CRYSTALBALL TEAM REPORT
A VISUAL ANALYTIC SYSTEM FOR FUTURE EVENT DISCOVERY

CSE 578 Course Project

Suraj Jyothi Unni
Computer Science
Arizona State University
Tempe, Arizona, USA
sjyothiu@asu.edu

Riya Salian
Computer Science
Arizona State University
Tempe, Arizona, USA
rsalian1@asu.edu

Preksha Gandhi
Computer Science
Arizona State University
Tempe, Arizona, USA
pgandhi6@asu.edu

Sangeetha Ramaswami
Computer Engineering
Arizona State University
Tempe, Arizona, USA
stramasw@asu.edu

Sai Likhith Yadav Seera
Computer Science
Arizona State University
Tempe, Arizona, USA
sseera@asu.edu

Keval Shah
Computer Science
Arizona State University
Tempe, Arizona,USA
kshah80@asu.edu

INTRODUCTION

Social media is one of the most popular platforms for people to share and communicate their ideas, opinions on different topics from the past, present and future. One of the common topics is discussing and informing others about future events. Posts are channels that can be used to share information such as the scale and spread of discussions related to various events, the lengths of discussion threads on the events, as well as people's emotions towards different events.

In recent times, there is a new form of advertisement drive among social media users. They have started using social media to organize, advertise, and inform others about future events. Many events are organized spontaneously and in an unplanned manner. This can lead to a scalability issue and many users don't know about the event. There is also another issue, few events which are organized spontaneously generally don't have a large attendance. But it can fluctuate until the event day. We can resolve these issues by gathering, analyzing, and representing information related to future events in the form of visualizations. This will empower individuals to recognize and even be prepared for the events. Consequently, If there are activities of interest taking place nearby, people may make plans or choose whether to join. To promote calm and orderly crowd gatherings or protests, authorities like police agencies or local event planners can organize and assign resources in advance.

The motivation for this paper is to create a system that can detect and characterize future occurrences from social media feeds. Characterization can be performed using terms of time, place, subject, and social network allowing for considerably more accurate detection of a future occurrence than just one or two of these factors alone. This is essential because even major and meticulously planned future actions may remain concealed in the vast flow of social media until they are released. Another source of motivation is input from authorities and future users who would want to be able to immediately identify upcoming occurrences of any protests in their area that they should be prepared for. We can also explore future occurrences that may suggest developing dangers that are of great interest to banks and other financial organizations. Finally, we have worked with businesses that are interested in a version of the streaming analysis skills mentioned here that can predict when people would be gathering in places

like malls, where timely, targeted online advertising might be beneficial. We developed CrystalBall to find different forms of future events and to let users concentrate on certain event types, including demonstrations and marches, based on our interactions with various stakeholders who have their own special demands in recognizing pertinent future events.

VISUALIZATION DESIGN

The visualization system is developed in javascript's d3.js library and the version 7 of the d3.js library is utilized. The visualization is displayed in the HTML page. To know about the events happening in future, we built a system with the following visualizations:

[D1] **Event Calendar View**

[D2] **Map View**

[D3] **Word Cloud View**

[D4] **Event List view** and

[D5]. **HeatMap View**

Event Calendar View(Fig 1): The Event Calendar View uses Twitter data to depict upcoming events by date and their relationships. The view is divided into several rows, with the date scale displayed on the left and the corresponding day on the right. Several events that occur within each row correspond to that date, which is displayed in the shape of circles. Location and keywords are used to represent the relationship between future events.

A solid red line connecting the two points denotes future events that will take place in the same place. A blue dotted line linking two points indicates upcoming events with at least eight keywords in common. When we hover over these lines, we will see future events and shared keywords within these events. When clicking on the blue dotted line, a list of hashtags between the two connecting events appear in the word cloud. Six emotions are represented by the colors in these event circles (anger in red, disgust in purple, fear in orange, joy in green, sadness in blue, and surprise in yellow).

The opacity of the circle, which goes from 0 to 1, represents the uncertainty of emotion. We would acquire event information by hovering over the circular marks (events). While clicking on the event, the Event List View(using Rank SVM model) appears for the respective event in the form of a circular radar chart with three bar charts associated with the event.

Map View (Fig 2): The map view depicts the location of the upcoming event. Each event location is represented by a donut chart, which accounts for the number of events taking place at that location. The donut chart is divided into five types for more detailed information about the event: tomorrow, <=week, <=2 weeks,< =30 days, and >30 days.These five types are colored differently. The view includes a spiral chart that loads when the event location is clicked and shows the order in which events occur at that specific location.

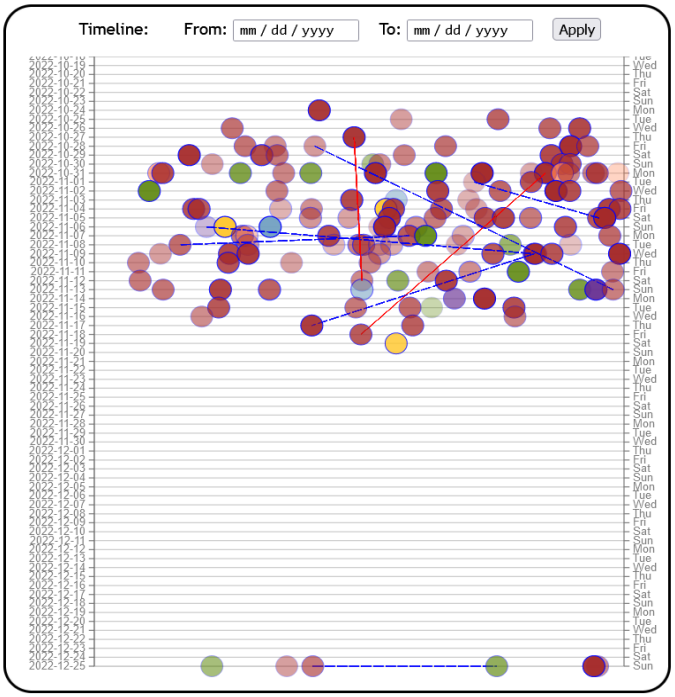

Figure 1: **Event Calendar View**

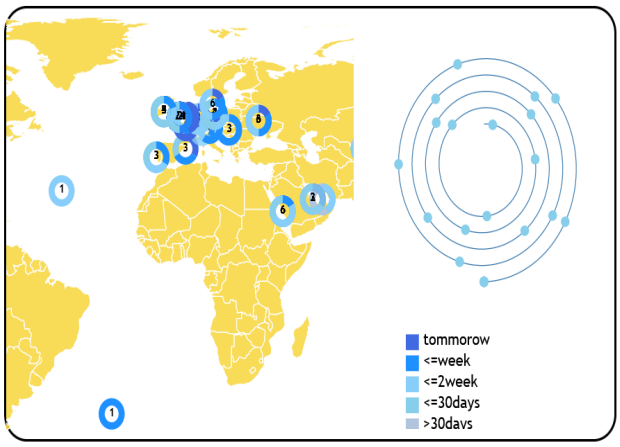

Figure 2: **Map View**

Word Cloud View (Fig 3): The word cloud view shows the keywords across all the events. The size of the words are an indicator of the frequency of those words in the tweets. On clicking on the donut chart in the map view we can see the keywords used for events in that location. When we click on the dotted line in the event calendar view, the common keywords with their relative frequency can also be analyzed from the word cloud. The word cloud layout was taken from the word cloud library [2].

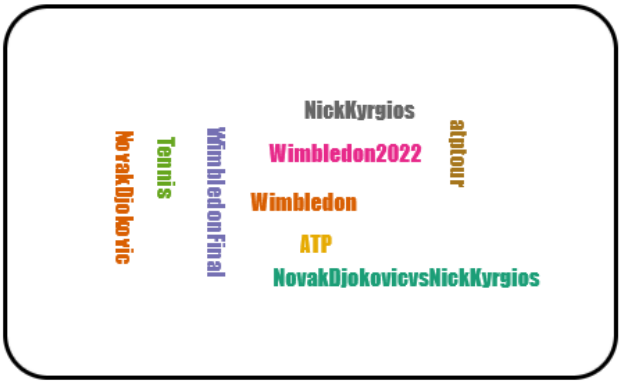

Figure 3: **Word Cloud View**

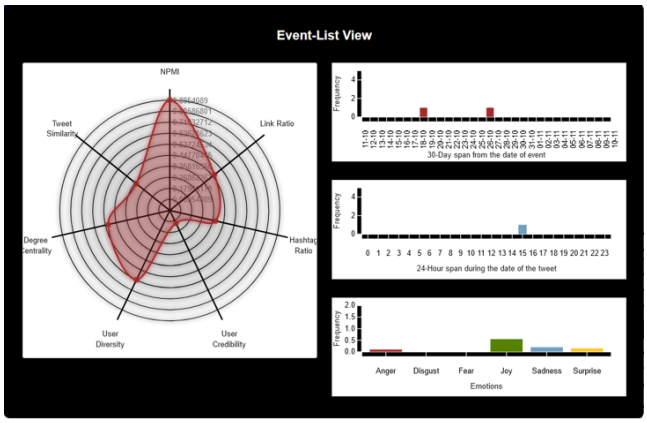

Figure 4: **Event List View**

Event List View(Fig 4): The event list view consists of two timeline bar charts, one emotion analysis bar chart and the circular radar chart.The three bar chart updates when each event is selected on the event calendar view.

1) The first timeline bar chart shows the number of tweets in a 30 day span. It provides information regarding the count of tweets on a daily basis in the last 30 days.

2) The second timeline bar chart shows the number of tweets in each hour in a day. This helps to analyze when the tweets regarding a future event were posted on a particular day.

3) The third emotion chart depicts the probability of different emotions like joy, surprise, sadness, anger, disgust, and fear. The emotions are based on the tweets of the particular event.

4) The circular radar chart is used to represent the measures of future events. These measures include Link ratio(LR), Hashtag ratio(HR), Normalized Pointwise Mutual Information(NPMI),User credibility(UC), User Diversity(UD), Degree Centrality(DC) and Tweet Similarity(TS). NPMI is the feature used for event identification. LR,HR,UC and UD are measures which provide event tweet informativeness. The event tweet cohesiveness is produced by DC and TS values. The circular radar chart is produced when an event is clicked on the event calendar view.

HeatMap View(Extension)(Fig 5): The Heat Map View is an extension of our work that uses Twitter data to compare the number of tweets for each of the eight categories for all six emotions. The color hue in the heat map cells indicates the number of tweets present for a particular group on an emotion. The higher the intensity of the hue, the more tweets would fall under that pair (category and emotion), and vice versa.

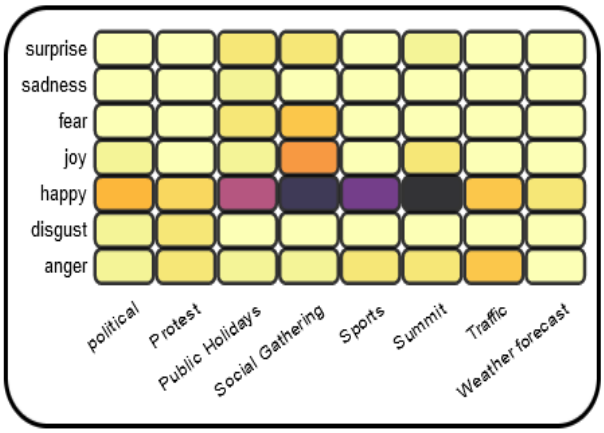

Figure 5: **Heat Map View**

DATASET DESCRIPTION

Event View data:

The dataset obtained from the Twitter API for the event calendar view includes attributes such as UserID, Event Name, Latitude and Longitude of event location, Tweet and Event Timestamps, and Hashtags. We have derived more parameters, like the category of event tweet, emotion, uncertainty, PMI, NPMI, link ratio, hashtag ratio, user credibility, user diversity, degree centrality, and tweet similarity, using Pandas and some machine learning techniques in Python.

Map View data:

The dataset for the mapview uses preprocessed data obtained from Twitter API. Attributes like location, tweet timestamp and event timestamp are retrieved from the event tweets. Parameters like latitude, longitude, uncertainty, emotions, data are attributes derived using python scripts and machine learning techniques.

Word Cloud View data:

The data used for word cloud view consisted mainly of event name, the tweet words associated with the event and the count of each unique term. The count is used to define the size of the word in the word cloud view.

Attribute types: Categorical, Quantitative

Event List view data:

- **Tweets/hour in a Day**: The count of tweets in each hour on a day the tweet was posted is calculated. This dataset is used to visualize the first bar chart which represents the tweet posting time.

  Attribute types: Sequential, Quantitative, Categorical

- **Tweets/day in the last 30 days**: The tweets/day in the last 30 days from a chosen date is represented in the second bar chart to show the distribution of tweets in the previous 30 days.

  Attribute types: Sequential, Quantitative, Categorical

- **Emotions Dataset:** The dataset has probabilities of the emotions anger, disgust, fear, joy, sadness, and surprise for each event.

  Attribute types: Sequential, Quantitative, Categorical

- **Circular radar chart dataset:** The dataset was retrieved from python. The different measures NPMI,LR,HR,UC,UD,TS and DC used in the visualization were calculated based on the ratios.

  Attribute types: Quantitative, Categorical

CASE STUDY

We'll talk about the Olympic event that was derived from the preprocessed data.

The Event View displays events based on the date they will take place.There are other Olympic events scheduled between February and November 2022, but we will focus on the competition that took place on November 13.

It is evident that the event and another one taking place on October 29 both have similar keywords and are depicted by blue dotted lines.

Circles in this event view stand in for marks and colors for channels.

Next, the map view shows where the Olympic event is taking place; in this case, Berlin, Germany. Six events are being observed in Berlin over the course of two weeks. When the donut chart is clicked, a spiral chart with information about the occurrences appears, where the events are organized by date and time.

The Word Cloud View extracts the trending keywords related to the event and for Olympics it is:

Olympics,gold,run,sports,Basketball

The new extension added by us is Heatmap View.Here the view depicts the relation of event to the emotions.The olympic event is part of Sports and in that category it is observed that happiness is the most happening emotion which signifies the majority of people are happy with the event.

The darker the color the more the score is.

Three bar graphs and a circular radar chart comprise the Event List View.

The first bar chart shows the count of tweets for each day in a 30 day span from the date of the event. There are tweets on 21st, 24th, 25th and 31st of October and on 11th November.

The second bar chart shows the count of tweets for each hour in a day. There are tweets around 12am- 1am and 10pm-11pm

The third chart shows the probability of emotions for the Olympic Event.It can be seen that Joy and Fear are the most probable emotions for the Olympic Event.

The circular radar chart shows the metrics for each event which is triggered by the onclick event in the Event View of the visualization.

The metrics for Olympic event is as follows:

Tweet Similarity:0.18

NPMI:0.62

Link Ratio:0.69

Hashtag Ratio:0.18

User Credibility: 0.62

User Diversity: 0.64

Degree Centrality:0.13

DISCUSSION

A. Lessons Learned

The project motivated the group to understand and implement five visualizations and add various interactions between them. It helped each of us to learn the process of creating a dataset using python and preprocessing data for each visualization. We learned to develop an interactive visualization which can be utilized by users to understand when a future event occurs, in which location it occurs, what category the event belongs to, how the event is discussed on social media, and what are the main features of the event. We were also able to add features regarding the tweets related to the event and views of other people about the event by considering emotions.

B. Future Work

The current crystalball system can be extended by including other innovative visualizations. A visualization to represent the weather conditions at a chosen location helps people to understand which events would be better to attend at each time of the year across different weather conditions.The system only helps to analyze data from one social media platform. Other social media platforms will also have information regarding these future events and their characteristics. This can be added to the system as an extension. The size of the data utilized can be increased which helps to analyze and visualize more features.

TEAM MEMBERS

- Suraj Jyothi Unni -1219664641
- Riya Salian - 1225625882
- Keval Shah - 1225553979
- Sai Likhith Yadav Seera - 1224791490
- Preksha Gandhi - 1219585913
- Sangeetha Ramaswami -1224264691

REFERENCES

[1] Isaac Cho,Ryan Wesslen,Svitlana Volkova, William Ribarsky, Wenwen Dou. CrystalBall: A Visual Analytic System for Future Event Discovery and Analysis from Social Media Data. In Proceedings Of IEEE Conference on Visual Analytics Science and Technology(VAST),2017.

[2] Word Cloud Generator, https://www.jasondavies.com/wordcloud/

[3] Spiral for John Hunter, http://bl.ocks.org/syntagmatic/3543186

[4] Radar Chart Redesign, http://bl.ocks.org/nbremer/21746a9668ffdf6d8242

[5] For the building blocks of Event-Calendar view, https://blog.logrocket.com/using-d3-to-create-a-calendar-app/