

PReAmacy: 소셜 네트워크 서비스에서 콘텐츠와 사용자의 친밀도를 고려한 개인화 추천 알고리즘

(PReAmacy: A Personalized Recommendation
Algorithm considering Contents and Intimacy
between Users in Social Network Services)

서 영 덕 [†]

김 정 동 ^{**}

백 두 권 ^{***}

(Young-Duk Seo)

(Jeong-Dong Kim)

(Doo-Kwon Baik)

요 약 ‘실시간성’, ‘사람들 간의 관계정보’, ‘빅 데이터’와 같은 다양한 특성을 갖는 소셜 네트워크 콘텐츠는 개인화 추천 시스템의 성능 향상에 큰 도움이 되고 있다. 그 중 ‘사람들 간의 관계정보’가 가장 중요한 역할을 하기 때문에, 이를 활용한 다양한 연구가 진행되고 있다. 하지만 기존의 연구에서는 사람들 간의 친밀도를 고려하지 않고 있어서 개인의 성향을 반영하기 어렵고 다양한 도메인에서 정확한 추천이 불가능하다. 본 논문은 기존 연구의 문제를 해결하기 위해 사용자간 친밀도를 측정하는 친밀도 알고리즘과 소셜 네트워크의 다양한 특성에 기반한 개인화 추천 알고리즘인 PReAmacy를 제안한다. 실험을 통해 PReAmacy가 기존의 알고리즘에 비해 높은 성능을 가지며 친밀도가 PReAmacy 알고리즘에 큰 비중을 차지한다는 것을 보였다.

키워드: 소셜 네트워크 서비스, 친밀도, 개인화 서비스, 개인화 추천 알고리즘

Abstract Various characteristics of social network contents such as real-time, people relationship and big data can help to improve personalized recommender systems. Among them, ‘people relationship’ is a key factor of recommendation, so many personalized recommender systems utilizing it have been proposed. However, existing researches can not reflect personal tendency and are unable to provide precise recommendations in various domains, because they do not consider intimacy among people. In this paper, to solve these problems, we propose PReAmacy, a Personalized Recommendation Algorithm, considering intimacy among users and various characteristics of social network contents. Our experimental results indicate that not only the precision of PReAmacy is higher than that of existing algorithms, but intimacy is of great importance in PReAmacy.

Keywords: social network services, intimacy, personalized service, personalized recommender algorithm

· 논문은 2011년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(NRF-2011-0025588), 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단-차세대정보-컴퓨팅기술개발사업의 지원을 받아 수행된 연구임(No.2012M3C4A7033346)

· 이 논문은 제40회 추계학술발표회에서 ‘PReAmacy: 소셜 네트워크 서비스 콘텐츠와 사용자의 친밀도를 고려한 개인화 추천 알고리즘’의 제목으로 발표된 논문을 확장한 것임

[†] 학생회원 : 고려대학교 컴퓨터 전자통신공학과
seoyoungd@korea.ac.kr

^{**} 정 회 원 : 고려대학교 컴퓨터 전자통신공학과 교수
kjd4u@korea.ac.kr
(Corresponding author임)

^{***} 종신회원 : 고려대학교 융합소프트웨어전문대학원 교수
baikdk@korea.ac.kr

논문접수 : 2013년 12월 16일
(Received 16 December 2013)

논문수정 : 2014년 4월 10일
(Revised 10 April 2014)

심사완료 : 2014년 4월 16일
(Accepted 16 April 2014)

Copyright©2014 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회논문지: 데이터베이스 제41권 제4호(2014.8)

1. 서 론

최근 페이스북, 트위터, del.icio.us, Last.fm과 같은 소셜 네트워크의 발전과 함께 수많은 사용자들이 소셜 네트워크를 사용하고 있다. 이를 통해 다양한 특성을 갖는 방대한 양의 콘텐츠가 생성되고 있고, 여러 분야에서 소셜 네트워크 콘텐츠를 이용한 연구가 진행되고 있다. 그 중에서 사용자의 요구와 선호도에 따라 적합한 정보를 제공하는 개인화 추천에 대한 연구가 활발히 이루어지고 있다.

전통적인 개인화 추천 시스템에서는 사용자의 과거 구매이력이나 사용자와 비슷한 구매 성향을 갖는 다른 사용자의 구매이력을 통해 사용자 개인의 특성을 분석하고, 이를 시스템에 적용시키는 방법을 사용하고 있다 [1]. 소셜 네트워크를 개인화 추천 시스템에 활용할 경우, '실시간성', '사람들 간의 관계정보', '빅 데이터'와 같은 다양한 특성을 적용할 수 있기 때문에 기존의 개인화 추천 방식에 비해 추천의 정확도를 향상시킬 수 있다 [2]. 이러한 이유로 소셜 네트워크 정보를 활용한 개인화 추천 서비스들이 등장하였다. 아마존은 페이스북과의 제휴를 통해 'Amazon+Facebook = A Perfect Storm for Recommendation'이라는 서비스를 제공한다. 이 서비스를 통해 아마존에서 상품을 구매하는 사람들은 페이스북 친구들에게 상품을 직접적으로 추천 받을 수 있음은 물론이고 페이스북 공개 정보에 기반하여 상품을 추천 받는 것이 가능하다. [3]과 [4]는 유명한 음악 포털 사이트인 Last.fm을 이용하여 음악재생정보와 사용자간 관계정보를 통해 사용자에게 음악을 추천해주는 서비스를 제공하고, [5]는 FilmTrust라는 영화 추천 웹사이트를 개발하여 사용자에게 영화를 추천해주는 서비스를 제공하고 있다.

개인화 추천 시스템에서 사용되는 소셜 네트워크 콘텐츠 중 가장 중요한 요소는 '사람들 간의 관계정보'이다. 사람들은 지인과 같이 자신과 관련이 있는 사람들의 의견을 신뢰하는 경향이 있는데, 이는 소셜 네트워크 사용자의 구매결정에 미치는 영향에 대한 조사결과를 통해 알 수 있다. 세계최고의 여론조사기관인 Harris Interactive에 따르면 구매자의 제품 및 브랜드 선택 의사결정에 있어 가장 많은 영향을 미치는 원인으로 71%가 '가족과 친구'의 의견이라고 조사되었다. 소셜 네트워크가 발전함에 따라 링크로 이어져 있는 소셜 네트워크상 사람들간의 관계를 통해 자신과 직·간접적으로 연관된 사용자의 콘텐츠를 사용할 수 있게 되었고, 실제로 개인화 추천 시스템에 적용하였을 경우 추천의 성능이 향상된다 [6].

이러한 장점 때문에 '사람들 간의 관계정보'를 활용한

개인화 추천 시스템과 관련한 연구가 활발히 진행되고 있지만, 기존의 연구에서는 사람들간의 가까움 정도를 고려하지 않고 있다. 사람들 간의 관계가 가깝고 먼 정도를 고려하지 않고 동일하게 취급하여 추천이 이루어지기 때문에 개인의 성향을 충분히 반영하지 못하고 다양한 도메인에서 정확한 추천이 불가능하다. 따라서 영화와 음악과 같은 한정된 도메인 내에서 비슷한 성향을 갖는 사람들이 사용하는 소셜 네트워크를 통해 추천이 이루어지고 있다. 본 논문에서는 트위터를 활용하여 사용자간 가까움 정도를 친밀도라 정의하고, 친밀도와 소셜 네트워크의 다양한 특성을 고려한 개인화 추천 알고리즘 PReAmacy를 제안하여 기존 연구의 문제점을 해결하려 한다.

이 논문의 구성은 다음과 같다. 제2장에서는 관련연구로 소셜 네트워크 콘텐츠 특히, '사람들 간의 관계정보'를 활용한 개인화 추천 기법에 대해 언급한다. 제3장에서는 본 논문에서 제안하는 소셜 네트워크 사용자간 친밀도를 구하는 친밀도 알고리즘과 PReAmacy에 대해 자세히 서술한다. 제4장에서 PReAmacy의 성능을 평가하기 위한 개인화 추천 시스템을 설명하고 개인화 추천 시스템을 통해 PReAmacy를 검증한다. 마지막으로 제5장에서는 결론 및 향후 연구에 대해 제시한다.

2. 관련 연구

기존의 개인화 추천 시스템에서는 상품에 대한 사용자의 평점 정보나 웹사이트 방문기록과 같은 사용자의 행동양식을 통해 사용자의 선호도를 측정하고 이를 통해 사용자가 원하는 정보를 추천해 준다. 소셜 네트워크를 개인화 추천 시스템에 활용할 경우 '실시간성', '사람들 간의 관계정보', '빅 데이터'와 같은 다양한 특성들을 적용할 수 있기 때문에 기존의 개인화 추천 방식에 비해 정확도가 향상된다. 따라서 소셜 네트워크를 활용한 개인화 추천에 대한 연구가 다양하게 진행되고 있다. [3]은 음악 전문 소셜 네트워크인 Last.fm데이터를 활용하여 사용자에게 음악을 추천해주는 연구를 진행하였다. 사용자 프로필에 기반한 태그정보를 통계적으로 분석하고 추천에 활용하여 음악 재생 정보를 통해 이루어지는 추천방식에 비해 높은 성능을 가치를 증명하였다. [7]은 소셜 북마킹 사이트인 del.icio.us에서 생성되는 사용자의 태그정보를 통해 사용자들의 공통관심사를 추출해주는 연구를 진행하였고, [8]은 소셜 네트워크 상에서의 사용자 선호도와 사용자간 영향력이 추천에 중요한 영향을 미친다는 사실을 증명하였다.

소셜 네트워크의 다양한 특성 중 '사람들 간의 관계정보'는 개인화 추천 시스템에서 가장 중요한 요소이다. 기존의 개인화 추천 시스템에서는 불특정 다수의 정보

를 사용하여 추천이 이루어졌지만, ‘사람들 간의 관계정보’를 사용함으로써 지인과 같은 자신과 관련이 있는 사람들을 통해 추천이 가능하게 되었다. 이로 인해 개인화 추천 시스템의 성능이 크게 향상되었으며[6], 이를 활용한 다양한 연구가 진행되고 있다. [4], [5]는 협업적 필터링 기법을 사용하여 사용자에게 음악과 영화를 추천해주는 시스템을 제안한다. [4]는 사용자의 음악 재생 정보와 태그정보, 친구정보를 활용하여 피어슨 상관계수를 통해 사용자간 유사도를 계산하여 추천을 제공한다. [5]는 유용한 추천 결과를 이끌어 낼 수 있는 정보를 제공하는 사용자를 신뢰관계에 있다고 보고 신뢰관계에 있는 사용자의 정보를 통해 추천을 제공하는 웹사이트인 FilmTrust를 개발하였다. [2]는 협업적 필터링 기법과 내용 기반 필터링 기법을 혼합한 하이브리드 형식의 추천방식을 제안한다. ‘사람들 간의 관계정보’와 사용자가 작성한 태그정보를 모두 고려하여 사용자에게 적합한 소셜 미디어를 추천해주며, 설문조사를 통한 이용자 연구를 통해 제안하는 시스템에 대한 평가를 진행하였다. [2], [4], [5] 모두 ‘사람들 간의 관계정보’를 이용하고 있지만 사용자간의 친밀도는 고려하지 않고 모든 사용자간의 관계를 동일하게 취급하여 추천이 이루어지기 때문에, 영화와 음악과 같은 비슷한 성향을 갖는 사람들의 집합을 가진 한정된 도메인의 소셜 네트워크 상에서만 추천이 이루어지고 있다.

친밀도는 사람들 간의 가까움 정도를 의미한다. 친밀도는 개인화 서비스에서 중요한 역할을 하기 때문에[9], 개인화 추천 시스템에 친밀도를 반영할 경우 사용자간의 관계를 동일하게 취급하는 추천 시스템에 비해 높은 성능의 추천이 가능하다. 본 논문에서는 소셜 네트워크가 가지고 있는 특성과 사용자간의 친밀도를 고려하여 개인의 성향이 반영된 개인화 추천 알고리즘인 PReAmacy를 제안한다.

3. 개인화 추천 알고리즘

친밀도를 고려한 개인화 추천 알고리즘은 협업적 필터링에 기반한 알고리즘이다. 협업적 필터링의 경우 피어슨 상관계수를 통해 사용자간 유사도를 계산하여 추천이 이루어지는 방식으로 ‘사람들 간의 관계정보’를 이용할 수 있기 때문에 소셜 네트워크를 활용한 개인화 추천 시스템에 많이 사용이 된다. 하지만 기존의 연구에서는 모든 사용자간 관계를 동일하게 보고 접근하고 있기 때문에 개인의 성향을 반영하기가 힘들고 추천의 정확성이 떨어진다. 이러한 문제를 해결하기 위해 기존 알고리즘과 달리 사용자간 유사도를 계산할 때 친밀도를 반영한다. 사용자간 친밀도를 친밀도 알고리즘을 통해 구하고 사용자가 작성한 콘텐츠의 작성 시간, 콘텐츠의

극성정보를 이용하여 개인화 추천에 적합한 PReAmacy를 제안한다.

3.1 친밀도 알고리즘

소셜 네트워크 상에서 사용자는 일반적으로 오프라인에서 자신과 친하거나 같은 소속에 속한 사람 혹은 관심 분야가 비슷한 사람을 친구로 등록하기 때문에, 소셜 네트워크 사용자의 친구정보는 사용자의 소속이나 관심 분야 등을 대변해주는 중요한 정보가 된다. 다시 말해서 두 사용자의 친구 관계가 유사하다는 것은 그들의 관심사가 비슷하다는 것이므로 두 사용자가 친할 확률이 높다는 것을 의미한다[10].

본 논문에서 사용되는 ‘사람들 간의 관계정보’를 활용한 친밀도 알고리즘은 친구를 공유하는 비율을 통해서 구해진다. 사용자 a 가 친구 $f_{a,1}, \dots, f_{a,n}$ 을 갖는다면 사용자 a 의 친구 집합 F_a^1 는 다음과 같이 정의된다.

$$F_a^1 = \{f_{a,1}, \dots, f_{a,n}\}$$

기존의 친밀도 알고리즘에서는 직접적인 관계에 있는 친구 관계만을 이용하여 친밀도를 계산하는데[11], 제안하는 PReAmacy에서는 보다 정확한 친밀도를 계산하고 사용자의 성향을 정확하게 파악하기 위해 직접적인 관계에 있는 친구 집합인 F_a^1 뿐만 아니라 보다 넓은 범위의 친구 집합까지 고려해 준다. 직접적으로 연결된 사용자의 정보뿐만 아니라 간접적으로 연결된 사용자의 정보를 이용할 경우, 자신이 가지는 사회적 범위(Social Circle)내에서 얻는 정보보다 많은 정보를 얻을 수 있기 때문에[12], 직접적으로 연결되어 있는 사용자뿐만 아니라 간접적으로 연결되어 있는 사용자 또한 고려해주는 것이 중요하다. 이를 위해 제안하는 친밀도 알고리즘에서는 그림 1과 같이 친구관계에 레벨(Level)을 적용하였다.

사용자 a 와 직접적인 관계에 있는 친구집합 F_a^1 는 레벨1로 두고, a 의 친구의 친구와 같이 간접적인 관계를 레벨2로 정의한다. 필요한 경우 레벨3 이상의 친구 관계까지 고려하여 친밀도를 계산한다. 레벨2 친구집합 F_a^2 는 다음과 같이 정의된다.

$$F_a^2 = F_{f_{a,1}}^1 \cup \dots \cup F_{f_{a,n}}^1$$

본 논문에서는 친밀도를 구하기 위하여 Jelinek-Mercer Smoothing 기법[13]을 활용하였다. 레벨1 친구집합 F_a^1 와 레벨2 친구집합 F_a^2 를 모두 고려한 사용자 a 와 $f_{a,i}$ 간 친밀도 $i_{a,f_{a,i}}$ 는 다음과 같다.

$$i_{a,f_{a,i}} = (1 - \lambda) \frac{|F_a^1 \cap F_{f_{a,i}}^1|}{|F_a^1|} + \lambda \frac{|F_a^2 \cap F_{f_{a,i}}^1|}{|F_a^2|}$$

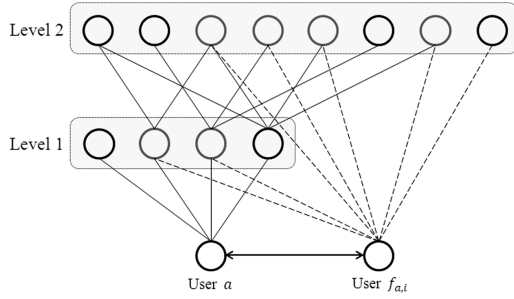


그림 1 사용자의 친구관계에서 정의한 레벨의 개념
Fig. 1 Concept of friend level

λ 는 상수로 TREC 실험결과를 통해 쿼리의 길이에 따라 0.1과 0.7이 사용된다. 제안하는 친밀도 알고리즘 $i_{a,f_{a,i}}$ 에서는 직접적인 친구관계 F_a^1 에 보다 많은 가중치를 두어야 하기 때문에 λ 를 0.1로 두었다. $|F_a^1 \cap F_{f_{a,i}}^1|$ 는 사용자 a 와 $f_{a,i}$ 간 공유하는 친구의 수를 의미하고, $|F_a^2 \cap F_{f_{a,i}}^1|$ 는 사용자 a 의 친구들과 $f_{a,i}$ 간 공유하는 친구의 수를 의미한다.

3.2 PReAmacy

제안하는 PReAmacy 알고리즘에서는 사용자 간의 친밀도를 계산하여 개인의 성향을 반영한 상품을 사용자에게 추천한다. 친밀도 정보뿐만 아니라 콘텐츠가 작성된 시간정보와 콘텐츠의 극성과 같은 소셜 네트워크의 다양한 특성을 나타내는 데이터 또한 사용된다.

우선 콘텐츠의 극성을 측정하기 위해 SentiWordNet[14]에 기반하여 각 도메인에 맞게 한국어 감정어 사전을 구축하였다. 감정어 사전은 긍정어와 부정어를 포함하며, 이를 기반으로 사용자가 작성한 콘텐츠를 *Good*, *None*, *Bad*로 분류하고 각각 점수를 5, 3, 1점씩 부여하였다.

콘텐츠가 작성된 시간정보는 소셜 네트워크의 특성상 최근에 작성된 콘텐츠 일수록 신뢰도가 높음으로 최근에 작성된 콘텐츠에 가중치를 주어야 한다[15]. 또한 시간정보가 추천 결과에 너무 많은 영향을 주지 않기 위해 0.5와 1.5사이의 값으로 값을 조정하였다. 콘텐츠가 작성된 시간을 t_{post} 라 하고 모든 콘텐츠 중 가장 최근에 작성된 콘텐츠의 시간정보와 가장 오래 전에 작성된 콘텐츠의 시간정보를 각각 t_{recent} , t_{old} 라 할 때, PReAmacy에 적용되는 시간함수 t_i 는 다음과 같이 구해진다.

$$t_i = 2^{\frac{t_{post}-t_{old}}{t_{recent}-t_{old}}} - 2^{-1}$$

PReAmacy는 사용자간 친밀도를 고려한 사용자 유사도와 상품의 선호도를 통해 구해지며 선호도 값은 콘텐츠의 극성과 시간함수의 곱으로 구해진다. 사용자의 친구집합 $F_a^1 \cup F_a^2$ 내에서만 계산이 되며, 상품 i 에 대한

사용자 a 의 랭킹스코어 $r_{a,i}$ 는 다음과 같다.

$$r_{a,i} = \overline{p_a t} + \frac{\sum_{u=1}^{|F_a^1 \cup F_a^2|} (p_{u,i} t_i - \overline{p_u t}) s_{a,u}}{\sum_{u=1}^{|F_a^1 \cup F_a^2|} s_{a,u}}$$

$p_{u,i} t_i$ 는 상품 i 에 대한 사용자 u 의 선호도를 의미하며, 사용자 u 가 i 에 대해 언급한 모든 콘텐츠 선호도의 평균이다. $\overline{p_a t}$ 와 $\overline{p_u t}$ 는 모든 상품에 대한 사용자 a 와 u 의 선호도 값 평균을 의미하며, $s_{a,u}$ 는 친밀도를 고려한 사용자 a 와 u 의 유사도 값이다. 기존의 협업적 필터링 기법에서는 사용자간 선호도의 유사도를 측정하기 위해 피어슨 상관계수를 사용한다[16]. σ_a 와 σ_u 를 사용자 a 와 u 의 선호도 값 표준편차라고 할 때, PReAmacy에서 사용되는 피어슨 상관계수 다음과 같다.

$$w_{a,u} = \frac{\sum_{i=1}^M (p_{a,i} t_i - \overline{p_a t})(p_{u,i} t_i - \overline{p_u t})}{\sigma_a \sigma_u}$$

선호도가 유사한 사용자를 통해 추천이 이루어질 경우, 개인화 추천의 성능이 향상된다[17]. 하지만 피어슨 상관계수만 사용하여 사용자간 유사도를 계산할 경우, 소셜 네트워크 상의 ‘사람들 간의 관계정보’를 사용할 수 없기 때문에 소셜 네트워크의 콘텐츠를 고려한 추천 시스템에는 적절하지 않다. 따라서 PReAmacy에서 사용하는 유사도 $s_{a,u}$ 는 피어슨 상관계수와 친밀도 $i_{a,u}$ 를 함께 고려하여 사용자간 유사도 $s_{a,u}$ 를 계산한다. $\alpha + \beta = 1$ 이고, 실험을 통해 각각 0.2와 0.8로 설정하였다.

$$s_{a,u} = \alpha w_{a,u} + \beta i_{a,u}$$

4. 실험 및 평가

본 논문에서의 실험은 스마트폰을 도메인으로 한달간 크롤링한 트위터 데이터를 이용하고 개인화 추천 시스템을 구축하여 실험을 진행하였다. 또한 PReAmacy의 추천 성능을 증명하기 위해 기존의 협업적 필터링과 친구집합 내의 정보만 고려한 기존의 협업적 필터링, 친밀도를 고려하지 않은 PReAmacy를 적용한 경우와 비교평가 하였다. 기존의 협업적 필터링의 경우 소셜 네트워크의 특징을 사용하지 않기 때문에, 상품에 대한 사용자의 선호도를 측정할 때 실시간성을 고려하지 않고 불특정 다수를 통해 추천이 이루어진다. 친구 집합 내의 정보만 고려한 기존의 협업적 필터링의 경우는 상품에 대한 사용자의 선호도는 기존의 협업적 필터링과 똑같이 측정되나, 불특정 다수가 아닌 친구의 정보를 통해서 추천이 이루어진다. 친밀도를 고려하지 않은 PReAmacy의 경우는 상품에 대한 사용자의 선호도를 측정할 때 실시간성과 콘텐츠의 극성을 모두 고려하고, 친구의 정보를 통해 추천이 이루어지지만 사용자간 친밀도는 고

려하지 않는다. PReAmacy와 3가지 비교대상 알고리즘에 대한 정확도를 각각 측정하여 친구의 정보를 이용하는 것이 개인화 추천에 얼마나 영향을 미치는지를 확인하고, 소셜 네트워크의 특징을 협업적 필터링에 적용하였을 경우 기존의 협업적 필터링 보다 나은 성능을 보이는지 알아보려 한다. 마지막으로 친밀도가 소셜 네트워크를 활용한 개인화 추천에 큰 비중을 차지한다는 것을 확인하려 한다.

성능을 평가하기 위한 방식으로는 개인화 추천 시스템의 성능평가지표로 많이 사용되는 Mean Average Precision(MAP), Mean Reciprocal Rank(MRR), Normalized Discounted Cumulative Gain(NDCG) 방식을 이용한다. 사용자는 자신에게 적합한 추천 결과를 찾기 위해 상위 랭크된 결과만을 보는 경향이 있기 때문에, 첫 번째로 추천된 상품이 추천 리스트 중에서 가장 높은 중요성을 갖는다. 따라서 개인화 추천의 성능을 평가하기 위해서는 랭킹된 결과의 정확성을 측정해야 하고 MAP, MRR, NDCG가 대표적으로 사용된다[1].

MAP, MRR, NDCG를 측정하기 위해서는 추천된 상품의 적합성을 판단해야 한다. 본 논문에서는 정확도를 측정한 실험을 통해 상품에 대한 사용자의 선호도($p_{u,i}$)가 특정 임계값을 넘으면 적합하다고 정의하였고, 임계값을 모든 상품에 대한 사용자의 선호도($\overline{p_u t}$)로 결정하였다. 또한, 사용자의 친구 정보를 통해 구해진 예측 선호도($r_{u,i}$)를 통해 상품의 랭킹을 결정하고, 랭킹된 결과의 적합성을 판단하여 MAP, MRR, NDCG를 측정하였다.

4.1 데이터 집합

본 논문에서는 2012년 7월 1일부터 2012년 7월 30일까지 크롤링한 한달 치의 트위터 데이터를 사용한다. 크롤링한 데이터 중 시중에 판매되고 있는 7가지의 스마트폰 iphone, galaxy, optimus, vega, blackberry, xperia, HTC와 관련된 256,176개의 콘텐츠를 따로 추출하여 저장하였다. 정확한 실험평가를 위해 개인화 추천에 사용되는

표 1 실험데이터

Table 1 Data Set

Time		2012-07-01 ~ 2012-07-30
Contents	Domain	Smart Phone
	Items	Iphone, galaxy, optimus, vega, blackberry, xperia, HTC
	#	259,176
Users	All Users	21,640
	Active USers	212

친구 집합은 사용자와 팔로워(Followee)관계이면서 스마트폰과 관련된 언급을 한 사용자들로 구성되며, 스마트폰과 관련한 언급을 한 21,640명의 사용자중 스마트폰에 대해 활발히 언급한 사용자 212명을 대상으로 실험을 진행하였다. 실험 데이터로 사용한 212명의 사용자는 본 논문의 3.1절에서 언급한 친구집합의 정의에 따라 level2에서 대부분의 사용자가 서로 친구관계를 형성한다. 본 논문에서 사용된 실험데이터는 표 1과 같다.

4.2 개인화 추천 시스템

PReAmacy를 기반으로 사용자에게 개인화 추천을 제공하기 위해 본 논문에서는 그림 2와 같이 개인화 추천 시스템을 설계하여 구현하였다. 개인화 추천 시스템은 크게 초기 데이터(Pre-stored data)와 개인화 추천 엔진(Personalized Recommendation Engine)으로 이루어져 있다. 초기 데이터는 트위터 콘텐츠와 사용자간 인맥 정보로 구성된 크롤링 데이터와 감정어 사전으로 이루어져 있다. 감정어 사전은 SentiWordNet에 기반하여 한국어에서 긍정과 부정을 나타내는 대표어휘를 선정하고 각 도메인에 맞게 확장 및 수정하였다. 본 논문에서 사용된 스마트폰 감정어 사전은 긍정어 384개, 부정어 509개로 최종적으로 생성된 감정어 사전은 표 2와 같다.

개인화 추천 엔진에서는 초기 데이터에 의해 가공된 데이터를 PReAmacy에 적용하여 사용자에게 적합한 추천 결과를 제공한다. 가공된 데이터는 친밀도와 도메인

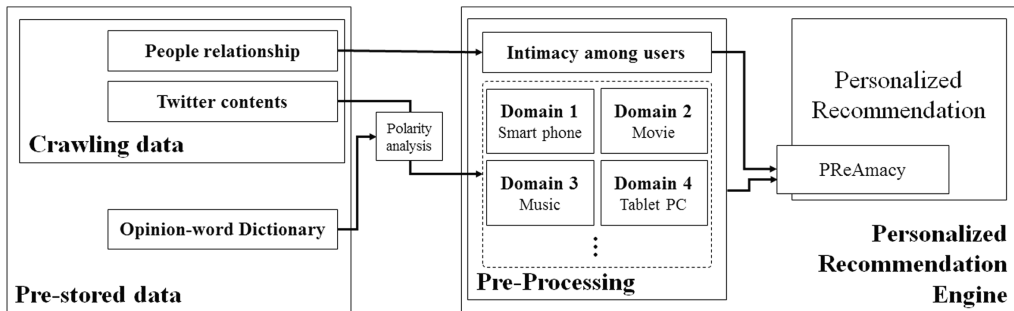


그림 2 개인화 추천 시스템

Fig. 2 Personalized Recommender System

표 2 스마트폰 감정어 사전
Table 2 Opinion-word dictionary for smart phone

Polarity	Opinion-word Dictionary	
	contents	#
Positive	간지나다, 간편하다, 감동받다, 강렬하다, 갖고 싶다, 개통 가능하다, 걱정 없다, 건강하다, 검색 잘되다, 견고하다, 고급스럽다, 고맙다, 고장나지않다 and so on	384
Negative	구동느리다, 구리다, 구매하기어렵다, 귀찮다, 그림감 나쁘다, 꺼졌다켜지다, 글자잘리다, 기분나쁘다, 까다롭다, 깨끗하지 않다, 그지같다, 꺼지지않다 and so on	509

별로 분류된 트위터 데이터로 이루어지는데, 사용자간 친밀도는 초기 데이터에 저장되어 있는 트위터 사용자 인맥정보를 친밀도 알고리즘에 적용하여 구해진다. 다음으로 크롤링한 트위터 콘텐츠를 도메인 별로 정해진 키워드를 통해 분류하고 감정어 사전을 통해 트위터 콘텐츠의 극성을 분석하여 PReAmacy에 적용이 가능한 콘텐츠로 가공한다.

4.3 정확도 측정

제안하는 PReAmacy의 성능을 검증하기 위해 기존의 협업적 필터링과 친구집합 내의 정보만 고려한 기존의 협업적 필터링, 친밀도를 고려하지 않은 PReAmacy와 비교평가를 수행하였다. 우선 MAP, MRR, NDCG를 이용하여 개인화 추천의 정확도를 측정하였고, 친밀도 가중치 β 에 따른 정확도를 측정하여 친밀도가 개인화 추천 알고리즘에 미치는 영향을 분석하였다.

표 3에서는 기존의 협업적 필터링 기법, 친구집합 내의 정보만 고려한 기존의 협업적 필터링, 친밀도를 고려하지 않은 PReAmacy와 제안하는 PReAmacy를 정량적으로 비교 평가한 결과를 나타낸다. 우선 PReAmacy가 기존의 협업적 필터링 기법에 비해 높은 정확도를 갖는데, 이를 통해 소셜 네트워크의 콘텐츠를 이용할 경우 추천의 질이 향상됨을 확인할 수 있다. 또한 소셜 네

표 3 정량적 비교평가 결과
Table 3 Result of comparison evaluation

Algorithm	Comparison Result		
	MAP	MRR	NDCG
Existing CF	0.7173	0.7370	0.8250
Existing CF using friends data	0.7352	0.7532	0.8261
PReAmacy without intimacy ($\alpha = 1, \beta = 0$)	0.8106	0.8058	0.8647
PReAmacy ($\alpha = 0.2, \beta = 0.8$)	0.9067	0.9195	0.9451

트워크 상의 '사람들 간의 관계정보'를 사용하였을 경우 추천 성능에 미치는 영향을 확인할 수 있다. 불특정 다수를 통해 추천이 이루어지는 기존의 협업적 필터링 기법이 친구의 정보만을 사용하여 추천결과를 제공하는 나머지 기법에 비해 낮은 성능을 보이는 것으로 보아 '사람들 간의 관계정보'를 개인화 추천 시스템에 사용할 경우 추천결과와 질이 향상됨을 확인할 수 있었다. 마지막으로 친밀도를 고려하지 않은 PReAmacy와 친밀도를 고려한 PReAmacy를 비교하여 친밀도가 개인화 추천 시스템에 미치는 영향을 확인하였다. 모든 사용자간의 관계를 동일하게 보고 접근하는 것이 아니라 PReAmacy는 친밀도 알고리즘을 통해 나와 좀더 친밀한 사용자에게 가중치를 주어 추천을 제공하기 때문에 기존의 협업적 필터링 기법이나 친밀도를 고려하지 않은 PReAmacy에 비해 추천의 정확도가 높음을 확인할 수 있다.

추천결과와 정확도뿐만 아니라 친밀도가 제안하는 알고리즘에 미치는 영향을 분석하기 위해 사용자의 유사도 계산에 사용되는 친밀도 가중치 β 에 따른 정확도를 측정하여 비교하였다. 그림 3에서 보는 것과 같이 정확도가 점점 증가하다가 β 가 0.8일 때, 가장 높은 정확도를 갖는다. 친밀도에 높은 가중치가 부여될 때 정확도가 증가함을 통해 친밀도가 개인화 추천 알고리즘에서 중요한 역할을 함을 확인하였다.

5. 결론 및 향후 연구

본 논문에서는 소셜 네트워크에서 사용된 콘텐츠, 특히 사용자간 친밀도 정보를 이용하여 개인의 성향이 반영된 개인화 추천 알고리즘인 PReAmacy를 제안한다. PReAmacy는 사용자의 친구 정보를 통해 사용자간 친밀도를 고려하여 추천을 제공하기 때문에 개인의 성향을 반영함과 동시에 추천의 정확도를 향상시킬 수 있다. PReAmacy의 성능을 측정하기 위하여 개인화 추천 시스템을 설계하여 시스템을 구현하였으며 스마트폰을 도메인으로 하여 실험을 진행하였다. 실험을 통해 PReAmacy가 기존의 협업적 필터링 기법보다 높은 성능을 보임을 증명하였고, PReAmacy에서 친밀도가 높은 비중을 차지 한다는 것을 확인하였다.

본 논문에서는 제안 알고리즘을 검증하기 위하여 스마트폰 도메인을 이용하였다. PReAmacy는 도메인에 따라 알고리즘에 적용되는 친구의 집합이 변화하기 때문에 다양한 도메인에서도 높은 성능을 갖는 추천이 가능할 것이다. 이를 검증하기 위해 다양한 도메인의 콘텐츠 집합과 각 도메인 별 감정어 사전을 구축하여 PReAmacy가 스마트폰 도메인뿐만 아니라 다양한 도메인에서도 높은 성능을 갖는다는 것을 보이고자 한다. 또한 소셜 네트워크의 대용량 데이터를 처리하기에 적합한지 여부

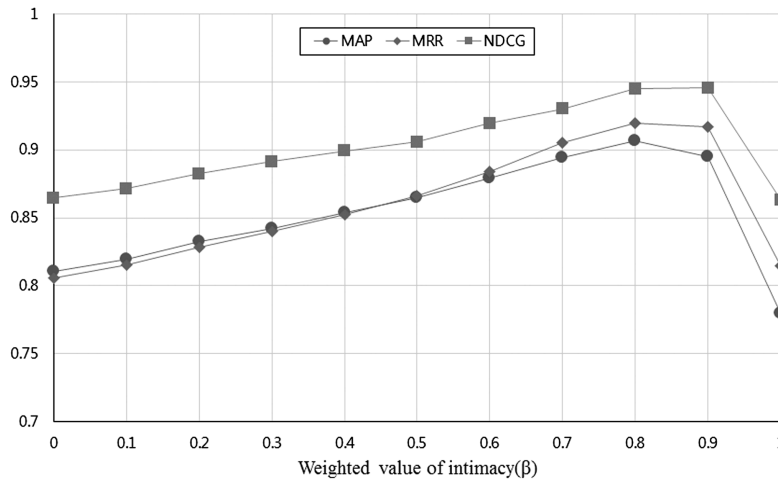


그림 3 친밀도 가중치에 따른 정확도 비교

Fig. 3 Comparison of precision according to change of weighted value of intimacy

를 PReAmacy의 계산 시간 성능을 측정하여 검증할 예정이다.

References

- [1] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutierrez, "Recommender systems survey," *Knowledge-Based Systems*, vol.46, pp.109-132, 2013.
- [2] I. Guy, N. Zwerdling, I. Ronen, D. Carmel, and E. Uziel, "Social media recommendation based on people and tags," *SIGIR. ACM*, pp.194-291, 2010.
- [3] C. S. Firan, W. Nejdl, and R. Paiu, "The Benefit of Using Tag-Based Profiles," *Proceedings of the Latin American Web Conference*, pp.32-41, 2007.
- [4] I. Konstas, V. Stathopoulos, And J. M. Jose, "On Social Networks and Collaborative Recommendation," *SIGIR. ACM*, pp.195-202, 2009.
- [5] J. Golbeck, "Generative Predictive Movie Recommendations from Trust in Social Networks," *Proceeding of 4th international conference on Trust Management*, pp.93-104, 2006.
- [6] H. Kautz, B. Selman, and M. Shah, "Referral Web: Combining Social Networks and Collaborative Filtering," *Communications of the ACM*, vol.40, pp.63-65, Mar. 1997.
- [7] X. Li, L. Guo, and Y. E. Zhao, "Tag-based Social Interest Discovery," *WWW*, pp.675-684, 2008.
- [8] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang, "Social Contextual Recommendation," *CIKM*, pp.45-54, 2012.
- [9] T. P. Liang, Y. W. Li, and E. Turban, "Personalized Services As EmPathic Responses: The Role of Intimacy," *Proceeding of PACIS*, 2009.
- [10] D. H. Lee, P. Brusilovsky, "Does Trust Influence Information Similarity?," *ACM RecSys'09 Workshop on Recommender Systems & the Social Web*, pp.71-74, 2009.
- [11] K. S. Seol, J. D. Kim, H. N. Shin, D. K. Baik, "Intimacy Measurement Method and Experiment between Social Network Service Users," *Journal of KIISE*, vol.39, no.4, pp.335-341, 2012.
- [12] M. S. Granovetter, "The Strength of Weak Ties," *American Journal of Sociology*, vol.78, no.6, pp.1360-1380, 1973.
- [13] F. Jelinek, and R. L. Mercer, "Interpolated estimation of Markov source parameters from sparse data," *Proceeding of the Workshop on Pattern Recognition in Practice*, 1980.
- [14] A. Esuli, and F. Sebastiani, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining," *LREC*, pp.417-422, 2006.
- [15] N. Dai, and B. D. Davison., "Freshness Matters: In Flowers, Food, and Web Authority," *Proceeding of SIGIR*, pp.114-121, 2010.
- [16] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News," *Proceeding of Communication of the ACM*, vol.40, Issue. 3, pp.77-87, 1997.
- [17] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," *Proceeding of SIGIR*, pp.230-237, 1999.



서 영 덕

2012년 고려대학교 컴퓨터통신공학부(공학사). 2012년~현재 고려대학교 컴퓨터·전파통신공학과(석·박통합과정). 관심분야는 소셜 네트워크 서비스, 자가적응소프트웨어



김 정 동

2008년 고려대학교 컴퓨터학과(이학석사) 2012년 고려대학교 컴퓨터전파통신공학과(공학박사). 2012년~현재 고려대학교 컴퓨터전파통신공학과(연구교수). 관심분야는 메타데이터, 데이터 통합, 시맨틱 웹, 온톨로지, 소셜 네트워크, 상황인지

백 두 권

정보과학회논문지 : 데이터베이스
제 41 권 제 3 호 참조