

The 4th International Conference on Electrical Engineering and Informatics (2013)

Automatic Summarization of Tweets in Providing Indonesian Trending Topic Explanation

Yosef Ardhito Winatmoko*, Masayu Leylia Khodra

*School of Electrical Engineering and Informatics, Institut Teknologi Bandung
Bandung 40132, West Java, Indonesia*

Abstract

Trending topic is one of Twitter's features to provide its users with the current discussion theme and hold big potential to give recent news and information. However, Twitter only gives a list of topics and often people do not understand what the meaning of each topic is. This paper will describe the characteristic of trending topic and try to automatically give explanation about trending topic according to their tweet collections. Although several tweet summarization techniques exist now, none of them is focusing on trending topic. We evaluate sentence similarity as indicator for trending topic's category and propose a new method to automatically generate explanation for trending topic. We found that there is a correlation between trending topics' characteristic with the explanation generation method.

© 2013 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Selection and peer-review under responsibility of the Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia.

Keywords: Twitter; Summarization; Trending Topic; Explanation

1. Introduction

Massive number of tweet received by Twitter each day had been exploited as news and event detection source [1][2]. For this information spreading purpose, Twitter provides its users with a list which contains 10 current discussion themes of tweets from any regions, called trending topic. Trending topic can be a word, phrase, or hashtags, which is a combination of word preceded by a number sign (#). The problem of trending topic is that this list is not accompanied by explanation about topics included whilst it can give insight about breaking news and interesting detail. We can see from Fig. 1a that it is quite hard to determine what several topics meant. For example,

* Corresponding author.

E-mail address: yosefardhitowin@gmail.com

the last topic in the list, “Habibie & Ainun”, was actually a film played in cinema and had received positive tweets. People were discussing about how the film inspire them and encouraged who had not seen the film yet. By providing this kind of explanation, one should find trending topic more useful and informative. People who previously have not heard anything about “Habibie & Ainun” would read the explanation and might then become excited to watch the film.

Currently, Twitter collaborates with third party application called WhatTheTrend (<http://whatthetrend.com>) to give explanation about trending topics. This application depends on users' help to provide manual summary for every topic. Because of the manual approach, many topics do not get any explanation. Another weakness of this approach is the generalization of topics across all regions. This means every topic will get the same explanation beside they appear in different region. In example, “Evan Almighty” was one of trending topic in Indonesian region, but WhatTheTrend used the same explanation as when “Evan Almighty” was listed as worldwide trending topic, shown in Fig. 1b. The given summary is not relevant to the exact reason why “Evan Almighty” had become Indonesian trending topic, which was actually because the film was airing in Indonesian television. Despite of the presence of WhatTheTrend as the source of explanation for topics, users often still have to manually search tweets and read it one by one to find out the meaning of any topics. People will surely be exhausted if they had to read all of the searching results. On the other hand, Automatic summarization can be used to give comprehensive explanation and help users understand the meaning of trending topics.

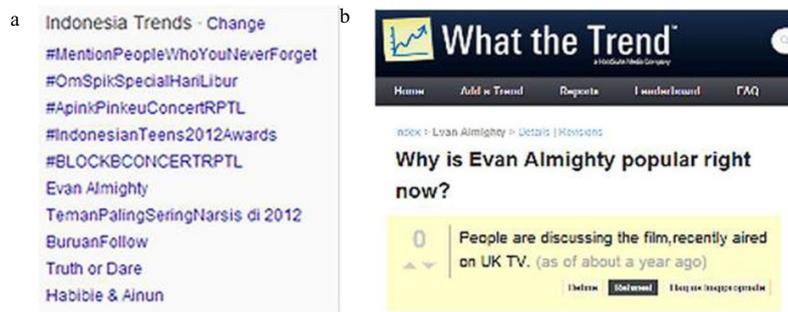


Fig. 1. Example for (a) trending topic from Indonesian region; (b) WhatTheTrend explanation of a trending topic

There are already attempts in summarizing tweets but none of those is focusing on trending topics. By taking trending topic's characteristic into account, people might get better summary for explanation. Our research aims to determine what kind of trending topic needs explanation and how to generate summary of tweets for topic's explanation. We use sentence similarity as indicator for trending topic's characteristic. After we find relevant topics, we adapt existing tweet summarization methods and sentence compression technique to build explanation. We choose to build explanation for Indonesian trending topic which none similar research exist now. Adding other work in this area [3][4], our main contributions are: (1) Evaluating the performance of sentence similarity as a feature to identify trending topics; (2) Implementing the current tweet summarization method to give explanation for identified trending topic.

In the following section, we present related works in the same area (Sec. 2) and different kind of trending topic (Sec. 3). We then define our approach to automatic trending topic explanation (Sec. 4). The rest of the paper includes evaluation (Sec. 5) and conclusion remarks (Sec. 6).

2. Related work

In this section, we discuss relevant experiments about tweet summarization and how well they perform. Despite the absence of trending topic analysis in these related works, summarization is the primary focus in providing explanation for topics. Inouye and Kalita [5] had evaluated previous works in text summarization applied for tweet summarization. Among these works were SumBasic [6], MEAD [7], Phrase Reinforcement (PR) [3], and Hybrid TF-IDF [4]. All of these algorithms will be described in the following paragraph.

SumBasic is a notable text summarization algorithm which performs well in tweet summarization task. The primary idea of SumBasic comes from the tendency of manual summary to include more frequent words in documents as part of the summary. In contrast to SumBasic, MEAD and other clustering approach does not perform as well as expected for tweet summarization. Phrase Reinforcement (PR) is indeed quite unique algorithm because it use word graph to summarize tweets. The main concept behind PR is to determine and combine partial summaries from tweets. Another summarization algorithm specifically for tweets is hybrid tf-idf, which based on tf-idf weighting method. The “hybrid” keyword comes from the adaptation of tf-idf weighting for short documents, particularly in its different definition of one document when calculating tf and idf value of terms. The performance of hybrid tf-idf comes second to SumBasic in ROUGE scores, but slightly better in human evaluation.

Other methods in tweet processing are TwitterStand [1] and TweetMotif [2]. TwitterStand does not explicitly mention summarization, but their intention to find current event and provide several representative tweets are very similar to tweet summarization task. In contrast, TweetMotif explicitly mention that their intention is to summarize topic. However, they do not limit their topic to only trending topic and are focusing on topic modelling.

3. Trending topic characteristic

According to Crane and Sornette [8], there are two factors which affect topics in social media. The first factor is where the topic comes from; they can be either endogenous or exogenous. Another factor is people’s reaction to the topic, critical or sub-critical. Twitter as a social system also has this kind of characteristic [9]. Because of the two factors described above, there are four kind of topic period:

1. Exogenous critical: rarely discussed and appear on the trending topic as a result of a certain event and people tend to join the discussion. An example of topic belong to this category is ‘Habibie & Ainun’.
2. Exogenous subcritical: this kind of topic is also rarely discussed and may become trending topic because of an event. The primary difference with the previous category is the reaction of people. People are not that excited to join the discussion as the previous category. One example of this category is '#backintheday'.
3. Endogenous critical: the third category is topics which usually become lasting topic and may become trending topic because incidentally many people discussed this topic together. Endogenous critical topic example is 'Manchester United'.
4. Endogenous subcritical: the last category is those topics seldom discussed by Twitter user and might become trending topic for no particular reason. There is no peak point of talking (i.e., burst point) in this category. Example for this category is 'facebook'.

Dealing with trending topic is rather trivial for human. Some trending topics are self-explaining or need not any explanation. However, this is a difficult task for machines with existing NLP tools because trending topic, similar with tweets, often does not follow formal language rules. However, we can use the predefined four categories of trending topic to determine which of those are worthy to be explained. Critical topics are definitely need explanation because they have particular event which encourages people to talk about it. Subcritical topics do not attract that much of attention, only those exogenous topic need an explanation of why it suddenly become trending topic while rarely become discussion theme. Therefore, topics which fall into endogenous subcritical do not need explanation, but the others are suitable for explanation.

4. Providing trending topic explanation method

Twitter also defines several features to optimize conversation functionality such as hashtag, mention, and retweet. In addition, orthographic and syntactic errors are also commonly found in tweets. This unique feature problem might be solved with text normalization technique such as machine translation [10]. However, these kinds of technique require decent NLP tools and good training sets which is now still unavailable for Indonesian language. Our proposed method consists of 4 processes visualized by Fig. 2.

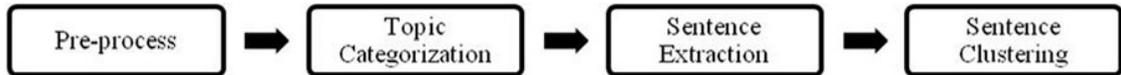


Fig. 2. Processes and flow of proposed trending topic explanation method

Pre-processing Twitter features is important especially if we want to do automatic summarization because we will use parts of the tweet as summary. To overcome this tweet-specific text problem, we propose several steps in pre-processing Twitter feature as follow:

- RT: retweets are very common in Twitter. This feature enables users to somehow “forward” other user's tweet and also sometimes give additional comment. To produce novel summary, we have to erase retweets format and divide retweets from its additional comments. Therefore, one tweet may result more than one sentence.
- Mention: this feature is very common in Twitter and may hold significant meaning about the relevant topic. However, we still strip this feature in identification of trending topic category because they may disturb the calculation of sentence similarity.
- Topic: as we have mentioned above, topic are not only word or hashtag, but also phrase. Tokenization of topic in tweets will make phrases become separated. In addition, the topic itself needs special treatment because they will appear in all of the tweets related to that topic. To prevent this problem, we replace topics inside tweets with one predefined token, such as: “__TOPIC__”. This enables algorithm to treat topic as one token.
- Emoticon: emoticon holds no significant meaning in detecting important sentence and might disturb the legibility of summary produced. We erase any emoticon appear in tweets.
- URL (Uniform Resource Locator): people usually tweet a shorter version of longer documents and attach the complete document as URL in their tweets. Twitter will shorten this URL independently although they might refer to same page. This feature does not have any importance in summarization and thus can be stripped from tweets.

We try to find what indication may approximate which category a topic belongs to. We need to approximate because Twitter API limit the number of received tweets for each topic. One major characteristic is the similarity between the sentences which discussed the topic. Tweets which talk about endogenous subcritical topic tend to have very low similarity because everyone actually discusses about different aspects of the topic or the topic itself is just a very common term. To prove this hypothesis, we use vector space model technique to represent each sentence and calculate the similarity within each sentence using cosine similarity method. The average similarity of given set of sentences will then evaluated if they do indicate the category of the topic. In conclusion, topics which result very low similarity do not need an explanation and therefore the following processes are not necessary.

According to the aforementioned performance evaluation by [5], sumbasic and hybrid tf-idf perform well for tweet summarization. Both sumbasic and hybrid tf-idf need length parameter for input, i.e. how many sentences they have to extract as the summary. We apply sumbasic and hybrid tf-idf algorithm to produce several sentences which are the most representative of topic. Rather than giving reader all of those n sentences as summary, we try to cluster these representative tweets first into subtopics. In order to do this, we use tf-idf weighting and distance-based method, i.e. cosine similarity, to cluster similar sentence. In addition, we need to define a specific threshold of which can differentiate whether one sentence is similar to the centroid of a cluster or not. If a sentence does not satisfy the given threshold for all available centroid, it will be used to form a new cluster.

Cluster of subtopics may contain any number of sentences. This might also be used to filter irrelevant sentences by defining a threshold for minimum number of member of a cluster, i.e. a cluster whose number of member is below given threshold, they are not representative or even irrelevant sentences. Therefore we will get a number of clusters which contains only a number of similar sentences above the given threshold. These clusters will be then represented as one sentence using sentence compression described in [11]. However, the algorithm described in [11] requires a Part-of-Speech (POS) tagger and standard tagger will not perform well for tweets as they do not follow formal language rules. Therefore, we modify the algorithm from which previously might differentiate two same words whose POS are different into assuming that they are all have the same POS. This is a safe modification because tweets usually use only popular words.

5. Evaluation

5.1. Experiment

First of all, we will prove our hypothesis about sentence similarity significance as trending topic's category indicator. We collected 8 Indonesian trending topics in 1 – 2 January 2013, each with corresponding 300 random tweets. Among 8 topics, 3 of them do not need explanation: Follback, #OmSpikSpecial2013, and #nowplaying. And then, we apply pre-process as described above and try to calculate the average sentence similarity for every trending topic. We will try to find the correct threshold for which each topic can be correctly categorized, if such threshold exists. Table 1 shows the average similarity result for each topic in dataset.

Table 1. 8 topics used for dataset

Topic	Need Explanation?	Average Similarity (%)	Vocabulary Size
Follback	No	10.33	309
#HBD27thMitaTheVirgin	Yes	10.01	371
#10BuahFavoritGue	Yes	7.90	223
Arisan 2	Yes	6.01	618
Jakarta	Yes	4.28	1129
Habibie & Ainun	Yes	3.9	651
#nowplaying	No	2.74	1248
#OmSpikSpecial2013	No	0.82	98

The second experiment scenario is to generate explanation for some topics from the tweet collection. For evaluation purpose, we define manual summary for those topics. The manual summary is built through combining 2 factors as indicated by trending topic's characteristic: what is the meaning of the topic and what is the reason for people tweeting about them. We assume tweet collection as the only information source. Because 3 topics do not need explanation as mentioned in the previous paragraph, there are 5 topics which will be used for the explanation experiment. For example, topic “#HBD27thMitaTheVirgin” explanation is:

@mitaenglandmuse merayakan ulang tahun. Tweet terkait ucapan selamat ulang tahun yang diberikan oleh pengguna lain. [*@mitaenglandmuse is celebrating birthday. Most of the tweets are saying happy birthday.*]

We use evaluation calculation as described in [5] and first try to find the optimum value of similarity threshold used in clustering and number of extracted sentences from sumbasic and hybrid tf-idf algorithm. Fig. 3 shows test results where x-axis is the similarity threshold in percent, i.e. 100 means completely similar, and y-axis is the result f-measure. Fig. 4 also shows test results where x-axis is the amount of sentences extracted.

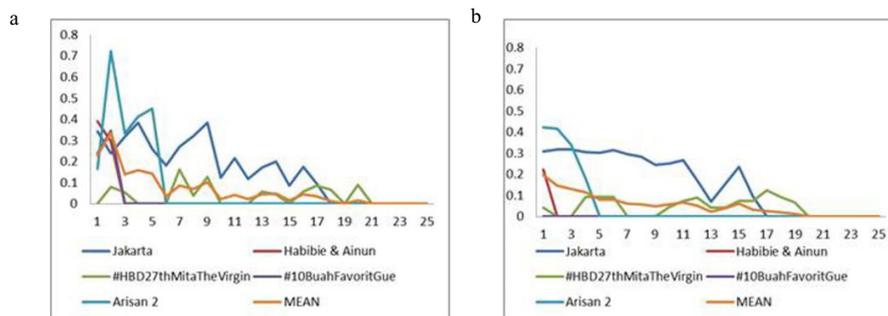


Fig. 3. F-measure result for clustering similarity threshold with (a) sumbasic and (b) hybrid tf-idf

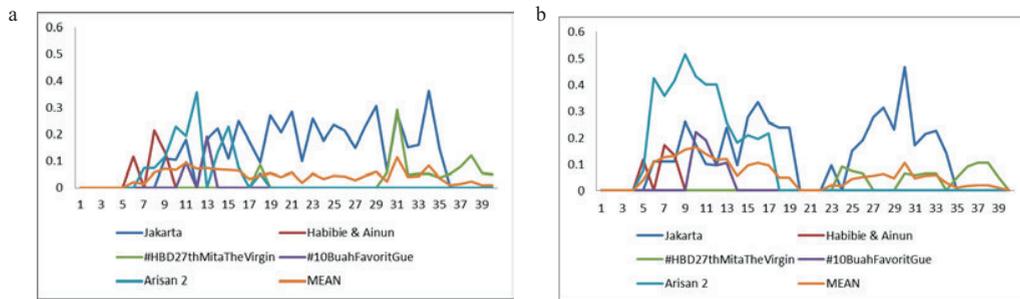


Fig. 4. F-measure result for number of extracted sentences threshold with (a) sumbasic and (b) hybrid tf-idf

From Fig. 3 we can approximate the optimum clustering similarity for sumbasic is 0.02 and hybrid tf-idf is 0.01. In addition, Fig. 4 shows that the the highest f-measure mean for number of sentence is 31 for sumbasic and 10 for hybrid tf-idf. Therefore, we used the given optimum threshold and try to evaluate the output explanation. Table 2 shows evaluation scores for the summary result.

Table 2. Explanation for trending topics evaluation scores

Topic	Sumbasic				Hybrid TF-IDF			
	precision	recall	F1	length	precision	recall	F1	length
#HBD27thMitaTheVirgin	0.04	0.07	0.05	26	0	0	0	12
#10BuahFavoritGue	0.18	0.13	0.15	12	0.33	0.07	0.11	5
Arisan 2	0.27	0.23	0.25	13	0.23	0.23	0.23	15
Jakarta	0.23	0.50	0.32	31	0.24	0.36	0.29	22
Habibie & Ainun	0.30	0.48	0.37	39	0.36	0.17	0.24	11
Mean	0.20	0.28	0.23	24	0.23	0.17	0.17	13

5.2. Analysis

From the experiment about determining topic category with sentence similarity we can conclude that there are 3 regions of topic character according to their average similarity:

1. Topics with average similarity < 3%. This is quite understandable because these topics contain sentences with very broad subtopics and therefore no specific relevant summary may be produced.
2. Topics within range 3 - 10% are topics which need explanation. We may say that topics with this amount of similarity may be considered a normal trending topic.
3. Topics with average similarity > 10% are topic which is not a discussion. Follback does not need explanation but #HBD27thMitaTheVirgin does. This might be a result of tweets about Follback is usually very short, therefore resulting very high similarity. Also, #HBD27thMitaTheVirgin is quite a self-explaining topic but still may produce an explanation with informative detail. However, users tend to tweet only “#HBD27thMitaTheVirgin” because the topic itself already explain the user's intention to tweet.

From the explanation generation experiment we can see that the proposed method do not perform as good as expected. This might be caused by the incomplete noisy text pre-processing and still leave many language inconsistencies. For example, we leave foreign language in tweets because they are minor errors and will then be filtered. However, foreign languages still appear in summary because some common phrases are frequently used for some topic. One example is for topic “#HBD27thMitaTheVirgin” where “Happy birthday” appears in the explanation.

Hybrid tf-idf produce averagely shorter summary than sumbasic, therefore they score higher precision and lower recall naturally. This characteristic may also cause extracted sentences from hybrid tf-idf to be totally different with the manual summary as shown with topic “#HBD27thMitaTheVirgin”. However, sumbasic’s score for the same topic also not that better. If we connect this to the sentences similarity calculated in the previous experiment, we might conclude that highly similar sentences tend to result worse explanation. As we have explained before, there are 2 components of manual summary. We can infer that similar sentences may only discuss part of the component or maybe not at all, producing irrelevant explanation.

6. Conclusion

We have proposed a method to automatically give explanation about trending topic in Twitter. There are 2 main parts of the method which are the topic categorization and explanation generation. The identification experiments give us insight about sentence similarity and concluding 3 regions of which a topic might belong to. However, it still needs improvement because the exact region is not identified yet. More topics with corresponding category can be used to determine the correct region of normal trending topic.

The second main part is the sentence generation which uses distance-based clustering to make subtopics and generate the explanation using sentence compression. The main problem will be the noisy texts which cause bad evaluation of automatic summarization algorithm. This leaves some room for improvement either on pre-processing stage or evaluation method. Also, we have concluded that topic with highly similar tweets give worse score for explanation due to the incomplete explanation. There should be method modification for such topics so the extracted sentences also cover required components to make a decent explanation.

References

- [1] Sankaranarayanan, J., Samet, H., Teitler, B. E., Lieberman, M. D., & Sperling, J. Twitterstand: news in tweets. Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM; 2009. p.42-51.
- [2] O’Connor, B., Krieger, M., & Ahn, D. Tweetmotif: Exploratory search and topic summarization for twitter. Proceedings of ICWSM; 2010. p. 2-3.
- [3] Sharifi, B., Hutton, M. A., & Kalita, J. Summarizing microblogs automatically. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics; 2010. p. 685-688.
- [4] Sharifi, B., Hutton, M. A., & Kalita, J. K. Experiments in microblog summarization. In Social Computing (SocialCom), 2010 IEEE Second International Conference on IEEE; 2010. p. 49-56.
- [5] Inouye, D., & Kalita, J. K. Comparing twitter summarization algorithms for multiple post summaries. In Privacy, security, risk and trust (passat), 2011 IEEE third international conference on and 2011 IEEE third international conference on social computing (socialcom) IEEE; 2011. p. 298-306
- [6] Radev, D., Allison, T., & Blair-Goldensohn, S. MEAD-a platform for multidocument multilingual text summarization. Proceedings of LREC; 2004.
- [7] Nenkova, A., & Vanderwende, L. The impact of frequency on summarization. Microsoft Research, Redmond, Washington, Tech. Rep. MSR-TR-2005-101; 2005.
- [8] Crane, R., & Sornette, D. Robust dynamic classes revealed by measuring the response function of a social system. Proceedings of the National Academy of Sciences, 2008; 15649-15653.
- [9] Kwak, H., Lee, C., Park, H., & Moon, S. What is Twitter, a social network or a news media? Proceedings of the 19th international conference on World wide web; 2010. p. 591-600.
- [10] Kaufmann, M., & Kalita, J. Syntactic normalization of Twitter messages. In International Conference on Natural Language Processing, Kharagpur, India; 2010.
- [11] Filippova, K. Multi-sentence compression: Finding shortest paths in word graphs. In Proceedings of the 23rd International Conference on Computational Linguistics (pp. 322-330). Association for Computational Linguistics.; 2010.