

소셜 네트워크 분석을 통한 소외계층 발견 방법

최동진*, 김정인*, 김판구*

*조선대학교 컴퓨터공학과

e-mail: dongjin.choi84@gmail.com, pkkim@chosun.ac.kr

A Method for Detecting Outlier Communities in Social Networks

Dongjin Choi*, Jeongin Kim*, Pankoo Kim*

*Dept of Computer Engineering, Chosun University

요 약

본 논문은 소셜 네트워크 서비스 내에 존재하는 사용자간의 인맥 네트워크 정보와 사용자들 간에 주고받는 메시지의 긍정, 부정 성향 분석을 통하여 상대적으로 고립되거나 소외된 사용자를 발굴하는 방법에 대한 내용을 다룬다. 현재 소셜 네트워크 서비스를 이용하여 인맥 네트워크를 분석하거나 특정 인물의 영향력을 분석 또는 화두가 되는 이슈 및 트렌드를 추출하는 연구에 초점이 맞춰지고 있다. 사회적으로 크게 대두되고 있는 학교폭력, 자살, 왕따와 같은 문제점을 해결하기 위하여 소셜 네트워크 분석을 통한 고립 및 소외계층 발견 기법이 필요하다. 따라서 본 논문에서는 기존의 소셜 네트워크 분석기법으로 해결할 수 없는 소외계층 발견 방법을 제안한다.

1. 서론

소셜 네트워크 서비스(SNS)는 페이스북, 트위터, 싸이월드 등과 같이 사용자 간의 자유로운 소통과 정보 공유, 인맥관계 형성을 통해 온라인상에서 사회적 관계를 생성하고 유대 시키는 온라인 플랫폼을 의미한다. SNS는 스마트폰의 활발한 보급과 무선인터넷망의 급격한 발달로 인하여 기하급수적으로 확산되어 페이스북 기준으로 그 사용자는 약 10억 명 이상에 달한다. 사용자 자신이 겪는 일상적인 내용에서부터 지식 경험에까지 다양한 내용을 SNS에 기록하게 되는데 대부분의 사용자는 자신의 심리적 성향을 직접 또는 간접적으로 표현하게 된다. 또한, 주변 지인들과의 텍스트기반 의사소통을 통해 관심사를 공유하고 정보를 전달한다. 이렇듯 사용자와 주변 인물들 간의 사회적으로 밀접한 데이터를 내포하고 있는 SNS 데이터는 최근 많은 연구자들의 관심을 독차지 하고 있다. 그 결과 다양한 분야에서 SNS를 대상으로 많은 연구가 진행되고 있는데 그 중 가장 대표적으로 인맥관계 네트워크 분석을 통한 네트워크 모니터링 연구[1]와 사용자의 영향력을 분석하는 연구[2], 화두가 되는 이슈 및 트렌드를 분석하는 연구[3]가 있다. 최근 사회적으로 학교폭력 및 자살, 왕따와 같은 고립 또는 소외계층에 대한 문제점이 크게 대두되고 있지만 오프라인 상에서 사후 조치에 급급할 뿐 직접적인 예방책과 해결방안은 아직 두각을 보이지 못하고 있다. 본 논문은 SNS 내에 존재하는 사용자 인맥 네트워크 정보와 의사소통 정보 분석을 통하여 메시지의 긍정 혹은 부정 성향을 판단하여 상대적으로 고립되거나 소외된 계층을 발견하는 방법을 설계한다. 본 논문은

SNS상에서 소외계층 발견이라는 아직 연구되지 않은 새로운 분야에 관한 내용으로 실험 내용을 다루고 있지 않지만 본 논문에서 제안하는 방법을 다양한 예시와 함께 기술한다. 논문의 구성으로 SNS 데이터를 이용한 타 연구의 동향과 소외계층 발견 기술의 필요성을 제 2장에서 기술하고 3장에서는 본 논문의 핵심내용인 소외계층 발견 방법에 대하여 설명한다. 마지막으로 4장에서 결론과 향후 연구를 제시하며 맺는다.

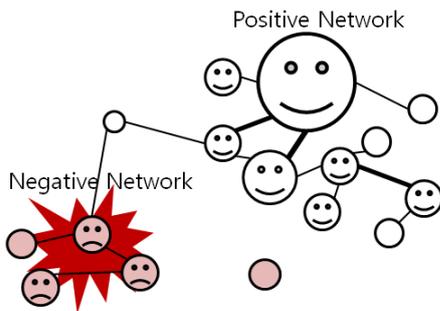
2. 관련연구

급속도로 빠르게 확산되고 있는 소셜 네트워크 서비스는 사회 과학자에서부터 물리학자, 컴퓨터과학자 등 다양한 분야에서 활발히 연구가 수행되고 있다. 대표적으로 SNS 데이터를 통하여 인맥 네트워크를 분석하는 연구를 바탕으로 친구, 상품, 전문가 등의 다양한 추천서비스[4, 5]를 제공하고 있고 더불어, 인맥 네트워크간의 밀접도 또는 그 강도를 분석하는 연구[1, 6]도 수행되고 있다. 또한, 소셜 미디어 데이터 분석을 통하여 대중이 관심 갖는 정보를 추출하고 정보의 흐름을 파악하는 연구가 다음(Daum)과 솔트룩스(Saltlux)를 중심으로 국내에서 진행되고 있고 국외의 경우, 구글(Google)과 페이스북을 중심으로 대용량의 웹문서와 소셜 데이터를 분석하여 트렌드 및 이슈를 추출하고 동향을 파악하는 연구가 진행되었다. 이렇듯, 소셜 네트워크 데이터는 국내외를 불문하고 많은 학자들에게 폭발적인 관심을 갖으며 과거 시도하지 못하였던 다양한 연구를 수행하고 있다. 최근, 국내에서 학교폭

력으로 인한 자살 및 왕따와 같은 고립 또는 소외계층에 대한 문제가 크게 대두되고 있는 상황에서 SNS 데이터는 사회문제 해결에 큰 가능성을 내포하고 있음은 분명하다. 그 대표적인 예로 모바일 가디언 서비스[7]는 문자메시지 내에 폭력성 글이 포함되어 있을 경우 해당 메시지를 부모 휴대폰으로 자동 전송 하여 미연에 방지할 수 있는 해결책을 제시하고 있다. 자신의 인맥 네트워크와의 상호 주고받는 메시지의 긍정과 부정 성향을 분석할 수 있고, 위치정보와 시간정보와 같은 세부적 고급 태그정보를 확보할 수 있는 SNS는 분명, 현대사회의 문제를 해결할 수 있는 기초 데이터로 활용할 수 있다.

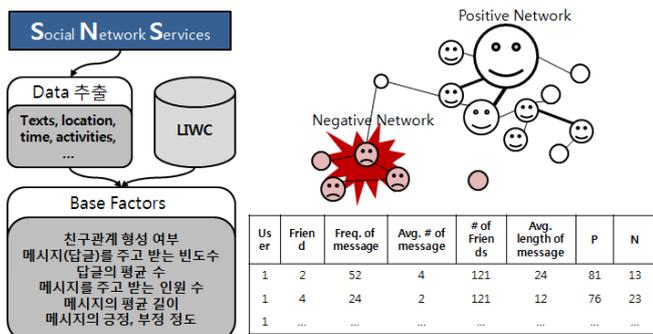
3. 소셜 네트워크 분석을 통한 소외계층 발견 방법 설계

본 장에서는 SNS 데이터 분석을 통하여 소외계층을 발견하는 방법을 설계한다. 소외계층은 상대적으로 고립되거나 그 영향력이 다소 낮은 그룹을 일컫는 것으로 그림 1과 같이 표현될 수 있다.



(그림 1) 일반 계층과 소외계층 구성 예

위 그림에서 알 수 있듯이 일반계층은 사회적 영향력이 큰 인물을 중심으로 긍정적인 메시지를 주고받는 그룹이 형성될 수 있다. 반면, 소외계층의 경우는 일반계층과의 사회적 연결망이 빈약하고 상대적으로 부정적인 메시지를 주고받는 경향이 나타날 수 있다. 기존의 연구들이 Positive 네트워크를 발견하고 분석하는 연구에 초점이 맞추어져 있다면 본 논문은 Negative 네트워크를 발견하는 방법에 초점을 맞춘다.



(그림 2) 소외계층 발견 방법 및 흐름도

소외계층 발견을 위하여 그림 2와 같이 전체 흐름도가 구성될 수 있는데 우선적으로 SNS 데이터 수집을 통하여 데이터 추출이 선행되어야 한다. 추출된 데이터를 바탕으로 친구관계 여부를 판단하고 지인들과 메시지 또는 답글을 주고받는 빈도수를 산출한다. 또한, 자신이 작성한 글에 지인이 답글을 올리는 평균 글 수와 메시지를 주고받는 지인의 수, 메시지의 평균 길이, 그리고 메시지의 긍정, 부정의 정도를 LIWC(Linguistic Inquiry and Word Count)를 이용하여 측정할 수 있다. 위 6가지 항목을 적용하는 것을 매우 까다로운 작업으로 차후 실험을 통하여 각 항목들이 지닐 수 있는 가중치를 산정해야만 한다. 이와 같은 방법으로 SNS 내에서 고립 또는 소외계층을 발견할 수 있고 그 데이터는 user, friend, Freq. of message, Avg. number of message, number of friends, Avg. length of message, positive, negative로 구성될 수 있다. 공통된 관심사를 공유하는 노드들 간에 긍정의 메시지 정도가 높을 경우, 노드들 간의 친밀도가 높은 Positive 네트워크로 구성된다는 전제조건을 따르고, 공통된 관심사를 공유하지만 부정의 메시지 정도가 높고, 메시지 교환 횟수가 낮으면 이 그룹을 소외그룹으로 판단한다. 위 전제조건은 향후 사회과학 분야 전문가의 조언으로 수정 보완이 요구된다.

4. 결론

본 논문은 현재 다양한 분야에서 화두가 되고 있는 소셜 네트워크 서비스 데이터를 바탕으로 기존에 실시되지 않았던 소외계층이라는 부정적 고립형 그룹을 발견하는 방법을 설계하였다. 트위터 데이터를 이용하여 영향력을 분석하고 화제가 되는 이슈를 추출하는 연구가 대부분을 이루지만 약 10억여 명이 사용하는 페이스북 데이터를 사용하였을 경우 그 잠재력은 대단할 것으로 예측된다. 또한, 페이스북 내에는 “좋아요”라는 부가적 태깅 기능과 개방적 댓글 서비스로 인하여 보다 적극적이고 활발한 인맥 네트워크가 형성되어 있다. 하지만 페이스북의 사내방침에 따라 데이터를 함부로 사용할 수 없다는 제약점이 있지만 연구목적으로 일부분 개방가능하기에 본 논문의 제안점을 실험하기에는 충분할 것으로 판단된다. 3장에서 기술하는 전제조건과 6가지 항목을 바탕으로 향후 1차적으로 트위터 데이터를 대상으로 소외계층을 발견하고 이후, 페이스북 데이터를 이용하여 보다 신뢰성 높은 실험을 실시할 예정이다.

감사의 글

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(No.2010-0011656) 또한, 교육과학기술부와 한국연구재단의 지역혁신인력양성사업으로 수행된 연구결과임(2011-04-기-03-023).

참고문헌

- [1] 강윤섭, 최승진, “공통 이웃 그래프 밀도를 사용한 소셜 네트워크 분석”, 정보과학회논문지제16권 제4호, 2010.04, 432-436
- [2] 박건우, 임선열, 이태양, 김정훈, 최웅철, “감성 단어 분석 모듈을 이용한 모바일 소셜 네트워크 서비스에서의 성향 판단 시스템 구현”, 한국정보과학회 학술발표논문집 제38권 제2호, 2011.11, 97-100
- [3] Jeongin Kim, Byeongkyu Ko, Pankoo Kim, “News Twitter-Based Topics Extraction and Comparison,” International Conference on SMA 2012, p.42, Aug 2012.
- [4] 박상원, 최은정, 박민수, 김정규, 서은석, 박영택, “온톨로지 기반 소셜 네트워크 분석을 이용한 전문가 추천 시스템”, 정보과학회논문지 제15권 제5호, 2009.05, 390-394
- [5] 조도은, 김시정, 곽윤식, “사용자 선호도 학습을 이용한 개인화 콘텐츠 추천 방법 연구”, 한국정보기술학회지 제9권 제9호, 2011, 229-235
- [6] 설광수, 김정동, 심형남, 백두권, “소셜네트워크 서비스 사용자 간의 친밀도 측정 기법 및 실험”, 정보과학회논문지, 제39권 제4호, 2012, 335-341
- [7] <http://www.mguardian.co.kr/>