DALHOUSIE UNIVERSITY

FACULTY OF COMPUTER SCIENCE

Project Updates

In
The Class of

CSCI6406: Visualization

Topic

Visualization of Emotions by geographic location
from Tweeter Data Analysis

By

Deep Prashantkumar Patel
Master of Applied Computer Science
Dalhousie University
(deep.patel@dal.ca)
[B00865413]

Submitted to
Dr. Stephen Brooks

Department of Computer Science
Dalhousie university.

Date:4th March, 2022

Topic

Visualization of Emotions by geographic location from Tweeter Data Analysis

Introduction

In this project, the main focus is on representing human emotions over the map to get the knowledge of overall reactions from the area. Certain filtering options can be used to generate the map on certain data such as province-wise, city-wise, date-wise polarity, and also with keywords. Using all those data there will be a chance to visualize the data as per requirements to understand human emotions in a better way. Considering one scenario about the election can help determine which candidate has a better chance to win the election based on sentimental analysis of tweets. Visualization of time series can be also used to represent the popularity of candidates to determine the peak period of a particular candidate.

Completed Implementation Portion

The implementation task for this project can be carried out in three steps as given below;

1. **Data Gathering**

   The data gathering was one important task of this project as extracting real-time data and using that data can be a challenging task. As a first step, I emailed Twitter explaining my project to access the tweet. They provided a Twitter developer account and using access token, access secret token, consumer key, consumer secret token I was able to access the tweets.

   The data from the tweeter has been extracted using the tweepy library. Using Python programming I have achieved a functionality where 10000 tweets from the past two months can be extracted on a specific word. Inside each tweet's data section, we can find text written in a tweet which will help in determining the polarity of the text to understand the emotion of the person.

   Here is the sample extracted data from python code for tweets as shown in figure 1. The link for python code that can extract tweets is provided as GitLab project (tweets extraction.py [function get_tweets]) with a link to the extracted data.

2. **Data Processing**

   The data processing in terms of tweet data analysis is making the text inside tweets clear by removing effecting values such as hashtags, URLs, retweet details, and some improper values. Data processing has been achieved by the following steps;

   Step-1: Cleaning Data
   - The data cleaning has major importance in the task of data processing as the data needs to be cleaned before calculating the polarity of the text data to determine whether the tweet was happy in that context or not.

   1) RemoveURL: URL starts with HTML, www

2) Remove Noise Words: To remove emojis, hashtags, tagged names, images
3) Remove Numbers

Step-2: Polarity Calculation

To showcase the emotions based on geographical location, I used the Text Blob python library which provided the polarity of each tweet.

1) Polarity > 0: Positive tweet
2) Polarity < 0: Negative tweet
3) Polarity = 0: Neutral tweet

I studied the internal functionality of the calculation of polarity and found there is a default bag of positive words and negative words are defined. Tweets are being split into words and then there is a calculation of several positive, negative words in the sentences, and based on the maximum number the polarity is being decided.

For the currently ongoing event on "Ukrain", I filtered tweets with "Ukrain" and found below mentioned polarity shown in (figure 1) for tweets in Canada.

Figure 1: Tweets Extraction and Polarity Calculation

3. Visualization

The Twitter data has been fetched for either of 3 events – "Ukrain", "Work from Home", "Covid-19" from the data gathering and data cleaning section. The final aim is to provide options to the user for choosing events to see emotions by geographic location from Tweeter Data Analysis and eventually based on the selection of province the time-series graph will be plotted for happiness index. Also, for the current date, the province-wise donut chart will be displayed based on the total happiness index.

[1] Added Map of Canada – Figure 2

The complete map of Canada has been generated using d3 version 7 using the coordinated locations of each province. The entire map has been generated using NodeJS and functionality of d3 with SVG and displayed on chrome browser with a view box of a particular size.

[2] Provided mouse hover effect – Figure 2

The mouse hover effect has been provided with a simple functionality of fetching the id of that particular block and changing its color. The final plan is to show details of the province and processed values from tweet data on the tooltip.

[3] Provided mouse click effect – Figure 3

The on-click event has been provided with the normal functionality of changing the color of that region for now. The final plan is to generate a time-series graph of time series (zoom, select datelines, display data on hover) and donut chart (on mouse hover effect, on Click display data) with proper visualization effects.

Figure 2: Choropleth Map

Figure 3: Mouse hover effect

[4] Donut chart – Figure 4

The time series and donut graphs will reflect the positive, negative, and neutral effects on some events based on the tweeter data. Currently, data has been extracted from the tweeter data using python code and the Node page has been created with a visualization of Canada's map and some visualization effects.

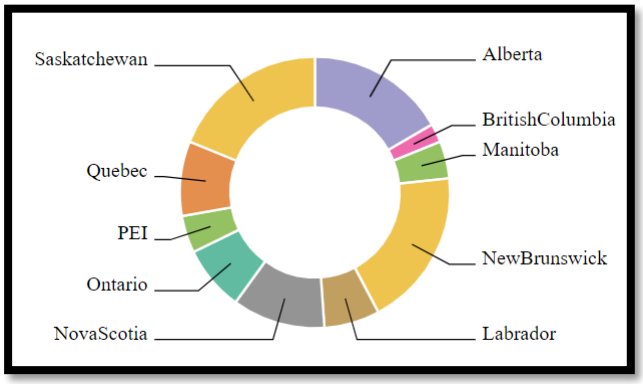

Figure 4: Donut chart based on the province on the event, polarity

Tasks need to be completed

- Bubble Map: Geographic provincial plots based on the polarity value and the size, color of the bubble
- Based on polarity count set the color of the geographic chart
- Based on provincial selection time series graph and donut chart to check the trend of happiness index (WIP)
- Zoom effect
- Integrate the visual code with python code for Twitter data

Git Project Link: https://git.cs.dal.ca/dppatel/csci6406_project.git

References

[1]. A. -L. Alten, G. Gadre, S. Kulkarni and C. -S. Wu, "Analyzing Happiness Index Based on Geographical Locations," 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), 2019, pp. 45-51, doi: 10.1109/ICAIBD.2019.8837010.

[2]. I. Costa, R. Lima, C. G. R. dos Santos, B. S. Meiguins, A. G. M. Soares and R. Y. da SilvaFranco, "EmojiText: An Information Visualization Technique for Analyzing Phrases and Sentiments," 2021 25th International Conference Information Visualisation (IV), 2021, pp. 114-119, doi: 10.1109/IV53921.2021.00027.

[3]. Twitter Sentiment Analysis using Python - GeeksforGeeks. (2017). Retrieved 5 March 2022, from https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/

[4]. Holtz, Y. (2022). Donut chart with group label in d3.js. Retrieved 5 March 2022, from https://www.d3-graph-gallery.com/graph/donut_label.html

[5]. canada-map-d3 - CodeSandbox. (2020). Retrieved 5 March 2022, from https://codesandbox.io/s/q87yj

[6]. GitHub - P7h/StormTweetsSentimentD3Viz: Computes and visualizes the sentiment analysis of tweets of US States in real-time using Storm. (2022). Retrieved 5 March 2022, from https://github.com/P7h/StormTweetsSentimentD3Viz

[7]. G. Kavitha, B. Saveen and N. Imtiaz, "Discovering Public Opinions by Performing Sentimental Analysis on Real Time Twitter Data," 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), 2018, pp. 1-4, doi: 10.1109/ICCSDET.2018.8821105.

[8]. S. Raheja and A. Asthana, "Sentimental Analysis of Twitter Comments on Covid-19," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 704-708, doi: 10.1109/Confluence51648.2021.9377048.