Describing the Levels of Detail for the Analysis of Spatio-temporal Events

Ricardo Almeida Silva

Abstract Spatio-temporal events are collected at high levels of detail (LoDs) in many phenomena. Both spatial and temporal characteristics of data can be expressed at different LoDs. Depending on the level of detail, different spatio-temporal patterns can be detected, and in some specific cases spatio-temporal patterns are just detected in some LoDs [1]. It is crucial to model spatio-temporal phenomena having in mind that different LoDs can be useful. We proposed a granularity theory devised to model spatio-temporal phenomena at multiple LoDs [2]–[4]. We aim to enhance the granularity theory in order to reason with different LoDs for a specific phenomenon. The goal is to moving towards an approach capable of identifying the appropriate level(s) of detail to look for a spatio-temporal pattern.

Keywords Spatio-temporal data · Multiple levels of detail · Granularity

1. Introduction

Crimes, forest fires, accidents, respiratory infections, human interaction with mobile devices (e.g., tweets), among others, are producing numerous amounts of spatio-temporal events with high levels of detail (LoDs). By spatio-temporal event, we mean a summary of what has happened in reality: a homicide occurs in some latitude and longitude at eight o'clock resulting in two victims; a fire incident starts in a particular latitude and longitude on 4th August 2006 at 17:00 hours leading to 130 hectares of burnt forest area. By spatio-temporal events, we mean

R.A. Silva
NOVA-LINCS Lab, Universidade Nova de Lisboa, Portugal
e-mail: ricardofcsasilva@gmail.com

Copyright (c) by the paper's authors. Copying permitted for private and academic purposes.
In: A. Comber, B. Bucher, S. Ivanovic (eds.): Proceedings of the 3rd AGILE Phd School, Champs sur Marne, France, 15-17-September-2015, published at http://ceur-ws.org

data with the following structure: (S, T, A_1, …, A_N) where S describes the location of the event, T specifies the time moment, and A1, …, AN are attributes detailing what has happened.

Looking at spatio-temporal events, both spatial and temporal components of data can be expressed at different LoDs that can range, for instance, from grids with different cell sizes to cities, countries; from seconds to months or years [3], [5]. The LoD reflects the size of the units in which phenomena are observed and often aggregated/summarized [5], likely changing our understanding of them. Consequently, different spatio-temporal patterns can be identified at different LoDs and some spatio-temporal patterns can be just detected in some them [1], [4]. It is crucial to model spatio-temporal phenomena having in mind that different LoDs can be useful.

A granularity theory devised to model spatio-temporal phenomena at multiple LoDs was proposed [2]–[4]. This theory provides the necessary formalism to look at a phenomenon at different LoDs. More particularly, it defines the concept of predicate, level of detail of predicate and a relationship between levels of detail called *is more detailed than*. Based on these concepts, we have a phenomenon representation for each LoD.

The granularity theory allow to conduct analyses in multiple phenomena's LoDs. However, the recent analytical approaches work on a single user-driven LoD (e.g., [6]–[10]), leaving the user with the difficult task of determining which LoD is suitable to analyze the data [17]. To understand what LoD(s) would be adequate to look for a spatio-temporal pattern, users often have to use "trial-and-error" approaches. The identification of the right LoDs is an open issue [1].

2. Characterizing Phenomena's Levels of Detail

Our goal is to extend the granularity theory proposed in order to describe each phenomenon's LoD based on a wide set of descriptive statistics which must be comparable between LoDs. Statistics measures have been widely used in many contexts with different purposes. The analysis of the wide set of statistics measures might suggest the presence or not of spatio-temporal patterns concerning a phenomenon's LoD. Let us provide some examples.

Let's assume that we are looking for spatial hotspots of crimes concerning narcotics, and their onset and/or disappearance over time. In this scenario, the average nearest neighbor index [11] (ANN) can give some hints. If ANN's value is less than one, the pattern exhibits clustering. Otherwise the trend is toward dispersion. This measure can be computed throughout time which might indicate variations between dispersed and clustered spatial distributions. Alternatively, it may reveal constant dispersed or clustered distributions.

Let's assume that we are studying how the number of victims, resulting from car accidents, is distributed in space and how this characteristic evolves throughout time. Getis-Ord General G [12] measures how concentrated the high or low

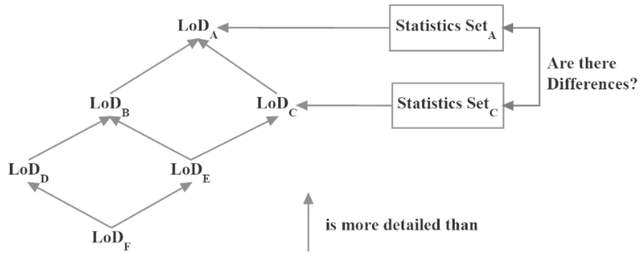

Fig. 1. A sketch of the approach proposed.

values are for a given study area. Positive scores indicates that the spatial distribution of high values is spatially clustered and the negative ones indicates the spatial distribution of low values is spatially clustered. The Getis-Ord General G measure might suggest unexpected spikes in the number of victims in a particular zone, for instance.

Space-time interaction arises when nearby cases occur at about the same time. This type of effect is very common in infectious diseases. The statistics methods like Knox [13], Mantel [14] and k-nearest neighbor test [15] measures the level of space-time interaction embedded in a phenomenon. These statistics can point to the presence of spatio-temporal clustering patterns.

As mentioned, different LoDs of phenomena may provide different perceptions. In these cases, the values of statistics will likely differ among different phenomenon's LoDs. The analysis of variations in the statistics measures of each LoD can provide the needed information to identify the proper LoDs to look for spatial or spatio-temporal patterns as sketched in Fig 1.

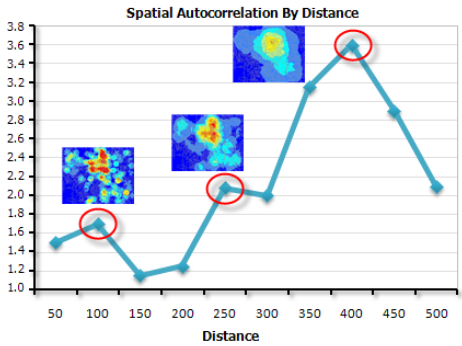

Fig. 2. Global Moran's I peaks reflect distances (i.e, LoDs) where the spatial processes promoting clustering are most pronounced [16].

For example, Global Moran's I [17] measures the spatial autocorrelation based on feature locations and an associated attribute. When the spatial distribution of high values and/or low values in the phenomena is more spatially clustered than would be expected if underlying spatial processes were random, the Global Moran's I value will be positive. In many density tools, a distance needs to be specified like happens with Global Moran's I. The distance you select implies the LoD of analysis.

ArcGIS[1] provides the *Incremental Spatial Autocorrelation* tool which applies the Global Moran's I for a series of a distances (i.e., different LoDs). Significant peak values suggest the LoDs where spatial processes promoting clustering are most pronounced, and therefore, the LoDs more appropriate for investigating hotspots (see Fig 2).

In short, we propose to extend the granularity theory in order to describe and reason about each phenomenon's LoD. Ultimately, our goal is to moving towards a systematic approach capable of identifying the appropriate level(s) of detail to look for a spatio-temporal pattern.

References

[1] G. Andrienko, N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S. I. Fabrikant, M. Jern, M.-J. Kraak, H. Schumann, and C. Tominski, "Space, time and visual analytics," *Int. J. Geogr. Inf. Sci.*, vol. 24, no. 10, pp. 1577–1600, Oct. 2010.

[2] J. M. Pires, R. A. Silva, and M. Y. Santos, "Reasoning about Space and Time: Moving towards a Theory of Granularities," in *Computational Science and Its Applications--ICCSA 2014*, Springer, 2014, pp. 328–343.

[3] R. A. Silva, J. M. Pires, and M. Y. Santos, "A granularity theory for modelling spatio-temporal phenomena at multiple levels of detail," *Int. J. Bus. Intell. Data Min.*, vol. 10, no. 1, p. 33, 2015.

[4] R. A. Silva, J. M. Pires, M. Y. Santos, and R. Leal, "Aggregating Spatio-temporal Phenomena at Multiple Levels of Detail," in *{AGILE} 2015*, Springer Science \mathplus Business Media, 2015, pp. 291–308.

[5] L. Lins, J. T. Klosowski, and C. Scheidegger, "Nanocubes for real-time exploration of spatiotemporal datasets," *Vis. Comput. Graph. IEEE Trans.*, vol. 19, no. 12, pp. 2456–2465, 2013.

[6] R. Silva, J. Moura-Pires, and M. Y. Santos, "Spatial Clustering in SOLAP Systems to Enhance Map Visualization," *Int. J. Data Warehous. Min.*, vol. 8, no. 2, pp. 23–43, Jan. 2012.

[7] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, and D. S. Ebert, "A visual analytics approach to understanding spatiotemporal hotspots," *Vis. Comput. Graph. IEEE Trans.*, vol. 16, no. 2, pp. 205–220, 2010.

[8] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips," *Vis. Comput. Graph. IEEE Trans.*, vol. 19, no. 12, pp. 2149–2158, 2013.

[1]http://www.arcgis.com/features/

[9] T. von Landesberger, S. Bremm, N. Andrienko, G. Andrienko, and M. Tekusova, “Visual analytics methods for categoric spatio-temporal data,” in *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, 2012, pp. 183–192.

[10] A. M. MacEachren, A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford, “Senseplace2: Geotwitter analytics support for situational awareness,” in *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, 2011, pp. 181–190.

[11] D. Ebdon, *Statistics in geography*. 1985.

[12] A. GETIS, “The Analysis of Spatial Association by Use of Distance Statistics,” *Geogr. Anal.*, vol. 24, no. 3, pp. 189–206, 1992.

[13] E. G. Knox and M. S. Bartlett, “The detection of space-time interactions,” *Appl. Stat.*, pp. 25–30, 1964.

[14] N. Mantel, “The detection of disease clustering and a generalized regression approach,” *Cancer Res.*, vol. 27, no. 2 Part 1, pp. 209–220, 1967.

[15] G. M. Jacquez, “A k nearest neighbour test for space--time interaction,” *Stat. Med.*, vol. 15, no. 18, pp. 1935–1949, 1996.

[16] ArcGIS, “Incremental Spatial Autocorrelation.” [Online]. Available: http://resources.arcgis.com/en/help/main/10.1/index.html#/Incremental_Spatial_Autocorrelation/005p0000004z000000/.

[17] P. A. P. Moran, “Notes on continuous stochastic phenomena,” *Biometrika*, pp. 17–23, 1950.