CSE 578
Data Visualization
MCS – Portfolio

Twitinfo - Aggregating and visualizing microblogs for event exploration

Sam Sunder Golla
Arizona State University
sgolla2@asu.edu
(1213172335)

INTRODUCTION

Many users try to query services like Twitter to follow or understand a particular event they are interested in. But the results of such queries make it hard for the user to understand an event since the results are basically a chronological log of posts. Our system allows a user to browse large collection of tweets which are visualized elegantly and thus help the user to summarize an event. We incorporated peak detection of twitter stream data related to a keyword or a hashtag associated with politics & sports which allows a user to understand how an event unfolded. Sentiment analysis is also portrayed in the form a visualization to aid a user in understanding the nature of the tweet.

EXPLANATION OF THE SOLUTION:

We are implementing a dashboard summarizing the event that the user has created (on the basis of the static dataset we have chosen, in this case it is soccer dataset). The dashboard displays a timeline for this event, raw tweet text sampled from the event, an overview graph of tweet sentiment, and a map view displaying tweet sentiment and locations. By aggregating metadata views, we are portraying the overall sentiment of a topic as well. By incorporating flux flow analysis, we tend to implement anomaly detection. We implemented outlier detection by using one-class svm and taking into account the number of followers, likes and retweets a particular user's tweet has received.

DESCRIPTION OF THE RESULTS:

Based on the keyword/hashtag, twitter API logs the required tweets. Using a line graph visualization, we created a central timeline interface, with the y-axis corresponding to tweet volume. Event-related tweets with geolocation are displayed using a map visualization. Tweets for the currently selected event or peak (subevent), are colored differently based on whether the sentiment is positive or negative. Using a simple visualization, we have also displayed the most

popular links in the currently-selected event or subevent and the Aggregated sentiment of all tweets in the event or subevent.

CONTRIBUTION TO THE PROJECT:

The contributions I made and the aspects that I am currently working on are:

1. Extracted data by using twitter APIs like Tweepy
2. Implemented sentiment analysis using Textblob library
3. Worked on timeline series visualization
4. Assisted in incorporating anomaly detection into our system
5. Hosted project online on Github

LESSONS LEARNED:

I was able to implement the concepts taught in class in this project and I enjoyed the task since front end designing was a thing that I was new to. I was able to correlate how certain datasets go along well with certain visualizations. I also worked on Datascraping, a new skill that I can add to my inventory. I also got a gist of how a story can be told using appropriate visualizations.

This course helped me expand my skillset to various machine learning algorithms and I ended up implementing one class SVM algorithm for outlier detection. Now, at the end of this semester, I'm really confident about the front-end design and development.

PROJECT MEMBERS:

On this project, I am collaborating with the following team members:

• Saumya Priya - saumya@asu.edu

• Saurabh Goyal - sgoyal20@asu.edu

• Anupriya Gupta – agupt224@asu.edu

ACKNOWLEDGEMENT:

I would like to extend my sincere thanks Dr. Chris Bryan for teaching and inspiring us in learning diverse technologies associated with Data Visualization. I would also like to thank the teaching assistant, Mr. Utkarsh Soni, for helping us during our project by giving his insights. I'm really thankful to my teammates Saurabh and Saumya for their role in the completion of this project.

REFERENCES:

[1]. E. Adar, M. Dontcheva, J. Fogarty, and D. S. Weld. Zoetrope: interacting with the ephemeral web. In UIST ’08. ACM Press, 2008.

[2]. P. Andre, M. L. Wilson, A. Russell, D. A. Smith, ´ A. Owens, and m. schraefel. Continuum: designing timelines for hierarchies, relationships and scale. In UIST ’07. ACM Press, 2007.

[3]. N. Bansal and N. Koudas. Blogscope: a system for online analysis of high volume text streams. In VLDB ’07. VLDB Endowment, 2007.

[4]. M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi. Eddi: Interactive Topic-based Browsing of Social Status Streams. In UIST ’10, 2010. [5]. danah boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In HICSS. IEEE, 2010.