

반자동으로 구축된 의미 사전을 이용한 한국어 상품평 분석 시스템¹

명재석 이동주 이상구
서울대학교 컴퓨터공학부
{justake, therocks, sglee}@europa.snu.ac.kr

A Korean Product Review Analysis System Using a Semi-Automatically Constructed Semantic Dictionary

Jaeseok Myung, Dongjoo Lee, Sang-goo Lee
School of Computer Science and Engineering, Seoul National University

요 약

웹 2.0 시대에 사용자가 작성한 리뷰는 다양한 활용성을 갖는 가치있는 데이터이다. 특히 온라인 쇼핑몰에서의 상품평은 사용자의 구매 결정에 직접적인 영향을 미치는 중요한 정보이다. 본 논문에서는 실제 쇼핑몰 사이트에 있는 상품평을 분석하여 각 상품의 특징과 이에 대한 사용자의 의견을 요약하는 상품평 분석 시스템을 설계하고 구현하였다. 각 상품평을 분석하는 과정에서는 자연언어처리 기법과 의미 사전을 사용한다. 의미 사전에는 상품의 특징을 표현하는 어휘와 각 어휘들의 극성(Polarity) 정보들을 반자동으로 정의할 수 있도록 구현하였다. 이에 더하여 문맥에 따라 다른 의미를 갖는 어휘에 대한 처리 방법에 대해서도 논의한다. 실험은 2개 상품 분류의 63개 실제 리뷰를 대상으로 수행하였으며 결과로 평균 88.94%의 정확률, 47.92%의 재현율을 나타냈다.

1. 서론

웹 2.0 시대의 흐름에 따라 사용자는 자신이 제공받는 서비스에 대하여 더 적극적으로 자신의 의사를 표현할 수 있는 환경을 맞이하고 있다. 많은 기업들이 자신들이 제공하는 온라인 서비스에 블로그, 태그, 리뷰 등의 기능을 추가하여 사용자의 참여를 유도하고 이를 활용하려 노력하고 있기 때문이다. 특히 온라인 쇼핑몰의 경우 더욱 적극적으로 사용자의 상품평을 수집하고 있다. 상품평의 경우에는 다른 사용자의 구매 의사에 결정적인 역할을 하기 때문에 이를 분석하여 사용자의 요구를 파악하는 것은 기업의 입장에서 매우 중요한 일이다.

그 동안 상품평에 담긴 사용자의 의견을 분석하려는 다양한 시도가 있었다. 특히 영어, 일본어 등에 대한 연구가 많이 존재하였으나 아직 한글에 대한 상품평 분석은 활발히 이루어지지 않고 있다. 이에 따라 본 논문에서는 한국어 상품평에서 사용자의 의견을 추출할 수 있

는 시스템을 구현하고 이 때 발생하는 문제점에 대하여 논의하려고 한다. 좀 더 세부적으로는 상품의 특징에 대하여 사용자의 선호 여부를 확인할 수 있는 부분들을 추출해서 이를 요약함으로써 사용자로 하여금 전체 상품평을 다 보지 않고도 상품평들이 가진 의미를 알 수 있도록 하는 것이 구현 시스템의 목표이다.

본 연구에서는 실제 쇼핑몰 사이트에 있는 상품평을 대상으로 자연언어처리 기법을 활용하여 분석 시스템을 구축하였다. 1장에서는 이와 관련하여 수행된 연구들의 개괄적인 내용을 소개한다. 2장에서는 구현한 시스템인 The Highlights의 특징과 구성, 시스템의 처리 흐름에 대해 논의한다. 3장에서는 구현 시스템에 대한 실험 및 결과 분석, 4장은 결론 및 향후 과제에 대해 논의한다.

1.1. Opinion Mining

Opinion Mining은 글쓴이가 서술 대상에 대하여 말하고자 하는 의견을 임의의 문서로부터 찾아내는 분야로서 최근 여러 연구에 의하여 성숙되고 있으며 Sentiment Analysis, Sentiment Classification, Opinion Extraction으로 표현되기도 한다. Opinion Mining의 대상이 되는 문

¹ 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 육성 지원 사업(IIITA-2007-C1090-0701-0031)의 연구결과로 수행되었음

제의 영역은 매우 넓으며 다음과 같은 간단한 예를 생각해 볼 수 있다.

- 최근의 세금 제도 개편에 대한 여론은?
- 선거에 당선될 것 같은 유력한 후보는?
- 제품에 대한 고객들의 불만사항은?
- 혼자 여행하기 좋은 장소는?

Opinion Mining은 기존의 다양한 연구 분야와도 깊은 연관성을 가지고 있다. 사용자가 원하는 정보를 검색, 요약, 분류한다는 점에서 기존의 정보 검색, 문서 요약, 문서 분류 등의 주제에서 사용하는 많은 방법이 적용 가능하다. 그러나 앞서 언급한 분야들이 주로 문서 단위로 사용자의 필요 정보를 추출했다면 Opinion Mining은 대상의 특징을 기준으로 검색, 요약, 분류를 수행한다는 다른 점이 있다. 데이터 단위의 관계에 주목한다는 점에서 데이터마이닝 분야와도 깊은 연관성을 맺고 있으며 Opinion Mining은 여기에 언어적인 기반과 사용자의 선호 여부를 나타내는 극성을 부여함으로써 좀 더 세부적인 분야로 자리잡았다. 또한 사용자의 의견을 분석하는 데 언어학적인 자원을 활용함으로써 전산언어학 분야와도 깊은 관계를 맺고 있다.

1.2 관련 연구

Opinion Mining을 수행하기 위해서는 다양한 하위 작업이 필요하다. 크게 살펴보자면 Opinion Mining은 주로 다음과 같은 세가지 하위 단계를 필요로 하며 연구의 주제도 이러한 기준으로 분류해 볼 수 있다.

첫째, 대상의 어떠한 특징에 대하여 사용자의 긍정적 혹은 부정적인 자세를 나타내는 극성을 판별하는 방향이다. 이를 위하여 간단히 PMI(Pointwise Mutual Information) 방법을 이용해 볼 수 있다. PMI는 비슷한 성질의 어휘는 가까운 위치에서 함께 나타나는 빈도가 높을 것이라는 가정에 근거하여 단어 사이의 관계를 측정하는 방법이다. 기존의 연구^[1]에서는 긍정적인 어휘의 PMI와 부정적인 어휘의 PMI 지수의 차이를 기준으로 극성을 판별하기도 하였다. 다른 방법으로 기계학습 분야에서 사용하는 기법들을 이용하는 연구^[2]도 소개되었다. 그러나 이러한 방법들은 문맥에 따른 정확한 극성 정보를 얻어내기 어렵다는 한계를 가지고 있다. 이에 따라 자연언어처리 기법을 활용하여 극성을 판별하는 연구^[3]도 수행되어 왔다.

둘째, 대상의 특징을 중점으로 문서를 요약하여 의견 표현 부분을 추출하는 연구 방향이 있다. 이러한 연구들은 대체적으로 1)대상의 특징을 식별하는 단계, 2)특징에 대한 의견을 식별하는 단계, 3)극성과 같이 의견이 갖는 의미를 부여하는 단계, 4)요약된 의견을 시각화하는 단계를 수행한다. 이러한 과정을 통계적인 방법을 사

용하여 구현한 기존의 연구들이 계속 발전하고 있다.^{[4][5][6]} 기존의 연구 결과들이 대체적으로 좋은 결과를 보이고 있지만 문맥을 고려한 정확성 측면에서는 아직 만족스러운 결과를 보이지 못하고 있다. 이에 따라 통계적인 방법을 사용할 때도 단순한 unigram 보다는 bigram, n-gram으로 확장하여 문맥에 따른 극성 정보를 판별하도록 노력하고 있으며 자연언어처리 기법을 이용하여 정확도를 향상시키려는 방법도 시도되고 있다.^{[7][8][9]}

셋째, 대상에 대한 사용자의 의견을 나타내는 언어학적 자원을 정의하고 구축하는 연구도 활발히 논의되고 있다. 위의 두 가지 연구의 진행 과정에는 대상의 특징을 나타내는 주제어는 무엇이고 이에 따른 서술어는 무엇인지 식별하는 과정이 반드시 포함되어야 한다. 또한 각 어휘 조합이 갖는 극성 정보도 필요하다. 따라서 사전의 어휘를 언어적인 특성을 고려하여 어떻게 구성하고 각 어휘가 갖는 의미적 정보를 어떻게 정의하고 구축할 것인가에 대한 논의가 필요하다. 추가적으로 분석을 마친 결과의 형태에 대해서도 생각해볼 수 있다. Opinion Mining 시스템의 출력이 어떠한 형태와 의미 정보를 갖는가에 따라 다양한 시각화 방법 및 활용이 생각될 수 있다. 예를 들어, 출력의 극성 정보가 긍정, 중립, 부정이라는 정보만을 담고 있는가, 혹은 이와 같은 정보를 실수 형태로 가지고 있는가에 따라 시각화 방법 및 순위 시스템 도입과 같은 응용 과정이 큰 영향을 받게 된다. 따라서 양질의 언어 자원 구축은 Opinion Mining 전체의 성능을 좌우할 중요한 주제어이며 이에 대한 연구가 활발히 진행되고 있다. 대표적으로 Appraisal Theory^[10]에서는 이와 같은 언어 자원을 이론적으로 정립하려는 시도가 이루어지고 있다. 또한 통계적인 방법을 이용하여 자동으로 사전을 구축하거나 WordNet을 확장하여 언어 자원을 구축하려는 시도도 존재하였다.^{[11][12]}

이와 같이 많은 연구가 활발하게 진행되고 있으나 한국어에서 사용자의 의견을 추출하려는 시도는 아직 미미한 실정이다. 본 연구는 한국어에서 사용자의 의견을 추출하는 시스템에 대해서 논의하고 있고 특히 사용자의 의견이 많이 나타나 있는 상품평 데이터에 대하여 Opinion Mining 시스템을 설계하고 구현하였다.

2. 상품평 분석 시스템: The Highlights

이번 장에서는 본 연구에서 제안하는 상품평 분석 시스템인 The Highlights에 대해 논의한다. 1절에서는 시스템의 개괄적인 구성과 처리 흐름을 살펴본다. 2절에서는 시스템의 주요 설계 목표와 특징에 대해 논의한다. 3절에서는 정형화된 의미 분석 결과인 Semantic Clause, 4절에서는 분석 과정에서 사용되는 의미 사전에 대해 논의하고 5절에서 예제를 통하여 실제 시스템의 분석 과정을 자세히 살펴본다.

2.1. 시스템 개요

그림 1은 The Highlights의 전체 시스템 구성 및 처리 흐름을 보여준다. 먼저 상품 분류와 그에 맞는 상품 품목, 각 상품에 작성된 상품평을 저장하고 있는 메인 시스템의 데이터베이스가 존재한다. 저장된 상품평은 형태소 분석 단계를 거치면서 문장 구분, 띄어쓰기 보정 등의 과정을 거친다. 이 과정에서 기 분석된 형태소 사진이 사용된다. 형태소 분석 과정에서 추출된 어휘들은 관리자가 언어 자원을 구축할 때 후보 어휘로 사용되기도 한다. 형태소 분석 단계를 마친 상품평은 문장 단위로 구문 분석 단계를 거친다. 구문 분석 단계에서는 의존 문법을 이용하여 주제어와 서술어간의 관계를 정의한다. 구문 분석 트리는 의미 분석 단계의 입력으로 사용된다. 의미 분석 단계에서는 관리자에 의해 구축된 의미 사전의 어휘를 이용하여 구문 분석 트리를 순회하면서 일치하는 어휘를 찾아 상품 분류에 맞는 주제어-서술어가 포함된 의견 표현 부분을 얻어낸다. 각 의견 표현 부분은 해당 어휘의 의미에 맞게 극성이 계산된 Semantic Clause 형태로 변형되어 최종 결과가 데이터베이스에 저장된다.

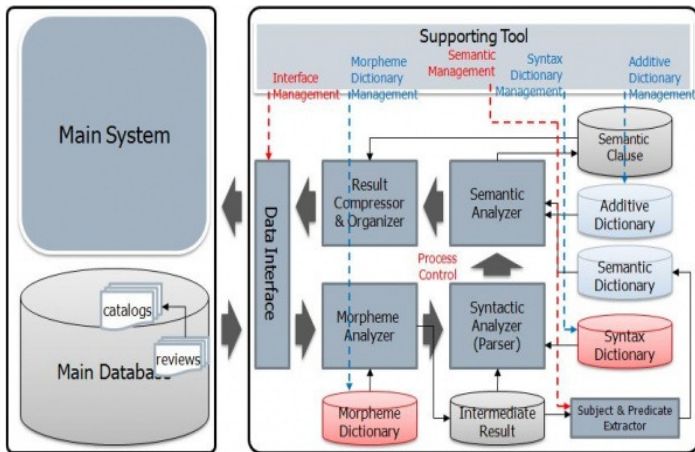


그림 1 시스템 구성 및 흐름

2.2. 제안 시스템의 접근 방향

Opinion Mining의 대부분의 연구는 사용자가 작성한 리뷰를 대상으로 이루어졌다. 리뷰는 의견 표현의 대상이 명확하기 때문에 사용자의 의견에 집중할 수 있는 좋은 실험 데이터이다. 특히 상품에 대한 리뷰는 상품 자체가 가지고 있는 특징이 뚜렷하고 사용자의 선호 여부가 명확하게 드러나기 때문에 Opinion Mining의 좋은 연구 대상이 되고 있다. 본 시스템은 상품평을 분석하는 목적을 가지고 있으며 설계 과정에서 다음과 같은 특징에 중점을 두었다.

첫째, **한국어 상품평에 대해 자연언어처리 기법을 도입한다.** 사용자의 의견을 추출하고 요약할 때, 많은 의

견을 찾는 것보다는 정확한 의견을 식별해 내는 것이 의견이 가지고 있는 원래 의미를 손상시키지 않는다는 점에서 더욱 유용하다. 특히 상품평은 요약된 의견이 갖는 원래 의미가 손상될 경우 치명적인 판단 오류를 야기할 가능성이 높다. 따라서 기존의 연구들이 n-gram으로 어휘를 확장하여 문맥상의 극성을 식별하거나 자연언어처리 기법을 도입하고 있으나 다른 외국어에 비하여 한국어에 대한 연구는 아직 부족한 실정이다.

둘째, **관리자에 의해 정제된 의미 사전을 구축한다.** 추출된 의견의 정확도는 처리 과정에서 사용되는 언어 자원의 질과 높은 연관 관계를 갖는다. 기존의 연구들이 완전 자동화 시스템을 목표로 특징 추출 단계를 거치는 것에 비하여 제안 시스템은 관리자에 의해 정제가 가능한 반자동화 방식의 특징 추출 단계를 통하여 사전을 구축한다. 시스템이 추천하는 후보 어휘를 관리자가 선택하게 함으로써 사전 구축의 용이성과 언어 자원의 질적 향상을 동시에 고려하였다. 후보 어휘는 실제 리뷰에서 사용된 어휘의 빈도를 이용하는 간단한 방법을 사용하여 추천하였다.

셋째, **동일한 어휘라 하더라도 쓰임에 따라 의미가 달라지는 언어적인 특징을 고려한다.** 특히 상품평의 경우에는 동일한 서술어가 상품 분류에 따라 다른 의미로 사용되는 경우가 많다. 예를 들어 ‘디지털 카메라’에서의 ‘크기’는 ‘작다’라는 서술어에 의해 좋은 의미를 가질 수 있지만 ‘모니터’에서의 ‘크기’가 ‘작다’는 나쁜 의미로 사용된다. 제안 시스템은 이러한 특성을 고려하여 상품 분류 별로 서술어에 대한 의미를 따로 관리하고 원래 의미에 덧붙여 언어 자원을 구축할 수 있도록 구현하였다

넷째, **정형화되고 의미 있는 결과 형태를 정의한다.** 추출한 의견 정보를 어떤 형태로 요약하여 저장하는가에 따라 결과의 시각화 방법 및 활용 범위가 달라질 수 있다. 제안 시스템은 Semantic Clause라는 의미 있는 단위로 의견 정보를 요약하여 다양한 활용이 가능하게 하였다. Semantic Clause에 대해서는 다음 절에서 자세히 논의한다.

2.3. Semantic Clause

일반적으로 의견 표현 부분은 대상의 어떤 특성을 나타내는 주제어, 해당 특성이 어떻다고 기술하는 서술어, 의미상의 긍정 및 부정을 나타내는 극성 등의 정보를 갖는다. 그러나 추출 결과의 세부적인 의미는 이보다 더 복잡한 형태를 가질 수 있으며 본 논문에서 구현한 시스템은 Appraisal Theory에 기반하여 다음과 같은 Semantic Clause라는 결과 형태를 정의 하였다.

$$SC = \{S, P, SL, Pt, St, Ms\}$$

- *S* (Subject, Feature) : 주체어, 대상의 특성을 나타낸다. 어떤 상품의 배송, 색상, 크기 등을 의미한다. 보통 상품평에서 *S* 는 생략 가능하거나 단일 체언, 복합어 등 여러 가지 언어학적 조합에 의해 구성될 수 있으나 본 시스템에서는 단일 체언으로 구성된 *S* 에 대해서만 처리하였다.

- *P* (Predicate, Opinion Word) : 서술어, 대상의 특성을 서술한다. 예를 들어 ‘배송’ 은 ‘빠르다’, ‘느리다’, ‘좋다’, ‘나쁘다’ 와 같은 어휘에 의해 서술된다. 언어학적 관점에서 서술어는 용언, 체언+서술격조사, 부사어+서술어와 같은 여러 가지 형태로 성립되지만 본 시스템에서는 단일 용언으로 이루어진 서술어에 대해서만 처리하였다.

- *SL* (Standard Label) : *SL* 은 *S*, *P* 의 표준화된 형태의 결합으로 이루어진다. 예를 들어, ‘색깔이 이뻐요’, ‘색상이 예뻐요’ 와 같은 표현들은 (색상, 예쁘다)와 같은 표준화된 *SL* 을 갖는다.

- *Pt* (Polarity) $\in \{Negative, Neutral, Positive\}$: *Pt* 는 *P* 가 *S* 를 긍정적, 부정적 혹은 중립적으로 서술하는지를 나타낸다. 시스템에서는 $\{-1.0, 0, 1.0\}$ 의 실수 값으로 각각을 나타내었다.

- *St* (Strength) : *St* 는 *P* 가 *S* 를 서술하는 의미적인 강도를 나타내는 실수 값을 의미한다. 예를 들어, ‘싫다’ 라는 표현을 1.0 정도의 강도로 표현했다면 ‘혐오스럽다’ 와 같은 표현은 그보다 강한 의미적 강도를 나타내도록 한다.

- *Ms* (Modifiers) = $\{m\}$: *m* 은 *P* 의 의미를 반의, 강조, 약화 시키는 부사 혹은 부사구, 부정 술어 형태를 의미한다. 예를 들어, ‘안 좋다’, ‘너무 좋다’ 와 같은 표현은 ‘좋다’ 는 서술어에 부가적인 의미가 부여되어 원래 의미의 강도와 극성이 변하는 형태로 볼 수 있다.

분석 결과의 활용은 결과가 얼마나 유용한 형태로 가공되는가에 따라 달라질 수 있다. 깊고 자세한 분석을 위해서는 그에 맞는 가공 형태가 필요하다. 얼마나 자세하게 결과에 의미를 부여하는가는 각 시스템이 정의할 문제가 될 수 있겠지만 기존의 연구들이 추출하는 주체어, 서술어, 극성 정보 이외에도 여러 가지 중요한 정보가 존재하며 시스템 설계 시에는 이러한 사항을 고려해야 한다.

2.4. 의미 사전 구조 및 구축

상품평에 대해 형태소 분석, 구문 분석 단계를 수행하면 구문 분석 트리가 생성된다. 생성된 구문 분석 트리는 의미 분석 단계의 입력으로 이용된다. 의미 분석 단계에서는 구문 분석 트리를 순회하며 *S*, *P*, *Ms* 의 조건에 맞는 서브 트리를 추출하게 되는데 이때 사용되는 어휘들은 적절한 형태로 정의되어야 한다. 본 시스템은 이러한 어휘 정보의 저장소를 의미 사전으로 정의하였다. 이

절에서는 본 시스템에서 사용한 의미 사전의 구조 및 특징, 구축과 관련한 사항에 대해 논의한다.

2.4.1. 상품에 대한 의미 사전 구조

본 시스템에서 사용하는 어휘는 크게 두 가지 형태로 나누어 볼 수 있다. 한 가지는 상품 자체의 특징을 기술하는데 사용하는 어휘이며, 다른 하나는 상품 자체와는 상관이 없으나 언어적인 측면에서 부가적인 정보를 부여하는 어휘이다. 두 가지 분류는 서로 다른 특성을 가지고 있기 때문에 저장되는 사전의 구조 또한 달라져야 한다. 본 연구에서는 전자를 기술 어휘 사전, 후자를 부가 어휘 사전으로 구분하였다.

기술 어휘 사전에는 다음과 같은 정보가 저장되며 그림 2에서는 기술 어휘 사전의 개괄적인 구조를 그림 형태로 표현하고 있다.

- 상품 분류(Product Category) : 각 상품 품목들은 상품 분류에 귀속된다. 예를 들어, ‘캐논 EOS 350D’ 와 같은 상품은 ‘디지털 카메라’ 라는 상품 분류에 속한다.

- 주체어 범주(Subject Category) : 상품의 특성을 기술하는 정보들은 특정한 범주에 속할 수 있다. 예를 들어, ‘크기’, ‘무게’ 등은 ‘규격’ 이라는 범주에 포함되어 있다고 볼 수 있다. 본 시스템에서는 각 상품 분류에 ‘배송’, ‘성능’, ‘가격’, ‘디자인’ 등 다양한 범주를 각각 부여할 수 있다. 즉, 전자제품은 성능, 디자인, 내구성과 같은 주체어 범주를 가질 수 있으나 의류는 디자인, 품질과 같은 다른 범주를 가질 수 있다.

- 주체어 범주에 따른 주체어(Subject) : 각 주체어 범주에는 그에 해당하는 주체어가 속할 수 있다. 예를 들면, ‘배송’ 이라는 범주에는 ‘배송’, ‘포장’, ‘택배’ 등의 주체어가 속할 것이다. 또한 각 주체어는 그에 맞는 표준형을 가질 수 있다. ‘상품’, ‘물건’, ‘제품’ 등은 모두 ‘제품’ 이라는 표준형으로 서술할 수 있다.

- 주체어에 따른 서술어(Predicate) : 각 주체어들은 그에 맞는 서술어를 갖는다. 배송은 빠르거나 느리지만 사이즈는 크거나 작은 속성을 갖는다. 따라서 의미 사전은 각 주체어에 대한 서술어를 따로 관리한다. 서술어 또한 표준형으로 통합될 수 있다. 이에 더하여 주체어와 서술어 간에 의미 관계를 표현하는 극성 및 의미 강도와 같은 정보를 포함해야 한다.

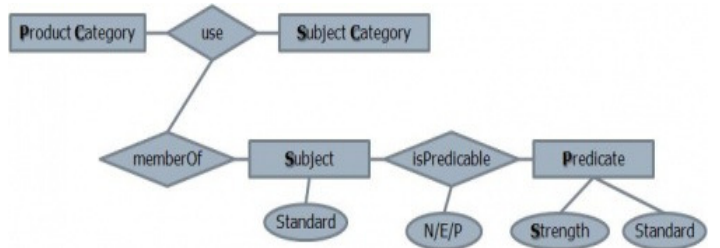


그림 2 기술 어휘 사전 기본 구조

부가 어휘 사전에는 다음과 같은 정보가 포함된다.

- 부가 어휘(Additive Voca) : 상품에 대한 서술 정보와 상관 없이 언어적인 측면에서, 표현에 부가적인 정보를 부여하는 어휘들이 존재한다. ‘매우’, ‘조금’, ‘~ㄴ 것 같다’와 같은 어휘들은 의미적인 강도에 영향을 주는 어휘로서 의미 사전에서 따로 관리되어야 한다. 이때 ‘반의’, ‘강조’ 등의 의미를 표현하는 Label과 영향력을 나타내는 실수 값인 Affecting Power 정보를 가질 수 있다.

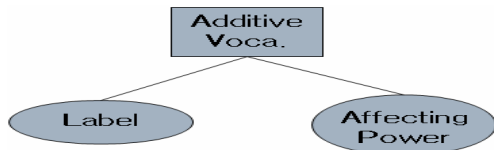


그림 3 부가 어휘 사전 구조

2.4.2. 사전 구축

잘 정의된 의미 사전은 Opinion Mining에서 중요한 역할을 한다. 완전 자동화를 목표로 하는 기존 시스템은 문서를 대상으로 특징 추출 단계를 수행하여 주제어와 서술어를 추출하지만 직접 구축한 언어 자원에 비하면 정확성을 보장하기 어렵다. 그러나 사람이 모든 어휘를 일일이 등록하는 것은 현실적으로 어려운 방법이며 관리자의 편의성을 위하여 후보 어휘들을 추천하는 단계가 있다면 사전 구축에 도움이 될 수 있다. 본 시스템은 실제 상품평에서 어휘가 나타난 빈도수를 기준으로 관리자께 후보 어휘를 추천하는 간단한 방법을 사용하였다. 사전 구축 단계에서는 관리자가 개입하는 반자동화 시스템이지만 일단 한 번 구축된 사전의 어휘는 등록된 분류에 포함된 모든 상품 품목에 대하여 적용 가능하기 때문에 사전 구축 단계를 지나 실제 분석을 수행하는 적용 단계에서는 관리자의 개입 없이 빠른 수행 성능을 보일 수 있다.

2.4.3. 어휘의 쓰임에 따른 의미 변화

문서에서 외형적으로는 동일한 어휘라 하더라도 쓰임에 따라 의미가 달라지는 경우가 많다. 특히 상품평 데이터에서는 동일한 표현이라 하더라도 상품의 분류에 따라 긍정적, 부정적 의미가 변화하는 경우가 존재한다. 앞서 언급한 예처럼 ‘디지털카메라’에서 ‘작다’는 어휘의 경우 좋은 의미로 사용될 수 있지만 ‘모니터’ 분류에서는 나쁜 의미로 사용될 수 있기 때문이다. 이러한 문제점은 의견 표현 부분이 갖는 극성 정보에 결정적 영향을 미치게 되어 전체 결과의 정확성을 떨어뜨리는 요인이 된다.

본 연구에서는 이러한 문제점을 해결하기 위해서 상품 분류에 따라 S-P 쌍에 대한 극성 정보와 의미 강도를 덮어 쓸 수 있도록 구현하였다. 상품 분류에 따라 덮어 쓸 의미 사전이 없을 경우에는 기본 값을 적용한다. 예를 들어, ‘크기’ - ‘작다’는 표현이 기본적으로는 긍정적인 극성 정보와 1.0의 의미 강도를 갖는다면 ‘모니터’라는 상품 분류에서는 부정적으로 분류되고 2.0의 의미 강도를 갖도록 설정할 수 있다. 실제 상품평 분석 시에는 기본적으로 정의된 의미 사전을 로딩한 이후에 해당 상품 분류에 특화된 의미 사전이 있는지 검색한다. 만약 재정의된 의미 사전에 해당 어휘가 있다면 극성 정보와 의미 강도 정보 등을 다시 로딩하여 사용한다.

2.5. 상품평 분석 과정 예제

이번 절에서는 ‘이 물건은 배송이 빨라서 정말 좋네요’라는 예제 문장에 대하여 의견 표현 부분을 추출하는 시스템의 분석 과정을 단계별로 살펴본다. 사용자가 작성한 상품평은 먼저 형태소 분석 단계를 거치게 된다. 형태소 분석기는 원형 복원을 위한 기 분석 사전을 활용하여 복잡한 음원 복원 과정을 생략하였다. 또한 모든 음절에 대해 분석 가능한 후보를 생성하는 동적 프로그래밍 기법을 활용하고 각 음절 단위에 대한 사전을 해시 형태로 구현하여 탐색 성능을 고려하였다. 분석 가능한 정보를 생성하고 나면 여러 가능 후보 중 최적의 후보를 선택해야 할 필요가 있으며, 이 부분은 선호 조건에 대한 규칙을 표현하고 이를 사전에 저장하는 방식으로 구현하였다. 형태소 분석기의 구현은 기존 방법^{[13][14]}을 참조하였으며 자세한 내용은 생략한다. 상품평에 대한 형태소 분석을 수행하고 나면 상품평을 문장 단위로 구분하고 띄어쓰기를 보정하는 과정을 수행한다^[15]. 이 과정에서 문장 부호, 종결형 어미, 다음 어절과의 관계를 고려한다. 다음은 이러한 과정을 거친 형태소 분석의 결과이다.

이	=> [이/DT/S]
물건은	=> [물건/NN/S+은/JO/S/CL/SB]
배송이	=> [배송/NN/S+이/JO/S/CL/SB]
빨라서	=> [빠르/AJ/S+아서/EM/S/CN/DP]
정말	=> [정말/AD/S]
좋네요	=> [좋/AJ/S+네요/EM/S/ED/NM]

표 1 형태소 분석 결과

형태소 분석 단계를 마친 중간 결과는 어절의 리스트로 이루어진 문장 단위로 구문 분석 과정을 거치게 된다. 구문 분석은 어순이 자유로운 한국어의 특성에 적합한 의존 문법을 이용한 구문 분석기를 사용하였다.^[16] 일반적으로 의존 문법을 이용한 구문 분석기는 하나의 의존소가 여러 개의 지배소를 갖지 않도록 모호성을 해결해 줄 필요가 있다. 기존의 연구들에서는 어휘 공기 빈도를 이용하거나 어휘 별로 수식 거리를 적용하는 방법을 통

하여 이러한 문제를 해결하는 방법이 소개되어 왔다.^{[17][18]} 본 연구에서도 마찬가지로 수식 거리를 이용하여 적절한 지배소를 결정하는 방법을 사용하였다. 그러나 각 어휘에 수식 거리를 할당하는 방법 대신에 의존 규칙 자체에 수식 거리를 포함하도록 설계하여 특정 어휘에 대한 수식 거리 정보가 없을 경우에도 규칙을 통하여 적절한 지배소를 선택하고자 노력하였다. 예제 문장에 대하여 생성된 구문 분석 트리는 다음과 같다.

[Node] 좋네요
[ARC] 좋네요 -> 정말{ 수식어 }
[Node] 정말
[ARC] 좋네요 -> 빨라서{ 연결 }
[Node] 빨라서
[ARC] 빨라서 -> 배송이{ 수식대상 }
[Node] 배송이
[ARC] 좋네요 -> 물건은{ 수식대상 }
[Node] 물건은
[ARC] 물건은 -> 이{ 수식어 }
[Node] 이

표 2 구문 분석 결과

시스템이 사용하는 의미 사전에서는 각 어휘를 단일 체언 혹은 용언의 기본형으로 저장하고 있다. 따라서 의미 분석 과정에서는 먼저 구문 분석 트리의 각 어절을 기본형으로 변형한다. 의미 분석 과정에서는 기본형으로 변형된 구문 분석 트리를 순회하며 의견 표현 부분을 찾는다. 이 때 순회 알고리즘은 기본적으로는 BFS(너비 우선 탐색)를 이용한다. 즉, 루트 노드부터 차례대로 큐에 삽입한 후 각 노드를 방문하여 의미 사전의 어휘와 비교한다. 일반적으로 BFS는 빠른 성능을 보장하지는 못하지만 하나의 문장을 구성하는 구문 분석 트리의 어절 노드의 개수는 많지 않기 때문에 이러한 방법을 사용할 수 있다. BFS 도중 상품 분류에 속하는 *S*, *P*, *Ms* 등을 만날 경우에는 해당 노드의 하위 서브 트리를 DFS(깊이 우선 탐색)로 순회하며 연결된 *S*, *P*, *Ms* 를 찾는 과정을 거친다. 이러한 과정을 통하여 모든 노드를 한번씩 방문한다. 순회 도중 *S* 와 *P* 의 연결 관계를 찾으면 기술 어휘 사전으로부터 해당 어휘 조합의 극성 정보와 의미 강도를 가져오고 부가 어휘 사전으로부터 *Ms* 의 Label과 Affecting Power를 가져온다. 이 결과로부터 Semantic Clause를 생성하고 시각화 할 수 있도록 저장한다. 예제 문장에서는 표 3과 같은 형태의 분석 결과를 얻을 수 있다. 결과의 NEP(Negative/nEutral/Positive)는 Semantic Clause의 *Pt*를 긍정, 중립, 부정에 따라 1, 0, -1로 변환한 값에 *St* 값을 곱하고, 다시 *Ms* 의 Affecting Power를 곱셈하여 계산한 값이다. 본 연구에서는 간단한 곱셈을 통하여 극성 정보 및 의미 강도, 부가 어휘의 영향력을 통합하려 시도하였으나 응용 시스템의 쓰임에 맞도록 다른 형태의 통합을 시도하거나 각각의 정보를 따로 활용하는 것도 생각할 수 있다.

[Semantic] S:배송, P:좋다, NEP:2.0, Modifier:정말
[Semantic] S:배송, P:빠르다, NEP:1.0, Modifier:
[Semantic] S:물건, P:좋다, NEP:2.0, Modifier:정말

표 3 의미 분석 결과

그림 4는 실제 상품평을 분석하여 추출한 의견 표현 부분의 일부를 보여주고 있다.

[80753593] [커리어룩 정장 3종 set] 국내산 벨트자켓+랩스카 야에서 34가지의 Feature/Opinion이 추출되었습니다.	
Category	
가격	"가격이 저렴하다"라고 총 2번 평되었습니다. "가격이 좋다"라고 총 2번 평되었습니다.
기타	"벨트가 이쁘다"라고 총 1번 평되었습니다. "벨트가 작다"라고 총 1번 평되었습니다. "벨트가 짧다"라고 총 1번 평되었습니다.
디자인	"디자인이 괜찮다"라고 총 1번 평되었습니다. "디자인이 예쁘다"라고 총 1번 평되었습니다. "색상이 다르다"라고 총 1번 평되었습니다. "색상이 예쁘다"라고 총 1번 평되었습니다.
묘사	"길이가 짧다"라고 총 1번 평되었습니다. "상의가 크다"라고 총 1번 평되었습니다. "치마가 난감하다"라고 총 1번 평되었습니다. "핑크색이 이쁘다"라고 총 1번 평되었습니다.
배송	"배송이 빠르다"라고 총 1번 평되었습니다. "배송이 적당하다"라고 총 1번 평되었습니다.
품질	"단추 구멍이 작다"라고 총 1번 평되었습니다. "바느질이 좋다"라고 총 1번 평되었습니다. "사이즈가 괜찮다"라고 총 1번 평되었습니다. "사이즈가 넉넉하다"라고 총 3번 평되었습니다. "사이즈가 크다"라고 총 3번 평되었습니다. "상품이 괜찮다"라고 총 1번 평되었습니다. "상품이 좋다"라고 총 1번 평되었습니다. "옷이 얇다"라고 총 1번 평되었습니다. "옷이 오그라들다"라고 총 1번 평되었습니다. "옷이 이쁘다"라고 총 2번 평되었습니다. "옷이 좋다"라고 총 1번 평되었습니다. "옷이 저그러지다"라고 총 1번 평되었습니다. "옷이 차분하다"라고 총 1번 평되었습니다. "옷이 크다"라고 총 2번 평되었습니다. "옷감이 얇다"라고 총 1번 평되었습니다. "옷감이 좋다"라고 총 2번 평되었습니다. "제품이 이쁘다"라고 총 1번 평되었습니다. "제품이 좋다"라고 총 1번 평되었습니다.

그림 4 추출된 의견 표현 부분

3. 실험

3.1. 수행 방법 및 결과

시스템의 분석 결과를 확인하기 위해 다음과 같은 실험을 수행하였다. 먼저 실제 운영중인 온라인 쇼핑몰 사이트인 '베스트바이어^[19]'에서 스커트 분류의 상품 1개에 대한 32개의 리뷰를 수집하였고 디지털카메라 분류의 1개 상품에 대한 31개의 리뷰를 수집하였다. 시스템의 실험 데이터가 실제 상품평을 대상으로 하여 비문이 많은 점, 한국어 시스템이라는 점, 관리자에 의해 정제된 특징 추출 과정을 거쳐서 언어 자원을 구축하는 점 때문에 다른 시스템과의 직접 비교는 어려운 문제였다. 따라서 수집한 리뷰에 대해서는 사람이 직접 보고 의견 표현 부분을 찾아내었다. 그 결과 스커트는 105개, 디지털카메라는 46개의 의견 표현 부분을 찾아내었다. 실험은 시스템이 의견 표현 부분을 얼마나 정확히 찾아내는가를 기준으로 수행하였다. 실험을 수행한 결과는 다음과 같다.

	Person 0	Person X
System 0	(A) 39	(B) 7
System X	(C) 66	(D) X

표 4 스킨트 상품평 분석 결과

	Person 0	Person X
System 0	(A) 27	(B) 2
System X	(C) 19	(D) X

표 5 디지털카메라 상품평 분석 결과

Precision = A / (A+B) : 전체 추출된 결과 중 맞는 결과
Recall = A / (A+C) : 맞는 결과 중 실제 추출된 결과
스킨트 : P = 39 / (39+7) => 84.78%
R = 39 / (39+66) => 37.14%
디지털카메라 : P = 27 / (27+2) = 93.10%
R = 27 / (27+19) = 58.69%

표 6 분석 결과의 정확률과 재현율

3.2. 결과 분석

실험 결과는 비교적 높은 정확률을 보여주고 있지만 재현율은 아직 보완할 점이 많음을 보여주고 있다. 일반적으로 상품평에서 사용자의 의견이 표현된 부분을 뽑아내는 것은 상품의 특성을 기준으로 상품평을 요약하는 과정과 유사하다고 볼 수 있다. 이러한 측면에서 볼 때 정확한 사용자의 의견을 요약해야 한다는 점에서 높은 정확률은 필수적이라고 볼 수 있다. 실험에 사용된 상품평이 실제 쇼핑몰에서 작성된 비문이 많은 데이터라는 점을 감안했을 때 시스템의 정확률은 받아들일만한 수준으로 볼 수 있다.

정확률에 비교하여 재현율은 현저히 낮은 수치를 보였다. 이러한 결과는 여러 가지 원인에 의해 나타날 수 있겠지만 첫째로는 시스템에서 정의한 Semantic Clause의 모델 구조상 한계로 보여진다. 앞 장에서 논의한 Semantic Clause에서 본 시스템은 S, P에 대하여 단일 체인, 단일 용언이라는 제약을 부여하였다. 따라서 사람이 찾아 낸 의견 표현 부분 중에는 시스템에서 원칙적으로 처리하지 못한 부분도 존재하였다. 실제로 시스템이 처리할 수 있는 의견 표현 부분을 분류해본 결과 스킨트의 경우 59(56.19%)개, 디지털카메라는 33(71.73%)개가 있었다. 이 데이터를 기준으로 실험 결과를 보이면 다음과 같다.

	Person 0	Person X
System 0	(A) 39	(B) 7
System X	(C) 20	(D) X

표 5 스킨트 SC 제약 적용 데이터 실험 결과

	Person 0	Person X
System 0	(A) 39	(B) 7
System X	(C) 6	(D) X

표 6 디카 SC 제약 적용 데이터 실험 결과

이를 기준으로 결과를 다시 분석하면 정확률은 변동이 없으며, 재현율은 스킨트가 66.10%, 디지털카메라는 81.81%의 결과를 보인다. 결국 언어학적 모델을 Semantic Clause에 어떻게 적용하는가에 따라 성능에 많은 차이를 보임을 알 수 있다.

정확률과 재현율에 있어서 영향을 주는 또 다른 요인은 자연언어처리의 성능으로 볼 수 있다. 시스템은 실제 상품평에 대하여 문장 구분, 형태소 분석, 구문 분석과 같은 여러 자연언어처리 과정을 거치며 각 단계의 결과가 다음 단계로 연결되는 구조를 가지고 있다. 따라서 선행 단계의 오류가 이후의 단계들에도 영향을 미칠 수 있다는 점을 생각해 볼 수 있다. 따라서 정확률과 재현율 또한 각 단계의 오류에 영향을 받은 수치로 볼 수 있다. 이러한 측면을 고려했을 때 시스템의 정확률은 높은 수준으로 볼 수 있으며 상대적으로 낮은 재현율 또한 이러한 영향을 많이 받은 것으로 생각된다.

4. 결론

참여와 개방, 공유를 지향하는 웹 2.0의 흐름은 인터넷의 문화를 크게 바꾸어 놓았다. 정보를 일방적으로 소비하기만 했던 사용자들의 참여는 리뷰라는 형태로 활성화되고 있으며 이로부터 유용한 정보를 뽑아내는 것은 매우 중요한 일이다. 이러한 측면에서 The Highlights 시스템은 상품평에 대한 실용적인 의미 분석 시스템으로 가치가 있다. 상품의 특징에 대한 사용자의 의견을 추출하거나 서비스 센터 등에 도입하여 고객의 불만 사항을 추출하여 정리하는 등 활용할 수 있는 분야는 매우 많을 것으로 보인다.

The Highlights 시스템은 상품평을 분석하기 위하여 자연언어처리 기법들을 활용하였고 Semantic Clause라는 의미 있는 결과 형태를 정의하였다. 또한 분석 과정에서 사용될 의미 사전의 어휘들을 반자동화 방식으로 구축하고 이에 따라 분석을 수행하였다. 또한 대상의 분류에 따라 어휘의 의미가 변화하는 특성에 대하여 상품 분류별 사전 관리 방식으로 대응하였으며 실제 온라인 쇼핑몰에서 작성된 리뷰 데이터에 대해서도 비교적 높은 정확도를 보였다.

그러나 높은 정확성을 유지하면서 재현율을 높이는 것은 어려운 작업이었으며 향후에는 Semantic Clause 모델의 확장, 자연언어처리 기법의 개선을 통하여 정확률과 재현율을 동시에 높일 수 있는 방안에 대해 연구할 필요가 있다.

참고 문헌

- [1] Turney,P.D., Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia, Pennsylvania, pp. 417-424. (NRC #44946) 2002
- [2] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volumn 10, 2002
- [3] Theresa Wilson, Janyce Wiebe, Paul Hoffmann, Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis, HLT/EMNLP, 2005
- [4] Kushal Dave, Steve Lawrence, David M. Pennock, Mining the peanut gallery : opinion extraction and semantic classification of product reviews, WWW 2003: 519-528
- [5] Bing Liu, Mingqing Hu and Junsheng Cheng, "Opinion Observer: Analyzing and Comparing Opinions on the Web" Proceedings of the 14th international World Wide Web conference(WWW-2005), May 10-14, 2005, in Chiba, Japan
- [6] Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhail Felker, Herman Ng, Chun Jin, Red Opal:product-feature scoring from reviews, ACM Conference on Electronic Commerce 2007:182-191
- [7] Hiroshi Kanayama, Tetsuya Nasukawa, Hideo Watanabe, Deeper sentiment analysis using machine translation technology, Coling 2004: 20th International Conference on Computational Linguistics, 23-27 August 2004, University of Geneva, Switzerland, Proceedings; 7pp
- [8] Jeonghee Yi, Wayne Niblack, Sentiment Mining in WebFountain, ICDE 2005:1073-1083
- [9] Ana-Maria Popescu, Oren Etzioni, Extracting Product Features and Opinions from Reviews, HLT/EMNLP 2005
- [10] Martin J. R., White P. R., The Language of Evaluation: Appraisal in English, 2005
- [11] Hatzivassiloglou V., Mackeown K., Predicting the Semantic Orientation of Adjectives, In Proc. Of 35th ACL/8th EACL, 1997
- [12] Esuli A., Sebastiani F., SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining , In Proceedings of LREC-06, 5th Conference of Language Resources and Evaluation, Genova, IT, 2006, pp. 417-422
- [13] 심광섭, 양재형, 인접 조건 검사에 의한 초고속 한글 형태소 분석기, 정보과학회논문지: 소프트웨어 및 응용 제 31권 제1호, 2004. 1, pp.89-99
- [14] 양승현, 김영섭, 부분 어절의 기분석에 기반한 고속 한국어 형태소 분석 방법, 정보과학회논문지: 소프트웨어 및 응용 제 27권 제 3호, 2000.3, pp.290-301
- [15] 강승식, 한글 문장의 자동 띄어쓰기를 위한 어절 블록 양방향 알고리즘, 정보과학회논문지: 소프트웨어 및 응용 제 27권 제4호, 2000.4, pp.441-447
- [16] 나동렬, 한국어 파싱에 대한 고찰, 정보과학회지 제12권 제8호, 1994. 9, pp. 33-46
- [17] 홍영국, 이종혁, 이근배, 의존문법에 기반을 둔 한국어 구문 분석기, 한국정보과학회 1993년도 봄 학술 발표논문집 제 20권 제1호, 1993. 4, pp. 781-784
- [18] 우연문, 송영인, 박소영, 임해창, 지배가능 경로 문맥을 이용한 의존 구문 분석의 수식 거리 모델, 정보과학회논문지: 소프트웨어 및 응용 제34권 제2호, 2007.2
- [19] <http://www.bb.co.kr>