

장애인의 비유적 표현을 위한 오피니언 마이닝 시스템

김창기[○], 서정민^{*}

[○]한국교통대학교 사회복지학과

^{*}(주)디커뮤니케이션즈 기술연구소

e-mail: cgkim@ut.ac.kr[○], jmseo@web-d.co.kr^{*}

An Opinion Mining System for A Figurative Representation of Disabilities

Chgan Gi Kim[○], Jeong Min Seo^{*}

[○]Dept. of Social Welfare, Korea Univ. of Transportation

^{*}Research Center of DCommunications Co., Ltd.

● Abstract ●

사회복지 영역의 확대로 복지서비스 수혜자들의 사례관리가 매우 중요한 영역으로 자리매김하고 있다. 이는 사례관리를 이용하여 새로운 서비스를 발굴하고, 실행결과를 평가하여 중요한 패턴을 추출 후 다른 유사 대상자들에게 적용하는 것이 실패를 줄이는 방법이기 때문이다. 그러나 현재 대부분의 사례관리시스템은 서비스를 입력하여 저장/관리하는 측면만을 제공하여 체계적인 분석이 안되고 있다. 이에 본 논문에서는 사례자들의 상담 및 서비스 결과에 관한 오피니언을 분석하여 마음속에 내포하고 있는 사례(비유적 표현)에 관한 실제적인 평가와 오피니언을 추출하는 시스템을 제안한다. 제안하는 시스템을 실험하기 위해 자기의 오피니언을 외부로 노출하기 꺼려하는 장애인을 대상으로 한 상담 사례를 이용하여 실험하였다.

키워드: 오피니언 마이닝(Opinion Mining), 장애인(Disabilities), 비유(Figurative), 상담(Counsel)

I. Introduction

오피니언 마이닝(Opinion Mining)은 텍스트 마이닝의 한 분야로 텍스트 내에서 오피니언을 수집하여 수집한 대상에 대한 오피니언을 추출하는 기술이다. 오피니언 마이닝은 일반적으로 특징 추출과 오피니언 분석의 과정을 통해 이루어진다. 오피니언 마이닝의 대상이 되는 글들은 대부분이 사실들과 오피니언들이 혼합되어 있는 형태이다. 특징 추출 과정에서는 사실과 구분지어 오피니언을 추출하고, 오피니언 분석 과정에서는 추출된 오피니언이 긍정인가 부정인가의 방향성(Orientation)을 판별한다. 그러나 기존의 오피니언 마이닝 시스템들의 특징 추출은 오피니언이 분명하게 드러나 있는 명확한 문장들만을 추출하였다[1]. 그러나 비유적인 문장과 같은 함축적인 문장에도 많은 오피니언이 존재한다. 비유적인 문장에서 오피니언을 추출한다면, 기존 오피니언 마이닝 시스템의 전체적인 분석 성능을 높이는 효과가 있다.

II. Extraction of Figurative Words

비유적 표현의 오피니언 마이닝을 하기 위해서는 두 가지 과정이 필요하다. 첫째는 텍스트 안에서 비유적인 표현이 쓰인 문장을 찾아내는 과정이며, 둘째는 찾아낸 비유적 문장을 분석하여 오피니언을 추출하는 과정이다. 첫 번째 과정은 비유 단어 사전에 태깅된 텍스트에서 비유적인 문장을 구성하는 형태소들의 패턴을 이용하여 비유적 문장을 구별해낸다. 두 번째 과정은 찾아낸 비유적 문장들을 오피니언 단어를 포함한 문장과 그렇지 않은 문장으로 분류하여, 오피니언 단어를 포함한 문장은 오피니언 단어를 포함하지 않는 또 다른 비유적 문장에서 동일한 비유한 표현이 등장하였을 때 그 문장의 참조 대상이 되며, 참조 대상이 없었던 오피니언 단어를 포함하지 않는 문장은 비유한 표현을 자체 구성한 어휘 데이터베이스에서 검색을 하여 오피니언 단어를 찾는다[2, 3].

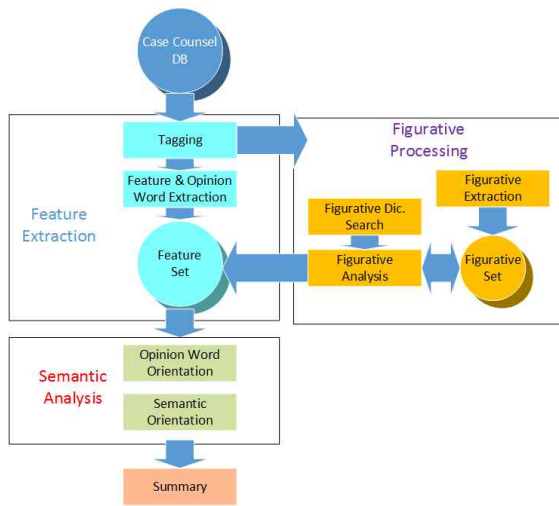


Fig. 1. Opinion Mining System Architecture

III. Implementation & Experiment

본 논문에서 제안한 시스템은 Java와 JSP를 사용하였고, DB는 MySQL 5.6을 사용하였다. 형태소 분석을 위한 시스템은 KLT(Korean Language Technology)[4]를 사용하였다. 입력된 상담 사례 자료는 HWP로 되어 있는 파일들을 PDF로 변환하여 입력하여 분석하였다. 그림 1은 실 예로 든 사회복지 서비스에 대한 평가를 위한 218명의 상담 사례 분석 결과 기록 파일을 제안하는 시스템에 넣은 결과를 그래프 형식으로 보여준 예이다.



Fig. 1. Analysis Result Example

그림에서 보면 매우 강한 긍정이 78, 긍정이 75, 중립이 13으로 나왔으나 중립은 제안하는 시스템에서는 사회복지사들의 전문적 의견을 따라 부정으로 포함시켰다.

Table 1. Classification Accuracy

	긍정	부정	총계
자료수	180	38	218
정확분류	153	30	183
오류분류	27	8	35
정확도	85%	79%	84%

그러나 실제 상담 사례를 기록한 내용을 보면 기록자인 사회복지사들의 주관적인 내용이 많은 부분을 차지하고 있어서 객관적인 기록을 할 수 있는 제도적 장치가 필요할 것으로 사료된다.

IV. Conclusions

본 논문에서는 오피니언 마이닝 기술을 이용하여 상담 사례자의 주관적 의견을 자동으로 분류할 수 있는 감성분석 알고리즘을 제안하였다. 제안된 알고리즘은 한글의 특성을 고려하여 사례 기록파일을 분석하여 속성과 감성어휘 등을 추출하고, 감성의 극성정도를 판단하는 기능을 한다. 제안한 시스템은 장애인 상담 사례 파일을 대상으로 사용자의 의견을 자동으로 분석 요약하여 그 결과를 제공하는 기능을 한다. 제안된 시스템의 성능을 평가하기 위해서 실제 상담사례를 대상으로 실험을 실시한 실험결과 평균적으로 84%의 정확도를 보였다.

References

- [1] X. Ding, B. Liu, and P.S. Yu., "A Holistic Lexicon-Based Approach to Opinion Mining," WSDM'08, 2008.
- [2] Narayanan, R., Liu, B., and Choudhary, A., "Sentiment Analysis of Conditional Sentences," Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-09), August 6 ~ 7, 2009, Singapore.
- [3] Liu, B., "Sentiment Analysis and Subjectivity," Invited Chapter for the Handbook of Natural Language Processing, Second Edition, To appear in Oct/Nov, 2009.
- [4] KLT, <http://cafe.daum.net/nlpk>