MTDES: Multi-dimensional Temporal Data Exploration System; Strong Support for Exploratory Analysis Award in VAST 2018, Mini-Challenge 2

Vung V. Pham*
Computer Science Department, Texas Tech University

Tommy Dang†
Computer Science Department, Texas Tech University

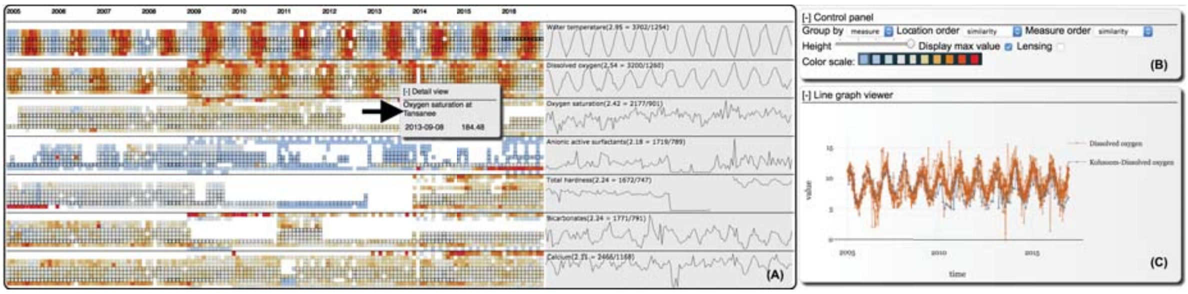

Figure 1: Application of the *MTDES* on the *VAST 2018 - MC2* dataset. The main view is a heat-map (A) to discover general pattern. A cell of the heat-map represents several aspects of the data (color for the value, position for the time, border thickness for the number of samples). A control panel (B) to set the view with various aggregation and sorting options. A Line graph viewer (C) to show the details distribution of values over time. Interactive features are also added to assist the analysis such as drag-drop between views, zooming into specific region, clicking on a individual cell to view individual values (at the black arrow).

Abstract

This work proposes a visual analytic solution which is well-designed to provide investigative functions with fluent interactions to analyze multi-dimensional temporal data. The solution allows users to view different dimensions of the data at different levels of details with a well-designed mixture of different visualizations and smooth interactions. At the general/overview level, various aggregation strategies are used to reduce data to be visualized, and different sorting procedures are used to cluster correlated data together to help discover patterns. Detail views are provided to explore and confirm/reject the identified patterns. Interaction and smooth transition between views are implemented to enable natural actions while performing analysis tasks. This work also presents the result of applying the solution to the *VAST 2018 - Mini-Challenge (MC) 2* dataset, which led to the *Strong Support for Exploratory Analysis* award for the challenge.

Index Terms: Multi-dimensional, temporal data analysis—Heat-map—Line-graph—Interaction Design; VAST Challenge 2018—Mini-Challenge 2—Award: Strong Support for Exploratory Analysis—

1 Introduction

Our solution is based on an exploration work-flow as 1) finding general patterns at a broader view by using different aggregating and sorting strategies, then 2) exploring and confirming (or rejecting) the patterns by checking at the finer level of details. Specifically, in this case, the user first finds a pattern using a heat-map with different aggregations (by month, by location, by measure) and sorting (by correlation/similarity) strategies. The user then could focus on analyzing the specific (smaller) group of the related data at the details views (line-graphs) to either confirm or reject the pattern.

*e-mail: vung.pham@ttu.edu
†e-mail: tommy.dang@ttu.edu

2 Design decisions

Line-graphs and heat-maps are commonly used to visualize temporal data [2]. Line-graphs are good at showing trends of the temporal data, however, using line-graphs to envision big datasets may lead to cluttering due to a large number of lines used. Heat-maps are useful to visualize temporal data without cluttering. However, it requires large display spaces for large datasets. Therefore, aggregations are used to reduce the amount of data to be displayed by the heat-map. This combination provides uncluttered, general overviews of the large data set. Also, different sorting strategies are used to cluster correlated elements to help discover patterns. Once the overall view highlights a pattern, the user can explore a smaller group of the data (the pattern) in details, in this case, line-graphs are well fit to show the trends and would not have cluttering issue since the data group to study is smaller.

There are many approaches to deal with the high number of dimensions of the data. They could be classified into two broad categories as either reducing the number of dimensions or representing several dimensions in one individual unit of space. Strategies of the first category include using traditional statistics or advanced techniques such as *Scagnostics* [4] to reduce the dimensions, one example of such approaches is the work presented by *Timeseer* [1]. For the latter category, the approaches tend to use several ways (color, size, width, glyphs) to represent different dimensions in an individual, available space. This work falls to the second category, for instance, an individual heat-map cell could represent several aspects in the data such as using glyphs (rectangles for normal values and circles for extremely high values), the color of the cell expresses its value, the thickness of the border represents the number of data items enclosed within that single cell, position of the cell could also represent aspects of the data such as left-right position serves its time information, similar approach could also be found from *MalViz* [3].

3 System model

Figure 1 shows the main components of our system designed for the *VAST 2018 MC2* dataset[1] with a heat-map, a control panel, and a line-graph viewer.

[1]http://vacommunity.org/VAST+Challenge+2018+MC2

IEEE Conference on Visual Analytics
Science and Technology (VAST)
21-26 October 2018, Berlin, Germany
978-1-5386-6861-0/18/$31.00 ©2018 IEEE

The first visualization is a heat-map (panel (A)) which is defined as a grid of cells. A row represents a measure at a location over time. A cell is a representation of the measured values in a particular month for a specific place. There are several individual extremely large values but hard to confirm if they are outliers, so circles are used to represent these values. The color of a cell represents the averaged measured value with the color scale shown in panel (B). The thickness of the border of a cell represents sampling frequency (number of samples per element per month per location), the thicker the border is, the higher the frequency.

There is a control panel in the panel (B) to set the aggregation and sorting strategies. For instance, user could group the rows by location or measure, user could order the locations (alphabetical, similarity, down-stream, distance from the dumping place, or sampling frequency), and/or order the measures (alphabetical, similarity, or sampling frequency).

There are also some other floating panels to assist the analysis at the further levels of details, such as a Line-graph viewer and Detail view. The Line graph viewer (C) on which user could drag and drop one or more rows to show the distribution of measure(s) over time for further investigation when needed. It also allows user to zoom into a specific region of interest to investigate at the finer level of details to explore and confirm or reject the pattern. The Detail view, at the black arrow in the panel (A), allows the user to drill down to individual data item.

4 Implementation

The source codes, the demo page, and the report of the patterns discovered for the VAST 2018 - MC2 dataset can be found from the Github page[2] of this project. Section 5 describes the steps to explore the data and two of the discovered patterns for this dataset.

5 Analyzing VAST 2018 - MC2 dataset

The first pattern that the general view (heat-map) noticeably highlights as shown in Figure 2 (a) is the sudden drop of Chlorodine in 2015. This behavior is somewhat consistent among all locations and could further be confirmed by dragging an dropping them to the Line-graph viewer (b).

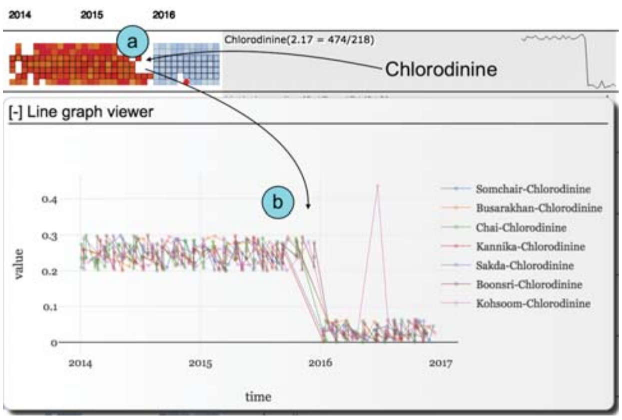

Figure 2: The drop of Chlorodine at Kohsoom in 2015

As shown in Figure 3, the *MTDES* highlights an anomaly for Magnesium with a region contains many red cells from Feb 2015 to May 2015 (a) and suggests a similar pattern as missing data or low values in the same period for Calcium (b). Further exploration, by dragging and dropping these two groups into the Line-graph viewer and zooming into the period, confirmed the misreport of Magnesium as Calcium (a1) and Calcium as Magnesium (b1).

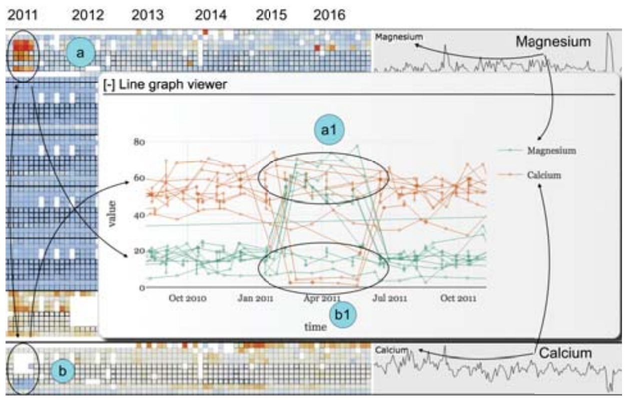

Figure 3: The misreport of Megesium as Calcium and vice versa.

While the first pattern (drop of Chlorodine) could relatively be easy to confirm by just the heat-map view, for the latter, the heat-map view only acts as a sign to suggest the pattern, and the line-graph helps to confirm this with confidence. Some may argue that we could discover the pattern using the Line-graph view only, however, in that case, there will be a long series of data over time leading to clutters and users would not know that users should zoom in the Oct 2010 to Oct 2011 (suggested by the heat-map) and have a more unobstructed view. This second pattern serves the purpose to show the complement of the two views at two exploration stages of *MTDES*.

6 Conclusions and future works

This work proposes a visualization solution to analyze multi-dimensional, temporal data with two main stages. The first is to explore the general patterns using the heat-map with different aggregation and sorting strategies. The second is to use details views (line-graph) to explore at further details to confirm/reject the identified pattern. The application of this solution to VAST 2018-MC2 dataset also confirms the design decisions as one view helps as a general view to identify the pattern while the other helps to focus into the smaller group of data with unobstructed view to confirm/reject the identified pattern. The VAST 2018-MC2 dataset also provides spatial information, but this work doesn't explore much in this direction, so in the future more work would be extended to analyze multi-dimensional, spatiotemporal data.

References

[1] T. N. Dang, A. Anand, and L. Wilkinson. Timeseer: Scagnostics for high-dimensional time series. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):470–483, March 2013. doi: 10.1109/TVCG.2012.128

[2] S. Kumatani, T. Itoh, Y. Motohashi, K. Umezu, and M. Takatsuka. Time-varying data visualization using clustered heatmap and dual scatterplots. *Proceedings of the International Conference on Information Visualisation*, 2016-Augus:63–68, 2016. doi: 10.1109/IV.2016.50

[3] V. T. Nguyen, A. S. Namin, and T. Dang. Malviz: An interactive visualization tool for tracing malware. In *Proceedings of the 27th ACM SIGSOFT International Symposium on Software Testing and Analysis*, ISSTA 2018, pp. 376–379. ACM, New York, NY, USA, 2018. doi: 10.1145/3213846.3229501

[4] L. Wilkinson, A. Anand, and R. Grossman. High-dimensional visual analytics: Interactive exploration guided by pairwise views of point distributions. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1363–1372, Nov 2006. doi: 10.1109/TVCG.2006.94

[2] https://idatavisualizationlab.github.io/VAST2018