

# 오피니언 마이닝을 통한 스마트폰 어플리케이션 이용 후기 분석

유하나\*, 윤재열\*\*, 김응모  
\*성균관 대학교 정보통신 공학부  
e-mail : hanayoo89@gmail.com

## Analyzing review of the smart phone application through opinion mining

\*Ha-Na Yoo, Jae-Yeol Yoon, Ung-mo Kim  
\*School of Information and Communication Engineering  
Sungkyunkwan University

### 요 약

스마트폰 시장이 커지면서, 사람들이 하루에 업로드하고 다운로드하는 어플리케이션의 수 또한 급격히 증가하고 있다. 앱스토어와 안드로이드마켓에 등록된 어플리케이션의 종류는 어마어마하며, 사람들은 자신의 생활을 편리하게 해줄 어플리케이션 혹은 재미를 위한 어플리케이션을 다운로드하고자 한다. 하지만 현재 어플리케이션에 대한 평가는 점수로만 이루어져있기 때문에 어느 부분에서 뛰어난지, 어떤 부분의 기능이 떨어지는지는 사용자가 알 수 없고, 특정 기능을 중요시하는 사용자일 경우 별점이 높아도 해당기능이 만족스럽지 않으면 만족감의 정도는 대단히 떨어지게 된다. 그러면 다른 어플리케이션을 받아 같은 작업을 반복해야하는데, 이 경우가 반복될 경우 비용적인 문제뿐만 아니라 사용자에게 매우 번거로운 일이다. 따라서 본 논문에서는 기존 사용자들이 자신이 사용한 어플리케이션에 대해 작성한 후기를 오피니언 마이닝 기술을 적용시켜 각 키워드별, 즉 속성별로 평가하고 긍정/부정 여부를 데이터베이스에 저장하여, 해당 어플리케이션을 검색한 미래의 어플리케이션 사용자에게 시각적으로 정보를 알려주어 사용자의 수고를 덜어주고자 한다. 어플리케이션 다운로드가 매우 단순한 작업이지만, 다운로드 수가 많기 때문에 본 논문의 제안을 적용한다면 비용을 절감시켜 줄 뿐만 아니라 매우 효율적인 작업이 될 것이라 기대한다.

### 1. 서론

최근 스마트폰을 사용하는 사람의 수가 증가하면서, 마켓에 배포되는 어플리케이션의 수 또한 증가하고 있다. 스마트폰의 어플리케이션을 통해서 무료전화나 문자를 할 수 있을 뿐만 아니라 게임, SNS, 금융, 도서, 교육, 건강, 의학, 음악, 뉴스, 내비게이션, 날씨 등 다양한 항목들로 사람들의 실생활에 사용되고 있고, 매우 유용하다. 기존 휴대폰처럼 전문적인 개발자들만 어플리케이션을 개발하는 것이 아니라 개발 툴만 있다면 일반 사용자들도 개발이 가능하기 때문에 어플리케이션 시장은 매우 커지고 있으며, 2011년 앱스토어에 등록된 어플리케이션의 경우 50만개를 돌파했다.

스마트폰 유저가 증가하면서 다운로드 수 또한 급속도로 늘고 있는데 안드로이드 어플리케이션 다운로드 수는 2008년 10월 안드로이드마켓이 열린 이후 2년 6개월 만인 2011년 3월 30억 개를 돌파하였고, 아이폰의 경우 마켓 개장 2년 후인 2011년 1월 100억 개를 돌파하였다. 앱스토어의 일 다운로드 수는 3000만건에 달하며 이는 우리나라 성인 인구 수와 비슷한 수치이다.

이것은 어플리케이션 시장이 커졌다는 의미일 수도 있

지만 그만큼 다운로드를 했다가 지우는 어플리케이션의 수 또한 크다는 의미이기도 하다. 좋은 어플리케이션도 많지만 상업적이고 자극적이며, 필수적이고 유용한 기능은 없는 어플리케이션도 상당수가 있으며 사용자에게 지불한 비용에 대비해 만족도가 매우 떨어질 뿐만 아니라, 무료 어플리케이션의 경우에도 설치했다가 지워버리는 경우가 더 많다. 이는 비용, 시간 측면에서 매우 낭비이다.

우리나라의 경우 인터넷에서 블로그나 카페를 통해서 여러 가지 주제들에 대한 의견 공유가 활발한 편이다. 영화, 상품, 맛집, 레시피 등 각자의 취미 생활과 상품에 대한 호불호를 사람들과 공유하고 있으며, 어플리케이션에 대한 평가도 마찬가지이다. 사람들이 자신이 다운로드한 어플리케이션 후기를 통해서 항목별 만족도를 분석하고 성능이 좋은 항목과 나쁜 항목을 사람들에게 보여준다면, 이후의 이용자들은 자신이 다운로드하려는 어플리케이션이 자신의 필요와 맞는지 알뿐만 아니라, 엉뚱한 어플리케이션을 다운로드하는 수고를 덜어주어 시간과 비용을 덜어줄 것이다.

이에 본 논문은 기존 사용자가 작성한 어플리케이션 이용후기에 오피니언 마이닝을 적용, 항목별로 분석하여 키워드를 추출하고, 또한 그 것의 긍정/부정 여부를 색상

으로 구분하여 어플리케이션을 다운로드 하려는 사용자가 어플리케이션의 항목별 특성을 한눈에 파악할 수 있는 시스템을 제안하고자 한다.

## 2. 관련 연구

### 2.1 오피니언 마이닝

최근에는 오피니언 마이닝에 대한 연구가 활발히 진행되고 있으며, 오피니언 마이닝은 웹 문서에서 사용자들의 의견, 평가를 분석하여 주제에 대한 평판을 도출하는 기술이다. 여러 웹사이트에 산재되어 있는 의견들을 일일이 찾아야 하는 수고를 덜어주기 위해 다양한 웹 사이트들에 기록된 사용자들의 의견 정보들을 자동 추출/ 분석하여 검색하고 통계를 내는 연구[1]와 검색의 정확률을 위해 추천문장이 담긴 추천문서를 분류하는 방법[2], 사용자의 검색 의도에 따른 상품평 검색을 위해 상품평에 우선순위를 부여하여 능률적인 의견 탐색이 가능해지도록 하는 방법 등의 연구[3]가 진행되고 있다.

상품평의 요약을 위한 방법으로는 크게 자연어 처리 기법과 통계학적 접근법으로 나뉘지는데 자연어 처리 기법의 경우 정확한 의견을 식별해 내어 원래의 의미를 손상시키지 않는 장점을 갖지만 자연어를 처리하는 데에는 많은 시간과 전문가의 노력이 필요하기 때문에 통계학적 접근법이 많이 쓰이고 있는 추세이다. 문장 구조와 문장 사이의 관계, 패턴 정보를 추출하여 통계를 내어 자동으로 사전을 구축하는 기법[4], 단어 사이의 관계를 추측하는 방법인 PMI를 사용하여 긍정/부정의 극성을 판별하는 기법[5], 기계학습을 사용하는 기법[6,7] 등이 연구되고 있다.

#### 2.1.1 Evaluative Texts의 분류 방법

Sentiment classification은 문서에서 특정 객체에 대한 전반적인 의견이나 개인적인 주관이나 어떠한가에 대한 분류를 하는 방법이다. 즉 문서가 의견을 지닌 문서인지를 판별하는 과정으로 전체적인 긍정/부정성을 판별할 수 있다. Comparative sentence and relation mining은 다른 객체의 특정한 특징에 비해 해당 객체가 상대적으로 어떠한지에 대해 분석하는 방법으로 문장단위로 이루어진다. Featured-based opinion mining and summarization은 제품, 서비스, 기관 등의 객체에 대한 특징을 추출하여 긍정/부정을 판별하는 방법으로 문장단위에서 이루어진다. 각 속성에 대해 같은 단어도 긍정/부정의 의미가 다르기 때문에 속성에 대한 극성 판단 또한 분류에 있어서 중요하다.

#### 2.1.2 PMI(Pointwise mutual information)

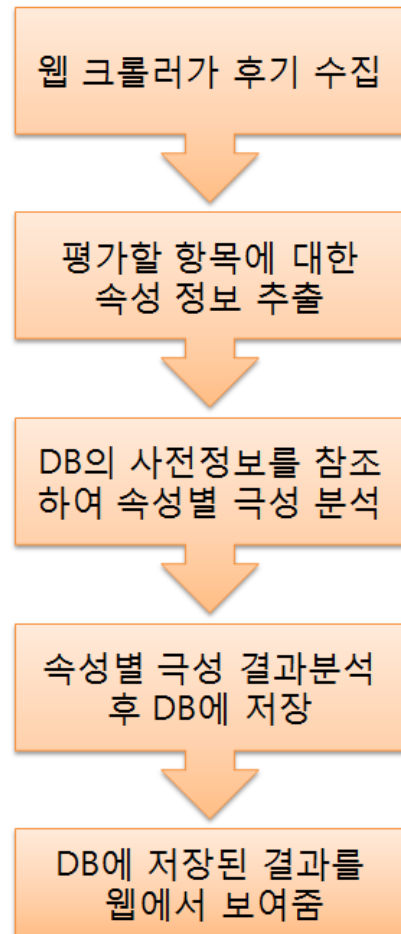
PMI는 term과 term 사이의 co-occurrence relation 정보를 통하여 두 용어 사이의 관계가 얼마나 종속적인지를 판단하는 것이다. 문맥에 관계하지 않고 보편적으로 긍정적인 단어들의 집합과 부정적인 단어들의 집합으로 나누

어 PMI를 계산한다. [5]에서는 가까운 거리에 쓰여진 단어가 연관성이 있을 것이라 생각하여, 가까운 거리에 쓰여진 웹페이지를 검색하여 PMI를 계산한 후 긍정 PMI에서 부정 PMI의 차를 구하여 극성을 판별하는 아이디어를 제시했다. PMI의 값이 0일 경우, 두 단어의 관계는 독립적이지만, 0이 아닐 경우 차가 연관성의 정도를 나타내게 된다.

본 논문에서는 속성별로 PMI를 계산하여 극성을 판별하여 어플리케이션의 속성별 평가를 보여주려고 한다.

## 3. 제안 시스템

스마트폰 사용자가 다운로드하려는 어플리케이션의 이름을 검색하면 웹 크롤링을 통해 해당 주소에서 상품에 대한 의견 정보를 가져오게 되고, 데이터베이스의 사전정보를 참조하여 사람들의 의견을 분석하고 평가하여, 이 결과를 다시 데이터베이스에 저장하여 웹에서 어플리케이션에 대한 전체적인 사람들의 의견을 보여주게 된다. 시스템의 전체 흐름은 다음의 (그림 1)과 같다.



(그림 1) 시스템의 흐름

### 3.1 웹 크롤러가 어플리케이션 후기를 수집

웹 크롤러는 많은 웹페이지에서 수집한 다양한 정보를 검색이 쉬운 형태로 인덱싱하여 서버에 보존하는데, 사용자가 어플리케이션의 이름을 검색엔진에서 검색하면 서버

안의 정보를 꺼내어 검색결과를 제공해준다. 이렇게 웹문서로부터 어플리케이션에 대한 다양한 의견 정보들을 수집해오면, 텍스트 항목만을 추출해낸다. 텍스트 중에서도 모든 문장이 어플리케이션 이용 후기에 대한 내용이 아니기 때문에, 속성이 포함된 텍스트를 문장단위로 추출한다. 영어 학습과 관련된 어플리케이션을 검색하여 평가와 관련된 텍스트를 추출하면 아래와 같은 글이 나오게 된다.

어플리케이션을 사용하는 **활용법이 매우 쉽고 편해요. 다양한 기능을 제공해주어 영어 공부를 쉽게 할 수 있어요. 교육적일 뿐만 아니라 오락성도 높아 꾸준한 재미를 준다**는 점에서 굉장히 훌륭한 어플리케이션이라고 생각합니다. 제공해주는 기능에 비한다면 **가격도 정말 저렴해요. 강력하게 추천합니다.**

위의 텍스트에서 “활용법이 굉장히 쉽고 편하다.”, “다양한 기능을 제공해준다.”, “오락성이 높아 꾸준한 재미를 준다.”, 가격이 정말 저렴하다.” 라는 속성이 포함된 문장으로 추출한다.

### 3.2 평가할 항목에 대한 속성 정보 추출

[8]에서는 상품평 분석을 위해 속성 문장을 정의하였는데, 본 논문에서는 [8]에서 정의한 개념을 사용한다. 속성 문장은 속성 어휘와 감성 어휘로 구성되어 있다. 한 문장에 여러 개의 속성이 있을 수 있어 여러 개의 속성문장으로 표현할 수 있으며, 한 속성 당 다수의 감성 어휘들이 나타날 수 있다. 위의 텍스트를 예를 들어보면 “활용법이 매우 쉽고 편해요.”에서는 활용법이라는 속성은 “쉽다, 편하다”라는 두 개의 감성 어휘로 표현되어 있는 것을 알 수 있다.

위의 텍스트에서 속성모델의 정의대로 속성어휘만 분리하면, 활용법, 기능, 오락성, 가격이 추출된다.

### 3.3 DB의 사전정보를 참조하여 속성별 극성 분석

문장의 극성을 판별하기 위해서는 단어의 의미를 나타내는 DB 사전이 구축되어 있어야 한다. 의견의 극성을 분석하는 데에 있어서, 그 정확성은 DB 사전이 얼마나 잘 정의되어 있는가에 달려있는데, 본 논문에서는 정확도를 위해 시스템이 기계학습을 통해 분류를 하면 시스템 관리자가 정확도 있게 재정립하여 만드는 DB사전을 사용한다. 구축된 사전정보를 참조하여 PMI의 개념을 적용시켜 속성별 극성 분석을 시행한다.

PMI(속성, 감성어휘)값을 계산하는데 긍정 PMI값과 부정 PMI값의 차이를 구하여 결과 값이 양수일 경우 해당 속성에 있어서 긍정적인 평가, 음수일 경우 부정적인 평가가 대체적으로 이루어진 것으로 판단할 수 있다.

속성	감성 어휘	PMI(속성, 감성어휘)
활용법	매우 쉽다, 편하다	+
기능	다양하다	+
오락성	높다, 재미를 준다.	+
가격	저렴하다	+

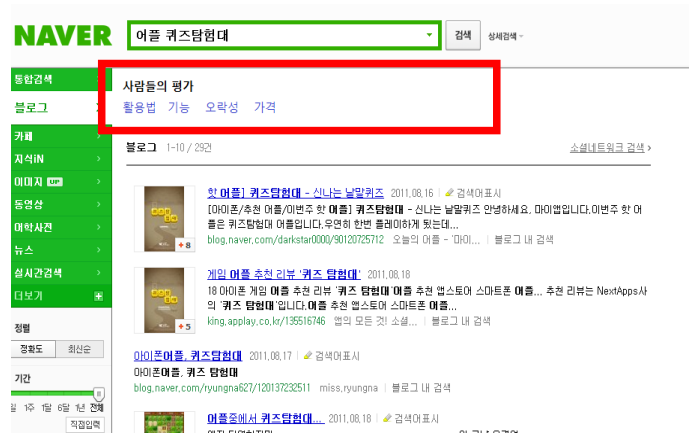
(그림 2) 각 속성에 대한 PMI값

### 3.4 속성별 극성결과 분석 후 DB에 저장

각 속성에 대한 PMI값이 (그림 2)와 같이 나오면 이 결과를 DB에 저장한다. 본 논문에서는 속성별 점수의 극성정도를 보여주기보다 속성에 대한 평가가 긍정인지 부정인지를 나타내는 게 목적이므로 PMI값의 결과가 긍정이면 1, 부정이면 0으로 속성별로 DB에 저장한다.

### 3.5 DB에 저장된 결과를 웹에서 보여줌

3.4 에 저장된 DB에서 값을 가져와 그 결과를 웹에서 사용자에게 보여준다. 사용자가 관심이 있어 하는 어플리케이션을 웹에서 검색하면 해당 어플리케이션에 대한 속성을 단어로 보여준다. 이 때 긍정일 경우 글자의 색을 파란색, 부정일 경우 빨간색으로 나타내어 간단하게 해당 속성이 어떠한 평가를 받고 있는지 보여줄 수 있다. 검색 후 상단에 간단하게 항목별로 어떤 평가가 이루어지고 있는지 보여준 후, 검색 결과를 나타낸다면 사용자들은 각각의 블로그나 카페를 직접 들어가 보지 않아도 사람들의 대체적인 평을 알 수 있고 더 자세한 결과를 보고 싶을 경우 검색결과를 이용하면 될 것이다. 그 결과는 (그림 3)과 같다.



(그림 3) 결과 화면

## 4. 결론

스마트폰 시장이 커지면서 어플리케이션 시장 또한 매우 커지고 있다. 시장에 나와 있는 어플리케이션의 수는 어마어마하며 다양한 기능들을 가지고 있지만 각각의 어플리케이션이 서로 다른 품질을 가지고 있기에 사람들의 평을 보고 다운로드 하는 것이 시간과 돈의 비용을 절감

시켜준다. 하지만 모든 평을 단시간에 볼 수 없으므로 오피니언 마이닝을 통해 의견분석을 하고 검색 후에 속성별 평가를 색깔로 알려준다면 사용자들은 많은 시간을 아낄 수 있을 것이다.

### 참고문헌

- [1] 양정연, 명재석, 이상구, "상품 리뷰 요약에서의 문맥 정보를 이용한 의견 분류 방법", 정보과학회논문지 : 데이터베이스 제 36 권 제 4 호 (2009.8)
- [2] 김문현, 구영현, 유성준, "추천키워드 및 기계학습을 이용한 문서 분류", 정보과학회논문지 : 소프트웨어 및 응용 제38권 제1호(2011.1)
- [3] 윤홍준, 김한준, 장재영, "오피니언 마이닝 기술을 이용한 효율적 상품평 검색 기법", 2009 한국컴퓨터종합학술대회 논문집 Vol.36, No.1(A)
- [4] 명재석, 이동주, 이상구, "반자동으로 구축된 의미 사전을 이용한 한국어 상품평 분석 시스템", 정보과학회논문지 : 소프트웨어 및 응용 제 35 권 제 6 호(2008.6)
- [5] 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.
- [6] T. Mitchell, Machine learning, McGraw-Hill, 1977.
- [7] T. Joachims, Making large-scale SVM learning practical. Advances in Kernel Methods-Support Vector Learning, B.Schlkopf and C. Budes and A. Smola(ed.), 1999.
- [8] 장재영, "온라인 쇼핑몰의 상품평 자동분류를 위한 감성분석 알고리즘", 한국전자거래학회지 제14권 제4호, 2009.11