Multimed Tools Appl
https://doi.org/10.1007/s11042-017-5513-0

A news-topic recommender system based on keywords extraction

Zihuan Wang[1] · Kyusup Hahn[2] · Youngsam Kim[3] · Sanghyup Song[4] · Jong-Mo Seo[1]

Received: 31 October 2016 / Revised: 16 October 2017 / Accepted: 6 December 2017

Abstract In recent years, internet news has become one of the most important channels for information acquisition, as more and more people read news through internet connected computers, tablets, and smart phones, etc. Owing to the constantly reproduced news, the number of online media increases dramatically and the volume of news also expands rapidly. Consequently, obtaining primary information from the internet is of great interest. This paper presents a news-topic recommender system based on keywords extraction. It is shown that the proposed system is very effective in acquiring specific topics within any specific period of time.

Keywords Internet news · Recommender system · Keywords extraction · Topic extraction

✉ Jong-Mo Seo
callme@snu.ac.kr

Zihuan Wang
keith228@snu.ac.kr

Kyusup Hahn
kyuhahn@snu.ac.kr

Youngsam Kim
youngsamy@gmail.com

Sanghyup Song
willsong@snu.ac.kr

1 Department of Electrical and Computer Engineering, Seoul National University, Gwanak-gu, Seoul, South Korea

2 Department of Communication, Seoul National University, Gwanak-gu, Seoul, South Korea

3 Department of Linguistics, Seoul National University, Gwanak-gu, Seoul, South Korea

4 Big Data Institute, Seoul National University, Gwanak-gu, Seoul, South Korea

Published online: 19 December 2017

1 Introduction

Topic detection and tracking (TDT) automatically organizes related material in stream of data. As the primary task of TDT, topic tracking aims to associate the incoming stories with known topics, where the topic is a seminal event with all directly related events. Based on the understanding of contents and keyword extraction techniques, the task of topic tracking is to monitor a stream of news data and acquire some data that discuss the same topic. Compared to information retrieval, information indexing, contents classification and data mining, topic tracking has lots of specialties of its own. First of all, topic tracking is a content-based comprehension for a special event. Secondly, topics actually refer to an event, something in the real world that has already happened, so that the best test data is internet news. A keywords extraction algorithm such as Term Frequency-Inverse Document Frequency (TF-IDF), TextRank, Rapid automatic keyword extraction (RAKE) has been successfully utilized in text mining, social network and recommendation systems.

There are many areas of the industry where topic tracking can be applied. For example, [23] investigated a method that uses text retrieval and clustering techniques for topic detection. The resulting cluster authorities highly informative for reflective detection of previously undefined topics and temporal distribution patterns. James et al. [7] defined a system with clustering algorithms solving related problems of new event detection and event tracking within a news stream. By applying wavelet analysis to the frequency-based raw signals of words, [8] attempted to detect events by analyzing text stream in Twitter with Event Detection with Clustering of Wavelet-based Signals (EDCoW). Yang et al. [22] studied the effectiveness of using information retrieval and machine learning techniques in topic tracking. Based on both information contents and temporal features of topics, it extends existing unsupervised learning and supervised clustering algorithms to do document classification. Topic tracking can also be used to remind companies whenever a topic is in the news. Dai et al. [2] applied the TDT technology to financial fields where TDT can effectively detect and track online financial topics. Masaki et al. [13] proposed a new approach to observe, summarize and track events from a collection of news. Jianshu and Lee [8] introduced an application of topic tracking for the social network.

In traditional methods of keyword annotation, the advantage of the graph method represented by TextRank is to consider the semantic relation between words and words in the document. Statistical methods represented by TF-IDF considers the statistical properties of the word only. However, methods such as TF-IDF and TextRank do not take into account the coverage problem of the extracted keywords, often leading to the keywords that focus on a large topic without considering other topics of the document. Unsupervised and supervised methods have their own advantages and disadvantages. Specifically, unsupervised methods do not require manual labeling for preparing the training data set. Whereas, the supervised method can be trained to adjust a variety of information to determine the impact of the degree of the keywords, and hence the effect is better than the unsupervised method. However, manually marking the training set is very inefficiency, especially in recent big-data years, and processing efficiency is getting more and more important. Therefore, the recent study of keyword extraction mainly focused on unsupervised methods. Rada and Tarau [14] introduced TextRank–a graph-based keyword extraction algorithm which has achieved very promising results and has aroused great interest in the field of research. Stuart et al. [16] proposed a keyword extraction method that can efficiently operate individual documents to enable an application to dynamic collections, is easily applied to new domains and operates well on multiple types of documents. Sungjick and kim [1, 16–18, 20, 21] introduced the algorithms which are mostly based on the frequency of certain words or greedy-search based method.

In this paper, we use the RAKE keywords extraction method to acquire keywords from internet news. We presents a news-topic recommender system based on extracting keywords. It is shown that the proposed system is very effective in acquiring specific topics within any specific period of time. The organization of this paper is as follows. In Section 2, we discuss three keyword extraction algorithms and their limitations. In Section 3, we present a news-topic recommender system based on existing keyword extraction algorithm. Finally, in Sections 4 and 5, we discuss experimental results, limitation of this experiment, and our future work.

2 Related work

2.1 TF-IDF

The importance of a word in an article can be measured by its frequency of occurrence. Important words often appear many times in an article. However, the frequent occurrence does not necessarily make a word important because some words are frequently seen in all kinds of articles. Those words that frequently occur in specified articles are important. From a statistical point of view, it is better to give larger weights to those words that are less frequently occurred and to reduce the weights of frequently occurred words. Inverse Document Frequency (IDF) can be utilized as the weight to be added or reduced, Term Frequency (TF) is referred to as word frequency. In [17], a new TF-IDF based algorithm is presented, where three TF variants are proposed as follows,

$$\mathrm{BTF}_i = \sum_{j=1}^{|D|} n_{i,j}\,, \tag{1}$$

$$\mathrm{NTF1}_i = \frac{\mathrm{BTF}_i}{\max\{\mathrm{BTF}_1, \mathrm{BTF}_2, ..., \mathrm{BTF}_{|T|}\}}, \tag{2}$$

$$\mathrm{NTF2}_i = \sum_{j=1}^{|D|} \frac{n_{i,j}}{\sum_{k=1}^{|T_j|} n_{k,j}}. \tag{3}$$

Here, BTF means a frequency that represents how often a term occurs in given dataset. Hence, the TF-IDF variants can be given by

$$\text{TF-IDF}_i = (\mathrm{TF}variant) \times \mathrm{IDF}_i \tag{4}$$

and

$$\mathrm{NTFIDF}_i = \log\left(\mathrm{TF}variant + 1.0\right) \times \mathrm{IDF}_i. \tag{5}$$

The above calculation is explicit. Indeed, we have some fundamental metric to extract the most descriptive terms in a document and also can quickly compute the similarity between two documents, which is a advantage. On the other hand, TF-IDF is based on the bag-of-words (BoW) model and does not capture position in the text, semantics, and co-occurrences in different documents. Thus, it is only useful as a lexical level feature and cannot capture semantics, compared to topic models, word embeddings.

2.2 TextRank

TextRank algorithm are Graph-based, as discussed in [14]. The basic idea consists of "voting" and "recommendation". When one vertex links to another, the former casts a vote for the latter. The higher the number of votes that are cast for a vertex, the higher its importance. Moreover, the importance of the vertex casting the vote determines the importance of the vote itself, and this information is also taken into account by the ranking model. Therefore, the score associated with a vertex is determined by the number of votes that cast for it. Formally, let $G = (V, E)$ be a directed graph with a set of vertices V and a set of edges E, where E is a subset of $V \times V$. For a given vertex V_i, let $In(V_i)$ be the set of vertices that point to it. $Out(V_i)$ is the set of vertices. In [14], vertex V_i is defined as follows:

$$S(V_i) = (1 - d) + d \sum_{j \in In(V_i)} \frac{S(V_j)}{|Out(V_j)|}, \tag{6}$$

where d is a damping factor that can be set between zero and one, which plays the role of integrating the model. The probability of jumping is from a given vertex to another random vertex in the graph. In the context of web surfing, this graph-based ranking algorithm actualizes the "random surfer model", where a user can click on links randomly with probability d, and the links jump to an entirely new page with probability $1 - d$. The factor d is usually set as 0.85. Our topic extraction model also considers a damping factor related to the number of media and daily keywords. TextRank examines the relationship between words and words in the document, but still, tends to choose the more frequent words in the document as a keyword. Our method will first remove noise which can be seen as stop words and proper name, also give priority to those daily, wide appeared keywords to fill the gaps.

2.3 RAKE(Rapid automatic key word extraction)

RAKE is one of the keywords extraction algorithms, it is popular because of its processing efficiency. Stuart et al. [16] compared its processing efficiency with TextRank. In parallel with the increasing data, the increment of the processing time of RAKE is slower than that of TextRank. Therefore, RAKE is suitable for processing large amounts of data. In order to measure how essential a keyword is, [16] defined $ess(k)$ as:

$$ess(k) = \frac{edf(k)^2}{rdf(k)}, \tag{7}$$

where the referenced document frequency of a keyword, represented by $rdf(k)$, is the number of documents in which the keyword occurred as a candidate keyword. $edf(k)$ is the number of documents from which the keyword is extracted.

3 Keyword extraction

This paper offers a recommender system based on RAKE, which is a keyword extraction method that operates on individual documents. Since we focus on internet news data where each news article can be viewed as an individual document, RAKE is suitable for our topic extraction method. However, the keywords in each news article extracted by RAKE are not enough to cover the period topic, and we believe that the period topic keywords are hidden in the daily keywords list. Therefore, the key point of this topic extraction method is to find topic keywords in daily keywords list. We solve this problem by re-scoring the keywords extracted by RAKE. This re-scoring algorithm can be used on hot topic tracking for internet news of given dates. Based on the processing efficiency, our method uses RAKE to extract keywords. In addition, we utilize the concept of $edf(k)$ to improve our algorithm; see details in the next section.

3.1 Keyword scoring

We aim to extract topic keywords for a specific period of time. We considers that the topic keywords should appear frequently and widely in a certain period of time. Since it is difficult to consider the frequency only, at first we introduce a scoring formula. Let D_{max} be the greatest number of days that a word continuously appears for as a keyword. Then the scoring formula can be given by

$$Score(k) = \frac{1}{T} \cdot \sum_{i=1}^{T} edf(k)_i \cdot D_{max}. \tag{8}$$

As mentioned in the previous section, $edf(k)$ is the number of documents from which the keyword is extracted. Through empirical experiments, we found that D_{max} increases with the weight of consecutive words. Figures 2, 3 and 4 show the topic tracking results. Moreover, we consider that an important event will be reported by more media. Thus, we incorporate another variable N_{media} with a intention of weakening the affect of frequency. N_{media} is defined as a damping factor representing the topic range of influence. Our scoring equation can be presented as

$$Score(k) = \frac{1}{T} \cdot N_{media} \cdot \sum_{i=1}^{T} edf(k)_i \cdot D_{max}. \tag{9}$$

The damping factor N_{media} is defined by

$$N_{media} = \frac{0.5 + \log(N_{kwz})}{\log(N_t)}, \tag{10}$$

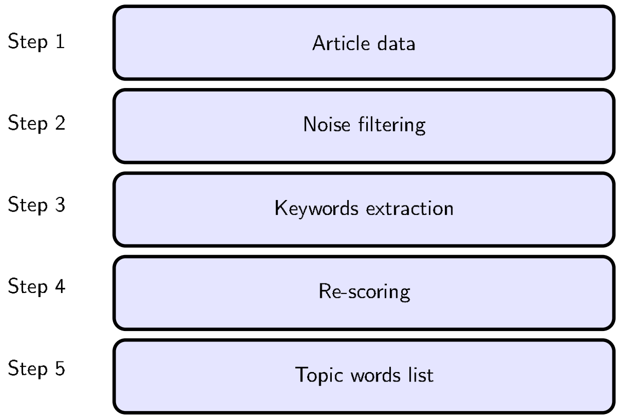
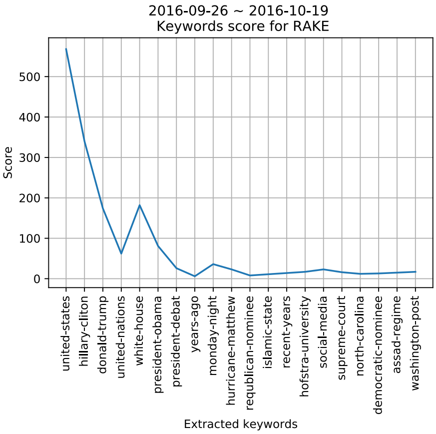
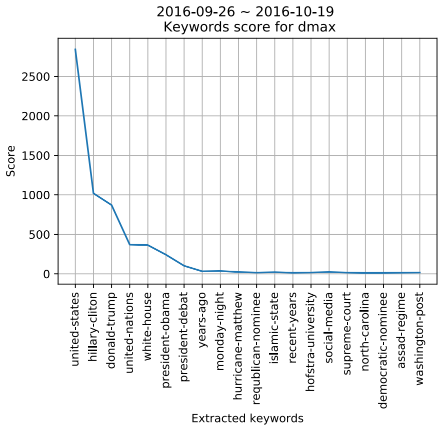


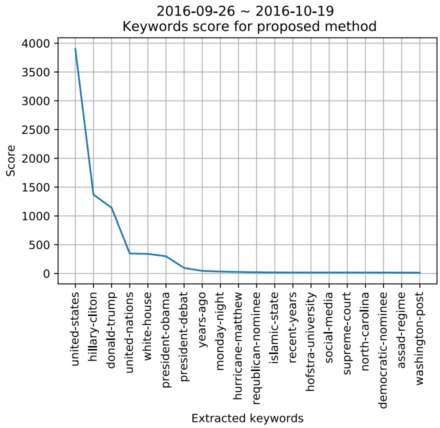
Fig. 1 Topic extraction procedure



(a) Score of rake



(b) Score of dmax



(c) Score of proposed method

Fig. 2 Topic tracking result 1

where $N_{media-total}$ is the number of media, N_{kwz} is the number of media that a keyword comes from. By (9) and (10), we get:

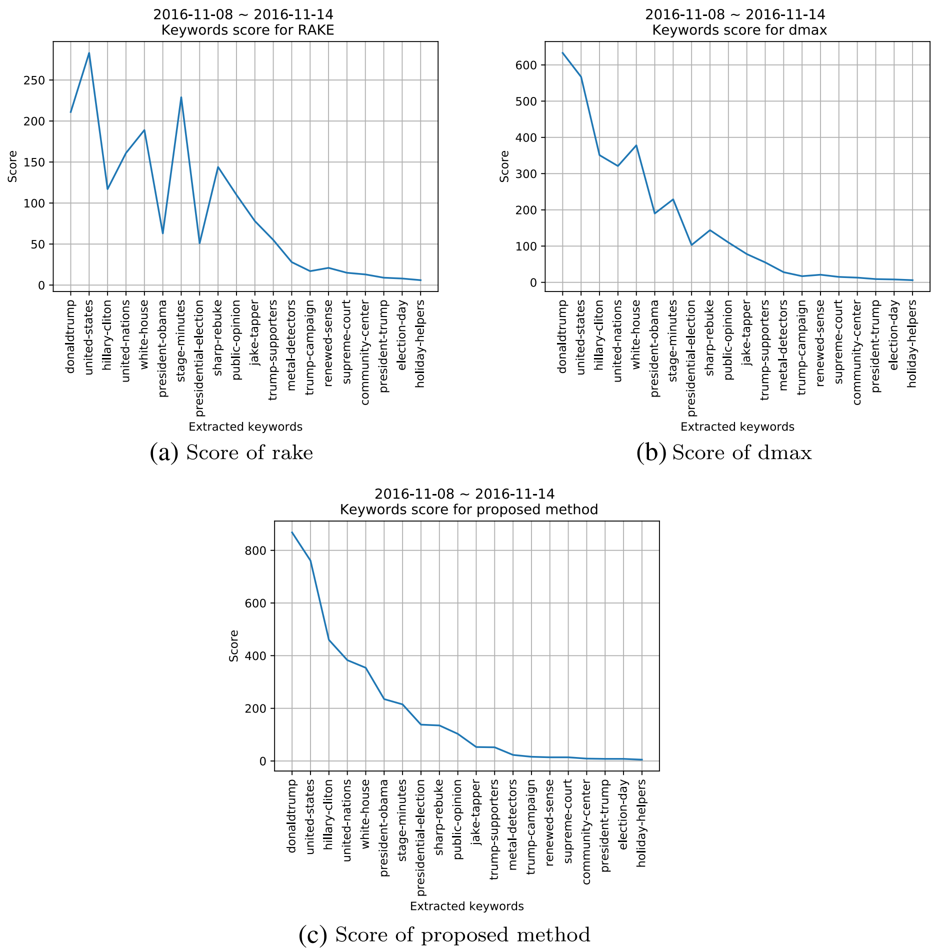
$$Score(k) = \frac{1}{T} \cdot \frac{0.5 + \log(N_{kwz})}{\log(N_t)} \cdot \sum_{i=1}^{T} edf(k)_i \cdot D_{max}. \tag{11}$$

This scoring method weakens the effect of frequency and improves the weight of continuously and broadly appeared keywords. In Fig. 5, we compare RAKE, D_{max}, and the proposed method. Detailed instructions will be explained later.

3.2 Topic words extraction process

Now we introduce the entire process of topic extraction.

Figure 1 shows the entire topic extraction process. News article data we use in experiment comes from the internet. We made some spider programs to crawl target news articles



(a) Score of rake

(b) Score of dmax

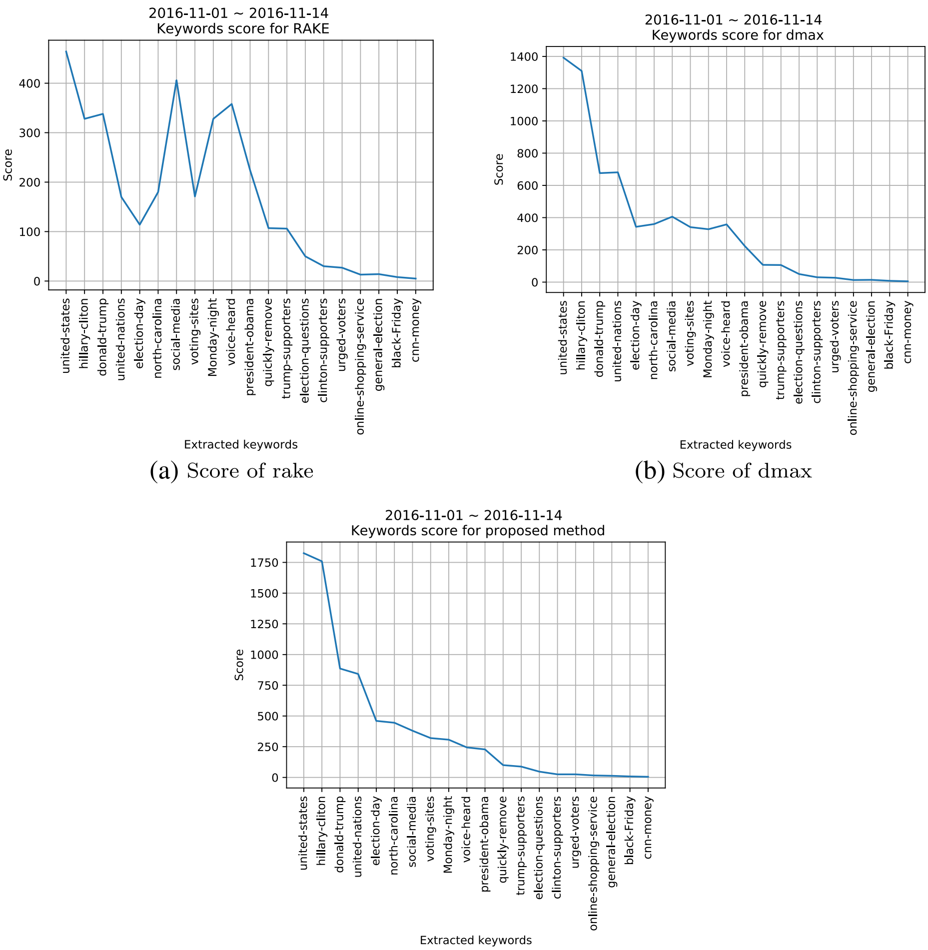
(c) Score of proposed method

Fig. 3 Topick tracking result 2

every day. Some noise such as conjunction, media name, specific proper name, etc. will be removed during crawling. After that RAKE will automatically extract keywords. The next step, our proposed method will rescore extracted keywords and define which can be considered as a topic.

4 Experiment

We set the start date on September 26, and October 19 be the end date as the data range. News data collected in this data range as the test data try to carry on the hot topic extraction used to determine whether the method as mentioned above is effective. If the presidential debate in the hot topic listing argues that works, the higher the sequence, the greater the accuracy. In this period, we collected 20,193 news data from CNN, Fox News, MSNBC, The New York Times, The Washington Post, The Wall Street Journal, USA TODAY, The Los



(a) Score of rake

(b) Score of dmax

(c) Score of proposed method

Fig. 4 Topic tracking result 3

Table 1 Dataset structure

Column	Description
Aid	Article id
Date	News date
Title	News title
Contents	News contents
Url	News url
Agency	Media, newspaper company
Category	Policy, Sports, Entertainment etc.
Keywords	Extracted keywords from title and contents

Angeles Times, The Philadelphia Inquirer, The St. Louis Post-Dispatch and Chicago Daily Herald. On November 8, 2016, which was the day of U.S. election day, we set November 1 as the start date and November 14 be the end of the data range. News data collected in this data range as the test data try to carry on the hot topic extraction used to determine earlier is effective. Similarly, if the election day in the hot topic listing argues that works, the higher the sequence, the greater the accuracy. In this period, we collected 14,528 news data also from CNN, Fox News, MSNBC, The New York Times, The Washington Post, The Wall Street Journal, USA TODAY, The Los Angeles Times and The Philadelphia Inquirer, The St. Louis Post-Dispatch and Chicago Daily Herald. Also, we set the start date on November 8, and November 14 be the end date as the data range, right after the election day. 8038 news data collected from same media is set as test data.

From Figs. 2, 3 and 4, we can see that the distribution of scores is very uneven, the top five ranked keywords occupy most of the components, that is to say, our method effectively indicating the topics with a given period of time.

To improve topic keyword extraction efficiency, we extracted keywords of each article when collecting an article. For one news data, we need its title, contents, publish date, URL, media and category. Also, we give it a unique ID to avoid duplication of crawling. Table 1 shows a database structure to store this information.

For a given specific period, we use the RAKE algorithm, the D_{max} method mentioned in this article, and the proposed method to extract the topic. From the results of experiment 1 in Fig. 2, we can see that the score curves of three methods are roughly the same, mostly

Table 2 Topic tracking result 1

2016-09-26 – 2016-10-19		
Rank	Keyword	Score
1	united-states	3900
2	hillary-cliton	1370
3	donald-trump	1142
4	united-nations	347
5	white-house	341
6	president-obama	299
7	presidential-debat	96
8	years-ago	45
9	monday-night	34
10	hurricane-matthew	22
11	republican-nominee	21
12	islamic-state	20
13	recent-years	19
14	hofstra-university	16
15	social-media	16
16	supreme-court	15
17	north-carolina	15
18	democratic-nominee	15
19	assad-regime	15
20	washington-post	14

concentrated in few of the top-ranked keyword (See Table 2). Figure 3 and Table 3 shows the results of experiment 2 and Fig. 4 and Table 4 shows the result of experiment 3. From (a) we can see the score of keyword "presidential-election" extracted by RAKE is not high. $D_{\max}$ relatively improved this phenomenon. If we pay attention to the score of "presidential-election" in (b), it is easy to find that the score is lower than the score of "stage-minutes". It is worth noting that the presidential candidate's debate happened in the chosen period. Therefore, from the "presidential-election" should be extracted as a topic and ranked beyond "stage-minutes". (c) shows our proposed method works well.

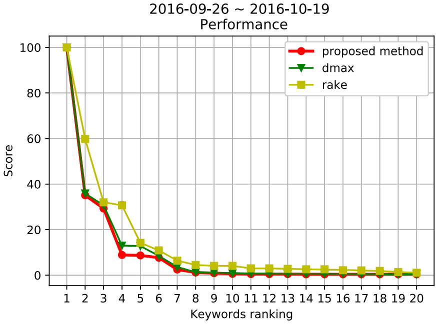
In topic tracking result 1, according to our calculation of the average rank of "presidential-debate", its value is 9. "Presidential-debate" appeared for 6 days out of 24 days in the top 20 keywords list and still ranked the seventh. The average ranking values of "election-day" and "donald-trump" is 9 and 7, respectively. In topic tracking result 3, one can notice that "election-day" appeared 3 days out of 14 days (i.e., low frequency of appearance) but still ranked the fifth in the top 20 keywords list, demonstrating the accuracy of our news-topic recommender system. However, "donald-trump" appeared 5 days out of 7 days, which does not prove the effectiveness of our recommender system. This is because "donald-trump" occurs frequently in the given period of time, which is relatively short. Additional experiment results show that when the given period of time is too short (less than one week), our proposed system does not work very well, which seems to be a limitation of our system. To facilitate the performance comparison, we use a relative comparison methodto convert the keywords' score, and we set the maximum value of the keywords

Table 3 Topic tracking result 2

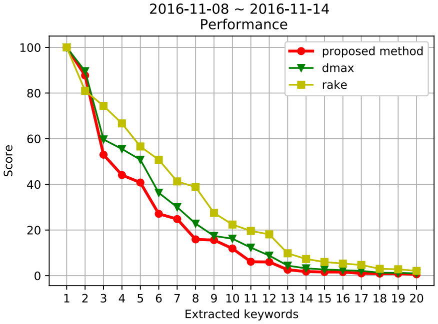
2016-11-08 – 2016-11-14		
Rank	Keyword	Score
1	donald-trump	868
2	united-states	761
3	hillary-cliton	460
4	united-nations	383
5	white-house	354
6	president-obama	235
7	stage-minutes	215
8	presidential-election	138
9	sharp-rebuke	135
10	public-opinion	103
11	jake-tapper	53
12	trump-supporter	52
13	metal-detectors	23
14	trump-campaign	16
15	renewed-sense	14
16	supreme-court	14
17	community-center	9
18	president-trump	8
19	election-day	8
20	holiday-helpers	5

Table 4 Topic tracking result 3

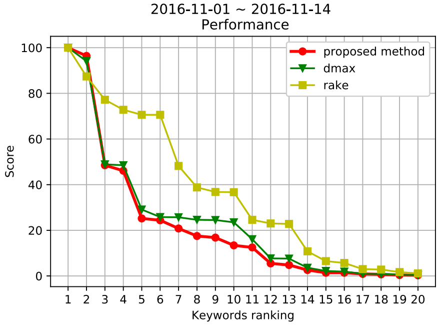
2016-11-01 – 2016-11-14		
Rank	Keyword	Score
1	united-states	1825
2	hillary-cliton	1759
3	donald-trump	886
4	united-nations	842
5	election-day	460
6	north-carolina	445
7	social-media	380
8	voting-sites	320
9	monday-night	307
10	voice-heard	244
11	president-obama	228
12	quickly-remove	100
13	trump-supporters	88
14	election-questions	47
15	clinton-supporters	25
16	urged-voters	25
17	online-shopping-service	16
18	general-election	13
19	black-friday	8
20	cnn-money	5



(a) Experiment 1



(b) Experiment 2



(c) Experiment 3

Fig. 5 Performance comparison

extracted by each method as 100. After conversion in proportion, we will get converted scores.

Figure 5 shows the results of the comparison of three methods. We can clearly see that the score curve of the proposed method decreases faster than D_{max} and RAKE. Which means it shows the best performance. Compared with (b) and (c), (a) has more keywords with small scores. Most scores concentrated at the few top-ranked key words because of the time span reached to 24 days. Which means it may make a topic get a higher score. As a result, a recurring topic will be efficiently extracted within the specified period of time. However, with a longer time span, our proposed method may mistake some proper names for topics. Because of proper names always appear in many news articles, it is a challenge to distinguish them as being a topic or not. If the time span is too short, the real topic may not be identifiable. To solve this problem, the noise filtering stage needs further optimization.

5 Conclusion and future work

In this paper, we have proposed a news-topic recommender system based on keywords extraction, in which on the basis of the experimental results. We have confirmed the topic words come from daily keywords. By theoretical analysis and numerical experiments, we have demonstrated that given a specified period of time, our recommender system improves the weight of the expected keyword and the keyword appears successfully. However, finding and removing those repeatedly-appeared words, such as proper nouns, still require further research. Also, the dataset used in this work is relatively limited to political categories. Therefore, it would be interesting to study the behaviour of our recommender system using massive data acquisition scheme in the future.

Acknowledgements This work was supported by Seoul National University Big Data Institute through the Data Science Research Project 2015.

References

1. Bruno P, Ralf S, Camelia I, Emilia K, Irina T (2004) Multilingual and cross-lingual news topic tracking. In: Proceedings of the 20th international conference on computational linguistics
2. Dai XY, Chen QC, Wang XL, Xu J (2010) Online topic detection and tracking of financial news based on hierarchical clustering. In: 2010 international conference on machine learning and cybernetics (ICMLC), vol 6. IEEE, pp 3341–3346
3. Hong Y, Zhang Y, Liu T, Li S (2007) Topic detection and tracking review. J Chinese Inform Process 21.6:71–87
4. Hsu WH, Chang SF (2006) Topic tracking across broadcast news videos with visual duplicates and semantic concepts. In: 2006 IEEE international conference on image processing. IEEE, pp 141–144
5. Iwata T, Watanabe S, Yamada T, Ueda N (2009) Topic tracking model for analyzing consumer purchase behavior. In: IJCAI, vol 9
6. James A (2012) Topic detection and tracking: event-based information organization, vol 12. Springer Science and Business Media, Berlin
7. James A, Papka R, Lavrenko V (1998) On-line new event detection and tracking. In: Proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval. ACM, pp 37–45
8. Jianshu W, Lee B-S (2011) Event detection in twitter. In: ICWSM 11, pp 401–408

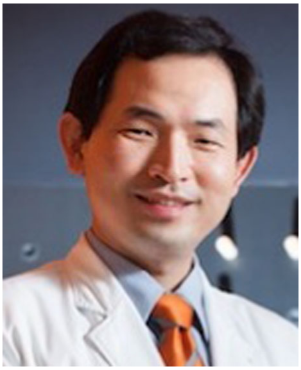
9. Jin Z, Lin HF, Zhao J (2005) Study on topic tracking and tendency classification based on HowNew (in Chinese). J China Soc Sci Tech Inform 24(5):555–561
10. Kuan-Yu C, Luesukprasert L, Chou S-cT (2007) Hot topic extraction based on timeline analysis and multidimensional sentence modeling. IEEE Trans Knowl Data Eng 19(8):1016–1025
11. Li S, Lv X, Li Y, Shi S (2010) Study on feature selection algorithm in topic tracking. In: 2010 2nd international conference on software engineering and data mining (SEDM). IEEE, pp 384–389
12. Li H, Li CH, Wang X (2009) Research on the algorithm of feature selection based on difference and multiple features (in Chinese). Microcomput Appl 30(10):1–5
13. Masaki M, Miura T, Shioya I (2006) Topic detection and tracking for news web pages. In: Proceedings of the 2006 IEEE/WIC/ACM international conference on web intelligence, pp 338–342
14. Mihalcea R, Tarau P (2004) Textrank: Bringing order into text. In: Proceedings of the 2004 conference on empirical methods in natural language processing
15. Mikel R, Ali S, Kanade T (2009) Tracking in unstructured crowded scenes. In: 2009 IEEE 12th international conference on computer vision. IEEE, Piscataway
16. Stuart R, Engel D, Cramer N, Cowley W (2010) Automatic keyword extraction from individual documents. In: Text mining, pp 1–20
17. Sungjick L, Kim H-j (2008) News keyword extraction for topic tracking. In: Fourth international conference on networked computing and advanced information management. NCM'08, vol 2. IEEE, Piscataway, pp 554–559
18. Sung-Jick L, Han-Joon K (2009) Keyword extraction from news corpus using modified TF-IDF. The Journal of Society for e-Business Studies 14(4):59–73
19. Thorsten B, Chen F, Farahat A (2003) A system for new event detection. In: Proceedings of the 26th annual international ACM SIGIR conference on research and development in informaion retrieval. ACM, pp 330–337
20. Wang J (2015) Support recovery with orthogonal matching pursuit in the presence of noise. IEEE Trans Signal Process 63(21):5868–5877
21. Wang J, Li P (2017) Recovery of sparse signals using multiple orthogonal least squares. IEEE Trans Signal Process 65(8):2049–2062
22. Yang Y, Carbonell JG, Brown RD, Pierce T, Archibald BT, Liu X (1999) Learning approaches for detecting and tracking news events. IEEE Intell Syst Their Appl 14(4):32–43
23. Yiming Y, Pierce T, Carbonell J (1998) A study of retrospective and on-line event detection. In: Proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval. ACM, pp 28–36
24. Zheng W, Zhang Y, Hong Y, Fan J, Liu T (2008) Topic tracking based on keywords dependency profile. In: Asia information retrieval symposium. Springer, Berlin
25. Zi-Yan J, Qing H, Hai-Jun Z, Jia-You L, Zhong-Zhi S (2004) A news event detection and tracking algorithm based on dynamic evolution model [J]. J Comp Res Dev 7:032

Zihuan Wang received the B.S. and M.S. degree in Electrical and Computer Engineering from Seoul National University, Seoul, South Korea in 2012 and 2016. He is currently working towards his Ph.D. degree in Electrical and Computer Engineering at Seoul National University, South Korea. His current research interests include text mining, healthcare and medical bigdata analysis.

Kyusup Hahn is currently Associate Professor of Communication at Seoul National University. His research area is political communication. His works have appeared in the Journal of Communication, the Journal of Applied Statistics, PlosOne, and Public Opinion Quarterly. He holds Ph.D. in Communication from Stanford University.

Youngsam Kim received M.S. degree in Cognitive Science from Seoul National University. And he is currently researching in the field of computational linguistics for the Ph.D. degree at the same university. His main interests include distributional semantic modeling, sentiment analysis and computational models for language learning.

Sanghyup Song currently works as a researcher at the Seoul National University Big Data Institute. He received his M.S. degree in Electrical Engineering and Computer Science from Seoul National University in Korea. His current interests include data visualization, in-memory data management and hardware accelerated data processing



Jong-Mo Seo graduated Seoul National University (SNU) School of Medicine, took resident course and retina fellowship in Seoul National University Hospital (SNUH) and got his M.S. and Ph.D. degree in biomedical engineering at SNU. He worked as a retina surgeon in SNUH, and in the year 2008, moved to the department of Electrical and Computer Engineering (ECE) of SNU as an assistant professor. He also served as a staff for the Electronic Medical Record (EMR) system of SNUH from the 1999 to 2003, designing the computer user interface and developing medical machine-computer data transfer modules. He joined the Seoul Artificial Retina Project from the beginning in the year 2000, and has been working as a moderator and a surgeon for the animal experiment. He is an associate professor at SNU ECE and conjoined professor at SNUH biomedical research institute and SNU medical school.