[ECS 272] Final Project: PEARL

Vaishnavi Karanam, Kathleen Blanck

I. Project Goal

The goal of our project was to reproduce some of the key features of PEARL, a timeline-based visual analytic tool designed to explore the emotional style of an individual over time based on text analysis of their social media posts. We specifically aimed to fully implement the following components and interactive features of the presented tool:

1. Emotional profile overview: This visualization includes an area chart that represents the volume of tweets over time and a stream-graph that shows mood variations over time. A sliding window is provided to select a time segment to further explore in the detail views.
2. Emotional profile detail view: This visualization renders a zoomed-in version of the area chart and stream-graph in the overview based on the selected time range. The bands of the stream-graph represent Pluchick's basic emotions. Hovering over a certain point renders a bubble chart that captures the proportions of each emotion at that time point; the white arrow within the bubbles represents the dominance of the emotions. Clicking on a certain point opens a view displaying the raw tweets at that time point.
3. Raw tweets view: This visualization renders the text and timestamp of all the tweets posted on the selected time point (from the detail view).
4. Trigger word view: The paper uses a word cloud to display the emotional words within tweet segments that lead to their derived moods. Since word clouds tend to look very cluttered, we decided to use an ordered list (ordered by tf-idf scores) to showcase the trigger words of a selected mood segment. We also decided to color the words based on the emotions they represent to make it easier for users to form a connection between text and emotion.
5. Interactive Legend: This feature allows users to highlight individual emotion bands on the stream graph by selecting the emotion boxes. We decided to also highlight the trigger words corresponding to the selected emotion box.

II. Coding Experience

A. Text Preprocessing

I decided to use python's NLTK library so it was fairy straightforward to clean, tokenize, and stem the tweets. It was also not too difficult to derive emotion matrices for the tweets using the NRC lexicons. The most challenging part of the preprocessing stage was to cluster the tweets into emotionally, temporally and semantically cohesive segments; We had to make a lot of assumptions (that will be mentioned in section 4.1) as the paper does not provide enough details about their technique.

B. Visualizing

The visualizations implemented in the program were generally simple variations of common visualization techniques. I was able to code the basis of each visualization fairly easily, but had to guess at how the visualizations were changed to match the data representation in the paper. The brushing and zooming between the overview of the data and the focused port was critical to make sure every layer of data was in sync. Overall, the coding needed to be done with a simple and easy-to-follow naming scheme because of the amount of dependency between the different view ports and representations. The most difficult part was creating clean and coherent interactions between the visualizations, which required a significant amount of data passing and filtering.

III. Paper Limitations

A. Text Preprocessing

The paper gives a very broad overview of the preprocessing stages; this is fair when it comes to the standard NLP text cleaning processes, but it makes it very difficult to reproduce more complex text analysis techniques such as the co-clustering algorithm they used. In the mood analysis section, they reference a paper that references another paper containing high-level implementation details of a constraint co-clustering algorithm; they never mentioned the type of clustering (k-means, spectral etc:-) or any other parameters (cluster size, number of clusters).

B. Visualizing

The paper did not have many details on the visualization techniques or the way the data was specifically captured in each visualization. For example, the dominance trait was used for a scatter plot of arrows along the stream graph, but we were unsure of what the arrow specifically referred to within the mood and why it was placed where it was. It was difficult to discern what interactions to implement and how they should appear because this tool is not available online, so we made our best guess based on the paper's depictions and descriptions. The paper was not specific with its range of colors and hue changes for representing the arousal of an emotion. It also did not discuss how its action menu selected or manipulated the data to show any of its 3 options: emotional outlook, emotional resilience, or extreme emotions.

IV. Implementation Differences

A. Text Preprocessing

Since the paper does not provide enough details about the clustering technique they used, we decided to extend the scikit library's k-means clustering algorithm to perform clustering.

Trump's tweets on a given day tend to cover a wide array of topics so we decided to only cluster on temporal and emotional dimensions. Although our tweet segments aren't exactly like those mentioned in the paper, we ended up deriving tweet segments that were temporally and emotionally coherent.

B. Visualizing

The paper did not specify how the dominance arrows were placed along the stream graph, so we decided to place one every week. While this is a bit cluttered in the overall view, we found it made a nice time step throughout the more focused views. We also decided to average the dominance dimension for every emotion for 3 days leading up to the arrow and 3 days after to find the orientation of the arrow. We decided not to include the scatter plot of the valence and arousal of each trigger word because it did not contribute to the goal stated in the paper or to the goal we were looking to accomplish. We used slightly different colors to make the system more visually appealing, but followed the guidelines set out by the Plutchik model. The paper did not define what "filtering" meant for the color key at the top, so we interpreted that to mean "highlighting." Rather than using a tooltip word cloud, we used the space from the omitted scatter plot to include an ordered list of trigger words, color coded by the emotion they trigger. These could also be highlighted by hovering over the emotion key at the top. Rather than have the bubble chart pop up while hovering over an emotion, we decided to have it appear on clicking. Clicking also activated the detail view at the bottom with the trigger words and raw text so that the emotion bubble could be compared to the tweets and triggers with more stability. In the paper, the word cloud tooltip would appear when the user hovered over the profile of the number of tweets path and the detail word view would appear when that is clicked. We concatenated these actions by clicking on the stream itself to display the tweets, ordered trigger words, and emotion bubble. We also included a y-axis to help users see exactly how many tweets per day were being sent. Lastly, when plotting the stream graph, the paper did not specify what valence value was being used for the graph's middle y-coordinate, so we offset it by the mean valence for all emotions for the day.

V. Dataset

The paper uses twitter posts of the subjects in their ground truth gathering study aimed at evaluating the accuracy of PEARL; each subject manually labeled tweet segments with emotion and VAD scores; the labeled segments were compared with those generated by PEARL to evaluate the tool's accuracy. As mentioned in our proposal, we decided to use Trump's tweets from 2015 through 2018 as our dataset. Trump is an avid twitter user who often posts multiple times a day so our dataset is fairly dense containing upwards of 18,000 tweets. We wanted to see if there were any drastic mood differences between his pre-presidency and presidency days. We also figured it would be interesting to see the overall volatility of his moods.

VI. Analysis Results

We noticed that the prominent emotions throughout the time span of the dataset were trust and anticipation. We didn't expect this result but attribute it to Trump's focus to market his personal brand. He tries he's best to embellish his views using positive language; a great example is his catch phrase "make America great again." While the words he uses are positive, the implication is that America has regressed over the past few years. Additionally, he uses twitter as a platform to advertise future policies and events; these phrases tend to contain anticipatory emotions. This system fails to encapsulate the nuances of Trump's intentions when he tweets about "crooked Hillary" or "building a wall." The portrayal of Trump's emotions does not match our perception for two additional reasons: human perception relies on experience and the meaning of compound phrases cannot be found by a lexicon-based approach. We were surprised when looking through the results at how effectively Trump was able to construct a positive narrative and outlook through his manipulation of phrases and intentions.

VII. Strengths & Limitations of our Implementation

A. Strengths

- Our implementation is less cluttered than theirs as we decided to omit visualizations (valence-arousal scatterplot, split-view and word cloud) that didn't add significant value to the goal of the tool.
- Our trigger word view is not a word-cloud inside a tooltip. If the trigger word view was a tooltip, it would be very difficult to view the trigger words synchronously with any of the other views as it would disappear when the mouse is moved. Additionally, a numbered, ordered list is more intuitive than a word cloud when it come to visualizing words associated with a weight.
- All the views are well connected and can be analyzed simultaneously

B. Limitations

- Like the paper, we used a lexicon based approach to assign emotions to tweets. This approach doesn't take into account compound words or sarcasm. We could extend this work by using a context-based approach to assign emotions (like a pre-trained RNN model).
- The raw tweets and trigger word view and defined in the same SVG. While scrolling horizontally, later tweets cannot be viewed in tandem with the trigger words. We had to use a fixed width SVG, so not all tweets can be rendered for days with a high tweet volume. This could be corrected with more time and experience in the web design development field.
- The data set analysis was limited because of Trump's use of slang and his frequent misspellings. However, Trump is the most relevant subject of emotional analysis in the public eye, so a good future work would be to implement an algorithm to correct spelling.