제36회 한국정보처리학회 추계학술발표대회 논문집 제18권 제2호 (2011. 11)

# 오피니언 마이닝을 이용한 친구 추천 시스템

황수진*, 윤재열*, 김이준*, 김응모*
*성균관대학교 정보통신공학부
e-mail:sj1228@skku.edu

# Friend Recommendation System Using Opinion Mining

Su-Jin Hwang*, Jae-Yeol Yoon*, Iee-Joon Kim*, Ung-Mo Kim*
*School of Information Communication Engineering, Sungkyunkwan Univ

요 약

오피니언 마이닝은 웹에 있는 문서를 분석하여 작성자의 의견을 요약된 형태로 보여주는 기술이다. 오피니언 마이닝을 이용해 문서 작성자의 주관적 의견을 알 수 있고 이를 통해 작성자의 성향이나 관심사와 같은 정보를 얻을 수 있다. 많은 네티즌들은 소셜 네트워크 서비스를 통해 자신의 의견이 담긴 글을 타인과 공유 하며 네트워크상의 인맥을 넓혀 나간다. 오피니언 마이닝을 통해 개인이 작성한 글들을 분석하여 관심사를 파악하고 비슷한 관심사를 가진 친구를 추천하는 친구 추천 시스템을 제안 한다.

## 1. 서론

인터넷 상에서 블로그, 미니홈피(싸이월드), 마이크로블로그(트위터), 프로필페이지(페이스북)와 같은 소셜 네트워크 서비스(Social Network Service; SNS)의 이용률이 증가하면서 인터넷은 개인의 의견을 표현하는 대표적인 공간이 되었다. 2010년 7월 기준으로 전 세계 인터넷 이용자의 약 71%, 국내에선 65.7%가 SNS를 이용 하는 것으로 나타났다.[1] SNS의 이용자는 불특정다수의 글에 자유롭게 접근할 수 있고 직접 작성한 글도 수많은 네티즌들에게 자유롭게 공개할 수 있기 때문에 타인과 정보를 공유 하고 인맥을 넓히기 위한 목적으로 이용 된다.

상당수 SNS가 친구를 팔로우 혹은 일촌 등록과 같은 기능을 통해 이용자끼리 직접 커뮤니케이션을 취하거나 빠른 교류를 할 수 있도록 하는 서비스를 제공 한다. 그러나 수많은 이용자들 중에 자신에게 알맞은 친구를 일일이 수동으로 찾기란 불가능 하다. 따라서 많은 SNS가 이용자에게 어울리는 타인을 추천해주는 친구 추천 기능을 제공 한다. 예로 싸이월드나 페이스북과 같은 SNS는 등록된 친구의 친구나 이용자의 프로필과 공통된 프로필을 가진 다른 회원을 친구로 추천해주는 기능을 제공 한다. 이러한 친구 추천 기능은 넷 상에서 타인과의 관계를 넓혀가려는 네티즌에게 유용하게 활용 될 수 있지만 자신의 개인 프로필을 웹상에 공개하길 거부하는 네티즌들에게는 불편한 서비스가 될 수 있다. 때문에 SNS 이용자가 직접 작성한 글을 분석하여 친구를 추천해주는 오피니언 마이닝을 이용한 친구 추천 기능을 제안하고자 한다.

오피니언 마이닝은 웹 문서를 분석하여 작성자의 의견을 파악 하는 기술이다. 분석하고자 하는 대상과 감정어를 함께 추출 하여 대상의 이미지가 긍정적인지 부정적인지를 분석 한다.[2] 이를 통해 웹에 문서를 작성한 개인이 문서에서 언급한 대상에 대해 어떠한 의견을 지니고 있는지 분석할 수 있다. 개인의 의견 정보를 이용하여 웹 문서의 작성자가 생각하는 긍정적 이미지를 가진 대상과 부정적 이미지를 가진 대상을 파악하고 비슷한 성향을 가진 또 다른 인터넷 이용자를 친구로 추천할 수 있다.

논문의 구성은 다음과 같다. 2장에서는 오피니언 마이닝과 관련된 연구를 소개하고 3장에서는 오피니언 마이닝을 이용한 친구추천 기능을 제안 한다. 4장에서는 제안한 기능을 평가하고 향후 연구에 대한 소개를 하면서 결론을 맺는다.

## 2. 관련 연구

### 2.1. 오피니언 마이닝 (Opinion Mining)

오피니언 마이닝은 텍스트 마이닝의 한 분야로서 주어진 텍스트의 주제가 아닌 주제에 대해서 작성자가 가지는 의견을 파악한다. 다른 말로 Sentiment Analysis나 Sentiment Classification이라 하며 비즈니스 인텔리전스(business intelligence)나 추천 시스템 등에 적용 가능 하다.[3] 또한 온라인에 있는 상당한 수의 문서들을 요약하는데 적합한 기술이며 90년대 말부터 학문적, 상업적인 방향으로 오피니언 마이닝의 연구와 적용은 꾸준히 진행되어왔다.[4] 특히 영화리뷰나, 상품평과 같은 문서에 오피니언 마이닝을 적용하는 연구가 활발하다.[5-8]

오피니언 마이닝은 세 가지 세부 작업인 주관성 분석, 극성 분석, 극성 강도 계산으로 이루어진다. 먼저 주어진 텍스트가 객관적인지 혹은 주관적인지를 판단하는 주관성 분석이 이루어진다. 다음으로 주관적인 텍스트의 내용이 긍정인지 부정인지를 판단하는 극성(polarity) 분석을 한

다. 마지막으로 긍정 혹은 부정으로 판단된 텍스트의 극성 강도를 계산 한다.[4] 오피니언 마이닝을 이용하여 분석한 결과는 최종적으로 사용자가 보기 쉬운 요약된 결과로 제공 된다.

### 2.2. 극성 분석

극성을 분석하기 위해 널리 사용되는 개념 중 하나가 PMI(Point-wise Mutual Information)이다. 확률 모델을 기반으로 두 단어가 얼마나 밀접한 관계를 가지는지를 계산 한다. PMI는 다음 (1)과 같은 방법으로 구한다.

$$PMI(w_1, w_2) = \log_2 \frac{P(w_1, w_2)}{P(w_1)P(w_2)} \quad (1)$$

(1)에서 $P(w_1, w_2)$는 분석 대상이 되는 전체 문서들 중에서 두 어휘 $w_1$과 $w_2$가 동시에 나타날 확률이다. $P(w_1)$과 $P(w_2)$는 문서에서 어휘 $w_1$과 $w_2$가 나타날 각각의 확률을 말한다. PMI가 0이면 두 어휘는 전혀 관련이 없다는 것을 의미하고 양수 값은 긍정, 음수 값은 부정적인 관계를 의미 한다. 긍정적인 어휘들의 집합과 부정적인 어휘들의 집합을 미리 정의해 놓은 후 각각의 어휘 집합들과 PMI을 계산하여 두 집합의 PMI 차를 계산 하면 최종 극성을 알아낼 수 있다.[5] 그 외에 극성을 판단하는 방법에는 Score Function 계산식을 이용하거나 기계 학습 알고리즘을 적용하는 방법이 있다.[6]

### 2.3 극성 강도 계산

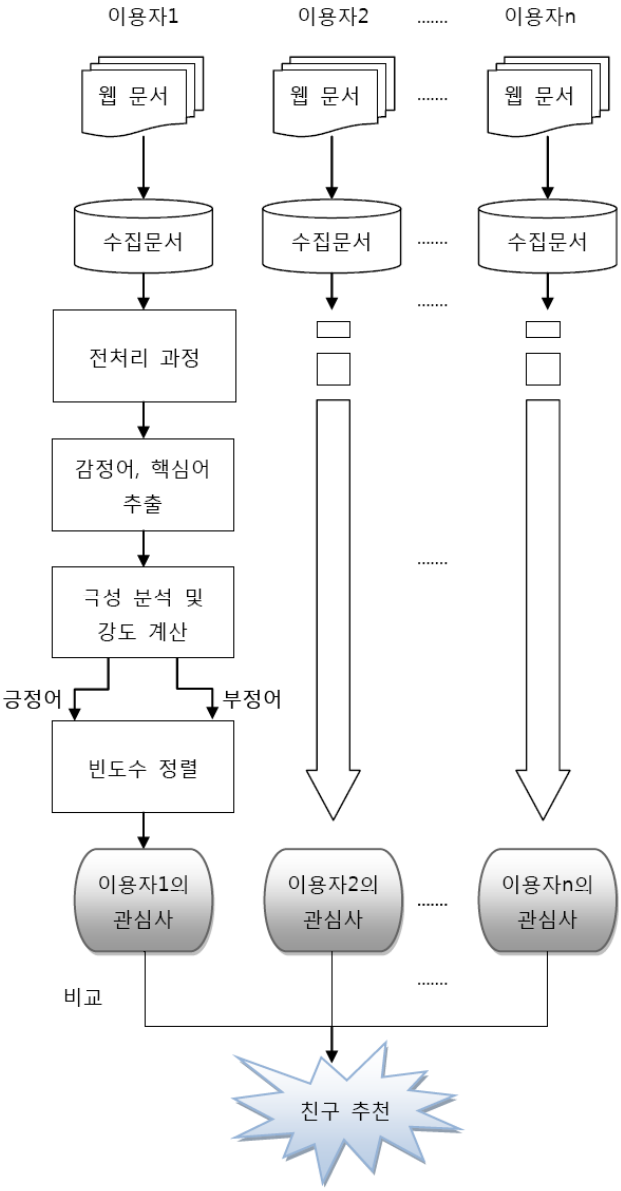
극성 강도는 사람이 직접 주관적으로 값을 부여해도 되지만 객관적 수치를 계산하기 위한 방법으로 Fei, Liu, Wu가 사용한 기계 학습 방법 (2)가 있다.

$$W_i = \begin{cases} \log(\frac{T_{p_i}}{T_{n_i}}) & T_{p_i} \neq 0 \text{ and } T_{n_i} \neq 0 \\ C + \log(\frac{T_{p_i}+1}{T_{n_i}+1}) & T_{p_i} = 0 \text{ or } T_{n_i} = 0 \end{cases} \quad (2)$$

$T_{p_i}$는 어떠한 어휘 패턴i가 긍정 문서에 나타난 빈도수, $T_{n_i}$는 부정 문서에 나타난 빈도수를 말한다. 극성강도 $W_i$는 양수일 때는 긍정, 음수일 때는 부정을 나타낸다.[7]

## 3. 친구 추천 시스템

친구 추천 시스템의 동작은 (그림 1)의 과정과 같이 이루어진다. 분석 대상은 SNS와 같은 웹 사이트에서 이용자들이 자유롭게 의견을 표현한 글들이다. 작성한 글들을 회원별로 웹 크롤링을 통해 수집한다. 인터넷상의 한국어 문서에는 신조어나 줄임말, 자음으로만 된 언어, 띄어쓰기, 철자법 오류와 같은 오류 데이터가 상당히 많다. 오류 데이터(noisy data)를 최대한 배제하기 위하여 띄어쓰기 모듈이나 맞춤법 검사기를 통해 수집한 문서의 전처리 과정을 거친다. 자음으로만 이루어진 언어나 중복된 문장부호는 삭제하거나 하나의 문장부호로 대체하여 처리 한다. 또한 한글파괴 현상이 일어난 언어들은 올바른 표준어로 대체하도록 한다.[8] 전처리 과정은 시스템이 도출할 결과의 정확도를 높이는 역할을 한다.



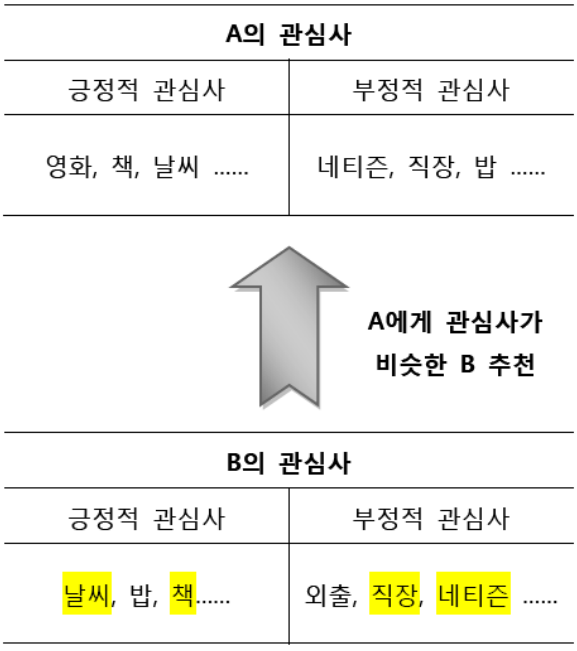

(그림 1) 친구 추천 시스템의 동작 과정

전처리 과정을 거친 문서에 있는 각 문장들은 형태소 분석과 품사 태깅을 거치고 문장에서 의견을 나타내는 언어인 감정어와 감정어의 대상이 되는 언어를 함께 추출한다. 이때 감정어의 대상이 되는 언어를 핵심어라 정의 한다. 감정어를 찾는 방법에는 수작업 분석을 통해 찾는 방법, 영어 시소러스를 기반으로 찾는 방법, 한국어 구문 분석기를 기반으로 감정어를 찾는 방법이 있다.[6]

선정한 핵심어의 극성을 분석하고 극성 강도를 계산하여 긍정적 핵심어와 부정적 핵심어로 분류 한다. 문서에서 언급된 횟수가 많을수록 작성자가 관심을 많이 기울이고 있는 대상이라 가정하고 핵심어를 문서에 나온 빈도수에 따라 정렬한다. 긍정적 핵심어와 부정적 핵심어에서 각각

상위 빈도수 언어 5개를 선별하고 이를 분석대상이 된 문서 작성자의 관심사로 지정 한다. 긍정적 핵심어에서 선별한 관심사는 작성자가 긍정적인 의견을 가지는 관심사가 되고 부정적인 핵심어에서 선별한 관심사는 작성자가 부정적인 의견을 가진 관심사가 된다. 관심사를 다른 이용자들의 관심사와 비교하여 공통된 관심사를 가진 이용자를 친구로 추천 한다.

제안한 친구 추천 시스템을 이용하여 얻은 결과의 예시는 (그림 2)와 같다.




(그림 2) 친구 추천 시스템의 추천 결과 예시

제안한 친구 추천 시스템은 이용자들이 작성한 문서에서 언급한 어휘 중에 빈도수가 많은 어휘를 선별하여 이용자의 현재 관심사로 지정한다. 또한 오피니언 마이닝을 통해 관심사를 긍정적인 것과 부정적인 것으로 분류해서 같은 관심사를 가질 뿐 아니라 그 관심사에 대해 같은 견해를 가진 또 다른 이용자를 친구로 추천한다.

## 4. 결론 및 향후 연구

오피니언 마이닝을 이용하여 온라인상에서 같은 관심사와 의견을 가지는 다른 이용자를 친구를 추천하는 시스템을 제안하였다. 제안한 시스템은 기존의 친구 추천 시스템이 등록한 친구 정보나 프로필 정보를 기반으로 하는 것과 다르게 지금까지 웹에서 작성된 개인의 글들을 바탕으로 한다. 그 결과 생일, 거주지, 출신 학교, 직장 등의 개인 정보를 공개하지 않고도 이미 불특정 다수에게 공개한 정보만으로 다른 이용자와의 공통점을 파악하고 알맞은 친구를 찾아낼 수 있다. 프로필에 수동적으로 작성한 관심사를 이용하여 친구 찾기를 하는 것과 비교하면 제안한 시스템은 이용자가 작성한 글들을 바탕으로 관심사를 자동으로 파악하므로 프로필의 별다른 수정 없이 최근의 관심사를 친구 찾기에 반영할 수 있다. 또한 이미 알고 있거나 익숙한 사람뿐만 아니라 관심사가 비슷하다면 전혀 몰랐던 사람까지 친구로 추천 받을 수 있다. 향후에는 빈도수에 따라 핵심어를 정렬할 뿐만 아니라 글의 작성 날짜를 반영하여 시간에 따라 변화하는 관심사를 반영할 수 있도록 시스템을 개선할 계획이다.

## 참고문헌

[1] 한국인터넷진흥원, "인터넷&시큐리티 이슈", pp.12, 2011 01

[2] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval Vol. 2 Nos. 1-2, pp. 1-12, 2008

[3] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques", In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP-02), pp.79-86, July 2002

[4] Jack G Conrad, Frank Schilder, "Opinion Mining in Legal Blogs", In Proceedings of the International Conference on Artificial Intelligence and Law (ICAIL), pp.231-236, 2007

[5] 양정연, 명재석, 이상구, "상품 리뷰 요약에서의 문맥 정보를 이용한 의견 분류 방법", 정보과학회논문지 : 데이터베이스 제 36권 제 4호, pp.254-262, 2009. 8

[6] 강한훈, 유성준, 한동일, "k-Structure를 이용한 한국어 상품평 단어 자동 추출 방법", 정보과학회논문지 : 소프트웨어 및 응용 제 37권 제 6호, pp.470-478, 2010. 6

[7] 김정호, 차명훈, 김명규, 채수환, "어미 변화를 고려한 감성 구문 패턴을 이용한 상품평 의견 분류", 한국컴퓨터종합학술대회논문집 Vol.37 No.1, pp.285-290, 2010

[8] 이우철, 이현아, 이공주, "효율적인 상품평 분석을 위한 어휘 통계 정보 기반 평가 항목 추출 시스템", 한국정보처리학회, pp.497-502, 2009