New Event Detection Using Public Tweets

Proposal Presentation

F. Tuğba Doğan - M. Ali Yeşilyaprak - Simge Yücel

Description of the problem

- Twitter is a microblogging social media which users can share their daily life struggles, local and global events etc.
- We will try to derive the events from the tweets and determine the location of the event, using these tweets' word distribution, spatial and temporal features.

Motivation/Importance

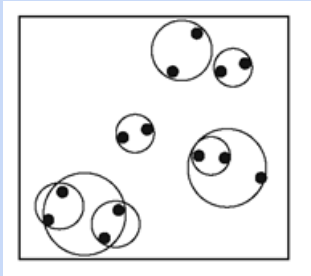
- Everyday nearly 170 million tweets are created.
- It becomes important to identify the events and bring event related tweets to forward to inform other users about the emerging events.
- Additionally, users may not want to be informed about local events which are occurring far from the user. Thus, deriving locality information of the event becomes important.

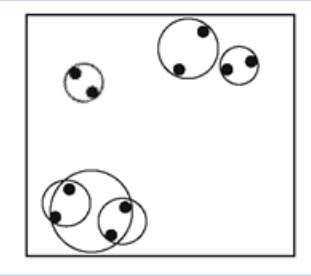
Methodology

- If the tweets are referring to the same topic we referred them as a clusters.
- We want to identify the tweets referring to events.
- We will fetch the words in tweets to derive their word distributions.
- From unique words, a dictionary will be generated.
- We will compute the word distribution matrix by counting the words in each tweet.

Methodology

- We will cluster the data with agglomerative clustering techniques.
- Clusters with high standard deviation values and low number of points will be eliminated as they will be considered as outliers.
- Cluster’s location will be determined according to its tweets’ user location information.





Methodology

- We’ll divide the data into frames, each frame consists of 5 days.
- Compare the first frame and the last frame.
- We define the similarity of the clusters according to their euclidean distances.
- Distances which are smaller than a threshold will be eliminated.

Input Data

- The data set contains 115,886 Twitter users and 3,844,612 updates from the users (collected from September 2009 to January 2010). All the locations of the users are self-labeled in United States in city-level granularity.
- Link: https://archive.org/details/twitter_cikm_2010

Expected Result

- We will distinguish event tweets and non-event tweets.
- We will create clusters according to both twitter content and users' location.
- Local and global events are identified by using these clusters. We will compare our result with the gold standard.

References

[1] Becker, Hila, Mor Naaman, and Luis Gravano. "Beyond trending topics: Real-world event identification on twitter." (2011).
[2] Abdelhaq, Hamed, Christian Sengstock, and Michael Gertz. "Eventweet: Online localized event detection from twitter." *Proceedings of the VLDB Endowment* 6.12 (2013): 1326-1329.
[3] Ozdikis, Ozer, Pinar Senkul, and Halit Oguztuzun. "Semantic expansion of hashtags for enhanced event detection in Twitter." *Proceedings of the 1st International Workshop on Online Social Systems*. 2012.
[4] Li, Rui, et al. "Tedas: A twitter-based event detection and analysis system."*Data engineering (icde), 2012 ieee 28th international conference on*. IEEE, 2012.

Thank You For Listening

Any Questions?