

A Facebook Page Ranking and Highlight Contents Selection Scheme based on User Interests

Jehyeok Rew[†] · Young-Hwan Choi[†] · Eenjun Hwang^{††}

ABSTRACT

As browsing and sharing multimedia contents through various social network services have become very common these days, there has been a rising demand for effectively provide appropriate multimedia contents to users depending on their interest and characteristics. Especially, considering the enormous volume of multimedia contents created daily, user satisfaction heavily depends on effective selection of appropriate multimedia contents. For instance, if the acquisition of contents is based on friends or related people, then it is usually considered to be more trustworthy. Facebook page is one of the social spaces where people can obtain various information of interest. Still, users have difficulty in finding proper information because Facebook generates tremendous amount of data daily. In this paper, we proposed a scheme to extract and select Facebook page contents effectively. In our proposed scheme, users are grouped based on mutual intimacy and Facebook pages of interest are selected based on the ranking of importance. We show the performance of our scheme through experiments.

Keywords : Facebook Page, Ranking, Highlight Contents Selection

사용자 관심 정보 기반의 Facebook 페이지 순위 매김 및 대표 콘텐츠 선택 기법

유 제 혁[†] · 최 영 환[†] · 황 인 준^{††}

요 약

최근 들어, 소셜 네트워크 서비스를 통한 멀티미디어 콘텐츠의 열람과 공유가 보편화됨에 따라, 사용자의 관심이나 특징에 부합되는 콘텐츠의 효과적 제공에 대한 욕구가 높아지고 있다. 특히, 생성되는 콘텐츠의 방대한 양을 고려할 때 적합한 콘텐츠의 효과적 선택이 사용자의 만족도를 크게 좌우하게 된다. 예를 들어, 사람간의 관계에 기반하여 관심 콘텐츠를 제공하는 방법의 경우 신뢰성이 어느 정도 보장되기 때문에, 사용자의 효과적인 열람이 가능하다. 대표적인 소셜 네트워크 서비스인 Facebook에서 제공하는 Facebook 페이지는 관심사에 대한 콘텐츠를 획득할 수 있는 공간이지만, 사용자는 제공되는 콘텐츠의 양이 너무 많아 원하는 목적에 맞게 콘텐츠를 획득하기 어렵다. 본 논문에서는, 개인 사용자의 관심에 따라 Facebook 페이지 콘텐츠를 순위별로 선정하여 대표 콘텐츠를 제공하는 기법을 제안한다. 사용자 간의 친밀도를 이용해 사용자 관심 분야에 따라 그룹화하고, 그룹화 된 사용자들에게 Facebook 페이지 콘텐츠를 중요도 순위별로 선택하여 사용자에게 제공한다. 마지막으로 실험을 통해 제안된 기법의 성능을 보인다.

키워드 : 페이스북 페이지, 순위 선정, 대표 콘텐츠 선택

1. 서 론

소셜 네트워크 서비스(Social Network Service, SNS)의 사용이 보편화 되면서, SNS 상의 정보에 대한 효과적 활용

의 욕구가 커지고 있다. 매일 엄청난 양의 데이터가 SNS 상에서 생성되고 있고, 사용자들은 자신들의 관심 정보를 얻을 만한 사람들을 통해 획득하고자 한다. 이러한 추세에 따라, 데이터 분석 처리 분야에서는 사용자 정보를 수집하고

※ 이 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 기초연구사업지원(NRF-2013R1A1A2012627)과 미래창조과학부 및 정보통신산업진흥원의 IT/SW 창의연구과제(NIPA-2013-H0502-13-1119)의 연구결과로 수행되었음.

※ 이 논문은 제40회 한국정보처리학회 추계학술발표회에서 'Facebook에서의 효과적인 소셜 검색을 위한 사용자 그룹 분류 기법'의 제목으로 발표된 논문을 확장한 것임.

† 준 회 원 : 고려대학교 전기전자전파공학과 석·박사 통합과정

†† 종신회원 : 고려대학교 전기전자전파공학과 교수

논문접수 : 2013년 12월 2일

수정일 : 1차 2014년 1월 2일

심사완료 : 2014년 1월 2일

* Corresponding Author : Eenjun Hwang(ehwang04@korea.ac.kr)

의미 있는 콘텐츠를 선정하여 제공하는 연구를 활발하게 진행하고 있다. 대표적인 SNS인 Facebook은 사용자 간의 관계에 기반하여 관심 정보를 손쉽게 획득할 수 있다. Facebook은 Facebook 페이지라는 특수한 웹 페이지 방식을 제공하여, 기업이나 조직, 브랜드, 음악 등 다양한 주제에 대한 정보의 공유를 가능하게 한다. 하지만, 뉴스피드를 통한 페이지 소식의 경우, 관심 정보가 누락되는 경우가 발생하고, 관심 정보 탐색에 불필요한 시간이 걸리며, 여러 페이지에 산재한 관심 콘텐츠의 경우 효과적으로 보여주지 못하고 있다. 이러한 문제를 해결하기 위해, 본 논문에서는, 개인 사용자의 관심 정보와 믿을 만한 사람들을 선정하는 친밀도 산정방법을 사용하여 Facebook 페이지 콘텐츠를 효과적으로 선택하고 제공하는 방법을 제안한다. 구체적으로, 우선 Facebook이 제공하는 사용자의 게시물을 내포하고 있는 타임라인과 사용자와 친구 관계인 게시물을 열람할 수 있는 뉴스피드의 정보를 획득하고 이를 토대로 사용자 간 친밀도를 계산한다. 둘째로, 계산된 친밀도에 기반하여 Facebook이 제공하는 관심 항목 정보, 카테고리 키워드를 이용해 친밀도가 높은 사용자들을 관심 정보별로 선정하여 그룹화 한다. 마지막으로 친밀도와 관심 정보를 토대로 그룹화 된 사용자들에게, 유사성이 강한 Facebook 페이지 내 콘텐츠 정보를 Tree map 형식의 어플리케이션 구조로 제공한다.

본 논문의 구성은 다음과 같다. 2장에서는 소셜 검색을 위해 제안된 순위 선정 기술 및 관련 연구에 대해 서술하고 3장에서는 친밀도와 공통 관심 항목을 기반으로 사용자들을 그룹화 하는 방법에 대해 기술한다. 4장에서는 친밀도와 공통 관심 항목으로 인해 그룹화 된 사용자들에게 Facebook 페이지 내의 대표 콘텐츠를 선정하여 제공하는 기법에 대해 기술하고 5장에서는 실험을 통해 앞서 제안된 방법들이 효과적임을 검증한다. 6장에서는 결론 및 향후 연구 방향을 제시한다.

2. 관련 연구

이 장에서는 SNS 상에서 콘텐츠를 효과적으로 제공하기 위한 콘텐츠 정보의 순위 선정에 관련된 연구에 대해 기술한다. 순위 선정 기술의 경우, 관계 기반의 방식과 콘텐츠 기반의 방식이 대표적이다[1].

2.1 관계에 기반 한 순위 선정

PageRank 알고리즘[2]은 콘텐츠간 우선 순위를 결정하거나, 콘텐츠를 이용하는 사람들간 관계를 정의하는 데 사용되는 대표적인 모델이다. PageRank 알고리즘을 기반으로 한 방식은 다양한 특성을 갖는 구조와 결합되어 연구가 진행되고 있다. Haveliwala[3]는 전통적인 PageRank 알고리즘에 사용되는 관계를 추론하는 벡터들의 특성을 사용하여 특정 주제의 중요도를 산정하는 방식을 제안하였다. 제안된

방식은 이슈 키워드와 사용자 질의의 중요도를 계산하여 가장 연관성이 높은 순위를 가지는 결과를 중요도 순으로 환산한다. J. Wang과 J. Liu, C. Wang[4]은 PageRank 알고리즘과 WordNet 방식이 적용된 시맨틱 검색을 이용하여 문서 내에 중요 키워드를 추출하는 방법을 제안하였다. 최근, SNS 이용자들이 제공하는 데이터 양이 급진적으로 증가하면서 PageRank 방식을 기반으로 빅 데이터 마이닝의 연산 시간을 줄이는 연구도 활발히 진행되고 있다. B. Bahmani과 K. Chakrabarti, D. Xin[5]은 방대한 양의 데이터들 중에서 PageRank가 가지는 벡터들의 특성을 Monte Carlo 방법에 의해 추론하였다. MapReduce 방법을 기반으로 하여 측정된 벡터 특성의 예측 및 연산시간은 기존의 기계학습 방법보다 뛰어난 성능을 보였다. 이러한 순위 결정 방법을 소셜 네트워크의 구조에 적용시키려는 시도도 있었다. 관계를 기반 한 방식의 경우, 사용자를 순위 결정의 대상으로 삼으며 자신과 가까운 사람이나 영향력이 큰 사람을 신뢰할 만한 정보원으로 검색하는 데 유용하다. PageRank 알고리즘을 SNS 구조에 적용한 PeopleRank 방식이 Mtibaae[6]등에 의해 제안되었다. PeopleRank는 페이지 링크가 갖는 속성들과 메시지 전달 구조 및 내용을 이용해 이웃간의 순위를 결정짓는 관계 구조를 갖는다. 순위 결정을 위한 사용자 간의 영향력과 친밀도를 계산하여 적용한 소셜 검색 엔진 Aardvark[7]등이 있다. Aardvark의 경우, 사용자간 프로필 등록정보의 유사성 비교, 메시지 유형 분석, 키워드 분석 등을 사용하여 친밀도를 계산하며 얼마나 사용자에게 영향력이 있는지에 대한 정보를 환산한다. 이를 통해 콘텐츠를 선정하기 위한 시스템에 효과적이라는 것이 실험을 통해 증명하였다.

2.2 콘텐츠에 기반 한 순위 선정

콘텐츠 기반의 순위 선정 기법에서는, 콘텐츠 내의 공통 키워드를 이용하여 웹 페이지의 순위를 결정한다. 순위 선정 기법이 반영된 대표적인 검색엔진으로 FolkRank[8]나 SocialSimRank[9], SNDocRank[10] 등이 있다. FolkRank는 PageRank 알고리즘을 폭소노미 환경에 적합하도록 변형시킨 형태를 지니고 있다. 태그 키워드 집합, 사용자 집합, 리소스 집합의 데이터들로 이루어지는 콘텐츠는 노드와 엣지의 형태로 변환하여 표현하였다. 이를 통해 관계를 정의하는 그래프를 나타내고, 각 노드와 엣지를 인접행렬로 산정하여 리소스 검색이나 리소스 추천, 태그 추천, 트렌드 추출 등에 활용한다. SocialSimRank의 경우, 웹 페이지 상의 태그와 콘텐츠가 포함된 웹 페이지, 사용자 질의 등의 정보를 토대로 사용자와 가장 유사도가 높은 콘텐츠의 순위를 환산한다. SNDocRank는 사용자에게 적합한 콘텐츠를 제공하기 위해 소셜 네트워크 기반으로 제안된 순위 선정 모델이다. 사용자 정보와 질의를 이용해 사용자와 관련된 웹상의 문서 정보를 도출하고 합산된 문서 순위를 보여준다. 콘텐츠에 기반 한 순위 선정 기술은 자신의 주변 친구들과 관심사를 고려한 문서 순위 결정에 유용하다.

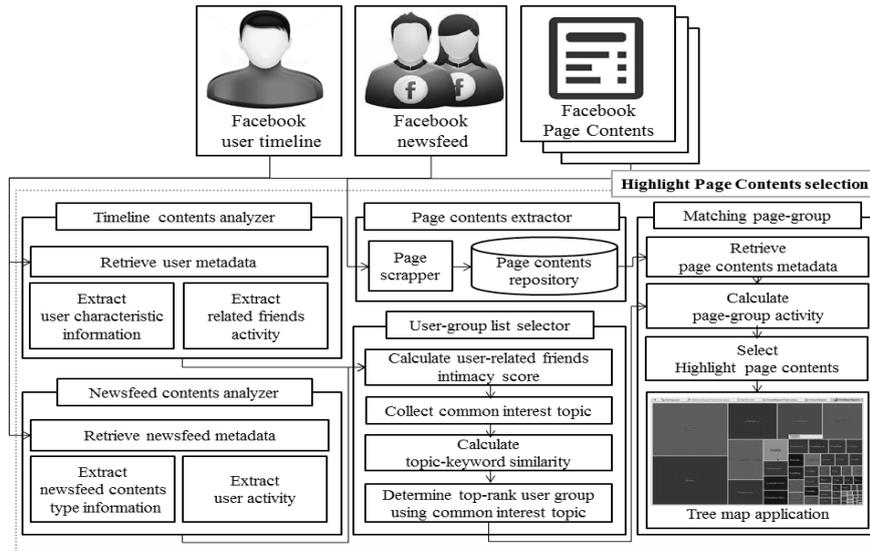


Fig. 1. System architecture

3. 친밀도 기반의 사용자 그룹 선정

이 장에서는 사용자 그룹 선정을 위해 Facebook에서 제공하는 타임라인과 뉴스피드를 이용하는 방법에 대해 기술한다. 개인 사용자와 관계를 형성한 사용자들 간의 관계를 이용하여 친밀도로 정의하는 방법과 친밀도로 정의된 관계를 이용해 사용자들을 관심항목별로 분류하는 방법에 대해 서술한다.

3.1 관계에 기반 한 순위 선정

사용자 그룹 선정을 위해 제안하는 전체 시스템은 Fig. 1과 같다. Facebook 사용자의 경우, 사용자 자신이 Facebook에 남긴 기록들과 관계를 형성하고 있는 다른 사용자들이 자신에게 남긴 기록, 사용자의 정보를 열람할 수 있는 타임라인과 관계를 형성한 사용자들의 기록을 열람할 수 있는 뉴스피드로 나뉜다. 타임라인과 뉴스피드로 분류된 페이지에서 Facebook 개인 사용자는 본인의 게시물이나 관계를 형성한 사용자의 기록물에 대해 관심을 표현하는 기능인 'Like', 'Comment', 'Tag', 'Share' 등의 행동을 취할 수 있으며 이를 활동(Activity)이라 정의한다.

실제 SNS 상에서 사용자는 관심이 있는 게시물을 자신의 영역에 공유하거나 댓글을 다는 행동을 취한다[12]. 특히, Facebook 내에서는 게시물을 공유하거나 댓글을 다는 행동을 관계를 형성한 사용자들끼리 표현하는 Facebook만의 구조를 가지고 있다. 앞서 정의한 활동을 취하기 위해선, 기본적으로 사용자간의 '친구' 관계가 형성되었다고 가정한다. '친구' 관계가 형성되면, 활동을 통해 서로의 게시물에 대해 관심을 표현할 수 있으며, 이에 따른 관심 정도를 친밀도라고 정의할 수 있다. 일반적으로, SNS 상에서 사용자나 문서 간 연결 관계를 그래프 형태로 표현하고 있다. 그래프 $G=(V, E)$ 에서 G 는 소셜 네트워크, V 는 사용자나 문서로

대표되는 노드들의 집합을 의미하며, E 는 노드들의 연결 강도를 나타낸다. PeopleRank 및 SONAR[11]등은 이러한 그래프 모델을 바탕으로 제안되었으며, 사용자 간 관계 강도인 E 를 SNS 플랫폼에 따라 선정하는 연구가 진행되고 있다. 특히, SONAR에서는 식 (1)과 같은 방식으로 E 의 값을 산정하였다.

$$Score_i = \frac{\sum_{y=1}^{|users|} (Score_{iy} \times R_{xy})}{\sum_{y=1}^{|users|} R_{xy}} \quad (1)$$

식 (1)에서 $Score_{iy}$ 는 사용자 y 가 관심 있는 콘텐츠 i 를 의미하며, R_{xy} 는 사용자 x 와 사용자 y 의 연결 강도를 의미한다. 본 논문에서는 Facebook만이 가질 수 있는 특수한 계시물 열람 구조와 순위 선정 모델에 근거하여, 관계를 형성한 사용자들이 개인 사용자의 타임라인에 남긴 활동을 'Timeline-Friends-Activity'로 정의하고 뉴스피드의 노출된 게시물들에 대해 사용자가 남긴 활동들을 'Newsfeed-User-Activity'으로 정의한다. Fig. 1에 제시된 구조에서, Timeline contents analyzer와 Newsfeed contents analyzer를 3.2와 3.3에서 설명한다. 앞서 활동으로 정의된 정보를 이용해 '친구' 관계를 형성한 사용자들 간의 친밀도, 즉, 그래프 모델 상의 E 값을 추론할 수 있으며, 또한 같은 관심사와 친밀도로 연관된 사용자들의 그룹 리스트를 획득할 수 있다. 이를 Fig. 1에서 User-group list selector라 하여 3.4에 설명한다. 획득한 정보로 Facebook 내의 사용자 등록정보와 공통 관심 항목을 추출하여, 공통 관심 항목별로 사용자를 선정하여 그룹화한다. Facebook이 제공하는 페이지콘텐츠를 특정 관심 정보로 그룹화 된 사용자에게 Tree map을 이용하여 효과적으로 보여준다.

3.2 Timeline-Friends-Activity

앞서 언급한 Timeline-Friends-Activity을 기반으로 Facebook 개인 사용자 x 의 타임라인에 등록된 게시물에 대해 사용자 y_i 가 행한 활동들을 분석하면 그들 간 친밀도를 정의할 수 있다. 본 논문에서 고려하는 활동 타입의 종류와 그 비중은 Table 1과 같다. 이를 기반으로 친밀도 점수 RT_x 은 식 (2)에 따라 정의할 수 있다.

$$RT_x(y_i) = \frac{\sum_{a_i \in A_i} \sum_{d_j \in D} cnt(a_i, d_j, u_i) \times AW(a_i) \times \frac{AT(d_j, y_i)}{T}}{\sum_{u_i \in U} \sum_{a_i \in A_i} \sum_{d_j \in D} cnt(a_i, d_j, u_i) \times AW(a_i)} \quad (2)$$

Table 1. Activities on the timeline

Activity type(a_i)	Explanation	Activity Weight
Like	y_i performed 'Like' action to user x 's timeline contents d_j	0.3
Comment	y_i left comments on user x 's timeline contents d_j	0.6
Publish	y_i published contents to user x 's timeline	0.5
Tagging	y_i tagged user x	0.5
Share	y_i shared contents to user x 's timeline	0.6

A_i 는 Timeline-Friends-Activity의 y_i 가 x 의 게시물 d_j 에 대해 취할 수 있는 전체 활동타입들을 나타내며 a_i 는 A_i 에 속한 하나의 활동을 의미한다. cnt 는 x 가 등록한 게시물 d_j 에 대해 관계를 형성한 사용자 y_i 가 행한 활동 타입 a_i 의 횟수를 나타낸다. 또한, u_i 는 활동을 취한 특정 사용자이며, 이는 전체 집합 U 로 나뉜다. x 의 타임라인에 등록된 전체 게시물의 집합을 D 라고 하며, d_j 는 그 중 하나의 게시물을 의미한다. y_i 는 x 와 관계를 형성한 i 번째 사용자를 나타내며 Y 는 관계를 형성한 전체 사용자 집합을 나타낸다. 또한 개인 사용자 활동 목록에 포함된 각 활동은 나름의 가중치 (AW , Activity weight)를 가지며, 친밀도 점수 RT_x 은 시간 정보에 따른 시간 가중치 값(AT , Activity time)을 내포한다. $AT(d_j, y_i)$ 는 y_i 가 d_j 에 활동을 수행한 시간을 나타낸다. 활동 정보 가중치는 Facebook 내에서 x 의 타임라인 게시물 d_j 에 y_i 가 취할 수 있는 모든 형태의 활동에 따라 부여하였다. 사용된 가중치 값은 기존의 콘텐츠 추천 시스템의 가중치 값을 Facebook 상의 친밀도 계산에 맞게 변환하였다[11]. 시간 가중치 값은 실험을 진행한 전체 기간 T 내에 y_i 가 타임라인에 활동을 취해준 시간인 AT 에 비례하여 값을 산정하였다. 최근 활동을 취해준 사용자들에게는 높은 가중치를 부여하여 시간 가중치에 따른 친밀도 점수를 산정하였다. Table 1에 표기된 활동 정보와 이를 통해 x 와 y_i 간 시간의

흐름에 연관된 친밀도 점수 RT_x 을 계산할 수 있다.

3.3 Newsfeed-User-Activity

Newsfeed-User-Activity은 사용자 x 가 열람할 수 있는 뉴스피드에서, y_i 가 등록한 게시물 p_i 에 취한 활동에 대한 정보를 나타낸다. p_i 에 취한 x 의 활동 타입 a_n 은 Table 2와 같다. 사용자는 Table 2를 통해 얻을 수 있는 정보를 기반으로 사용자 x 와 y_i 간의 뉴스피드에서의 친밀도 점수 RN_x 을 식 (3)에 의해 산출 할 수 있다.

$$RN_x(y_i) = \frac{\sum_{a_n \in A_n} \sum_{p_i \in P} cnt(a_n, u_i, p_i) \times AW(a_n) \times \frac{AT(x, p_i)}{T}}{\sum_{u_i \in U} \sum_{a_n \in A_n} \sum_{p_i \in P} cnt(a_n, u_i, p_i) \times AW(a_n)} \quad (3)$$

A_n 는 Newsfeed-User-Activity의 x 가 y_i 의 게시물 p_i 에 대해 취할 수 있는 활동들의 집합이며 a_n 는 집합에 포함된 하나의 활동 정보를 의미한다. cnt 는 y_i 의 게시물 p_i 에 대한 사용자 x 의 활동 a_n 의 횟수를 나타낸다. Timeline-Friends-Activity과 같이, 활동 정보와 각각의 활동 정보 가중치 AW , 그리고 시간정보에 의존한 시간 가중치인 AT 에 의해 친밀도 점수인 RN_x 의 정보를 계산할 수 있다. $AT(x, p_i)$ 은 x 가 뉴스피드 내 y_i 의 게시물 p_i 에 활동을 수행한 시간이다. 활동 정보 가중치는 x 가 y_i 에게 취할 수 있는 모든 형태의 활동에 따라 부여하였으며, 시간 가중치 값은 앞에서와 유사한 방법으로 산정하였다.

Table 2. Activities on the newsfeed

Activity type(a_n)	Explanation	Activity Weight
Like	x performed 'Like' action to newsfeed contents p_i	0.3
Comment	x left comments on newsfeed contents p_i	0.6
Tagging	x tagged user y_i in p_i	0.5
Share	x shared y_i 's newsfeed contents p_i to user x 's timeline	0.6

3.4 공통 관심 항목을 이용한 유사 사용자 선정

Timeline과 Newsfeed를 통해 획득한 친밀도를 기반으로, 사용자 x 와 y_i 간 공통 관심 항목을 선출한 후 유사도 측정 방법을 적용한다. Facebook이 제공하는 사용자들의 기본 정보에서, 사용자에게 의해 설정된 관심사의 속성 정보들을 토대로 x 와 y_i 간 관심 항목의 집합을 I_x, I_{y_i} 로 나타내면 그들의 공통 관심 항목 C 는 아래의 식 (4)와 같다.

$$C \in (I_x \cap I_{y_i}) \quad (4)$$

C 에 대한 x 의 키워드를 수집하여 키워드 그룹 w_x 로 표현한다. 같은 방법으로, y_i 의 C 에 관련된 키워드를 수집하여 키워드 그룹인 w_{y_i} 로 나타낸다. 키워드 그룹인 w_x 와 w_{y_i} 는 숫자로 표현하지 않고 문자로 표현 가능하기 때문에, 비교를 위해서 이진화 속성 매칭이 아닌 문서-속성간의 유사도 측정 방법인 Cosine Similarity를 이용한다. 유사도 측정에서 두 개체가 닮은 정도에 대한 수치적인 척도를 계산할 수 있으며, 비교하는 대상의 개체가 닮을수록 유사도가 1에 가까워지며 반대로 비교하는 대상의 개체가 다를수록 유사도가 0에 가까워진다.

$$S_T(x, y_i) = \frac{w_x \cdot w_{y_i}}{\|w_x\| \|w_{y_i}\|} \times \text{Norm}(RT_x(y_i) + RN_x(y_i)) \quad (5)$$

유사도 측정을 위해, 키워드 그룹 w_x 와 w_{y_i} , 그리고 총 비교할 대상 벡터들의 정보를 이용한다. 여기서 w_x 와 w_{y_i} 는 ‘Musician’의 경우 {Eric’, ‘SWV’, ‘LSG’}, {‘Eric’, ‘Kim’, ‘Evan’} 등과 같은 예로 선출된다.

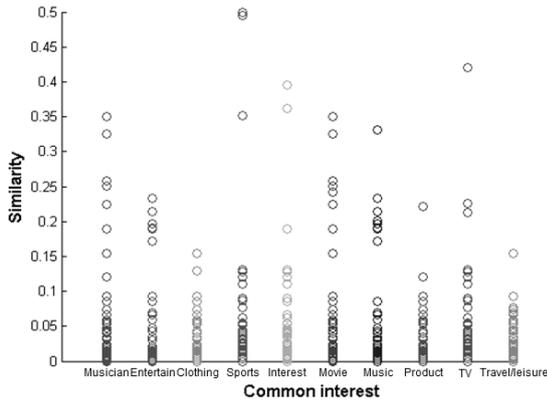


Fig. 2. Similarity measurement of common interest

식 (2), (3)에서 계산된 RT_x 와 RN_x 의 합이 최소 0, 최대 1로 정규화 시킨 값을 가중치로 계산해 줌에 따라 친밀도가 반영된 관심 분야 유사도 측정이 가능하다. 즉, 식 (5)에서 x 와 y_i 간 관심 분야 그룹을 결정짓는 유사도를 결정지을 수 있다. Fig. 2는 식 (5)에 의해 계산된 유사도 결과를 산점도 형태로 보여주고 있다. 가로 축에 나타난 정보는 공통 관심 항목, 즉 ‘Musician’, ‘Clothing’, ‘Sports’, ‘Movie’ 등이 될 수 있다. 세로 축은 개인 사용자와 관계를 형성한 사용자들 간의 키워드 그룹 유사도를 나타내며, 분포된 점은 관계를 형성한 사용자들을 나타낸다. 나타난 그림에 의하면, x 와 친밀도 점수에 기반하여 관심 정보가 유사한 사용자들이 높은 유사도 점수를 획득하였다. 또한, 유사도가 높은 순의 상위 15명을 선정해 최상 순위 그룹으로 정의한다. 최상 순위 그룹은 친밀도가 반영된 관심 분야에 다른 사용자들보다 유사한 경향을 많이 띄고 있다고 추론 할 수 있다.

4. Facebook 페이지 순위 및 콘텐츠 선정

4.1 Facebook 페이지

앞서 기술한 바와 같이, Facebook은 Facebook 페이지라는 특수한 웹 페이지 형식을 가지고 있다. 페이지는 회사, 조직 및 브랜드가 자신들의 소식을 공유하고 사람들과 연결할 수 있도록 도와준다. Timeline과 마찬가지로, 앱을 추가하거나 이벤트를 여는 등의 활동을 행할 수 있고 이러한 페이지의 소식을 주기적으로 뉴스피드에서 업데이트 받을 수 있다[14]. Facebook 페이지는 특정 관심사의 정보를 공유하는 공간이 될 수 있으며, 회사, 조직 및 브랜드의 정보를 공유하는 공간이 될 수 있다. 3.4에서 제시된 관심 그룹 사용자와 공통 관심 항목, 친밀도 정보를 이용하여 관심 그룹에 가장 유사한 페이지와 해당 페이지 콘텐츠의 순위를 환산한다.

4.2 Page-Group-Activity를 통한 콘텐츠 선정방법

콘텐츠 선정을 위해 사용자들의 기본 정보에 등록된 각 페이지 이름과 페이지의 프로필 정보, 페이지 내등록된 게시물에 대한 정보를 수집한다. 수집된 페이지 내 콘텐츠 집합 $G = \{g_1, g_2, \dots, g_m\}$ 이라 정의하며, 식 (4), (5)에 의해 선정된 최상 순위 사용자 $TR = \{tr_1, tr_2, \dots, tr_f\}$ 을 정의한다. 페이지 내의 콘텐츠에 취할 수 있는 최상위 사용자들의 활동을 Table 3에 제시했다. 페이지 내 콘텐츠 G 에 최상 순위 그룹 사용자들 TR 이 취한 활동 값은 식 (6)과 같이 정의할 수 있다. 식 (6)에서 S 는 Page-Group-Activity의 활동들의 집합이며 s_n 는 집합에 포함된 하나의 활동 정보를 의미한다. cnt 는 연관된 페이지 게시물 g_m 에 대한 최상 순위 그룹 사용자 tr_f 의 활동 s_n 을 나타낸다.

$$RP_{TR}(g_m) = \frac{\sum_{s_n \in S} \sum_{g_m \in G} cnt(s_n, tr_f, g_m) \times AW(s_n)}{\sum_{tr_f \in TR} \sum_{s_n \in S} \sum_{g_m \in G} cnt(s_n, tr_f, g_m) \times AW(s_n)} \quad (6)$$

Table 3. Activities on the page contents

Activity type(s_n)	Explanation	Activity Weight
Like	tr_f performed ‘Like’ action to page contents g_m	0.3
Comment	tr_f left comments on page contents g_m	0.6
Tagging	tr_f tagged certain user tr_f on page contents g_m	0.5
Share	tr_f shared page contents g_m to certain user tr_f	0.6

계산된 RP_{TR} 값에 의해, 페이지 내 콘텐츠 g_m 의 순위점수를 환산할 수 있으며, 이를 토대로 최상위 사용자들이 유용하다고 판단할 수 있는 Facebook 페이지 내의 콘텐츠의 순위를 결정할 수 있다. 높은 순위 점수를 받은 페이지 내 콘텐츠를 “Highlight Page Contents”라 하며 다른 관심 항목간의 점수를 비교하여 높은 점수의 페이지 콘텐츠를 우선적으로 선정하게 된다.

4.3 선택된 대표 콘텐츠의 시각화 방법

앞에서, 공통 관심 항목으로 생성된 최상위 사용자 그룹과 Facebook 페이지 간의 활동 정보를 획득하였다. 예를 들면, 최상위 순위 그룹 사용자들이 활발히 활동을 취한 Facebook 페이지 내 콘텐츠는 최상위 그룹 사용자들에게 신뢰성 있고 유용한 정보를 제공한다. 이를 효과적으로 나타내기 위해 Tree map[13]을 활용한다. Tree map은 신뢰성 있고 유용한 정보를 최상위 그룹 사용자들에게 효과적으로 나열하는 데 유용하다. Tree map은 전체 사이즈가 1로 고정된 화면상에 콘텐츠의 중요도에 따라 화면을 분할하여 나타낸다. 예를 들면 ‘A’의 콘텐츠의 점수가 0.3이면 전체 사이즈가 1인 화면에 30 %의 비율로 화면상에 나타내어진다. Facebook 페이지 대표 콘텐츠의 효과적인 시각화를 위해, 페이지 내 콘텐츠 g_m 에 최상위 순위 그룹 사용자들 TR 의 활동 값 RP_{TR} 를 0에서 1사이로 정규화 한다.

$$Z = \frac{A - E[A]}{\sigma(A)} \tag{7}$$

A 는 페이지 내 콘텐츠 g_m 에 취한 하나의 활동 값을 나타내며, $E[A]$ 는 페이지 내의 전체 콘텐츠 g_m 에 취한 활동의 평균을 나타낸다. 식 (7)을 통해 정규화 된 값 Z 값을 계산한다. Z 는 최소 값 0과 최대 값 1을 가지며 값이 클수록 사용자에게 큰 화면의 비율로 선택된 콘텐츠를 웹 어플리케이션 상에 나타낸다.

5. 실험 및 결과

5.1 실험 데이터

실험은 Facebook 사용자의 타임라인 기록물과 사용자의 기본 등록 정보 및 관계를 형성한 사용자들을 대상으로 진행하였다. 타임라인에 등록된 게시물은 약 400개, 관계를 형성한 사용자들은 300명이며 활동 기록에 대한 모든 정보를 수집하였다. Facebook page는 약 271개의 그룹과 관련된 페이지 내 콘텐츠 정보를 일주일 기간 동안 수집하여 사용하였다.

5.2 친밀도 계산 결과

Timeline-Friends-Activity에 의해 구해진 친밀도와 Newsfeed-User-Activity에 의해 구해진 친밀도를 각각 Fig.

3과 Fig. 4에 나타내었다. 그림에서 ‘Lee2’의 경우 x 에게 가장 많은 활동을 하였다. Timeline-Friends-Activity 방법에 의한 친밀도 선정을 위해 x 의 게시물 정보를 활용하였다. Newsfeed-User-Activity의 친밀도 점수는 최근 뉴스피드 항목에 노출된 약 1500개의 무작위 게시물에 대해 x 의 뉴스피드에 취한 활동 기록을 기반으로 계산하였다. Newsfeed-User-Activity의 경우 마찬가지로, ‘Lee2’의 경우 개인 사용자가 ‘Lee2’가 뉴스피드에 남긴 콘텐츠에 많은 관심을 보이는 것으로 나타났다.

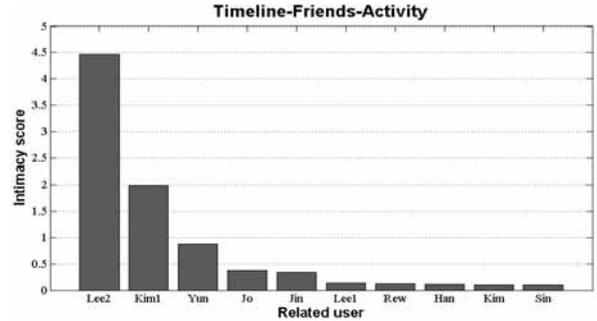


Fig. 3. Intimacy score based on Timeline-Friends-Activity methods

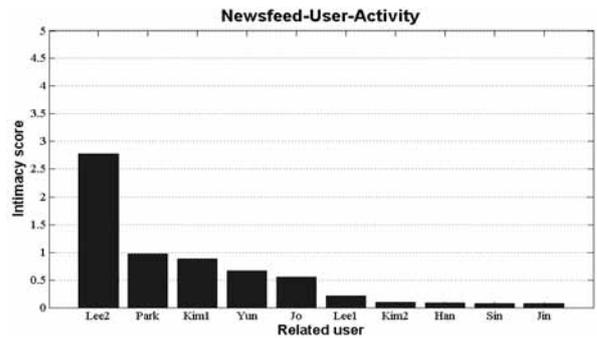


Fig. 4. Intimacy score based on Newsfeed-User-Activity methods

5.3 사용자 그룹 선정 결과

결정된 친밀도 점수를 기반으로 한 그룹은 사용자 y_i 의 특정 성향과 서로 밀접한 연관관계를 가지며, 이는 뉴스피드의 노출된 정보에 신뢰성을 가진다. x 에 설정된 관심 정보, 즉 관심 분야 정보에 대해 신뢰성을 가지는 최상위 그룹을 친밀도에 의해 선별할 수 있다. 친밀도 점수와 사용자 그룹 선정 항목의 유사도 측정에 의해 선별된 그룹의 예는 Table 4와 같다. 친밀도가 반영된 유사도 평균은 최상위 그룹의 공통 관심 항목 C 에 대해 식 (5)로 계산된 평균을 기록하였다.

Table 4. Example of user rank group

Rank Group	Top-rank Group User	Similarity
‘Musician’	Lee2, Park, Jin, Sin, Jo etc.	0.2151
‘Clothing’	Park, Kim1, Lee1, Han, Sin etc.	0.1877
‘Entertainer’	Jin, Lee2, Kim1, Lee1, Han 외 etc.	0.1046

Table 5은 친밀도에 의한 사용자 선정의 성능 평가를 위해 Naive Bayes와 k-nearest neighbor 기법과 비교하고 있는데 공통 관심항목별로 선별된 결과의 정확도를 보이고 있다. 친밀도를 반영한 그룹 선정 결과인 F-measure 값이 가장 높게 나타남을 실험을 통해 확인하였다.

Table 5. Comparison of grouping methods

Methods	Precision	F-measure
Naive Bayes	0.56	0.513
k-nearest neighbor	0.58	0.536
User grouping based on Facebook intimacy	0.62	0.647

5.4 Facebook 페이지 대표 콘텐츠 선정 결과

개인 사용자는 공통 관심 항목과 친밀도에 따라 최상위 그룹 사용자로 설정하고, 연관성이 있는 Facebook 페이지를 순위정보를 이용해 선정하였다. Table 6에서 하나의 정보는 페이지 분류 내에 해당하는 각 페이지 이름을 가리키며, 괄호 안의 숫자는 페이지 내의 해당 “Highlight Page Contents”와 식 (7)에 의해 환산된 값을 나타낸다. 공통 관심 항목 ‘Sports’로 그룹화 된 최상 순위 그룹의 활동 정보를 토대로 매긴 순위점수 결과는 ‘Wanye Rooney’, ‘Manchester United’, ‘Arsenal’ 이었으며, 해외 축구에 관련된 페이지의 정보가 대다수였다. 특히 실험데이터가 수집된 기간 내에, 관련된 페이지의 소식 업데이트가 활발하였고, 많은 사용자들이 ‘Like’, ‘Tag’, ‘Comment’의 활동을 함에 따라 높은 순위로 매겨질 수 있는 점수를 획득하였다. 또한 Facebook 구조상에 이와 같은 활동 정보를 통해, 정보의 확산성이 커짐에 따라, 그룹에 관련된 사용자들 외에 많은 사람들이 정보를 획득하고 많은 활동을 취하였다. Fig. 5에서는 Table 6의 정보를 이용하여 Tree map 형식의 웹 어플리케이션을 통해 효과적으로 사용자에게 제공하였다. ‘Wanye Rooney’의 Facebook 페이지에서 가장 높은 점수를 획득한 15번째로 게시된 콘텐츠가 식 (7)에 의해 0.301로 환산되었다. 따라서 전체가 1인 크기에서 30%에 가까운 부분을 이미지와 페이지 내 콘텐츠 제목을 통해 나타내었다.

Table 6. Highlight page contents using page selection

Common interest ‘Sports’			
Highlight Page Contents	‘Wanye Rooney’ (15, 0.301)	‘Manchester United’ (3, 0.070)	‘Arsenal’ (18, 0.022)
Common interest ‘Music’			
Highlight Page Contents	‘Snoop Dogg’ (12, 0.098)	‘2pac’ (10, 0.075)	‘Eric’ (41, 0.068)
Common interest ‘Travel’			
Highlight Page Contents	‘Travel Lovers’ (8, 0.157)	‘I love Travel’ (24, 0.075)	‘Filson’ (11, 0.028)



Fig. 5. Highlight page contents visualization

5.5 정보 연관성 평가

아래의 Fig. 6은 친밀도가 주는 정보의 신뢰성을 평가하기 위해 최상 순위 그룹 사용자와 비순위 그룹 사용자를 대상으로 한 정보 제공의 연관성 실험을 진행하였다. 비순위 그룹은, 최상 순위 그룹 사용자와 같은 관심 분야 주제를 가지는 임의의 Facebook 사용자를 최상위 그룹 사용자의 수만큼 선정하여 비순위 그룹 사용자로 설정하였다. 비순위 그룹은 Un-rank 그룹이라 하며, 최상 위 그룹은 Top-rank 라 명명하였다. 연관성 점수는 5점에 근접할수록 관심 분야 주제와 일치하는 항목이 나타나는 비율이 높으며, 0점에 가까워질수록 그 비율이 낮아진다. 연관성 점수는 개인 사용자의 관심 항목을 10개를 선정하고 관심 항목과 연관된 전체 키워드의 전체 수에서, 최상위 그룹과 비순위 그룹에서 선별된 키워드의 비율로 점수를 측정하였다. 실험 결과로 나타난 비순위 그룹 사용자의 평균 연관성 점수는 2.223, 최상 순위그룹 사용자의 평균 연관성 점수는 2.99로 측정되었다. ‘Product’와 ‘Travel/leisure’의 경우, 관심 분야 주제에 대해 집중적으로 내용을 언급하는 비순위 그룹으로 사용자가 많이 포함 될수록 최상 순위의 연관성 점수보다 높게 나타나는 경우가 발생하였다. 하지만 이를 제외한 나머지 경우, 관심 분야 주제에 대해 최상 순위 그룹의 연관성 점수가 평균적으로 높음을 확인 할 수 있다. 즉, 최상 순위 그룹 사용자가 제공할 수 있는 정보는 비순위 그룹 사용자가 제

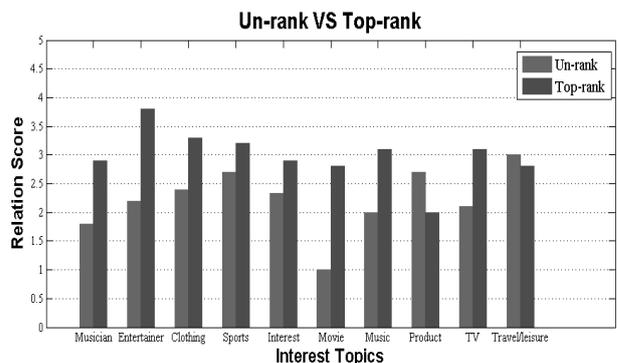


Fig. 6. Un-rank and Top-rank relation score

공하는 정보보다 관심 분야 주제와 일치하는 정보를 약 25%정도 더 내포한다. 이는 친밀도 점수가 반영된 유사도에 따라 특정한 관심 분야로 그룹화 된 최상위 사용자들이 주는 정보가 효율적임을 나타낸다.

6. 결 론

본 논문에서는 Facebook에서의 친밀도를 이용하여 Facebook 페이지의 콘텐츠를 효과적으로 선정하여 사용자에게 제공하는 방법을 제안하였다. 기본적인 사용자 정보와 현재 사용자의 관심 정보와 관계를 형성된 사용자간의 친밀도 점수를 실시간으로 반영하여 문서 검색의 신뢰성을 높이고, 정의된 순위 산정 방법에 의해 선정된 Facebook 페이지 콘텐츠를 효과적으로 가시화 하였다. 또한 관심 분야 별로 생성된 최상위 그룹 목록은 신뢰성 있는 정보를 제시함을 실험을 기반으로 제시하였고, 이를 토대로 Facebook 페이지 내의 콘텐츠 정보를 효과적으로 선정하여 제공하는 것이 가능함을 보였다. 하지만, 실제 사용자들이 Facebook의 페이지 콘텐츠를 이용하는 경우가 적거나, 관련된 사용자의 메타데이터를 수집하는 과정에서의 보안 설정 및 공개 여부 정책으로 실험을 진행하는 데 제약조건이 되었다. 따라서 향후 연구과제로 실험 데이터의 보안적인 측면을 고려하는 방안이 필요하다.

참 고 문 헌

[1] 박병선, 그래프서치(Graph Search)를 통해 본 소셜검색의 발전과 시사점, 방송통신정책, 제25권 3호, 2013.

[2] A. N. Lagnvile, C. D. Meyer, "Google's PageRank and beyond: The science of search engine rankings," Princeton University Press, 2006.

[3] H. T. Haveliwala, "Topic-sensitive pagerank," *Proceedings of the 11th conference on World Wide Web*, New York, pp. 517-516, 2002.

[4] J. Wang, J. Liu, C. Wang, "Keyword extraction based on pagerank," *Advanced in Knowledge Discovery and Data Mining*, Nanjing, pp.857-864, 2007.

[5] B. Bahmani et. al, "Fast personalized pagerank on mapreduce," *Proceeding of the 2011 ACM SIGMOD International Conference on Management of data*. New York, pp.973-984, 2011.

[6] M. Abderrahmen et al, "Peoplerank: social opportunistic forwarding," *In INFOCOM, 2010 Proceedings IEEE*, San Diego, pp.1-5, 2010.

[7] D. Horowitz and S.D. Kamvar, "The Anatomy of a LargeScale Social Search Engine," *Proceedings of 19th international conference on World Wide Web*, Raleigh, pp.431-440, 2010.

[8] J. Robert et al, "Tag recommendations in folksonomies," *Knowledge Discovery in Databases: PKDD 2007*, Warsaw, pp.506-514, 2007.

[9] B. Shenghua et al, "Optimizing web search using social annotations," *Proceedings of the 16th international conference on World Wide Web*, Banff, pp.501-510, 2007.

[10] G. Liang et al, "SNDocRank: document ranking based on social networks," *Proceedings of the 19th international conference on World Wide Web*, Raleigh, pp.1103-1104, 2010.

[11] F. Carmagnola et al, "Sonars: A social networks-based algorithm for social recommender systems," *User Modeling, Adaptation, and Personalization*, pp.223-234, 2009.

[12] A. Lambert, "Intimacy and friendship on Facebook," Palgrave Macmillan, 2013.

[13] B. Shneiderman, "Tree visualization with tree-maps: 2-d space-filling approach," *ACM Transactions on Graphics (TOG)*, Vol.11, Issue 1, pp.92-99, 1992.

[14] Facebook page, www.facebook.com/help, 2013.11.



유 세 혁

e-mail : rjh1026@korea.ac.kr

2012년 경희대학교 전자정보학부(학사)

2012년~현 재 고려대학교 전기전자전파
공학과 석·박사 통합과정

관심분야 : 바이오 인포메틱스, 영상처리,
소셜 네트워크 데이터 분석



최 영 환

e-mail : work48@korea.ac.kr

2007년 고려대학교 전자공학과(학사)

2012년~현 재 고려대학교 전기전자전파
공학과 석·박사 통합과정

관심분야 : 바이오 인포메틱스, 영상처리,
영상분할



황 인 준

e-mail : ehwang04@korea.ac.kr

1988년 서울대학교 컴퓨터공학과(학사)

1990년 서울대학교 컴퓨터공학과(석사)

1998년 Univ. of Maryland at College
Park 전산학과(박사)

1998~1999년 Bowie State Univ., Assistant
Professor

1999년~1999년 Hughes Research Lab. 연구교수

1999년~2004년 아주대학교 정보통신전문대학원 조/부교수

2004년~현 재 고려대학교 전기전자전파공학과 교수

관심분야 : 데이터베이스, 멀티미디어 검색, 빅 데이터 처리,
소셜 네트워크 데이터 분석