

# 소셜 네트워크에서 사용자의 관심 분야, 인적 관계 및 응답 품질을 고려한 분야별 전문가 추천 기법

## Expert Recommendation Scheme by Fields Using User's interesting, Human Relations and Response Quality in Social Networks

송희섭\*, 유승훈\*\*, 정재윤\*\*, 박재열\*\*, 안지환\*\*, 임종태\*\*, 복경수\*\*, 유재수\*\*  
충북대학교 빅데이터학과\*, 충북대학교 정보통신공학과\*\*

Heesub Song(heesub@chungbuk.ac.kr)\*, Seunghun Yoo(yshun2487@naver.com)\*\*,  
Jaeyun Jeong(jjeong@chungbuk.ac.kr)\*\*, Jaeyeol Park(yeols@chungbuk.ac.kr)\*\*,  
Jihwan Ahn(jasonahn1202@chungbuk.ac.kr)\*\*, Jongtae Lim(jtlim@chungbuk.ac.kr)\*\*,  
Kyoungsoo Bok(ksbok@chungbuk.ac.kr)\*\*, Jaesoo Yoo(yjs@chungbuk.ac.kr)\*\*

### 요약

최근 인터넷과 스마트 폰의 발달로 사용자들 사이의 관계를 통해 다양한 정보를 생성하고 공유할 수 있는 소셜 미디어 서비스가 활발히 이용되고 있다. 특히 정보의 양이 방대해지고 신뢰할 수 없는 정보가 증가함에 따라 사용자에게 필요한 정보를 제공해 줄 수 있는 전문가 추천 기법에 대한 연구들이 진행되고 있다. 본 논문에서는 사용자의 관심 분야, 인적 관계, 응답 품질을 고려한 전문가 추천 기법을 제안한다. 사용자의 관심 분야는 사용자가 소셜 네트워크상의 활동을 분석해 최신의 사용자의 관심 분야 지수를 판단한다. 사용자의 인적 관계는 소셜 네트워크상의 같은 관심분야의 사용자만을 추출하여 인적 관계를 구축하여 인적 관계 지수를 판단한다. 사용자의 응답 품질은 사용자의 응답 속도와 응답 내용을 고려하여 응답 품질 지수를 판단한다. 마지막으로 사용자의 관심 분야, 인적 관계, 응답 품질을 합하여 사용자의 전문가 지수를 판단하고 사용자의 질의를 분석하여 질의와 전문가 그룹을 매칭하여 전문가를 추천한다. 다양한 성능평가를 통해 제안하는 기법이 기존 기법에 비해 성능이 우수함을 보인다.

■ **중심어** : 소셜네트워크 | 전문가 추천 | 전문가 검색 | 관심 분야 | 인적 관계 | 응답 품질 |

### Abstract

Recently, with the rapid development of internet and smart phones, social network services that can create and share various information through relationships among users have been actively used. Especially as the amount of information becomes enormous and unreliable information increases, expert recommendation that can offer necessary information to users have been studied. In this paper, we propose an expert recommendation scheme considering users' interests, human relations, and response quality. The users' interests are evaluated by analyzing their past activities in social network. The human relations are evaluated by extracting the users who have the same interesting fields. The response quality is evaluated by considering the user's response speed and response contents. The proposed scheme determines the user's expert score by combining the users' interests, the human relations, and the response quality. Finally, we recommend proper experts by matching queries and expert groups. It is shown through various performance evaluations that the proposed scheme outperforms the existing schemes.

■ **keyword** : SNS | Expert Recommendation | Expert Search | Interesting Field | Human Relations | Response Quality |

\* 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업(IITP-2017-2013-0-00881), 2015년도 정부(교육부)의 재원으로 한국연구재단(No.2015R1D1A3A01015962), 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단(No. 2016R1A2B3007527), 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 사회맞춤형 산학협력 선도대학(LINC+) 육성사업의 연구결과입니다.

\* 본 논문은 한국콘텐츠학회 2017 춘계 종합학술대회 우수논문입니다.

접수일자 : 2017년 10월 11일

심사완료일 : 2017년 10월 30일

수정일자 : 2017년 10월 30일

교신저자 : 유재수, e-mail : yjs@chungbuk.ac.kr

## I. 서론

인터넷의 발달로 인해 사용자들 사이에 인적 관계를 통해 정보를 생성하고 공유하는 트위터, 페이스북, 인스타그램과 같은 소셜 네트워크 서비스가 활발하게 활용되고 있다[1]. 소셜 네트워크 서비스는 인터넷을 통해 사용자 모두와 인적 관계를 형성할 수 있고 스마트폰을 통해 시간과 장소에 제약 없이 빠르게 정보를 공유, 수집할 수 있다. 최근 소셜 네트워크 서비스는 타인과의 의사소통 및 정보 공유뿐만 아니라 사용자가 궁금한 내용이 담긴 질문을 작성하고 다수의 다른 사용자들에게 응답을 얻는 등 능동적으로 정보를 얻기 위한 하나의 매체로도 이용하고 있다[2]. 그러나 질문을 한 사용자의 인적 관계의 한계로 인해 질문한 내용이 전문가에게 도달하여 고품질의 답변을 얻기 힘들다는 단점이 있다. 다양하고 방대한 정보 때문에 사용자들은 악의적인 정보, 신뢰할 수 없는 정보, 필요치 않는 정보에 쉽게 노출이 되어 정보의 양에 비해 양질의 정보를 얻기 힘들다는 단점이 있다. 이러한 문제점들을 해결하기 위해서 사용자가 질문한 정보에 대해 신뢰할 수 있고 양질의 정보를 제공해 줄 수 있는 전문가를 추천하고 사용자가 필요한 정보를 요청하거나 볼 수 있도록 할 수 있는 서비스가 필요하다[3].

기존의 전문가 추천 기법은 소셜 네트워크 사용자들의 직접 입력한 프로필을 기반으로 추천하는 기법과 사용자들의 활동내역을 분석하여 추천하는 기법을 사용하였다. 사용자들의 프로필을 이용하는 기법은 질의의 내용과 사용자들의 프로필의 유사도를 판별하여 가장 높은 유사도를 가진 사용자를 추천해주는 기법으로 비용은 적게 들지만 전문가 정확도가 낮은 단점이 있다[4-10]. 사용자들의 활동 내역을 분석하여 추천하는 기법은 사용자가 남긴 글과 댓글 좋아요 친구관계 등의 행위를 고려하여 사용자의 질의와 유사도를 판별하여 가장 높은 유사도를 가진 사용자를 추천해주는 기법으로 질의의 내용의 전문가를 추천하지만 응답 품질을 보장할 수는 없다[11-15].

기존 기법의 문제점들을 해결하기 위해 본 논문에서는 사용자의 관심 분야, 인적 관계, 응답 품질을 추출하

여 전문가를 추천한다. 또한 사용자가 질의한 내용을 분석하여 단어의 계층적 구조를 이용하여 전문가 그룹을 매칭하는 방법을 사용하고 있다. 사용자의 관심 분야는 사용자의 소셜 활동을 분석하여 최근 사용자가 관심을 가지고 있는 분야에 대하여 분석하고 사용자 관심 분야를 계산한다, 사용자의 인적 관계는 소셜 네트워크 상의 사용자들의 인적 관계를 이용하여 같은 분야의 사용자들이 사용자와 관계를 얼마나 맺고 있는지 분석하여 사용자의 영향력을 계산한다. 사용자의 응답 품질은 사용자가 응답한 것을 고려하여 얼마나 빠르고 정확하게 다른 사용자의 질의를 답해 줄 수 있는지에 대하여 분석하여 사용자의 응답품질을 계산한다. 또한 사용자 질의를 단어의 계층적 구조를 이용하여 질의의 내용뿐만 아니라 하위어에 대한 전문가를 포함시켜 전문가 추천의 정확도와 신뢰성을 높인다. 제안하는 기법의 우수성을 보이기 위해 기존 기법과의 다양한 성능평가를 수행한다.

본 논문의 구성은 다음과 같다. II장에서는 관련 연구와 기존 기법에 대하여 소개하고 문제점을 제시한다. III장에서는 제안하는 기법의 특징과 과정을 상세히 기술한다. IV장에서는 성능평가를 통해서 제안하는 기법의 우수성을 입증한다. V장에서는 결론과 향후 연구 방향에 대해 기술한다.

## II. 관련 연구

### 1. 사용자 프로필을 이용한 전문가 추천 기법

Duchateau은 소셜 네트워크 내의 사람들 사이에서 주어진 질의에 응답할 수 있는 사용자를 찾는 방법을 제안하였다[3]. 사용자는 소셜 네트워크에 가입할 때 프로필을 작성하게 되고 또한 사용자들끼리 인적 관계를 맺어 이루어져 있다. 이러한 프로필 정보를 이용하여 사용자의 질의에 맞는 사용자를 전문가로 추천한다. 전문가 추천을 위해 단순 매칭, 클러스터 매칭, 트리 매칭으로 세 가지 방법을 제안한다. 단순 매칭 방법은 질의와 사용자들의 프로필을 단순 비교 하는 방법이다. 질의를 단어로 구분하고 인적 관계를 맺은 사용자들의 프

로필을 수집하여 질의와 프로필이 가장 일치하는 사용자를 추천해 주는 방법이다. 클러스터 매칭 방법은 인적 관계를 맺고 있는 사용자 중 비슷한 프로필을 가진 사용자를 클러스터로 구성하고 유사어와 동의어를 추가하여 확장된 클러스터로 유지한다. 질의 또한 유사어와 동의어를 추가하여 질의 클러스터를 구축한다. 질의 클러스터와 사용자 클러스터를 비교하여 가장 비슷한 사용자를 전문가로 추천한다. 트리 매칭 기법은 사용자의 프로필을 단어의 계층적 관계를 이용하여 트리로 구축하고 질의 또한 트리로 구축한다. 질의와 프로필이 일치되는 트리 구조가 많을수록 높은 유사도 점수를 부여하고 전문가로 추천한다.

소셜 네트워크상의 사용자들의 프로필을 분석하는 방법은 다른 방법에 비해 검색 시간은 빠르지만 시간이 지남에 따라 사용자의 관심 분야나 전문 분야가 바뀌게 되는데 대부분의 사용자는 프로필을 갱신하지 않는다. 사용자가 프로필을 갱신하지 않아 사용자의 최신 관심 분야를 추출할 수 없다는 단점이 있다. 또한 일부의 사용자들은 프로필을 허위로 작성하여 입력하는 경우도 존재한다. 하지만 프로필을 기반으로 전문가 검색을 한다면 이러한 허위 정보를 입력한 사용자도 전문가로 검색할 가능성이 있어 검색 결과의 신뢰성을 저하시키는 원인이 된다. 본 논문에서 제안하는 기법은 사용자의 프로필을 기반으로 하는 것이 아니라 사용자의 소셜 활동을 분석하여 사용자의 최신의 관심분야를 추출하여 사용자의 프로필이 갱신되지 않았을 때의 문제점을 해결 하였고 또한 사용자의 인적 관계를 이용하여 사용자의 허위 정보를 판단할 수 있다.

## 2. 사용자 소셜 활동을 이용한 전문가 추천 기법

프로필 기반의 전문가 추천의 문제점을 해결하기 위해 최근에는 소셜 네트워크상의 사용자의 활동을 고려하여 전문가를 추천하는 기법이 연구되었다[11-15]. I. B. Jeon[11]은 소셜 네트워크로 연결된 사용자 중 전문 지식을 가진 사용자를 추천 해주는 기법을 제안 하였다. 소셜 네트워크 전문가를 추천하기 위해 각 사용자의 최근 활동을 수집하고 분석한 다음 동적 프로필을 생성한다. 사용자의 질의가 요청되면 제안된 기법은 질

의 결과의 정확성을 높이기 위해 온톨로지를 사용하여 질의의 클러스터를 생성한다. [11]에서 제안된 기법은 질의 클러스터를 사용자의 동적 프로필과 비교하여 가장 유사한 사용자를 전문가로 추천한다.

A. Bozzon[15]는 소셜 네트워크상의 사용자 활동 내역과 사용자들의 인적 관계를 분석하여 전문가를 추천 해주는 기법을 제안하였다. 소셜 네트워크상의 사용자 활동 내역을 분석하기 위해 사용자의 프로필, 남긴 글 등을 분석하여 질의와 유사성이 높은 사용자를 전문가로 추천한다. 사용자가 질의를 요청하면 서버는 질의를 전문가 후보셋과 매칭하기 위해 준비한다. 사용자와 인적 관계를 맺고 있는 사용자들을 모두 전문가 후보셋으로 포함하고 인적 관계를 맺고 있는 사용자들의 프로필이나 남긴 글을 분석하여 전문 지식을 판단한다. 전문 지식 점수에 따라 순위를 부여하고 질의를 요청한 사용자에게 전문가를 추천하게 된다.

사용자의 활동을 기반으로 전문가를 추천하는 기법은 사용자의 최근 관심 분야를 이용하여 전문가를 추천 해준다. 그러나 이러한 전문가 추천 기법은 사용자의 인적 관계를 맺은 사용자의 수가 충분치 못하거나 사용자들이 활동을 하지 않는 관심 분야의 경우에는 정확한 전문가를 추천 해줄 수 없다. 또한 사용자의 인적 관계가 모두 비슷한 분야의 사용자일 경우 사용자가 궁금해 하는 다양한 질의에 적합한 사용자를 찾을 수 없다. 본 논문에서 제안하는 기법은 사용자가 직접 맺고 있는 인적 관계뿐만 아니라 전체 소셜 네트워크상의 모든 사용자 중의 같은 관심 분야의 인적관계만을 고려하기 때문에 사용자의 인적 관계가 부족하거나 한 가지 관심분야의 인적 관계만 가지고 있더라도 정확한 전문가를 추천 해 줄 수 있다.

## III. 제안하는 전문가 추천 기법

### 1. 전문가 추천 시스템 구조

기존의 전문가 추천 기법 중 사용자의 프로필을 이용하여 전문가를 추천하는 기법은 사용자가 가입할 때 직접 작성한 프로필을 기반으로 전문가를 추천해주는 방

식으로 검색시간과 비용은 적게 들지만 정확도와 신뢰도가 매우 낮은 문제점을 가지고 있고, 사용자의 활동을 기반으로 전문가를 추천하는 기법은 사용자의 응답 품질이 낮은 문제점을 가지고 있다. 본 논문에서는 기존의 전문가 추천 기법의 문제점들을 해결하기 위하여 제안하는 기법에서는 소셜 네트워크에서 사용자의 소셜 행위, 인적 관계, 응답 품질을 고려한 분야별 전문가 추천 기법을 제안한다.

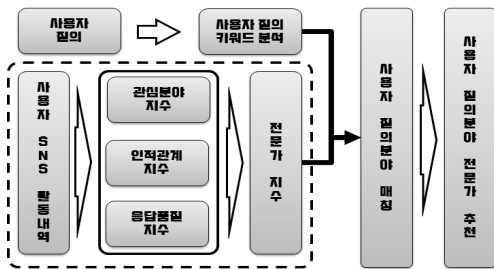


그림 1. 전문가 추천 시스템 구조

[그림 1]은 제안하는 전문가 추천 시스템 구조를 나타낸다. 전문가를 추천하기 위해 먼저 소셜 네트워크에서 사용자의 활동 내역을 수집한다. 사용자의 활동 내역은 사용자가 남긴 글, 댓글, 좋아요 수, 공유, 작성 시간 등이다. 사용자의 활동 내역을 이용하여 전문가 지수를 계산하는데 사용된다. 첫 번째로 수집된 사용자의 활동 내역을 분석하여 사용자의 키워드를 추출하여 관심분야 지수를 계산한다. 키워드를 추출하는 방법으로는 TF-IDF 기법을 이용하여 추출한다. 추출된 키워드 별로 시간, 좋아요 수, 공유하기, 댓글 등을 이용하여 각 활동별 가중치를 통하여 사용자의 관심분야 지수를 계산한다. 두 번째로 사용자별 소셜 네트워크의 인적 관계를 이용하여 인적 관계 지수를 계산한다. 인적 관계 지수는 각 사용자의 관심분야와 소셜 네트워크상의 인적 관계를 분석하여 관심분야별 인적 관계를 구축하고, PageRank 기법을 사용하여 인적 관계 지수를 계산한다. 세 번째로는 사용자의 응답 속도와 응답 신뢰도를 이용하여 응답 품질 지수를 계산한다. 응답 품질 지수는 사용자가 다른 사용자의 질문에 대한 답변의 속도, 질의한 사용자의 만족도 등을 분석하여 응답 품질 지수를 계산한다. 마지막으로 관심분야 지수, 인적 관계 지

수, 응답 품질 지수를 이용하여 분야별 전문가 점수를 계산한다. 사용자의 질의가 요청되면 사용자 질의를 분석하여 질의 분야를 판단한다. 질의 분야를 단어의 계층적 구조를 사용하여 매칭하고 사용자에게 전문가 그룹을 추천한다.

## 2. 사용자 관심 분야 지수

소셜 네트워크를 이용하는 사용자는 가입할 때 프로필을 작성한다. 하지만 대부분의 사용자는 자신의 관심분야가 변화하더라도 초기에 작성한 프로필을 갱신하지 않는다. 즉, 사용자의 소셜 행위를 분석하지 않는다면 사용자의 변화하는 관심분야나 전문 분야를 정확히 파악할 수 없다.

사용자 관심 분야 지수를 계산하는 방식은 [그림 2]와 같다. 사용자 관심 분야 지수는 사용자가 소셜 네트워크상에 작성한 글이나 공유한 데이터를 기반으로 한다. 사용자가 작성한 글이나 공유한 데이터는 사용자의 관심사를 표출한 것으로 해석한다. 사용자의 소셜 네트워크 활동 데이터를 수집한 후 형태소 분석기를 통해 단어를 추출하고 TF-IDF 기법을 이용하여 키워드를 추출한다. 또한 사용자의 소셜 행위의 남긴 시간, 좋아요 수, 코멘트 수, 공유 수를 이용하여 소셜 행위 신뢰도를 계산하여 TF-IDF 기법에 합하여 사용자 관심분야 지수를 계산한다.

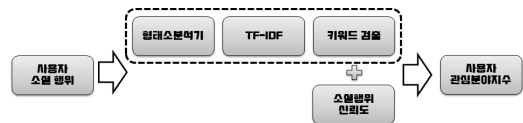


그림 2. 관심분야 지수 계산 방법

TF-IDF (Term Frequency - Inverse Document Frequency) 기법은 정보 검색과 텍스트 마이닝에서 이용하는 가중치이다. 여러 문서로 이루어진 문서 군이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한지를 나타내는 통계적 수치로 사용된다. 특히 문서의 핵심어를 추출하는 하는 용도로 많이 사용된다. TF는 단어 빈도로 특정 단어가 문서 내에 얼마나 자주 등장하는지를 나타내는 값이다. IDF는 역문서 빈도로 전체

문서에서 특정 단어를 가지고 있는 문서가 얼마나 되는지를 나타내는 값이다. TF-IDF는 TF와 IDF를 곱한 값이다. 예를 들어, 사용자가 활동한 문서가 100개라 하고 그중 “하둑”이라는 단어를 포함하는 문서가 8개라고 가정한다. 그 중 하나의 문서에서 “하둑”이라는 단어가 5번 나왔다고 가정하면 그 문서에 대한 TF 값은 5이고 IDF는 2.52573으로 TF-IDF 값은 12.6286 이다. 사용자 관심 분야를 추출하기 위해 사용자 활동 내역에 대해 TF-IDF 방식으로 키워드를 추출한다. TF-IDF 값이 높을수록 중요한 키워드로 추출한다. 추출된 키워드가 나타난 문서가 얼마나 신뢰도가 있는 문서인지를 판단하게 된다. 문서의 신뢰도를 평가하는 방법은 문서의 최신성에 가중치를 주어 최신의 글 일수록 시간 가중치를 주고, 문서의 좋아요 수, 댓글 수, 공유 수를 고려하여 문서의 신뢰도를 고려한다. 식 (1)은 사용자의 관심 분야 지수를 도출하는 방법을 기술한 식이다.  $dn$ 은 사용자가 소셜 네트워크상에 남긴 문서이고,  $dnT$ 는 사용자가 소셜 네트워크상에 남긴 문서의 신뢰도이다.

$$Keyword_i = \left( \frac{1}{dn} \sum_{dn=1}^{dn} TF_i * dnT \right) * IDF \quad (1)$$

식 (2)은 사용자가 소셜 네트워크상에 남긴 문서의 신뢰도에 대한 가중치를 계산하는 식이다.  $Time_w$ 은  $Time_w = \sqrt{-0.002739t + 1}$  로 사용자가 남긴 문서의 시간 가중치로 작성한지 오래된 문서일수록 낮은 가중치를 가지게 된다.  $Like_w$ 는 사용자가 남긴 문서의 좋아요 평균값이고,  $Comment_w$ 는 사용자가 남긴 문서 평균값,  $Share_w$ 는 사용자가 남긴 문서의 공유 평균값이다. 가중치  $\alpha, \beta, \gamma, \delta$ 의 합은 1로 표현된다.

$$dnT = \alpha(Time_w) + \beta(Like_w) + \gamma(Comment_w) + \delta(Share_w) \quad (2)$$

이렇게 얻어진 키워드를 오름차순으로 정렬하여 상위 Top-n 키워드를 추출한다. 하지만 이렇게 만들어진 키워드 점수만을 가지고 사용자 관심분야 지수를 계산한다면 키워드별로 문서의 양이 다르기 때문에 최신의

키워드인 경우에는 매우 낮은 점수가 전문가가 될 경우가 있다. 이러한 문제점을 해결하기 위해 소셜 네트워크상의 같은 관심분야를 가지고 있는 사용자 그룹의 합으로 사용자 관심분야 지수를 나누어 계산한다.

### 3. 사용자 인적 관계 지수

전문가 추천 기법에서 인적 관계의 연결 관계를 고려하는 것은 중요하다. 전문가는 다양한 분야의 인적 네트워크를 구성하고 있는 것 보다 같은 분야의 많은 사람들과 관계를 맺는 경향이 있다. 제안하는 기법은 사용자의 소셜 네트워크상의 활동을 분석하여 사용자별 관심분야를 추출한 결과를 가지고 관심분야 별로 인적 관계를 재구축한다. PageRank 알고리즘은 검색엔진의 품질을 향상시키기 위해 만들어진 알고리즘으로 웹페이지의 링크 관계를 통하여 웹페이지 당 점수를 계산하는 알고리즘이다. PageRank 알고리즘을 인적 관계에 적용하면 한 분야의 전문가는 같은 분야의 많은 사용자들과 인적 관계를 맺고 있어 PageRank 점수가 높고, 전문성이 떨어지는 사용자는 같은 분야의 사용자들과 인적 관계가 적어 PageRank 점수가 낮아진다. 본 논문에서는 PageRank 알고리즘을 인적 관계에 적용하여 사용자의 인적 관계 지수를 판단하는데 사용한다.

[그림 3]은 관심분야별 인적 관계의 구축을 나타낸다. 소셜 네트워크의 인적 관계가 A와 B라는 다양한 관심분야로 섞여 있다고 가정한다. A라는 관심분야의 인적 관계 점수를 계산하기 위해 B의 관심분야는 배제하고 관심분야 A만을 이용하여 인적 관계를 재구축하고 PageRank 알고리즘을 사용하여 인적 관계 지수를 계산한다. 같은 방식을 통해 관심분야 별 인적 관계 지수를 계산한다.

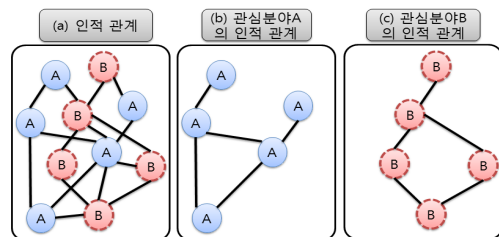


그림 3. 관심분야 별 인적 관계 지수

식 (3)은 제안하는 기법에서 관심분야별 사용자 인적 관계 지수를 계산하는 방법이다.  $PR(U_2)$ 의 값은 관심 분야  $i$ 의 사용자  $U_2$ 의 키워드 점수 값,  $C(U_2)$ 는 사용자  $U_2$ 의 관심분야  $i$ 인 친구의 수,  $N$ 은 관심분야  $i$ 을 가지는 총 사용자 수이다.  $d$ 은 덤핑 벡터로 0과 1사이의 수로 본 논문에서는 0.85로 설정하였다.

$$PageRank_i(U_1) = (1-d)/N + d(PR(U_2)/C(U_2) + \dots + PR(U_n)/C(U_n)) \quad (3)$$

#### 4. 사용자 응답 품질 지수

소셜 네트워크 전문가 추천 서비스는 사용자가 전문가로 추천 받은 사용자에게 질의를 할 수 있고 전문가는 질의에 맞는 답변을 할 수 있다. 하지만 전문가 추천 시스템에서 사용자의 질의에 맞는 지식이 많은 전문가를 추천해주더라도 전문가가 답변을 하지 않는다면 사용자는 질의에 대한 전문가의 답변을 얻을 수 없다. 이러한 문제를 해결하기 위해 본 논문에서는 분야별 전문가의 응답 품질을 고려하여 질의한 사용자에게 신속하고 정확한 답변을 줄 수 있는 전문가를 추천한다.

식 (4)은 제안하는 기법에서의 응답 품질 지수를 나타낸다.  $Ar$ 은 사용자의 응답 양이고,  $An$ 은 사용자의 답변 $n$ ,  $SimAn$ 은 사용자의 답변 $n$ 과 관심분야의 유사도,  $likeAn$ 은 사용자의 답변 $n$ 의 좋아요 평균값,  $Atn$ 은 사용자의 평균 응답시간이다. 가중치  $\alpha, \beta$ 의 합은 1로 표현된다.

$$QAR = \alpha(Ar) + \beta\left(\frac{1}{n} \sum_{An=1}^n SimAn(likeAn + Atn)\right) \quad (4)$$

#### 5. 전문가 지수

본 논문에서는 사용자의 관심분야 지수, 인적 관계 지수, 응답 품질 지수를 종합하여 전문가 지수를 도출한다. 식 (5)은 응답 품질과 인적 관계를 고려한 분야별 전문가 지수를 계산하는 방법이다.  $keyword$ 는 관심분야 지수이고,  $PageRank$ 는 인적 관계지수,  $QAR$ 은 응답 품질 지수 이다. 가중치  $\alpha, \beta, \gamma$ 의 합은 1로 표현된다.

$$ExR = \alpha(keyword) + \beta(PageRank) + \gamma(QAR) \quad (5)$$

식 (5)를 이용하여 계산된 전문가 지수를 이용하여 상위  $N$ 명의 사용자를 관심분야, 인적 관계, 응답 품질을 고려한 분야별 전문가 그룹으로 구성한다. 구성된 전문가 그룹을 단어의 계층적 관계인 워드넷을 이용하여 상위어와 하위어의 구조로 저장한다.

#### 6. 사용자 질의와 전문가 그룹 매칭

기존의 전문가 추천 기법들은 사용자가 요청한 검색 질의와 단순히 매칭 하여 결과를 매칭 된 결과만의 전문가만 추천해주는 방식 이었다. 하지만 단순히 질의와 전문가만을 매칭 할 경우 포함되는 질의에 대한 하위에 대한 전문가는 추천 결과에 포함되지 않아 전문가 추천에 대한 정확도가 저하된다. 본 논문에서는 질의에 대한 전문가 그룹뿐만 아니라 단어의 계층적 관계를 구축하여 하위에 포함되어 있는 전문가들도 포함하여 사용자 질의에 정확도를 증가시킨다. 사용자 질의와 전문가 그룹의 매칭은 먼저 사용자 질의와 매칭 된 전문가 그룹을 선택한다. 매칭 된 전문가 그룹의 하위 전문가 그룹을 판단하여 매칭 된 전문가 그룹 결과에 포함시켜 최종 전문가 그룹을 도출한다. 예를 들어, [그림 4]와 같이 사용자 질의가 전문가 그룹B에 매칭 되었을 때 전문가로 추천되는 그룹은 전문가 그룹 B와 전문가 그룹 B의 하위 전문가 그룹인 C, D, E 그룹을 선택한다. 전문가 그룹 C, D, E 에 포함되어있는 전문가를 모두 추천 결과에 포함하는 것이 아닌 상위어에서 분기한 개수인  $\frac{1}{3}$  만큼의 전문가 그룹 C, D, E의 상위 전문가만을 포함 한다.

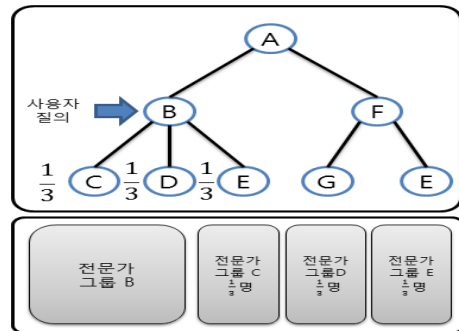


그림 4. 사용자 질의와 전문가 그룹 매칭 방법

#### IV. 성능평가

제안하는 전문가 추천 기법과 기존의 기법의 성능 비교 평가를 통해 제안하는 기법의 우수성을 입증한다. 사용자의 활동 정보를 분석하여 전문가를 추천하는 [11]의 기법과 성능 비교를 수행한다. [11]은 각 사용자의 최근 활동을 수집하고 분석 한 다음 동적 프로필을 생성한 뒤 질의가 요청되면 온톨로지를 사용하여 질의 클러스터를 생성한 뒤 확장된 질의와 사용자 동적 프로필을 비교하여 가장 유사한 사용자를 전문가로 추천하는 방법이다. Intel Core i5 4440 @ 3.10GHz, 8.00GB Dual-Channel DDR3, Windows 7 Ultimate K 64-bit 환경에서 추천된 결과의 정확도를 측정하기 위해 재현율, 정확률, F-Score을 사용하여 제안하는 전문가 추천 기법의 우수성을 입증하였다. 실험 데이터로는 실제 데이터 중에서 의미 있는 데이터의 양이 적기 때문에 임의의 사용자 데이터를 정규분포로 생성하였다. 실험 데이터는 [표 1]과 같다. 사용자의 수는 1000명, 사용자 인적 관계 수는 0명에서 200명까지, 사용자가 남긴 문서 수는 0에서부터 100개, 문서에서 추출된 키워드의 수는 3개, 문서의 신뢰성 점수는 0부터 1점까지, 키워드의 종류는 100개, 응답 품질은 0부터 1점까지, 각 사용자는 정규분포를 만족하도록 임의로 생성하였다.

표 1. 성능 평가 파라미터

파라미터	값
사용자의 수	1000
인적 관계 수	0 ~ 200
사용자 문서의 수	0 ~ 100
키워드의 수	3
문서의 신뢰성 점수	0 ~ 1
키워드의 종류	100
응답 품질 점수	0 ~ 1

전문가 추천 기법의 성능 비교를 위해 재현율(recall), 정밀도(precision), F-score을 사용하였다. 재현율, 정밀도, F-score를 평가하기 위한 데이터는 관심분야지수, 인적 관계 지수, 응답 품질 지수 세 가지의 방법으로 전문가를 랭킹한 후, 세 가지 랭킹결과에 모두 속한 전문

가를 전문가 집합으로 사용한다. 기존 기법으로는 사용한 전문가 추천 기법은 사용자의 활동만을 고려한 기법을 사용하였다.

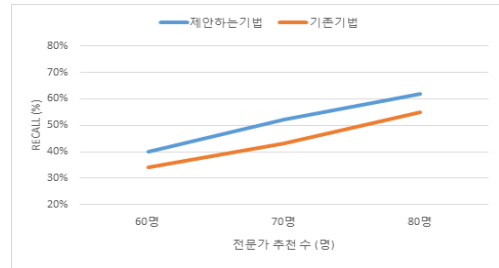


그림 5. 전문가 추천 수 변화에 따른 재현율

[그림 5]는 전문가 추천 수 변화에 따른 재현율을 나타낸다. 재현율은 전문가로 분류된 항목들 중 실제 전문가로 추천된 항목의 비율이다. 전문가 추천 수는 60명에서 80명까지 10명 단위로 변화시키며 측정하였다. 성능 평가 결과 기존기법에 비해 제안하는 기법이 우수한 성능을 보이는 것을 확인 하였다. 기존기법에 비해 평균 10.6%정도 높은 재현율을 보였다. 그 이유는 기존 기법은 사용자 인적 관계와 응답 품질을 고려하지 않았다. 또한 사용자 질의에 대한 계층적 관계를 포함하지 않았기 때문에 전문가 지수가 낮게 평가되었기 때문이다.

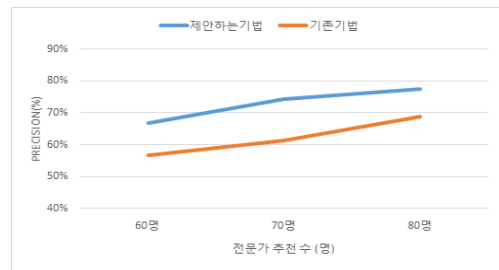


그림 6. 전문가 추천 수 변화에 따른 정밀도

[그림 6]는 전문가 추천 수 변화에 따른 정밀도를 나타낸다. 정밀도는 전문가로 추천된 결과들 중 전문가로 분류된 결과물의 비율이다. 전문가 추천 수는 60명에서 80명까지 10명 단위로 변화시키며 측정하였다. 성능 평가 결과 기존기법에 비해 제안하는 기법이 우수한 성능을 보이는 것을 확인 하였다. 기존기법에 비해 평균

7.3%정도 높은 정밀도를 보였다. 기존 기법은 사용자 인적 관계와 응답 품질을 고려하지 않았다. 또한 사용자 질의에 대한 계층적 관계를 포함하지 않았기 때문에 전문가 지수가 낮게 평가되었다.

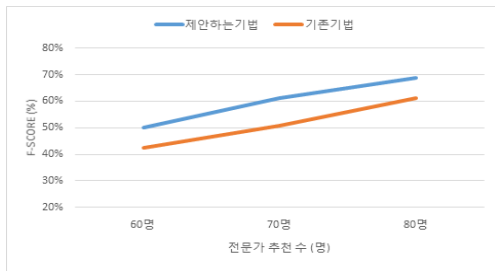


그림 7. 전문가 추천 수 변화에 따른 F-score

[그림 7]는 전문가 추천 수 변화에 따른 F-score를 나타낸다. F-score는 재현율과 정밀도 두 가지 지수를 통계적으로 종합하여 계산한다. 전문가 추천 수는 60명에서 80명까지 10명 단위로 변화시키며 측정하였다. 성능 평가 결과 기존기법에 비해 제안하는 기법이 우수한 성능을 보이는 것을 확인하였다. 기존기법에 비해 평균 8.3%정도 높은 F-score를 보였다. 그 이유는 기존기법은 사용자 인적 관계와 응답 품질을 고려하지 않았다. 또한 사용자 질의에 대한 계층적 관계를 포함하지 않았기 때문에 전문가 지수가 낮게 평가되었기 때문이다.

기존의 전문가 추천 기법은 사용자의 문서만을 고려해 추천하기 때문에 정확한 전문가를 판단하기 어렵다. 하지만 제안하는 전문가 추천 기법은 사용자의 관심 분야, 인적 관계, 응답 품질을 고려하여 전문가를 추천하기 때문에 재현율, 정밀도, F-score 모두 향상되었음을 볼 수 있다.

## V. 결론

본 논문에서는 소셜 네트워크 환경에서 사용자의 관심분야, 인적 관계, 응답 품질을 고려하여 전문가를 추천하는 기법을 제안하였다. 사용자의 관심분야 지수는 사용자 활동을 분석하여 키워드를 추출하고 신뢰도를 계산하여 도출하였고 인적 관계 지수는 사용자의 인적

관계를 이용하여 인적 관계 지수를 도출하였다. 응답 품질 지수는 사용자가 얼마나 응답을 빠르고 정확하게 했는지를 판단하여 도출하였다. 관심분야 지수, 인적 관계 지수, 응답 품질 지수를 합하여 전문가 지수를 도출하고 단어의 계층적 관계를 이용하여 사용자 검색 질의가 요청되었을 때 전문가 그룹과 매칭하여 결과를 도출하는 방법을 제안하였다. 기존의 기법에 비해 재현율은 10.6%만큼 향상 되었고 정밀도는 7.3% 만큼 F-score는 8.3% 향상된 것을 확인할 수 있었다. 제안하는 기법을 통해 상품이나 새로운 연구에 대한 전문가의 의견을 들을 수 있는 질의응답 시스템으로 활용 가능하다. 본 논문에서는 정규분포를 이용한 임의로 생성한 데이터를 사용하여 실험평가를 진행하였다. 실험에 진행된 데이터의 신뢰도를 증명하기 위하여 실험 데이터에 대한 추가 성능 평가를 진행할 필요가 있고, 또한 다양한 기법과 비교하여 실험할 필요가 있다. 본 논문에서 사용된 TF-IDF 방식으로는 실제로 사용자가 작성한 글의 의도한 내용을 파악하기에는 부족하기에 사용자의 최근 활동 내용 분석에 LDA(Latent Dirichlet Allocation) 기법을 사용하여 의미 있는 단어를 찾아 분석하는 방법을 연구할 예정이다.

## 참고 문헌

- [1] M. R. Morris, J. Teevan, and K. Panovich, "What do people ask their social networks, and why?: a survey study of status message q&a behavior," Proc. International Conference on Human Factors in Computing Systems, pp.1739-1748, 2010.
- [2] Morris, Meredith Ringel, Jaime Teevan, and Katrina Panovich, "A Comparison of Information Seeking Using Search Engines and Social Networks," ICWSM 10, pp.23-26, 2010.
- [3] Duchateau, Fabien, "Who can best answer a query in my social network?," Data Engineering Workshops (ICDEW), 2011 IEEE 27th



- International Conference on. IEEE, 2011.
- [4] Y. Joung, S. M. Chen, C. Wu, and T. H. Chiu, "A Comparative Study of Expert Search Strategies in Online Social Networks," Proc. International Conference on Advanced Information Networking and Applications, pp.960-967, 2013.
- [5] D. Liu, L. Wang, J. Zheng, K. Ning, and L. Zhang, "Influence Analysis Based Expert Finding Model and Its Applications in Enterprise Social Network," Proc. International Conference on Services Computing, pp.368-375, 2013.
- [6] S. D. Gollapalli, P. Mitra, and C. L. Giles, "Ranking experts using author-document-topic graphs," Proc. ACM/IEEE-CS Joint Conference on Digital Libraries, pp.87-96, 2013.
- [7] T. Vu and A. Baid, "Ask, don't search: A social help engine for online social network mobile users," Proc. IEEE Samoff Symposium, pp.1-5, 2012.
- [8] A. Omidvar, M. Garakani, and H. R. Safarpour, "Context based user ranking in forums for expert finding using WordNet dictionary and social network analysis," Information Technology and Management, Vol.15, No.1, pp.51-63, 2014.
- [9] 송창우, 김종훈, 정경용, 류중경, 이정현, "시맨틱 웹에서 개인화 프로파일을 이용한 콘텐츠 추천 검색 시스템," 한국콘텐츠학회논문지, 제8권, 제1호, pp.318-327, 2008.
- [10] 이서희, 박윤정, 한진수, 최도진, 임종태, 복경수, 유재수, "논문의 중요성 및 품질을 이용한 학술 전문가 검색 기법," 한국콘텐츠학회논문지, 제16권, 제12호, pp.458-467, 2016.
- [11] I. B. Jeon, J. T. Lim, H. Li, K. S. Bok, and J. S. Yoo, "A Reliable Expert Search Scheme in Social Network Environments," Proc. International Conference on Convergence Content, pp.245-246, 2013.
- [12] 전인배, 임종태, 복경수, 유재수, "소셜 네트워크 환경의 신뢰성 향상을 위한 전문가 검색 기법,"

정보과학회:컴퓨팅의 실제 및 레터, 제20권, 제5호, pp.311-315, 2014.

- [13] S. Wang, S. Xie, X. Zhang, Z. Li, P. S. Yu, and X. Shu, "Future Influence Ranking of Scientific Literature," Proc. SIAM International Conference on Data Mining, pp.749-757, 2014.
- [14] S. K. Rani, K. Raju, and V. V. Kumari, "Expert Finding System using Latent Effort Ranking in Academic Social Networks," International Journal of Information Technology and Computer Science, Vol.7, No.2, pp.21-27, 2015.
- [15] A. Bozzon, M. Brambilla, S. Ceri, M. Silvestri, and G. Vesci, "Choosing the right crowd: expert finding in social networks," Proc. International Conference on Extending Database Technology, pp.637-648, 2013.

#### 저자 소개

##### 송희섭(Heesub Song)

준회원



스 시스템 등

- 2016년 2월 : 한림대학교 전자물리학과(이학사)
  - 2016년 3월 ~ 현재 : 충북대학교 빅데이터 협동과정(석사과정)
- <관심분야> : 소셜 미디어, 소셜 네트워크, 빅데이터, 데이터베이

##### 유승훈(Seunghun Yoo)

준회원



템, 빅데이터 등

- 2016년 2월 : 충북대학교 정보통신공학과(공학사)
  - 2016년 3월 ~ 현재 : 충북대학교 정보통신공학과(석사과정)
- <관심분야> : 데이터베이스 시스템, 분산 컴퓨팅, 인-메모리 시스템, 빅데이터 등

정 재 윤(Jaeyun Jeong)

준회원



- 2016년 2월 : 충북대학교 정보통신공학과(공학사)
- 2016년 3월 ~ 현재 : 충북대학교 정보통신공학과(석사과정)

<관심분야> : 그래프 분석, 빈발 패턴 마이닝, 데이터베이스 시스템, 분산 컴퓨팅 등

박 재 열(Jaeyeol Park)

정회원



- 2014년 2월 : 충북대학교 정보통신공학과(공학사)
- 2016년 2월 : 충북대학교 정보통신공학과(공학석사)
- 2017년 3월 ~ 현재 : 충북대학교 정보통신공학과(박사과정)

<관심분야> : 데이터베이스 시스템, RDF, 실체화 뷰, 빅데이터 등

안 지 환(Jihwan Ahn)

준회원



- 2016년 2월 : 충북대학교 정보통신공학과(공학사)
- 2017년 3월 ~ 현재 : 충북대학교 정보통신공학과(석사과정)

<관심분야> : 소셜 미디어, 소셜 네트워크, 추천 시스템, 빅데이터 등

임 종 태(Jongtae Lim)

정회원



- 2009년 2월 : 충북대학교 정보통신공학과(공학사)
- 2011년 2월 : 충북대학교 정보통신공학과(공학석사)
- 2015년 8월 : 충북대학교 정보통신공학과(공학박사)

<관심분야> : 데이터베이스 시스템, 시공간 데이터 베이스, 위치기반 서비스, 모바일 P2P 네트워크, 빅데이터 등

복 경 수(Kyungsoo Bok)

종신회원



- 1998년 2월 : 충북대학교 수학과(이학사)
- 2000년 2월 : 충북대학교 정보통신공학과(공학석사)
- 2005년 2월 : 충북대학교 정보통신공학과(공학박사)

▪ 2005년 3월 ~ 2008년 2월 : 한국과학기술원 전산학과 Postdoc

- 2008년 3월 ~ 2011년 2월 : (주)가인정보기술 연구소
- 2011년 3월 ~ 현재 : 충북대학교 정보통신공학과 초빙교수

<관심분야> : 데이터베이스 시스템, 이동객체 데이터베이스, 소셜 네트워크, 빅데이터 등

유 재 수(Jaesoo Yoo)

종신회원



- 1989년 2월 : 전북대학교 컴퓨터공학과(공학사)
- 1991년 2월 : KAIST 전산학과(공학석사)
- 1995년 2월 : KAIST 전산학과(공학박사)

▪ 1995년 3월 ~ 1996년 8월 : 목포대학교 전산통계학과(전임강사)

▪ 1996년 8월 ~ 현재 : 충북대학교 정보통신공학부 및 컴퓨터정보통신연구소 교수

▪ 2009년 3월 ~ 2010년 2월 : 캘리포니아주립대학교방문교수

<관심분야> : 데이터베이스시스템, 빅데이터, 센서 네트워크 및 RFID, 소셜 네트워크 서비스, 분산 객체 컴퓨팅, 바이오인포매틱스 등