Trend Analysis for Online Comments

Jiangiang Hou* and Yupin Cheng
Department of Industrial Engineering and Engineering Management
National Tsing Hua University
E-mail:*adamhou@ie.nthu.edu.tw

Abstract. To know the trend of a specific news related to school, the staff firstly tend to search for the comments related to the news, then read and analyze all the comments to find the characteristics of comments related to educational news. Secondly, they will review all the comments to shape the topics and categorized all the comments into the above topics. However, there are two problems exists currently. First of all, current social media platform cannot automatically analyze the trend of comments related to specific school news. Second, staff needs to waste a lot of time reading and judging the topics of comments related to a specific news so as to get the trend of a specific school news. To solve the above problem, we proposed a preparation stage to figure out the characteristics of comments related to educational news, the principles of generating topics related to comments and the principles of judging the topics of comments. After that, we proposed a Trend Analysis of Comment Topics Methodology, it includes four stages, which are collecting comment characteristics, judging comment topics, judging the topic of each comment, analyzing comment trend. This methodology we proposed can assist user to get the trend of specific news related to school efficiently and accurately.

1. Introduction
As readers want to know the trend of comments related to a specific issue, they firstly need to search for the news related to the issue and collect the comments into the statistical application, such as Excel, etc. After collecting all the comments, readers need to review all the comments to subjectively decide the topics of these comments. In order to solve the above problems, we proposed a To-Be Model to assist readers accurately and quickly acquire the trend of comments related to his/ her interested issue. Once readers finish collecting the comments of his/her interested issue, they can upload the comments to the system. Then, the system will automatically retrieve six characteristics, which are people characteristics, object characteristics, issue characteristics, criticism characteristics, praise characteristics and other characteristics, of each comment, decide the topics of comments, classify each comment to the topic it belonged, and provide multiple visualization graphs. As a result, readers can capture the trend of comments by the multiple visualization graphs. In all, the To-Be Model we proposed is to help readers reduce wasting time in doing the routine work and capture the trend of comment topics accurately and quickly.

2. Literature Review
To assist readers quickly and accurately capture the trend of comment topics, our study developed a methodology, which is called “Trend Analysis of Comment Topics”, to automatically execute three works, which are retrieve the critical characteristics of comments, using Self-Organization featuring Method (SOM) method to cluster the comments, and visualizing the trend of comment topics in multiple graphs. We will introduce three related research works as following...
2.1. Retrieval of Critical Characteristics of Comments
To retrieve critical characteristics of comments, the related works focus on evaluating the importance of words by their traits in comment. As for the traits of words, some studies have retrieved characteristics of comments based on the semantic of words (Elena et al., 2013), some studies have retrieved characteristics of comments based on the position of words (Shabina et al., 2016), other studies have retrieved characteristics of comments based on the amounts of words (Amrutha et al., 2015). As a whole, the related researches can be applied to retrieve different characteristics by the trait of words; therefore, these researches can help readers efficiently capture the characteristics of words in comments.

2.2. Cluster/Classification of Comments
To determine the topics of comments or compare the difference among comments, related works have applied the cluster or classification method for it. As for determine the topics of comments, related studies mainly group the comments with similar words by specific cluster or classification method. Furthermore, some of these studies used the semantic of words to compare the similarity among comments (Williams et al., 2016), while others used the amounts of words to compare the similarity among comments (Supatta et al., 2015). On the other hand, few studies applied cluster or classification method to group comments by multiply similar characteristics (Wang et al., 2017).

2.3. Visualization of Critical Information Related to Comments
To visualize the critical information of comments, related works focus on visualize the critical information of comments of demonstrate the relationship among comments. As for visualizing the critical information of comments, some studies have demonstrated the characteristics of comments by multiple visualization graphs, such as line chart and bar chart (Hu et al., 2017). On the other hand, some studies have illustrated the relationships among comments in multiple visualization graphs, such as network graphs (Wen et al., 2016).

3. Methodology
Our research proposed a “Trend Analysis of Comment Topics” methodology, including a pre-operation section, which is called “Analysis of comments related to educational news”, and a methodology itself. As for the pre-operation section, our research defines the six characteristics of comments related to educational news, aggregates the principles of determining topics of comments and aggregates the principles of determining the topic of each comment belonged. After getting the results from previous section, our research develops a trend analysis of comments topics methodology to retrieve the characteristics of comments, cluster comments with similar characteristics by SOM method, determine the topic of each group by Association Rule method, classify new comments into the topic it belonged by calculating the similarity, and visualize the critical information of topics by multiple graphs.

3.1. Retrieval of Critical Characteristics of Comments
In order to develop a systematical method of automatically determining the topics of comments, our research defines six characteristics of comments, aggregates principles of determining the topics of comments, and aggregates principles of determining the topic of each comment belonged.

3.2. Retrieval of Characteristics from Comments
In this phase, the trend analysis of comment topics model can be applied to identify clarify critical information of each comment by retrieving six characteristics, which are “people characteristics”, “object characteristics”, “issue characteristics”, “criticism characteristics”, “praise characteristics”, and “other characteristics”.

3.3. Analysis of Topics from Comments
In this phase, the trend analysis for comment topics model is applied to cluster the comments with similar characteristics by SOM method and determine the topic of comments by Associate Rules.
3.4. Determination of Topic of Each New Comment Belonged
This phase aims at calculating the similarity among the matrix of new comments and the matrix of topics to clarify topic of each new comment belonged.

3.5. Analysis of Trend of Comments
This phase aims at demonstrating the trend of comments by four visualization graphs, which are line chart based on the numbers of comments of each topic contains, mutually relationship chart based on the similarity between two topics and the numbers of comments of each topic contains, bubble chart based on each time interval that the numbers of comments of each topic contains and pie chart based on the similarity among a comment and each topic.

4. Conclusions
Due to lack of systematic method, information seekers must spend a lot of time manually analyzing comments and subjectively determining the topics of comments to capture the trend of comments. This paper proposed a To-Be Model to assist readers accurately and quickly acquire the trend of comments related to his/her interested issue. Once readers finish collecting the comments of his/her interested issue, they can upload the comments to the system. As a whole, the To-Be Model we proposed is to help readers reduce wasting time in doing the routine work and capture the trend of comment topics accurately and quickly.

5. References
[1] Amrutha, B. and Mintu, P., 2015, “Keyword based Tweet extraction and detection of related topics,” Procedia Computer Science, Vol. 46, pp. 364-371.
[2] Elena, L. and Manuel, P., 2013, “Towards automatic Tweet generation: A comparative study from the text summarization perspective in the journalism genre,” Experts Systems with Applications, Vol. 40, pp. 6624-6630.
[3] Hu, M., Wongsuphasawat, K. and Stasko, J., 2017, “Visualizing social media content with sentence tree,” IEEE Transcations on Visualization and Computer Graphics, Vol. 23, pp. 1-10.
[4] Shabina, D., Harmanish, T. and Kavita, T., 2016, “NLP and ontology based clustering – An integrated approach for optimal information extraction from social Web,” International Conference on Computing for Sustainable Global Development, pp. 1765-1770.
[5] Supatta, V., Parinya, S., Pisit, C. and Choochart, H., 2015, “A comparison of similarity measures for online social media thai text classification,” Electrical Engineering / Electronic, Computer, Telecommunications and Information Technology, 2015 12th International Conference, pp. 1-6.
[6] Wang, T., 2017, “Social identity dimensions and consumer behavior in social media,” Asia Pacific Management Review, pp. 1-7.
[7] Wen, C. Teng, Z., Chen, J., Wu, Y., Gong, R. and Pu, J., 2016, “Socialradius: visual exploration of user check-in behavior based on social media data,” International Conference on Cooperative Design Visualization and Engineering, pp. 300-308.
[8] Williams, E., Gray, J., Morris, E., Bradshaw, B., Williams, K. and Brandon, D., 2016, “A comparison of two methods for the topical clustering of social mediaposts,” Ubiquitous Computing, Electronics & Mobile Communication Conference, pp. 1-7.