International Journal of Research in Engineering, Science and Management
Volume-2, Issue-3, March-2019
www.ijresm.com | ISSN (Online): 2581-5792

Searching and Ranking News Topics using Twitter

Srushti Yewale[1], Pooja Walunj[2], Nita Sahane[3], Anushka Sonawane[4]
[1,2,3,4]*Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India*

***Abstract*: In modern times, social media services such as Twitter provide an enormous amount of user-generated data, which have great potential to contain informative news-related content. For these assets to be helpful, we should find an approach to filter clamor and just catch the substance that, in light of its comparability to the news media, is viewed as important. In any case, even after clamor is evacuated, information over-burden may at present exist in the rest of the information-thus, it is advantageous to organize it for utilization. To accomplish prioritization, data must be positioned arranged by evaluated significance considering three variables. To begin with, the transient prevalence of a specific point in the news media is a factor of significance, and can be viewed as the media center (MF) of a subject. Second, the worldly pervasiveness of the point in online networking demonstrates its client consideration (UA). Last, the collaboration between the online networking clients who specify this theme indicates the quality of the group talking about it, and can be viewed as the client connection (UI) around the subject. We propose an unsupervised structure which identifies news subjects predominant in both web-based social networking and the news media, and afterward positions them by utilizing their degrees of MF, UA, and UI**

***Keywords*: Information filtering, social computing, social network analysis, topic identification, topic ranking.**

1. Introduction

Twitter, is utilized by a large number of individuals around the globe, star viding tremendous measures of client created information. One may accept that this source possibly contains data with equivalent or more noteworthy incentive than the news media, however one should likewise expect that due to the unverified idea of the source, quite a bit of this substance is futile. For web-based social networking information to be of any utilization for point identification, we should find an approach to filter uninformative data and catch just data which, in light of its substance comparability to the news media, might be viewed as helpful or important. The news media introduces professionally verified events or occasions, while online networking presents the interests of the group of onlookers in these territories, and may in this way give understanding into their fame. Online networking administrations like Twitter can likewise give extra or supporting data to a specific news media point. The system will focus on providing filtered news to the users. The system will first use twitter data set and filter out all the twits that are related to news. After filteration of news related twits, news topics are ranked based on factors like MF,UA, UI. In addition, the system will also focus on classifying news topics based on the location of user. It is expected that through the providing of filtered news, instead of reading unnecessary data user gets to read quality news depending on his interests and location. Even though this paper focuses on news topics, it can be easily adapted to a variety of fields, from technology to culture and sports. Our system has a framework for keyword extraction, graph clustering and location wise filteration of news. To achieve its goal, our system uses keywords from news media sources to identify the overlap with Twitter from that same period. We then build a graph whose nodes represent these keywords and whose edges depict their co-occurrences in social media. The graph is then clustered to clearly identify distinct topics. After obtaining well-separated topic clusters (TCs), the factors that signify their importance are calculated: MF, UA, and UI. Finally, the topics are ranked that combines these three factors.

2. Literature survey

A. Generating event storylines from microblogs (2012)

Microblogging service has emerged to be a dominant web medium for billions of individuals sharing and spreading instant news and information, therefore monitoring the event evolution on microblog sphere is crucial for providing both better user experience and deeper understanding on real-time events. In this paper we explore the problem of generating storylines from microblogs for user input queries. This problem is challenging due to the sparse, dynamic and social nature of microblogs. Given a query of an ongoing event, we propose to sketch the real-time storyline of the event by a two-level solution. We first propose a language model with dynamic pseudo relevance feedback to ob tain relevant tweets, and then generate storylines via graph optimization. Comprehensive experiments on Twitter data sets demonstrate the effectiveness of the proposed methods in each level and the overall framework

B. Summarizing sporting events using twitter (2012)

The status updates posted to social networks, such as Twitter and Facebook, contain a myriad of information about what people are doing and watching. During events, such as sports games, many update saresent describing and expressing opinions about the event. In this paper, we describe an

algorithm that generates a journalistic summary of an event using only status updates from Twitter as a source. Temporal cues, such as spikes in the volume of status updates, are used to identify the important moments within an event, and a sentence ranking method is used to extract relevant sentences from the corpus of status updates describing each important moment within an event. We evaluate our algorithm compared to human-generated summaries and the previous best summarization algorithm, and find that the results of our method are superior to the previous algorithm and approach the readability and grammaticality of the human-generated summaries.

C. Two sides to every story: Subjective event summarization of sports events using Twitter (2014)

Researchers have shown how different observers can describe events from very different perspectives, and how these perspectives can be discovered and analyzed. It is demonstrated that this allows story telling via automated community-discovery and automated topic detection. The focus has been on the difference between comments from fans of the two teams over course of a match, and we have shown how the volume and focus of topics of discussion vary overtime. In particular, supporters are more vocal and focused when their team has an advantage, especially towards the end of a match: they only tweet when they're winning.

3. Existing system

If Two traditional methods for detecting topics are LDA and PLSA. LDA is a generative probabilistic model that can be applied to different tasks, including topic identification. PLSA, similarly, is a statistical technique, which can also be applied to topic modeling. In these approaches, however, temporal information is lost, which is paramount in identifying prevalent topics and is an important characteristic of social media data. Matsuo et al. employed a different approach to achieve the clustering of co-occurrence graphs. They used Newman clustering to efficiently identify word clusters. The core idea behind Newman clustering is the concept of edge betweenness. The betweenness measure of an edge is the number of shortest paths between pairs of nodes that run along it. If a network contains clusters that are loosely connected by a few inter cluster edges, then all shortest paths between different clusters must go along one of these edges. Consequently, the edges connecting different clusters will have high edge betweenness, and removing them iteratively will yield well-defined clusters. If you have more sub points you can use as per the requirement.

A. Disadvantages

- Even after the removal of unimportant content, there is still information overload in the remaining news-related data, which must be prioritized for consumption.
- LDA and PLSA only discover topics from text corpora; they do not rank based on popularity or prevalence.
- The main disadvantage of the algorithm was its high computational demand.
- The existing work, however, only considers the personal interests of users, and not prevalent topics at a global scale.
- These methods, however, only use data from microblogs and do not attempt to integrate them with real news. Additionally, the detected topics are not ranked by popularity or prevalence.

4. Proposed system

In this project we propose k means model in which we gather news information from different news portal and on the basis of news popularity we categories news in different domain and then rank it using k-means algorithm put important news always top of the dashboard of social media site. Also displays news with respect to user domain so they get only those news which they want it.

A. Advantages

- User get result in his/her respective domain only.
- Popular news always displays on top.
- Detection and removal of fake news on social media.
- User can easily choose different domain and get trending news.

5. Architecture

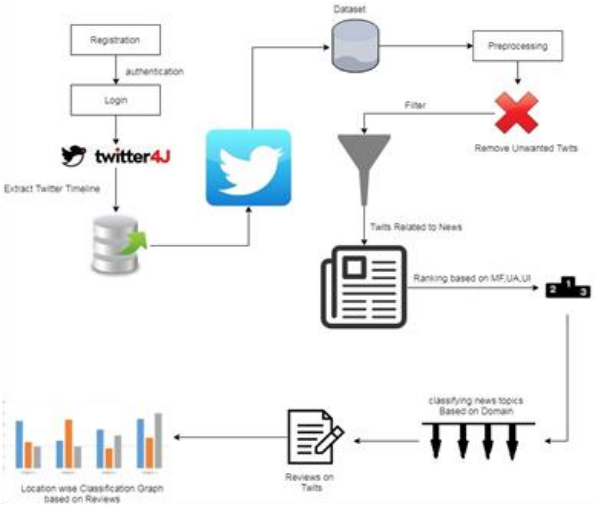

Fig. 1. Architecture

6. Conclusion

This paper presented an overview on searching and ranking news topics using twitter.

Acknowledgement

This survey is supported by the Sinhgad Academy of Engineering, under the Dept. of Computer Engineering and is supervised by Prof. Gauri Bhange.

References

[1] Lin Chenand Chun Lin, "Generating event story lines from microblogs," Proceedings of the 21st ACM international conference on Information and knowledge management, pp. 210-216, 2012.

[2] Nichols Jeffrey, Jalal Mahmud and Clemens Drews, "Summarizing sporting events using twitter," Proceedings of the 2012 ACM international conference on Intelligent User Interfaces, pp. 122-128, 2012.

[3] Corney David, Carlos Martin and Ayse Gker, "Two sides to every story: Subjective event summarization of sports events using Twitter," in ICMR2014 workshop on Social Multimedia and Storytelling, pp. 662-672, 2014.

[4] Brendon O. Connor and Bala Subramanyan, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," Proceedings of the International AAAI Conference on Weblogs and Social Media, Washington, DC, pp. 511519, May 2010.

[5] Sakaki Takeshi, Makoto Okazaki and Yutaka Matsuo, Earth quake shakes Twitter users: real-time event detection by social sensors," in Proceedings of the19th international conference on World Wide Web, pp. 851- 860, 2010.

[6] K. Shubhankar, A. P. Singh, and V. Pudi, "An efficient algorithm for topic ranking and modeling topic evolution," in Database Expert Syst. Appl., Toulouse, France, 2011, pp. 320330.

[7] K. Sarkar, M. Nasipuri, and S. Ghose, "A new approach to key phrase extraction using neural networks," Int. J. Comput. Sci. Issues, vol. 7, no. 3, pp. 1625, Mar. 2010.

[8] H. Yin, B. Cui, H. Lu, Y. Huang, and J. Yao, "A unified model for stable and temporal topic detection from social media data," in Proc. IEEE 29th Int. Conf. Data Eng. (ICDE), Brisbane, QLD, Australia, 2013, pp. 661-672.

[9] C. Wang M. Zhang, L. Ru, and S. Ma, Automatic online news topic ranking using media focus and user attention based on aging theory," inProc.17th Conf. Inf. Knowl. Manag., Napa County, CA, USA, 2008, pp. 1033-1042.

[10] E. Kwan, P. L. Hsu, J. H. Liang, and Y.-S. Chen, Event identification for social streams using keyword-based evolving graph sequences," in Proc. IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal. Min., Niagara Falls, Canada, 2013, pp. 450-457.