Analysis the Influence of Official Information on Network Public Opinion based on Visualization Method

Chunjiang Dong[1]

Tianjin University
Tianjin, 300350, China
E-mail: `dongcj@tju.edu.cn`

Geng Sun

Tianjin University
Tianjin, 300350, China
E-mail: `tosungeng@gmail.com`

In order to sway the public opinion effectively, Latent Dirichlet Allocation(LDA) topic model is used in this paper to visualize and analyze the influence of official information on online public opinion. Taking the event "Terminal High Altitude Area Defense(THAAD) in South Korea" as a case study, this paper takes the topics of relevant data of the event as an example to calculate the similarity between the topics and analyze the key events. The visual analysis of the examples shows that the official press conference has a significant effect on the online public opinion in two days to four days, pushing the related topics to the climax or bottom. Moreover, in the process of event development, the similarities tend to be basically the same. This visualization can provide reference to relevant departments on the regulation of online public opinions, that is, it is most appropriate to guide and regulate online public opinion within 2 days to 4 days before the expected time.

[1]Speaker

1.Introduction

With the boom of the Internet, any emergencies may cause media sensation. m. The wide spread of public opinion on the Internet means that if there is a lack of effective regulation, network public opinion is likely to provoke rumors and even cause panic. Thus, analysis and supervision of network public opinion in the social management are particularly important. However, traditional processing of Internet social networking data will involve high human resources costs, and the overall grasp of the event and the development trend of the analysis are not comprehensive. In this paper, we choose the microblogging as a research carrier to visualize and analyze the microblogging hotspot "THAAD in South Korea" and thus reflect the network of public opinion. This method can free us from reading a lot of text and analyze the trend of network public opinion, thereby facilitating the relevant departments to find problems in the network of public opinion to guide the right.

2.Relation Work

In recent years, scholars both at home and abroad have made many contributions to the research of topic visualization, putting forward a variety of effective visualization projects[1,2]. These projects can be roughly divided into two categories: topic visualization projects based on static and dynamic methods. In the static topic visualization researches, the major visualization project is word cloud, and many studies have provided methods to improve it. Wordle can generate different word layouts in accordance with various permutations. William Williams[3] applied Wordle into research on management education. The dynamic topic visualization researches mostly used rivers as the metaphor to show the evolution of topics. TextFlow[4] is the perfect combination of topic information mining and visualization technology, which is mainly used to analyze the variation patterns among multiple topics. With the increase of topic numbers, the simple description of topics could be visually confusing. In order to sort out this problem, HierarchicalTopics[5] set up a layered structure to avoid semantic interference in the large text collection. TopicLens[6] introduced the visualization project for topic model real-time analysis from the perspective of lens. They put forward dynamic hierarchical rank-2 nonnegative matrix factorization(DH-NMF) Topic model algorithm. TopicLens is able to help the users to find out real-time rules and contents of sub-topics in the data set. Topicpanorama[7] builds hierarchical structure of topics with semi-automatic method, and then constructs panorama of multiple data sources by clusters of similar topics. PhenoLines[8] analyzes the evolution of disease types by visualization method on the basis of topic model for clinical data, and provides a new solution to the depiction and optimization for the topic model.

There are many studies that combined topic model with visualization analysis on the rise. Firstly, to meet different needs, document data set has been organized into meaningful multi-dimensional form[9]、 cluster[10] or relational network[11], then used to analyze the rules by visualization tools and methods.

3.Overview of System

The system mainly consists of data collection and processing, LDA[12] topic extraction, visualization and visual analysis. As shown in Figure 1.
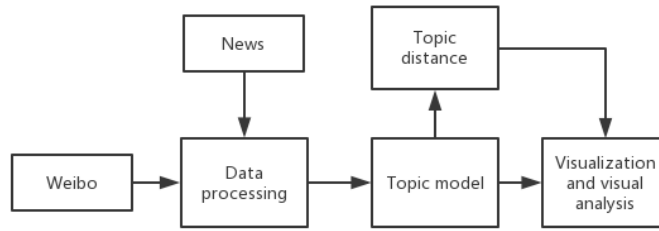
Figure 1: Diagram of System Components

3.1 Data Processing

The data collection and processing are mainly composed of web crawler and word segmentation tools. The web crawler can obtain micro-blog information posted by the users of Sina micro-blog, but the micro-blog information gathered here is non-structural textual data, mainly including the users' basic information, textual content, post time, number of shares and comments of micro-blogging. After the textual information is obtained, IKAnalyzer is used to split the textual content in micro-blog into character stream. Through the segmentation processing of IKAnalyzer, words with confusing meaning can be removed, and the mistake of splitting the words like the names of the characters and some professional terms can be avoided as well. The processed data can improve the accuracy of topic analysis, and the content of the topic will be more distinctive.

3.2 Topic Relation Model

After LDA extraction, we will get the distribution of each topic on topic words. This paper uses the Jensen-Shannon(JS) divergence between the two distributions. In order to calculate the distance between the two topic distributions, the topic words of the two topics shall be the same. If topic t does not contain the keyword w_i, that is, the relevancy of topic t and w_i is 0, then

$$P(w_i \mid t) = 0 \tag{3.1}$$

Thus topical words of the two topics to be calculated shall be unified first, setting the distribution of keywords not included as 0. Then, to satisfy the sum of topic distribution possibility to 1, we normalize the data. The core formula is as follows:

$$P(w_i \mid t) = \frac{p(w_i \mid t)}{\sum_{i=1}^{n} p(w_i \mid t)} \tag{3.2}$$

Among them, $P(w_i|t)$ denotes the probability value after normalization, and $p(w_i|t)$ denotes the probability value before normalization. For both distributions, the similarity between the distributions can be calculated according to the JS divergence. For the two discrete distributions of P and Q, p_i and q_i represent the probability of the ith element of distribution P and Q, respectively. The divergence JS is calculated as follows:

$$KL(P \parallel Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \tag{3.3}$$

$$JS(P \parallel Q) = \frac{1}{2} KL(P \parallel \frac{P+Q}{2}) + \frac{1}{2} KL(Q \parallel \frac{P+Q}{2}) \tag{3.4}$$

After the above processing, we can get the distance between the topics. According to the characteristics of JS divergence, the larger the distance between the two topics is, the higher the similarity between the two topics is because JS divergence have characteristic of symmetry.

4.Case Study

Through specific cases, our paper finds out rules and problems by using the method of visualization analysis which has shown an edge in dealing with the hot events in micro-blog.

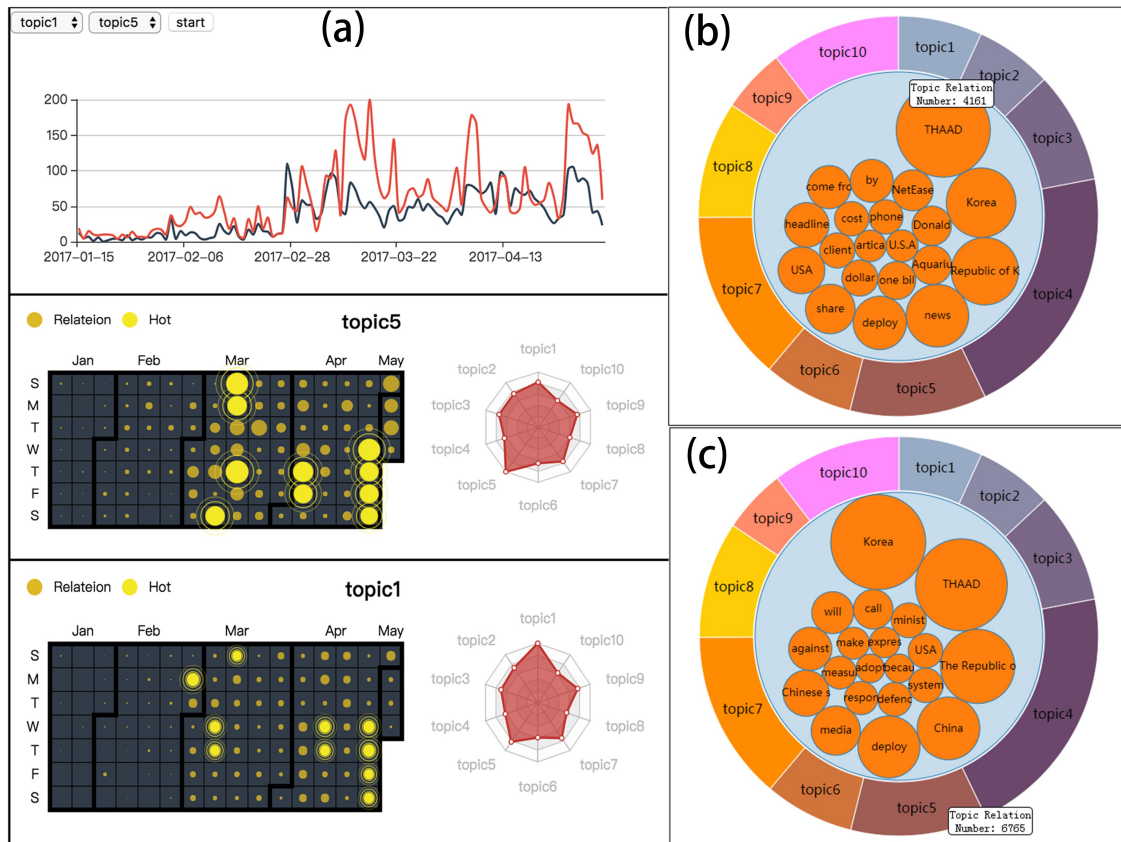## 4.1 Topic Selection of THAAD in South Korea

Figure 2: Topic Relation Analysis View

THAAD is the transliteration of terminal high altitude area defense system, known as the THAAD system at the early stage. The United States Army used a kinetic kill terminal defense system to intercept short-range and medium range ballistic missiles, which is a part of the US National Missile Defense Agency, aiming at intercepting Scud missiles and other tactical ballistic missiles. THAAD in South Korea has been a major concern of China, since the deployment of THAAD in South Korea has posed a threat to China. Ever since the beginning of 2017, the event has drawn the attention of netizens in China. On February 27, 2017, Lotte Group decided to provide premises for the deployment of THAAD in South Korea. Since then, the Chinese netizen's worries about THAAD in South Korea have kept rising and have become a topical event in Sina micro-blog, so we take the research data of THAAD in South Korea in the micro-blog as case study.

Due to the need of research, we use "THAAD", "THAAD Lotte", "Lotte THAAD", "Lotte" as keywords to extract data from Sina micro-blog. we collect the data during the timeperiod between 2017-01-15 and 2017-05-03, about 70,000 data of micro-blog. Then according to the news, we get 4 key events from THAAD inSouth Korea. Taking the THAAD in South Korea event data as an example, we combine the visualization and visual analysis to analyze the influence of official information on the network public opinion.

4.2 Case 1: The Analysis of Development Trend of Topic

When focusing on the development of an event, the relationship between the event and topic is very important to the analysis of the whole event. In accordance with the calculation

results of the topic relational model in 3.2, we compare the development trend of the two topics with high similarity. Taking the Figure 2 (a) as the entry of system analysis, when we select the topic1 in Figure 2 (b) we will get calendar and radar graphs in Figure 2 (a). As we know from the radar graph, topic1 has the highest similarity with topic5, so we select topic5 in Figure 2 (c), and get calendar and radar graphs in Figure 2 (a), and then select topic1 and topic5 in the topic view, draw the trend curve of two topics. As shown in Figure 2 (a). From Figure 2 (a) we can get the information that two similar topics have almost the same trend at climax. Compared Figure 2 (b) with (c), it is easy to see that the time period for heated discussion of two topics is from March 8, 2017 to March 13, 2017, and April 26, 2017 to April 30, 2017 respevtively. When we analyze the development trends of 10 topics, we can get the changing of all topics.

4.3 Case 2: The Impacts of Key Events on the Development Trend of Topics

In the development process of THAAD event, the key events play a vital role in the whole event. In this paper, we take the official statement as the key event to observe the impact of key events on micro-blog public opinion. Trough the collection of news information, we have obtained 4 key events between January 15, 2017 and May 4, 2017, and the details are in Table 1.

DATE	CONTENT	DAYS
2017-02-17	Lotte Board of Directors decided to provide land for the deployment of THAAD	After 4 days
2017-03-07	Part equipment of THAAD system arrived in the US military base in Korea	After 3 days
2017-03-17	China against THAAD, asked to stop the relevant deployment	After 2 days
2017-04-26	U.S. Department of Defense said, will continue to upgrade the THAAD system in South Korea	After 2 days

Table 1: Details of Key Events and Days of Impact from Figure 3

Four key events are corresponding to Event 1 to Event 4 in Figure 3. Taking Event 1 and Event 4 as examples, before Lotte Group showed its support for the the deployment of THAAD, the attention paid to THAAD has been relatively low, but when Lotte Group released the news of providing premises for the deployment of THAAD, the number for the response to THAAD in micro-blog increased dramatically in 4 days, among which the topic 3 has the most significant change, which has put the event into the first peak phase. After the United States Department of Defense released the news of upgrading support for the THAAD in the future, the total number of response to the THAAD event in micro-blog has increased significantly, reaching the climax of THAAD event. The topic relation view in visualization system can be combined to analyze the influence of key event on single topic. According to the development trend of topic1 and topic5 in the case 1, after the news release of upgrading support for THAAD in the future from the United States in April 26, the heated discussion of Topic 1 and Topic 5 between April 26, 2017 and April 30, 2017 has been sublimated. The contents of the two topics can be found from Figure 2 (b) and Figure 2 (c), which are mainly about the opposition and resistance for THAAD events. Thus the influence of official information on the network public opinion can be shown.
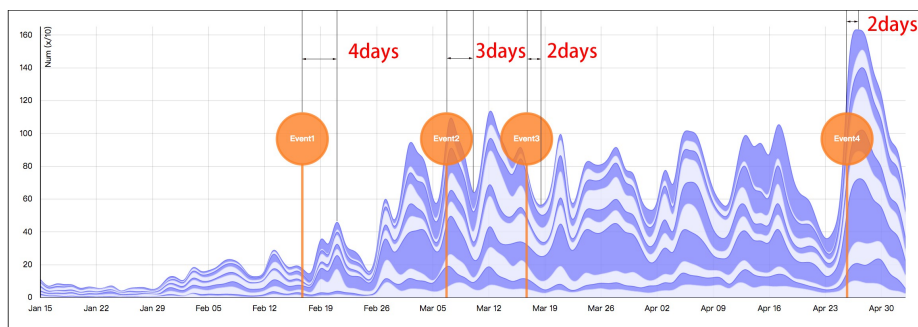
Figure 3: The Impact of Key Events on Topic Development

Through Figure 3 and the above case analysis, it can be shown clearly after the news release from the official authority, network public opinion will have obvious changes, after 1 to 2 days of fermentation period, the event will reach a climax within 2 to 4 days. The conclusion here is of great significance. When the network public opinion outbreaks, relevant departments could carry out appropriate guidance to achieve expected results via the release of official information. It will achieve the best guiding effects within 2 to 4 days.

5.Discussion

In research on the analysis of network public opinion, most researchers focus on the emotion and attention of users, and pay no attention to the influence of the publication of official information on network events. Henning Hamer [13] et al. used convolutional neural networks to visualize Twitter data. In their paper, we can see the change and competition between the topics. Compared to their work, we analyze both the trend of topics in public opinion online and the role of official information in swaying online public opinion. With the assistance of our method, relevant departments can not only timely guide the trend of online public opinion, but also learn how to effectively guide the direction of online public opinion.

6.Conclusion

In this paper, we provide a visualized system to analyze the influence of official information on the topic of the micro-blog, and the trend of the topic change over time for hot events in micro-blog. In the selection of topics, we choose the real-time event "THAAD into South Korea". It is one of the most representative topics in micro-blog in recent days. After data collection and cleaning, we adopt LDA algorithm to extract topics from the text information, calculating the similarity among those topics, so as to analyze the trend of the topic change with time and relationship among the topics of "THAAD into South Korea". Through the design of visualization system, the node of key events for "THAAD into South Korea" was added into the system to analyze the impact of official information on the development trend of hot events. Based on the above case analysis and visualization system, we can draw the conclusion as follows: in a micro-blog case, the changing trend for topics with high similarity is basically similar, while changing trend for topics of low similarity is basically different. Affected by the official information, public opinion in micro-blog shows obvious change and suffers from the greatest impact within 2 to 4 days after release of the official information, and the number of followers will peak or push down. If relevant departments hope to carry out the guidance and adjustment of network public opinion through official information, the most appropriate time to take measures is within 2 to 4 days before the event is out of control.

References

[1] G.-D. Sun, Y.-C. Wu, R.-H. Liang, and S.-X. Liu. *A Survey of Visual Analytics Techniques and Applications: State-of-the-Art Research and Future Challenges*[J]. Journal of Computer Science and Technology, 28(5): 852-867(2013)

[2] S. Liu, W. Cui, Y. Wu, and M. Liu. *A survey on information visualization: recent advances and challenges*[J]. The Visual Computer, 30(12): 1373-1393(2014)

[3] W. Williams, E. L. Parkes, and P. Davies. *Wordle: A method for analysing MBA student induction experience*[J]. The International Journal of Management Education, 11(1): 44-53(2013)

[4] Cui W, Liu S, Tan L et al. *TextFlow: Towards better understanding of evolving topics in text*[J]. IEEE Transactions on Visualization and Computer Graphics, 17(12): 2412- 2421(2011)

[5] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky. *HierarchicalTopics: Visually exploring large text collections using topic hierarchies*[J]. IEEE Transactions on Visualization and Computer Graphics, 19(12): 2002–2011(2013)

[6] M. Kim, K. Kang, D. Park, J. Choo, and N. Elmqvist. *TopicLens: Efficient Multi-Level Visual Topic Exploration of Large-Scale Document Collections*[J]. IEEE Trans Vis Comput Graph, 23(1): 151-160(2017)

[7] S. Liu, X. Wang, J. Chen, J. Zhu, and B. Guo. *TopicPanorama: A full picture of relevant topics*[C]. In Proceedings of the IEEE Conference on Visual Analytics Science and Technology. 183–192(2014)

[8] Michael Glueck, Mahdi Pakdaman Naeini, Finale Doshi-Velez, Fanny Chevalier, Azam Khan, Daniel Wigdor, Michael Brudno. *PhenoLines: Phenotype Comparison Visualizations for Disease Subtyping via Topic Models*[J]. IEEE Transactions on Visualization and Computer Graphics, PP(99), pp. 1-1(2017)

[9] Y. Chen, L. Wang, M. Dong, and J. Hua. *Exemplar-based visualization of large document corpus*[J]. IEEE Transactions on Visualization and Computer Graphics, 15(6): 1161-1168(2009)

[10] H. Lee, J. Kihm, J. Choo, J. Stasko, and H. Park. iVisClustering: *An interactive visual document clustering via topic modelling*[J]. In Computer Graphics Forum. Wiley Online Library. 31: 1155–1164(2012)

[11] B. Gretarsson, J. Odonovan, S. Bostandjiev, T. Hollerer, A. Asuncion, D. Newman, and P. Smyth. *TopicNets: Visual analysis of large text corpora with topic modelling*[J]. ACM Transactions on Intelligent Systems and Technology. 3(2):23, 2012.

[12] Ian Porteous, David Newman, Alexander Ihler, Arthur Asuncion, Padhraic Smyth, and Max Welling. *Fast collapsed gibbs sampling for latent dirichlet allocation*[C]. InProceeding of the 14th ACM SIGKDD inter-national conference on Knowledge discovery and data mining. New York, NY, USA, KDD'08, 569–577(2008)

[13] Henning Hamer; Andreas Merentitis; Nikolaos Frangiadakis; Sergey Shukanov. *Automatic visual analysis of real-world events covered by social media using convolutional neural networks*[C].2015 IEEE International Conference on Data Mining Workshop (ICDMW). 1-6(2015)