

CosMovis: Analyzing Semantic Network of Sentiment Words in Movie Reviews

Hyoji Ha*

Gi-nam Kim†

Wonjoo Hwang‡

Hanmin Choi§

Kyungwon Lee¶

*† Life Media Interdisciplinary Program
Ajou University

‡§¶ Department of Digital Media
Ajou University

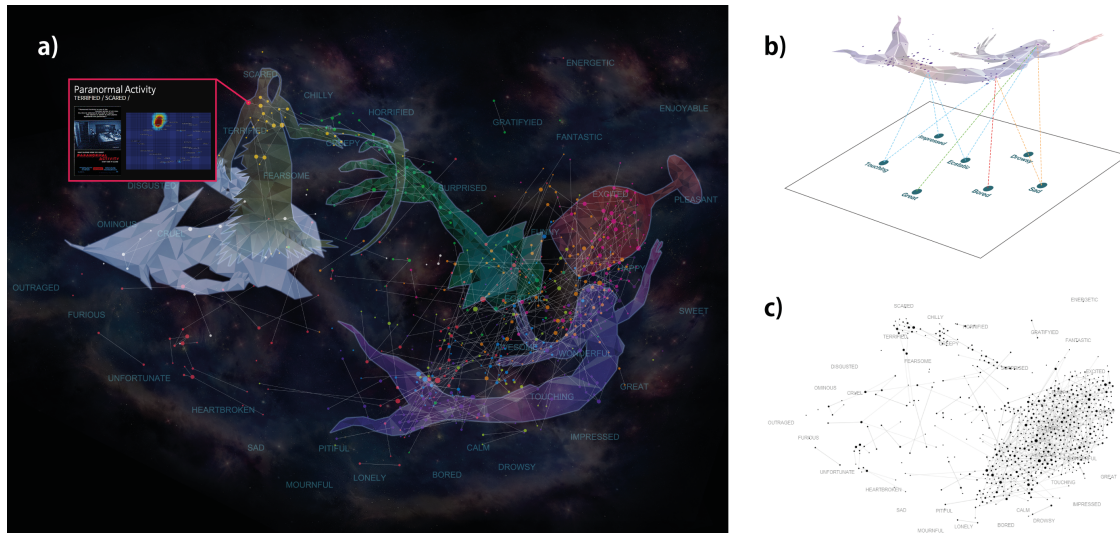


Figure 1: (a) Comprehensive CosMovis constellation map of sentiment word based movies. (b) Basic structure of the Sentiment-Movie Network. (c) The extreme position of node and cluster.

ABSTRACT

In this paper, we present the new method to easily recognize intricate network and cluster by connecting Multidimensional Scaling(MDS) Map and Social Network Graph, and comprehend the feature of each node using Heatmap Visualization. For the sake of this method, this paper used netizen's movie review data. Also, the process is as follow : 1) We calculated the frequency of sentiment word from each movie review. 2) We designed Heatmap Visualization which easily apprehends main emotion of netizens which appear in each movie review. 3) We made location of each node reflect the frequency of sentiment word by designing Sentiment-Movie Network made up of combining MDS Map and Social Network Graph, as well as we embodied Network Graph to make same nodes form cluster. 4) By granting meaning in accordance with characteristic of clustering, we applied asterism graphic to facilitate cognitive interpretation. Our demonstration is available to <http://idlab.ajou.ac.kr/cosmovis/>.

Keywords: Semantic networks, Sentiment word analysis, Review mining.

*e-mail: hjha0508@ajou.ac.kr

†e-mail: gnkim@ajou.ac.kr

‡e-mail: wonjoo93@ajou.ac.kr

§e-mail: hanmini@ajou.ac.kr

¶e-mail: kwlee@ajou.ac.kr

Index Terms: I.3.6 [Computing Methodologies]: Computer Graphics—Methodology and Techniques; H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces;

1 INTRODUCTION

Social Network Analysis(SNA) performs a significant role in understanding and finding solutions of society-functional problems by examining the original structure and relationship of network. Therefore, SNA Visualization is applied in extensive fields including Network Analysis based on data similarity, Network Analysis about social-scientific situation, and Graph Theories. Particularly, Force-directed, typical algorithm which draws network graph, is used to combine relative nodes to create Cluster, so it is highly useful to draw graph for this analysis[1]. However, as the entry value of node location from that graph drawn by Force-directed is random and the extreme position is determined by relative connection between nodes, the location of node changes whenever the graph is redrawn.

To solve this problem, this paper suggests several methods as follows. First, for the purpose of making node distributed in semantic location when designing similarity-based Network Analysis, we analyzed correlation between main sentiment words, needed in network interpretation, using Multi-dimensional Scaling(MDS), and then designed 2-dimensional distribution map. As each node is influenced by Semantic Point depending on attribute value, we made the absolute position of node reflect attribute of node[2]. Also, we made it easy to comprehend the frequency of sentiment word from each movie on by using Heatmap of sentiment word on the basis of distribution map[3]. Second, we applied constellation map on

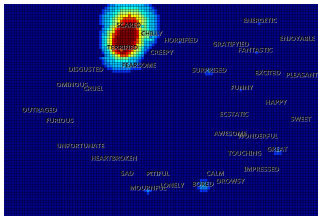


Figure 2: Heatmap Visualization and positioning of node on the Sentiment-Movie Network.(Movie: Paranormal Activity)

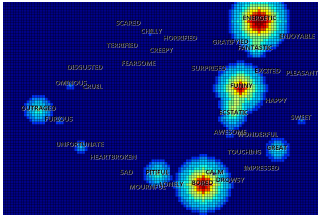


Figure 3: Heatmap Visualization and positioning of node on the Sentiment-Movie Network.(Movie: Breaking Dawn part1)

the basis of node and edge of network community structure for labelling characteristic of each cluster when network-based nodes on the 2-dimensional distribution map formed clustering.

2 VISUALIZATION PROPOSAL

2.1 Heatmap Visualization

Movie review data were collected from the movie information service in NAVER, a web portal site with largest users in Korea. As a result, we obtained 4,107,605 reviews on 2289 movies from 2004 to 2013 by Crawling method. Contents of the reviews were divided into morphemes, and were classified to 68 main sentiment words. Therefore, we subsequently organized sentiment word groups consisting of 1331 words. We calculated each ratio consist of sentiment word cluster from each movie and found out maximum ratio. Furthermore, we selected sentiment word cluster as a less influential cluster which maximum TF-IDF score ratio is less than 0.10 referring to the result of previous process. As a result, 36 sentiment word clusters were chosen and they were divided into Happy, Surprise, Boring, Sad, Anger, Disgust, Fear. Correlation index between each sentiment word were estimated based on these 36 sentiment words, and we designed 2-dimensional distribution map in association with the index.

Heatmap is a visualizing method consisted of rectangular tiling shaded in color scale[5]. It is used to search pattern or anomaly of data metrics. In this research, we showed the TF-IDF score of sentiment word into Heatmap utilizing coordinate space in 2-dimensional distribution map of each sentiment word for visualizing sentimental distribution graphic of each movie node consisting of network. Using Heatmap made it possible to easily compare movie nodes which have sentiment words contrasted one another or similar.

2.2 Sentiment-Movie Network

Our proposed graph is similar to Artefact Actor Network(AAN) which is a type of Multi-Layered Social Network. AAN connects between Artefact Network and Social Network using semantic connection, so it expresses semantic relation between two networks[4]. In our proposed graph, we connected sentiment words on 2-Dimensional Scaling Map with Movie Network(Figure 1(b)) and we called this network Sentiment-Movie Network.

Figure 2 and figure 3 show the example that location of node on

graph can be altered depending on the frequency of sentiment word indicated in Heatmap Visualization. As every node connected by network made up of suggested methods is located in the graph, clustering is formed by combined similar movies in the space of sentiment word with high frequency considering connection between movies and between related sentiment words(Figure 1(c)).

2.3 Constellation Visualization

This chapter facilitates a cognitive understanding of the process to design constellation image visualization, based upon specific nodes and edges with significant sentiment word frequency to clarify the semantic parts of each clustering.

We created a constellation graphic of each cluster network, considering the significant sentiment words, information on movies and synopses in each cluster. In order to realize constellation images, we referred to the labeling data of 11 different clusters yielded from k-means clustering, most significant categories of sentiment words in each cluster and their following information on movies and synopses. A comprehensive network map is shown in Figure 1(a).

3 DISCUSSION

Pilot research is designed on target to 20 participants to prove how much network nodes, suggested in this lecture, are well located based upon 2-dimensional distribution map of sentiment word. This research is headed from selecting one movie node in network and showing related heatmap to selecting which is a heatmap of node on neighboring or reversing place of first-selected node. As the percentage of correct answers from questions about basic location structure of network was about 93%, it was measured that most participants comprehended overall network structure well.

4 CONCLUSION

This paper suggested Heatmap Visualization useful to comprehend the characteristic of each node, expression which shows one node on 2-dimensional graph according to Heatmap and constellation visualization for the purpose of semantic interpretation about clustering.

Proposed method in this research can be applied to alteration of target movie or number of review data, as well as applied as a recommendation system in that it compose clustering by combining movies of similar characteristic with analyzing sentiment word from movie review data. This research method is expected to be applied in various fields using asterism graphics as it is applied in movie data and asterism graphics are used to analyze data as a metaphor.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012S1A5A2A01020132).

REFERENCES

- [1] T. M. J. FRUCHTERMAN and E. M. REINGOLD. Graph drawing by force-directed placement. *Software-practice and experience*, 21:1129–1164, 1991.
- [2] H. Ha, G. Kim, and K. Lee. A study on analysis of affective words in movie reviews and the situation of watching movies. *Society of Design Convergence*, 43:17–32, 2013.
- [3] D. Kim, H. Park, J. Sim, S. Park, and K. Lee. Visualization of movie recommender system using distribution maps. *IEEE Pacific Visualization Symposium 2012*, 2012.
- [4] W. Reinhardt, M. Moi, and T. Varlemann. Artefact-actor-networks as tie between social networks and artefact networks. *Collaborative Computing: Networking, Applications and Worksharing*, 2009.
- [5] L. Wilkinson and F. M. The history of the cluster heat map. *The American Statistician*, 63:2, 2009.