MINING TEXTS TO UNDERSTAND CUSTOMERS' IMAGE OF BRANDS

Hyung Jun Ahn
Hongik University
94, Wausan-ro, Mapo-gu, Seoul, 121-791, Korea
hjahn@hongik.ac.kr

ABSTRACT

Text mining is becoming increasingly important in understanding customers and markets these days. This paper presents a method of mining texts about customer sentiments using a network analysis technique. A data set collected about two global mobile device manufactures were used for testing the method. The analysis results show that the method can be effectively used to extract key sentiments in the customers' texts.

Keywords: Text Mining, Customer Sentiments, Semantic Network Analysis, Social Network Services

1. INTRODUCTION

With the wide acceptance of social network services (SNS), there is increasing amount of text data online that represent customers' perception of brand images. Users share their experiences and sentiments about products and brands online rigorously, which in turn provides a good opportunity to companies for understanding the market in real-time by analyzing the data[1,2].

This study aims to show that the network analysis method can be effectively used to analyze such data about customers' perception of brands[3]. Network analysis method has been chosen for several reasons. First, it can easily identify important key words in the texts by applying various measures of node importance. Second, it can not only extract keywords, but also cluster the keywords at the same time. Third, the method is very effective in visualizing the results of text analysis. Fourth, there are many software packages readily available for network analysis.

In order to show the effectiveness of the approach, this study applies the method to two sets of data collected through a structured interview of participants.

2. METHODOLOGY

This study used the Zaltman Metaphor Elicitation Technique (ZMET) interview technique to interview 171 university students and collect sentences about the participants' perception of target brands. With the ZMET technique, each participant is asked to select a few photos before an interview that best represent their perception about a target brand or a product[4]. By involving photos, participants are given strong visual stimulus which enables them to provide sentences that describe their true perception much better, compared to interviews that solely rely on verbal conversation. There were two target brands in the study, both global manufacturers of mobile devices. Each participant was assigned one of the brands randomly.

The collected sentences were preprocessed so that only appropriate candidate keywords are passed to the network analysis stage[5]. The preprocessing involves: stop-words removal and stemming. Stop words removal filters out those words that are too common or meaningless to be used as keywords (e.g. *a, the, is, to*, etc.). Stemming standardizes keywords so that the words with the same meaning can be indexed identically (e.g. both *get* and *got* are indexed as *get*).

After this, a word network is constructed using the keywords. In the network, each word becomes a node. Words appearing close to each other within a certain distance in the sentences are linked in the network. This study used word distances ranging from 1 to 7 to assign links to the words. The most effective distances, 3 to 4, were used for the final analysis. Then, a network analysis measure was applied to find the most important nodes, or keywords, in the network. Among many measures, this study found the Eigenvector centrality most effective in selecting the keywords, and hence, the results in the following section are all based on the measure.

3. FINDINGS

Figure 1 and 2 show the visualized results of the analysis for the two brands respectively. In the result, node size shows the importance of a keyword; node color represents clusters. Because the analysis was conducted in Korean, English translation was added to each node for readers. Note that only the most important nodes were selected for the simplicity of visualization.
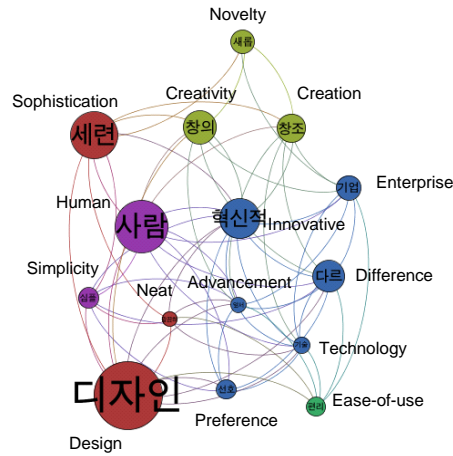
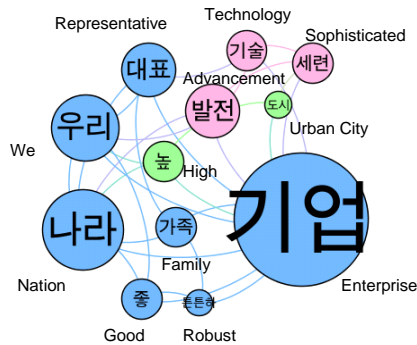**Figure 1.** Analysis of the texts of brand A

Figure 2. Analysis of the texts of brand S

In Figure 1, we can see that there are about five clusters. The three notable clusters can be interpreted to represent *novelty*, *innovative entrepreneurship*, and *design and sophistication*. This is in accordance with the generally accepted view of the brand, which is a very creative enterprise of mobile devices that focus on human-oriented sophisticated design.

Figure 2 shows a different result. In the figure, we can see that the concepts such as *nation, we,* and *representative* outstand, which is clearly different from the results of brand A. This is because brand S is regarded as a national brand and, hence, a national pride in Korea. It also contains clusters that represent sophistication and advancement, similar to brand A. Again, these views are in good congruence with the general view of the brand in Korea.

4. CONCLUSION

This study showed that the network analysis method can be successfully used to extract keywords in texts that describe customers' perception of brands. The advantages of the method include that it can cluster the keywords as well when extracting keywords, and that it is very strong in visualizing the results as shown in the figures. Based on the promising results, more applications of the method to other brands and products are needed as further research to test and refine the method.

5. ACKNOWLEDGEMENT

6. REFERENCES

[1] M. Cooke, and N. Buckley, Web 2.0, social networks and the future of market research. *International Journal of Market Research*, 50(2), p267-292, 2008.

[2] N. Glance, M. Hurst, K. Nigam, M. Siegler, R. Stockton, and T. Tomokiyo, Deriving marketing intelligence from online discussion. *Paper presented at the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, Chicago, IL, USA, August 21-24, 2005. http://dx.doi.org/10.1145/1081870.1081919.

[3] G. Palshikar, A. Ghosh, R. De, and S. Pal, Keyword extraction from a single document using centrality measures. In A. Ghosh, R.K. De, and S.K. Pal (Eds.), *Proceedings of Pattern Recognition and Machine Intelligence* (p503-510). Springer Berlin, Heidelberg: Springer-Verlag, 2007.

[4] G.L. Christensen, and J.C. Olson, Mapping consumers' mental models with ZMET. *Psychology and Marketing*, 19(6), p477-501, 2002. http://dx.doi.org/10.1002/mar.10021.

[5] R. Feldman, and J. Sanger, *The text mining handbook: Advanced approaches in analyzing unstructured data*. New York: Cambridge Univ Pr, 2007. http://dx.doi.org/10.1017/CBO9780511546914.