distinct states (or counties)		
	Mean distance to nearest neighbor		Population	max			
Local	std. of distance to nearest neighbor Global Moran's I			mean	Entropy of states (or counties)		
Local	Kernel density (range)			std.	Number of adjacent states (or counties) that have the come feature time		
	Kernel density (bandwidth)			min of shortest distance			
	Ripley's K (range)	oley's K range)		max of shortest distance	same feature type		
	Ripley's K (mean deviation)			mean of shortest distance	Number of distinct feature types for nearest neighbor		
	std. ellipse (rotation)	Semivariogram		std. of shortest distance			
	std. ellipse (std. along x-axis)	(first distance lag)		entropy of nearest road types	Entropy of feature types for nearest neighbor		
	std. ellipse (std. along y-axis)			$mean \\ precipitation$			
	Intensity	Semivariogram (median distance lag)		std. $precipitation$		Mean KL Divergence of the topic	
Clobal			Climate	mean temperature max	LDA-based		
Giobai				std. temperature max	approach	aistriotion	
	Kernel density (range)			mean temperature min			
		Semivario qram		std. temperature min		Entropy of	
	Kernel density (bandwidth)	(last distance lag)		mean water		distribution	
				std. water vapor pressure	-		

SpatialAccuracy2018



#### Experiment

1. Similarity of place types





#### Experiment

#### 2. Coreference resolution



Property

abo abstract



#### GeoNames Home | Postal Codes | Download / Webservice | Abou

SpatialAccuracy2018

	Kobani	all count	tries :			
	sear	ch show on map [advanced search]				
					13 records	found for "Kobani"
	Name	Country	Feature class		Latitude	Longitude
1 🖲	Kobani Kobani Kohani	Mali, Sikasso	intermittent stream		N 11° 5' 23"	W 6° 49' 47''
2 🕐	( <u>'Ayn al 'Arab</u> 🥥 Aarab Peunar, Aarab Peunàr, Ain el Aarab, Arab Peunar, Ain el Aarab, Ein-al-Arab, Kobane, Kobani, ' Arab Bina	Syria, Aleppo	seat of a second-orde population 50,000	f a second-order administrative division on 50,000		E 38* 21' 12"
3 19	Mkoani 🎱 Kobani, Mkoani	Tanzania, Pemba South Mkoani District > Mbuyuni	populated place		S 5° 22' 0"	E 39° 39' 0''
4 🖤	Nàhiyat Markaz 'Ayn al 'Arab Kobane Kobani, Kobani, Kobani, Kubane, Kubani, Kubani, Kubani, Kūbāni, Nahiyat Markaz 'Ayn al 'Arab, Nāhiyat	Syria, Aleppo	third-order administra	ative division	N 36° 48' 17"	E 38° 23' 27"
5 🖲	Kobani	Ivory Coast, Savanes	intermittent stream	💐 DBpedia 🛛 💿	Browse using -	Formats -
6 🖲	Kobani	Ivory Coast, Denguélé	intermittent stream			
7 🖲	Kobani	Ivory Coast, Woroba	intermittent stream	About: Kobanî		
8 🖲	Kobani	Ivory Coast.	intermittent stream			
9 🖲	Kobani	Ivory Coast, Savanes	stream	An Entity of Type : settlement, from Named Graph : http://dbpedia.org, wit		
10 🖲	Kobani	Ivory Coast, Santiago Metropolitan Region	stream			
11 🖲	Kobani	Ivory Coast, Denguélé	stream	Kobanî (Kurdish: كۆبانى pronounced [koˈbaːni North Levantine pronunciation: [ʕe عين العرب		nced [ko'ba:ni:], als
12 🛡	Kobani	Ivory Coast, Denguélé	stream			nunciation: [Se:n el'
13 🖲	Razvaliny Kobani Razvaliny Kobani	Georgia,	ruin(s)	immediately south of the border with Turkey. As a c of the Kurdish YPG militia since 2012.		

#### Which Kobani?

C Faceted Browser C Sparql Endpoint

Space : dbpedia.org

Value **7.0** 

ndered Kobanê [ko'ba:ne]), also known as Ayn al-Arab (Arabic: b]), is a city in the Aleppo Governorate in northern Syria, lying equence of the Syrian Civil War, the city has been under control

> Kobani (Kurdish: كوبانى pronounced [ko'ba:ni:], also rendered Kobanê [ko'ba:ne]), also known as Ayn al-Arab (Arabic: عين العرب North Levantine pronunciation: [Se:n el'Sorob]), is a city in the Aleppo Governorate in northern Syria, lying immediately south of the

Have to use the feature types in addition to string and spatial distances.

	border with Turkey. As a consequence of the Syrian CuVI War, the city has been under control of the Kurdieh YPG millis since 2012. In 2014, it was underlaugh declared to be the administrative enter of the Kobani Chanon of Bojaar. Prove 2014 to January 2015, the city was under siege by latamic State of Iraq and the Levant. Most of the city was destroyed and most of the population field to Turkey. In 2015, many returned and reconstruction began. Prior to the Syrian CiVI War. Kobani was recorded as having a population of close to 45,000. The majority of inhabitants were Kurds, with Arab, Turkmen, and Armenian minorities. (w)
do:areaTotal	<ul> <li>7000000.000000 (xsd.souble)</li> </ul>
abo:country	■ œr/Syria
abo:elevation	<ul> <li>\$20,000000 (xst.double)</li> </ul>
do:IsPartOf	exrAyn_al-Arab_District     exrAtepo_Governorate





## Summary & Discussion

- Semantic uncertainty of VGI has to be understood and quantified
- Semantic signatures are introduced to quantify the semantic uncertainty

However ...

- Need a framework/guideline of using semantic signatures;
- To use semantic uncertainty to infer other types of uncertainties
- From exploratory study to solving emergent VGI challenges: federated geographic information retrieval, place alignment, data cleaning...
- More complex spatial patterns could be quantified and incorporated into the signature set;
- To design new spatial-semantic interaction statistics: modified LISA, Ripley's K ...



#### Thanks a lot!

#### Any questions / comments?

