VISUAL TREND ANALYSIS METHOD FOR ONTOLOGY BASED OPINION MINING ON MOVIE REVIEWS

Seongmin Mun*, Ginam Kim**, Gyeongcheol Choi* and Kyungwon Lee***
Life Media Interdisciplinary Program, Naver corp**, Department of Digital Media ****
Ajou University, 206, World cup-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 443-749, KOREA

ABSTRACT

The rapid development of the Web 2.0 era has generated a huge amount of online word-of-mouth information, which has influenced the society in various fields. Effects of Word of Mouth on movies, in particular, has become the new standards to evaluate a movie that affects potential audiences' decision. Therefore, this research proposes a methodology to analyze the responses of movie audiences in two ways; overall regardless of time flow and trending which reflects the changes of audience opinions over time. Review analysis was conducted through ontology followed by opinion mining, further visualized via polar chart network, clustering and timeline. Visualization designed in this process is presented in "http://54.255.190.140/index/v0".

KEYWORDS

Clustering, Data Visualization, Network, Ontology, Opinion, Polarity.

1. INTRODUCTION

As film industry is growing every year with the advance of media technology, online movie reviews are getting becoming predominant on potential audience behaviors to decide whether to be a consumer. We thus designed the ontology based upon such audience reviews in order to propose a methodology to analyze movie reviews in two ways; overall reviews regardless of time flow, and trending reviews reflecting how the opinions have changed over time. To demonstrate, we first constructed ontology to reflect the contents of individual reviews. Second, we conducted opinion mining based on the ontology. We next visualized the results and analyzed such results, presenting a scenario through the methodology of this research to analyze the changes of opinions in a movie review.

2. ONTOLOGY CONSTRUCTION

In order to construct an ontology for opinion mining, we first selected the categories of ontology and its evaluation keywords associated with each category. After extracting the index terms with high frequency in review data, we screened them as potential keywords and classified the finally selected terms. We further verified such process with topic modeling. Keywords utilized in our ontology in consequence of this progress are presented in Table 1.

Table 1. Final sets of keywords applied in our ontology

Element class	Feature words
Director	Director, Direction, Composition, Editing, DirectionSkill, Attraction
Movie	Movie, Work, Overall, Scale
Actor	Performance, Actor, Character, MainCharacter, PerformanceSkill, Lines, SupportingActor, Casting, Voice, Style, Person, Appearance, Dubbing, Actress, Performer, Role, CastMember
Vision	Vision, Sequence, Scene, Spectacle, Atmosphere, Expression, Screen, Visual, Background, Graphic, MakeUp, Visuality, ImageBeauty

Story	Story, Content, Material, Telling, Tale, Closing, Ending, Scenario, Probability, Synopsis, Setting, OriginalWork, Finish, Subject, Accident, Structure, ContentTelling, StoryLine, Reading, StoryTelling, CastMember, StoryComposition
Sound	Sound, Song, Volume, Soundtrack, Melody, Music

3. VISUALIZATION

As well as providing the positive/negative tendency of each movie review, our new visualization method enables users to compare the evaluations of different movies since it forms clusters between movies with similar reviews, by comparing most heavily evaluated elements of each movie. After positive/negative tendency information visualization was performed, we further visualized clustering data in order to form the clusters of movies with similar reviews. We also designed an interface to facilitate more various visualization analyses through interaction with users.

3.1 Individual Movie Analysis

Mining results of individual movies in this visualization are presented as one set of a pie chart and a bar graph. A pie chart, containing six directions of 12, 2, 4, 6, 8 and 10 o' clock, indicates the frequency and polarity of each evaluation element. Size of a sector would increase when the elements are considered important as evaluation elements depending on its frequency, and vice versa.

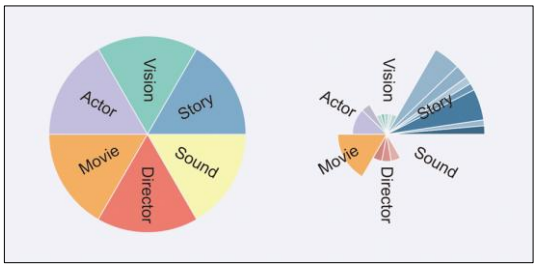

Figure 1. Opinion Mining Visualization for the each movie

3.2 Analysis on Node Position

Our finding that the location of nodes is determined according to the weight of evaluation elements when visualizing the clusters in Opinion Analyzer also implies that the location of nodes provides information on the weight of each element suggesting its importance. Figure 2 indicates the weight of evaluation elements of New World, Secretly Greatly and Tower. It is revealed that while the Actor/Actress was of similar importance, Plots have appeared more frequently suggesting its importance, as the nodes are located to the right.

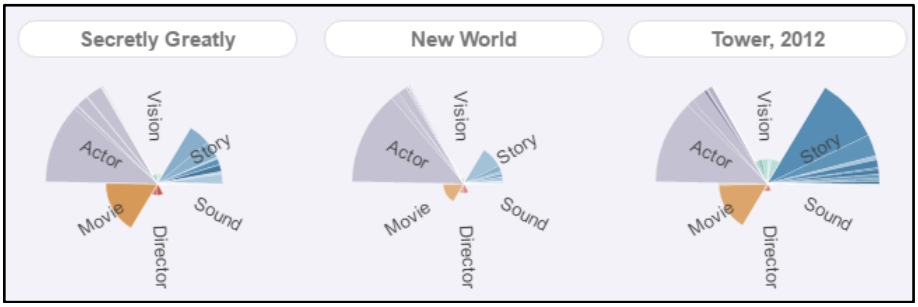

Figure 2. Evaluation elements of *Secretly Greatly*, *New World* and *Tower*

3.3 Analysis on Clustered Movie Groups

Since it is general to analyze the relationships between the pre-clustered data groups if the amount of data is too large, sample movies of this research were divided into three groups.

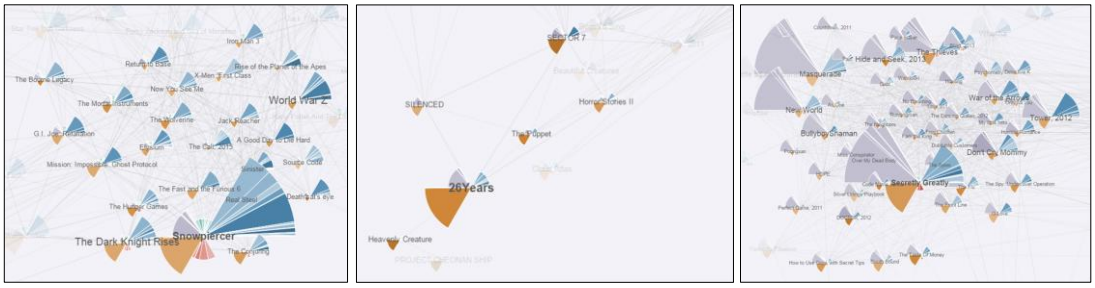

Figure 3. Overall Visualization of the 130 movies in this research, indicating three different clusters

3.4 Network Analysis between Movies

Network Analysis represents a method to signify similar nodes, determining which node shares similar data with other nodes located in different clusters. For instance, movies located at the center of cluster groups such as Meet the In-Laws in Figure 4, is linked to other movies sharing similar attributes. All the movies of which links are connected belong to different groups, suggesting that the network contributes to discovering new clusters.

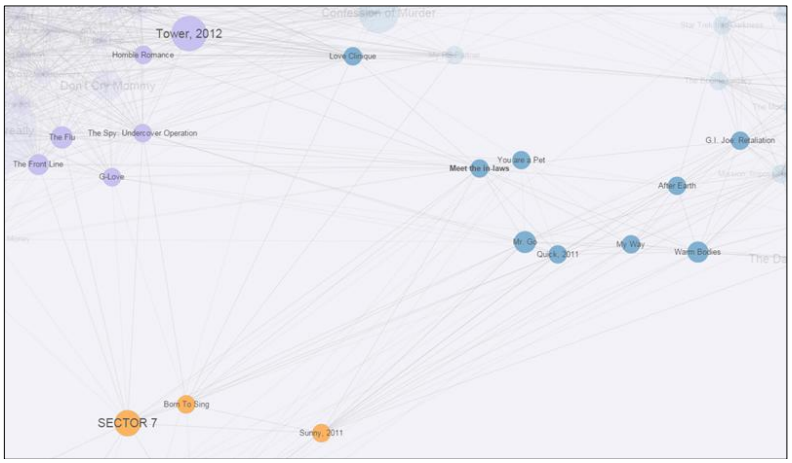

Figure 4. Network of movies relevant to *Meet the In-Laws*

4. TREND ANALYSIS FOR THE MOVIE REVIEWS

In addition to analysis and visualization of the overall reviews of a movie, this methodology can reflect the time flow as suggested in Figure 5 which is a pilot data visualization of a movie Snowpiercer. In 2013 when the movie was first released, its plot was a major evaluation element, positive and negative reviews accounting for approximately equally. However it is found that Actors/Actresses gained more popularity in the movie review as time went on, with high positivity. It is also noticed that locations of nodes thereby changed according to the importance weight of evaluation elements

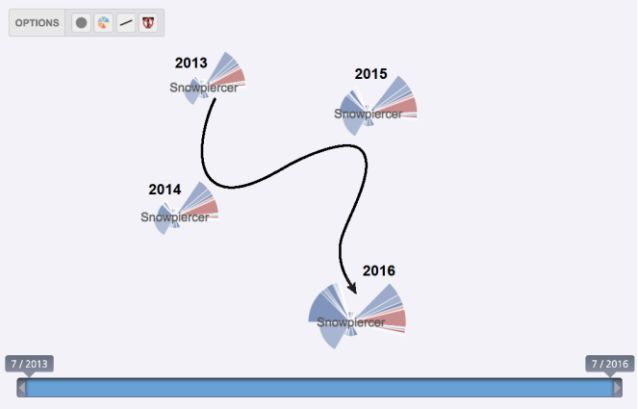

Figure 5. Changes of reviews on *Snowpiercer* over time

5. CONCLUSION

This study aimed to propose a methodology to analyze and visualize the movie reviews in two ways; overall and trending. We thus constructed an ontology of movie reviews and suggested a visualization analysis methodology in order to analyze the movie reviews. Major conclusions of this research are as follows. First, visualization analysis on each movie revealed that overall tendency of reviews of each movie differ from each other. Second, 130 films as samples in this research were clustered as three groups depending largely on the feedback from the audience. Third, it is suggested that this visualization analysis methodology can be applied to review trend analysis that reflects the changes of audiences' opinion over time.

In addition to these findings, this research also proposes a new framework of ontology on movie domain applying keyword extraction and topic modeling, and that argument structure identification was applied in opinion miming.

REFERENCES

Journal

1. Fruchterman, T. M. and Reingold, E. M, 1991. Graph drawing by force-directed placement. Software: Practice and Experience, Vol. 21, No. 11, pp. 1129-1164.
2. Gruber T. R, 1993. A translation approach to portable ontology specifications. Knowledge acquisition, Vol. 5, No. 2, pp. 199-220.
3. Jaeseok Myeong et al, 2008. A Korean Product Review Analysis System Using a Semi-Automatically Constructed Semantic Dictionary. Journal of KIISE, Vol. 35, No. 6, pp. 392-402.
4. Yunjeong Lee et al, 2009. Analysis and Visualization for Comment Messages of Internet Posts. The Journal of the Korea Contents Association, Vol. 9, No. 7, pp. 45-56.
5. Moon Yul Huh and Woon Ock Cha, 2008. Information Theory and Data Visualization Approach to Poll Analysis. The Korean journal of applied statistics, Vol. 20, No. 1, pp. 61-78.

Conference paper or contributed volume

1. Anaïs Cadilhac et al, 2010. Ontolexical resources for feature based opinion mining : a case-study, Proceedings of the 6th Workshop on Ontologies and Lexical Resources(Ontolex 2010). Beijing, China, pp. 77-86.
2. Bjørkelund, E. et al, 2012. A study of opinion mining and visualization of hotel reviews. In Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services. Bali, Indonesia, pp. 229-238.
3. Borth, D. et al, 2013. Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content. Proceedings of the 21st ACM international conference on Multimedia. Barcelona, Spain, pp. 451-460.