Cross-scale Spatial Enrichment of Trajectories for Speeding Up Similarity Computing

Cheng Fu, Robert Weibel

cheng.fu@geo.uzh.ch, robert.weibel@geo.uzh.ch

University of Zurich

Abstract. Different types of cross-scale analytics have been applied to spatial enrichment and aggregation of trajectories using geographical context sources as both subjects can present different spatial patterns at different scales. This paper clarifies the taxonomy of different types of "cross scale". A conceptual framework is then proposed on summarizing the key components for spatial enrichment of trajectories supporting different cross-scale types. Following a workflow guided by the proposed framework, POIs are used for enrichment of GPS waypoints, in a proof-of-concept case study. The preservation of pairwise trajectory similarity between the raw trajectory and the enriched trajectories is investigated. Empirical results show a good preservation while the time on computing the distance/similarity matrix is significantly reduced. This shows the potential for applications relying on an efficient trajectory clustering strategy.

Keywords. Cross-scale, Spatial Enrichment, Trajectory Modeling

Published in "Adjunct Proceedings of the 15th International Conference on Location Based Services (LBS 2019)", edited by Georg Gartner and Haosheng Huang, LBS 2019, 11–13 November 2019, Vienna, Austria.

This contribution underwent double-blind peer review based on the paper. https://doi.org/10.34726/lbs2019.65

1. Introduction

Spatial enrichment integrates geographical context and spatial semantics into trajectory modeling, which can simplify querying, analyzing, and mining trajectories. Roads, points of interest (POIs), and places extracted from the raw trajectories are common geographical context resources for trajectory enrichment. However, geographical context commonly involves the issue of *scale* (Goodchild, 2011), that is, the same geographical features may present different patterns while being aggregated at different granularities, leading to one type of modifiable areal unit problems (MAUP, Openshaw, 1977). As the term *scale* refers to different definitions, cross-scale modeling and analysis essentially also refers to different scenarios.

The question thus is how the spatial enrichment and aggregation at different scales can benefit trajectory modeling and analysis. For example, if the enrichment by an external geographical context, such as POIs, at a coarse scale can achieve a similar result as another external geographical context at a fine scale, the former resource may be selected because it may take less computing resource and time.

However, existing frameworks for spatial enrichment of trajectories, such as Yan et al., (2011) and Soares et al. (2019), do not address integrating cross-scale analytics, even though they have otherwise developed a comprehensive understanding of data sources, models, and applications.

This paper starts by clarifying different scenarios of general cross-scale modeling based on one common definition of scale. Two main research objectives are conduced:

- A conceptual framework is proposed for fitting the cross-scale modeling scenarios into spatial enrichment of trajectory analysis.
- Using the framework, a proof-of-concept case study empirically shows that the cross-scale approach can benefit trajectory similarity computing by saving computing time.

2. A Theory of Cross-scale Analysis

2.1. Scale as Spatial Granularity

Based on the taxonomy by Atkinson & Tate (2000), this paper uses the definition of scale as the *spatial extent* rather than the amount of detail. The term further narrows to the aspect of *spatial granularity*, rather than the scales of spatial variance.

2.2. Three Scenarios of Cross-scale

The granularity of spatial subdivision units for spatial analysis can be further categorized into *fixed granularity* and *variable granularity*. Subdivision units with fixed granularity have the same or similar shape and areal size, such as grid tessellations, image pixels, etc. Subdivision units with variable granularity have significantly different areal size from each other. Examples of variable-granularly units are subdivision units by fractal geometry (Jiang & Brandt, 2016) and urban-countryside subdivisions.

The term "*cross-scale analysis*" usually refers to comparing results aggregated at different scales. Regarding the context of two types of granularity, we conceptualize cross-scale analysis as three types of comparisons: *fixed-granularity multi-pass subdivisions, variable-granularity single-pass subdivisions,* and *variable-granularity multi-pass subdivisions*.

Fixed-granularity multi-pass-subdivision modeling as cross-scale analysis (Scenario a, *Figure 1*) uses different fixed-granularity subdivision units in different modeling processes and compares the results, such as empirical studies by Lloyd (2014). Variable-granularity single-pass-subdivision modeling as cross-scale analysis compares results from groups with significantly different geographic areas within the same subdivision. For example, statistical comparisons between cities and rural areas can be categorized as one type of this cross-scale analysis. Variable-granularity multi-pass-subdivision modeling compares results from different passes of variable-granularity single-pass-subdivision modeling by aggregating subdivision units (Scenario b, *Figure 1*), such the modeling by Soleymani et al. (2014).

This paper conducts cross-scale analytics in trajectory modeling applying the variable-granularity multi-pass-subdivision scenario.

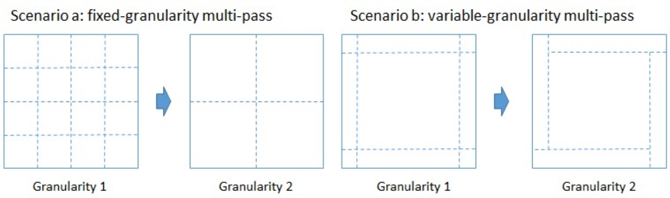

Figure 1. Theoretical models of the existing cross-scale analysis.

3. Methodology

3.1. Variable-granularity Cross-scale Subdivision for Trajectory Modeling

A GPS trajectory can be simplified by aggregating raw waypoints into less representative waypoints after being enriched by places, such as POIs or land use parcels. However, not all POIs or land uses are equally important for trajectory modeling. For example, to model the movement of a truck, small grids can be used for aggregating waypoints within the city centers while larger grids are used for modeling suburban and countryside area.

3.2. Conceptual Framework Overview

We conceptualize cross-scale enrichment as a key processing component for trajectory applications: A subdivision model is extracted from the geographical context and is applied to enrich a trajectory; Waypoints of a trajectory are then enriched by the subdivision units (*Figure 2*). The enriched waypoints can be aggregated into fewer, representative waypoints based on the enriched spatial proximity or semantic proximity for further applications, such as trajectory clustering.

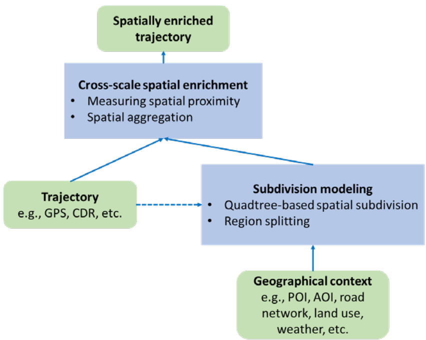

Figure 2. A framework for cross-scale-based spatial enrichment on trajectory

The quadtree scheme is selected for modeling variable-granularity subdivision from a geographical context. Commonly used for spatial indexing, a quadtree splits space into hierarchical, mutually exclusive grid meshes (*quadnodes*) at different levels. Areas with more data are split into smaller quadnodes. An existing quadtree can easily support multi-pass cross-scale analysis, by controlling the threshold of data records for splitting quadnodes into smaller quadnodes.

4. Preliminary Results for Evaluation

As a proof of concept, a pilot empirical study was conducted to illustrate the feasibility of the proposed framework, as well as to explore the tradeoff by using POIs as the geographical context to enrich and aggregate waypoints with the variable-granularity multi-pass cross-scale analysis and investigate if the pairwise similarity between trajectories can still be preserved, towards its further application to cluster big trajectory data sets in an efficient way.

500 sample truck trajectories from a Greece-based fleet management company were randomly selected. The mean trip length regarding the number of waypoints is 263 (SD = 215, MIN = 51, MAX = 1889) and the median time interval of two waypoints is 20 seconds. 4,641,857 OpenStreeMap POIs over Europe were collected, regarding the potential to extend the study for the whole truck trajectory database. Longest Common Substring (LCSS) and Dynamic Time Warping (DTW) are used as the trajectory similarity metrics. The mean center is used as the location of points with the same spatial proximity after enrichment.

Three thresholds were used in quadtree building: 1, 5, and 10 POIs per quadnode. As the benchmark, the cost for calculating 124,750 pairs of LCSS and DTW for the 500 raw trajectories was 7,748.0 seconds and 7,076.0 seconds, respectively. While the pairwise distance is well preserved, the computing time is significantly reduced (*Table 1*).

	LCSS			DTW		
Quadtree splitting threshold	Pearson's r	Spearman's rho	Time (sec.)	Pearson's r	Spearman's rho	Time (sec.)
1	0.65	0.93	878.0	0.82	0.80	1,581.0
5	0.60	0.87	58.9	0.77	0.77	254.0
10	0.56	0.85	28.8	0.79	0.81	157.0

Table 1. Correlations between the pairwise LCSS and DTW of the simplified trajectories and the raw trajectories, respectively. All correlations have a 0.00 p-values significance.

To conclude, our cross-scale waypoint enrichment and aggregation framework enables to speed up related trajectory similarity computing and thus benefit trajectory clustering computing.

References

Atkinson, P. M., & Tate, N. J. (2000). Spatial Scale Problems and Geostatistical Solutions: A

Review. *The Professional Geographer*, *52*(4), 607–623.

Goodchild, M. F. (2011). Scale in GIS: An overview. *Geomorphology*, *130*(1–2), 5–9.

Jiang, B., & Brandt, S. A. (2016). A Fractal Perspective on Scale in Geography. *ISPRS International Journal of Geo-Information*, *5*(6), 95.

Lloyd, C. (2014). *Exploring Spatial Scale in Geography*. Chichester, UK: Wiley.

Openshaw, S. (1977). Optimal Zoning Systems for Spatial Interaction Models. *Environment and Planning A: Economy and Space*, *9*, 169–184.

Soares, A., Rose, J., Etemad, M., Renso, C., & Matwin, S. (2019). VISTA : A visual analytics platform for semantic annotation of trajectories. In *EDBT: 22nd International Conference on Extending Database Technology* (pp. 1–5).

Soleymani, A., Cachat, J., Robinson, K., Dodge, S., Kalueff, A. V, Weibel, R., … Allan, V. (2014). Integrating cross-scale analysis in the spatial and temporal domains for classification of behavioral movement. *Journal of Spatial Information Science*, *8*(8)

Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., & Aberer, K. (2011). SeMiTri. In *Proceedings of the 14th International Conference on Extending Database Technology - EDBT/ICDT '11* (p. 259). New York, New York, USA: ACM Press.