RendezView: An Interactive Visual Mining Tool for Discerning Flock Relationships in Social Media Data

Melissa Bica
Department of Computer Science
430 UCB
University of Colorado
Boulder, CO 80309-0430
melissa.bica@colorado.edu

Kyoung-Sook Kim
Artificial Intelligence Research Center
National Institute of Advanced Industrial Science and Technology
Tsukuba, Ibaraki 305-8568, Japan
ks.kim@aist.go.jp

***Abstract*—Social media data provide insight into people's opinions, thoughts, and reactions about real-world events. However, this data is often analyzed at a shallow level with simple visual representations, making much of this insight indeterminable. Our approach to this problem was to create a framework for visual data mining that enables users to find implicit patterns and relationships within their data, focusing particularly on flock phenomena in social media. RendezView is an interactive visualization framework that consists of three visual components: a spatiotemporal 3D map, a word cloud, and a Sankey flow diagram. These components provide individual functions for data exploration and interoperate with each other based on user interaction. The current version of RendezView can represent local topics and their co-occurrence relationships from geo-tagged Twitter messages.**

Keywords—three dimensional visualization, spatiotemporal data analysis, flock phenomena, geosocial media

I. Introduction

Social media generates huge amounts of data, and researchers or organizations who analyze this data in search of social phenomena need tools to accomplish this efficiently and effectively. The purpose of this project is to create an interactive framework for visual data mining to enable finding flock patterns and relationships in complex social media data, particularly from Twitter, where "flock" refers to the clustering of similar data.

The motivation behind RendezView is to develop a web-based interface for visual data mining[3] to extend the functionality of the previous Sophy[4] framework which structures geo-tagged social media data (Figure 1). Our contributions are to create multiple, interactive data visualizations which represent geo-spatial, time, and topic data, and to improve user interactions with the data visualizations coupled with data mining processes. This makes the framework more usable and functional, allowing users to conduct more thorough data analysis.

II. Data Visualizations

The RendezView interface is centered on three types of data visualizations: a spatiotemporal 3D map, a word cloud, and a Sankey flow diagram (Figure 2).

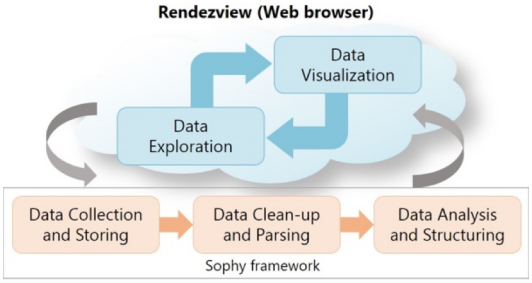

Figure 1: RendezView overview.

A. Spatiotemporal 3D Map

The 3D map is the primary visualization in the RendezView interface, implemented using Three.js[5] and GeoJSON[2] data. Geospatial data is represented in the X-Y dimensions, while temporal data is represented in the Z dimension. Each of the cubes mapped on top represents a matching row from the database. Different colored boxes correspond to each keyword the user searches for, and different shades represent the measure of popularity.

B. Word Cloud

The word cloud shows the frequency of keywords and Twitter hashtags used in conjunction with the search keyword. It updates to include the aggregate of all word frequency information when multiple boxes are selected. This is implemented using D3.js[1].

C. Sankey Flow Diagram

The Sankey diagram consists of nodes and links, where nodes are keywords and links are geospatial-temporal intersections between keywords. The width of the links shows the flow quantity, which is the measure of the intersection of the boxes in the 3D map. This is based on a D3 Sankey plugin.

III. Interactive Data Visualizations

Visual data mining is the process of detecting patterns within big data using visualizations. The value of RendezView comes from its ability to support *interactive* visual data mining. As shown in Figure 3, the visualizations are not only user interactive individually, but interoperate with each other. The filtered data is first visualized in the 3D map. When the user selects boxes in the map, the word cloud appears and displays

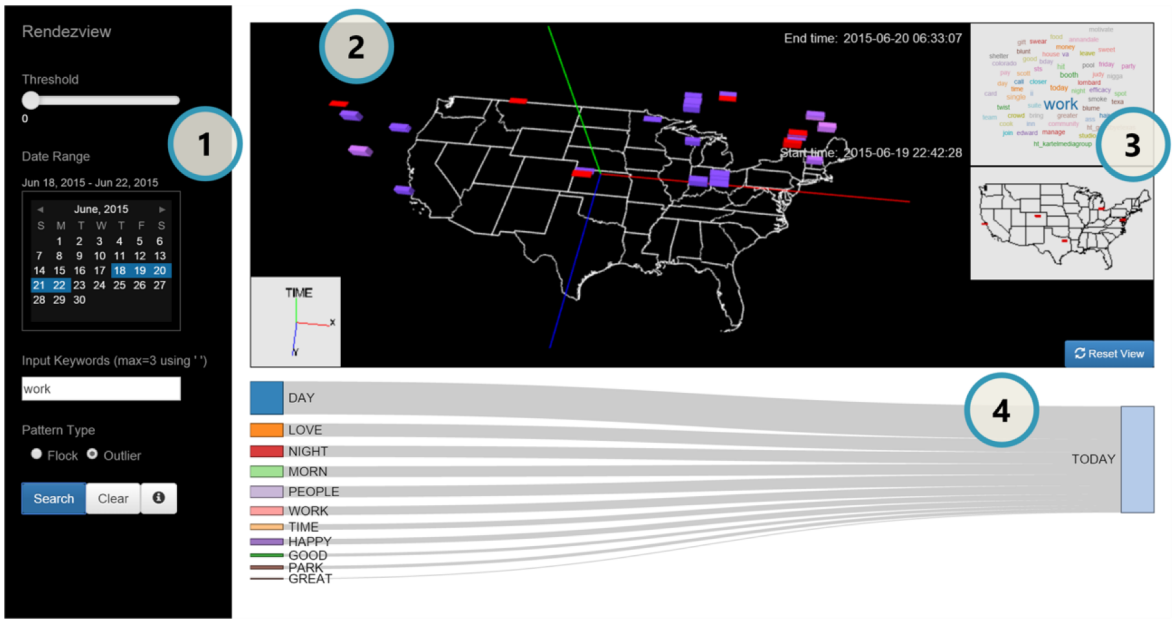

Figure 2: RendezView UI layout, design, and functionalities. The primary components of the interface are: **(1)** sidebar for user input to filter data, **(2)** spatiotemporal 3D map, **(3)** detail information which appears for selected boxes, including a word cloud and 2D projection, and **(4)** a Sankey diagram for the co-occurrence relationships.

data from the user-selected spatiotemporal area. The word cloud shows the relative frequency of words, which indicates to the user which keywords to filter the data on in subsequent searches. This also determines which keyword links the user should select from the Sankey diagram in order to display the most common keyword co-occurrences on the 3D map. These interactions allow the user to conduct a thorough analysis and find patterns hidden deep within the data.

With this process of interactive data exploration, RendezView can be used by researchers and scientists to investigate social phenomena over a geographic region, such as information flow during disasters, patterns in work and hiring, or trends in political discourse. As an example, a social scientist can use the 3D map to compare the flock pattern of job postings between the east and west coasts in the US by searching on the keyword "work." The researcher then identifies keywords in the word cloud with a similar pattern, such as "apply" and "hire." The Sankey diagram informs the next keyword search based on strong spatiotemporal co-occurrence frequency with the original keyword, and the user iterates with this process to thoroughly analyze nationwide work and hiring patterns.

IV. Conclusions

We successfully implemented an interactive data visualization framework which allows users to view implicit flock patterns and relationships in social media. Future work includes making the system dynamic by connecting the interface to a live database. Additionally, improved interaction between the Sankey diagram and 3D map will improve user interaction and depth of visualization. This includes adding an option to view link widths according to various aggregation types and updating the Sankey diagram based on selected boxes on the map.

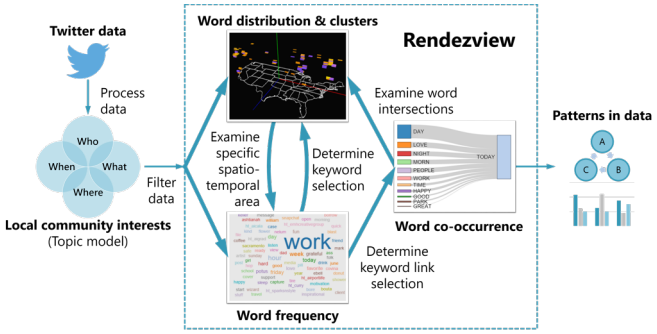

Figure 3: Interactive visual data mining process.

Acknowledgments

The authors of this project thank the Partners for International Research and Education program, the Open Science Data Cloud, the National Institute of Advanced Industrial Science and Technology, and the National Science Foundation for their support. Additionally, we thank Dr. Bob Grossman, Dr. Maria Patterson, and Dr. Jason Haga for their guidance.

References

[1] D3.js. http://d3js.org/.

[2] GeoJSON. http://geojson.org/.

[3] D. Keim et al. Information visualization and visual data mining. *Visualization and Computer Graphics, IEEE Transactions on,* 8(1):1-8, 2002.

[4] K.S. Kim, H. Ogawa, A. Nakamura, and I. Kojima. Sophy: a morphological framework for structuring geo-referenced social media. 2014.

[5] Three.js. http://threejs.org/.