

DETECTING, RANKING AND DOMAIN WISE SORTING OF NEWS ON SOCIAL MEDIA

Komal Mahamuni¹, Aishwarya Kadam², Shrutika Deokar³, Prof. Kavita Jadhav⁴

¹Computer Engineering, Pune university,

²Computer Engineering, Pune university,

³Computer Engineering, Pune university,

⁴Computer Engineering, Pune university,

Abstract— Now a day's many people are using social Media for getting information. For example Twitter is providing huge amount of user generated data, which have great importance to contain news-related content. So for finding news related data first of all we need to remove or filter noise from the all data. After removing the noise there can be the data which is not related news. So we again we need to use data prioritization. For data prioritization we are going to use MF, UA and UI as factors. After detecting news related data we are going to rank that news using MF,UA and UI as well as we will categorize all the news location wise using comments or reviews. We are going to use Twitter data set for performing all this operations.

Keywords — Disseminating news, User Interest, User interaction, Mass Identification, relevance factor, news, social computing, analysis of social network, topic ranking.

INTRODUCTION

Now a days, internet and social media platforms have become most popular ways of spreading and broadcasting information and news as compared to traditional media like newspapers. For example, Twitter which is used by millions of people all over the globe for micro blogging services. Huge and enormous amount of user data is generated by this. Apart from twitter, there are other social media platforms which also generate the similar amount of data. All these generated data which in turns can be converted into information and thus can be useful to identify similar trends and generate news related data.

Due to the enormous amount of data being generated daily, even after removal of unwanted data, there still remains lot of overloaded information to use. Hence to help prioritize relevant data, news must always be ranked in order of conjecture importance. While estimating relevant topics, the popularity of a particular news or article, users interest in particular topic and the number of users or people discussing and interacting about topic are three most important factors while choosing the topical importance. This paper deals with all these three aspects in details and provides rank of news accordingly

Literature survey

Title – Blog News Rank: Finding and Ranking Frequent News Topics Using Social Media Factors

Authors: Harshitha H, Mohammed Rafi

Description:

In early days, mass media sources such as news media used to inform us about daily events. Now a days, social media services such as Twitter huge amount of user-generated data, which has a great potential to contain informative news-related content. For these resources to be useful, we have to find a way to filter noise and capture the content that, based on its similarity to the news media, is considered valuable. Information overload stills exist even after noise and unwanted data is removed. This paper proposes and gives a solution to build an unsupervised framework i.e. Blog News Rank. This framework basically identifies and ranks topic by their relevance of frequency.

Title : Automatic Online News Topic Ranking Based on Aging Theory

Authors: Canhui Wang, Min Zhang, Liyun Ru, Shaoping Ma

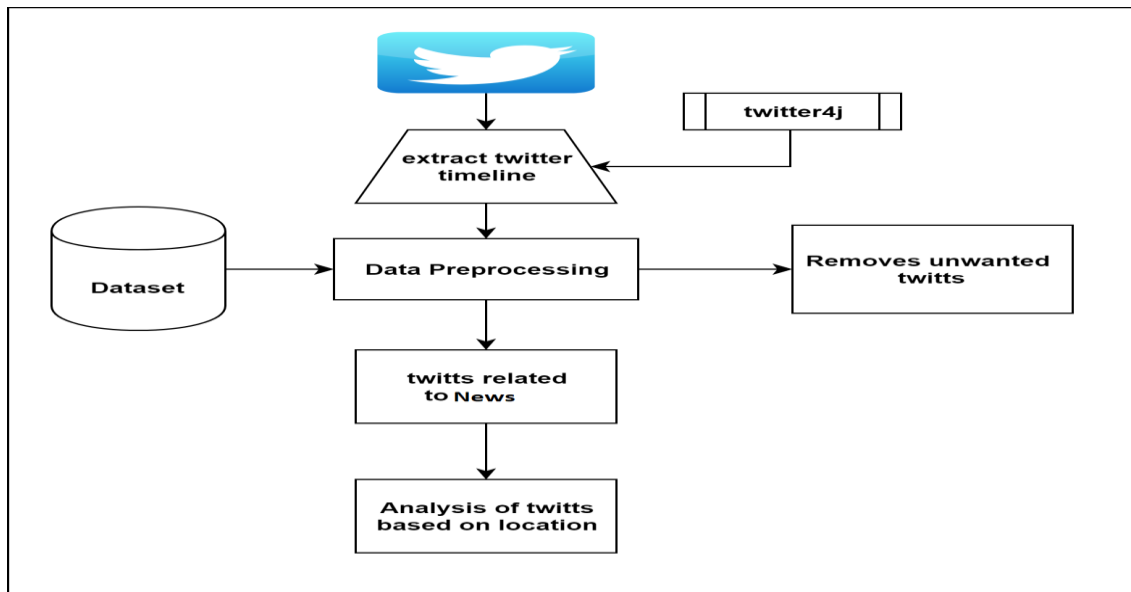
Description:

In today's world where so many news are being generated every minute, it is almost impossible to view all generated topics. Hence it's a necessity to rank these topics with their respective priority. This paper does topic ranking using quantitative measure of inconsistency between user attentions and media focus. This serves as a base of topic ranking. Also it provides evidence that there is a gap between what media shows and what users view. Thus this system is based on ageing theory.

PROPOSED SYSTEM

In this project we are going to gather news related data from Twitter and using the reference popularity of news we are going to categories news in different domain and then we will rank the news using k-means algorithm. Our system will put news top of the dashboard of social media site which is popular always. Also we will displays the news as per the user domain so user get only those news in which users are interested. Apart from this we are classifying tweets based on few pre-existing approaches. We are classifying the news based on the location. The system also solves the problem of population concentration. Like for example a state like Maharashtra which is so populated will have a more dominant feelings or sentiments over particular News topic in the sentiment analysis and would silence down the voice of the state with less population like Goa. The main purpose is to determine the overall sentiment region wise.

SYSTEM ARCHITECTURE



MATHEMATICAL MODEL

Let W be the set of whole system which consists of the input, process and output of the system.

$$W = \{input, process, output\}.$$

where,

input = is the set of inputs given to the system to achieve the problem statement.

process = is the procedure or the algorithm applied to the system which gives the expected output.

output = is the output of the system.

$$W = U, T, NT, R, L, D.$$

Where,

1. U is the set of number users.

$$U = \{U_1, U_2 \dots \dots U_n\}.$$

2. T is set of twits on timeline

$$T = \{t_1, t_2, \dots \dots t_3\}.$$

3. NT be the set of news related tweets.

4. R be the rank.

5. D is the dataset required for stop words removal.

6. L be the location of tweet.

Process:

Step 1: user U will submits the twits on timeline or re-twits or comments on particular twit.

At the time of tweet, user need to add his location to that tweet. This tweets may be related to social, entertainment, educational or anything, but we will consider the tweets related to news.

Step 2: the admin will request for analysis of data.

Step 3: the system will fetch or extract the timelines data by using twitter 4j jar of various twits T.

Step 4: the system will perform the data preprocessing in which the stops words are removed and as well as twits related to news are extracted separated by using the dataset D.

Step 5: After this, system will classify the data into region wise and rank R.

ADVANTAGES OF PROPOSED SYSTEM

- 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.

CONCLUSIONS

In this paper we have successfully demonstrated and proposed a method called Social Rank. This has been created to identify news topics from various generated sources and social media platform and then eventually rank them by considering their MF, UA and UI factors and relevancy. Filtered and Ranked news related topics from both professional news providers and individuals have several benefits. Our main goal is increasing the quality of news recommender systems, as well as finding hidden, popular topics. Our system we are classifying and clustering the news location wise and positive and negative feedbacks, so we can get proper result or location wise opinion of peoples for particular Twits.

REFERENCES

- [1] O. Phelan, K. McCarthy, and B. Smyth, "Using Twitter to recommend real-time topical news," in Proc. 3rd Conf. Recommender Syst., New York, NY, USA, 2009, pp. 385–388.
- [2] 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, ON, Canada, 2013, pp. 450–457.
- [3] K. Sarkar, M. Nasipuri, and S. Ghose, "A new approach to keyphrase extraction using neural networks," Int. J. Comput. Sci. Issues, vol. 7, no. 3, pp. 16–25, Mar. 2010.
- [4] 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. [11] C. Wang, M. Zhang, L. Ru, and S. Ma, "Automatic online news topic ranking using media focus and user attention based on aging theory," in Proc. 17th Conf. Inf. Knowl. Manag., Napa County, CA, USA, 2008, pp. 1033–1042.
- [5] C. C. Chen, Y.-T. Chen, Y. Sun, and M. C. Chen, "Life cycle modeling of news events using aging theory," in Machine Learning: ECML 2003. Heidelberg, Germany: Springer Berlin Heidelberg, 2003, pp. 47–59.
- [6] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling, "Twitter Stand: News in tweets," in Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst., Seattle, WA, USA, 2009, pp. 42–51.
- [7] O. Phelan, K. McCarthy, and B. Smyth, "Using Twitter to recommend real-time topical news," in Proc. 3rd Conf. Recommender Syst., New York, NY, USA, 2009, pp. 385–388.
- [8] 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. 320–330.
- [9] S. Brin and L. Page, "Reprint of: The anatomy of a large-scale hypertextual web search engine," Comput. Network., vol. 56, no. 18, pp. 3825–3833, 2012.