

GeoHighlight: Interactive Guidance-based Visualization for Spatiotemporal Data

Behrooz Omidvar-Tehrani
Department of Computer Science
The Ohio State University
omidvar-tehrani.1@osu.edu

Plácido A. Souza Neto, Gustavo Guerino
Federal Institute of Rio Grande do Norte
IFRN, Brazil
placido.neto@ifrn.edu.br, gustavo.guerino@academico.ifrn.edu.br

Abstract—Spatiotemporal data is becoming increasingly available in various domains such as transportation and social science. Discovering patterns and trends in this data provides improved insights for planning and decision making for smart city management, disaster management and other applications. However, exploratory analysis of such data is a challenge due to its huge size and diversity of spatiotemporal data. It is often unclear for the analyst *what to see next* during an analysis process, i.e., lack of guidance. In this paper, we introduce GEOHIGHLIGHT, an efficient interactive guidance framework for visualizing spatiotemporal data. We provide several scenarios that illustrate the usability of GEOHIGHLIGHT in areas such as urban planning, marketing and aviation.

I. INTRODUCTION

Nowadays, there exists huge amounts of spatiotemporal data in various fields of science. Understanding patterns and trends through visualizing spatiotemporal data improves user planning and decision making. Some instance applications of spatiotemporal data are smart city management, disaster management and autonomous transport. Traditionally, an exploratory analysis scenario on spatiotemporal data is described as follows: the analyst visualizes the data using an off-the-shelf product (e.g., Tableau¹, Spotfire²). Then she looks at different parts of data for interesting patterns and trends. With the growing size of spatiotemporal datasets, this classical approach is not practical anymore: geographical points are scattered everywhere and the analyst cannot effectively observe insights.

To overcome this challenge, visualization environments offer a plethora of operations to manipulate data (filter, aggregate, etc.). In practice, this duplicates the problem: the analyst is left alone in a huge space of data and operations. In an exploratory context, the principled challenge for the analyst is “*what to see next*” during the analysis process. A *guidance mechanism* is necessary to point out potential future directions of analysis.

In this paper, we argue the need for a guidance-based visualization framework for spatiotemporal data. We consider following desiderata for this framework.

Genericness. The framework component should be agnostic (making no assumption) about the dataset type, attributes and distribution.

Limited Options. The guidance approach should provide a limited set of recommendations because too many options distract the analyst. [?]

Relevance. The guidance approach should deliver results which have similar characteristics to what the analyst has already liked.

Diversity. Recommendations should represent distinct regions so that the analyst can observe different aspects of data and decide for the next analysis iteration.

Interactivity. The exploratory nature of the analysis requires the guidance component to be involved in an interactive process. Hence the analyst can investigate and refine different aspects of spatiotemporal data in iterative steps. For being interactive, the guidance component should be efficient so that the train of thought of analyst would not be broken during the analysis process. Despite progress in efficient spatiotemporal processing [6], sub-second interactivity is still missing.

We inspire from both recommendation [4] and visual highlighting [3], [5] methodologies and propose GEOHIGHLIGHT as a solution to aforementioned challenges. GEOHIGHLIGHT is a visualization framework for spatiotemporal data which guides the analyst throughout the process towards interesting points. Given a geographical point of interest, the question is then how to recommend other points to be considered in future analysis steps in form of guidance. In this paper, we focus on one specific guidance approach, i.e., highlighting k -best points given a point of interest. Those k points should have high quality. Quality is formulated as optimization of two dimensions: *relevance* and *diversity*. Optimizing relevance ensures that recommended points are in-line with what the analyst has already liked. Optimizing diversity results points which are as different as possible from each other and unveil different aspects of analysis. In this paper, we propose a generic interactive analysis approach for guiding analysts towards potential interesting points. The analyst considers the guidance and picks a direction for the next analysis iteration.

II. FRAMEWORK

In this paper, we address the problem of *generic guidance* in spatiotemporal data: “what is the process of guiding analysts in iterative analysis steps on any spatiotemporal dataset?” In other words, we are interested in an approach which highlights a set of k points that the analyst should consider in the next analysis iteration. This should not be a heuristic-based data-dependent highlighting, but a generic approach which

¹<http://www.tableau.com>

²<http://spotfire.tibco.com>

is applied on any spatiotemporal dataset. We describe the desiderata of generic guidance approach as follows.

GEOHIGHLIGHT operates in two steps: PREPARATION and HIGHLIGHTER. In order to speed up computing relevance in online execution, we pre-compute an inverted index for each single geographical point in \mathcal{P} in the offline PREPARATION step (as is commonly done in Web search). Each index \mathcal{L}_p for the point p stores all other points in \mathcal{P} in decreasing order of their relevance with p . Thanks to the parameter σ , we only partially materialize the indexes.

Figure XXX illustrates a screenshot of GEOHIGHLIGHT. We observe that ...

III. DEMO PLAN

Demo. What datasets do we use in our demo?

Scenario 1. Lucas example on taxis.

Scenario 2. Marketing example

Scenario 3. Aviation example

REFERENCES

- [1] A. M. Aly, W. G. Aref, and M. Ouzzani. Spatial queries with k-nearest-neighbor and relational predicates. In *SIGSPATIAL*, page 28. ACM, 2015.
- [2] J. Freire, A. Bessa, F. Chirigati, H. T. Vo, and K. Zhao. Exploring what not to clean in urban data: A study using new york city taxi trips. *IEEE Data Eng. Bull.*, 39(2):63–77, 2016.
- [3] J. Liang and M. L. Huang. Highlighting in information visualization: A survey. In *2010 14th International Conference Information Visualisation*, July 2010.
- [4] B. Omidvar-Tehrani, S. Amer-Yahia, and A. Termier. Interactive user group analysis. In *CIKM*, pages 403–412. ACM, 2015.
- [5] A. C. Robinson. Highlighting in geovisualization. *Cartography and Geographic Information Science*, 38(4):373–383, 2011.
- [6] J. Yu, J. Wu, and M. Sarwat. Geospark: A cluster computing framework for processing large-scale spatial data. In *SIGSPATIAL*, page 70. ACM, 2015.