International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 30th June 2020. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-09&issue=ISSUE-06

Title: **DETECTING LEGITIMATE INFORMATION BY IDENTIFYING AND PRIORITIZING PREVALENT NEWS USING SOCIAL MEDIA FACTORS**

Volume 09, Issue 06, Pages: 147-152

Paper Authors

GUGGULLA SIVANANDA REDDY, C.BALAJI

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

DETECTING LEGITIMATE INFORMATION BY IDENTIFYING AND PRIORITIZING PREVALENT NEWS USING SOCIAL MEDIA FACTORS

GUGGULLA SIVANANDA REDDY, C.BALAJI
PG SCHOLAR.DEPT OF CSE, SIR C.V. RAMAN INSTITUTE OF TECHNOLOGY & SCIENCE, AP, INDIA
ASSOCIATE PROFESSOR, DEPT OF CSE, SIR C.V. RAMAN INSTITUTE OF TECHNOLOGY & SCIENCE,, AP, INDIA

ABSTRACT: Mass media sources, specifically the news media, have traditionally informed us of daily events. 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 resources to be useful, we must find a way to filter noise and only capture the content that, based on its similarity to the news media, is considered valuable. However, even after noise is removed, information overload may still exist in the remaining data—hence, it is convenient to prioritize it for consumption. To achieve prioritization, information must be ranked in order of estimated importance considering three factors. First, the temporal prevalence of a particular topic in the news media is a factor of importance, and can be considered the media focus (MF) of a topic. Second, the temporal prevalence of the topic in social media indicates its user attention (UA). Last, the interaction between the social media users who mention this topic indicates the strength of the community discussing it, and can be regarded as the user interaction (UI) toward the topic. We propose an unsupervised framework—SociRank—which identifies news topics prevalent in both social media and the news media, and then ranks them by relevance using their degrees of MF, UA, and UI. Our experiments show that SociRank improves the quality and variety of automatically identified news topics.

1. INTRODUCTION

THE mining of valuable information from online sources has become a prominent research area in information technology in recent years. Historically, knowledge that apprises the general public of daily events has been provided by mass media sources, specifically the news media. Many of these news media sources have either abandoned their hardcopy publications and moved to the World Wide Web, or now produce both hard-copy and Internet versions simultaneously.

These news media sources are considered reliable because they are published by professional journalists, who are held accountable for their content. On the other hand, the Internet, being a free and open forum for information exchange, has recently seen a fascinating phenomenon known as social media. In social media, regular, nonjournalist users are able to publish unverified content and express their interest in certain events.

Microblogs have become one of the most popular social media outlets. One microblogging service in particular, Twitter, is used by millions of people around the world, providing enormous amounts of user-generated data. One may assume that this source potentially contains information with equal or greater value than the news media,

but one must also assume that because of the unverified nature of the source, much of this content is useless. For social media data to be of any use for topic identification, we must find a way to filter uninformative information and capture only information which, based on its content similarity to the news media, may be considered useful or valuable.

The news media presents professionally verified occurrences or events, while social media presents the interests of the audience in these areas, and may thus provide insight into their popularity. Social media services like Twitter can also provide additional or supporting information to a particular news media topic. In summary, truly valuable information may be thought of as the area in which these two media sources topically intersect. Unfortunately, even after the removal of unimportant content, there is still information overload in the remaining news-related data, which must be prioritized for consumption.

To assist in the prioritization of news information, news must be ranked in order of estimated importance. The temporal prevalence of a particular topic in the news media indicates that it is widely covered by news media sources, making it an important factor when estimating topical relevance. This factor may be referred to as the MF of the topic. The temporal prevalence of the topic in social media, specifically in Twitter, indicates that users are interested in the topic and can provide a basis for the estimation of its popularity. This factor is regarded as the UA of the topic. Likewise, the number of users discussing a topic and the interaction between them also gives insight into topical importance, referred to as the UI. By combining these three factors, we gain insight into topical importance and are then able to rank the news topics accordingly.

Consolidated, filtered, and ranked news topics from both professional news providers and individuals have several benefits. The most evident use is the potential to improve the quality and coverage of news recommender systems or Web feeds, adding user popularity feedback. Additionally, news topics that perhaps were not perceived as popular by the mass media could be uncovered from social media and given more coverage and priority. For instance, a particular story that has been discontinued by news providers could be given resurgence and continued if it is still a popular topic among social networks. This information, in turn, can be filtered to discover how particular topics are discussed in different geographic locations, which serve as feedback for businesses and governments.

A straightforward approach for identifying topics from different social and news media sources is the application of topic modeling. Many methods have been proposed in this area, such as latent Dirichlet allocation (LDA) [1] and probabilistic latent semantic analysis (PLSA) [2], [3]. Topic modeling is, in essence, the discovery of “topics” in text corpora by clustering together frequently co-occurring words This approach, however, misses out in the temporal component of prevalent topic detection, that is, it does not take into account how topics change with time. Furthermore, topic modeling and other topic detection techniques do notrank topics according to their popularity by taking into account their prevalence in both news media and social media.

We propose an unsupervised system—SocialRank—which effectively identifies

news topics that are prevalent in both social media and the news media, and then ranks them by relevance using their degrees of MF, UA, and UI. Even though this paper focuses on news topics, it can be easily adapted to a wide variety of fields, from science and technology to culture and sports. To the best of our knowledge, no other work attempts to employ the use of either the social media interests of users or their social relationships to aid in the ranking of topics. Moreover, SocialRank undergoes an empirical framework, comprising and integrating several techniques, such as keyword extraction, measures of similarity, graph clustering, and social network analysis. The effectiveness of our system is validated by extensive controlled and uncontrolled experiments.

To achieve its goal, SocialRank uses keywords from news media sources (for a specified period of time) to identify the overlap with social media 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 by an overall measure that combines these three factors.

2. EXISTING SYSTEM

Wartena and Brussee [4] implemented a method to detect topics by clustering keywords. Their method entails the clustering of keywords—based on different similarity measures—using the induced k-bisecting clustering algorithm. Although they do not employ the use of graphs, they do observe that a distance measure based on the Jensen–Shannon divergence (or information radius) of probability distributions performs well.

More recently, research has been conducted in identifying topics and events from social media data, taking into account temporal information. Cataldi et al. [7] proposed a topic detection technique that retrieves real-time emerging topics from Twitter. Their methods uses the set of terms from tweets and model their life cycle according to a novel aging theory.

Additionally, they take into account social relationships—more specifically, the authority of the users in the network—to determine the importance of the topics. Zhao et al. [8] carried out similar work by developing a Twitter-LDA model designed to identify topics in tweets. Their work, however, only considers the personal interests of users, and not prevalent topics at a global scale.

Disadvantages

There is no Information filtering for social computing. There is anonymous topic ranking.

3. PROPOSED SYSTEM

In the proposed system, the method proposed an unsupervised method— SociRank—which identifies news topics prevalent in both social media and the news media, and then ranks them by taking into account their MF, UA, and UI as relevance factors.

The temporal prevalence of a particular topic in the news media is considered the MF of a topic, which gives us insight into its mass media popularity. The temporal prevalence of the topic in social media, specifically Twitter, indicates user interest, and is considered its UA.

Finally, the interaction between the social media users who mention the topic indicates the strength of the community discussing it, and is considered the UI. To the best of our

knowledge, no other work has attempted to employ the use of either the interests of social media users or their social relationships to aid in the ranking of topics.

Advantages

There is effective Topic Identification and keyword extraction.

Efficient Outlier Detection

4. IMPLEMENTATION

- **Admin**

  In this module, the Admin has to login by using valid user name and password. After login successful he can perform some operations such as Authorizing users, Login ,View all users and authorize, give click option to view all users locations in GMap using Multiple Markers ,View all Friend Request and Response ,View all users time line tweet details with Soci rank, rating and give tweet ,View all tweets by clustering based on tweet name and show tweeted details,Soci_Rank,rating and View all Relevant Term Identification on all tweets and group together(similar tweeted details for each and every created tweet) ,View all users outlier detection tweet with its tweeted details,Soci_Rank,rating and View all term frequency on all tweets count(Display the tweets which is getting tweet regularly) based on tweet name, View all tweet news Socirank in chart and View all tweet term frequency count in chart based on date and time, View all tweets tweeted socirank in chart

  Friend Request & Response

  In this module, the admin can view all the friend requests and responses. Here all the requests and responses will be displayed with their tags such as Id, requested user photo, requested user name, user name request to, status and time & date. If the user accepts the request then the status will be changed to accepted or else the status will remains as waiting.

- **User**

  In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user can perform some operations like Register with Location with lat and login using GMap and Login, View Your Profile with location ,Search Friend and Find Friend Request, View all Your Friends Details and Location Route path from Your Location, View all your time line tweets with Soci rank, rating and give tweet, Create tweet for News like Tweet name, tweet uses, Tweet desc(enc),tweet image and View all your tweet with re tweet details,Socirank,rating,Search tweet and list all Tweets and view its details and give re tweet, give rank by hyper link and View all your friends Tweets and give Tweet

  Searching Users to make friends

  In this module, the user searches for users in Same Site and in the Sites and sends friend requests to them. The user can search for users in other sites

to make friends only if they have permission.

5. CONCLUSION

In this paper, we proposed an unsupervised method— SociRank—which identifies news topics prevalent in bothsocial media and the news media, and then ranks them by taking into account their MF, UA, and UI as relevance factors. The temporal prevalence of a particular topic in the news media is considered the MF of a topic, which gives us insight into its mass media popularity. The temporal prevalence of the topic in social media, specifically Twitter, indicates user interest, and is considered its UA. Finally, the interaction between the social media users who mention the topic indicates the strength of the community discussing it, and is considered the UI. To the best of our knowledge, no other work has attempted to employ the use of either the interests of social media users or their social relationships to aid in the ranking of topics.

Consolidated, filtered, and ranked news topics from both professional news providers and individuals have several benefits. One of its main uses is increasing the quality and variety of news recommender systems, as well as discovering hidden, popular topics. Our system can aid news providers by providing feedback of topics that have been discontinued by the mass media, but are still being discussed by the general population. SocialRank can also be extended and adapted to other topics besides news, such as science, technology, sports, and other trends. We have performed extensive experiments to test the performance of SocialRank, including controlled experiments for its different components. SocialRank has been compared to media focus-only ranking by utilizing results obtained from a manual voting method as the ground truth. In the voting method, 20 individuals were asked to rank topics from specified time periods based on their perceived importance. The evaluation provides evidence that our method is capable of effectively selecting prevalent news topics and ranking them based on the three previously mentioned measures of importance. Our results present a clear distinction between ranking topics by MF only and ranking them by including UA and UI. This distinction provides a basis for the importance of this paper, and clearly demonstrates the shortcomings of relying solely on the mass media for topic ranking.

REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” J. Mach. Learn. Res., vol. 3, pp. 993–1022, Jan. 2003.

[2] T. Hofmann, “Probabilistic latent semantic analysis,” in Proc. 15th Conf. Uncertainty Artif. Intell., 1999, pp. 289–296.

[3] T. Hofmann, “Probabilistic latent semantic indexing,” in Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, Berkeley, CA, USA, 1999, pp. 50–57.

[4] C. Wartena and R. Brussee, “Topic detection by clustering keywords,” in Proc. 19th Int. Workshop Database Expert Syst. Appl. (DEXA), Turin, Italy, 2008, pp. 54–58.

[5] F. Archetti, P. Campanelli, E. Fersini, and E. Messina, “A hierarchical document clustering environment based on the induced bisecting k-means,” in Proc. 7th Int. Conf. Flexible Query Answering Syst., Milan, Italy, 2006, pp. 257–269. [Online]. Available:http://dx.doi.org/10.1007/11766254_22.

[6] C. D. Manning and H. Schütze, Foundations of Statistical Natural Language Processing. Cambridge, MA, USA: MIT Press, 1999.

[7] M. Cataldi, L. Di Caro, and C. Schifanella, "Emerging topic detection on Twitter based on temporal and social terms evaluation," in Proc. 10th Int. Workshop Multimedia Data Min. (MDMKDD), Washington, DC, USA, 2010, Art. no. 4. [Online]. Available: http://doi.acm.org/10.1145/1814245.1814249.

[8] W. X. Zhao et al., "Comparing Twitter and traditional media using topic models," in Advances in Information Retrieval. Heidelberg, Germany: Springer Berlin Heidelberg, 2011, pp. 338–349.

[9] Q. Diao, J. Jiang, F. Zhu, and E.-P. Lim, "Finding bursty topics from microblogs," in Proc. 50th Annu. Meeting Assoc. Comput. Linguist. Long Papers, vol. 1. 2012, pp. 536–544.

[10] 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.

[12] 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.

[13] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling, "TwitterStand: News in tweets," in Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst., Seattle, WA, USA, 2009, pp. 42–51.

[14] 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.

[15] 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.

[16] S. Brin and L. Page, "Reprint of: The anatomy of a large-scale hypertextual web search engine," Comput. Netw., vol. 56, no. 18, pp. 3825–3833, 2012.

[17] 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.

[18] K. Kireyev, "Semantic-based estimation of term informativeness," in Proc. Human Language Technol. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguist., 2009, pp. 530–538.

[19] G. Salton, C.-S. Yang, and C. T. Yu, "A theory of term importance in automatic text analysis," J. Amer. Soc. Inf. Sci., vol. 26, no. 1, pp. 33–44, 1975.

[20] H. P. Luhn, "A statistical approach to mechanized encoding and searching of literary information," IBM J. Res. Develop., vol. 1, no. 4, pp. 309–317, 1957.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org