International Journal of Technology (2016) 7: 1239-1245
ISSN 2086-9614

A SOCIAL NETWORK NEWSWORTHINESS FILTER BASED ON TOPIC ANALYSIS

Chaluemwut Noyunsan[1], Tatpong Katanyukul[1], Yuqing Wu[2], Kanda Runapongsa Saikaew[1*]

[1]*Department of Computer Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, 40002, Thailand*
[2]*Department of Computer Science, Pomona College, Claremont, CA 91711, USA*

(Received: August 2016 / Revised: October 2016 / Accepted: October 2016)

ABSTRACT

Assessing trustworthiness of social media posts is increasingly important, as the number of online users and activities grows. Current deploying assessment systems measure post trustworthiness as credibility. However, they measure the credibility of all posts, indiscriminately. The credibility concept was intended for news types of posts. Labeling other types of posts with credibility scores may confuse the users. Previous notable works envisioned filtering out non-newsworthy posts before credibility assessment as a key factor towards a more efficient credibility system. Thus, we propose to implement a topic-based supervised learning approach that uses Term Frequency-Interim Document Frequency (TF-IDF) and cosine similarity for filtering out the posts that do not need credibility assessment. Our experimental results show that about 70% of the proposed filtering suggestions are agreed by the users. Such results support the notion of newsworthiness, introduced in the pioneering work of credibility assessment. The topic-based supervised learning approach is shown to provide a viable social network filter.

Keywords: Credibility measurement; Social media analysis; Topic analysis

1. INTRODUCTION

Social media is user-generated content. It is a stream of various kinds of content posts. Not all content is trustworthy. Many studies attempted to provide credibility measurement on Twitter (Gupta et al., 2014; Aggarwal et al., 2012). The credibility measurement of social media posts can help users to decide whether to believe that a post is true. However, all previous studies ran credibility assessments on all posts. There are various types of posts, e.g., news, announcements, soliciting, personal opinions, 'selfies', status updates, and small talks. Some types are unsuitable for credibility assessment. This practice may lead to user confusion and unnecessary computation by the assessment system. This paper attempts to filter out the posts that are unnecessary for credibility computation before credibility assessment process. We propose an approach to automatically filter out posts that are non-essential for credibility assessment using supervised learning and topic analysis.

Gupta et al. (2014) reported as follows: TweetCred, one of the pioneer deployments on Twitter credibility assessment implemented as a Chrome extension, provides a credibility score ranging from 1 (low credibility) to 7 (high credibility) for every post. They reported that the percentage of tweets that the users agreed with TweetCred score was 42.95%. They further discussed that

[*] Corresponding author's email: krunapon@kku.ac.th, Tel. +6643362160, Fax. +6643362160

Permalink/DOI: https://doi.org/10.14716/ijtech.v7i7.5072

symbol. The notion of posts that do not need credibility scores is not new. Castillo et al. (2011) denoted such posts as conversation/non-newsworthy, and they also employed human workers via Amazon's Mechanical Turk (AMT), to separate non-newsworthy posts from newsworthy ones.

Gupta and Kumaraguru (2012) studied assessment credibility in high impact events. They retrieved tweets via Twitter API and they used the crowdsourcing label data. Additionally, Buckley et al. (1993) used the Pseudo Relevance Feedback (PRF) for classification and they performed these steps to improve the prediction result. Joachims et al. (2002) classified data by using rank Support Vector Machines (SVMs) and re-ordered results by using the Okapi Best Matching (BM25) weighting scheme. BM25 is a 'bag of words' retrieval function Robertson et al. (1999).

Ikegami et al. (2013) investigated assessment credibility about the Great East Japan Earthquake in 2011. They collected tweets and used an opinion classifier. They grouped topics and applied sentiment analysis (opinion mining) to extract subjective information. Topics with high positive opinion rankings were high credibility topics. Kawabe et al. (2015) enhanced this idea by adding expert users to compute credibility.

Alonso et al. (2010) studied interestingness of micro-blogging content. Their definition of interestingness/uninterestingness was similar to newsworthiness/non-newsworthiness (Castillo et al., 2011). Furthermore, Alonso et al. (2010) also employed human workers, AMT to label interestingness of the posts and found that 89% of uninteresting posts did not have any hyperlink. This inspired a use of a simple rule-based system to filter out uninteresting posts.

Later, Yang and Rim (2014) took a topic analysis approach based on a Latent Dirichlet allocation (LDA) topic model, as described by Blei et al. (2003) to develop an automatic system to assess post interestingness. Additionally, Yang and Rim (2014) defined interesting as possibly being "of potential interest to not only the authors and their followers but a wider audience" and uninteresting as being "only interesting to the authors and their friends due to personal interests." However, they associated "general and mundane topics appear any time spans" to uninteresting posts. This notion is clearly seen in the development of their system, but also it sets their perception of interestingness apart from our newsworthiness. For example, common, mundane news, e.g., a missing child, may appear to be uninteresting, but it is newsworthy.

This paper proposes a topic-analysis based approach that employs a simpler topic model, TF-IDF to identify unimportant posts. Our work is distinct from non-newsworthiness filtering in Castillo et al. (2011) in that we developed an automatic filtering system, in contrast to relying on human workers. Our system design and development can be used to analyze data of typical social networks. The system was implemented and experimented on using Facebook data. Facebook (FB) is the largest social network with 1.5 billion users. Retrieving Facebook data is challenging in its own right, when compared with processing data from Twitter, the most studied social network platform in the previous works. While Twitter favors an open environment, Facebook values a privileged sense of selected members. There are many restrictions to access data on Facebook. Therefore, developing a system to analyze Facebook data requires several workarounds and practical contrivances.

2. METHODOLOGY

Figure 1 illustrates the system overview, which consists of the FB filter, implemented as a Chrome Extension, and a server partly connected to a text corpus (a large and structured set of texts or a document in a collection). The FB filter identifies if the post is newsworthy. The FB filter does so by reading the post's content, analyzing associated topics, and comparing cosine similarity of TF-IDF vectors to a corpus. The corpus contains a large number of non-newsworthy posts. If the post under question is similar to any of the non-newsworthy posts, the FB filter will label it as non-newsworthy. If there is no post in the corpus similar to the post under the question, then the post is assumed to be newsworthy.

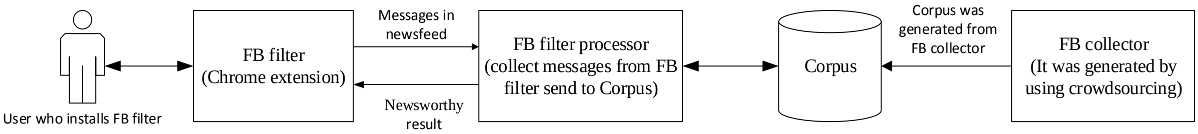

User who installs FB filter
FB filter
(Chrome extension)
Messages in newsfeed
Newsworthy result
FB filter processor
(collect messages from FB filter send to Corpus)
Corpus
Corpus was generated from FB collector
FB collector
(It was generated by using crowdsourcing)

Figure 1 The system overview

To measure post similarity, we first map text-based posts to numerical feature vectors using TF-IDF and then calculate cosine similarity between the corresponding feature vectors of two posts.

TF-IDF can be viewed as a mapping mechanism mapping text-based posts of various character lengths to a fixed-size matrix of numerical values (Salton & McGill, 1983). Each post will be represented as a vector of TF-IDF values. TF-IDF is designed to quantify a message by frequencies of words it contains in relation to other messages. The premise is that words that well represent a message (or an inherent topic of a message) appear with unusually high frequencies in the message compared to their appearance frequencies in other messages.

Given corpus, ={m_1,m_2,..,m_N} , the TF-IDF value, tfidf, of word w_j in message m_i can be calculated from Equation 1:

$$tfidf = \frac{f(m_i,w_j)}{\sum_{t\in Q_i} f(m_i,w_k)} \cdot \log \frac{N}{\sum_{n=1}^{N} \delta(m_n,w_j)}, \tag{1}$$

where $f(m_i,w_j)$ represents a number of occurrence of word w_j in message m_i, Q_i represents the number of distinct words in message m_i, N is the number of messages in the corpus M, and $\delta(m_n,w_j) = 1$ if message m_n contains word w_j and $\delta(m_n,w_j) = 0$ otherwise.

To measure similarity between two TF-IDF feature vectors, the cosine similarity method quantifies the degree of similarity by a cosine value of an angle between two vectors. The degree of similarity between vectors $\vec{p}$ and $\vec{q}$ can be computed as shown in Equation 2:

$$\cos(\theta) = \frac{\vec{p}\cdot\vec{q}}{|\vec{p}|\cdot|\vec{q}|}. \tag{2}$$

The corpus was built by crowdsourcing. A human worker was a person whose age was between 18–60, had studied in a college, and had lived in Thailand. We also asked the crowdsourcing human workers to install a FB collector and a Chrome extension for collecting Facebook messages. The workers were asked to go through Facebook posts. While the workers viewed any post, the FB collector would ask the human workers to decide if the post was newsworthy. The human workers' assessment would be collected into the corpus. After about a month of data collection, we had received 1,886 records from 42 human workers. These records would be used as training data. These records included 881 newsworthy records and 1,005 non-newsworthy records.

After receiving training data, we attempted to find the best cosine similarity to determine whether a message was non-newsworthy. Figure 2 shows the process finding the best cosine similarity in the corpus. First, we set zero to “S” variable which was the cosine similarity. We varied the similar cosine values from 0.1 to 1. Next, the corpus was created by randomly selecting non-newsworthy data and test data was created from non-newsworthy and newsworthy messages. Then, we computed “S_sim” which was cosine similarity between corpus and t1 (message in test data). If “S_sim” was greater than “S”, we classified m1 to a non-newsworthy message. The “S_sim” was the threshold value used to determine whether a message was non-newsworthy or not in our preliminary experiments.

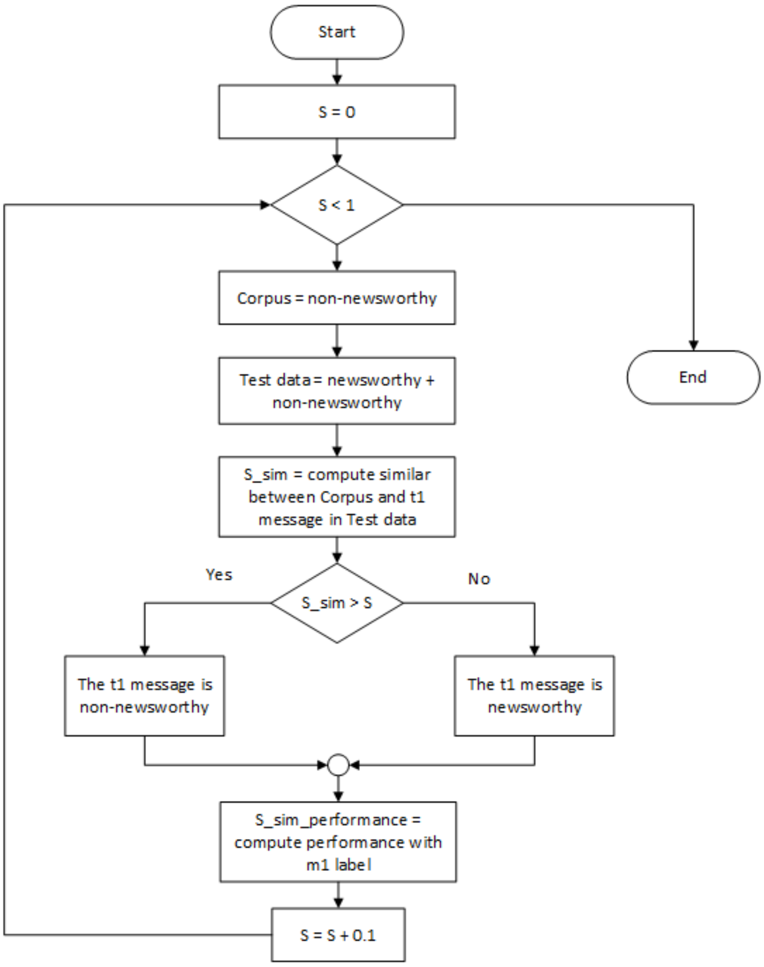

Start
S = 0
S < 1
Corpus = non-newsworthy
End
Test data = newsworthy + non-newsworthy
S_sim = compute similar between Corpus and t1 message in Test data
Yes
No
S_sim > S
The t1 message is non-newsworthy
The t1 message is newsworthy
S_sim_performance = compute performance with m1 label
S = S + 0.1

Figure 2 Flow chart of the FB collector

To determine whether a post was newsworthy or not, the post was compared with non-newsworthy posts in the corpus. The post was assumed to be non-newsworthy when there was at least one similar post found in the corpus. The degree of similarity of the two posts was measured by the cosine similarity of two TF-IDF vectors corresponding to the two posts.

Figure 3 shows the FB filter processing. The vector_message TF-IDF vectors were created from the m1 message. The m1 message was labeled as non-newsworthy if its cosine similarity was greater than 0.5

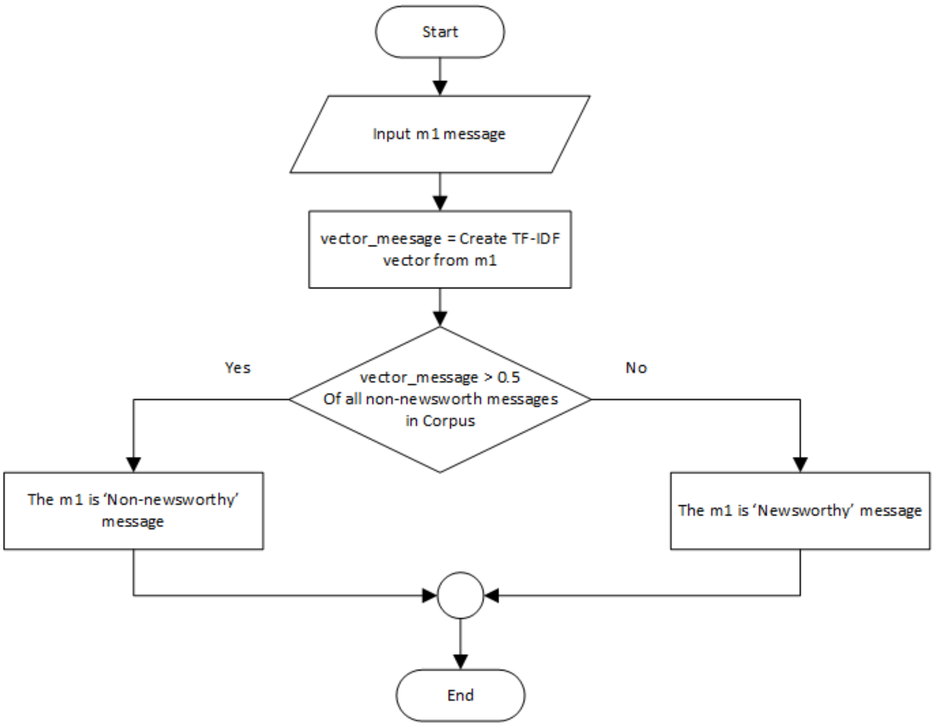
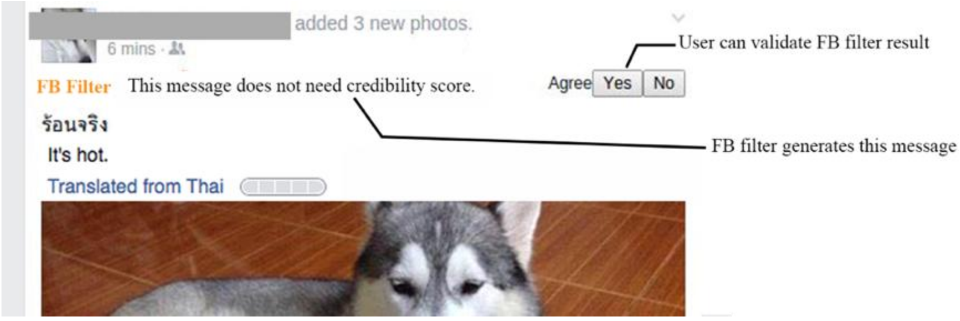

Start
Input m1 message
vector_meesage = Create TF-IDF vector from m1
vector_message > 0.5 Of all non-newsworth messages in Corpus
Yes
No
The m1 is 'Non-newsworthy' message
The m1 is 'Newsworthy' message
End

Figure 3 Flow chart of the FB filter

3. EXPERIMENTAL ANALYSIS

We also employed crowdsourcing to evaluate our non-newsworthiness filter system. The human workers were asked to install an evaluation program, implemented as a Chrome Extension. While going through a Facebook post, the program labeled the post either newsworthy or non-newsworthy and asked if the human worker agreed.



[redacted] added 3 new photos.

6 mins ·

FB Filter This message does not need credibility score.

ร้อนจริง

It's hot.

Translated from Thai

Agree
Yes
No
User can validate FB filter result
FB filter generates this message

Figure 4 Facebook post with FB filter installed

Figure 4 shows an example of a Facebook post with automatic newsworthiness decision and buttons to get validation from a worker.

Out of 1,184 records received from crowdsourcing of 30 human workers, about 70.26% of FB filter results were agreed by the human workers. Table 1 shows a confusion matrix. It should be noted that these 1,184 records were used as test data.

Table 1 Evaluation results

Newsworthiness filter	Agreed by users	Disagreed by users
Newsworthy posts	510	304
Non-newsworthy posts	322	48
Total	832	352

Note that the human workers disagreed with about 304 newsworthiness labels. This large number of mislabeling might be attributed to the fact that the filter was designed to label non-newsworthiness only when it had enough confidence. The filter only searched for a non-newsworthy post that was similar to the post in question. The negative effect of mislabeling a non-newsworthy post was less than that of mislabeling a newsworthy post. Based on the experimental result, human evaluators thought that 52.87% of the posts were non-newsworthy. This supports the results which show non-newsworthy in social media posts of several previous works, including conversation posts of Castillo et al. (2011) and uninterestingness posts of Alonso et al. (2010).

Since the observations occurred in the filter using topic-based approach, it cannot be applied to any post without a text message, such as a post containing only a photo or a post containing only a link. The quality of the filter vitally depends on the variety of collected posts in the corpus. Rather than using a corpus of non-newsworthy posts, an extension to use the corpus of both labels could provide more accurate filtering. Investigation of such extension and issues that may arise, such as how to regulate its behavior according to pre-decided rationale, may lead to a more practical filter. In addition, incorporating unsupervised scheme, such as the Yang and Rim (2014) approach, could lead to a more robust filter.

The experimental analysis reveals that: (1) people agree that some posts should not be assigned credibility scores; and (2) there exists a simple approach to develop a filter for non-newsworthy posts. This kind of filter could and should be a part of preprocesses to credibility assessment. Having a non-newsworthiness filter will allow the credibility assessment system to provide fewer confusing results as well as reduce unnecessary computation that may incur. We also speculate that filtering out non-newsworthy posts may lead to a more efficient credibility assessment of the remaining posts in some systems. For example, a data-driven adaptive system may work better, since filtering may remove noisy data. The implication of the proposed approach, as a part of a more comprehensive system, may assist people to consume more reliable digital content and thus lead to a safer online society.

4. DISCUSSION

Data is central to success of system implementation and evaluation. Regarding data acquisition, accessing non-public Facebook data even from one’s own account is challenging. It requires tweaking, in our case, by using a Chrome extension to capture the HTML source code on currently accessed Facebook data on a web browser. The mitigation requires users’ cooperation: installing our Chrome-extension program. This limits a number of users and amount of data acquired in terms of both quantity and diversity. Therefore, any study on Facebook, though possible, comes with a higher price in terms of data acquisition difficulty.

Regarding the corpus for TF-IDF as well as TF-IDF and cosine similarity were used to implement the proposed system which was heavily based on size and timeliness of a non-newsworthy corpus.

5. CONCLUSION

This paper proposes a supervised approach to filter non-newsworthy posts from social post data streams. Non-newsworthy posts are irrelevant to credibility measurement. We developed a content-based filter system on Facebook using a corpus of non-newsworthy posts, TF-IDF, and cosine similarity. We used crowdsourcing to prepare a corpus and to evaluate the system. The evaluation results indicated that about 70% of non-newsworthiness labels were agreed by human evaluators. In the future, we plan to integrate an effective non-newsworthiness filtering method into a complete credibility assessment system so that system will report credibility scores only for relevant newsworthy posts, which will hopefully reduce users' confusion and lead to a highly efficient system.

6. REFERENCES

Aggarwal, A., Rajadesingan, A., Kumaraguru, P., 2012. PhishAri: Automatic Realtime Phishing Detection on Twitter. *In*: eCrime Researchers Summit (eCrime), IEEE, pp. 1−12

Alonso, O., Carson, C., Gerster, D., Ji, X., Nabar, S.U., 2010. Detecting Uninteresting Content in Text Streams. *In*: Proceedings of the SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation (CSE 2010) - July 23, 2010, Geneva, Switzerland

Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, Volume 3, pp. 993−1022

Buckley, C., Salton, G., Allan, J., 1993. Automatic Retrieval with Locality Information Using SMART. *In*: Proceedings of the First Text REtrieval Conference (TREC-1), pp. 59−72

Castillo, C., Mendoza, M., Poblete, B., 2011. Information Credibility on Twitter. *In*: Proceedings of the 20th International Conference on World Wide Web, March 28–April 1, 2011, Hyderabad, India, pp. 675−684

Gupta, A., Kumaraguru, P., 2012. Credibility Ranking of Tweets during High Impact Events. *In* Proceedings of the 1st Workshop on Privacy and Security in Online Social Media, Article No. 2, Lyon, France, April 17, 2012

Gupta, A., Kumaraguru, P., Castillo, C., Meier, P., 2014. Tweetcred: Real-time Credibility Assessment of Content on Twitter. *Social Informatics–Lecture Notes in Computer Science,* Volume 8851, pp. 228−243. Springer International Publishing Switzerland

Ikegami, Y., Kawai, K., Namihira, Y., Tsuruta, S., 2013. Topic and Opinion Classification Based Information Credibility Analysis on Twitter. *In*: Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, pp. 4676−4681, October 13-16, 2013, IEEE Computer Society Washington, DC, USA

Joachims, T., 2002. Optimizing Search Engines Using Clickthrough Data. *In*: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 133−142, Edmonton, Alberta, Canada, July 23-26, 2002, ACM New York, NY, USA

Kawabe, T., Namihira, Y., Suzuki, K., Nara, M., Sakurai, Y., Tsuruta, S., Knauf, R., 2015. Tweet Credibility Analysis Evaluation by Improving Sentiment Dictionary. *In*: 2015 IEEE Congress on Evolutionary Computation (CEC), pp. 2354−2361

Robertson, S.E., Walker, S., Beaulieu, M., 1999. Okapi at TREC-7: Automatic Ad Hoc, Filtering, VLC and Interactive Track. In E.M. Voorhees & D.K. Harman (Eds), *The Seventh Text REtrival Conference (TREC-7)*, pp. 253−264, Gaithersburg, MD: U.S. Department of commerce, NIST (National Institute of Standards and Technology)

Salton, G., McGill, M.J., 1983. *Introduction to Modern Information Retrieval*, Mcgraw-Hill College

Yang, M-C., Rim, H-C., 2014. Identifying Interesting Twitter Contents using Topical Analysis. *Expert Systems with Applications*, Volume 41(9), pp. 4330−4336