

DV Personal Reflection Report - 2022

CrystalBall: A Visual Analytic System for Future Event Discovery and Analysis from Social Media Data

Tanisha Tembhre
Arizona State University
ttembhre@asu.edu

INTRODUCTION

In today's world, social media has become a platform for exchange of tremendous amounts of data and information. It is a place for people to exchange news, start discussions and express their opinions and thoughts in unprecedented ways. A lot of research has already been done on social media data to identify past and present trends as well as public opinion and sentiments on a wide variety of topics. While social media is more popular for discussions and networking, a new trend of organizing, hosting and marketing events through social media has started becoming popular as well. Examples of such events could be concerts, technical workshops, business talks, sports matches etc.

With a motivation to create a use-case for the above discussed data, I undertook implementation of the paper, CrystalBall: A Visual Analytic System for Future Event Discovery and Analysis from Social Media Data [4], as my group project for Data Visualization course as part of my MCS portfolio. The main goal of the project was to implement the CrystallBall visual analytic system which summarizes future events in an interactive dashboard. In addition to this, a meaningful extension to enhance user experience was also to be added.

This paper implementation leverages Twitter data to find all the future events, show it to the user in a meaningful way, and allow them to get an overview as well as comprehensive information about all these future events. This system can be used by city officials, event managers, marketing managers, sales people, facility owners and anybody else who wants to contribute to or attend these events. It can help them explore and be prepared for upcoming events.

DESCRIPTION OF PROJECT AND SOLUTION

To summarize the Twitter data in a visual analytic system, tweets were extracted using python library 'tweetPy' and tweets having location and date in them were filtered. Data was pre-processed using techniques such as stop word removal, lemmatization, tokenization and stemming [2]. 'Yake' python library was used to identify prominent keywords. Further, Bert model and Logistic Regression technique were used for sentiment analysis and category prediction respectively. As a result, date, location, trending keywords, category and sentiment [3] related to the event were obtained. Multiple visualization panels were then created to depict this data in a meaningful and intuitive way. The system is designed to have 6 views: Event Calendar View, Map View, Word Cloud View, Social Network

View, Tweets Panel View, and Event Category View. Each view focuses on different aspects of the event information like date, location, summary, popularity, tweets and category. The technologies used for creating the visualizations were D3.js, HTML, CSS, JavaScript, Tag Cloud and Leaflet. The interface can be seen in figure 1.

Event Calendar View

Event Calendar View shows the distribution of events across different dates. Each date has a row of circles where the circles and their colors represent events and the sentiment of public towards those events respectively. Red, yellow and green colors of the circles represent positive, neutral and negative emotions respectively. The weekend dates are marked in red to allow the user to filter out events happening on a weekend. This will be helpful as people prefer to attend events on a weekend. To show events sharing the same location, solid lines are drawn to link all such events. Similarly, events having same keywords are linked by dotted lines. The solid lines and keyword lines are colored based on the number of geo hops and keyword hops as shown in the legend at the bottom. Geo hops represent the number of links between events having the same location while keyword hops represent the number of links between events having the same keywords. This view has interactions with other views as well. Hovering over a circle or solid line shows associated location, highlights the location in map view, keywords in word cloud and events in social network view. Hovering over a dotted line shows associated keywords. Clicking on the solid lines shows related keywords in word cloud, related tweets in tweet panel and related location in map view. Clicking on date and event circles also updates the word cloud and tweet panel. This helps the user see all related information about events in detail.

Map View

This view [1] shows the distribution of events across different locations. Each location has a donut chart, which is divided into segments of different colors representing how far into the future will the event occur. The number in the centre of the donut chart represents number of events occurring on that location. Clicking on donut chart zooms in to show a radial lollipop chart to show details of individual events. This view is highly interactive with other views. Hovering over donut chart and event circles highlight the corresponding information in events calendar view, social network view and updates

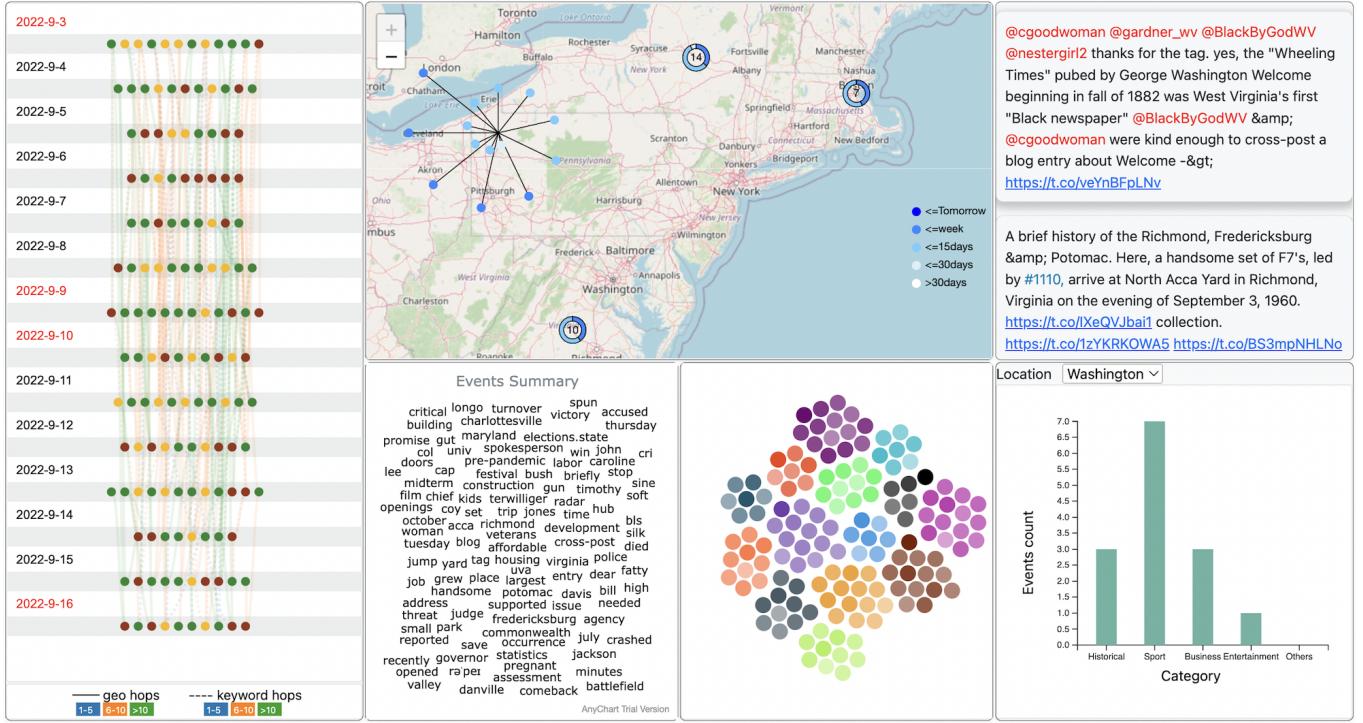


Figure 1. CrystalBall interface. The interface has 6 views: A) Event Calendar View, B) Map View, C) Word Cloud View, and D) Social Network view, E) Tweet panel View, F) Event Category view.

the word cloud. Clicking on the donut chart or event circles updates the tweet panel as well.

Word Cloud View

This view can be used to see the prominent keywords related to the events. The size of the keyword is proportional to its frequency in the tweets. This can help the user get an overview of what the events are about. Clicking on the keywords in this panel highlights the word in red and the related information in event calendar view and map view.

Social Network View

The main purpose of this view is to judge the popularity of the events. All the events occurring at the same location are clustered together and have the same color hue. The color luminescence represents the popularity of the event. Darker color luminescence implies more tweets and hence more popularity. Hovering over the events show the location and number of tweets for the event, as well as highlights related information in event calendar view, map view and word cloud view. Clicking on events updates tweet panel as well.

Tweets Panel View

This panel shows the tweets related to the events or locations selected in other panels. This could be useful to understand the context of the event and get external links containing more information about the events. This panel gets updated when events or locations are clicked in event calendar view, map view or social network view.

Event Category View

This view is the extension we implemented to the paper. It shows distribution of events across different categories. The user can choose a location from the drop-down menu to see number of events for different categories. This can help them pick out their preferred location and then explore other panels.

RESULTS DESCRIPTION

In figure 1, it can be seen that each view has different aspects of events and event related information. The user can choose any view as a starting point for exploration and navigate through other views for corresponding information.

- In figure 1, (A) Event Calendar View, different dates show different events. The use-case supported by this view is that of a user planning to attend events of interest in a limited time period.
- Figure 1, (B) Map View, supports the use-case of a user trying to plan their itinerary in a specific location.
- Figure 1, (C) Word Cloud View, supports the use-case of a user trying to locate events pertaining to a specific topic. This was confirmed by a case study where words like attack, tributes, memorial etc. directed the user to the dates near 9/11 in the event calendar view. This showed correlation between discussion of 9/11 attacks with the memorial events happening on its anniversary.
- Figure 1, (D) Social Network View, supports the use-case of a user wanting to attend the most-talked about and popular event which expects to have more footfall.

- Figure 1, (E) Tweet Panel View, provides more reading material for selected events.
- Figure 1, (F) Event Category View, supports the use-case of a user focused on attending a particular type of event. Examples include sports enthusiasts or music enthusiast can start their exploration from this view and further navigate to other views.

PERSONAL CONTRIBUTION

I created the basic structure of the entire interface using CSS grid. Then, I majorly worked on creating Event Calendar View from scratch. To achieve this, I first worked on data wrangling. From the csv files, I sorted the events according to the dates and stored them in a python dictionary. Using D3.js, I made groups of date, circle row and the gray bounding box for the circles. I then centred the circle row [5] using calculations based on the total width of the panel and the length of the circle row which was distinct for each date. Further, I coloured the circles based on the emotions associated with the corresponding events using d3.scaleOrdinal() function. I also identified the dates which were weekends and colored them red. I developed a tool-tip showing event location when user hovered over the circles as can be seen in Figure 2 (A). For circles sharing the same location, I drew curved solid lines using D3.js linkHorizontal() method [6]. Based on the lists I created of events sharing same location and events sharing same keywords, I colored the solid and dotted lines by counting their geo hops and keyword hops respectively. I then added a legend at the bottom of the panel describing the color scheme for the lines.

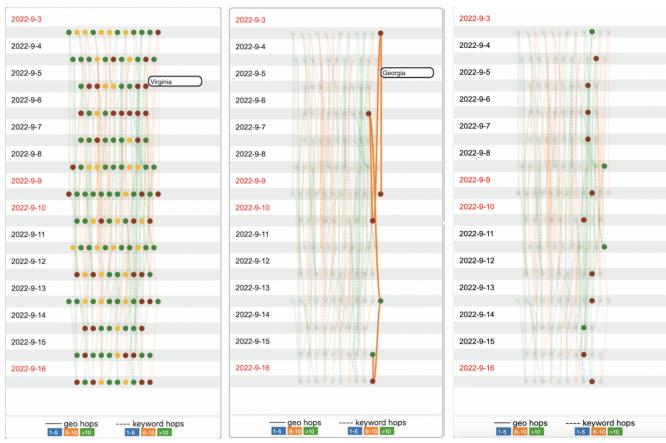


Figure 2. A) Event Calendar View, **B)** Solid lines connecting events sharing same location, **C)** Highlighted events based on interactions with other panels

Then, I worked on adding interactions to the elements. I made a function to highlight solid lines and connected events, when user hovers over them and show a tool-tip displaying the common location. This can be seen in Figure 2 (B). Additionally, I wrote multiple functions to highlight events corresponding to a particular location, event ID and keywords. These functions were integrated in other panels to establish inter-panel interaction. This can be seen in Figure 2 (C). I also integrated

functions that can update other panels when user performs actions like clicking on date, event circles or solid lines and hovering on solid or dotted lines.

CHALLENGES AND LESSONS LEARNT

- In Event Calendar View, while drawing the circle rows, bounding boxes and solid lines, I struggled with how to place different groups on top of or below each other based on the requirement. I resolved this issue by restructuring my code and creating d3 groups which need to be on top, at the end. This helped me realize the importance of understanding the DOM structure and writing well-structured code.
- This project helped me learn technical aspects of D3.js in depth, along with allowing me to hone my skills in HTML, CSS and JavaScript.
- Implementing this project helped me see the power of data visualization to effectively portray information for the target audience. I also learnt best methods and techniques to transform unstructured data into useful information using data visualization.

TEAM MEMBERS

- Bharath Kumar Bandaru - bbandaru@asu.edu
- Karthik Chowdhary Nelapati - Knelapa1@asu.edu
- Keerthana GoluSula - kgoluSul@asu.edu
- Nikhil Sagar Miriyala - nmiriya1@asu.edu
- Sandhya Balaji - sbalaj17@asu.edu
- Tanisha Tembhre - ttembhre@asu.edu

REFERENCES

- [1] Volodymyr Agafonkin. 2012. Leaflet Quick Start Guide. (2012). <https://leafletjs.com/examples/quick-start/> [Accessed: 2022-10-25].
- [2] Khoula Al-Kharusi. 2018. Semi-Supervised Document Classification. (2018). <https://www.kaggle.com/code/khoulaalkharusi/semi-supervised-document-classification> [Accessed: 2022-10-24].
- [3] Luis Bronchal. 2017. Sentiment analysis with SVM. (2017). <https://www.kaggle.com/code/lbronchal/sentiment-analysis-with-svm> [Accessed: 2022-11-03].
- [4] Isaac Cho, Ryan Wesslen, Svitlana Volkova, William Ribarsky, and Wenwen Dou. 2017. CrystalBall: A Visual Analytic System for Future Event Discovery and Analysis from Social Media Data. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 25–35. DOI: <http://dx.doi.org/10.1109/VAST.2017.8585658>
- [5] Yan Holtz. 2018. Building shapes in d3.js. (2018). <https://d3-graph-gallery.com/graph/shape.html> [Accessed: 2022-10-22].
- [6] Shivas Jayaram. 2018. linkHorizontal. (2018). <https://bl.ocks.org/shivasj/b3fb218a556bc15e36ae3152d1c7ec25> [Accessed: 2022-10-30].