Comparative Search by Using Technologies to Extract Superior Targets from Comparison Sentences

Yeong Hyeon Gu[1], Seong Joon Yoo[1]

[1]School of Computer Engineering, Sejong University,
Gunja, Gwangjin, Seoul, Korea
sjyoo@sejong.ac.kr

Abstract. Studies on extracting technology of structure information from comparison sentences are to extract comparison target, features, and relations from given sentences and it is a very important technology in comparing search. There have been studies in English speaking countries, but this is the first study on Korean documents. In this study[1], the comparison targets and relations were focused among the structure information of comparison sentences to find the superiority and inferiority determination rules. Then, the comparison keywords for superiority and inferiority as well as the predicate were used to extract superior targets among the comparison targets. Then, the comparative search results and comparative opinions are provided in a single view through the comparative search system.

Keywords: comparative sentence, comparative opinion mining, comparative relationship, online opinion mining, comparative search

1 Introduction

Studies on Comparative Sentence Mining, which shows comprehensive comparison results by analyzing review documents that contain user opinions, can be applied in various fields. For companies, they can find what to improve by analyzing the users' opinions on products and services. For individuals, being able to see huge amount of systematically organized and analyzed reviews on webs rather than a small amount of reviews gathered by a portal site will help them to make decision of purchasing and this can be utilized as the opinion mining factor technologies [1, 2, 3, 4, 5, 6].

There are many studies in comparative sentence classification methods in English speaking countries, and recently there are studies applied to Korean. However, there have no studies on comparison targets and superior targets which are very useful information in comparative search in Korean so more studies on broad field and more researchers are required.

The methods to find targets that are in superior position than the comparing target

[1] This work was supported by the Korea Research Foundation(KRF) grant funded by the Korea government(MEST) (No. 2010-0015842)

from different levels of comparison sentences by using rules are described in this study.

Chapter 2 describes the related works and Chapter 3 describes the method by which superior targets were extracted from comparative sentences using the Comparative Sentence Method to show the comparative search result. Finally chapter 4 describes the results acquired from experiments.

2 Related Works

Studies classifying sentences including comparative opinions using some keywords in English speaking world include [7, 8]. In this study, 83 keywords were identified and used in those sentences. Likewise, the classification of the comparative opinion sentences using keywords showed a high recall rate (98%), however, the precision was very low (32%). In order to solve this problem, Class Sequential Rule mining was used.

In [9] the structural factors of comparative sentences were analyzed and the comparison targets and attributes were extracted to extract the superior targets in the comparison. For this, the comparative sentences were classified into 4 types and appropriate rules for each type were applied to extract the superior targets in the comparison.

In [10], the structure factors of the Chinese comparative sentences were classified such as comparison target, comparison attributes and emotions.

The extracting methods for relation items from comparative sentences are highly used in English and Chinese documents, however, there have been no studies targeting Korean documents.

3 Extracting Superior Targets and Comparison Search

Comparative sentences have certain structures. Those are comparing subject, comparing target, comparing features, and comparing superiority. For example, a sentence “Restaurant A is more delicious than restaurant B” was examined. Restaurant A is a comparing subject and restaurant B is a comparing target. Although it was not specifically described here, the flavor of food is a comparing feature and A restaurant is actually superior target than restaurant B.

Korean comparison sentences are classified into equality and inequality comparison upon the first judgment for the similarity and difference of two targets. The inequality comparison is divided into simple inequality and degree inequality comparison upon its degree.

In the superiority/inferiority sentences, there are words frequently used to show the comparisons and these words are called superiority/inferiority comparison words. In this study, the superiority/inferiority comparison words were used to determine the superior targets and the study results on the methods of showing the comparative search results on the web was summarized.

In order to collect the comparison sentences, the web crawler is used to collect comments from restaurant information provision websites. In this model sentences are retrieved from the document and then those sentences containing any inequity comparative words are extracted. The extracted sentences are classified into comparative sentences or non-comparative ones by the Rule based comparative sentence analyzer, with comparative ones added to the database. Non-comparative ones go through another classification procedure provided by the analyzer, with comparative ones added to the database.

In order to classify the superior targets from the classified comparative sentences, the verb information that is closely located to the comparison word is used in a that sentence used the superiority/inferiority comparison proposition such as 'than', and 'compare to', if the closest verb from a comparison word has the positive meaning, the comparison subject becomes the superior target and if it has the negative meaning, the comparison target becomes the superior target. Oppositely, if the comparison word is used for adverb such as 'more' or 'far' and if the predicate is positive, the comparison target becomes the superior target and if it is negative, the comparison subject becomes the superior target. At this time, for the determination for positive and negative, the positive and negative verb list stored in DB needs to be used. This is summarized in Table 1.

Table 1. Rules to determine superior target from declarative sentences

Category		Details
Sentences using superiority /inferiority comparison words as proposition	POS	Comparison target + comparison words as proposition('보다(than)', '비해(compare to)' etc) + adjective
	Rule	If the predicate is positive, → The comparison subject becomes the superior target. If the predicate is negative, → The comparison target becomes the superior target.
Sentences using superiority /inferiority comparison words as adverb	POS	Comparison target + Proposition + comparison words as adverb('더(more)', '훨씬(far)' etc) + Adjective
	Rule	If the predicate is positive, → The comparison target becomes the superior target. If the predicate is negative, → The comparison subject becomes the superior target.

For general declarative sentences, superior target could be found by using comparison words and positivity and negativity of predicates, however, for negative sentences with negative words, the superior target needs to be determined oppositely from the previous rule. For example "A is not better than B", the proposition 'than(boda)' was used as a comparison word and since there is a positive predicate 'better(jota)', the comparison subject A is supposed to be superior according to the

rule, however, actually comparison target B is the superior target. Likewise, if a negative word ‘not’ is followed after a predicate, the superior target needs to be determined oppositely from the rule for general declarative sentence.

In this study, the technologies to extract comparative opinion sentences and superior targets were used to implement the comparison search system. The number of comparative opinions on certain targets can be found and several comparative opinions can be seen at the same time. Moreover, even if not all multiple comparative sentences are being read, the superior targets can be determined since it shows the superior targets.

4 Experiment

For experiment, web crawler was used for restaurant review websites Menupan.com and Wingbus to retrieve users’ comments. 299 Comparison sentences included the word ‘than (boda)” and superior comparison targets were found in 275 sentences which showed 91.9% of precision. 122 Comparison sentences included the word ‘compared to (bihae)” and superior comparison targets were found in 100 sentences which showed 81.9% of precision. 30 Sentences included the comparison word ‘far(Hwelsin) and all of them found the superior comparison target. Also 117 Sentences included the comparison word ‘more(deo)’ but 4 sentences couldn’t find the superior comparison target which showed 96.5% precision.

5 Conclusion

In this study, the methods to extract comparison superior targets in superiority/inferiority comparison sentences by using comparison words and sentence pattern rules are studied. In order to find the comparison subjects and targets, restaurant lists are compiled and the comments that may include comparison sentences were found by using the restaurant list, and then classified the comparison sentences through the method that integrated and applied the sentence structure rules and statistic methods. Then, the superiority/inferiority comparison keywords and predicates were used to extract the superior targets.

In the future, more comparison words need to be used for experiment, also various rules derived from general rules have be to found also the extraction methods for superior target information from adnominal phrases and complex sentences need to be studied. Moreover, the methods to process sentences that can’t determine the superior target from will be studied.

References

1. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment Classification using Machine Learning Techniques. In: the Conference on Empirical Methods in Natural Language Processing, pp.79--86 (2002)
2. Dave, K., Lawrence, S., Pennock, D.: Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In: 12th Intl. World Wide Web Conference, pp. 512--528 (2003)
3. Liu, B.: Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer (2007)
4. Peter, D.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: 40th Annual Meeting of the Association for Computational Linguistics, pp.417--424 (2002)
5. Hwang, J., Ko, Y.: A Korean Sentence and Document Sentiment Classification System Using Sentiment Features. Journal of Korean Institute of Information Scientists and Engineers (KIISE): Computing Practices and Letters, Vol. 14, No. 3, pp. 336-340, (2008)
6. Kim, Y., Jung, Y., Myaeng, S.: An Opinion Analysis System Using Domain-Specific Lexical Knowledge. In: 4th Asia Information Retrieval Symposium, AIRS 2008, LNCS 4993, pp.466--471 (2008)
7. Jindal, N., Liu, B.: Mining Comparative Sentences and Relations. In: AAAI'06, pp.1331--1336 (2006)
8. Jindal, N., Liu, B.: Identifying Comparative Sentences in Text Document. In: 29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval, pp. 244--251 (2006)
9. Liu, B., Ganapathibhotla, M.: Mining Opinion in Comparative Sentences. In: 22nd International Conference on Computational Linguistics, pp.241--248, Manchester (2008)
10. Huang, X., Wan, X., Yang, J., Xiao, J.: Learning to Identify Comparative Sentences in Chinese Text. In: 10th Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence, Vol.5351, pp.187 -- 198 (2008)