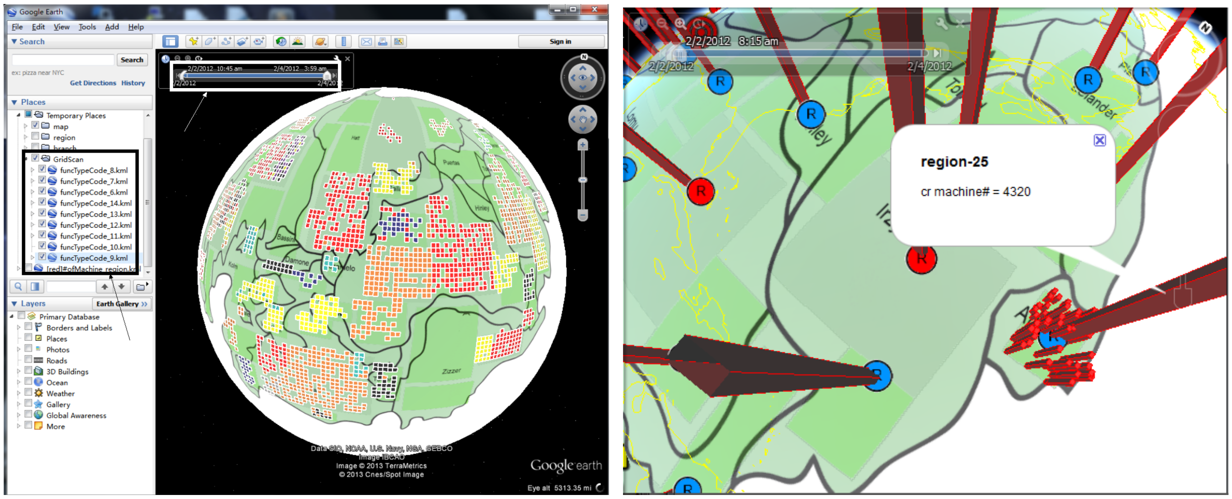
Spatiotemporal Anomaly Visualization for Large-scale Dynamic Networks

Tao Zhang[*]
Department of Computer Science
Central Michigan University

Qi Liao[†]
Department of Computer Science
Central Michigan University

Lei Shi[‡]
Institute of Software
Chinese Academy of Sciences

Weishan Dong[§]
IBM Research - China



(a) Main visualization on anomalous regions

(b) Zoomed-in view with anomaly bars

Figure 1: An overview of the interactive spatiotemporal anomaly analytic tool. In the main visualization interface, filters (black box) and time slider (white box) allow interactive exploration of the evolution of multi-dimensional attributes. Anomalous activities are visualized through 2D grids (left) and 3D bars (right) through geographic information system.

Abstract

As we move into big data era, many data grows not just in size but complexity with a rich set of attributes that contain location and time information, such as data from mobile devices (e.g., smart phones), natural disasters (e.g., earthquake and hurricane), epidemic spread, etc. We are motivated to build a visualization tool for exploring generic spatiotemporal data, i.e., records containing time location information and numeric values. Since the values evolve over time and across regions, we are particularly interested in detecting and analyzing the anomalous changes over time/space. Our visualization tool is developed based on one popular geographic information system (i.e., Google Earth) and uses a simple yet effective technique, i.e., 3D bars for representing the value dynamics. By combining data mining algorithms such as GridScan as an overlay 2D grids on top of map, the tool may guide users effectively to find potential anomalies.

1 Introduction

Data that contains location and time information exists everywhere. However, monitoring and understanding these spatiotemporal data is challenging because not only the data grows much larger in size but also more complex in nature. This is further complicated by the fact that the data values are usually very dynamic and they change not only across different areas but over time as well. It is difficult for humans to understand the dynamics and correlation of events among the time and space. While there have been data mining and machine learning approaches on spatial, temporal and spatiotemporal data [3,4], there is a gap between the data mining results and our interpretation of results, particularly in anomaly analysis and situation awareness, where users usually want a more intuitive interface to view these relationships.

On the other hand, there has been work on visualization on spatial data such as geographic visualization [2]. However, visualizing the spatial information alone does not take into consideration the causality relationship of events. As we are most interested in detecting the areas, where changes of values are abnormal compared to the past *and* neighbors, above solutions are less effective in completing the task. To that end, we develop a visual analytic system that allows users or domain experts to interactively explore spatiotemporal datasets and their anomalous changes. Our system is built on top of one popular geographic information system (GIS), i.e., Google Earth (GE) and uses a generic data format, i.e., Keyhole Markup Language (KML).

Unlike traditional geographic visualization, we introduce visual cues that can help users understand the correlation of anomalous events. In particular, we introduce a simple yet effective visual scheme, i.e., 3D bars of different colors and size for representing the value dynamics at different locations. Depending on the ways to construct the bars, one can calculate the anomalous scores that can be used to encode the bars (e.g., the higher the bar, the more anomalous that area). Colors of bars can represent different attributes/dimensions of data. Users can zoom, drag and pan, click for queries, or adjust time sliders to investigate events within a particular time window.

To bring intelligence into visualization, we allow the tool to take outputs from spatiotemporal data mining techniques, in particular detecting significant over-density and/or under-density clusters,

[*] e-mail: zhang3t@cmich.edu
[†] e-mail: liao1q@cmich.edu
[‡] e-mail: shil@ios.ac.cn
[§] e-mail: dongweis@cn.ibm.com

which are encoded as grids superimposed on the cartographic layer of Google Earth to guide users to investigate interesting areas that can potentially be anomalous. The anomaly grids and bars can be used together for better understanding of spatiotemporal anomalies. The interactive nature of the tool allows users to work on different levels of granularity.

2 Spatiotemporal Data Analysis and Visualization

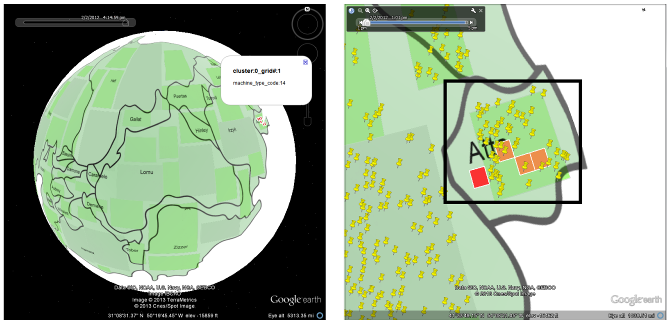
Most maps are 2D and therefore the third dimension is not utilized. To that end, we use 3D bars to take advantage of both geographic location and attribute values. Data measurements could be seen directly and accurately on their geo-positions. There can be multiple ways to construct and interpret the meanings of 3D bars. Bars indicate visualized items' three main attributes: longitude, latitude and the attribute values. The height of bars can be directly derived from the actual values of the attributes or dimensions of data. Summary statistics of those attributes, when used as bar heights, naturally represent the anomaly level. We use distinct colors for bars to distinguish different attributes in concern. Investigators could view single or multi attributes by checking the filter in Google Earth.

One limitation for understanding large-scale spatiotemporal data with 3D bars lies on degree of human perception, which is usually challenged by observing too much information concurrently. There must be a starting place for humans to look at. Therefore, we bring spatiotemporal data mining and visualization together. In general, clustering serves the purpose well for detecting areas where values change significantly over time as well as over neighbors. In particular, GridScan [1] is chosen as a spatiotemporal cluster detection algorithm. Given baseline information, GridScan can be applied to detect two types of clusters indicating anomalies, i.e., under-density and over-density. The area's boundary defines the location and the extent of a cluster. In addition to locating potential anomalies, GridScan also gives statistical evidence of the detected anomalies by their p-values. If the p-value of a detected cluster is smaller than a statistical level, say, 0.05, then the cluster is regarded statistically significant, which means that the anomaly observed is an unusual event that can barely happen. When applying to spatiotemporal datasets, GridScan can be used for detecting data changes over time. By comparing two adjacent temporally aggregated data slices D_t and D_{t+1} and treating D_t as baseline, GridScan can detect clusters in D_{t+1}, which reflects the significant changes between time t and $t+1$.

We use 2D colored grids to represent interesting zones detected by the GridScan algorithm described above. Like 3D anomaly bars, 2D grids are also intuitive to use for ordinary users, who can instantly visualize the regions that they should look at. Another benefit is that by viewing the grids' distribution, users may quickly understand the changing *trend* by considering another important data dimension, temporal dynamics. In order to show the overall trend and dynamics of value in one static view in addition to animation, we apply color encoding algorithms to represent 3D bars and 2D grids, i.e., $Color_i = \dfrac{Color_S \cdot (n-1-i) + Color_E \cdot i}{n-1}$,where $Color = \{R, G, B\}$ is a 3-tuple containing values of *red*, *green* and *blue*. $Color_S$ and $Color_E$ represent the *start* time and *end* time respectively.

3 Preliminary Results

We evaluate our visualization design on spatiotemporal anomalies by performing visual analytics on a publicly available dataset, i.e., VAST 2012 mini-challenge 1 data, which embodies a large-scale enterprise network with network traffic log and geographic information. The log data come from locations all over fictitious Bank of Money (BoM) facilities that contain close to one million IP addresses. The main goal of our visualization solution is to create large-scale cyber situation awareness.



(a) atm (b) area 'atta'_detail

Figure 2: Under-density / over-density cluster distribution of the number of online machines.

Figure 1a shows an overview of the data. Selecting KML files together may make a visualization combined with both regions and branches. By dragging *start* and *end* icons of the time slider, the view is refreshed to fit the specified time window. The small colorful grids are clusters resulting from the GridScan algorithm. The different colors represent the different time slices. After narrowing down by selecting each one function type, we could find interesting distribution differences between the 'workstation', 'server' and 'atm' type of machines. One common characteristic of these types of machines is the under-density cluster in an interesting area: 'region-25', as shown in the Figure 2.

For the next level, we apply the second visual analytic method with bars of three dimensions on the interesting area detected above. Figure 1b shows that with the significantly different tendency in the interesting area 'region-25', an anomaly is observed. In the graph, the bigger red bar contains the aggregation result for each region. The red bars' heights are directly related to the branches' numbers of changes in online machines. This could be caused by some geological incidents that might happen in that area, disasters such as hurricane or earthquake.

4 Conclusion

Data with additional dimensions such as time and space has grown increasingly complex. In order to understand and analyze the abnormal change patterns, we develop an interactive visual analytic tool based on Google Earth which adopts overview+detail investigation flow through using multiple-level 3D anomaly bars and 2D anomaly grids. The magnitude and granularity of bars/grids adjust dynamically based on zoom level while color spectrums transit based on selected time windows. The overview presentation may help investigators to quickly detect trend and pattern evolution over time and regions. Therefore, the tool is useful for situation awareness and investigation tasks in which administrators need to quickly identify spatiotemporal anomalies in interesting regions.

References

[1] W. Dong, X. Zhang, L. Li, C. Sun, L. Shi, and W. Sun. Detecting irregularly shaped significant spatial and spatio-temporal clusters. In *SDM*, pages 732–743. SIAM / Omnipress, 2012.

[2] M. Nöllenburg. Geographic visualization. *Human-Centered Visualization Environments 2006*, 4417:257–294, 2006.

[3] K. Rao, A.Govardhan, and K. Rao. Spatiotemporal data mining: Issues, tasks and applications. *International Journal of Computer Science and Engineering Survey*, 3:39–52, 2012.

[4] J. F. Roddick and M. Spiliopoulou. A bibliography of temporal, spatial and spatio-temporal data mining research. *SIGKDD Explor. Newsl.*, 1(1):34–38, June 1999.