Follower Behavior Analysis via Influential Transmitters on Social Issues in Twitter

Kwang-Yong Jeong, Kyung-Soon Lee

Chonbuk National University, CAIIT, Dept. of Computer Science & Engineering,
Korea

{kyjeong0520, selfsolee}@chonbuk.ac.kr

Abstract. A follower can be divided into supporter, non-supporter, or neutral according to a follower's intention to a target user. Even though a follower is identified as a supporter, an opinion may not be positive to the target user. In this paper, we propose a method to classify a follower as supporter, non-supporter or neutral. To expand information of a follower, influential transmitters who support a target user are detected by using a modified HITS algorithm. To detect a follower's specific opinion, social issues are extracted based on tweets of influential transmitters. The thread tweets are clustered based on Latent Dirichlet Allocation for social issues. Then, sentiment analysis is conducted for the clusters of a follower. To see the effectiveness of our method, a Korean tweet collection is constructed. As a result, we found that lots of supporting followers show opposite opinions depending on particular issues.

Keywords. Follower behavior, influential transmitter, opinion classification, supporting/non-supporting follower, social issue.

1 Introduction

Twitter has become one of the most popular social media platforms for people to share their attitudes or minds. It also allows people to communicate each other based on following or friend relationships [1]. As twitter users are rapidly increased, interactions among the users are markedly increased and influence of users who have lots of followers is greater. These interactions further create opportunities for business companies to conduct online marketing activities [2].

Traditional newspapers and the media are measured influence and far-reaching power of the media over the number of subscribers. However, subscribers in twitter are showing various opinions through a mention. Recent researches on social network analyze characteristics of tweets and infer occurrences, extract events, and detect influential users [3, 4, 5]. These researches measured the influence of followers in social network, and considered a follower as a supporter for a target user (a followee).

In our observation, however, followers can have different purposes in making following relationship. It results in making different responses for the followee's tweets. Some followers reply or retweet for particular issues and other followers reply, retweet or remain mentions for all of tweets. Besides, followers are not positive to all of opinions toward followee's opinions in the tweet. They show negative opinion toward a target user on the particular issues. Therefore, it is not appropriate to classify the follower to be a supporter.

On the other side, when a famous user writes tweets on Twitter, people tend to listen to opinions of influential users. These users play an important mediating role in the spread of tweets. We call these users as influential transmitters. An influential transmitter can be shown as a representative of the supporting followers.

In this paper, we propose follower polarity classification by detecting influential transmitters and follower opinion classification according to social issues by clustering social issues. In order to see the effectiveness of the proposed method, the experiments are conducted on a Korean tweet collection.

The paper is organized as follows: Section 2 describes related work. Section 3 presents our classification model. Section 4 shows experimental results. Finally, we conclude in Section 5.
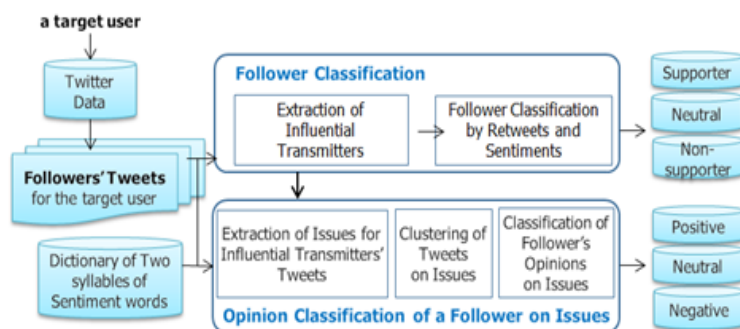
Fig. 1. Overall system architecture

2 Related Work

Generally, influence means changes of people's cognition, attitude and behavior. Socio-scientific studies on the influence were displayed from various angles such as related network analysis, expected theory, persuasive process research, and so on.

Research on measuring user influence in Twitter [3] applied and compared various standards such as the number of followers, the number of retweets, the number of mentions so as to examine who can be influential in twitter space. Weng et al. [4] proposed the TwitterRank, which was extended from PageRank to identify influential Twitter users by taking both the link structure and the topical similarity into consideration.

Social influence is measured generally based on the graph based algorithms such as PageRank and HITS graph. To find influential authors in brand-page communities, Purohit et al. [5] measured influence of twitter users by the number of retweets, replies and mentions based on the HITS algorithm. In our method, we focus on detecting influential transmitters not influential users to expand information for followers' activities to a target user for the polarity classification of a follower.

Researches on sentiment classification are divided into methods to utilize external resources and to utilize internal resources. The method using external resources such as sentiment dictionary and information searching snippet improves confidence of a collected corpus. The method using internal resources is based on syllable n-gram or sliding window [6]. The weakness is that it depends on corpus. In our method, we use bigram and trigram syllables to recognize positive and negative sentiments by analyzing Korean morpheme and use support vector machines to classify sentiments of tweets.

3 Analyzing Follower Behaviors

In order to analyze a follower's intention of the following relationship and opinion on the particular issue toward a target user, the followers are classified as a supporter, a non-supporter, or neutral. In order to make an anchor, influential transmitters are detected as strong supporters. Then an opinion of follower is classified for each social issue as positive, negative, or neutral toward the opinion of a target user. To detect social issues, issue clusters are constructed based on retweet lifetime. The overall system architecture is shown in Figure. 1.

3.1 Follower Polarity Classification Using Influential Transmitters

According to a follower's intention for the following relationship toward a target user, their behaviors might be different for tweets of a target user. A follower can be classified into supporter, non-supporter, or neutral toward a followee.

Even though a follower supports a target user, opinions of the follower can be negative to the opinion of a target user depending on the particular issues.

Based on our observation, there are important users with lots of followers who spread information in order to support a target user as strong
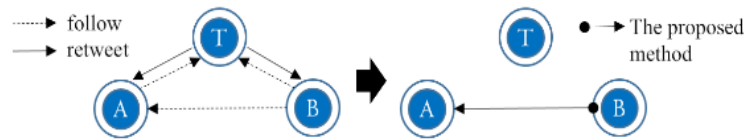
Fig. 2. Modified HITS according to retweet sequence to detect influential transmitters
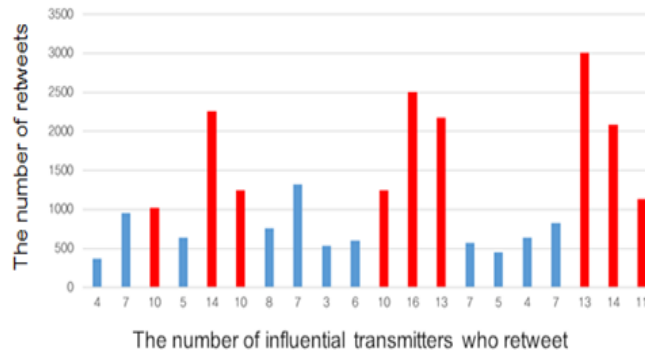

Fig 3. Social contagion by influential transmitters. The red colored bar represents that more than ten users retweet among 16 users
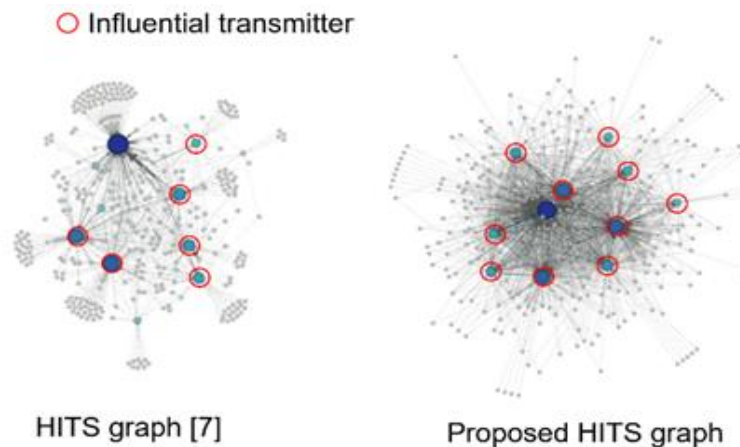

Fig. 4. Detecting Influential transmitters

supporters. We call this user as an influential transmitter for the target user. An influential transmitter plays an important mediating role in the spread of the opinion of a target user to influence their followers and people.

Besides, most followers tend to co-follow the target user and the influential transmitter. When a user retweets a tweet of the influential transmitter, it can be considered as the retweet of a tweet of the target user. To expand information of activities of followers, the activities among influential transmitters and a target user are considered to analyze polarity of a follower. Since influential transmitters can represent followers, we can

predict an opinion of all followers via a response for the influential transmitters.

3.1.1 Detecting Influential Transmitters

Influential transmitters are different from influential users. The extraction method of influential authors [5] used the most influential user by retweets, mentions and replies among all the users. However, influential transmitters to be extracted in this research means users who play an important role of spreading of the followee's tweets or opinions in order to support the followee.

Since the conventional HITS algorithm can only detect influential users, we revised an edge on the HITS graph to delicately express the relationship of retweeting by applying the concept of social contagion to detect influential transmitters. The social contagion refers to tendency for certain behavior exhibited by one user to be copied by other users who have been exposed to media coverage describing the behavior of the target user [12].

When a follower retweets a tweet of a target user, an edge is connected between the follower and only to the transmitter, not the original user of a tweet. The connecting condition of an edge is as follows.

1. A follows a target user T. B follows A and T.
2. Both A and B retweet a tweet of T.
3. The retweet sequence is that B retweets a tweet of T after A does.

The authority and hub score are calculated as follows on HITS graph:

$$Auth^{t+1}(v_i) = \frac{1}{\sum\limits_{j:e_{ji}\in E} w_{ji}} Hub^t(v_j) \tag{1}$$

$$Hub^{t+1}(v_i) = \frac{1}{\sum\limits_{j:e_{ij}\in E} w_{ij}} Auth^t(v_j) \tag{2}$$

where w_{ij} is for a weight of e_{ij} and e_{ij} is the number of followees who retweet a tweet of j before i retweets when i retweets an j's tweet.

When the conditions as described above are satisfied, the user A gives an effect to the user B

when B retweets a tweet of T. Therefore, the related work [5] creates edge of A->T and B->T, the proposed method, however, creates an edge of B->A so as to reflect the transmitter. By reflecting these flows of retweeting relationships on the HITS algorithm, the high ranked followers with high authority scores are detected as influential transmitters. Figure 4 shows the results of detecting influential transmitters.

3.1.2 Follower Polarity Classification

Influential transmitters deliver tweets of a target user to their followers and propagate an opinion of a target user. They can be considered as strong supporters with the same opinion of the target user. Not only retweeting action for a target user's tweet can be seen as the agreement of opinion of the target user, but also retweeting for an influential transmitter's tweet can be considered as the supporting action for the target user's opinion.

Based on the retweet relationships among the followers on social network, the supporters for the target user are classified with strong confidence.

Then, the followers who cannot be classified by the retweet structure are classified by using support vector machines (SVMs) which show high performance in classification. We used bigram and trigram syllables to identify positive or negative sentiments in tweets by constructing bi-gram syllables sentiment dictionary after Korean morphological analysis. The number of positive and syllables and negative syllables in the dictionary are 146 and 135, respectively.

3.2 Follower Opinion Classification on Social Issue Clusters

Even though a follower supports a target user, opinions of the follower can have different responses such as positive or negative toward the target user depending on the particular issues. Social issues should be extracted in order to analyze opinions of followers.

Based on our observation that all tweets of a target user do not show popularity, the tweets on social issues are extracted to analyze. In order to extract issue keywords, the conventional *tfidf* weighting scheme is slightly modified to detect issue keywords.
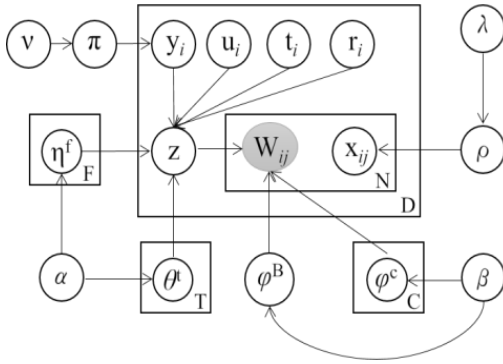
Fig. 5. LDA model for issue clustering on the tweet thread lifetime

$$weight(t) = tf(t,d) \cdot idf(t,N) \cdot sf(t), \qquad (3)$$

where t represents a term, $tf(t, d)$ for term frequency of t in a tweet d, $idf(t,N)= log(N/df(t))$. N represents the total number of tweets of a target user. $df(t)$ represents document frequency of the term t. $sf(t) = IM(t) / IM_{all}$. $IM(t)$ represents the number of transmitters who mention a term t among influential transmitters, and IM_{all} represents the total number of influential transmitters.

Two terms with the most high values of $weight(t)$ are extracted to represent an issue keyword. Here we only used tweets of a target user and influential transmitters to extract issues. Then all tweets of each follower are clustered according to the each issue.

Tweets are clustered by using topic modeling based on the retweet lifetime. Then the opinion is analyzed for the clustered tweets by sentiment analysis using machine learning.

3.2.1 Clustering Tweets on Social Issues Based on LDA

Latent Dirichlet Allocation (LDA) is a fully generative model for describing the latent topics of documents and a standard tool in topic modeling [8, 9, 10].

For the extraction method of topics by LDA in social data, TimeUserLDA model [11] is to extract topic which is explosively increased within particular time in tweet data. The TimeUserLDA is adapted for the our work to make issue clusters on each social issue for the followers' tweets of the

target user, by reflecting relations between tweet's lifetime and target user to topic model.

In order to make clusters of tweets related to a target user's tweet, two conditions are considered whether it belongs to the lifetime of a target user's tweet or not. The tweet lifetime of a target user is defined for the time from the creation of a tweet to the time of a final retweet.

Variables indicated in the figure are as follows.

α: Early prior probability each document belongs to topic K,

β: Early prior probability each word belongs to topic K,

T: Time $t \in \{1,\dots,T\}$,

F: Set of users writing at time t, $f \in \{1,\dots,F\}$,

D: All of tweet user u writes at time t, $d \in \{1,\dots,M\}$,

W: Total words appeared in document d, $w \in \{1,\dots,N\}$,

η^f: probability of topic k appeared in user f, $f \in \{1,\dots,F\}$, $k \in \{1,\dots,K\}$,

θ^t: probability of topic k appeared within the lifespan of a tweet of a target user $t \in \{1,\dots,T\}$, $k \in \{1,\dots,K\}$,

φ^c: probability of word w belongs to topic k, $k \in \{1,\dots,K\}$, $w \in \{1,\dots,V\}$,

z: Topic ratio of word w in document d, $d \in \{1,\dots,M\}$, $w \in \{1,\dots,N\}$,

K: the number of topic $k \in \{1,\dots,K\}$,

π: Probability of selecting a topic of a target user rather than individual topic.

A tweet posted by a follower from the time of a tweet created until the last retweet probability can be associated with the target user's tweet.

Follower Opinion classification for Each Issue. The clustered tweets on each issue are analyzed for sentiment whether to express positive, negative or neutral opinion toward the opinion of a target user. Here, retweeting activities for a tweet can be considered as a positive opinion toward the target user for the issue. Based on retweeting relationship, tweets are classified as positive.

Table 1. Korean Twitter test collection

Target Users (Followees)	The number of Followers	The number of Tweets
Moon, Jaein	43,321	128,626
Saenuri party	37,213	111,954
Ahn, Cheolsoo	628,635	324,315
Byun, Heejae	70,299	248,653
Total	-	813,548

Table 2. The issues selected for each target user and the answer tweets

Target Users (Followees)	The number of Issues	The number of Tweets
Moon, Jaein	6	600
Saenuri party	4	400
Ahn, Cheolsoo	5	500
Byun, Heejae	7	700

Table 3. Experimental results of follower bias classification

Method	Precision	Recall	F1
Baseline SVM	0.640	0.548	0.590
SVM with influential users	0.683	0.591	0.633
Proposed method	**0.709**	**0.623**	**0.663**

For the analysis of other tweets, we used SVMs to classify a follower's opinion based on the bigram and trigram syllables in a sentiment dictionary for Korean characters to consider a user's writing patterns by using a variation of a word on Twitter.

4 Experiments

4.1 Korean Twitter Test Collection

To see the effectiveness of the proposed method, we construct a test collection for Korean Twitter data. Four target users are selected and all their tweets and all followers and the followers' tweets are collected from May 1 to May 31, 2013 via Twitter API. The users selected stand on political

Table 4. Results of each class (in Precision)

Method	Supporter	Non supporter	Neutrality
Baseline SVM	0.64	0.52	0.65
Proposed method	**0.84**	**0.55**	**0.71**

Table 5. Experimental results of tweet clustering

Method	Precision
TimeUserLDA	0.73
Proposed Method	0.75

Table 6. Experimental results of opinion classification

Method	Precision	Recall	F1
SVM with the morphemes	0.724	0.467	0.567
SVM with bi-gram syllables	0.689	0.521	0.593
SVM with tri-gram syllables	**0.735**	**0.563**	**0.637**

issues and have lots of followers. The performance is measured in recall, precision, and F1.

Table 1 shows the statistics of followers and the followers' tweets for each target user.

Table 2 shows the gold standard data for opinion classification of followers. Issues are selected from the clustered issues. 100 tweets are randomly selected from each issue.

4.2 Experimental Results

4.2.1. Results of Follower Polarity Classification

Comparison methods for a follower's polarity classification are as follows.

− Baseline SVM : Classification by SVMs by applying Korean characteristics [7],

− SVM with influential users on HITS [5]: Classification by SVM with influential users by the related work [5],

Table 7. Results of follower classification and opinion classification on each issue for the target user

Target User	Follower Classification	Social Issue	Opinion Classification		
			Support	Non-support	Neutral
Moon, Jaein	Supporter: 75 followers	"Moon Jae-In", "Fund"	20	1	0
		"Kwang-Ju", "5.18"	43	0	0
		"Jin-ju National Medical Center", "Closure"	27	2	0
		"Park young-sook", "pass away"	23	0	0
		"Roh Moo-hyun", "Foundation"	31	6	0
	Non-supporter: 13 followers	"Moon Jae-In", "Fund"	7	1	2
		"Kwang-Ju", "5.18"	5	2	1
		"Jin-ju National Medical Center", "Closure"	3	3	1
		"Roh Moo-hyun", "Foundation"	4	7	0
	Neutrality: 112 followers	"Moon Jae-In", "Fund"	18	1	0
		"Kwang-Ju", "5.18"	21	1	0
		"Jin-ju National Medical Center", "Closure"	18	0	1
Saenuri party	Supporter: 78	"NIS", "election"	21	5	0
		"Kwang-Ju", "5.18"	32	2	1
		"Park Geun-hye", "government"	36	3	0
		"Yoon chang-jung", "Blue House"	5	1	0
	Non-supporter: 35	"NIS", "election"	10	5	0
		"Kwang-Ju", "5.18"	14	5	2
		"Park Geun-hye", "government"	10	5	0
		"Yoon chang-jung", "Blue House"	6	3	2
	Neutrality: 87	"NIS", "election"	2	1	2
		"Kwang-Ju", "5.18"	6	0	0
		"Park Geun-hye", "government"	9	0	0
		"Yoon chang-jung", "Blue House"	9	0	1
Ahn, Cheolsoo	Supporter: 82	"Democratic Party", "new leadership"	27	6	0
		"Nowon", "Concert"	49	0	0
		"Kwang-Ju", "5.18"	39	0	1
		"Jin-ju National Medical Center", "Closure"	26	5	1
	Non-supporter: 24	"Democratic Party", "new leadership"	4	4	0
		"Kwang-Ju", "5.18"	4	1	1
		"Jin-ju National Medical Center", "Closure"	3	2	1
	Neutrality: 94	"Democratic Party", "new leadership"	6	3	0
		"Nowon", "Concert"	16	0	0
		"Park young-sook", "pass away"	29	1	5
		"Kwang-Ju", "5.18"	13	0	0
		"Jin-ju National Medical Center", "Closure"	2	0	1
Byun, Heejae	Supporter: 86	"5.18", "riot"	47	9	0
		"Ilbe" "North Korea"	26	0	1
		"Nancy Lang", "father"	12	0	0
		"Park Chung-hee", "Park Geun-hye""	16	6	2
	Non-supporter: 56	"Yoon chang-jung", "truth"	15	11	6
		"Marches" "Celebration"	10	1	4
		"Nancy Lang", "father"	7	2	1
		"5.18", "riot"	18	4	0
		"Roh Moo-hyun", "Kim Dae-jung	11	6	0
	Neutrality: 58	"Yoon chang-jung", "truth"	1	1	3
		"Marches" "Celebration"	4	2	2
		"5.18", "riot"	9	3	2
		"Ilbe" "North Korea"	1	0	3

- The proposed method: Classification by SVM with the proposed influential transmitters on HITS.

Table 3 shows results of follower polarity classification. The classification with influential transmitters achieves higher performance compared to the related work [5].

Table 4 shows the results of the classification for each class in precision. The proposed method in the supporting follower detection shows higher performance as compared to the comparative method. It shows that the method using influential transmitters is effective to classify the supporting followers.

4.2.2. Results of Social Issue Clustering

Comparison methods for tweet clustering are as follows.

- TimeUserLDA: LDA with time and users to find social issues [11],
- Proposed method: LDA model by applying the lifespan of tweets.

Table 5 shows the result of the social issue clustering. The 200 tweets are randomly selected from each cluster which is judged to be an issue for a target user among the clusters. The experimental result shows that the model using the lifespan of tweets is effective.

4.2.3. Results of Opinion Classification of a Follower

Comparison methods for classifying opinions are as follows.

- SVM based on morpheme features: SVM with morpheme features after Korean morphological analysis,
- SVM with bigram syllables: SVM with bigram syllables features to represent sentiments,
- SVM with trigram syllables: SVM with trigram syllables features to represent sentiments.

Table 6 shows the results of opinion classification. This result shows that newly-coined words and the variation of words by users on social network can be considered by using trigram syllables.

4.2.4. Analysis of Opinions for Followers

Even though a follower supports a target user, the follower can have an opposite opinion toward the target user. On each social issue of the target user, the supporting and non-supporting ratio is analyzed.

Table 7 shows the results of follower polarity and opinion classification for tweets on each issue for the target user. According to opinion classification for the particular issues by each target user, results of opinion classification exist users who have opposing opinions unlike the followers classification. This analysis shows that a follower is not always supporting toward the target user. A target user can take an advice by the followers' responses for the future action.

5 Conclusion

Detecting influential transmitters and applying on HITS algorithm is effective for follower polarity classification as for supporting, non-supporting or neutral. In order to specifically analyze opinions of followers, tweets of each follower are clustered based on LDA model by considering the lifespan of tweets according to an issue mentioned by the target user. Then the tweets are classified as a positive, negative opinion toward the target user. From the analysis of the experimental results, we found that the supporting followers generally agree the opinion of the target user; however, some are opposed to the particular issues. This research showed that the proposed method of the follower polarity classification and opinion classification depending on the issues is effective. Future researches should focus on methods for better user classification and less dependency on followers.

Acknowledgements

References

1. **Chen, C., Gao, D., Li, W., & Hou, Y. (2014).** Inferring topic-dependent influence roles of Twitter users. *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval,* pp. 1203–1206. DOI: 10.1145/2600428.2609545.

2. **Anagnostopoulos, A., Kumar, R., & Mahdian, M. (2008).** Influence and Correlation in Social Networks. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* pp. 7–15. DOI: 10.1145/1401890.1401897.

3. **Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K.P. (2010).** Measuring User Influence in Twitter: The Million Follower Fallacy. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media,* pp. 10–17.

4. **Weng, J., Lim, E.-P., Jiang, J., & He, Q. (2010).** Twitterrank: finding topic-sensitive influential twitterers. *Proceedings of the third ACM international conference on Web search and data mining,* pp. 261–270. DOI: 10.1145/1718487.1718520.

5. **Purohit, H., Ajmera, J., Joshi, S., Verma, A., & Sheth, A.P. (2012).** Finding Influential Authors in Brand-Page Communities. *Proceedings of the sixth international conference on weblogs and social media.*

6. **Cui, H., Mittal, V., & Datar, M. (2006).** Comparative Experiments on Sentiment Classification for Online Product Reviews. *Proceedings of the 21st National Conference on Artificial Intelligence,* pp. 1265–1270.

7. **Lim, J.S. & Kim J.M. (2014).** An Empirical Comparison of Machine Learning Models for Classifying Emotions in Korean Twitter. *Journal of Korea Multimedia Society,* Vol. 17, No. 2, pp. 232–239. DOI: 10.9717/kmms.2014.17.2.232.

8. **Maskeri, G., Sarkar, S., & Heafield, K. (2008).** Mining business topics in source code using latent Dirichlet allocation. *Proceedings of the 1st India software engineering conference,* pp. 113–120. DOI: 10.1145/1342211.1342234.

9. **Hong, L. & Davison, B.D. (2010).** Empirical study of topic modeling in twitter. *Proceedings of the First Workshop on Social Media Analytics,* pp. 80–88. DOI: 10.1145/1964870.

10. **Diao, Q., Jiang, J., Zhu, F., & Lim, E.P. (2012).** Finding bursty topics from microblogs. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics,* Vol. 1, pp. 536–544.

11. **Tsolmon, B. & Lee, K.-S. (2014).** Extracting Social Events based on Latent Dirichlet Allocation with Time and User Analysis. *Proceeding of the 37th Annual International ACM SIGIR conference,* pp. 1187–1190.

12. **Levy, D. & Nail P. (1993).** *Contagion: a Theoretical and Empirical Review and Conceptualization Reconceptualization.* General Psychology Monographs, 119, pp. 233–284.

Kwang-Yong Jeong received his Master degree in computer science and engineering from Chonbuk National University in 2015. His scientific interest is in information retrieval and social data analysis.

Kyung-Soon Lee received her MS and PhD degree in Computer Science from KAIST (Korean Advanced Institute of Science and Technology) in 1997 and 2001, respectively. Her scientific interest is in information retrieval, natural language understanding and text data mining.