

소셜 네트워크 환경에서 사용자 행위 분석을 통한 신뢰성 높은 사용자 검색 기법

A Reliable User Search Scheme Considering User Behavior Analysis in Social Networks

노 연 우*, 김대윤**, 임 종 태*, 복 경 수*, 유 재 수*[†]
충북대학교 정보통신공학부*,
충북대학교 빅데이터학과**

Yeon-woo Noh*, Daeyun Kim**, Jongtae Lim*,
Kyoungsoo Bok*, Jaesoo Yoo*[†]
School of Information and Communication
Engineering Chungbuk National University*,
Department of Big Data, Chungbuk National
University**

요약

최근 소셜 네트워크 서비스의 이용이 증가함에 따라 서비스의 질을 극대화하기 위하여 신뢰할 만한 글들을 사용자에게 제공하기 위한 연구가 활발히 이루어지고 있다. 본 논문에서는 소셜 네트워크 환경에서 사용자의 행위를 기반으로 악의적 사용자들을 배제하고 신뢰성 높은 사용자들을 검색하는 기법을 제안한다. 신뢰성 높은 사용자들을 검색함으로써 사용자들에게 신뢰할 만한 글들을 제공할 수 있고 소셜 네트워크의 특성을 함께 고려함으로써 기존 연구보다 정확하게 검색 가능하게 하였다.

I. 서론

최근 모바일 기기의 발달로 인해 소셜 네트워크 서비스(SNS: Social Network Service)가 급속히 발전하고 있다. 수많은 사람들이 SNS를 활발히 이용함에 따라 시간으로 방대한 양의 데이터들이 생성되고 있다. 이러한 대량의 SNS 데이터로부터 신뢰할 만한 글들을 찾는 것이 중요한 문제로 제기되면서 신뢰성 높은 사용자들을 검색하는 연구가 활발하게 이루어지고 있다.

대표적인 사용자 검색 기법은 소셜 네트워크의 특성만을 고려하여 평점을 계산한 후 평점이 높은 사용자들을 검색한다[1]. 사용자를 검색하기 위해서 사용자의 프로필과 태그 정보를 이용하여 유사도를 측정하고, 이 유사도를 이용하여 질의와 매칭한다. 하지만 기존의 기법은 소셜 네트워크 사용자의 행위를 고려하지 않아 사용자 검색 결과의 정확도를 보장할 수 없다.

본 논문에서는 소셜 네트워크 환경에서 소셜 네트워크의 특성과 사용자의 행위 분석을 통한 신뢰성 높은 사용자 검색 기법을 제안한다. 신뢰성 높은 사용자가 작성하는 글들은 스팸성이 낮아 신뢰할 만한 글들이다. 따라서 제안하는 기법에서는 소셜 네트워크의 환경 자체의 특성과 소셜 네트워크에 가장 큰 영향을 미치는 요인인 사용자의 행위 분석을 함께 고려하여 사용자 검색 기법 결과의 정확도를 높인다.

[†] 교신저자 : yjs@chungbuk.ac.kr

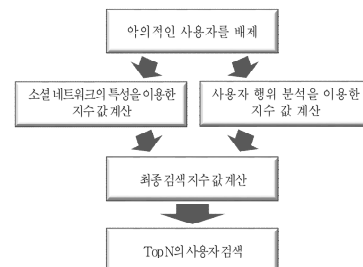
본 연구는 미래창조과학부 및 정보통신기술진흥센터의 대학 ICT연구센터육성 지원사업의 연구(IIIP-2016-H8501-16-1013)와 2015년도 산업통상자원부의 재원으로 한국에너지기술평가원(KETEP)의 에너지인력양성사업으로 지원받아 수행한 인력양성 성과(No. 20144030200450)임.

II. 제안하는 사용자 검색 기법

1. 개요

SNS 환경에서는 신뢰성과 연관되는 다양한 요인들이 존재하지만 SNS를 사용하는 사용자의 행위가 신뢰성과 가장 큰 연관성을 가진다[2]. 제안하는 기법에서 신뢰성 높은 사용자 검색을 위해 소셜 네트워크의 특성인 팔로워, 리트윗, 멘션뿐만 아니라 사용자의 행위 분석을 이용하여 사용자를 검색한다. 특정 사용자의 트윗 글에 대한 나머지 사용자들의 행위를 분석했을 때, 그 행위의 긍정성이 높다고 판단되면 사용자의 신뢰성이 높다고 판별한다.

그림 1은 제안하는 기법의 전체적인 처리 절차이다. 우선 스팸성 높은 글을 작성하는 악의적인 사용자들을 배제한다. 그리고 소셜 네트워크의 특성과 사용자의 행위를 분석하여 가중치 값을 계산하는데, 사용자의 행위는 긍정적 행위와 부정적 행위로 나뉜다. 행위 수에 가중치를 곱한 값을 최종적으로 사용자에게 부여한 후, 이를 랭킹하여 신뢰성 높은 사용자들을 검색한다.



▶▶ 그림 1. 사용자 검색 처리 절차

2. 사용자의 행위 분석을 통한 신뢰성 높은 사용자 검색

기존 사용자 검색 기법에서는 단순히 소셜 네트워크만을 기준으로 사용자를 검색하기 때문에 최종 검색된 결과의 높은 정확도를 보장하는 것이 어려웠다. 소셜 네트워크 사용의 주체자인 사용자의 행위를 함께 분석함으로써 결과의 정확도를 높이는 것이 중요하다. 본 논문에서는 악의적 사용자들을 사전에 배제시키고 소셜 네트워크의 특성에 기반한 사용자의 행위 분석을 통해 신뢰성 높은 사용자를 검색하는 기법을 제시한다.

스팸성 높은 글들을 작성하는 사용자를 사전에 배제시켜 글 자체의 신뢰성을 보장한다. 악의적 사용자는 일반 사용자에 비해 맹목적으로 다른 계정을 팔로우하는 경우가 많다. 따라서 팔로워에 근거하여 팔로워의 수가 팔로잉하고 있는 수보다 작게 되면 이는 악의적인 사용자일 확률이 높다고 판단하였다. 수식 1은 악의적 사용자를 배제하기 위한 수식이며, 0에 가까울수록 스팸성 높은 글을 작성할 확률이 높다.

$$P_{U_A} = \frac{\sum_{f \in U_A^f} Followers}{\sum_{f \in U_A^f} Followers + \sum_{f \in U_A^f} Followee} \quad (1)$$

소셜 네트워크의 특성인 팔로워, 리트윗, 멘션 요소를 이용하여 이 요소들의 지수 값이 높을수록 사용자 검색에 있어서 더 높은 정확도를 제공한다[3].

수식 2는 팔로워 수를 나타내며, 높은 팔로워 수를 가질수록 사용자의 트윗에 대한 신뢰도가 높다는 것을 의미한다.

$$I_{U_A^f} = \frac{\sum_{f \in U_A^f} Followers + 2}{\alpha} \quad (2)$$

수식 3은 리트윗 수와 전파력을 나타내며, 이 값이 클수록 사용자가 남긴 글에 대한 신뢰도가 크다고 측정하였다.

$$I_{U_A^r} = \frac{\left(\frac{\sum_{rt \in U_A^r} Retweets}{\sum_{t \in U_A^r} Tweets} \times \frac{\sum_{U_f \in U_A^f} Followers}{\sum_{f \in U_A^f} Followers} \right) + 2}{\beta} \quad (3)$$

수식 4는 멘션 수로써, 멘션 수가 많을수록 신뢰도가 높다.

$$I_{U_A^m} = \frac{\sum_{m \in U_A^m} Mentions + 2}{\gamma} \quad (4)$$

수식 5는 이 세 요소를 모두 합한 값으로 소셜 네트워크의 특성을 나타내는 지수이다.

$$I_{U_A} = \log(I_{U_A^f}) + \log(I_{U_A^r}) + \log(I_{U_A^m}) \quad (5)$$

소셜 네트워크 환경에서 사용자를 검색함에 있어서 사용자의 행위를 분석하는 것은 중요하다. 기존 연구에서는 소셜 네트워크적인 요소만 고려하여 사용자 검색 결과의 높은 정확도를 보장할 수 없었다. 제안하는 기법에서 활용하는 사용자의 행위는 트윗 답아가기, 트윗 링크

복사하기, 쪽지로 공유하기와 같은 긍정적 행위와 뮤트하기, 차단하기, 신고하기와 같은 부정적 행위로 구분할 수 있다. 각 행위 요소에 부여되는 가중치는 상대적 중요도에 의해 판단된다. 그림 2는 사용자 행위에 따른 가중치 값을 나타내고, α 는 0보다 큰 임의의 값이며 β 는 0보다 작은 임의의 값이다.

사용자 행위	가중치
트윗 답아가기	6α
트윗 링크 복사하기	2α
쪽지로 공유하기	α
뮤트하기	β
차단하기	2β
신고하기	6β

▶▶ 그림 2. 사용자 행위에 따른 가중치

수식 6은 각 행위 수에 각각의 가중치를 곱하여 나온 값을 의미하며 이 값이 높을수록 긍정적인 행위를 많이 한 사용자로 판단되어 신뢰성 높은 사용자 검색에 큰 영향을 미친다.

$$B_{U_A} = (6\alpha \times B_1) + (2\alpha \times B_2) + (\alpha \times B_3) + (\beta \times B_4) + (2\beta \times B_5) + (6\beta \times B_6) \quad (6)$$

최종적으로 악의적인 사용자를 배제하는 수식과 소셜 네트워크의 특성을 이용한 지수 값을 계산하는 수식과 사용자 행위를 분석하여 도출된 값을 모두 합하여 사용자 검색 지수 값을 계산하게 된다. 수식 7은 최종 사용자 검색 지수 값을 나타내며, 이 값을 기준으로 사용자를 랭킹화한 후 Top N의 신뢰성 높은 사용자를 검색한다.

$$Search_{total} = P_{U_A} + I_{U_A} + B_{U_A} \quad (7)$$

IV. 결론

본 논문에서는 트위터 사용자의 행위 분석을 통한 사용자 검색 기법을 제안하였다. 제안하는 기법에서는 소셜 네트워크의 특성과 트위터 사용자의 행위 분석을 통해 이를 가중치로 부여함으로써 신뢰할 만한 글들을 작성하는 사용자를 검색하였다. 사용자 행위 분석 시 긍정적인 행위를 많이 할수록 행위의 긍정성이 높다고 판단하고 그 사용자는 신뢰성이 높다고 판별하였다. 향후 연구는 제안하는 기법의 우수성을 보이기 위해 다양한 성능평가를 수행하는 것이다.

■ 참고 문헌 ■

- [1] F. Duchateau, "Who can best answer a query in my social network?," Proc. IEEE International Workshop on Data Engineering, pp.218-223, 2011
- [2] E. Lee, A Study on the Factors Influencing upon SNS Credibility, Sejong University, 2012. (in Korean)
- [3] 노연우, 전현욱, 육미선, 한지은, 임종태, 김연우, 복경수, 유재수 "소셜 네트워크에서 사용자의 영향력을 고려한 신뢰성 높은 핫 토픽 검출 기법", 한국콘텐츠학회 종합 학술대회 논문집, pp. 71-72, May 2015