PEARL: An Interactive Visual Analytic Tool for Understanding Personal Emotion Style Derived from Social Media

CSE 578: Data Visualization Project

Aditi Shashank Joshi
Master's in Computer Science
Arizona State University
1222838916
ajoshi64@asu.edu

Dhaval Maheshbhai Patodiya
Master's in Computer Science
Arizona State University
1222592176
dpatodiy@asu.edu

Harsh Patel
Master's in Computer Science
Arizona State University
1222872391
hpatel66@asu.edu

Nishtha Bhimte
Master's in Computer Science
Arizona State University
1219495667
nbhimte@asu.edu

Pranshu Pandey
Master's in Computer Science
Arizona State University
1221548692
ppande19@asu.edu

Vineetha Pattapagalu
Master's in Computer Science
Arizona State University
1225432468
vpattapa@asu.edu

INTRODUCTION

In the modern era, humans can be categorized as living two types of lives- a real life and a social life. Real life is the life which we have been living since ancient times. Social life is the life which we started living after the invention of the internet on various social media platforms like Instagram, Twitter, etc. In real life, people interact with each other physically whereas in social life, we do so through means of social platforms using mediums like text, tweets, etc. Humans are also emotional beings, and we feel different types of emotions from time to time, which can be easily detected while interacting with people face to face. But it is hard to detect emotions through text. Our reference paper [1] has tried to solve this problem and come up with a visualization tool named PEARL (Personal Emotion Analysis, Reasoning, and Learning), to understand human emotions based on tweets through multiple time intervals and periods. In addition to comprehending human emotions based on tweets over various time intervals, it also helps in interacting with the person's emotion profile to understand the intricacies present in the mappings and the details. It presents linguistic evidence which helps understand why and how the data is present as shown. With the flavor of multi-dimensional analysis, it helps understand the various aspects of the text in the context of the emotions presented by the subject with PEARL. The project implements all the visualizations mentioned in the paper along with all the interactions. We also added an additional visualization to show the average emotional intensities of the users through multiple time intervals and periods.

VISUALIZATION DESIGN

1 System Design

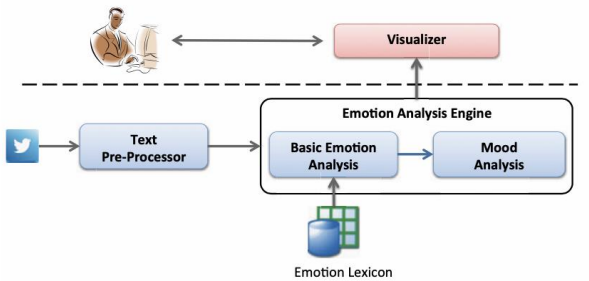

Figure 1: Overview of PEARL Architecture [1]

To perform the visualization, we extracted multidimensional emotional features from the tweets. This process consisted of 3 main parts, text pre-processing, emotion analysis, and mood analysis. It was performed beforehand to produce the data files for a selected set of users that was then fed into the visualization algorithms.

2 Graphs

There are 6 visualizations in this project that are:

1. Time Series Visualization
2. Valence Emotion Visualization
3. Tweets-Word Visualization
4. Valence-Arousal Visualization
5. Tweets Display Visualization
6. Average Emotion Visualization

2.1 Time Series Graph

The Time Series Graph gives an overview of **Valence Emotion Visualization, or the emotion profile of a user** presented as single line with the dominant emotion points shown as white dots on the line shown in Fig 2. It also gives some overview of the **Tweets-Word Visualization** by showing a smaller version in blue color as shown in Fig 2.

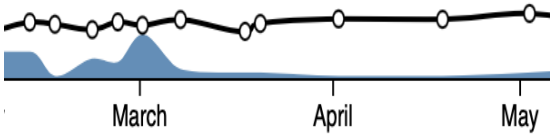

Figure 2: Time Series

2.2 Valence Emotion Visualization

This visualization represents the emotion variables encoded by an emotion band of different colors presenting emotions like anger (red), disgust (violet), joy (yellow), surprise (dark blue), anticipation (orange), fear (dark green), sadness (light blue), and trust (light green). The center position of the emotion band on y-axis represents the average valence score of all the tweets at that point in time. The white arrow and its orientation indicate the dominance and each emotion is represented by a different color. From looking at the visualization observer can know the emotion trend of the user over a particular time period and how the user can manage their emotions.

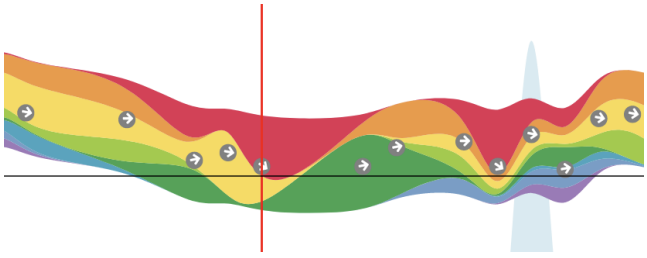

Figure 3: Valence Emotion Visualization

2.3 Tweets-Word Visualization

It is also called as emotion profile detail graph as it gives the words used by the user for the mood or emotion which is currently experienced. On hovering it shows a box as show in Figure 4 which gives a cloud view of the words.

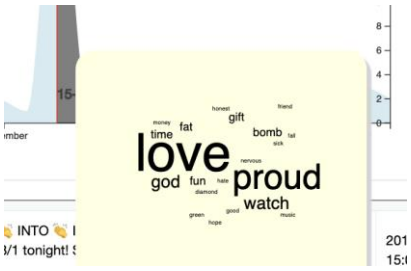

Figure 4: Tweets Word Visualization

2.4 Valence-Arousal Visualization

Also known as the mood-view graph, it gives a scatterplot of all the emotional words represented by small circles. Some words represent different emotions so the circle will be displayed in the form of a pie chart of different emotion colors. The radius of the circle represents the dominance, and the filling colors indicate the emotions as shown in Fig 5. The number associated with each point represents the mean value of arousal and valence on the X and Y axes respectively.

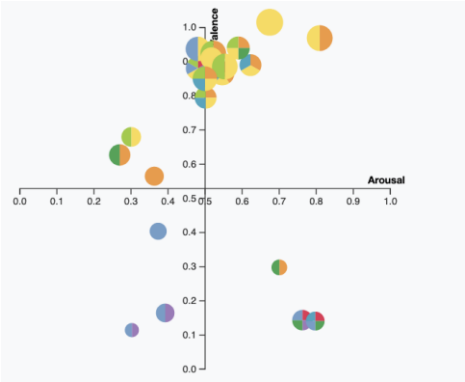

Figure 5: Valence Arousal

2.5 Tweets Display Visualization

The raw tweets of the user for the selected segment will be shown here. By hovering on the emotion bubble of the Valence-Emotion Visualization, the tweets for that emotion will be highlighted. The visualization is shown in Fig 6.

"Thank y'all SOOO MUCH for **watch**ing with me, this was magical! **love** y'all #NAVY @PrimeVideo @SavageXFenty #SavageXFentySHOW"	2019-09-23 02:33:00
"Forgot how much I **love**d this song Fab! Thank you man #SavageXFentySHOW"	2019-09-23 02:27:00
"**fat** Joooooeee!!! We **love** you!!!!! Thank you @djkhaled 🙏😩😩#SavageXFentySHOW"	2019-09-23 02:25:00
"ALEKKKKK!!! **diamond bomb** to the the kitty!!!! Yassssss #SavageXFentySHOW"	2019-09-23 02:24:00
"Also my fav **music** and choreography!! Them chicks look like the lil villains from Mario and Luigi that you gotta jump on before they get you"	2019-09-23 02:23:00
"WINDOWS was literally one of my favorite moments!!! And this soooonnnnggg! Brazilian	2019-09-23

Figure 6: Tweets Display

2.6 Average Emotion Visualization - Extension

This graph calculates the average intensity of all emotions like joy, anger, disgust, surprise, anticipation, fear, sadness, and trust for all users across the entire time period. The graph is shown in Fig 7. This helps the observer to know what the top emotions across all the users for that month or year are and can use that information for their own use.

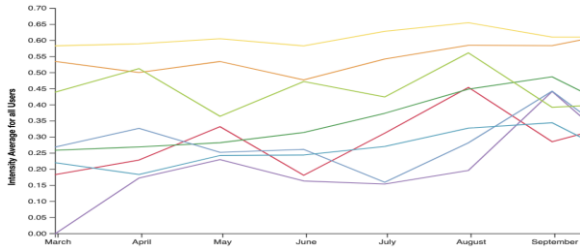

Figure 7: Average Emotions

3 Interactions

The interactions for graph are mentioned as below:

1. Observer can select the user from dropdown to see their emotion details over the time.
2. A box can be created on the **Time-Series Graph** to select the date range to show the emotions details for the user for selected date. The **Valence-Emotion Graph** and **Tweets-Word Graph** will be changed according to the selection.
3. Hovering on the streams on emotions shown in **Valence-Emotion Graph** will highlight that emotions stream in the graph.
4. Hovering on the arrows in **Valence-Emotion Graph** shows the emotions bubble of the user at that time. And hovering on the emotions bubble will display the information related to that bubble.
5. Clicking on the arrows in **Valence-Emotion Graph** will display the **Valence-Arousal Graph** showing the emotional words used by the user and it will also display the **Tweet-Display Graph** showing the tweets used by the user for that particular time frame.
6. Hovering on the word circle displayed in **Valence-Arousal Graph** will display the meta data information for that word.
7. Hovering on the **Tweets Word Graph** will display a word box showing different words used by the user in tweets.
8. Clicking on the emotion color circle displayed beside user dropdown will reduces the opacity of the stream in **Valence-emotion Graph**. Users are free to select or deselect any unwanted emotions in graph.

DATASET DESCRIPTION

The dataset consists of 5 different users whose tweet data from 2020 (and before) is obtained through a public dataset [2]. This consists of many different parameters, out of which we only use two for this project- Tweet Contents and Tweet Timestamps. Using these parameters, we derive multidimensional data for our visualizations.

We first extracted the tweet content and timestamps from the datasets. Then, we performed tokenization and stemming on these tweets to extract a list of keywords for every tweet. We used a standard NLP library in Python called NLTK [3] to perform this.

We then combined the NRC lexicon [4] and the ANEW lexicon [5] to get a new lexicon which had a list of emotional words with the specific emotion they represent on Plutchik's wheel of emotions, and their corresponding VAD score (Valence-Arousal-Dominance score). For words not common to both lexicons, we chose to exclude them from the new lexicon. Using this combined lexicon, we extracted a list of emotional words for every tweet with the emotions they belong to and their VAD scores.

Using this information, we extracted the intensity of each of the 8 emotions on Plutchik's wheel for every tweet. We did this by counting the number of words belonging to each emotion in a tweet and dividing that by the total number of emotional words in that tweet.

We also calculated the VAD score for each emotional category for the entire tweet by averaging the VAD scores

for all the words belonging to specific emotions in the tweet. So, for every tweet, we obtained 8 parameters for the intensities of each emotion, and a further 24 parameters for the average VAD scores of each emotion.

To combine this information from each tweet to extract information for a particular temporal segment, we performed constrained clustering. We clustered tweets with constraints on their temporal proximity from their timestamps. For each tweet segment, we then calculated the average intensity and average VAD score of each emotion. This again gave us 32 parameters for each tweet segment.

To calculate the trigger words for a particular segment, we calculated the tf-IDF score for each word considering one entire segment as a document. A higher score indicated a more triggering word for a segment.

Finally, for every tweet segment we obtained 32 parameters relating to emotional intensities and VAD scores, a list of trigger words, and the time period for the segment. These details were then used for visualization.

CASE STUDIES

1. Political Campaign

Elections are coming up and a political party wants to decide the candidates representing that party in those elections, decide an agenda or speech, assign some tasks to different people, and motivate their team members for work. But they have a hard time selecting an agenda for their campaign, and also a candidate who can control their emotions and take correct decisions. Our PEARL visualization tool can help them solve their problem. They can select any social media platform used by users let's say twitter and feed data to our model and the visualization will be available. To solve their problem of selecting a candidate, our tool will plot an emotion graph for different candidates for a selected time frame, and the party can go through the graph for each candidate and by viewing the visualization they can decide which users are best suited according to their requirements. The problem of selecting an agenda can also be solved with this tool. We can select a medium which is used by most of the targeted users, process the data for each user and an average emotion graph will be plotted for all users which will give us an idea about the peak emotions the targeted users are feeling and thus we can select topics that can target these emotions. They can also select the topics or words shown in the Tweets-Word Visualization and Tweets-Display Graph as well.

2. Software Company

Recently there has been a decline in the performance of the employees of the company. They want to find out the reasons behind the decline in the performance of all employees or a few targeted employees. They want to keep track of the emotions of all employees as well, so that they can improve the overall mood of the company and can come up with a method to improve their overall well-being. Our tool can help with this problem, they can provide data available from the team's channel, twitter, etc. used by employees in the company. Our tool will display the emotion graph for each employee, along with time, words, tweets, etc. Through the emotion graph of the employees and the Tweet Display graph, they can investigate the issues which caused those emotions to be present, and subsequently can try to cheer them up as well by resolving these issues. They can also see the overall emotions felt by the employees in the company through Average-Emotion Graph and can start or stop plans which may have caused them. They will also be able to see the overall effects of policies implemented in the company.

3. Other user cases

Users can use the graph in several cases which are as follows:

1. **Main Graph:**
    a. User can investigate the words which prompted the mood prediction
    b. User can see the social presence of the individual over a timeline
    c. User can see the dominant traits throughout year.
    d. A person's mood affects the overall performance of a person. This could be used to identify a person's performance for a particular job
    e. The person's mood could be an indication for the person's well-being. This could be used by the individual for self-reflection.
    f. The current mood of the person could help in better ways to interact with the person in a business setting, this could be helpful in directing the conversation which would consider the current mind frame of the customer.
2. **Extension Graph:**
    a. For the users who choose not to share their data can view the generalized advertisements presented to the audience

as per the most common emotion displayed by users in the previous time frame.

b. Politicians can use it for advertising campaigns

c. By selecting the date range from where the disasters or pandemic occurred, one can get the generalized emotion of users and try to come up with some methods to improve morals.

d. For the users who choose not to share their data can view the generalized advertisements presented to the audience as per the most common emotion displayed by users in the previous time frame.

TEAM MEMBERS

Task Distribution across team members:

1. Aditi Joshi : Ownership of Tweets Display Visualization
2. Dhaval Patodiya : Ownership of Extension Visualization, Team Report, Video, and Bug Fixes
3. Harsh Patel : Ownership of Poster and helped with Valence-Arousal Visualization
4. Nishtha Bhimte : Ownership of Valence-Emotion Visualization, Valence-Arousal Visualization, and Bug Fixes
5. Pranshu Pandey : Ownership of Data Pre-Processing, Time-Series Visualization, and Bug Fixes
6. Vineetha Pattapaglu : Ownership of Tweets-Word Visualization

ORIGINAL WORK AND MODIFICATION

The original work only produces an emotion graph for 2 users whereas we have extracted information about 5 users. We also added an average emotion graph which helps observers to see the general trends of emotions across months for all users. This may help observers to consider each user's emotion without bias for their studies. Some may also use this to get information regarding the trends of emotions across all users during stress inducing global events such as the COVID-19 pandemic.

DISCUSSION

1. Lesson Learned

We learned many things while implementing this project.

1. Teamwork: Coordination between various team members for graph and interactions requires teamwork.
2. D3: We gained more insights about the different graphs and power of d3 library for like zooming, streamgraph, area chart, bubble charts, etc.
3. Data Pre-processing: For preprocessing the data and extracting the details required for displaying the information knowledge on Python was required. We also gained knowledge about how to use popular NLP libraries in Python like NLTK.
4. JavaScript, HTML and CSS: We learned to implement the function-oriented system design in JavaScript. We created a different JS file for different graphs and integrated them in such a way that it doesn't overwrite the others graph. The project gave us an in-depth knowledge of HTML and CSS as well.
5. Agile development: The biweekly reports helped us in implementing agile methodologies, in which the scrum master wrote the biweekly report, and we prepared tasks for each team member to be completed in 2 weeks' time.
6. Soft Skills: Writing a report, preparing, and presenting the poster helped us in improving our soft skills which will help us in the future for tasks such as documenting code.

2. Further improvements

The visualization can further be improved by implementing in a real time environment where the data for the user is extracted in real time and processed and presented in the form of visualizations. The data can be further improved by using a more sophisticated model for data pre-processing. Currently a static dataset is used to perform the preprocessing. In the future, a way can be developed where we can capture human emotion in real-life interactions, which may be through text or speech. This will help in improving the visualization.

REFERENCES

[1] J. Zhao, L. Gou, F. Wang and M. Zhou, "PEARL: An interactive visual analytic tool for understanding personal emotion style derived from social media," 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), 2014, pp. 203-212, doi: 10.1109/VAST.2014.7042496.

[2] Top 50 Twitter profiles Twitter datasets

[3] NLTK

[4] Sentiment and emotion lexicons - National Research Council Canada

[5] Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings