Predicting the scale of trending topic diffusion among online communities

Dohyeong Kim[1], Soyeon Caren Han[2], Sungyoung Lee[1], and Byeong Ho Kang[2]

[1] Department of Computer Engineering
Kyung Hee University, Giheng-gu,Youngin, Korea
{dhkim,sylee}@oslab.khu.ac.kr
[2] School of Engineering and ICT
University of Tasmania, Sandy Bay, 7005, Tasmania, Australia
{Soyeon.Han, Byeong.Kang}@utas.edu.au

Abstract. Online trending topics represent the most popular topics among users in certain online community, such as a country community. Trending topics in one community are different from others since the users in the community may discuss different topics from other communities. Surprisingly, almost 90% of trending topics are diffused among multiple online communities, so it shows peoples interests in a certain community can be shared to others in another community. The aim of this research is to predict the scale of trending topic diffusion among different online communities. The scale of diffusion represents the number of online communities that a trending topic diffuses. We proposed a diffusion scale prediction model for trending topics with the following four features, including community innovation feature, context feature, topic feature, and rank feature. We examined the proposed model with four different machine learning in predicting the scale of diffusion in Twitter Trending Topics among 8 English-speaking countries. Our model achieved the highest prediction accuracy (80.80%) with C4.5 decision tree.

Keywords: Twitter, Twitter Trending Topics, Information Diffusion, Trending Topic Diffusion, Diffusion Prediction

1 Introduction

By using different types of web-based services, such as search engines, social media, and Internet news aggregation sites, internet users can share and search information throughout the world. These web services, including Google, Yahoo, and Twitter, analyse their social data and provide a Trending Topics service, which displays the most popular terms that are discussed and searched within their community. One of these web services, Twitter, monitors their social data, detects the terms (including phrases and hash-tags) currently most often mentioned by their users, and publishes these on their site. Abdur Chowdhury[1], a

[1] Twitter,Inc. 2009 https://blog.twitter.com/2009/top-twitter-trends-2009

chief scientist at Twitter Research Team, defined the Twitter Trending Topics as the objects showing us that people everywhere can be united in concern around important events. Trending topics are estimated to reflect the real-world issues from the people's point of view. Kwak et al.[13] demonstrated that over 85% of trending topics in Twitter are related to breaking news headlines, and the related tweets of each trending topic provide more detailed information of news and users' opinions. Hence, being able to know which topics people are currently most interested in on Twitter, and their point of view, may lead to opportunities for analysing the market share in almost every industry or research fields, including marketing, politics, and economics.

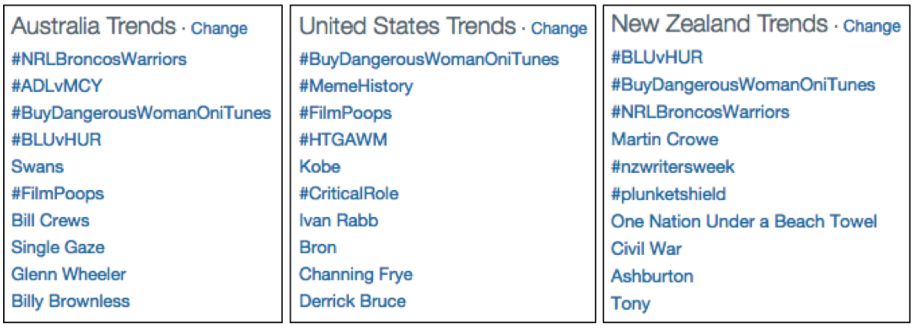

Fig. 1. Twitter Trending Topics for different countries

The 'Twitter Trending Topics' service includes the top 10 trending topics for different locations. This list enables recognition of the degree of current popularity of that topic in a specific geographic location, from individual cities, to countries, and worldwide[2]. Figure 1 shows the top 10 trending topics for Austaralia, United States, and New Zealand at 4pm, 11th March 2016. The trending topics in one country's community are different from other countries' communities. This is because peoples' interests differ across countries. The Trending Topics list for each country includes their local events and worldwide events.

The interesting phenomenon is that many trending topics tend to diffuse among multiple countries. As can be seen in Figure 2, we found that over 90% of trending topics for each country appeared in different countries trending topics list. 92.27% of the U.Ks trending topics also appeared in at least two different countries' trending topic lists, while only 7.73% of them appeared solely in the U.Ks trending topics list. This shows that the majority of trending topics appeared and diffused in multiple countries' trending topics. For example, on July 17th 2014, when a missile downed the Malaysia Airlines plane over Ukraine, it was breaking news around the world. During this time, the topic '#MalaysiaAirlines' initially appeared on the Malaysian trending topics list. After that,

[2] Twitter,Inc. 2014 https://support.twitter.com/articles/101125-faqs-about-trends-on-twitter

Twitter users in the Philipines, Singapore, Australia, and New Zealand started talking about the trending topic. Finally, 3 hours after starting in the Malaysian trending topics list, all 8 English speaking countries were discussing the topic '#MalaysiaAirlines' because it is world breaking news. If the trending topic is just about local events, it normally diffuses to less than two countries.

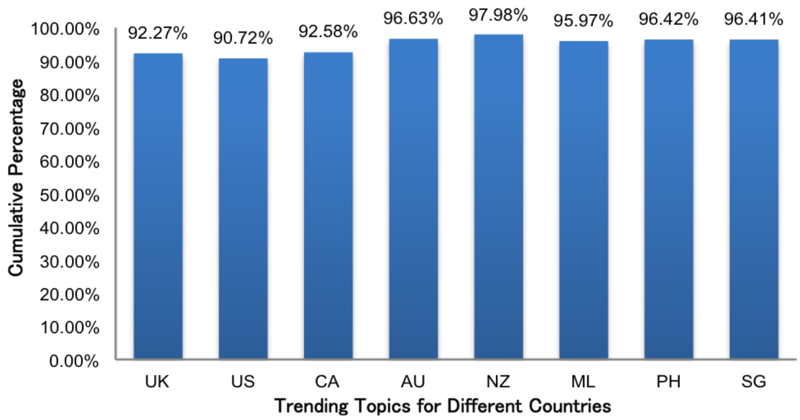

Fig. 2. Pecentage of trending topics that diffused to multiple countries

The number of countries that trending topics diffuse to can be the barometer of people's interests in trending topics. We defined the number of diffused countries as the scale of diffusion for trending topics. Predicting the scale of diffusion for trending topics can be helpful to identify the influence of the topic in the near future.

However, the 'Trending Topics' list displays only limited information, including the trending topic term, its rank, location, and updated date and time. We used only this available information to build the diffusion scale prediction model for trending topics among different online communities. Therefore, the reserach aim of our studty is to answer the following question: "How can we predict the scale of diffusion for trending topics with the available, limited information?" In this paper, we focused on finding the important features that can be related to the scale of trending topics, and building the successful prediction model using machine learning techniques.

The paper is structured as follows: Section 2 presents the related work, followed in Sections 3 and 4 by the methodology for collecting trending topics and building the Diffusion Scale Prediction Model for trending topics. In Section 5, we describe the evaluations conducted, and discuss the results. Finally, we conclude this paper in Section 6.

2 Related Works

In order to build the information diffusion model for the online trending topics, we conducted the literature review in several related research areas, including online trending topics and information diffusion. This section covers how the trending topics are extracted from the online environment, how trending topics are used in the real world, and the different types of information diffusion model proposed in other research areas.

Trending Topic Extraction in Online Environment Trending topics are the topic or issue that a great number of people is talking about in a certain period of time. In the early stage, the trending topics tend to be extracted from newspapers or paper-based documents. In 1992, Andersen et al.[2] proposed JASPER, which extracts the trending topics from newspapers and financial reports in order to support decision-making in a finance environment. Since 2000, Internet users can share and search information by using different types of web-based services and applications, such as Google, Yahoo, Twitter, or Weibo[9]. Those web-based services found that user activity data represents the peoples interests in the certain period, so they started detecting the peoples interests and presenting the most discussed and searched topics, Trending Topics, from their users data. For example, Twitter, one of the most popular social media websites, analyses postings from more than 41 million users and provides the real-time top 10 trending topics of different regions from cities to worldwide[13]. Weibo, a popular Chinese social media website, also presents the top 10 popular topics with the number of searches[4]. In addition to social media, search engines such as Google and Baidu, analyse their users search activities, and share the list of daily popular search keywords that represents peoples interests on a day[8]. Since people tend to make a decision relying on the mainstream interests and behaviors, the Trending Topic services received a lot of attention from researches and industries.

Research using Topics in Online Environment Many researchers applied various summarisation and extraction approaches aimed at revealing the exact meaning of trending topics. Sharifi [17] applied a phrase reinforcement algorithm to summarise related tweets of Twitter Trending Topics. Then, the author conducted evaluation for comparing hybrid TFIDF and phrase reinforcement in use of Trending topics summarising. Inouye [11] also conducted an experiment to compare twitter summarisation algorithms. They found that simple frequency-based techniques produce the best performance in tweets summarisation. Han and Chung [6] applied simple Term Frequency approach for extracting the representative keywords to disambiguate the approach. They also proved that the most successful approach to reveal the exact meaning of trending topics is simple Term Frequency, which is evaluated by 20 postgraduate students[7]. Some researchers examined classifying trending topics. Lee et al.[14] classifies trending topics into 18 general categories by labeling and applying machine-learning techniques. Zubiaga et al. [18] aimed to classify trending topics by applying several proposed features and used SVM to check the accuracy. Han et al.[10] proposed a temporal modeling framework for predicting rank change of trending topics using historical rank data. The proposed model achieved extremely high prediction accuracy (94.01%) with a C4.5 decision tree.

Information Diffusion Modelling We conducted reviews of several Information Diffusion Models used in predicting product diffusion, blog posting diffusion, and social media posting diffusion. The early diffusion models are derived from the Bass diffusion model [12], which calculates the diffusion using the relationship between the current adopters and potential adopters with a new product interaction. Rogers published diffusion of innovations to against a typographical error of Bass paper, which did not cover how, why, and at what rate new ideas and technology spread of a new idea. According to Roger [16], diffusion is a process by which an innovation is communicated through certain channels over time among members of a social system, and the member can be classified into five categories, including innovators, early adopters, early majority, late majority, and laggard.

3 Trending Topics Monitoring

In this paper, we focused on predicting the scale of trending topics diffusion across multiple countries. For our studies, we collected Twitter Trending Topics from 8 different countries communities, including U.S., U.K., Australia, New Zealand, Canada, Malaysia, Philippine, and Singapore. Twitter provides an API (Application Programming Interface) that allows developers or researchers to crawl and collect the data easily. Through this API service, we collected twitter trending topics for 3 years (until 31th December, 2015)

3.1 Trending Topics

Twitter monitors all users data and detects the popular trending topics that most people are currently discussing about. The detected popular trending topics are displayed on the service 'Twitter Trending Topics'. This trending topic service is located on the sidebar of Twitter interface by default so it is very easy for users to check the current trending topics and discuss about it. It provides top 10 trending topics in real time. Hence, we have collected those top 10 trending topics per hour using Twitter API. In total, we have collected 705354 unique trending topics in 3 years.

3.2 Related Tweets

Trending topics in Twitter consist of short phrases, words, or hash-tags. Twitter never provides any detailed explanation of trending topics so it is very difficult to identify the meaning of trending topics until you have a look related tweets of those topics. For example, when a missile destroys Malaysian Airlines, the trending topics were 'Malaysia Airlines', 'Malaysian', etc. It is almost impossible to realise what happened to the Malaysia Airlines by only checking the trending topics. In order to reveal the exact meaning of each trending topic, we need to collect not only the trending topics, but also the appropriate related tweets of a specific trending topic, and the related tweets should not contain contents that are irrelevant. If the trending topic is Malaysia Airline which is about a missile attack that happened on July 18th, we should not collect the related tweets about a missing Malaysia Airliner that occurred on March 8th. It is extremely important to distinguish the tweets that are related to specific trending topics. Twitter API provides the tweet/search crawling service that allows users to collect the tweets by using the search query. The concept of the tweet/search service is the same as a search engine. Users can search the tweets that contain the search keyword. The search results contain detailed information about each tweet, including content, username, location, created date-time, etc. We used the created date-time to extract the appropriate tweets for the trending topics. As we collect the top 10 trending topics on an hourly basis, we search and collect the related tweets that users upload in the last one hour. For example, when Malaysia Airline is on the trending topics list at 8pm, we search and collect the related tweets that users upload in the previous one hour, 7pm to 8pm. This collection approach prevents irrelevant tweets.

4 Trending Topics Diffusion Prediction Model

The goal of this research is to predict the scale of trending topics diffusion across multiple countries. We propose a model that predicts the number of countries the trending trending diffuses by using four different trending topic features and machine learning technique. The proposed model can be described using the following equation:

$$Scale(T_x) = ML(DP(T_x)) \tag{1}$$

$$DP(T_x) = [CI(T_X), CT(T_X), TP(T_X), RK(T_X)] \tag{2}$$

In order to predict the scale $Scale$ of the trending topic T_x diffusion, we extracted four different features of the topic T_x for the diffusion prediction DP model. As can be seen in equation 2, the four features of the specific trending topic T_x include community innovation level $CI(T_X)$, context feature $CT(T_X)$, topic feature $TP(T_X)$, and rank feature $RK(T_X)$. Then, machine learning techniques ML are applied for learning our model. $Scale$ represents the number of countries the trending topic T_x diffuses to. The outcome/result will be the number, from 1 to 8. For example, if the prediction is 1, it identifies that there will be no diffusion to other countries.

$$Scale(T_x) = \begin{cases} 1, \text{will be no diffusion} \\ ... \\ 8, \text{will be diffused to all countries} \end{cases} \quad (3)$$

4.1 Community Innovation Level Feature

Community innovation level feature of the trending topic determines an innovation level of online community that the topic initiates. We modified the concept of innovation of diffusion approach proposed by Roger [16]. This feature classifies communities based on the level of the community adopting the trending topic. There are four types of levels: 1) Innovator: Communities that start diffusing the trending topics, 2) Early Adopter: Communities that adopt the diffused trending topics in the early stage, 3) Late Adopter: Communities that adopt the diffused trending topics after the average participant, and 4) Laggards: Communities that are the last to adopt the diffused trending topic.

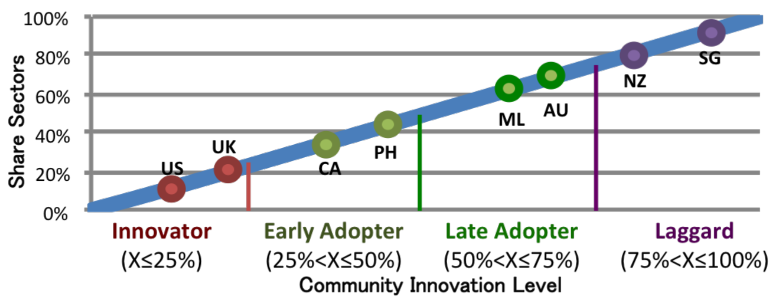

Fig. 3. Community Innovation Level of 8 different English speaking countries

The way we classify the community innovation feature can be seen from figure 3 the graph shows the innovation level of each country community. For example, U.K. and U.S. are the Innovator, which tend to start diffusing the trending topics to other countries' communities, and Canada (CA) and Philippines (PH) are the Early Adopter that usually adopt the trending topics in the early stage. The Y-axis shows the percentage of share sectors (market share the percentage of people who know the specific trending topics), and the X-axis represents the average time taken to adopt the specific trending topics. We classified the innovation levels based on the average percentage of time that a country spent on adopting the trending topics. For example, on average, when only 15% of 8 English-speaking countries communities adopts the certain trending topic, the topic is already extremely popular among majority of people in U.S. online community and appears in U.S. trending topic list.

4.2 Context Feature

This feature enables identification of the type of trending topics by using context patterns. We used three categories, including breaking news, meme, and commemorative day, on the context patterns. Table 1 shows the example pattern for classifying the type of trending topics. Based on this context pattern and the type, we created and used over 20 rules to classify the trending topics using context patterns. For example, if the trending topic contains Person Pronoun AND Noun, then the trending topic is classified into Meme category.

Table 1. Context Feature Classification Pattern Example

Patterns	Example	Category
Be + Noun	Nothing was the same	Meme
Verb + Noun	#HowToAskSomeoneOnADate	Meme
Person pronoun + Noun	#ILoveEXO	Meme
Noun + Commemorative days	Christmas	Commemorative
Noun + No Commemorative days	Galaxy	News

Table 2 shows that over 85% of trending topics collected in three years are talking about the news, which is matched with the results from Kwak et al [1]. They mentioned that around 80% of trending topics are related to the title of breaking news.

Table 2. Percentage of trending topics based on the context pattern

Context categories	Percentage
Commemoratives	4%
Meme	9%
News	87%

4.3 Topic Feature

This feature represents the semantic topic of the Trending Topics. As trending topics are mostly about new and real-time events, the traditional topic classification ontology cannot be applied. This is because traditional ontology usually does not contain new terms, and it normally classifies the topic into the semantically related category, rather than the category related to the real-time situation about the topic. Assume the topic ‘Samsung’ is on the ‘Trending Topic’ list now. In this case, traditional topic classification ontology will classify the topic into ‘Technology’. However, if the trending topic ‘Samsung’ is about the news, Samsung sponsoring British football team Chelsea, the topic should be classified into ‘Sports’ category.

Therefore, we classified the trending topics using the New York Times (NY times) classification service. The service provides nine (9) topic categories as follows: Sports, Entertainment, Politics, Business, World issue, Technology, Fashion, Obituaries, Health. The way we identified the category of a trending topic is as follows: first of all, we search the trending topic with the NY times topic classification service. We set the published time as the day that trending topic first emerged. Then we can locate any related articles that were published with that term, on that day. Finally, the trending topics related categories are supplied by the NY Times classification service. As can be seen in the table 3, we found that 80% of trending topics are classified in the following three categories, including entertainment, sports and politics. It represents that most people are interested in the issues/events of entertainment, sports, or politics.

Table 3. Topic disbribution in U.S. Trending Topics

No	Topic	Percentage
1	Entertainment	42%
2	Sports	28%
3	Politics	10%
4	Fashion	6%
5	World issue	5%
6	Obituaries	4%
7	Health	3%
8	Technology	2%

4.4 Rank Feature

Each trending topic has a popularity ranking (Rank 1 to 10). The ranking of trending topics is changing in real-time. For this feature, we used the ranking value of a trending topic when it was initiated/started from a certain country. However, there are some trending topics, which are starting in the multiple countries' communities. In this case, we checked the community innovation level of starting countries. Once a trending topic comes in, first check whether it started in a single country. When it starts in a single country, then we use the original rank of this trending topic. When it initiates in multiple countries, then check whether it starts from the country that is innovator level. If it starts from innovator, then check whether there are more than one innovator for this trending topic. If more than one, then choose the highest rank of this trending topic as the starting rank. If there is only one innovator, then choose rank from innovator as starting rank of this trending topic. If there is no innovator, then choose the highest rank of this trending topic as the starting rank.

5 Evaluation

5.1 Evaluation Data

As mentioned in section 3, we used Twitter API and collected trending topic terms, related tweets and ranking patterns for those topics for three years (from 1st January, 2013 to 31st December, 2015) in 8 different English speaking countries. We crawled the top 10 trending topics for each country every hour. The API returns the trending topic term, the rank, the location and time of the API request. The trending topics list displays only the topics terms with no detailed information so we searched the related tweets of each trending topic and used published date-time to extract the appropriate related tweets.

In order to achieve the equation 1 in section 4, we used the training data contains trending topics with extrated four different features (including Community Innovation level, Context, Topic, and Rank) as attributes and the number of diffused countries as a class/outcome.

5.2 Machine Learning Techniques

For building the prediction model using our training data, we applied machine learning techniques, which are initially introduced for predictions. We selected four machine learning techniques: Naive Bayes(NB)[5], Neural Networks(NN)[3], Support Vector Machines(SVM)[1], and Decision Trees(DT)[15]. The philosophies behind these techniques are very different, but each has been shown to be effective in several time-series prediction studies.

5.3 Evaluation Result

We used 10-fold cross validation on three years of training data, which is described in section 5.1. The main objective of this experiment is to examine the the scale prediction model of trending topics diffusion in four machine learning techniques. Experiment demonstrated that the proposed diffusion prediction model reasonably predicts the scale of diffusion of trending topics. Table 4 presents the scale of diffusion prediction accuracy with four machine learning techniques. The combination of context and topic features produce the lowest accuracy with the four techniques techniques. In contrast, using only community level or ranking feature, provides a better accuracy with both being over 0.5. The accuracy results increased from 0.04 up to 0.07 by using community level feature and ranking feature together. The greatest accuracy was achieved by using all four features of evaluation in the four machine learning techniques, this being over 0.7, which is much higher than using a single feature, or two features, to predict the scale of trending topic diffusion.

By using the combination of context and topic features, the scale of diffusion prediction accuracy just reached 0.28, which is not enough for predicting the scale of trending topics. This result can be explained in two reasons: First, more than 80% of trending topics are categorised into the ‘News’ type by using given context

Table 4. The prediction accuracy with four machine learning techniques

	Context + Topic	Rank	Community Level	Rank + Community Level	All
NB	0.212	0.582	0.545	0.632	0.722
NN	0.252	0.592	0.559	0.636	0.738
SVM	0.258	0.593	0.558	0.642	0.727
C4.5	0.281	0.609	0.658	0.736	0.808

pattern rules so that the context feature is not suitable to classify or predict the scale of trending topics. Secondly, topic feature has similar circumstances, which are usually classified trending topics into the following three groups, including sports, entertainments, and politics. Therefore, it is not suitable to use only context or topic feature as a parameter for predicting the scale of trending topic diffusion.

The prediction accuracy obtained by using only ranking or community level feature in NN and SVM are almost the same. However, the prediction result with the combination of ranking and community level feature in NN is lower than those in SVM, and the accuracy result of using all four factors in NN is higher than in SVM. SVM can solve the problem of structure selection in NN, which can improve the accuracy result of using ranking and community level factors. Comparing to NN, the accuracy of using context factor is better in SVM, while the performance outcome of using all four factors is lower in SVM. That means patterns for adding context and topic factors better fit the NN algorithm. Another interesting finding is that C4.5 decision tree have better accuracy than the other three machine learning techniques. Since C4.5 decision tree uses a tree to conduct the prediction/classification, the classifying results may better fit with the proposed diffusion scale prediction model. When we use context and topic factor with C4.5, the accuracy result is lower than the others, just reaching 0.281. However, using only community level factor or ranking factor, the accuracy result reaches over 0.6. When combining ranking factor and community level factor, the accuracy result increased at least 0.07 compared to using a single factor. The highest accuracy result almost reached 0.8 by using all four factors together.

Figure 4 represents the error rate based on the predicted scale of trending topics diffusion. The error rate shows that the prediction model is successfully predicting when the trending topic has less diffusion or diffuses to all communities. Based on our analysis, 50% of trending topics tend to diffuse 2 or 8 countries. Therefore, for the future work, this would require an analysis of trending topics in order to identify the important features in distinguishing the medium level of popular trending topic diffusion patterns.

6 Conclusion

In this paper, we proposed a diffusion scale prediction model for trending topics among different online communities. We developed the model with four important features that can be used in predicting the number of countries that trend-

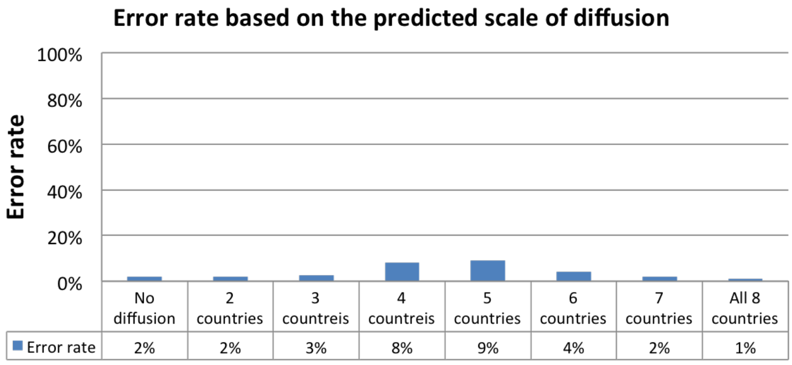

Fig. 4. Error Rate based on the predicted scale of diffusion

ing topics diffuses to. The four features include community innovation level, context, topic, and rank features. The proposed features are learned by four different machine-learning techniques: Nave Bayes, Neural Network, Support Vector Machine, and C4.5 decision tree. Based on the results, the prediction model learned by C4.5 decision tree achieved the highest prediction accuracy (80.8%) in scale prediction. Compared to traditional social data applied diffusion prediction model, our proposed prediction model works successfully. For the future work, we will focus on predicting range (depth) and speed of trending topic diffusion so it can forecast three dimensions of trending topic diffusion patterns. We hope the paper is the step forwards improving the performance for any researches using trending topics as a data.

Acknowledgement

This work was supported by the Industrial Strategic Technology Development Program, 10052955, Experiential Knowledge Platform Development Research for the Acquisition and Utilization of Field Expert Knowledge, funded by the Ministry of Trade, Industry & Energy (MI, Korea).

References

1. Abe, S. (2010). Support vector machines for pattern classification. Springer.
2. Andersen, P. M., Hayes, P. J., Huettner, A. K., Schmandt, L. M., Nirenburg, I. B., & Weinstein, S. P. (1992, March). Automatic extraction of facts from press releases to generate news stories. In Proceedings of the third conference on Applied natural language processing (pp. 170-177). Association for Computational Linguistics.
3. Araujo, P., Astray, G., Ferrerio-Lage, J. A., Mejuto, J. C., Rodriguez-Suarez, J. A., & Soto, B. (2011). Multilayer perceptron neural network for flow prediction. Journal of Environmental Monitoring, 13(1), 35-41.

4. Chen, L., Zhang, C., & Wilson, C. (2013, October). Tweeting under pressure: analyzing trending topics and evolving word choice on sina weibo. In Proceedings of the first ACM conference on Online social networks (pp. 89-100). ACM.
5. Efron, B. (2013). Bayes' theorem in the 21st century. Science, 340(6137), 1177-1178.
6. Han, S. C., & Chung, H. (2012). Social issue gives you an opportunity: Discovering the personalised relevance of social issues. In Knowledge Management and Acquisition for Intelligent Systems (pp. 272-284). Springer Berlin Heidelberg.
7. Han, S. C., Chung, H., Kim, D. H., Lee, S., & Kang, B. H. (2014). Twitter trending topics meaning disambiguation. In Knowledge management and acquisition for smart systems and services (pp. 126-137). Springer International Publishing.
8. Han, S. C., Liang, Y., Chung, H., Kim, H., & Kang, B. H. (2015). Chinese trending search terms popularity rank prediction. Information Technology and Management, 1-7.
9. Han, S. C., & Kang, B. H. (2012, June). Identifying the relevance of Social Issues to a Target. In Web Services (ICWS), 2012 IEEE 19th International Conference on (pp. 666-667). IEEE.
10. Han, S. C., Chung, H., & Kang, B. H. (2015). Trending Topics Rank Prediction. In Web Information Systems EngineeringWISE 2015 (pp. 316-323). Springer International Publishing.
11. Inouye, David, and Jugal K. Kalita. Comparing twitter summarization algorithms for multiple post summaries. Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom). IEEE, 2011.
12. Jeyaraj, A., & Sabherwal, R. (2014). The Bass Model of Diffusion: Recommendations for Use in Information Systems Research and Practice. JITTA: Journal of Information Technology Theory and Application, 15(1), 5.
13. Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World wide web (pp. 591-600). ACM.
14. Lee, K., Palsetia, D., Narayanan, R., Patwary, M. M. A., Agrawal, A., & Choudhary, A. (2011, December). Twitter trending topic classification. In Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on (pp. 251-258). IEEE.
15. Loh, W. Y. (2011). Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), 14-23.
16. Rogers, E. M. (2010). Diffusion of innovations. Simon and Schuster. Chicago
17. Sharifi, B., Hutton, M. A., & Kalita, J. (2010, June). Summarizing microblogs automatically. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 685-688). Association for Computational Linguistics.
18. Zubiaga, A., Spina, D., Fresno, V., & Martinez, R. (2011, October). Classifying trending topics: a typology of conversation triggers on twitter. In Proceedings of the 20th ACM international conference on Information and knowledge management (pp. 2461-2464). ACM.