

Sub-Event Detection and Sentiment Analysis in Social Networks

Sayan Unankard, Xue Li, and Mohamed A. Sharaf

Motivation : With the increasing volume of discussion on real-world events over social networks, event detection and monitoring are becoming compelling research issues. For a longer-running event like an election, users may want to **monitor sub-events** (i.e., hierarchically nested events which break down the event in more refined parts) such as the debate or campaign launch speech for easily understanding the key points of a discussion. Also, the policy-makers may want to **detect the feeling of users** during the course of the elections.

Challenge

- With a very large number of short and noisy messages and **unknown no of events in advance**, the sub-event detection performance is a critical issue.
- Traditional lexicon dictionary does **not cover opinion words** in micro-blogs. Also, micro-blog messages contain extensive use of **irony and sarcasm** which are difficult for a machine to detect.

Methods

- We deploy an **hierarchical clustering** method to group messages into clusters. **Slang conversion** and **synonym expansion** is utilized. A cluster can be considered as sub-event if there is strong correlation between the event location (location mentioned in the messages) and the user location.
- We used combination between a **lexicon-based** approach and **rule-based** approach to perform sentiment classification. We used the lexicon dictionary which was introduced in [Hu+KDD2004] and our annotated the Internet slang dictionary. Sarcasm identification method is proposed.

Evaluation results :

- We compare our method performance with temporal peaks detection method in [Macus+CHI2011]. Our approach outperforms the baseline by **31.43%** in F1-Score.
- We evaluate our sentiment analysis method by comparing the performance with aspect-based opinion summarization on Twitter data [Ringsquandl+AAAI2013]. Our method can effectively classify with a F1-score higher than the baseline in the same domain of politics by average **20.19%**.

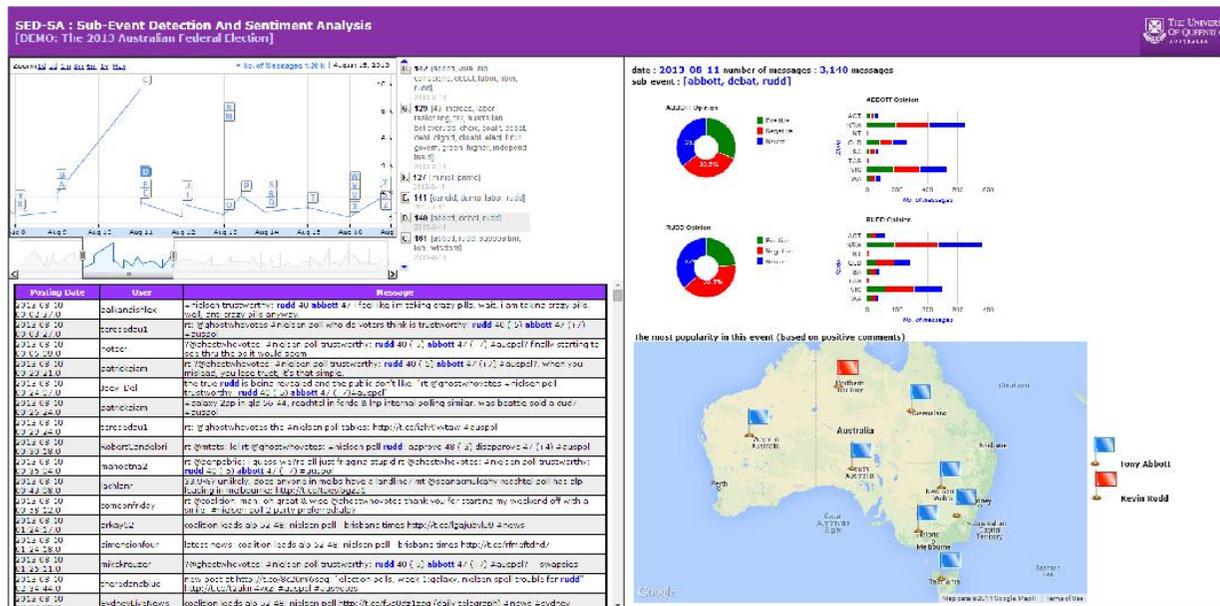


Fig 1 : A dashboard to display sub-event and sentiment of two specific candidates.

Dataset: The messages posted by Australian-based users related to **the 2013 Australian Federal Election** were collected by Twitter Search API. **808,661** messages (**4 Aug – 8 Sep 2013**) with the user's initial event query is used for our experiments.