Bairi Sandhya Rani
14CO205
(+91) 9035769725
sandhya.bairi154@gmail.com

Interpreting the Public Sentiment Variations on Twitter

Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Jiajun Bu, Chun Chen, and Xiaofei He

ESSENCE

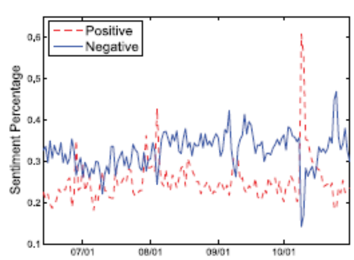
- Millions of users use Twitter to share their opinions which is used for sentiment analysis and decision making and to find their relation with real-world events.
- This paper proposes two Latent Dirichlet Allocation (LDA) based models, namely Foreground and Background LDA (FB-LDA) and Reason Candidate and Background LDA (RCB-LDA).
- FB-LDA helps to extract foreground topics and filter out longstanding background topics and can give potential interpretations of the sentiment variations.
- RCB-LDA ranks the most representative tweets with respect to their "popularity" within the variation period to further enhance the readability of the mined reasons.
- The proposed models can also be applied to other tasks such as finding topic differences between two sets of documents.

WORK DESCRIBED

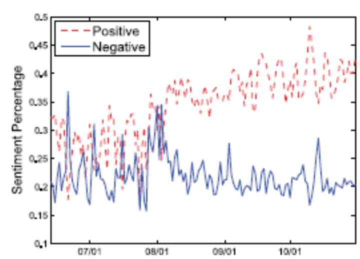
- Emerging topics within the sentiment variation periods could be highly related to the genuine reasons behind the variations.
- The work is to analyze and interpret the public sentiment variations in microblogging services, followed by the above stated generative models that are developed to solve the reason mining problem.

IMPORTANCE

- Finding possible reasons behind sentiment variation can provide important decision making information, for instance, if public sentiment changes greatly on some products, the related companies may want to know the reason behind such feedback for their products.
- Similarly, if negative sentiment towards Barack Obama increases significantly, their Administration office may want to know the reason for changed opinions of the people and then react accordingly to reverse this trend.



Sentiment variation tracking of "Obama".



Sentiment variation tracking of "Apple".

METHODOLOGY

Initially, Public Sentiment is tracked using the following three steps namely, tweets extraction and preprocessing, sentiment label assignment and sentiment variation tracking. Under sentiment label assignment, a sentiment label is assigned to each individual tweet by combining two state-of-the-art sentiment analysis tools SentiStrength and TwitterSentiment. Finally, based on the sentiment labels obtained for each tweet, the sentiment variation is tracked regarding the corresponding target using some descriptive statistics. As mentioned earlier, FB-LDA can filter out background topics and extract foreground topics from tweets in the variation period, with the help of an auxiliary set of background tweets generated just before the variation. RCB-LDA is an extension to FB-LDA that does tweet-candidate association. It first extracts representative tweets for the foreground topics (obtained from FB-LDA) as reason candidates. Then it will associate each remaining tweet in the variation period with one reason candidate and rank the reason candidates by the number of tweets associated with them.

PROS AND CONS

- Experimental results on real Twitter data show that this method can outperform baseline methods in finding foreground topics and ranking reason candidates, and effectively mine desired information behind public sentiment variations.
- In comparison with LDA and k-means, FB-LDA can work properly without depending on any manually set or fixed thresholds.
- These models are general and not limited to the possible reason mining problem. They can be applied to various tasks involving finding the topic differences between two sets of documents.
- To find the correlation between tweets and events, RCB-LDA model utilizes a background tweets set as a reference to remove noises and background topics. As a result, the interference of noises and background topics can be eliminated.

FUTURE WORK

- A summary generation model can be developed that analyzes trending topics by detecting the sub topics and summarizes the public views on them efficiently over time period.
- A sentiment-based rating prediction method can be developed to improve prediction accuracy in recommender systems that takes each user's sentimental attributes on different products, interpersonal sentimental influence and item's reputation similarity into consideration.

WORK FOR IMPLEMENTATION

- Sentiment Analysis of Twitter data and basic LDA models