Method for Construction of Spatial Sentiment Lexicon using Place Reviews: Case Study on Theme Parks

Youngmin Lee*, Mikyeong Moon**, Inyoung Chae*, Kiyun Yu*, Jiyoung Kim*

* Seoul National University; daldanka@snu.ac.kr; ciy5164@snu.ac.kr; kiyun@snu.ac.kr; soodaq@snu.ac.kr
 ** National Information Society Agency, mkmoon@nia.or.kr

Extended Abstract

With the development of GPS (global positioning system) technology and smart devices, the use of location-based services is constantly increasing. As the use of this technology increases, reviewing and grading places through the use of location-based services has also become a common practice among users. Previous reviews of a place can significantly affect their potential visitors. A user-created review is the result of a visitor's actual positive or negative sentiment expression, and the sentiment could be expressed as a positive, negative, or neutral opinion.

In order to perform sentiment analysis, each word should be separated by its POS (part-of-speech) through natural language processing. To do this, a database of features for place and spatial sentiment words should be constructed. In the past, sentiment analysis has been used mainly for product reviews (Chang 2009, Hu & Liu 2004, Myung et al 2008, Scaffidi et al). No spatial sentiment lexicon for sentiment analysis of places has been constructed yet. Specifically, Hangeul lacks in terms of research on sentiment word analysis, compared to English. Hangeul, unlike English, has complicated characteristics and is composed of complex adjectives and suffixes (Jang et al. 2015).

Therefore, in this study, we propose a method to construct a spatial sentiment lexicon using place reviews written in Hangeul. In this study, we focused primarily on a "theme parks' out of the many possible place categories. Although the general polarity (positive, negative, or neutral) of senti-

ment for a place is important, analyzing the polarity for a specific place property, according to certain place categories is also valuable. We construct a spatial sentiment lexicon in accordance with each category of a place because the properties of a place and the language around the place changes according to the characteristics of the place. Accordingly, in this study, we preferentially constructed a spatial sentiment lexicon for a theme park. Other types of spatial sentiment lexicon could be constructed using the same methodology.

Generally, sentiment words written in Hangeul are represented by adjectives such as 'good', 'nice', and 'beautiful'. However, in this research, we included verbs such as 'crowded', 'take a rest', and 'take a walk' in order to consider other types of words describing the places. In addition, nouns representing a sentiment for a place such as 'recommendation', 'satisfaction', and 'disappointment' were also included as candidates for spatial sentiment words.

The following three aspects were considered in the construction of a spatial sentiment lexicon. First, the polarity of the sentiment word is analyzed. To do this, we calculated sentiment polarity and probability using the results of our survey. Second, it should be taken into account that some sentiment words are associated with properties of a place. For example, 'many' or 'little' could be positive or negative sentiments depending on the place. For example, 'green' could have a positive implication when paired with 'many', but 'people' could have negative connotation with the same word. In this case, the survey was carried out using a combination of spatial features and predicate configuration. Lastly, the spatial features and predicate characteristics of a place should be classified. We defined the combination of the spatial feature and the predicate as the 'spatial feature sentence'. By considering these characteristics, we designed the spatial sentiment lexicon to manage the meaning of sentiment words separately, according to the place categories.

The work flow used in this study is shown in *Figure 1*. The list of POIs (points-of-interest) was collected through Naver map search. Then, the place reviews for the POIs were collected through Google maps API. The analysis of place reviews were collected from the Google map API as follows. First, morpheme analysis through natural language processing for the reviews was implemented. Second, the morpheme was converted to a POS-tagging word; at this time, noun, adjective, and verbs were extracted as the candidates for spatial sentiment words. Third, the polarity and probability of the sentiment words were calculated by the survey conducted using the candidate of spatial sentiment words. Finally, the combination spatial fea-

ture sentence was created and the probability of sentiment is calculated using the result of the survey.
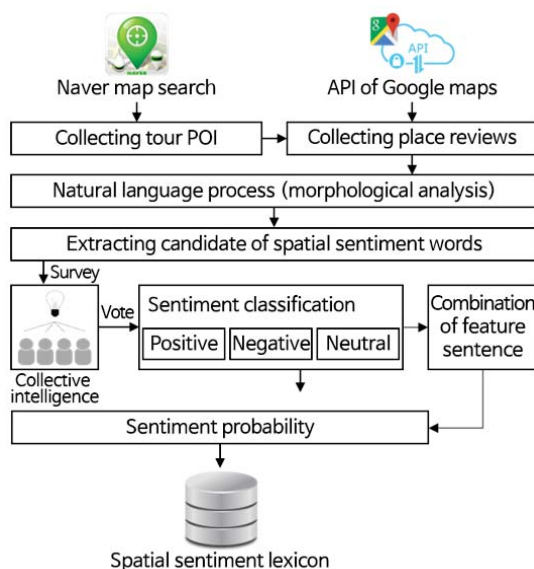
Figure 1. Research workflow.

For the experiment, we collected 118 search results of places within 'Seoul theme park' from the Naver map. We then collected 80 location IDs and matched them with 118 location names using the Google map API (application programming interface). Then, we collected 80 place reviews corresponding to the ID using details from the Google places API. As a result, we collected 204 reviews for 63 locations. We did not collect the remaining 17 reviews because they did not have reviews or there was no return value through the API.

We conducted a survey using the words extracted through morpheme analysis for location reviews collected through Google maps. The survey consisted of 150 questions. Possible responses were positive, negative, neutral, or inappropriate for words describing a place. We received more than 50 responses for the 150 questions. As a result, we calculated whether each word had a positive (+1), negative (-1), or neutral (0) polarity. The probability for sentiment of 16 nouns, 73 adjectives, 25 verbs, and 24 spatial feature sentences was found. The 12 words that were determined as inappropriate descriptors for representing a location were excluded from the spatial sen-

timent lexicon. Moreover, the spatial sentences were composed using the calculated results from the survey (*Table 1*).

Spatial sentiment words	POS	Sentiment polarity	Probability
'Recommendation'	Noun	Positive (+1)	1.000
'Inappropriate'	Noun	Negative (-1)	1.000
'Quiet'	Adjective	Positive (+1)	0.558
'Walk'	Verb	Neutral (0)	0.660
'People', 'Numerous'	Noun, Adjective	Neutral (0)	0.764
'Exercise', 'good'	Verb, Adjective	Positive (+1)	1.000
…	…	…	…

Table 1. Part of the spatial sentiment lexicon.

The spatial sentiment lexicon could be utilized as a reference when performing sentiment analysis on the contents of various social media platforms, and could offer useful information to those who want to visit a place. In future work, we will study a method used to extend the lexicon by adding synonyms for pre-constructed sentiment words, as well as a methodology to analyze syntax more precisely through a combination configuration of spatial feature sentences.

Acknowledgement

References

Chang J (2009) A sentiment Analysis Algorithm for Automatic Product Reviews Classification in On-Line Shopping Mall. 14(4):19-33

Hu M, & Liu B (2004) Mining and Summarizing Customer Reviews. In proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 168-177

Myung J, Lee D, & Lee S (2008) A Korean Product Review Analysis System Using a Semi-Automatically Constructed Semantic Dictionary. Journal of KIISE:Software and Applications. 35(6): 392-403

Jang K, Park S, & Kim W (2015) Automatic Construction of a Negative/positive Corpus and Emotional Classification using the Internet Emotional Sign. Journal of KIISE. 42(4):512-521

Scaffidi C, Bierhoff K, Chang E, Felker M, Ng H, & Jin C (2007) Red Opal: Product-Feature Scoring from Reviews. In proceedings of the 8th ACM conference on Electronic commerce. 182-191