

소셜 카테고리를 이용한 추천 방법[☆]

Social Category based Recommendation Method

유 소 엽¹ 정 옥 란²
So-Yeop Yoo Ok-Ran Jeong

요 약

최근 SNS가 이슈가 되고 다양한 분야에서 이를 이용한 연구가 활발하게 진행되고 있다. 특히 SNS 상에서 생성되는 여러 소셜 데이터를 기반으로 사용자의 관심사를 찾아내고 추천해 주는 시스템에 대한 연구가 대두되고 있다. 사용자의 관심과 선호도는 단순히 사용자가 작성한 글에서만 나타나는 것이 아니라, 친구와의 관계와 작성한 내용기반으로 분류되는 카테고리를 이용하여 파악될 수 있다. 본 논문에서는 사용자의 사회적 관계와 사용자가 작성한 소셜 데이터의 카테고리를 이용하여 사용자의 선호도를 자동으로 추출하고, 이를 기반으로 추천하는 방법을 제안한다. 그리고 실험을 통해 제안한 기법의 유효성을 검증한다.

☞ 주제어 : 추천, 사용자 관심, 사회적 관계, 소셜 카테고리, 소셜 네트워크 서비스

ABSTRACT

SNS becomes a recent issue, and many researches in various kinds of field are being done by taking advantage of it. Especially, there are many researches existed on the system that finds user's interest and makes recommendation based on multiple social data generated on the SNS. User's interest is not only revealed from the user's writing but also from the user's relationship with friends. This study proposes a recommendation method that extracts user's interest by using social relationship and its categorization applies it to the recommendation. In this way, it can recommend user's interest with category based on the writings by the user and furthermore it can apply the user's relationship with his/her friends for more accurate recommendation. In addition, if necessary, the recommendation can be made by extracting any interest shared between the user and specific friends. Through experiments, we show that our method using social category can produce satisfactory result.

☞ keyword : Recommendation, User Interest, Social Relationship, Social Category, SNS

1. 서 론

최근 트위터(Twitter)[1], 페이스북(Facebook)[2]과 같은 온라인상에서 관계를 기반으로 이루어지는 SNS(Social Network Service)가 생겨나고, 많은 사람들이 이용하고 있다. 특히 언제 어디서나 접근이 가능한 스마트폰의 보급으로 인해 수많은 정보의 공유가 SNS 내에서 실시간으로 이루어지고 있다 [3-6].

SNS에서 공유되는 정보는 그 종류 또한 다양하다. 친구 관계뿐만 아니라 텍스트, 사진, 동영상까지 많은 종류의 정보가 사용자들에 의해 만들어지고 공유된다. 또한 수많은 사람들이 실시간으로 정보를 만들고 공유하기 때문에 정보가 실시간으로 업로드되는 뉴스피드는 굉장히 빠른 속도로 업데이트가 된다[5].

일반적으로 사용자들은 SNS상에서 관계를 맺은 모든 친구들과 동일한 관심사를 공유하기도 하지만 친구에 따라서 각각 서로 다른 특정한 관심사를 공유하기도 한다. 그렇기 때문에 친구 전체와의 관계를 이용하여 '나'의 가장 큰 관심사를 찾아내는 것도 중요하지만, 특정 친구와의 관심사를 찾아내는 것 또한 충분히 의미 있는 일이 될 수 있다. 단순히 전체적인 사용자의 관심사(User Interest)만 추출하는 것이 아니라 특정 인물과의 관계와 관심사도 추출할 수 있다면 이를 이용하여 사용자에게 보다 세밀하고 적합한 추천이 가능하다.

¹ Dept. of Software Design & Management, Gachon Univ., Seongnam, 461-701, Korea

² Dept. of Software Design & Management, Gachon Univ., Seongnam, 461-701, Korea

* Corresponding author (orjeong@gachon.ac.kr)

[Received 21 August 2014, Reviewed 05 September 2014, Accepted 23 September 2014]

☆ 이 논문은 2014년도 가천대학교 교내연구비 지원에 의한 결과이며, 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단의 기초연구사업지원을 받아 수행한 것임. (NRF-2013R1A1A3A04008339)

본 연구에서는 사용자의 SNS상에서의 관계, 즉 사회적 관계(Social Relationship)를 분석하고, 사용자가 작성한 글이나 댓글 등의 활동(Action)을 통해 카테고리를 분석하여 사용자의 전체적인 관심사와 특정 친구와 공유하는 선호도를 자동으로 추출한다. 이를 기반으로 사용자의 관심사를 추천하는 방법을 제안한다.

기존의 연구에서 사용자의 선호도 추출 방법과 이를 이용한 추천 방법들이 많이 연구되어 왔지만, 본 연구에서는 사회적 관계와 사용자의 활동을 통한 소셜 카테고리 분석을 이용하였다. 여기에서 우리는 SNS특성을 반영한 카테고리를 소셜 카테고리라고 정의하였다. 그에 대한 내용은 3장에서 자세하게 설명한다.

2. 선행 연구

인터넷의 급격한 발전과 뛰어난 접근성을 갖고 있는 스마트폰의 보급의 확산으로 온라인상에서 관계를 기반으로 이루어지는 SNS 사이트 역시 활발하게 이용되고 있다. SNS 사이트의 확산과 동시에 그 사용자 수 또한 급격하게 증가하면서 수많은 사회적 관계와 데이터가 발생하게 됐다.

이러한 SNS 사이트의 발전으로 사회적 관계와 텍스트 등과 같이 사용자에게 의해 만들어진 데이터가 수없이 많이 만들어 지고 있다. 그에 따라 오프라인이 아닌 온라인 상에서 이루어지고 만들어지는 사용자의 관심사와 활동 패턴을 파악하기 위한 다양한 연구들[6-11]이 진행되고 있다.

최근 활발하게 이루어지고 있는 SNS, 특히 그 중에서도 트위터나 페이스북과 같은 SNS에 대한 다양한 연구들 중 사용자의 관심사를 찾아내는 연구가 굉장히 많이 존재한다. 이는 사용자의 관심사를 파악하는 것이 많은 영역에서 실용적으로 활용이 가능하기 때문이다. 특히 사용자의 관심사를 파악하여 그에 부합하는 상품을 추천해 줌으로써 사용자의 흥미를 높이고 구매욕을 상승시키는 등 기업의 마케팅에 있어 적극적 활용이 가능하다. 이러한 활용도 때문에 사용자의 관심사를 찾아내기 위한 활발한 연구가 이루어지고 있다.

기존 SNS 관련 연구들인 [6-8]에서는 SNS의 급격한 발전에 따라 증가한 소셜 데이터(Social data) 속에서 사용자의 관심사와 활동의 패턴을 연구했다. 여기서 소셜 데이터는 SNS상에서 만들어 지는 텍스트, 사회적 관계, 동영상 등의 모든 데이터라고 정의했다. 특히 [6]에서는

SNS에 존재하는 모든 요소들을 함께 고려하여 각각의 가중치를 구하는 확률 모델을 제안했다. 이 때, SNS에 있는 모든 요소들을 사회적 상관관계(Social correlation), 사용자(User), 아이템(Item), 희소성(Sparsity) 이렇게 크게 4개의 요소로 분류했다. 사회적 상관관계는 자신과 비슷한 사람과 관계를 맺고자 하는 동질성(Homophily)와 사회적인 영향으로 관계를 맺고자 하는 사회적 영향(Social Influence)가 포함된다. 이와 비슷하게 [8]에서는 사회적 상관관계의 2가지 요소의 이론을 기반으로 하여 사용자의 관심사와 행동을 고려하여 토픽 모델(Topic Model)을 제안했다.

또한 사용자의 관심사를 파악하기 위해서도 다양한 연구들이 진행됐다. [9]에서는 수많은 SNS의 정보를 통합하여 기존의 태그 기반 협력적 필터링(Tag-based Collaborative Filtering)을 향상시키는 방법을 제안했다. 이 논문에서 제안하는 방법은 SNS에 있는 정보들의 다양한 요소들 중 아이템 태그(Item-tag)와 친구 관계(Friendship), 단체의 구성원 관계(Membership)를 모두 고려하여 전체의 유사성(Similarity)을 구하고, 이를 통해 관심의 정도를 정해 추천을 하는 것이다.

[10]과 [11]은 SNS 중에서도 특히 트위터에서 사용자의 관심사를 찾아내는 방법을 제안했다. [10]은 트위터의 특정 사용자에게 대한 관심사의 주제(Interest topic)들을 발견하는 데 집중했다. 이 논문이 제안하는 방법인 Twopics는 트위터에서 작성한 글(Tweet)을 기반으로 사용자의 관심사가 목록으로 정리된 토픽 프로필(Topic profile)을 만들어 내게 된다. [11]은 HIG(Hierarchical Interest Graph)를 제안했다. 이는 사용자가 작성한 글의 내용을 기반으로 기초적인 관심사를 만들어 낸 후 WCG(Wikipedia Category Graph)를 매칭함으로써 만들어 지는 그래프를 통해 사용자의 관심사를 알아냈다.

물론 앞서 언급한 것처럼, 많은 연구들을 통해 보다 정확하고 효율적으로 사용자의 관심사 추출이 가능해졌고, 이를 개선하기 위한 연구들도 이미 많이 진행되고 있다. 하지만 대부분의 기존 연구들은 사회적 관계망에만 집중해 사용자의 사회적 관계를 파악하거나 사용자가 취한 행동을 통해서만 관심사를 추출하였다. 또 사용자의 관심사를 카테고리를 이용해서 추출한다고 하더라도 그 카테고리를 분류할 때 위키피디아에서 제공하는 분류만을 고려하여 SNS의 특징을 확연하게 표현하는 데에는 어려움이 있었다.

본 연구에서는 보다 정확한 사용자의 관심사 추출을 위해서 사용자가 직접 작성한 글, 사진, 동영상 등 사용

자가 취한 행동 뿐만 아니라, 사용자의 사회적 관계를 함께 적용하여 페이스북의 페이지 카테고리 분류를 기준으로 사용자의 관심사를 표현할 수 있는 카테고리를 추출한다. 이를 통해서 사용자가 특정 인물과 어떤 관심사를 공유하는지, 또 사용자 간의 관계를 가중치로 이용하여 관심 있는 카테고리를 기반으로 한 추천 방법을 제안하여 선행연구들과 차별화하였다.

3. 소셜 카테고리를 이용한 추천 방법

본 장에서는 소셜 데이터를 이용하여 사회적 관계를 추출하고, 카테고리 분류를 적용하여 사용자의 관심사를 찾아내고 추천하는 방법에 대해 설명한다. 우리가 제안하고자 하는 소셜 카테고리와의 사용자의 관심사를 기반으로 하는 추천 방법의 전체적인 구조와 방법을 설명하고, 각 단계별로 자세하게 기술하고자 한다.

3.1 제안하는 추천 방법

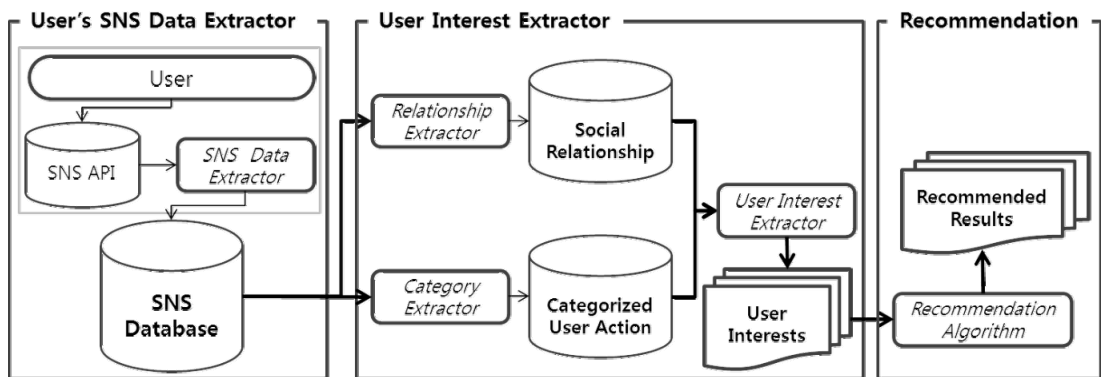
우리가 제안하는 추천 방법에서는 최종적인 추천을 위해 각 SNS에서 제공하는 API를 이용하여 SNS에 존재하는 수많은 데이터, 즉 소셜 데이터를 가져온다. 이 소셜 데이터에는 친구 관계와 사용자가 작성한 글, 사진, 동영상등 수많은 활동들이 포함된다. API를 통해 가져오는 데이터는 어떤 SNS를 기반으로 하느냐에 따라 차이가 나탄다. 본 논문에서는 페이스북을 기준으로 하여 실험을 진행하였기 때문에 자세한 설명은 페이스북을 기준으로 할 것이다.

API에서 가져온 소셜 데이터를 이용하여 사용자의 친구 목록을 가져오고, 각 친구와 얼마나 많은 상호작용(Interaction)을 가졌는가를 기준으로 랭킹을 매기게 된다. 본 논문에서 상호작용은 사용자와 친구 간의 SNS 상에서의 상호교류로, 특정 친구가 사용자에게 글을 작성하거나 사용자가 작성한 글에 대해 댓글이나 공유, 태그 등의 활동을 한 것으로 정의한다. 페이스북을 예로 들자면, 특정 친구가 사용자의 피드(Feed) 내에서 ‘좋아요’, 댓글, 태그 등의 활동을 사용자와 특정 친구 간의 상호작용이라고 볼 수 있다.

이렇게 서로의 상호작용을 기준으로 랭킹이 된 친구 목록에서 각 친구와 주고받은 모든 글과 댓글의 내용을 분석하여 빈도수가 많은 단어들을 각 친구 별로 추출한다. 각 친구 별로 추출된 단어들을 각 SNS에 따라 만들어 지는 카테고리를 기준으로 하여 카테고리 분류를 한다. 본 논문에서는 실험을 위해 페이스북의 카테고리를 기준으로 하여 카테고리 분류를 했다.

이러한 방법으로 카테고리 분류를 하면 각 친구 별로 사용자와 어떠한 카테고리에서 유사점을 갖고 있는지, 결국 어떠한 관심사를 해당 친구와 사용자가 공유하고 있는지를 알 수 있게 된다.

본 연구에서 제안하는 추천 방법은 (그림 1)처럼 크게 3가지의 단계로 분류할 수 있다. (1) SNS API를 이용하여 사용자의 소셜 데이터를 추출하는 과정, (2) 추출된 소셜 데이터를 이용하여 사회적 관계와 사용자의 활동을 기반으로 한 카테고리 분류를 통해 사용자의 관심사 목록을 추출하는 단계, 그리고 (3) 최종적으로 관심사의 빈도수와 가중치를 이용하여 그 용도에 따라 관심 있는 카테고리를 기반으로 하여 추천을 해 주게 된다.



(그림 1) 제안하는 추천 방법의 전체적인 구조

(Figure 1) Overall Structure of Proposed Recommendation Method

사용자의 관심사 목록이 카테고리로 분류되어 있기 때문에 빈도수를 계산하기에 용이하다. 또한 관계를 이용한 가중치를 적용하기 때문에 사용자의 요구에 따라 사용자의 전반적인 관심사들의 추천이 가능할 뿐만 아니라 특정 친구와만 공유하는 관심사의 추천도 가능하다.

다시 말해, 우리가 제안하는 방법을 이용한다면 사용자가 그 용도와 필요성에 따라 추천 방법을 다양하게 적용할 수 있다는 장점이 있다. 복잡한 알고리즘 없이 우리가 제안하는 추천 방법을 이용하여 모두와 공유하는 사용자의 관심사, 특정 친구와 공유하는 관심사, 두 가지 관심사를 기반으로 하는 추천이 가능하기 때문에 보다 세밀하고 상황에 맞는 추천이 가능한 것이다.

3.2, 3.3과 3.4에서는 제안하는 추천 방법 중 두 번째 단계에 해당하는 사용자의 관심사 목록을 추출하기 위한 사회적 관계 추출과 사용자 행동의 카테고리 분류, 그리고 사용자의 관심사 목록 추출에 대해 각각 기술할 것이다.

3.2 사회적 관계 추출

SNS에서 사회적 관계는 데이터 셋에 따라 다르지만 일반적으로 트위터처럼 Follower와 Followee로 구분되는 관계를 갖거나 페이스북처럼 친구 신청과 수락을 통해 관계를 갖게 된다. 친구의 친구, 혹은 연예인과 같은 유명 인사나 기업의 페이지까지도 관계를 맺을 수 있다는 특징 때문에 SNS 상에서 사용자들은 넓은 범위의 사회적 관계를 갖게 된다. 하지만 사용자가 이 많은 관계 속에서 모든 친구들과 동일한 관계를 갖고 있는 것은 아니다. 그렇기 때문에 어떠한 친구들과 더 긴밀한 관계를 갖고 있는지를 파악하는 것은 사용자의 사회적 관계 파악에 매우 중요한 요소가 된다.

사회적 관계를 파악하기 위해서는 우선 사용자의 친구 목록을 모두 파악해야 한다. 각 SNS API를 통해 친구 목록을 가져온 이후에는 각 친구들과 얼마나 많은 상호작용을 가졌는지 구해야 한다. 특정 사용자인 ‘나’와 교류가 많은 친구를 구하기 위해서 내가 작성한 글에 대해 친구들이 취한 행동의 비율을 계산한다. 이때 친구들의 행동을 페이스북을 기준으로 한다면 ‘좋아요’, 댓글, 사진이나 비디오, 혹은 글 등에 태그된 경우를 이야기 할 수 있다.

친구 중 한 명이 ‘나’에게 행동을 취한 수를 비율로 나타냈을 때, 큰 비중을 차지할수록 ‘나’와 친밀도가 높다고 볼 수 있게 되고, 이를 통해 친구 목록에 랭킹을 매길

수 있게 된다. 이 비율은 최종적으로 사용자의 관심사를 추출할 때 가중치로 이용된다.

$$\frac{NA_v}{NA_{all} - NA_u} \tag{1}$$

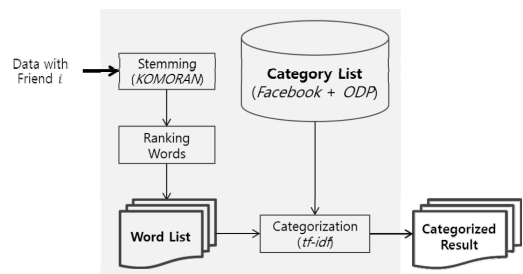
식 (1)은 사회적 관계를 추출할 때 계산되는 가중치에 대한 식이다. NA_{all} 은 사용자에게 대해 친구가 취한 전체 행동 수를 나타내고, NA_u 는 사용자(u)의 행동 수를, NA_v 는 사용자(u)와 친구(v) 간의 상호작용의 수를 나타낸다.

3.3 카테고리 추출

3.3.1 소셜 데이터의 카테고리 분류

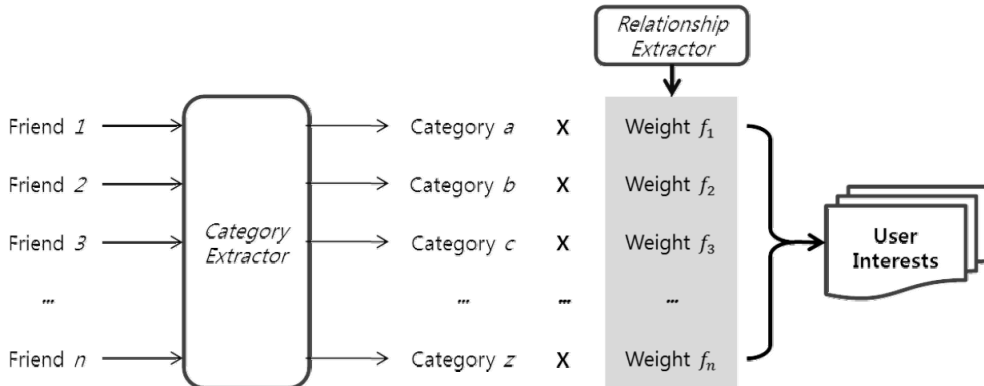
이 논문에서 제안하는 추천 구조에서는 카테고리를 이용하여 사용자의 관심사를 찾기 위해 소셜 데이터를 카테고리 기반으로 분류하는 과정이 필요하다. 카테고리를 분류하는 과정에서는 사회적 관계 분석을 통해 만들어진 친구 랭킹 목록과 각 친구들과 서로 상호작용을 했던 내용을 이용한다.

친구 랭킹 목록에 존재하는 친구들을 순서대로 한 명씩 선택하여 각의 친구와 나누었던 메시지나 댓글을 SNS API를 통해 추출하고, 추출한 문장들을 하나의 문서로 만들게 된다. 하나의 문서가 된 이 문장들을 Shineware Soft에서 제작하고 무료로 공개한 한국어 형태소 분석기인 ‘KOMORAN (v2.0.4)[12]’을 이용하여 형태소 분석을 수행하게 된다. 이 때, 가장 많이 사용된 단어들을 랭킹을 매겨 목록으로 추출해서 보여준다.



(그림 2) 카테고리 추출 과정
(Figure 2) The Process of Categorization

이렇게 형태소 분석을 통한 단어들의 랭킹이 완료되면, 각 친구와 많이 사용한 단어들을 기반으로 두 사람



(그림 3) 사용자의 관심사 목록 추출 방법
(Figure 3) The Method of Extracting User Interest List

사이의 카테고리를 추출하게 된다. 카테고리의 추출은 기존 SNS에 존재하는 카테고리 분류를 이용한다. 특히 페이스북에서는 페이지의 분류를 위한 카테고리가 존재하기 때문에 이 분류를 이용하면 카테고리의 추출이 가능하다. 다양한 페이스북의 페이지를 수집하여 각 페이지의 카테고리 별로 많이 쓰이는 단어들을 분류하여 수집하면, 정보 검색과 텍스트 마이닝 분야에서 많이 이용하는 방법인 TF-IDF(Term Frequency - Inverse Document Frequency)기법을 이용하여 카테고리 분류를 할 수 있다. 카테고리 분류의 기준은 3.3.2에 자세하게 기술한다.

수집된 SNS의 카테고리 분류를 기반으로 하면, 친구와 '나' 사이의 행동을 기반으로 추출한 단어들을 TF-IDF를 통해 분류가 가능하게 된다. 각 친구 별로 추출된 카테고리들을 앞서 계산한 사회적 관계에 적용하여 최종적으로 카테고리를 추출할 수 있다.

다시 말해서 우리가 제안하는 방법을 통해 '나'와 특정 친구와의 카테고리뿐만 아니라 전체 친구와의 카테고리도 추출이 가능해 진다. 이를 이용하면 결과적으로 '나'의 관심 분야, 즉 사용자의 관심사 추출이 가능한 것이다.

3.3.2 카테고리 분류 기준

페이스북의 페이지 분류[13]는 대분류와 중분류로 나뉘게 된다. 가장 큰 카테고리의 분류는 6개(지역 비즈니스 또는 장소 / 회사, 기관, 연구소 / 상표 또는 제품명 / 예술가, 밴드, 공인 / 엔터테인먼트 / 비영리, 자선단체)로 구성되고, 그 안에 각각 여러 개의 중분류들을 갖게 된

다. 하지만 페이스북의 대분류는 6개밖에 되지 않아 너무 적기 때문에 광범위한 분류가 된다. 또 중분류를 기준으로 할 경우, 총 156개의 카테고리(2014.06)로 너무 세세하고 많다. 그렇기 때문에 적절한 분류의 기준을 세우기 위해 DMOZ에서 제공하는 ODP(Open Directory Project)[14]의 최상위 카테고리를 함께 이용한다.

ODP의 카테고리는 총 16개로 이루어져 있지만 지역에 해당하는 'Regional'과 'World'는 페이스북의 페이지 분류와 거리가 있기 때문에 표 1과 같이 이 두 개의 카테고리를 제외한 총 14개의 카테고리를 기준으로 페이스북의 156개 카테고리를 다시 분류하여 사용한다.

SNS는 이용하는 사용자들에 의해 만들어 지는 데이터로 구성되기 때문에 카테고리를 분류할 때도 이러한 SNS의 특성을 반드시 반영해야만 한다. 그렇기 때문에 실제로 이러한 특성이 고려되어 있다고 볼 수 있는 페이스북의 카테고리 분류를 기반으로 분류된 카테고리 목록을 적용하는 것이 보다 효율적인 카테고리 분류를 가능하게 한다. 이러한 SNS 특성을 반영한 카테고리를 우리는 소셜 카테고리라고 정의하였다.

(표 1) 최상위 카테고리 목록
(Table 1) Category List (Highest Categories)

Arts	Business	Computers	Games
Health	Home	Kids & Teens	News
Recreation	Reference	Science	Shopping
Society	Sports		

3.4 사용자 관심목록 추출

사용자의 소셜 데이터를 기반으로 사회적 관계와 각 친구 별 공통적으로 많이 사용한 단어들을 통한 카테고리 분류가 완료된 다음에는 최종적인 사용자의 관심사를 추출하기 위해 카테고리가 분류된 결과에 사회적 관계를 계산한 결과를 기반으로 구한 가중치를 적용한다. 이러한 방법을 이용하면 실제 사용자의 관심사에 대한 결과를 도출할 수 있게 된다. 최종 결과인 사용자의 관심사를 추출하기 위한 방법은 (그림 3)과 같다.

사용자의 관심사 추출을 통해 최종적으로 사용자 관심사의 추출이 가능할 뿐만 아니라, 사회적 관계 추출을 통해 나온 가중치를 적용하여 계산하지 않을 경우 단순히 각 친구 별로 카테고리 추출만 적용하여 특정 친구와 공유하는 관심사의 추출도 가능하다.

4. 실험 및 결과

4.1 실험 개요

본 연구에서는 페이스북의 소셜 데이터를 기반으로 실험을 진행했다. 실험을 위해 페이스북 사용자의 정보에 접근 권한을 주는 액세스 토큰(Access Token)을 받아 페이스북에서 제공하는 Graph API를 이용하여 사용자의 친구 목록과 뉴스피드를 모두 가져온다. 이 소셜 데이터를 기반으로 사용자의 사회적 관계 찾아낸다. 사회적 관계를 찾아냄과 동시에 뉴스피드의 전체 내용 중에서 각 친구와 공유한 글과 댓글들을 찾아내어 형태소 분석기인 KOMORAN(v2.0.4)을 이용하여 형태소 분석을 한다. 형태소 분석을 통해 나온 단어들에 대해 랭킹을 매기고, 페이스북의 페이지 카테고리 분류를 기반으로 카테고리를 추출한다. 이렇게 해서 추출된 카테고리와 단순히 친구 관계를 기반으로 하여 추천되는 내용과 비교하여 실험을 할 수 있다.

(표 2) 페이스북 API에서 사용한 데이터와 Permission (Table 2) Used Data and Permissions from Facebook API

	Used data	Graph API permissions
Social Relationship Extractor	friend lists, likes, photos, post, comments	user_friends, read_streams, user_photos
Category Extractor	friend lists, newsfeed	user_friends, read_stream

페이스북의 페이지 카테고리 분류는 앞서 3.3.2에서 자세하게 말한 것처럼, 기본적으로 크게 6가지로 분류되어 있으나 이는 너무 광범위한 분류이고, 또 중분류로 들어가면 너무 세세하게 분류되어 있기 때문에 실험 결과를 도출하는 데에 적합하지 않다. 156개의 카테고리를 신뢰할 수 있는 분류로 나누기 위해 ODP의 대분류를 이용하였다. 총 14개의 카테고리를 사용했으며 그 분류는 표 3과 같다.

(표 3) 카테고리 분류표 (Table 3) Classified Category Table

상위 카테고리	하위 카테고리
Arts	Movie, TV/Movie Award, Album Song, Music Video, Concert Tour, Concert Venue, Record Label, Music Award, Music Chart, TV Show, TV Network, Actor/Director, Producer, Writer, Fictional Character, Movie Character, Musician/Band, Author, Artist, Comedian, Entertainer, Designer, Photographer, Studio, Movie Theater, Arts/Entertainment/Nightlife, Museum/Art Gallery
Business	Product/Service, Electronics, Building Materials, Commercial Equipment, Business Person, Entrepreneur, Local Business, Landmark, Airport, Automotive, Event Planning/Event Services, Bank/Financial Services, Pet Services, Professional Services, Business Services, Real Estate, Transportation, Company, Bank/Financial Institution, Insurance Company, Small Business, Energy/Utility, Retail & Consumer Merchandise, Automobiles & Parts, Industrials, Transport/Freight, Aerospace/Defense, Mining/Materials, Consulting/Business Services, Engineering/Construction, Telecommunication
Computers	Website, Computers, Software, Phone/Tablet, App Page, Computers/Technology, Internet/Software
Games	Games/Toys, Video Game, Board Game

상위 카테고리	하위 카테고리
Health	Health/Beauty, Vitamins/Supplements, Doctor, Spas/Beauty/Personal Care, Health/Medical/Pharmacy, Hospital/Clinic, Health/Medical/Pharmaceuticals
Home	Appliances, Home Décor, Kitchen/Cooking, Patio/Garden, Chef, Home Improvement
Kids & Teens	Baby Goods/Kids Goods
News	Magazine, Public Figure, Journalist, News Personality, Media/News/Publishing
Recreation	Wine/Spirits, Outdoor Gear/Sporting Goods, Drugs, Radio Station, Pet, Restaurant/Café, Bar, Club, Hotel, Food/Grocery, Attractions/Things to Do, Tours/Sightseeing, Food/Beverages, Travel/Leisure
Reference	Book, Book Store, Library, Book Series, Education
Science	Chemicals, Farming/Agriculture, Biotechnology
Shopping	Cars, Bags/Luggage, Camera/Photo, Clothing, Office Supplies, Furniture, Household Supplies, Tools/Equipment, Jewelry/Watches, Pet Supplies, Shopping/Retail
Society	Teacher, Lawyer, Community/Government, Public Places, Government Official, Non-Governmental Organization (NGO), Legal/Law, Community Organization, Political Organization, Church/Religious Organization, Organization School, University, Non-Profit Organization, Government Organization, Cause, Political Party, Middle School
Sports	Sports League, Professional Sports Team, Amateur Sports Team, School Sports Team, Athlete, Sports Venue, Sports/Recreation/Activities

4.2 실험 방법

실험 결과의 도출은 적중률(Hit ratio)을 계산하는 방법을 통해 이루어졌다. 우리가 제안한 사용자의 관심사 추천 방법이 실질적으로 사용자의 관심사를 충분히 반영했는지를 계산하기 위해 제안한 방법을 이용해 추출된 사용자의 관심사를 나타내는 카테고리들과 페이스북에서 사용자가 직접 ‘좋아요’를 한 페이지들의 카테고리를 이용했다. 사용자가 ‘좋아요’를 누른 특정 페이지들의 카테고리를 우리가 계산하고자 하는 적중률의 기준으로 삼는 이유는 사용자가 실제로 페이스북을 사용하면서 본인이 관심 있고 마음에 드는 페이지를 직접 선택하는 것이기 때문에 상당 부분 사용자의 관심사를 반영했다고 볼 수 있기 때문이다. 또한 페이지 별로 카테고리가 분류되어 있어 우리가 제안하는 방법의 최종 결과와 비교가 용이하기 때문에 선택했다.

적중률은 사용자가 ‘좋아요’를 한 페이지들의 카테고리를 기준으로 하여 우리가 제안한 추천 방법을 통해 추출된 카테고리가 얼마만큼 적중을 하는지를 계산하여 도출된다. 적중률이 높을수록 우리가 제안한 추천 방법을 통해 추출된 사용자의 관심사 카테고리가 실질적으로 사용자의 관심사를 높게 반영했다고 볼 수 있다.

실험을 위한 데이터를 모으기 위해서 먼저, 페이스북의 API를 이용하여 사용자들의 액세스 토큰을 받았다. 액세스 토큰을 받아야만 사용자의 정보를 갖고 올 수 있는 페이스북의 특성 때문에 토큰의 최소 2시간에서 최대 60일로 제한되어 있는 만료 기한이 있었고 가져올 수 있는 데이터의 양에 제한이 있었다. 액세스 토큰을 얻기 위한 임시 사이트를 통해 총 50명의 사용자에게 액세스 토큰을 얻었고, 친구 목록과 뉴스피드, 사진뿐만 아니라 실험의 기준을 위해 1330개의 ‘좋아요’ 한 페이지들의 정보를 데이터 셋으로 한다.

페이스북 페이지의 카테고리 분류 데이터를 얻기 위해 총 50명의 사용자에게 얻은 액세스 토큰을 이용하여 ‘좋아요’ 한 페이지 정보를 얻었다. 해당 데이터를 기반으로 페이스북 페이지의 카테고리 분류 기준을 정했으며, 실험에 사용했다.

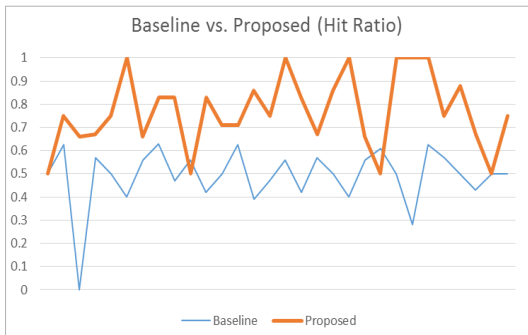
4.3 실험 결과

제안한 추천 방법이 어느 정도의 신뢰도를 갖고 있는지를 검증하기 위해 실험의 기준으로 같은 방법에서 사회적 관계를 제외하고 사용자의 관심사를 추출하는 방법

을 이용했다. 같은 조건 내에서 사회적 관계가 적용된 방법과 제외한 알고리즘의 적중률은 (그림 4)과 같은 결과를 나타냈다.

사회적 관계를 적용하지 않은 경우, 평균적으로 약 49%의 적중률을 갖고 있고, 적용했을 경우에는 약 77%의 적중률을 갖고 있어, 사회적 관계를 적용하여 사용자의 관심사를 추출하는 것이 상대적으로 좋은 결과를 나타낸다는 것을 알 수 있다.

본 연구에서 제안하고자 한 추천 방법의 신뢰도를 검증하기 위해서는 기존의 다른 연구들에서 제안된 추천 방법을 실험을 통해 비교하였다. 우리가 제안한 추천 방법은 기존 연구에서 제안하는 방법과는 실험 방법 및 실험에 이용되는 데이터의 양 등 다양한 부분에서 차이점이 있기 때문에 기존의 연구들 중 가장 유사한 연구와 비교하였다.



(그림 4) Baseline과 제안된 방법의 Hit Ratio 비교
(Figure 4) Comparison of Hit Ratio of Baseline and Proposed Method

기존에 SNS를 기반으로 사용자의 관심사를 추천하는 알고리즘에 대한 연구들 중에서 [11]에서 제시한 방법은 트위터에서 사용자가 작성한 피드만을 고려한다. 작성한 피드의 내용을 분석하고, 이를 위키피디아에서 제공하는 카테고리 분류인 WCG(Wikipedia Category Graph)와 서로 매치하여 사용자의 관심사를 찾아내게 된다. 매치할 때의 가중치 계산법에 따라 ‘Bell’과 ‘Bell log’, ‘Priority Interest’의 세 가지 실험 결과를 얻게 된다. [11]의 실험 방법으로 사용자들을 대상으로 설문조사를 했다. 각 답변은 추출되어 나온 관심사에 대해 본인 선호도와 비교한 결과에 따라 Yes/No/Maybe의 3가지 답변을 받게 된다. [11]에서는 이 실험을 총 37명의 사용자에게 시행했다.

(표 4) 제안하는 추천 방법과 기존 알고리즘([11])의 실험 결과 비교

(Table 4) Comparison of Experimental Result between Proposed Recommendation Method and the Existing Algorithm(11)

Compared Algorithms		Hit Ratio
Proposed Algorithm in [8]	Bell	0.53
	Bell Log	0.67
	Priority Interest	0.76
Our Proposed Method	Baseline	0.49
	Proposed	0.77

[11]에서 제안한 알고리즘들은 랭킹의 개수를 나누어 실험했다. 우리가 제안한 방법을 이용할 경우, 추천되는 사용자의 관심사가 사용자의 활동량에 따라 0개에서 최대 14개의 카테고리를 갖기 때문에, [11]이 실험한 결과 중에서 Top 10에 대한 실험 결과를 이용하여 비교가 가능하다. (표 4)에서 보이는 것처럼 [11]이 제안한 알고리즘의 가장 좋은 결과가 76%로 우리가 제안한 방법이 이와 유사하지만 조금 높은 결과를 나타낸다.

수치상의 비교 결과만 놓고 볼 경우, 우리가 제안한 추천 방법이 기존의 알고리즘과 비슷한 결과를 보여주지만 실질적으로 제안하는 방법과 알고리즘이 서로 다른 점도 있고 실험에 있어서도 방법과 양 등의 차이점을 고려할 때 우수한 결과를 보여주고 있다.

[11]의 실험 방법은 사용자에 대한 직접적인 설문조사로 이루어진 반면, 본 연구에서의 실험은 설문조사가 아니라 우리가 제안한 방법을 통해 추천한 사용자의 관심사와 사용자가 페이스북을 이용하면서 실제로 선택한 페이지의 카테고리 간의 적중률을 알고리즘을 적용하여 검증하였다. 수치상으로는 매우 근소한 차이로 우리가 제안한 알고리즘이 우수하지만, 다양한 요소들과 새로운 자동화 기법을 고려할 때, 대량의 소셜 데이터를 적용하기에 매우 유용한 방법이다.

5. 결 론

이 논문에서, 우리는 SNS에서 사용자의 사회적 관계와 소셜 데이터의 카테고리 분류를 통해 사용자의 관심사를 찾아내고 추천해 주는 방법을 제안했다. 사용자가 작성한 글에만 집중하여 관심사를 찾아내는 기존의 방식

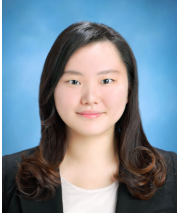
에서 벗어나 친구 관계와 해당 친구와 공유한 글들 또한 고려하여 추천을 할 때 사용자에게 보다 적합한 추천이 가능해 진다. 실험을 통해 사용자의 피드만 단순히 분석하는 것보다 사회적 관계를 함께 고려하는 것이 더 좋은 결과를 보여줄 수 있다는 것을 실험을 통해 검증하였다.

SNS에서 사회적 관계 이외의 또 다른 요소를 찾아 함께 고려한다면 더 정확하게 사용자의 관심사를 고려한 추천이 가능해 질 것이다. 특히 SNS는 사회적인 사건과 이슈에 굉장히 민감하게 반응하며 이러한 특성이 사용자에게 많은 부분에서 영향을 줄 수 있다. 중요한 변인이 될 수 있는 사회적 이슈와 사건들을 어떻게 특정화하고 어떤 카테고리에 가중치를 두어야 하는지에 대한 향후 연구도 필요할 것이다. 뿐만 아니라, 실제로 SNS를 기반으로 대량의 소셜 데이터를 이용하는 소셜 미디어 추천 시스템에 적용한다면, 신뢰도가 높은 추천 방법이 될 것이라고 기대한다.

참 고 문 헌 (Reference)

- [1] Twitter, <http://www.twitter.com>.
- [2] Facebook, <http://www.facebook.com>.
- [3] G.Lawton, "Knowledge Management: Ready for Prime Time?", IEEE, Computer, Vol. 34, No.2, pp216-244, 2001.
- [4] Katarzyna Musial, Przemyslaw Kazienko, and TomazKajdanowicz, "Social Recommendations within the Multimedia Sharing Systems", WSKS 2008, LNAI 5288, pp. 364-372, 2008.
- [5] Yu-Doo Kim, Il-Young Moon, "A Study on Algorithm for Selection Priority of Contents in Social Network Service", Korean Institute of Marine Information and Communication Sciences, 2011.
- [6] Bo Hu, Mohsen Jamali, and Martin Ester, "Learning the Strength of the Factors Influencing User Behavior in Online Social Networks", ASONAM, 2011 IEEE/ACM International Conference, p368-375, 2012.
- [7] Michael Durr, Valentin Protschky, and Claudia Linnhoff-Popien, "Modeling Social Network Interaction Graphs", ASONAM, 2012 IEEE/ACM International Conference, p660-667, 2012.
- [8] Bo Hu, Zhao Song, and Martin Ester, "User Features and Social Networks for Topic Modeling in Online Social Media", ASONAM, 2012 IEEE/ACM International Conference, p202-209, 2012.
- [9] Sogol Naseri, Arash Bohrehmand, Chen Ding, and Chi-Hung Chi, "Enhancing Tag-based Collaborative Filtering via Integrated Social Networking Information", ASONAM, 2013 IEEE/ACM International Conference, p761-765, 2013.
- [10] Matthew Michelson, Sofus A. Macskassy, "Discovering Users' Topics of Interest on Twitter: A First Look", AND '10, p73-80, 2010.
- [11] Pavan Kapanipathi, Preetee Jain, Chitra Venkataramani, and Amit Sheth, "User Interest Identification on Twitter Using a Hierarchical Knowledge Base", The Semantic Web: Trends and Challenges, p99-113, ESWC, 2014.
- [12] Shinware Soft, KOMORAN, <http://shinware.tistory.com/entry/>, 2014.
- [13] Facebook Creating Page, https://www.facebook.com/pages/create/?ref_type=bookmark, 2014.
- [14] DMOZ ODP, <http://www.dmoz.org/>.
- [15] Young-Sung Shin, Young-Man Oh, Byeong-Seok Oh, Hyeong-il Kim, and Jae-woo Chang, "An Expert Recommendation Technique using Hybrid Collaborative Filtering in SNS", Database Society, p53, KIISE, 2012.
- [16] Mohsen Jamali, Martin Ester, "Mining Social Networks for Recommendation", Tutorial in ICDM 2011, 2011.
- [17] Daniela Godoy, Analia Amandi, "Modeling User Interest by Conceptual Clustering", Information Systems, Volume 31, Issues 4-5, p247-265, ELSEVIER, 2006.
- [18] Krishnan Ramanathan, Komal Kapoor, "Creating User Profiles Using Wikipedia", Conceptual Modeling - ER 2009, p415-427, ISBN, 2009.
- [19] Ke Tao, Fabian Abel, Qi Gao, and Geert-Jan Houben, "TUMS: Twitter-Based User Modeling Service", ESWC 2011 Workshops, 2011.
- [20] FacebookAPI, <https://developers.facebook.com/>.

● 저 자 소개 ●



유 소 엽 (So-Yeop Yoo)

2014년 2월 가천대학교 소프트웨어설계·경영학과 학사

2014년~현재 가천대학교 일반대학원 소프트웨어설계·경영학과 (석사과정)

관심분야 : Social Network, Data Mining

E-mail : bbusso@gc.gachon.ac.kr



정 옥 란 (Ok-Ran Jeong)

2005년 이화여자대학교 컴퓨터공학과 (공학박사)

2005년~2006년 서울대학교 컴퓨터공학부 (박사후 연구원)

2007년 Univ. of Illinois of Urbana Champaign (visiting scholar)

2008년~2009년 성균관대학교 정보통신공학부 (연구교수)

2009년~현재 가천대학교 소프트웨어설계·경영학과 (조교수)

관심분야 : 웹 마이닝, 정보검색, 추천 시스템, 소셜 컴퓨팅

E-mail : orjeong@gachon.ac.kr