오피니언 마이닝을 이용한 음식점 리뷰 분석과 요약

김상욱*, 김원영*, 김응모*
*성균관대학교 정보통신공학부
e-mail : ghost0806@skku.edu

Restaurant Review Analysis and Summary using Opinion Mining Techniques

Sang-wook Kim*, Won-young Kim*, Ung-mo Kim*
*School of Information and Communication Engineering, Sungkyunkwan University

요 약

사용자의 참여를 강조하는 Web2.0 시대를 맞이하여 개인의 블로그나 까페에 올라오는 무수히 많은 리뷰들이 실제 소비자의 마음을 움직이는 데에 많은 영향을 미치고 있다. 하지만 많은 리뷰들이 상당히 길게 작성되어 있기 때문에 원하는 정보만을 찾아내는 것은 어려운 일이다. 본 논문에서는 다양한 종류의 리뷰들 중에서도 많은 부분을 차지하고 있는 음식점에 관한 리뷰들을 분석하여 사용자가 원하는 정보를 요약하여 제공하는 방법을 제안한다. 이러한 방법을 통해서 사용자는 객관적인 판단을 내릴 수 있고, 시간적인 측면에서의 효율성을 획득할 수 있을 것이다.

1. 서론

사용자의 참여를 강조하는 Web2.0 시대가 도래 하였다. 대표적인 예로 웹상에 자신의 블로그에 관심 있는 분야에 대한 다양한 리뷰를 올리고, 이러한 리뷰들이 동일한 관심사를 가진 사용자에게 유용한 정보가 되는 경우를 들 수 있다. 예를 들어, 어떤 상품을 구매하려는 사람의 경우에 이 상품에 대한 리뷰들을 읽어 봄으로써 상품에 대한 객관적인 평가를 할 수 있을 것이다. 또한, 이런 평가는 실제로 상품구매결정에 많은 영향을 미치게 된다. 하지만 많은 리뷰들이 실질적인 필자의 의견이 들어간 문장에 비해 상당히 길게 작성되어 있다. 따라서 장문의 리뷰에서 원하는 정보만을 찾아내는 것은 어려운 일이고, 수많은 리뷰들을 읽기에는 시간적인 제약이 따르게 된다. 그러나 이런 제약사항들로 인해 적은 수의 리뷰만 읽어보게 된다면 정보의 객관성이 떨어지는 문제가 발생한다. 이러한 문제점들을 해결하기 위한 방법으로 자동으로 리뷰의 의견들을 모으고 요약 해주는 연구에 대한 관심이 커지고 있다.

본 논문에서는 개인 블로그나 메뉴판닷컴(http://www.menupan.com)과 같은 음식점 웹포탈의 리뷰들을 분석하여 사용자가 원하는 정보를 손쉽게 얻을 수 있도록 하는 것을 목표로 한다.

본 논문의 나머지 부분은 다음과 같이 구성된다. 2장에서는 관련연구에 대해서 설명하고, 3장에서는 오피니언 마이닝(opinion mining)기법을 이용한 음식점 리뷰를 분석하고 요약하는 방법에 대해 설명 할 것이다. 4장 결론에서는 기대되는 효과와 향후 연구 과제에 대해 논한다.

2. 관련 연구

2.1 오피니언 마이닝

오피니언 마이닝이란 글의 주제가 아니라 주제에 대한 필자의 태도에 초점을 맞추고 있는 마이닝 기법을 말한다[1]. 오피니언 마이닝은 다음과 같은 세 가지의 하위 단계로 나눠볼 수 있다.

1. Subjectivity Analysis - 주어진 텍스트가 필자의 주관적인 견해(긍정적이거나 부정적인 감정 또는 의견을 표현하는 것)를 드러내는지 혹은 객관적 입장을 취하고 있는지에 대한 문제를 포함한다[2].
2. Polarity Analysis - 주관적인 견해를 포함하는 텍스트에 대해 그 견해가 긍정적인지 부정적인지 확인하는 과정이다[3].
3. Polarity Degree - 주관적인 견해가 들어간 텍스트에서 극성의 강도를 측정하는 과정이다[4].

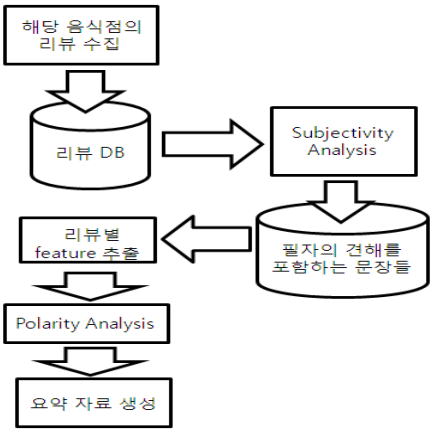
2.2 Feature-based Summary

현재까지 연구된 리뷰 마이닝과 요약에 관한 기법은 대부분 상품평에 관한 것이었다. Liu와 Hu에 의해 연구된 리뷰 요약 기법은 다음의 세 단계로 요약된다[5].

1. 고객이 자신의 의견을 나타낸 상품에 대한 특징을 찾아낸다.
2. 각각의 특징에 대해서 긍정적이거나 부정적인 의견을 포함하는 문장을 찾아낸다.
3. 찾아낸 정보들을 기반으로 요약 자료를 생성한다.

3. 제안 내용

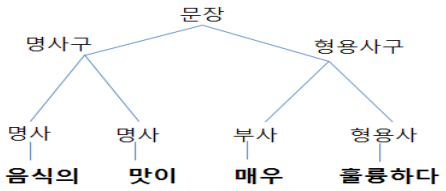
음식점 관련 리뷰를 분석, 요약하기 위해서 크게 다음의 세 단계를 거치게 된다.



(그림 1) 음식점 리뷰 분석과 요약에 대한 구조도

우선 음식점의 이름으로 검색하여 해당 음식점에 대한 모든 리뷰들을 수집한다. 수집된 각각의 리뷰에 대해서 Subjectivity Analysis를 통해 필자의 주관적인 견해가 들어간 문장만을 수집하고 중립적이거나 오피니언을 포함하지 않는 문장은 제외 한다 .

다음으로 전 단계에서 수집된 문장들 중에서 특징을 포함하는 문장을 추출하여 해당 문장에 대해 Polarity Analysis를 수행한다. 이 과정에서 각각의 특징에 대한 opinion word(주관적인 견해를 나타내는 핵심단어)를 찾기 위해 [6]의 영어에 대한 opinion word 추출 기법을 한국어에 적용하였다[그림2].



(그림2) 문장에서 오피니언을 추출하는 기법

예시 문장에서는 “음식의 맛”을 특징으로 추출해내고, “훌륭하다”를 opinion word로 추출하게 되는 것이다. 이러한 경우에는 opinion word를 긍정적인 의미로 해석할 수 있기 때문에 해당 특징에 대한 긍정적 문장의 카운트를 하나 증가시킨다.

○○식당

Feature: **음식의 맛**
긍정적: 259
<긍정적 의견의 문장들>
부정적: 12
<부정적 의견의 문장들>

Feature: **친절도**
긍정적: 132
<긍정적 의견의 문장들>
부정적: 15
<부정적 의견의 문장들>

⋮

(그림3) 요약 자료 예시

마지막으로 두 번째 단계의 과정을 수집된 모든 문장에 적용한 후 최종결과에 대한 요약자료를 생성한다[그림3]. 요약자료에는 사용자가 직접 확인할 수 있도록 긍정적이거나 부정적으로 분류된 문장들을 포함한다.

4. 결론

본 논문에서는 사용자가 원하는 음식점의 리뷰들을 분석하여 요약제시 함으로써 방문 경험이 있는 필자의 평가를 통해 사용자의 객관적인 판단을 도와준다. 또한, 방대한 양의 리뷰들을 일일이 읽어보지 않아도 된다는 점에서 시간적인 측면에서의 효율성을 제공한다.

이번 연구에서는 각각의 특징에 대한 Polarity Analysis를 통해 긍정적 또는 부정적의 두 가지로 분류된 결과만을 보여주지만, 세부적으로 Polarity Degree를 측정하여 리뷰에 나타난 필자의 의견을 점수화하는 것으로 확장할 수 있다.

감사의 글

이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임 (No. 2009-0075771).

참고문헌

[1] Jack G. Conrad, Frank Schilder "Opinion Mining in Legal Blogs", in Proceedings of the 11th international conference on Artificial intelligence and law, pages 231-236, 2007.

[2] Vasileios Hatzivassiloglou, Kathleen R. McKeown "Predicting the Semantic Orientation of Adjectives", in Proceedings of the 8th conference on European chapter of the Association for Computational Linguistics, pages 174-181, 1997.

[3] Bo Pang, Lillian Lee "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", in Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Article No. 271, 2004.

[4] Bo Pang, Lillian Lee "Seeing Stars: Exploiting class Relationships for sentiment categorization with respect to rating scales", in Proceedings of the 43rd Annual Meeting of the ACL, pages 115-124, 2005.

[5] Minqing Hu, Bing Liu "Mining and Summarizing Customer Reviews", in Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 168-177, 2004.

[6] Lipika Dey. S K Mirajul Haque "Opinion Mining from Noisy Text Data", in Proceedings of the 2nd workshop on Analytics for noisy unstructured text data, pages 83-90, 2008.