

# 전자 상거래의 온라인 감성분석처리가 필요한 이유

## Why do we need the On-Line Sentiment Analytical Processing in e-Business?

연종흠(Jongheum Yeon)\*, 이동주(Dongjoo Lee)\*,  
심준호(Junho Shim)\*\*, 이상구(Sang-goo Lee)\*\*  
jonghm@europa.snu.ac.kr, therocks@europa.snu.ac.kr  
jshim@sookmyung.ac.kr, sglee@europa.snu.ac.kr

### 초 록

전자 상거래 사이트의 상품 리뷰는 구매 예정자들에게 유용한 정보로 활용될 수 있지만, 방대한 양으로 인해 사용자가 모든 리뷰를 읽는 것은 불가능에 가깝다. 따라서 감성 분석(Sentiment Analysis)을 활용하여 의견 정보를 요약하는 연구가 이루어지고 있다. 하지만 기존의 감성 분석은 구매예정자에게 특징별로 요약된 평가 등을 제공하는 것에 초점을 맞추고 있어, 상대적으로 의견 정보를 다양한 각도로 분석하는 데이터 모델, 질의 및 시각화 방법과 같이 관리자의 의사 결정을 지원하는 것에 대한 연구는 부족한 실정이다. 따라서 의견 정보를 효과적으로 분석해주는 새로운 처리기법이 필요하며 이를 온라인 감성 분석 처리, OLSAP(On-Line Sentiment Analytical Processing)라 제안한다.

### 1. 서론

웹을 통한 전자 상거래와 정보 공유가 활발해짐에 따라 상품에 대한 리뷰 문서가 기하급수적으로 증가하였다. 사용자들은 상품을 구매한 사이트뿐만 아니라 트위터, 블로그, 페이스북과 같은 소셜 미디어를 통해 자연스럽게 의견을 공유하고 상품 구매시 실제적인 도움을 받는다. 하지만 리뷰 문서의 방대한 양으로 인해 구매 예정자가 모든 리뷰 문서를 읽고 제품에 대한 전체적인 평

가를 파악하는 것은 점점 더 어려워 지고 있다.

상품에 대한 주요 의견을 효과적인 정보의 형태로 전달하는 것이 실제 구매에 영향을 주기 때문에[1], 전자 상거래 사이트인 Amazon.com은 상품 자체의 정보뿐만 아니라 상품에 대한 리뷰를 효과적으로 전달하는 것에 높은 비중을 두고 있다. 구체적으로 상품 자체의 만족도 점수와 세부 특징별 점수를 구매자로부터 입력 받고 이들 점수를 종합하여 제공하며, 구매 예정자들이

---

\* 서울대학교 컴퓨터공학부 박사과정  
\*\* 숙명여자대학교 컴퓨터과학부 교수  
\*\*\* 서울대학교 컴퓨터공학부 교수

유용하다고 평가한 리뷰나 상품 특징을 강조하여 표시한다. 하지만 이 대부분의 과정은 사용자나 시스템 관리자에 의해 수작업으로 진행되기 때문에, 대량의 문서 리뷰를 대상으로 할 때는 효율성과 비용의 문제가 발생한다.

이러한 문제를 해결하고자 리뷰 문서에서 의견을 자동으로 추출하고 분석하는 연구가 시도되었다. 대표적으로 자연언어처리 기법이나 기계 학습에 기반하여 의견의 긍정, 부정을 판별하는 연구를 비롯하여 [2, 3, 4], 통계적 분석 기법에 기반하여 상품에 대한 사용자의 점수, 상품 특징 어휘 빈도 수로부터 특징 단위의 평가를 도출하는 연구[5, 6, 7, 8]가 진행되었다.

시스템에 모이는 의견 정보가 증가할수록 OLAP(On-Line Analytical Processing) 처럼 의견 정보를 다양한 각도로 분석하고, 의사 결정 지원에 활용하고자 하는 요구가 증가하고 있다. 앞서 살펴본 바와 같이 현재까지의 연구들은 구매 예정자를 대상으로 리뷰를 특징 단위의 점수로 요약하거나, 특정 키워드에 연관된 의견 성향을 판단하는 것에 주로 초점을 맞춰왔다. 이러한 방식의 처리도 필요하지만, 관리자가 의사 결정 지원에 활용하기 위한 데이터 형태로는 적합하지 못한 단점이 있다. 따라서, 본 논문에서는 의사 결정 지원을 위해 의견 정보를 데이터 웨어하우스에 저장한 후, 의견 정보를 온라인에서 동적으로 분석하고 통합하는 처리 기법인 온라인 감성 분석 처리(On-Line Sentiment Analytical Processing, OLSAP)과 이를 위한 의견 정보의 모델링 방안을 제시한다.

## 2. 의견 데이터 모델링

상품 리뷰는 비구조적 텍스트 데이터이기

때문에 상품 리뷰가 포함하고 있는 의견 정보를 활용하기 위해서는 일차적으로 의견 데이터의 추출 및 극성 판단의 과정을 거쳐야 한다. 본 논문은 추출된 의견 데이터의 활용에 초점을 맞추고 있으므로 이 과정을 간략하게만 다루며, OLSAP에 적합한 형태의 의견 데이터 모델을 제시한다.

### 2.1 의견 데이터 추출 및 극성 판단

의견 데이터는 “이 카메라는 화질이 정말 좋아요.”처럼 상품 그 자체나 상품의 일부 특징에 대해 사용자의 의견을 언급하는 텍스트 데이터로부터 추출한다. 이에 대해 형태소 분석이나 구문분석과 같은 적절한 자연어처리를 거친 후, (카메라, 화질, 좋다, 정말) 와 같이 표준화된 형태로 의견의 단위를 추출한다. 또한 이렇게 추출된 정보에서 “좋다”가 긍정/부정인지, “정말”의 경우 의미를 강화시키는지, 약화시키는지 등을 판단하여 최종적인 극성값을 구하게 된다.

[9]의 경우, 전문가에 의해 반자동으로 사전에 구축된 어휘 사전과 극성정보를 활용하여 한국어로 작성된 상품평에서 이러한 의견 정보를 추출하는 방법을 제시하였다.

### 2.2 의견 데이터 모델링

의견 정보의 추출 대상과 활용하는 목적에 따라서 의견 데이터의 모델은 상이할 수 있으나, OLSAP에서는 다음과 같은 의견 데이터로 모델링하여 사용한다.

$$(o_i, f_j, e_k, m_l, ve_{jk}, vm_{kl}, u_m, t_n, p_o)$$

$o_i$ 는 “아이폰”과 같은 의견이 표현된 대상이며,  $f_j$ 는 “LCD”와 같이  $o_i$ 의 세부 특징을 나타내며, 값이 없을 경우 제품 그 자체를 나타낸다.  $e_k$ 는 “좋다”와 같이 각 특징

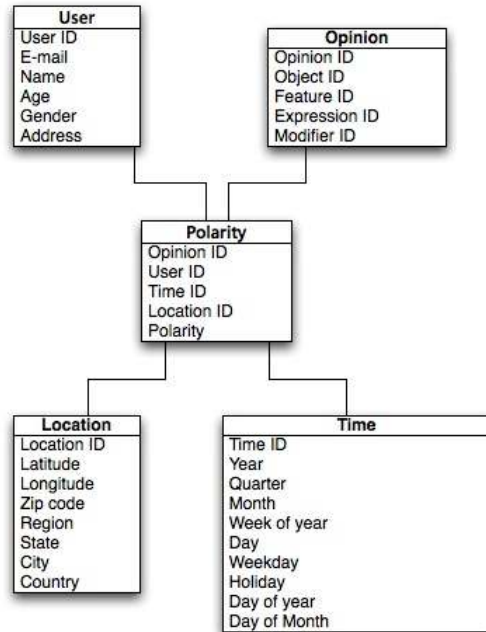
에 대한 어휘이며,  $m_l$ 는 “꽤”와 같은 의견의 강도를 나타내는 어휘이다.  $ve_{jk}$ 와  $vm_{kl}$ 는 각각 특징과 의견강도에 대한 실수 값이며 일반적으로 부정일 경우 음수, 긍정일 경우 양수이다.  $u_m$ 는 의견을 제시한 사용자,  $t_n$ 는 의견이 작성된 시각,  $p_o$ 는 의견이 작성된 위치를 나타낸다. 일반적인 구매 후기는 위치 정보가 없지만, 의견 데이터는 트위터와 같은 소셜 미디어도 나타나며, 이때 포함되는 위치정보는 데이터 분석의 중요한 요소가 된다.

### 3. 온라인 감성 분석 처리

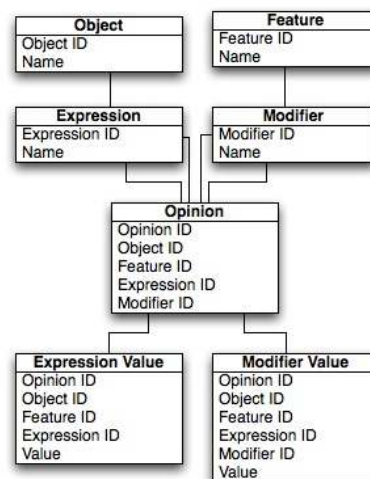
OLAP은 의사 결정 지원을 위해 데이터 웨어하우스에 기반하여 대규모의 다차원 데이터를 통합하고, 온라인에서 동적으로 분석하는 일체의 처리를 말한다. 여기에는 다차원 데이터에 대해 통계 정보, 시계열 분석, 집계 연산 등의 응용이 포함된다. OLAP은 사실 테이블(fact table)로 다차원 데이터를 구성하며, 사실 테이블은 차원을 결정하는 애트리뷰트와, 각 차원에서 측정된 값인 측정 애트리뷰트로 구성된다.

OLSAP은 OLAP에 기반하여 다양한 의견 정보를 중심으로 다차원 데이터를 구성하며, 이를 활용하여 각종 질의 및 응용을 처리한다. 이를 위해서는 우선적으로 OLSAP의 데이터베이스 스키마를 구성하는 것이 필요하다. 스키마는 도메인이나 응용에 따라 달라지며, OLAP에서 이용되는 대표적인 모델링 기법으로는 사실 테이블을 중심으로 차원 테이블이 결합된 스타 스키마(star schema)가 있다. OLSAP에 스타 스키마를 활용할 경우, 중심이 되는 사실 테이블에는 3장의 의견 데이터 모델을 사용하고, 이와 관련된 사용자, 시간, 장소, 어휘 사전 등의 정보는 차원 테이블로 구성하게 된다. 이때

사실 테이블의 측정 애트리뷰트는 의견 데이터의 극성 값이 된다. <그림 1>은 OLSAP의 스타 스키마를 나타내며, <그림 2>는 스타 스키마 내에서 의견 정보를 나타내는 Opinion 테이블의 연관 테이블들이다.



<그림 1> OLSAP의 스타 스키마



<그림 2> 의견 정보 연관 테이블

이렇게 구성된 데이터베이스를 이용하여 제품에 대한 사용자들의 전체적인 평가뿐만 아니라, 지역별 평가, 시간에 따른 의견의 변화 추세 등 복합적인 정보를 얻을 수 있게 된다.

#### 4. 결론

구매 예정자나 관리자에게 의견 데이터는 매우 유용한 정보이지만, 상대적으로 이를 효과적으로 활용하고자 하는 시도는 부족하였다. 이에 본 논문에서는 비구조적인 텍스트에서 추출한 의견 데이터를 다루기 위한 데이터 모델과, 축적된 의견 정보를 다각도로 분석하기 위한 처리기법인 OLSAP를 제안하였다. 이러한 방법으로 분석된 의견 정보는 제품의 평가뿐만 아니라, 향후 전자상거래 시스템과 결합하여 마케팅 등 다양한 목적으로 활용 할 수 있을 것으로 기대된다.

---

#### 참고문헌

---

[1] C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, H. Ng, and C. Jin, "Red Opal: Product-Feature Scoring from Reviews", ACM EC, pp.182-191, 2007.

[2] A. Esuli and F. Sebastiani, "Determining Term Subjectivity and Term Orientation for Opinion Mining", EACL, pp.193-200, 2006.

[3] N. Jindal and B. Liu, "Mining Comparative Sentences and Relations", AAAI, pp.1331-1336, 2006.

[4] E. Boiy, P. Hens, K. Deschacht, and M. Moens, "Automatic Sentiment Analysis in On-line Text", ELPUB,

pp.349-360, 2007.

- [5] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima, "Mining Product Reputations on the Web", SIGKDD, pp.341-349, 2002.
- [6] B. Liu, M. Hu, and J. Cheng, "Opinion observer: analyzing and comparing opinions on the Web", WWW, pp.342-351, 2005.
- [7] N. Godbole, M. Srinivasaiah, and S. Skiena, "Large-Scale Sentiment Analysis for News and Blogs", ICWSM, 2007.
- [8] X. Ding, B. Liu, and P. S Yu, "A holistic lexicon-based approach to opinion mining", WSDM, pp.231-240, 2008.
- [9] 명재석, 이동주, 이상구, "반자동으로 구축된 의미 사전을 이용한 한국어 상품평 분석 시스템", 정보과학회논문지 : 소프트웨어 및 응용 제 35권, 제6호, pp.392-403, 2008.