Giovani de Almeida Valdrighi

Integrating Space and Time for Visualizing Events in a 2D plot

Rio de Janeiro, Brasil

2021

Giovani de Almeida Valdrighi

Integrating Space and Time for Visualizing Events in a 2D plot

Project of monograph presented to School of Applied Math as partial requirement to the continuity of the monograph development.

Getúlio Vargas Foundation - FGV

School of Applied Math

Applied Math Undergraduate Course

Rio de Janeiro, Brasil

2021

Giovani de Almeida Valdrighi

Integrating Space and Time for Visualizing Events in a 2D plot

Project of monograph presented to School of Applied Math as partial requirement to the continuity of the monograph development.

Approved work. Rio de Janeiro, Brasil, June 25Th of 2021:

Professor Advisor: Jorge Poco

Rio de Janeiro, Brasil
2021

Contents

Introduction

Due to technological advances, the use of sensors with geolocation has increased. They are used in smart and self-driving cars, smartphones and apps such as Uber and Waze, remote sensors, and permitted collection of information about the spatial distribution of objects. The collected data may also contain other types of information, such as scientific measures (*e.g.* wind, humidity, temperature), text, and temporal information. In this last case, they are called spatio-temporal data and are widely used. Clustering techniques, such as ST-DBSCAN, help to analyze this type of data identifying groups that can show a general overview of the most essential events present. These methods are useful in many domains, such as human mobility, healthcare, seismology, and climate science.

Although these techniques help to understand the data, since the results are at least three-dimensional (two dimensions of space and one of time), the interpretation of the clustering results is a challenging task. Visualization techniques are used to help comprehend results, allowing an intuitive and fast data analysis, however, visualization techniques developed so far do not present direct ways to represent such data, and the methods used have limitations. Thus, it becomes necessary to develop a visualization method for events (groups) of spatio-temporal data.

In this document in the first chapter will be presented literature research on space-time clustering and space-time visualizations. The next chapter presents the expected results of this work, and in the last chapter will be presented the methodology of this work development.

1 Literature research

This chapter presents the literature research on the related topics to this project. It is divided into initial research on the methods of space-time clustering, and after that, research on the methods of space-time visualization.

1.1 Clustering

Clustering techniques are a proper unsupervised learning method and are among the most used data mining methods in many areas. Different methods have already been proposed, some of them are now presented.

SNN (Shared Nearest Neighbor) is a method proposed by (ERTÖZ; STEINBACH; KUMAR, 2003, p. 10) that can compute clusters of different sizes, shapes, and densities. It starts by identifying the neighbors of each point, for each pair of neighbors, the number of neighbors that they share; this number is then used to define a measure of similarity between points. (ARYAL; WANG, 2017, p. 10) extends SNN to spatio-temporal data with the method SNN+. It proposes an adaptation of the method that considers time distance to identify the nearest neighbors.

DBSCAN, (ESTER et al., 1996, p. 10), is a clustering method that uses two parameters: a max distance between points and a minimal number of points to a set be defined as a cluster. This method starts by for each point identifying the neighbors that are closer than the max distance defined; if this set of neighbors is bigger than the minimal number of points to be classified as a cluster. The method then iterates for each one of these neighbors, now called border points, and identifies the respective neighbors until there is no point closer than the max distance. The method STDBSCAN presented by (BIRANT; KUT, 2007, p. 10) is an adaptation of DBSCAN that uses different max distances for space and time to identify neighborhoods; it can also deal with situations where points have a time duration.

1.2 Visualization

One of the most common methods for visualizing spatio-temporal data is the Space-Time Cube (BACH et al., 2014; HÄGERSTRAND, 1970, p. 10), space is represented in the width and depth of a cube, and the height represents the time. As it is a 3D object to be represented in a screen, there is the need for transformations, *e.g.*, flattening, cutting, or projection. However, it can become cluttered when the number of features to be represented grows, and its presentation on a 2D screen can cause misinterpretation of placement.

Another method for the visualization is the use of animations, the time of the data is represented by the physical time, and at each frame, a conventional visualization method for space can be used. However, animation has limitations (HARROWER, 2007; AIGNER et al., 2011, p. 10), each time frame is visualized individually, and there are cognitive limits of how many objects human can keep track of simultaneously. In this manner, animations make comparison and interpretation difficult.

Another possible alternative is the use of small multiples; different timestamps are represented individually in concatenated plots. Although, the user must define each timestamp that will be selected, and small-multiples have limitations when the number of concatenated plots increases, generating a large visualization.

It is now present works make in the visualization of spatio-temporal data that uses some of the presented methods above.

(BACH et al., 2017) present a descriptive framework for visualizations using the space-time cube with the different transformations that can be used to represent the 3D cube in a 2D screen. Some of them are cutting or flattening in any of the cube dimensions, and dynamics methods can also be incorporated, such as animated or interactive plots. (GATALSKY; ANDRIENKO; ANDRIENKO, 2004, p. 10) use interactivity with a space-time cube linked to a map view for visualizing points. The user can interact with the plot to identify distinct patterns; however, they do not use grouped data. Working with grouped data, (ANDRIENKO et al., 2013, p. 10) use transformations to represent a "group" space to analyze the inner structure of groups and plot it with a space-time cube. However, the points observed present a trajectory.

Another method that has already been used is static plots where the time is mapped to the horizontal direction, and other attributes are represented in the vertical direction or with other primitives such as color or shape. This method can show the whole period of data and do not create occlusions when projecting on the screen, although it has a limitation in the number of attributes represented.

(BUCHMÜLLER et al., 2019) works with collective spatio-temporal data; an example is data from the movement of a group of fish that is observed for a time interval. It is used projection methods in each time frame. This projection is then used to generate a 1D ordering. At each time frame, points are positioned vertically according to this ordering. The projection can preserve most of the spatial distribution, and the change between frames can be visualized.

2 Expected Results

The expected result of this project is to present a method for the visualization of grouped spatio-temporal data that represents in a static 2D plot the general distribution of groups; *i.e.*, can show the spatial and temporal size of groups, spatial the neighborhood of groups and the distribution of points of the same group while showing the data for the all time interval observed.

With the method, the object is to produce a fast and intuitive web interface that implements the presented method and its possible variations and permit different datasets.

After that, produce an evaluation of the presented method using error metrics, such as the distance and neighborhood preservation of the 2D distribution in the plot and an evaluation done by users of the web interface.

The last objective is to present case studies of the proposed method to assess its applicability with real-world data. The datasets that will be used are traffic alerts from the city of Rio de Janeiro and criminality reports from the city of São Paulo. Different datasets may also be used.

3 Methodology

3.1 Method Development

The first step of development will be a deeper search for the current methods for visualizing spatio-temporal events, or in general, visualization techniques for grouped data that have more than two dimensions.

Based on the work of Buchmüller et al. (BUCHMÜLLER et al., 2019, p. 10), where space is represented by a 1D ordering of points obtained by a projection with tree methods, or space-filling curves, a search on the methods for projection of points that belong to groups must be done. A method that tries to adapt point projections to project groups of points must be considered. One alternative is to project the mean point of the points of a group and, based on this projection, present an algorithm to plot the points of a group (not only the central one) without loss of space information.

3.2 Evaluation

To evaluate how well the method represents the original data without loss of spatial information, a metric that may be considered is the preservation of neighborhood, for each point is calculated the k nearest points in the original 2D space and in the plot, the intersection of these neighborhoods is them evaluated. If for each point, the neighborhood in 2D and in 1D is the same, then our method was able to maintain the neighborhoods fully. Different metrics for the neighborhood and distance preservation must be considered. The metrics need to be compared with the results of naive techniques to verify the improvements of our proposed method.

3.3 Web Interface

The web interface will be created using *Javascript* for the interactivity; the plot will be created using the package *D3.js* that permits the manipulation of SVG objects based on data. Suppose the use o SVG is not possible due to its limitation on the number of objects. In that case, the plot will be developed using *WebGl*, a *Javascript API* for rendering 2D and 3D images using the GPU (Graphics Processing Unit).

3.4 Case Studies

The method will then be used with real-world data to demonstrate its applicability. Two initial datasets will be used: traffic alerts and criminality reports. The first is a dataset of Waze alerts from Rio de Janeiro created by users that indicate traffic impediments, accidents, and changes. This data contains latitude, longitude, time, time duration, category, and comments by the user. The second alert is criminality reports from the city of São Paulo; they include the type of crime, the latitude, longitude, and time. Other datasets may also be considered.

References

AIGNER, Wolfgang et al. **Visualization of time-oriented data**. [S.l.]: Springer Science & Business Media, 2011.

ANDRIENKO, Natalia et al. Space Transformation for Understanding Group Movement. **IEEE transactions on visualization and computer graphics**, v. 19, p. 2169–78, Dec. 2013. DOI: 10.1109/TVCG.2013.193.

ARYAL, Amar; WANG, Sujing. Discovery of patterns in spatio-temporal data using clustering techniques. In: p. 990–995. DOI: 10.1109/ICIVC.2017.7984703.

BACH, Benjamin et al. A descriptive framework for temporal data visualizations based on generalized space-time cubes. In: WILEY ONLINE LIBRARY. COMPUTER Graphics Forum. [S.l.: s.n.], 2017. P. 36–61.

________. A review of temporal data visualizations based on space-time cube operations. In: EUROGRAPHICS conference on visualization. [S.l.: s.n.], 2014.

BIRANT, Derya; KUT, Alp. ST-DBSCAN: An algorithm for clustering spatial–temporal data. **Data & Knowledge Engineering**, v. 60, n. 1, p. 208–221, 2007. Intelligent Data Mining. ISSN 0169-023X. DOI: https://doi.org/10.1016/j.datak.2006.01.013.

BUCHMÜLLER, Juri et al. MotionRugs: Visualizing Collective Trends in Space and Time. **IEEE Transactions on Visualization and Computer Graphics**, v. 25, n. 1, p. 76–86, 2019. DOI: 10.1109/TVCG.2018.2865049.

ERTÖZ, Levent; STEINBACH, Michael; KUMAR, Vipin. Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data. In: DOI: 10.1137/1.9781611972733.5.

ESTER, Martin et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: PROCEEDINGS of the Second International Conference on Knowledge Discovery and Data Mining. Portland, Oregon: AAAI Press, 1996. (KDD'96), p. 226–231.

GATALSKY, P.; ANDRIENKO, N.; ANDRIENKO, G. Interactive analysis of event data using space-time cube. In: PROCEEDINGS. Eighth International Conference on Information Visualisation, 2004. IV 2004. [S.l.: s.n.], 2004. P. 145–152. DOI: 10.1109/IV.2004.1320137.

HÄGERSTRAND, Torsten. What about people in Regional Science? **Papers of the Regional Science Association**, v. 24, n. 1, p. 6–21, Dec. 1970. ISSN 1435-5957. DOI: 10.1007/BF01936872. Available from: <https://doi.org/10.1007/BF01936872>.

HARROWER, Mark. The Cognitive Limits of Animated Maps. **Cartographica: The International Journal for Geographic Information and Geovisualization**, v. 42, n. 4, p. 349–357, 2007. DOI: 10.3138/carto.42.4.349. eprint: https://doi.org/10.3138/carto.42.4.349. Available from: <https://doi.org/10.3138/carto.42.4.349>.