

오피니언 마이닝을 이용한 인터넷 쇼핑몰 사이트 분석

김다정*, 윤재열**, 김이준***, 김응모*

*성균관 대학교 정보통신 공학부
e-mail:djdj88zz@naver.com

Analyzing Internet shopping mall through opinion mining

Da-Jung Kim*, Jae-Yeol Yoon**, Iee-Joon Kim***, Ung-mo Kim*
*School of Information and Communication Engineering
Sungkyunkwan University

요 약

전 세계적으로 인터넷 보급률이 높아지고 전자상거래가 활발해지면서 직접 발품을 팔아가며 보다 비싼 가격에 물건을 구입하는 일보다는 인터넷몰을 이용하여 제품을 구매하는 일이 늘어나고 있다. 특히 인터넷몰 중에서도 의류를 판매하는 인터넷 쇼핑몰은 새로운 산업으로 각광받으며 하루가 다르게 새로운 사이트가 우후죽순 생겨나고 있다. 자리에 앉아 직원의 직접적인 홍보에 대한 부담을 갖지 않고 차별화 제품을 싼 가격에 구입할 수 있다는 이점을 갖고 있는 인터넷 쇼핑몰이지만 직접 제품을 눈으로 확인하지 못하기 때문에 실제 제품과 동일한 제품인지 조작된 사진인지의 여부를 확인하기 힘들고 직원과의 상담 역시 전화나 인터넷을 통해 진행되기 때문에 환불이나 교환과 같은 의사소통에 문제가 생기고는 한다. 이에 본 논문에서는 오피니언 마이닝을 통해 웹 서버에 저장되어 있는 수많은 쇼핑몰의 리뷰를 정리하고 각 사이트의 서비스, 품질 등에 대한 평가를 카테고리 별로 분석하여 소비자가 현명하고 효율적인 소비 결정을 내릴 수 있도록 도움 방법을 제안하고자 한다.

1. 서론

최근 몇 년간 보다 싼 가격에 앉은 자리에서 편리하게 쇼핑을 할 수 있는 이점이 부각 되면서 많은 인터넷 쇼핑몰(Internet shopping mall) 사이트가 우후죽순 생겨나고 있다.

세계적 조사 통계업체인 입소스-라이드(Ipsos-Reid)에서 2002년 세계 12개국의 성인 6600명을 대상으로 한 인터넷 이용률 조사 결과 인터넷 접속목적 중 전자상거래(66%)가 사진, 동영상 송수신(68%)에 이은 두 번째로 나타나 우리 생활 속에 인터넷 쇼핑이 새로운 소비 패턴으로 자리매김하였음을 알 수 있다. [1]

우리 생활 속에서 빠질 수 없는 온라인 쇼핑 업계는 2010년에는 44조원이라는 어마어마한 시장규모를 형성할 것이라는 예측이 나오고 있다. [2]

이렇게 많은 사람들이 인터넷 쇼핑 물 서비스를 이용하지만 쇼핑 물을 이용한 모두가 서비스에 만족을 하는 것은 아니다. 온라인으로 거래가 되는 인터넷 쇼핑 물의 특성상 환불과 교환이 어렵고 실제 제품과 동일한 제품인지 조작된 사진지의 여부를 확인하는 것 또한 힘들다. 직원과의 상담 역시 전화나 인터넷을 통해 진행되기 때문에 환불이나 교환과 같은 의사소통에 문제가 생길 수 있다.

이러한 문제점을 극복하기 위해 대다수의 사람들은 보다 성공적인 인터넷 쇼핑을 위해 해당 인터넷 쇼핑 사이트 이용자들의 평가와 판매되는 제품의 질, 직원들의 친절도와 같은 서비스에 관한 정보를 수집하여 평가한 후 구매 결정 내린다.

이와 같은 정보는 높은 인터넷 보급률 덕분에 다양한 사람들이 자신의 블로그나 정보 공유 카페와 같은 온라인 커뮤니티 또는 게시판에 서비스나 제품의 후기를 공유하기에 쉽게 찾아 볼 수 있다. 기존의 이용자나 새로운 이용자들은 이러한 정보를 검색하고 수집하여 구매 결정에 참고하고자 할 것 이다.

하지만 웹상에 존재하는 수많은 후기를 모두 검색하고 분석한 후 구매를 결정하는 데에는 많은 시간과 노력이 필요할 것이다.

이에 본 논문에서는 인터넷 쇼핑 물 서비스를 통해 성공적인 쇼핑을 하고자 하는 사람들을 위해 인터넷 상에 준비해 있는 후기와 평가들을 오피니언 마이닝을 통해 분석하고 수치화하여 소비자의 구매결정을 도울 수 있는 방법을 제안하고자 한다.

2. 관련 연구

2.1 오피니언 마이닝

오피니언 마이닝(Opinion Mining) 또는 감정 분석(Sentiment Analysis)은 텍스트 마이닝의 일종으로 SNS, 블로그, 카페 등과 같이 인터넷에 산재한 모든 웹문서, 댓글 등에서 제품에 대한 평가나 의견 등을 참고하고 분석하여 해당 평가나 의견에서 의미 있는 정보를 찾아내는 방법이다.

이러한 오피니언 마이닝은 그 특징에 따라 Sentiment classification과 Feature-based opinion mining and summarization 그리고 Comparative sentence and relation mining의 세 가지로 구분 될 수 있다.

Sentiment classification은 Document Level의 분석법으로 텍스트 단위로 긍정적인지 부정적인지를 분석하는 기법으로, 예를 들어 특정 의견이 가지고 있는 성향이 전체적으로 긍정을 의미하는지 부정을 의미하는지 분석하는 기법이다.

‘이 제품은 가격은 비싸지만 성능도 좋고 디자인도 예뻐서 좋아요.’

와 같은 의견이 있을 때 Sentiment classification 기법으로 전체적인 의견이 긍정적임을 파악 할 수 있지만 ‘가격’, ‘성능’ 그리고 ‘디자인’에 관한 세부적인 항목에 관한 극성을 파악 할 수 없다.

두 번째로 언급된 Feature-based opinion mining and summarization은 sentence level의 분석법으로 하나의 텍스트에 대한 세부 속성이 긍정적인지 부정적인지를 분석하는 기법이다.

마지막 Comparative sentence and relation mining은 Feature-based opinion mining and summarization과 마찬가지로 sentence level의 분석법으로 두 가지 대상을 비교하여 상대적인 긍정, 부정을 분석하는 기법이다.

본 논문에서는 해당 사이트에 대한 세부적인 특성을 분석하고자 하므로 Feature-based opinion mining and summarization을 고려하고 있다.[3-6]

2.1.1. Feature-based opinion mining and summarization

이 디지털 카메라의 해상도는 굉장히 높고, 배터리도 오래가네요. 하지만 사이즈가 휴대하기에는 조금 크고 무게도 가볍지 않아요. 전반적으로 좋은 카메라라 생각합니다.

(그림 1)

특정 제품의 후기에는 그 제품의 여러 특성에 따른 평가가 존재하기 마련이다. 예를 들어 (그림 1)과 같은 후기에서 ‘해상도’, ‘배터리’, ‘사이즈’, ‘무게’라는 속성을 발견할 수 있다. 위 후기의 작성자는 Document level에서 전반적

으로 긍정적인 의견을 가지고 있지만 Sentence level에서 ‘화소’와 ‘배터리’에는 긍정을, ‘사이즈’와 ‘무게’에서는 부정을 표하고 있다. 이렇듯 제품의 세밀한 분석을 위해서는 Document level에서가 아닌 Sentence level에서의 분석이 필요하다.

2.2. 형태소 분석

인터넷에 산재한 글을 올바르게 분석하기 위해서는 문장을 형태소 단위로 분석할 필요가 있다. 형태소 분석이란 자연 언어 분석의 첫 단계로 단어를 구성하는 각각의 형태소를 인식하고 불규칙 활용이나 축약, 탈락 현상이 일어난 경우 원형을 복원하는 과정이다.

예를 들어 “이 카메라는 정말 커요.”라는 문장이 있을 때, 이 문장을 “이/카메라/는/정/말/크/어/요”와 같이 형태소 분석이 가능하다. 이렇게 분석된 문장은 오피니언 마이닝을 이용하여 극성 분석한다. [7]

2.3. 연관규칙 마이닝

제품과 사이트 후기에는 이모티콘과 같이 평가요소와 관련 없는 단어가 사용될 수 있다. 만약 특정 제품과 연관성을 가지고 있는 항목이라면 이는 빈번하게 나타날 태고 연관 규칙 마이닝을 이용하여 제품의 속성을 추출할 수 있다.

연관 규칙 마이닝은 트랜잭션 데이터베이스, 관계형 데이터베이스 그리고 다른 정보 저장소에서 목적물이나 항목의 집합 가운데 빈번하게 발생하는 패턴, 연관, 상호관계 또는 친화도를 찾아내는 규칙이다.

연관 규칙 마이닝은 Support와 Confidence를 이용한 항목 집합 분석이 가능하다. Support는 전체 집합 중 A와 B가 동시에 일어날 경우의 확률이고 Confidence는 A가 일어났을 때 B가 일어날 확률의 집합이다. 트랜잭션에 항목 A가 존재하면서 항목 B가 같이 존재하는 패턴이 있을 때, A와 B의 연관성을 A=>B로 표현 할 수 있다.

예를 들어 어느 할인매장에서 귀저기를 구입하는 남성 고객은 맥주도 같이 구매하는 패턴이 빈번하게 발생한다고 하면 귀저기=>맥주 [support=4%, confidence=77%]와 같이 표현 가능하다.[8]

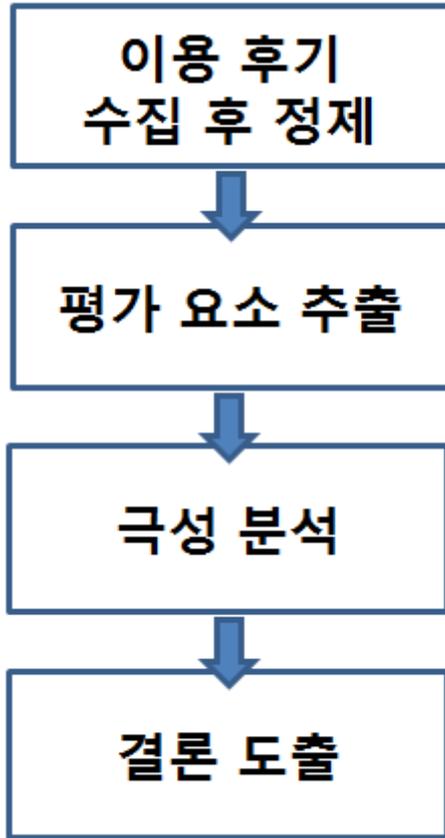
3. 오피니언 마이닝을 이용한 제품과 서비스 후기 분석

3.1. 제안 배경

현존하는 후기와 리뷰들은 다양한 곳에 널리 산재해 있어 많은 정보를 신속하고 정확하게 분석해 내기 힘들었

다. 그러나 오피니언 마이닝을 이용하여 해당 사이트에 대한 평가를 분석하면 더 빠르고 정확한 정보를 얻을 수 있어 해당 쇼핑 물의 신뢰성을 높일 수 있다.

3.2. 제안 내용



(그림 1) 시스템 흐름

3.2.1. 웹 서버에 저장된 이용후기 수집 후 정제

SNS, 블로그, 카페 등에 저장된 후기들을 수집하고 문장의 오류, 오타, 이모티콘 이나 의성어 같이 평가요소로부터 제외되는 항목을 제외한 후 형태소 분석을 통해 평가 요소를 포함하는 문장단위로 구분하여 추출한다.

(그림 2)와 같은 이용 후기가 있을 경우 “화면과 똑같다”, “색이 예쁘다”, “가격이 비싸다.” 와 같이 평가 요소를 포함하는 문장단위를 구분하여 추출한다.

실제로 받아보니 화면과 똑같아요. 색도 굉장히 예뻐요. 하지만 다른 사이트에서 조금 더 싼 가격에 팔고 있습니다.....

(그림 2)

3.2.2. 평가 요소 추출 및 극성 분석

연관 규칙 마이닝을 통하여 빈번하게 발생되어 해당 제품과 연관성이 높은 단어집합을 찾아내야 한다. 3.1에서 추출된 문장을 대상으로 ‘무게’, ‘크기’ ‘색’ 등과 같이 제품의 평가와 관련이 있는 후보명사를 추출해 낸다.

해당 후보명사들이 포함된 문장을 찾은 후 그 문장에 포함된 평가 항목을 통해 소비자들의 평가(극성)를 알 수 있다.[9]

3.2.3. 최종 결과 및 결론

3.2를 통해 극성 분석 결과를 요약하면 다음(그림 3)에서 나타내는 것과 같다. 각 제품 별로 극성 평가를 한 후 해당 제품의 평가 비율을 보여주며 그에 해당하는 점수를 나타낸다.

사이트 A	극성 평가	점수
제품 가격	나쁨 ————— 좋음	-36
품질	나쁨 ————— 좋음	+39
서비스	나쁨 ————— 좋음	+15
...

(그림 3)

최종결과와 같은 자료를 참고하여 쇼핑을 한다면 더 이상 웹에 산재해 있는 많은 후기들을 전부 읽어보고 확인하지 않아도 각 쇼핑물의 특성과 소비자, 사용자들의 해당 쇼핑물에 대한 전반적인 만족도를 알 수 있을 것이다.

악의가 있는 후기나 신뢰도가 높지 않은 후기를 극성 평가에 반영한다면 그 결과 역시 신뢰도가 높지 않으므로 평가 대상의 신뢰성을 확인하는 것 역시 중요하다.

이러한 점을 고려하여 자료를 선택하고 분석한 후 나온 결과를 이용하여 수많은 쇼핑물을 리뷰하고 실제 쇼핑 과정에 적극 반영하여 불합리한 지출과 불필요한 소비를 줄이는데 도움이 될 수 있을 것으로 기대된다.

참고문헌

[1] Sun-Ja Kye, In-Ok Kim “Usefulness of the 20’s and 30’s internet searching and consumer satisfaction for Internet shopping” 한국가족자원경영학회지, 9권 2호, 2005.
 [2] ‘2010년 유통산업 전망’, 대한상의보고서
 [3] Su-Min Kim, Eduard Hovy, “Determining the Sentiment of Opinions”
 [4] Namrata Godbole, Manjunath Srinivasaiah, StevenSkiena, “Large-Scale Sentiment Analysis for News and Blogs,” Int’l AAAI Conference on Weblogs

and Social Media (ICWSM 2007), 2007.

[5] E. Boiy, P. Hens, K. Deschacht, M. Moens, "Automatic Sentiment Analysis in On-line Text," ELPUB2007 Conference on Electronic Publishing, June 2007.

[6] J. Yi, W. Niblack, "Sentiment Mining in Web-Fountain," International Conference on Data Engineering (ICDE'05), pp.1073-1083, 2005.

[7] 강승식 지음, "한국어 형태소 분석과 정보검색", 홍릉과학출판사, 2002.

[8] InfoLab, in Stanford, <http://infolab.stanford.edu/~ullman/mining/assocrules.pdf>

[9] 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.