Sentiment Analysis Based on Fuzzy Propagation in Online Social Networks: a Case study on TweetScope[*]

Duc Nguyen Trung[1] and Jason J. Jung[1,2][**]

[1] Department of Computer Engineering
Yeungnam University
Gyeongsan, Korea 712-749
duc.nguyentrung@gmail.com
[2] Universiti Malaya
Kuala Lumpur, Malaysia
j2jung@gmail.com

Abstract. Understanding customers' opinion and subjectivity is regarded as an important task in various domains (e.g., marketing). Particularly, with many types of social media (e.g., Twitter and FaceBook), such opinions are propagated to other users and might make a significant influence on them. In this paper, we propose a fuzzy propagation modeling for opinion mining by sentiment analysis of online social networks. Thereby, a practical system, called *TweetScope*, has been implemented to efficiently collect and analyze all possible tweets from customers.

Keywords: Sentiment analysis; Opinion mining; Online social media; Fuzzy propagation; Information diffusion.

1. Introduction

It is important for businesses to collect customers' feedbacks about their products and services in direct and more importantly indirect manners [15,16,17]. Online users have been creating a large amount of information (e.g., personal experiences and opinions) in various forms (e.g., rating, reviews, comments, and articles). Such "personal opinions" among users have be efficiently processed by using various learning methodologies (e.g., decision tree, clustering, and so on) [21].

Since many social networking services (SNS) have been emerged, they have enabled the customers to share and exchange their personal opinions. Then, these customers can either make a significant influence on others or get influences from the others. For example, if some of friends (or family) have shown any positive (and negative) comments against a certain item (e.g., news and products), one will have a similar feeling regardless of their own personal opinion [7,8,14,6].

More importantly, the SNS has shown significant power on information diffusion. Once a new piece of information is generated, the information can be propagated to a very large number of other uses in a short time. Particularly, as shown in Fig. 1, in Twitter,

---

[*] This paper is significantly revised from an earlier version presented at the 5th Asian Conference on Intelligent Information and Database Systems (ACIIDS 2013) held in Kuala Lumpur, Malaysia in March 2013.
[**] Corresponding author

tweets (i.e., t_1, t_2, and t_3) generated by users U_A and U_B can be broadcasted to other users who are *following* the user. As some of the following users (i.e., U_X) have *retweeted* the tweets, the information can be exponentially exposed to many users. This Twitter-based information diffusion process will be explained in more detail in Sect. 3.
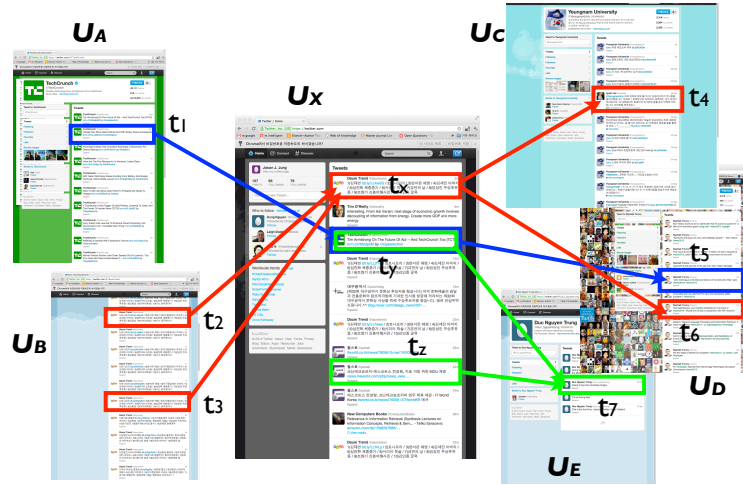

Fig. 1: A sample of informattion propagation on Twitter

In this context, it is important to detect and understand the so-called Word of Mouth (WoM) phenomena [1]. There have been many applications to be aware of how information is propagated through the social media [18,19]. Particularly, in terms of efficiently establishing *viral* marketing strategy, businesses need to understand customers' feedbacks and opinions [13,10]. In order to deal with this issue, we have focused on designing a mathematical model of information propagation on social media for sentimental analysis. The main research questions are

– Is there any relationship between emotional words and information propagation through social media?
– Is it possible for businesses to employ emotional words to increase effect of the information propagation?

Thereby, we have developed and evaluated a practical decision support system, called *TweetScope*, which is capable of extracting and visualizing the feasible information on marketing. We expect that the visual interface of this system can help decision makers to understand the diffusion patterns on tweets [9,11].

The outline of this paper is as follows. In the following Sect. 2, we show the previous work related to the social understanding systems. Sect. 3 describes the formalization of diffusion network on social networks and a practical sample about information propagation and the uncertain influence of friends opinion to a personal comments. Sect. 4 introduces a definition of fuzzy information propagation network base on a fuzzy relationship

between the microtexts diffused on Social network. In Sect. 5, we show the experimental data collection and data preprocessing. The experimental results and evaluations are discussed in Sect. 6. Sect. 7 draws our conclusion of this work and presents next research directions in the future.

2.   Related work

Understanding how information are diffused over on social networks has recently attracted much interest, especially in detecting and understanding the so-called Word of Mouth (WoM) phenomena [1]. There are already research works to be aware of how information is propagated through the social media [18,19]. Particularly, in terms of efficiently establishing *viral* marketing strategy, businesses need to understand customers' feedbacks and opinions [13,10].

Previous work [3], authors examined a dataset of political blog entries to determine relation between political opinions and increasing of feedback in term of quantity of comment; and the influence or diffusion of original political opinion to the following discussion is evaluated too. The result has shown that people are more interested in emotionally-charged discussion. Which blog entries being more emotional are received more feedbacks than another even if its sentiment is either positive or negative. Therefore, the experiment also pointed out that emotional statuses are potentially attract users interesting in information propagation environment on social networks. In case of Twitter, these statuses or tweets can be more diffused among users by sharing action such as retweeting, replies and so forth.

In order to evaluate relation between diffusion pattern and emotional statuses posted over social network, we have focused on designing a mathematical model of information propagation on social media for sentimental analysis. The main research questions are

- Is there any relationship between emotional words and information propagation through social media?
- Is it possible for businesses to employ emotional words to increase effect of the information propagation?

The model supports to detect the relation based on fraction between coverage and sensitivity value of the diffusion patterns, which called coverage rate. Statuses are classified as positive and negative depend on whether it contains emotional words or not. Such classification allows us to examine distribution of its coverage rate values to identify the difference between groups. For testing the model, we evaluate a case study on a practical system named TweetScope. The application monitors and analyzes data fetched from a text stream provided by Twitter by filtering tweets on the timeline of certain famous accounts who have a quite enough number of statuses and followers. It's capable of extracting and visualizing the feasible information on marketing. We expect that the visual interface of this system can help decision makers to understand the diffusion patterns on tweets [9,11].

3. Problem description and example

In order to formalize the information diffusion patterns, we have to define the following notations and show an example. In this paper, we call customers' comments on SNS as microtexts, since they are usually short.

Definition 1 (Microtext). *A microtext t is a piece of textual information. It is composed of three main features given by*

$$t = \langle TF, \tau, \Psi \rangle \tag{1}$$

which are i) term frequencies TF (how many times each term appears in the microtext), ii) timestamps τ (when the microtext was generated), and iii) a set of neighbors Ψ (who has been involved in the microtexts).

Given a microtext t, a set of term features $TF(t_i, w_k)$ can be extracted by measuring term frequencies in the vector-space model. It can be represented as

$$TF_t = \begin{bmatrix} w_1 & w_2 & \cdots & w_{|t|} \\ \frac{count(w_1)}{|t|} & \frac{count(w_2)}{|t|} & \cdots & \frac{count(w_{|t|})}{|t|} \end{bmatrix}^T \tag{2}$$

where $w_i \in t$ is a list of words in t, and function $count$ returns the number of occurrence of a term w_k. Also, $|t_i|$ is the length of t_i (i.e., the total number of words).

Definition 2 (Propagation network). *A propagation network $\mathcal{N}_\mathcal{P}$ is a network where information is diffused from one to other users via their relationship. It is represented as*

$$\mathcal{N}_\mathcal{P} = \langle \mathcal{U}, \mathcal{N} \rangle \tag{3}$$

where $\mathcal{U}$ is a set of users and $\mathcal{N} \subseteq |\mathcal{U}| \times |\mathcal{U}|$ is a set of relationship betweens the users.

Definition 3 (Linguistic Network). *A linguistic network $\mathcal{N}_\mathcal{L}$ is a network where people show their opinions about a subject or trend of news. Propagation information are restricted to microtexts have same subject w_k with term features $TF(t_i, w_k)$ equal or greater than a threshold α. A Linguistic network $\mathcal{N}_\mathcal{L}$ is represented as*

$$\mathcal{N}_\mathcal{L} = \langle \mathcal{U}, \mathcal{N}, S, \alpha \rangle \tag{4}$$

where S is a set of considered subjects, and $\alpha \in [0..1]$

In case of Twitter, the microtexts are called "tweets" showing opinions of users about a certain news. These tweets are diffused among users by retweet action. Retweet (in short, RT) is the powerful and unique function to propagate information to other users. Let consider an example of information propagation on Twitter shown in Fig. 1. We assume that users $U_A, U_B, ..., U_X$ have the directed relationships (by "following") among them. The users can share their opinion to their followers by posting a new tweet, or just retweet a tweet of their friends, e.g., U_A posts a tweet t_1 to show his or her opinion, user U_X is interested with the topic of t_1, he retweet it as a tweet t_y. The followers of U_X can continues share t_1 more deeper to their friends.

Rather than sharing their own feelings or friend's opinions, the user can summarize or comments a friend's tweet either intentionally or unintentionally, i.e. user U_x posts a

tweet t_x after observing the tweets t_2 and t_3 of user U_B to talk about an interesting topic. Hence, there exists an uncertain or fuzzy relationship between statuses of users in a group of friends [2].

In our previous work [4], we have defined a RT network $\mathcal{N}_{RT(t)}^{twt}$ to represent information propagation network of a specified tweet twt, and also the propagation pattern of twt by coverage rate ϕ^{twt}. We have extended these definitions of RT network over the linguistic network $\mathcal{N}_\mathcal{L}$, as following.

$$\mathcal{N}_{L,RT(t)}^{twt} = \langle U_{RT}^{twt}, N_{RT}^{twt}, \mathcal{T}_{RT}^{twt}, S, \alpha \rangle \tag{5}$$

where $U_{RT}^{twt} \subseteq U$, $N_{RT}^{twt} \subseteq U_{RT}^{twt} \times U_{RT}^{twt}$ and $\mathcal{T}_{RT}^{twt}$ is a set of timestamps when U_{RT}^{twt} have retweeted.

Each Retweet Network of a certain tweet twt has its own a propagation pattern $g(twt)$ by coverage rate ϕ^{twt} that indicates how many users twt is diffused to within a unit time. The interesting issue is that some $g(twt)$ can reach its own highest value of ϕ more quick than another; and their maximum value are distributed in various ranges, i.e. In Figure 2, $g_1(twt_i)$ has a peak later than $g_2(twt_j)$. In consider to compare between diffusion patterns, we represent the maximum value of a diffusion pattern $g(twt)$ and its time as following

$$\Phi^{twt} = [\phi_{max}, d] \tag{6}$$

where ϕ_{max} is highest value of $g(twt)$ and d is timestamp when it is happened. For easy accessing, we denote Φ_v^{twt} for ϕ_{max} value and Φ_t^{twt} for d value.

4. Fuzzy Propagation Model

By dividing the $Domain(\Phi_t^{twt})$ to a set of smaller ranges, we can map Φ_t^{twt} to linguistic variables which represent how quickly $g(twt)$ reaches its own highest value.

$$L^t = \left\{ l_i^t | l_i^t := [(i-1) \times \Delta^t, i \times \Delta^t] \right\} \tag{7}$$

where $i \in [1..m]$, $\Delta^t = \frac{max(\Phi_t^{twt}) - min(\Phi_t^{twt})}{m}$.

Beside that, when consider to all tweets of a certain subject, we also can divide the $\bigcup_{\forall twt} Domain(\Phi_v^{twt})$ to linguistic variables which represent how larger of Φ_v^{twt} as following

$$L^v = \left\{ l_i^v | l_i^v := [(i-1) \times \Delta^v, i \times \Delta^v] \right\} \tag{8}$$

where $i \in [1..n]$, $\Delta^v = \frac{max(\Phi_v^{twt}) - min(\Phi_v^{twt})}{n}$.

Definition 4 (Linguistic variables of Information propagation patterns). *Linguistic variables of Information propagation patterns are fuzzy sets used to cluster values Φ^{twt} to some similar groups. It is defined through linguistic variables of values and time of the highest peak of $g(twt)$ for all considered microtexts as following*

$$\mathcal{L} = \left\{ l | l \in L^v \times L^t \right\} \tag{9}$$
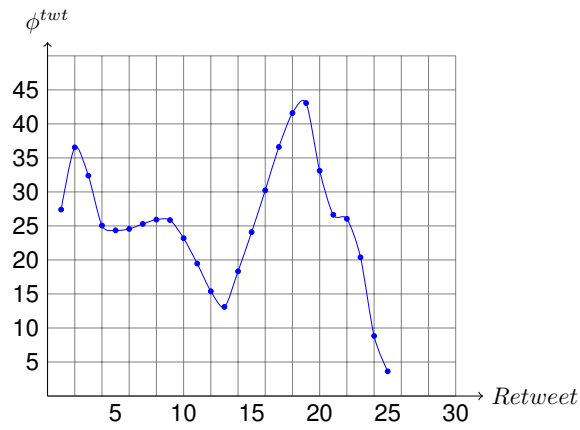
(a)
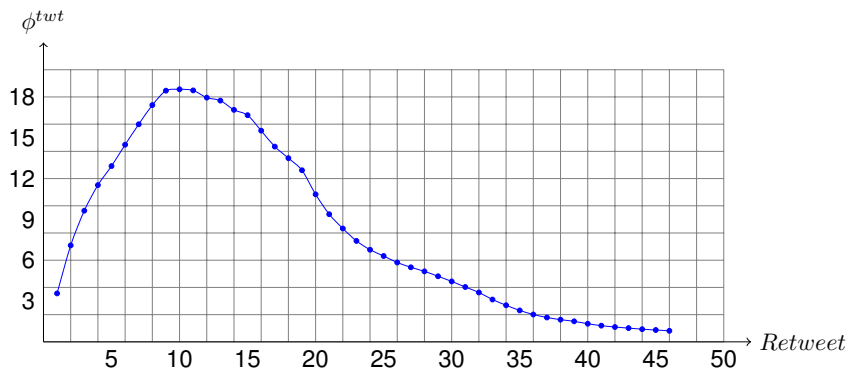
(b)

Fig. 2: Two diffusion patterns by (a) $g_1(twt_i)$ of @Windowsphone account, (b) $g_2(twt_j)$ of @SamsumMobile accounts

Definition 5 (Information diffusion relationship). *Information diffusion relationship* $f_{w_k}(M_u, t_i)$ *represents a semantic relationship between a set of microtexts and a specified one when we consider a certain subject. It is probability that a microtext* t_i *was posted by a user* u *after observing some microtexts* M_u *of his/her friends about a certain subject* w_k.

$$f_{w_k}(M_u, t_i) = p_{w_k}(M_u | t_i) \tag{10}$$

where $M_u = \{m \mid timestamp\ \mathcal{T}(m) < \mathcal{T}(t_i)\}$ *is a set of previous microtexts posted by friends of user u*

We can calculate $f_{w_k}(M_u, t_i)$ via $f_{w_k}(s_j, t_i)$ as following

$$f_{w_k}(M_u, t_i) = \frac{1}{|M_u|} \sum_{m \in M_u} f(m, t_i) \tag{11}$$

Definition 6 (Emotion transition network). *Emotion transition network* $\mathcal{N}_{\mathcal{E}}$ *where microtexts or opinions are diffused from one to other users over a Linguistic Network* $\mathcal{N}_{\mathcal{L}}$ *and the interaction between them (if any) is represented by a fuzzy relationship. Given a Linguistic network* $\mathcal{N}_{\mathcal{L}} = \langle \mathcal{U}, \mathcal{N}, S, \alpha \rangle$*, Emotion transition network* $\mathcal{N}_{\mathcal{E}}$ *is represented as*

$$\mathcal{N}_{\mathcal{E}} = \langle \mathcal{N}_{\mathcal{L}}, F, \beta \rangle \tag{12}$$

where $F = \{f_{w_k} \mid \forall w_k \in S \Rightarrow f_{w_k} : M x M \to [\beta..1]\}$ *is set of links between the microtexts with weight equal or greater than a threshold* β.
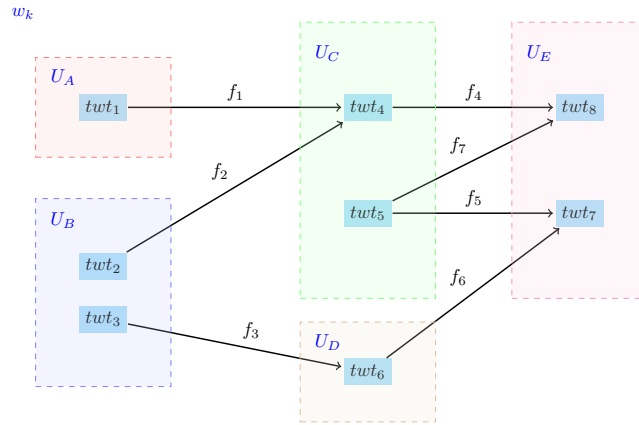

Fig. 3: A sample of emotion transition network of a specified subject w_k

Property 1 (Emotion status). Given a set of emotional words E and twt is a microtext. We denote twt^E to represent what emotional words does twt contains. Hence two able emotion status of a microtext twt are i) non-emotional correspond with $E = \emptyset$ ($|E| = 0$) and ii) emotional when $E \neq \emptyset$ ($|E| > 0$).

Property 2 (Emotion transition). Given an edge $(s_j, t_i)_{w_k}$ with weight $f_{w_k}(s_j, t_i)$, we call it is an emotion transition if emotion statuses of s_j, t_i are difference.

Property 3 (Source of a node). Given a microtext as a node t_i, $Source(t_i) = \{s | f(s, t_i) \in F\}$ is a set of nodes that has outgoing edges to t_i.

Property 4 (Derived of a node). Given a microtext a node t_i, $Derived(t_i) = \{s | f(t_i, s) \in F\}$ is a set of nodes that have incoming edges from t_i.

Mapping tweets diffused on $\mathcal{N}_\epsilon$ to space of information propagation linguistic variables will group tweets that have similar diffusion patterns. These information is used to track the dependence between information propagation patterns and emotional words. It's our next mission to build an evaluating method to clarify the relationship. Some researches already show the relationship between emotion [3][5] and information propagation in narrow fields, hence we assumed some hypotheses as following:

1. Tweets with emotional words are retweeted more quickly by following users of its owner, and also by followers of those more deeply over friendship network. Hence ratio between emotional tweets with small L_i^t value, and quantity of emotional tweets are higher than ones of non-emotional tweets.
2. Ratio between emotional tweets with large L_i^v value, and quantity of emotional tweets are higher than ones of non-emotional tweets.

5.   Experimental Results

5.1.   Implementation

For purpose to evaluate the propagation model and discovery relationship between emotional words and ability of information diffusion over the propagation network, we implemented a practical system called *TweetScope*, its architecture is described as Fig. 4 and Fig. 5 is a demo of *TweetScope* GUI (Graphical User Interface).

Tweet crawler is a basic component that has missions in i) listening on Twitter Stream Service to collect any tweets from monitored accounts and considered subjects; ii) storing the fetched information to a local database, it is preprocessed before iii) parsing by Entity Tagger to recognize any emotional words in content of tweets [20] [22] and also represent their subject in vector-space model.

Model constructor provides logical functions that help to built the diffusion model for next analytic steps. It includes some components i) RT Network builder create retweet network for each tweet base on friendship relationship between retweeting users; ii) Diffusion pattern builder calculate coverage and sensitivity values for the list of retweets; iii) Fuzzy propagation model constructor constructs terms of the information propagation model which is necessary for next steps.

Analyzer is an application layer in the framework of *TweetScope*. This layer is dynamic for multi-purposes of data analyzing, it contains some useful controller to calculate, extract and generate data for extend applications. In this work, the analyzer processes information diffusion patterns' parameters, maps them to linguistic variables space and provide a visual interface to help user to observe the result easily.
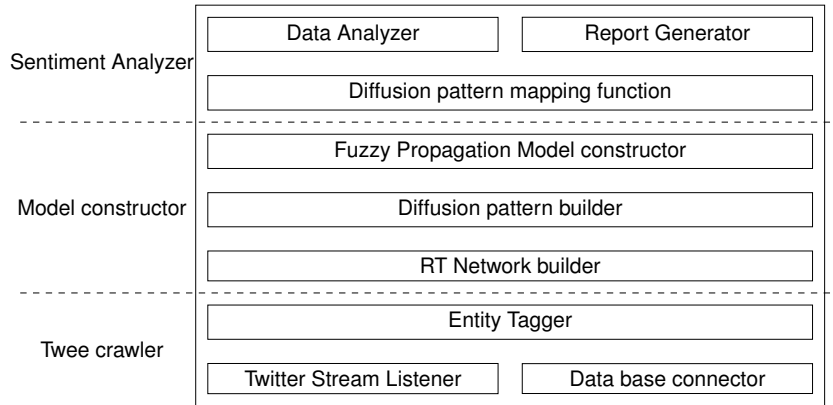
Fig. 4: System architecture of *TweetScope*

5.2. Data collection

We chose several Twitter accounts from 7 groups, i.e. IT, Mobile, Sport.... Their tweets and RT information are collected from 16th March 2012 to 27th November 2012. In the scope of evaluation, we have built their RT networks and classified tweets into two group i) emotional tweets what contain at least one emotional word and ii) non-emotional tweets. The statistical specification of the collected dataset is shown in Table 1

| Group	Number of Tweets	Number of Retweets	Number of Emotional Tweets
Fast food	6,553	29,603	979
IT	12,997	105,010	825
Luxury	9,489	15,845	1,078
Manufacturer	12,104	12,238	1,443
Mobile	9,959	19,363	1,450
Music tool	9,642	42,307	849
Sport	3,271	107,683	246
Grand Total	74,914	338,699	6,870

Table 1: Statistical of the collected dataset

6.   Discussion

Our *TweetScope* application can collect tweets from Twitter efficiently by using Twitter Stream service; user can access and generate useful data what represent how information propagate on Twitter using the above definition model with the collected data. However this study has several limits in practical collecting and processing. It covers only one SNS, lack in number of tweets and scale of friendship relationship network. Moreover,
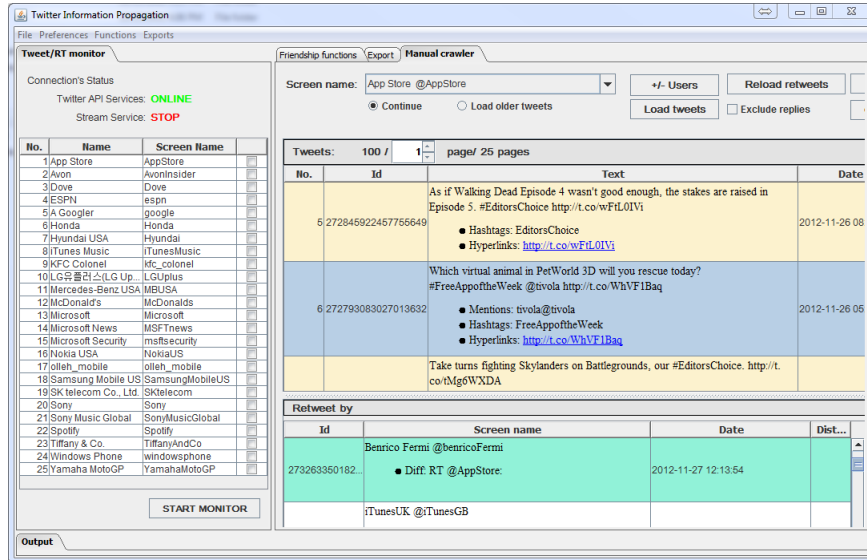
(a)
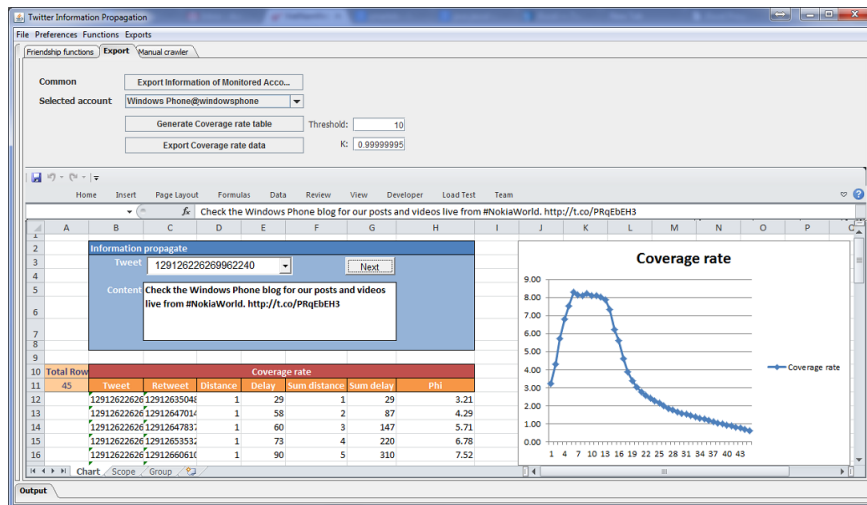
(b)

Fig. 5: GUI of *TweetScope* (a) the main interface (b) a sample interface of an export function

TweetScope cannot work online with dynamic information propagation, it has to fetch all data to a local database before processing and analyzing, hence it requires more on storage memory and performance ability of computer system.
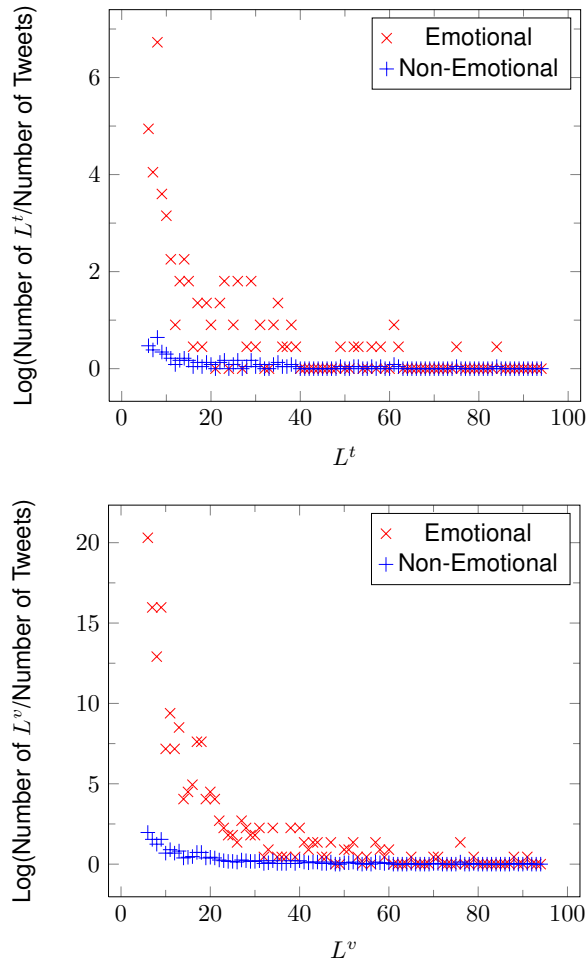

Fig. 6: A Distribution of ratio between number of tweets (or L_i) and tweet quantity of each kind (a) Distribution of L^t (b) Distribution of L^v with $i \in [5..95]$

We have used *TweetScope* to generate the calculated results as shown in Fig 6. In this processing, we only pick a set of tweets which have been retweeted over than a threshold 20 times, with quantity of linguistic variable of time and values respectively N = 100, M = 100. In fact, people can retweet any tweets with many reasons including subjects, location of followers, posting timestamp of tweet, and so on. Therefore, noise can exist

in the result if tweets reach highest ϕ value too soon or too late. To avoid the interference we ignored 5% of linguistic values from first and the end, i.e. We only consider L_i with $i \in [5..95]$. In plot chart (a), ratio of emotional tweets what reach highest ϕ more quickly, are larger than ones of non-emotional tweet. While the ratio of emotional tweets with each L_i^v are almost larger than ones of non-emotional tweet until to the end of axis L^v as shown in the plot chart (b), it suggest that emotional tweets can easily get higher value ϕ than non-emotional tweets.

7.   Conclusion

In this paper, we built a mathematical model of information propagation on SNS and a practical application named *TweetScope* to help answer the question about relationship of the effect of information propagation again emotional words used in the diffused tweets. For evaluating the model, we tested it with a sample dataset that are collected from Twitter using our application *TweetScope*. We found that tweets contain emotional words, are most frequently retweeted. The values of their linguistic variables are concentrated around the coordinate origin point in space of information propagation linguistic variables, where represent high frequency of retweeting. However, there many reason when user retweet a tweet hence the experiential result does not mention that non-emotional tweets are not more retweeted.

In future work, we are planning to extend our proposal by improving the model to track on multi properties of information propagation pattern other than the max value of coverage rate and its time; clarify further more about the relationship between positive/negative emotional words and efficiency of information propagation on Social Network Service. Besides that, we are going to implement our purposes by applying the model to a large scale dataset, analyzing and visualization to help user understand clearly how are information diffused and how to increase effect of the information propagation using emotional words in the advertising content on marketing and business field. Moreover, mash-up applications [12] will be implemented by using external open APIs.

References

1. Brown, J., Broderick, A.J., Lee, N.: Word of mouth communication within online communities: Conceptualizing the online social network. Journal of Interactive Marketing 21(3), 2–20 (2007)
2. Brunelli, M., Fedrizzi, M.: A fuzzy approach to social network analysis. In: Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining. pp. 225–230. ASONAM '09, IEEE Computer Society, Washington, DC, USA (2009)
3. Dang-Xuan, L., Stieglitz, S.: Impact and diffusion of sentiment in political communication – an empirical analysis of political weblogs. In: Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media. pp. 427–430. The AAAI Press, Dublin, Ireland (2012)
4. Duc, N.T., Jung, J.J., Lee, N., Kim, J.: Thematic analysis by discovering diffusion patterns in social media: An exploratory study with tweetscope. In: Proceedings of the 5th Asian Conference on Intelligent Information and Database Systems (ACIIDS 2013). Lecture Notes in Computer Science, Springer, Kuala Lumpur, Malaysia (2013)

5. Huffaker, D.: Dimensions of leadership and social influence in online communities. Human Communication Research 36(4), 593–617 (2010)
6. Jung, J.J.: An empirical study on optimizing query transformation on semantic peer-to-peer neworks. Journal of Intelligent & Fuzzy Systems 21(3), 187–195 (2010)
7. Jung, J.J.: Ontology mapping composition for query transformation on distributed environments. Expert Systems with Applications 37(12), 8401–8405 (2010)
8. Jung, J.J.: Reusing ontology mappings for query segmentation and routing in semantic peer-to-peer environment. Information Sciences 180(17), 3248–3257 (2010)
9. Jung, J.J.: Service chain-based business alliance formation in service-oriented architecture. Expert Systems with Applications 38(3), 2206–2211 (2011)
10. Jung, J.J.: Attribute selection-based recommendation framework for short-head user group: An empirical study by movielens and imdb. Expert Systems with Applications 39(4), 4049–4054 (2012)
11. Jung, J.J.: Computational reputation model based on selecting consensus choices: an empirical study on semantic wiki platform. Expert Systems with Applications 39(10), 9002–9007 (2012)
12. Jung, J.J.: Contextgrid: A contextual mashup-based collaborative browsing system. Information Systems Frontiers 14(4), 953–961 (2012)
13. Jung, J.J.: Discovering community of lingual practice for matching multilingual tags from folksonomies. Computer Journal 55(3), 337–346 (2012)
14. Jung, J.J.: Evolutionary approach for semantic-based query sampling in large-scale information sources. Information Sciences 182(1), 30–39 (2012)
15. Jung, J.J.: Semantic annotation of cognitive map for knowledge sharing between heterogeneous businesses. Expert Systems with Applications 39(5), 1245–1248 (2012)
16. Jung, J.J.: Semantic optimization of query transformation in a large-scale peer-to-peer network. Neurocomputing 88, 36–41 (2012)
17. Jung, J.J.: Cross-lingual Query Expansion in Multilingual Folksonomies: a Case Study on Flickr. Knowledge-Based Systems, 42, 60–67 (2013)
18. Kim, M., Xie, L., Christen, P.: Event diffusion patterns in social media. In: Breslin, J.G., Ellison, N.B., Shanahan, J.G., Tufekci, Z. (eds.) Proceedings of the 6th International Conference on Weblogs and Social Media (ICWSM 2012). The AAAI Press, Dublin, Ireland (2012)
19. Pham, X.H., Jung, J.J., Hwang, D.: Beating social pulse: Understanding information propagation via online social tagging systems. Journal of Universal Computer Science 18(8), 1022–1031 (2012)
20. Salmeron, J.L.: Fuzzy cognitive maps for artificial emotions forecasting. Applied Soft Computing 12(12), 3704–3710 (2012)
21. Sen, S., Lerman, D.: Why are you telling me this? an examination into negative consumer reviews on the web. Journal of Interactive Marketing 21(4), 76–94 (2007)
22. Strapparava, C., Mihalcea, R.: Learning to identify emotions in text. In: Proceedings of the 2008 ACM symposium on Applied computing. pp. 1556–1560. SAC '08, ACM, New York, NY, USA (2008)

Duc Nguyen Trung is a PhD student in Computer Engineering Department of Yeungnam University, Korea. He received his B.Eng. degree from Information Technology Department of Vietnam Maritime University in 2002, where his work as a lecturer since 2007. He received his M.S. degree in Mathematical Foundation for computer and computing systems from Science University of Vietnam National University, HaNoi in 2006. He is interested in Database systems, Software development, Intelligent computation and Machine learning. He can be contacted at: duc.nguyentrung@gmail.com.

Jason J. Jung is a corresponding author of this paper. He is an associate professor in Yeungnam University, Korea, since September 2007. He was a postdoctoral researcher in INRIA Rhône-Alpes, France in 2006, and a visiting scientist in Fraunhofer Institute (FIRST) in Berlin, Germany in 2004. He received the B.Eng. in Computer Science and Mechanical Engineering from Inha University in 1999. He received M.S. and Ph.D. degrees in Computer and Information Engineering from Inha University in 2002 and 2005, respectively. His research topics are knowledge engineering on social networks by using machine learning, semantic Web mining, and ambient intelligence.