

CS226 Big-Data Management
Project Proposal

Topics trend on social media

Group 8 members:

Baixi Sun 862123350
Longze Su 862188702
Yunfan Kang 862099709
Zhijie Qiao 862114253
Jiatai Chen 862120043

Project Overview:

This project aims to analyze the trend of different topics on social media like Twitter. Only a generation ago, keeping up to date with the latest topics was a pretty simple task – pick up the daily newspaper or watch the evening news on TV. The explosion in information and communication channels mean it's not quite so simple nowadays. Nowadays as social media is constantly evolving, people would prefer to look for news, share their points of view or feelings on social media. News organization also post breaking news on their official twitter account. If a majority of users are searching for and posting feelings about a certain news, it becomes a trending topic.

Topics trend is important for any business. For marketers and community managers, joining a conversation in a timely manner means you are able to add something to the conversation. There may be trends within your industry that dictate product development. Presenting trending contents to users can also convenience them for acquiring information and increase his app usage or increase selling of certain products [1]. Various attempts have also been made to identify the trending topics on social networks in the past decade [2][3][4].

Topic trend could vary with time and specific events. For example, on football nights the trending topics are probably about NFL. If some nation-wide breaking news occurred, the users will be searching and posting about it at the same time. Our aim will be investigating how the topics trend vary.

Planned Approach:

Approaches to investigate the trending topics on different social media varies. For Twitter or Instagram it could be crawling the Hashtag, which serves a purpose of exposing the post to a wider audience. Hashtags are a tool that social media platforms use to categorize their content. Same or similar hashtags could categorize the post into a certain topics.

We plan to use the data acquired from API provided by the social medias for developers, like Twitter streaming API.

For the analysis tasks we would use Hadoop MapReduce to count the frequency of Hashtags in a certain time interval and stand out the most frequent ones.

Milestones:

Week	Task
3	Proposal and Literature Survey
4	Project High-level Design and Methodology
5-6	Building models and Testing
7	Analysis of insights and Paper outline
8	Visualization and Result Analysis
9	Draft and finalize the paper.

References:

- [1] Zhao, W.X., Guo, Y., He, Y., Jiang, H., Wu, Y., Li, X.: We know what you want to buy: a demographic-based system for product recommendation on microblogs. In: KDD (2014)
- [2] TweetTracker: track, analyze, and understand activity on Twitter. tweettracker.fulton.asu.edu/, 2014.
- [3] Ceren Budak, Theodore Georgiou, Divyakant Agrawal, and Amr El Abbadi. GeoScope: Online Detection of Geo-Correlated Information Trends in Social Networks. In VLDB, 2014.
- [4] JingwenBian, YangYang, and Tat-SengChua. Multimediasum- marization for trending topics in microblogs. In CIKM, 2013.