

Visual Dashboard for Real-Time Analysis of Social Media

Huang Linya¹

linya.huang.2020@mitb.smu.edu.sg

Lim Jiahui¹

jiahui.lim.2020@mitb.smu.edu.sg

Zhang Ying¹

ying.zhang.2020@mitb.smu.edu.sg

School of Computing and Information Systems, Singapore Management University

Introduction

Detecting abnormal events, such as disaster or crisis, from microblog social media has become a trend, as social media has played a pervasive role in the way people behave and think. Nowadays, people are also using time-stamped, geo-located data to share live information about what's happening in their surroundings, which enables the public, government and researchers to identify abnormal events in community more quickly and take immediate actions.

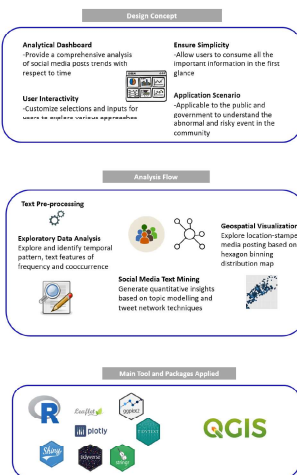
To better analyzing and visualizing social media texts, several text analytics techniques can be applied, such as word cloud, topic modeling, network analysis, geospatial analysis and so on. In real-world practices, researchers have built various social media text visualization using different techniques, but their solutions lack an integrated way of working.

In order to build a comprehend visualization dashboard with an interactive user interface, we built the application based on R shiny - a web application framework to create interactive web applications - and text analytical R packages.

Motivation

Our research is motivated by the lack of integrated and comprehensive real-time social-media dashboard. We aim to apply appropriate text analytics method and visually driven data analysis techniques in R language and R shiny to provide a handy analytic tool to help users understand the social-media posting through social media post analytics approaches.

Approach



Datasets

VAST Challenge 2021 Mini-Challenge 3 poses a social media and text analysis challenge to detect the meaningful event reports, evaluate risk level, public influence and suggest team of first response based on location-stamped posts. It provides the collection of microblogs and emergency calls from the event day and geospatial map of the corresponding areas.

No	Dataset	Information
1	Aliba.shp	Aliba geospatial map file with geometry information.
2	csv-1700-1830, csv-1831-2000, csv-2001-2131	Call Center and Blog information during time period from 1700 to 2131 on January 23, 2014, with 5 variables - date, author, message, hotspot location, latitude and longitude

Dashboard Design

Exploratory

Summary of datasets at glance

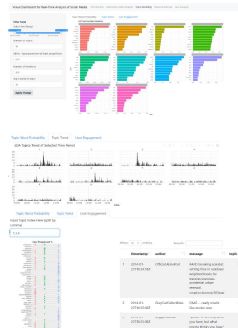


- Temporal Pattern: Temporal pattern distribution of the number of posts from various source such as microblog posts vs. call center messages, tweets vs. retweets given the specific time period.

- Word Cloud based on frequency and cooccurrence rate: straightforward text analysis provides a overview of content being discussed and help to highlight past events that occurred from various sources

Topic Modeling

Customized parameters for LDA topic modeling



By selecting time range and LDA topic modeling parameters, the text modeling will perform dynamically in the backend. With the LDA results, users can further explore top meaningful words in each topic, identify the peaks of key major events at certain point of time through topic trend distribution and profile users based on their posts contributions to the topics.

Network



With this VisNetwork object, we visualize the retweets network and draw the histograms of right skewed centrality distribution graph indicating patterns of the information dissemination. A variety of insights would be derived through different centrality measurements. For instance, most popular and active users could be identified by in-degree and out-degree centrality. And most influential users can also be detected by closeness and eigenvector centrality.

Geospatial



To visually display the social media contents, we first plotted location-stamped text and call data on the leaflet map (via extraction and `st_intersection` of cross junctions). Then, we extracted hexagon binning map by using MMQGIS grid layer function in QGIS application. Subsequently, we identified the location-stamped posts intersection with the hexagon and counted the number of posts within each hexagon areas. By plotting the hexagon layer with color gradient over Aliba map, we then visualized the hexagon binning distribution map.

Insights

EDA and Topic Modeling

From initial exploratory analysis, we observe from the spikes that the frequency of microblogs and calls received are in sync. The number of retweets and tweets also corresponds to each other, most likely by people affected during that period of time.

The top trending hashtags observed are "#POKRally" and "#KronosStar", while "#DancingDolphinFire" and "#standoff" interestingly appears towards the end of the timeline. A deeper analysis of tagged user shows that top users such as "@centralbulletin" and "@newsonlinetoday", which denotes the official accounts of Abila media outlets that are frequently tagged or retweeted by public.

Topic Modeling using microblog data and with 10 topics of interest across the entire evening shows that the main events occurred were: 1. Fire at Dancing Dolphin building (from 6:40pm to about 8pm) 2. Black Van incident and Shooting Standoff at gelato galore (from 7:40pm onwards) 3. POK Rally event (from 5pm to 6pm).

Network Analysis

We can tell from the in-degree distribution graph that the top 3 users which are being retweeted a lot by other users are news media AbilaPost, KronosStar and HomelandIlluminations.



From out-degree centrality, WiseWords, BlueVelvet, SaraJane and rnbwBrite are the most active users. Interestingly, for the eigenvector centrality distribution, the 6 out of top 10 tweet users are the same as that in in-degree centrality distribution. Only BlueVelvet, rnbwBrite and wiseWords are different.

Geospatial Analysis

- Shooting Standoff at Gelato Galore (7:40pm onwards)

The final event of the evening denotes another serious incident, given the high number of messages (both microblog and calls) received within the same area. We also noticed that the black van hit-and-run incident seems to be connected to the shooting as well, given the call messages locations tracked along the streets until around 7:39pm. Given the call message for SWAT team at 7:44pm, and that the estimated start time for this event is around 7:40pm, we noticed a slight lag in time but nonetheless almost in-synched.



Conclusion and Future Work

In this work we presented an integrated visual analytics dashboard for public and government analysis of social media data. We applied multiple text mining and visualization techniques for social events response based on text content visualization, network exploration and spatio-temporal analysis. For abnormal event response, we demonstrated a analyze flow to verify and examine the dashboard application. For future work, we will research and include more text mining techniques and provide more functional extension to the dashboard, and we hope to apply the concept of the framework design for more real-world text data analysis in response to time, entity relationship, and location.