MapSpace: POI-based Multi-Scale Global Land Use Modeling

Gautam Thakur
Geoinformatics Engineering Group, Geospatial Science and Human Security Division, Oak Ridge National Laboratory, USA.
thakurg@ornl.gov

Junchuan Fan
Geoinformatics Engineering Group, Geospatial Science and Human Security Division, Oak Ridge National Laboratory, USA.
fanj@ornl.gov

Abstract

Accurate and up-to-date land use maps are important to the study of human-environment interactions, urban morphology, environmental justice, etc. Traditional land use mapping approaches involve several surveys and expert knowledge of the region to be mapped. While traditional approaches generate accurate and authoritative maps, it is expensive and takes a long time to develop a new version of map. Besides, such maps have region-specific spatial embedding, making them difficult to benchmark and compare against other land use maps. This work introduces a scalable POI-based land use modeling approach to generate global land use maps at multiple spatial scales and different semantic granularities. In addition, our land use maps adhere to a unified land use categories and can be compared for accuracy and precision.

Acknowledgements This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

1 Introduction

Accurate development of land use maps of a geographic region is critically important for the informed policy-making of different stake-holders, such as urban planning, environmental justice, and economic development. Points of Interest(POI) within a geographic region are suitable indicators of how humans use the space, and land use characteristics are often reflected by the POIs within. The logical correspondence between POI features and land use types, coupled with the expanding coverage and timeliness of big POI data, offers researchers a unique opportunity to model land use dynamics based on POI features (e.g., category) at varying spatial scales and semantic granularities. In this research, land use means the socioeconomic functions for which a land is used, not the physical nature or form of the land surface. POIs' rich features can be projected in multiple dimensions, such as a spatial dimension (i.e., geographic coordinates, polygons), a semantic dimension (e.g., functional use) and sometimes a temporal dimension (e.g., time of existence, opening hours). A combination of spatial distribution, semantic characteristics, and temporal dynamics of POIs within an area of interest (AOI) reflects its unique land use characteristics. For example, an industrial factory usually has POIs like *warehouse*, *storage tank* whereas a recreational park would have amenities like *bench*, *playground*. Thus, the integration of spatial, semantic, and temporal aspects of POIs offers valuable insights into the land use dynamics at different spatial scales and semantic granularities. With the fast advancement of natural language processing research and growing geo-computing capabilities, the fusion and mining of spatial and semantic dimensions of crowd-sourced geographic information have gained lots of attention. Geo-referenced Wikipedia articles were used to identify spatial patterns of natural hazards like wildfire across US [2]. POI data was integrated with word embeddings to classify urban land use at TAZ level [6] . [1] trained land use classifier for Europe based on POI-related features at grid cell level. The advantage of POI-based land use modeling is its general applicability to different geographic regions, and land use maps can be generated at different spatial scales and semantic granularities by adapting to the POI data coverage.

This short paper introduces a scalable POI-based approach for global land use modeling at multiple spatial scales and different semantic granularities. We will first discuss important factors for POI-based land use modeling, including data fusion, spatial scale, and semantic granularity. Then, we explore the most appropriate scale and semantic granularity for POI-based land use modeling at three different geographic regions around the world.

2 POI-based land use modeling

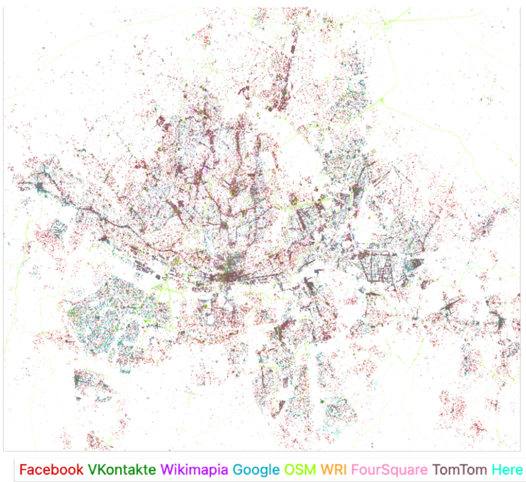


Figure 1 POI coverage of different POI data sources for Johannesburg

Different location-based platforms may partially overlap (or not) in their spatial data coverage. As seen in Figure 1, the spatial coverage of nine different online POI data sources is quite different from each other. Effectively fusing POI from these sources can significantly augment the coverage of an AOI. Besides, unlike remotely sensed images that have pixels covering the entire AOI, POIs are sample points scattered inside an AOI with rich information. There is no preset spatial resolution for the POI-based land use model. POI coverage could also vary from AOI to AOI. Therefore, the *spatial scale* and *semantic granularity* of POI-based land use modeling should be adaptive to specific AOI. *Data fusion*, *spatial scale* and *semantic granularity* are three critical factors for POI-based global land use modeling.

2.1 Data fusion

Data heterogeneity across different data sources is common for non-authoritative crowd-sourced geographic data. Therefore, the effective fusion of heterogeneous POI data from diverse sources is critically important. Furthermore, the data fusion process could serve as a foundation for further incorporation of other data sources like social media data, ground images. In [3], we developed a Semantic Ontology Network (SONET) to deal with the heterogeneity of POI categorization across multiple data sources. The POI categories from data sources were translated to *OSM tags* that serve as an intermediate semantic bridge. For example, *Restaurant* from Facebook, *food* from Google and *eatery* from Wikimapia are all mapped to OSM tags *amenity=restaurant;building=retail*. The OSM tags were effectively matched to data sources where POI categories vary not only in number but also in semantic granularity (e.g., general categories such as *restaurant*, and *hospital*, coexist with more specific categories such as *Italian Restaurant* and *Dental Hospital*). Furthermore, all the OSM tags are organized into a three-level land use category hierarchy. SONET is a graph database that contains 12,667 source linked categories from nine different data sources and is still growing.

2.2 Spatial scale

Spatial scale is a critical parameter for any spatial analysis. Concerning POI-based land use modeling, spatial scale refers to the spatial granularity at which we partition the space, and the resulting *spatial block* will be the smallest unit of which land use will be modeled. An AOI can be partitioned based on a simple grid and generate squared grid cells like image pixels, or it can be partitioned through Delaunay or Voronoi triangulation. These partitioning approaches emphasize the geometrical aspect of space. They do not pay enough attention to the geographical aspect of the space, i.e., how people interact with and use the space. One natural way to partition the space that can organically capture the way human activities interact with space is to partition the space based on the road networks. Road network is essential for people to carry out their daily activities and road network-based spatial blocks are organic spatial units for POI grouping and subsequent semantic clustering analysis. The hierarchy of segments in a road network naturally gives us the flexibility to partition space at different spatial granularities without losing spatial and semantic coherence.

2.3 Semantic granularity

Semantic granularity refers to the level of semantic specificity of land use category of a *spatial block*. For example, a spatial block can simply be labeled as *nonresidential*, or specifically *institutions/public services*, or more specifically *religious*. The semantic granularity of land use model is closely related to the spatial scale of the land use model. A small spatial block generated from lower-level roads (e.g., residential roads, service roads) can be assigned a fine-grained land use type, whereas spatial block generated from higher-level road network (e.g., tertiary roads) can only be characterized by coarser land use category.

2.4 Multi-scale semantic clustering of POI tags

With the data fusion process, POIs from different data sources are merged and translated into a collection of OSM tags. The data curation was achieved through the platform discuss in [5]. The spatial distribution and semantic composition of POIs within an AOI offer valuable insights into an AOI's land use. However, finding the appropriate spatial scale and semantic granularity for the land use model is critical. Figure 2 demonstrates the changes of land use maps as the spatial scale and semantic granularity change. We use topic modeling as a semantic clustering tool to explore the interactions between spatial scales and semantic granularity of POI-based land use modeling for three selected AOIs. Topic modeling is a collection of algorithms that uses probabilistic generative model to uncover the latent theme of a text corpus. In our experiment, an AOI is partitioned into spatial blocks with different levels of the road network. We model *POI tags* as words and group POI tags based on

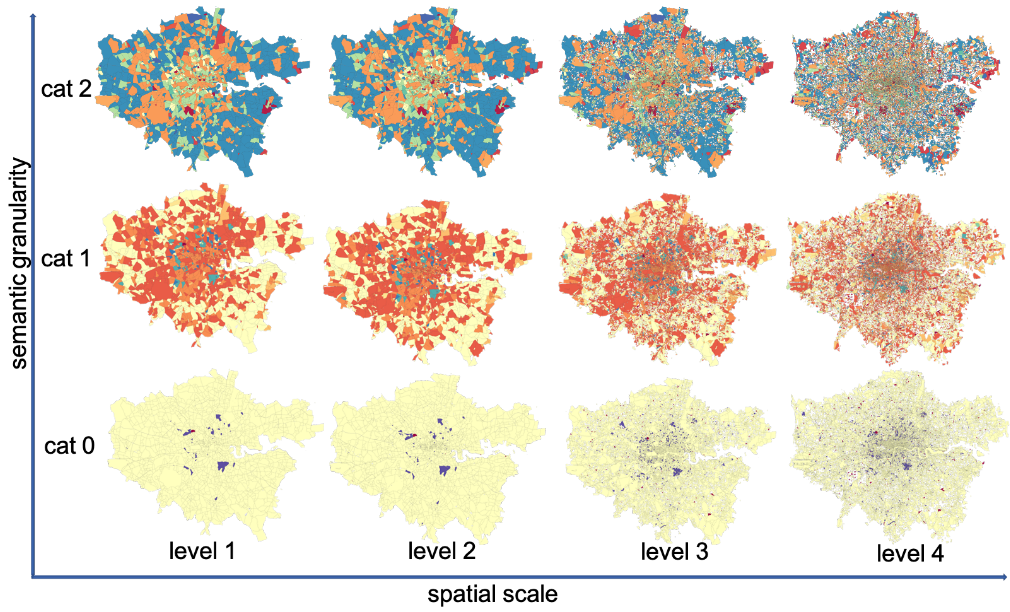


Figure 2 Land use maps change with spatial scale and semantic granularity

hierarchical spatial blocks generated from different levels of road networks. As a result, the POIs inside the same *level n* spatial block will form a *POI tag sentence*, the POIs inside *level n* spatial blocks that are inside the same *level n-1* spatial block will form a *POI tag document*. The hierarchical organization of POIs captures the spatial distribution patterns within a geographic region. The land use of a spatial block is modeled as the latent semantic topics over the *POI tag document* formed by POI tags.

3 Experiment

We test the efficacy of our approach on three AOIs from three different countries on three different continents, including *Seoul (South Korea)*, *Johannesburg (South Africa)*, and *London (UK)* (as shown in Figure-3). The road network is divided into four levels: ***level 1*** roads include *tertiary* and higher than *tertiary* level roads, including *motorway*, *primary*,*trunk*, *secondary*, and *tertiary*; ***level 2*** roads include *unclassified* level roads in addition to *level 1*; ***level 3*** roads include *residential* level roads in addition to *level 2*; and ***level 4*** roads include all the roads. For the POI data, nine different POI data sources were collected and semantically conflated for these AOIs.

3.1 Interaction of spatial scale and semantic granularity

For each of the three selected AOIs, we partitioned the space and generated spatial blocks at four different spatial granularities. By grouping POIs with spatial blocks of different granularities, we can create and evaluate the semantic clustering results at different spatial scales. In addition, we trained topic models with different pre-set number of latent topics at each spatial scale to test the effect of changing semantic granularity. By adjusting number of topics (3∼75), the topic modeling can detect semantic clusters with varying granularities. Ideally, the best way to evaluate the topic models trained at different spatial scales and different semantic granularities is through human topic ranking, but they are expensive to

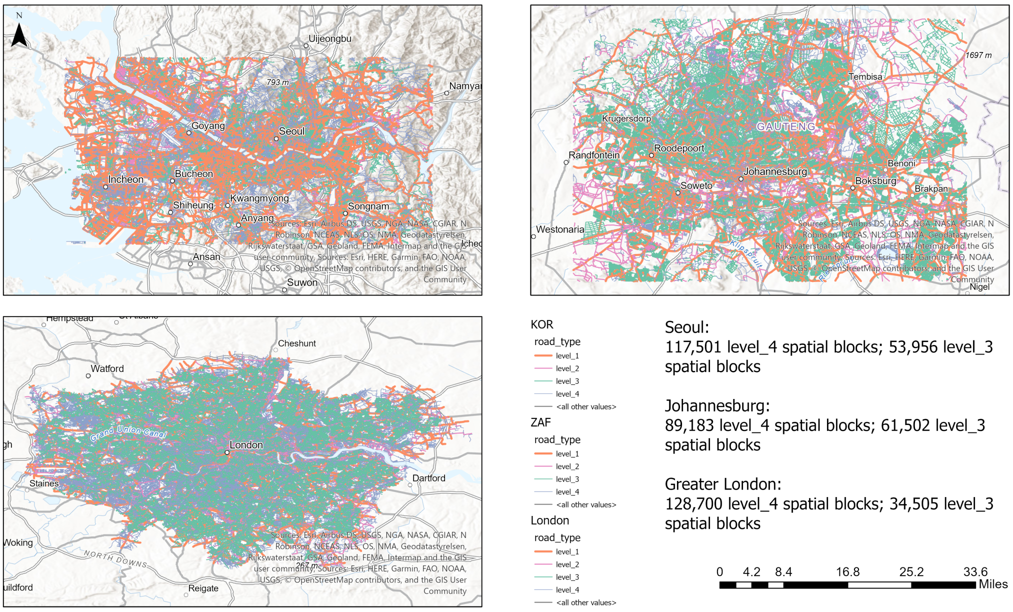
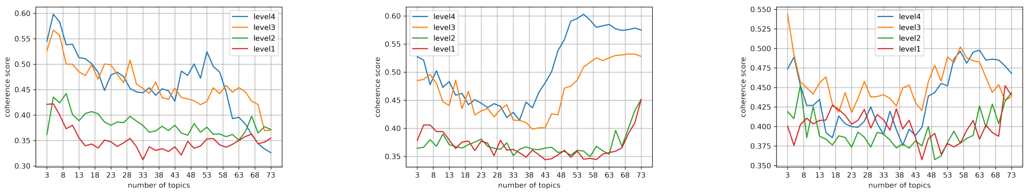


Figure 3 Network hierarchy of three study regions



(a) London **(b)** Seoul **(c)** Johannesburg

Figure 4 Coherence scores of different topic models at different spatial scales

produce. Here, the result topic models are evaluated through semantic coherence measures [4], which are calculated based on pointwise mutual information between word pairs. Figure 4 shows the semantic coherence scores of all the trained topic models. For all three AOIs, the spatial blocks generated from lower-level road networks generated more semantically coherent topic models, indicating that lower-level road network such as residential roads and service roads can partition space in a more semantically coherent way. In other words, POI groups at a residential or lower-level spatial block reflect a more natural organization of human activities in space and how human use the space. For both Seoul and London, the *level 3* and *level 4* spatial block generates semantic clusters with similar coherency when the number of cluster is smaller than 35. When the number of semantic cluster is set to be greater than 40, level 4 spatial blocks generates more coherent clusters. Both Seoul and London have two peak coherence scores (one at 3∼8 clusters and another at 50∼55 clusters) for level 3 and level 4 spatial blocks. For Johannesburg, level 3 spatial block has better semantic coherence and it has peak semantic coherence around the similar number of clusters with Seoul and London. Table-1 shows the top three most coherent topics for each of the AOI for topic models trained with level 3 spatial block and 10 pre-set topic.

Three AOIs have different land use themes. For Seoul, the top three are retail/restaurant, mixed residential and commercial, and tourism/entertainment. For London, its residential, retail, transportation; For Johannesburg, its retail, industrial and leisure/recreation.

AOI	Top 3 themes
Seoul	1. payment;pharmacy;route;cuisine;cafe;retail;smoking;market_route_ref;coins;amenity_fast_food 2. building;apartments;amenity;commercial;leisure;platform;highway_bus_stop;school;park 3. motel;sauna;hotel;capacity;level;bar;사우나;leisure_fitness_centre;cinema;농협
London	1. building;residential;house;amenity;leisure;parking;amenity_post_box;platform;highway_bus_stop;bench 2. shop;retail;restaurant;convenience;cafe;open_data;level;drive_through;hairdresser;church 3. route_bicycle;network_lcn;colour;cycle_network_gb;layer;housenumber;london_cycle;co_uk
Johannesburg	1. shop;parking;restaurant;surface;supermarket;cuisine;amenity_fast_food;convenience;amenity;pharmacy 2. railway;industrial;gauge;route;rail;electrified;service;bus;platform;contact_line_frequency 3. leisure;park;pitch;sport;leisure_swimming_pool;tennis;playground;soccer;leisure_sports_centre;golf_course

Table 1 Topic model results with highest semantic coherence scores.

4 Conclusion

Accurate and precise land use maps are essential to modeling finer-resolution population distribution, measuring the growth of urban sprawl, and studying the societal and cultural phenomenon. This concept paper introduced a globally applicable multi-scale land-use modeling technique based on the spatial and semantic dimension of POI data. Also, we demonstrated the selection of the most appropriate spatial scale and semantic granularity for land use modeling on three diverse and geographically apart regions of the world. In the future, we plan to introduce temporality, socio-economic, and other factors to generate vibrant land use maps at planet-scale and develop universal benchmarks to validate land use maps.

References

1 Renato et. al. Andrade. POI Mining for Land Use Classification: A Case Study. *ISPRS International Journal of Geo-Information*, 9(9):493, aug 2020. doi:10.3390/ijgi9090493.

2 J. Fan and K. Stewart. Detecting Spatial Patterns of Natural Hazards From the Wikipedia Knowledge Base. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-4/W2(July):87–93, 2015. doi:10.5194/isprsannals-II-4-W2-87-2015.

3 Rachel Palumbo, Laura Thompson, and Gautam Thakur. Sonet: A semantic ontological network graph for managing points of interest data heterogeneity. In *ACM GeoHumanities*, New York, NY, USA, 2019. doi:10.1145/3356991.3365474.

4 Michael Röder, Andreas Both, and Alexander Hinneburg. Exploring the space of topic coherence measures. In *Proceedings of WSDM*, 2015. doi:10.1145/2684822.2685324.

5 Gautam S. Thakur, Budhendra L. Bhaduri, Jesse O. Piburn, Kelly M. Sims, Robert N. Stewart, and Marie L. Urban. Planetsense: A real-time streaming and spatio-temporal analytics platform for gathering geo-spatial intelligence from open source data. In *ACM SigSpatial*, New York, NY, USA, 2015. doi:10.1145/2820783.2820882.

6 Yao Yao, Xia Li, Xiaoping Liu, Penghua Liu, Zhaotang Liang, Jinbao Zhang, and Ke Mai. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *IJGIS*, 31(4):825–848, 2017. doi:10.1080/13658816.2016.1244608.