

SNS 사용자간 다차원 상호관계분석

Analysis of multi-dimensional interaction among SNS users

이 경 민* 남 궁 현** 김 응 희*** 이 강 용**** 김 흥 기*****
Kyungmin Lee Hyun Namgoong Eung Hee Kim KangYong Lee Hong-Gee Kim

요 약

최근, 온라인상에서 사용자간 관계형성 및 커뮤니케이션을 지원하는 Social Network Service(SNS)들이 주목을 받고 있다. SNS에서 형성되는 사용자 정보와 그들의 활동에 대한 정보는 사용자에 대한 특징적인 정보로서 추천과 같은 외부서비스에서 재사용될 수 있는 유용한 정보로 여겨진다. 기존의 사용자 기본 프로필 정보를 통한 분석이나, 친밀도 및 유사도를 도출하기 위한 단편적인 분석 방법은 가공된 정보의 활용에 있어 한계점을 지닌다. 본 논문에서는 사용자의 SNS활동들로부터 사용자간 상호영향을 도출하고 다차원 상호관계(Multi-dimension interaction)를 표현 할 수 있는 통합 된 프로필 모델을 정의하고 이를 도출하기 위한 시스템을 제안한다. 제안된 시스템은 대표적인 SNS인 트위터(Twitter.com)부터 사용자 정보를 수집하여 정보의 재가공을 수행한다. 본 논문의 실험에서는 가공된 정보를 통해 얻을 수 있는, 기존 방식에서는 드러나지 않던 사용자간의 새로운 관계들에 대해 설명한다. 가공된 정보는 특정 사용자의 통합 프로필로서 콘텐츠 아이템 추천과 같은 외부서비스에서 활용가능 할 것으로 기대된다.

ABSTRACT

Social Network Service(SNS) has become a hot trend as a web service which helps users construct social relationships in the web and enables online communication. The information about user activities and behaviors obtained from the SNSs is expected to be an useful knowledge source for other services such as recommendation services. Most of previous researches on SNS rely on analyzing overall network topology and surveying the activities in a one-dimensional aspect. This paper propose a system for measuring multi-dimensional interaction through the activities in a SNS. The proposed system delivers an unified profile (consisting of profile and multi-dimension interaction) model from user-activities in Twitter.com. At the experimental section, some meaningful perspectives on a set of the unified profiles are described.

☞ keyword : Social Network Service(소셜 네트워크 서비스), Web(웹), Twitter(트위터), Social Network Analysis(소셜 네트워크 분석)

1. 서 론

Social Network Service(SNS)는 사용자들이 온라인에서 사회적 관계형성을 지원하고 이를 통한

커뮤니케이션 서비스를 제공하는 온라인 서비스이다. SNS는 (1) 지정된 특정 사용자 간의 정보 공유를 지원하는 폐쇄적 형태의 관계 그리고 (2) 모든 사용자와의 정보 교환을 허용하는 개방적인 관계를 제공한다. 관계 설정을 통해 사용자 간 사회적 관계를 명시하고 유지/관리 할 수 있다는 점으로 인해 SNS는 최근 대중으로부터 반향을 불러일으키고 있다[1].

최근 연구들에 따르면, 실세계(offline)에서 맺는 사람들의 관계 세기를 측정하는 친밀도, 유사도 등의 요소가 온라인상의 SNS에서도 적용가능하다고 밝혀졌다[2]. 또한, SNS사용자들 사이에서 작성된 정보를 분석하여, 사용자 관계성향을 도출하고자 하는 방법들이 국내/외에서 활발히 제안

* 정 회 원 : 서울대학교 협동과정 의료정보학과 석사과정
sephiroce@snu.ac.kr(투고자, 책임저자)

** 정 회 원 : 서울대학교 치의학과 박사과정
ngh@snu.ac.kr

*** 정 회 원 : 서울대학교 협동과정 의료정보학과
박사과정 eungheekim@snu.ac.kr

**** 정 회 원 : 한국전자통신연구원 방송통신융합연구부
연구원 kanglee@etri.re.kr

***** 정 회 원 : 서울대학교 의료정보학과 부교수
hgkim@snu.ac.kr(교신저자)

[2010/12/24 투고 - 2010/12/31 심사 - 2010/03/08 심사완료]

되고 있다. SNS로 부터 수집 된 사용자 프로필 정보를 분석하여 사용자간 친밀도를 산정하거나 [3,4], SNS 내에서 행하여지는 태그, 북마크, 친구, 블로그 등, 사용자의 활동으로부터 사용자 간 유사도를 측정하는 방법들이 연구되고 있다[5].

그러나, 기존의 연구들에서 산출된 정보는 SNS의 특징적인 정보인 (1) 사용자들의 관계와 (2) 실시간 배포 되는 단문 포스팅(Posting), (3) 단문 포스팅을 이용한 사용자간 정보교류 행위에 대한 다각적 분석 없이 주로 사용자의 프로필 정보를 이용하여 특정 요소만 산정한다. 이로 인해 (1) 시간, 관계의 방향성에 대한 개념을 내포하는 사용자 관계표현에 어려움이 있고, (2) 다양한 시각으로 사용자간 관계의 세기를 측정하는 ‘다차원 상호관계[2]’는 분석하는 방법이 제시되지 않아 추천과 같은 외부 서비스에 분석 결과를 성공적으로 적용함에 있어 그 한계를 보인다.

본 연구에서는 SNS 정보로 부터 사용자 간 상호영향을 다차원 상호관계로 표현 할 수 있는 통합된 프로필 모델과 구축 시스템을 제안한다. 제안 되는 모델은 사용자 간 영향의 정도를 표현할 요소로 친밀도, 관계구조 유사도, 용어 유사도를 산정/합산한 뒤, 해당 결과를 이용하여 ‘강한 영향을 가진 다른 사용자의 목록’을 작성한다. 또한, 제안된 모델을 기반으로 트위터 사용자 정보를 분석하고, 결과를 기술 및 해석한다.

본 논문은 2장에서 관련연구의 동향 및 한계점을, 3장은 데이터 수집, 분석과 유사도, 친밀도 등을 산정/통합 하는 방법을 단계적으로 설명한다. 4장에서는 해당 방법을 검증하기 위해, 트위터로부터 자료를 수집하여, 이를 제안 된 방법으로 수행한 실험과 그 결과를 설명한다. 5장은 이에 대한 결론 및 고찰을 기술한다.

2. 관련연구 분석

2.1 온라인 사용자 간 상호관계 분석

기존 연구에 따르면 실세계에서 통용되는 사람

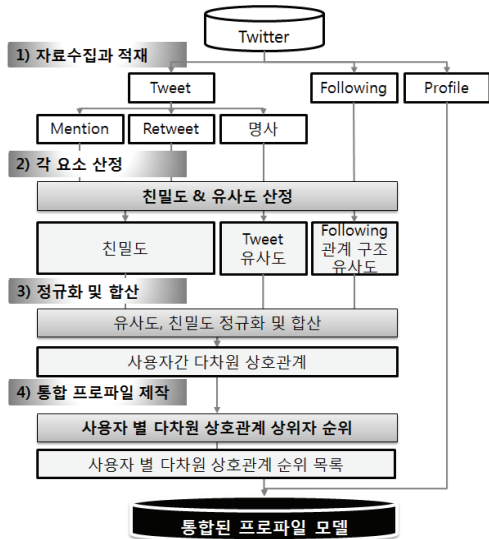
들 간 관계의 세기를 측정하는 요소가, SNS 같은 가상 커뮤니티(Virtual Community)에도 반영될 수 있다고 보고되고 있다[2]. 여기서 사람 간 관계의 세기는 연락 빈도, 친밀도, 관계에 대한 적극성, 충고, 동료가 되고 싶은 욕구, 다양한 주제, 관계 기간, 의존성, 정서적 강도, 신뢰성, 기분 좋은 상호작용과 같이 11가지 지표로 측정이 가능하다[2]. 같은 맥락의 지표라 할 수 있는 사용자 간 유사도는 코사인 유사도(Cosine Similarity), 엔트로피, 역사용자 빈도, 기본 선호도 평가[6] 등의 방법론을 통해 정량화 될 수 있으며, 이와 같은 정보들은 추천 서비스와 같은 외부서비스에 효과적으로 활용될 수 있다.

2.2 트위터 사용자 성향분석

차미영의 연구에서는 트위터 사용자간 메시지, 포스팅, 관계정보로 사용자들의 영향력을 분석하였다[7]. 연구에 따르면 5천만 명의 사용자 정보와 20억 개의 Following 관계, 17억 개의 Tweet 등을 분석한 결과, 많은 Follower의 수와의 Mention/Retweet을 가진 것 연관성 없이 별개의 요소로 영향력을 행사하고 있음을 보인다.

2.3 SNS 정보를 이용한 유사도, 친밀도 산정 방법

SNS로 부터 사용자간 유사도, 친밀도를 산정한 기존 연구들은 주로 사용자들의 프로필, 태그, 북마크 등 사용자가 작성한 정보를 이용하였다. 이와 같은 연구들은 사용자에 의해 작성된 정보를 공유하는 행위가 사용자 간 유사도를 측정하는데 중요한 요소임을 밝혀냈으며[5], SNS 안에서 친밀도에 영향을 주는 혈연, 신뢰도 등의 12개의 요소를 정의한 뒤 1000명을 대상으로 설문조사하여 상관관계를 연구하였다[3]. 또한, 친밀도 측정을 위해 한 사용자를 중심으로 형성된 사용자들의 관계를 분석하여 친밀도를 정의하고 사용자 간 관계를 다양화 하였다[4].



(그림 1) 통합된 프로필 모델 작성 절차

이와 같이, 기존 연구들에서는 SNS로 부터 사용자에 의해 작성된 정보를 이용하여 사용자의 성향을 분석하며, 친밀도 혹은 유사 등 단일 값을 산정하였다. 본 연구에서는 사용자 프로필, 단문포스팅과 같은 트위터 기본정보를 이용하여 친밀도, 용어유사도, 관계구조 유사도를 산정 후 이를 합산하고 다차원 적인 사용자간 상호관계를 분석하여 추천과 같은 외부서비스에서 가용한 형태로 정보를 가공한다.

3. 방 법

본 장에서는 SNS 사용자 간 다차원 상호관계를 표현 할 수 있는 프로필 모델 구축 과정을 설명하며, 다음의 과정을 따른다.

- 1) 자료수집과 적재 : 사용자들의 Tweet, Following 정보, Profile 적재 후, Tweet에서 Mention, Retweet, 명사를 추출.
- 2) 요소 산정 :
 - 친밀도 : Mention과 Retweet을 이용하여 빈도, 시간, 방향을 고려한 Mention점수와 Retweet 점수 산정 후 합산.

(표 1) 트위터 기본정보

저장정보	항 목	설 명
Profile	ScreenName	트위터 내 사용자 이름
	Description	사용자가 설정한 자신에 대한 설명
Tweet	Tweet	사용자가 트위터에 작성한 140내의 단문 포스팅
	작성시간	Tweet 작성 시간
Mention	Sender	메시지 보낸 사용자
	Receiver	메시지 받은 사용자
Reweet	Retweetee	Tweet이 참조된 사용자
	Retweeter	Tweet을 참조한 사용자
Following	Source	다른 사용자의 포스팅을 보도록 설정하는 것
	Target	다른 사용자가 자신의 포스팅을 보도록 설정되는 것

- 관심 유사도 : Tweet으로부터 추출 된 명사를 이용하여 코사인 유사도 산정.
 - 관계 구조 유사도 : Following 관계로 관계 구조 유사도(Structural similarity) 산정.
- 3) 정규화 및 합산 : 산정된 요소들을 각 요소 별 최대값으로 나누어 [0, 1]로 정규화 한 뒤, 합산하여 산정.
 - 4) 통합 프로필 제작 : 사용자별 관계 세기를 사용자의 프로필에 업데이트.

3.1 자료 수집과 저장

트위터가 제공하는 정보중, 사용자에게 널리 사용되는 (표 1)에 명시된 정보의 수집 및 저장을 통해 본 연구에 활용될 기초데이터(Seed Data)를 구축한다. 이를 이용하여 친밀도(Intimacy), 용어/관계구조 유사도(Term/Structural Similarity)를 산출/저장하며 세부 항목은 (표 2)와 같다.

본 연구에서는 명사 기반 용어 사용 유사도 분석을 위해 JAVA로 작성된 형태소 분석기 인 Korean Morpheme Analyzer V0.9.1 by therocks*를 사용하며, 트위터 자료를 수집하기 위해 JAVA로

(표 2) 산출된 저장정보

항 목 명	설 명
source	행위를 하는 사용자
target	행위를 받는 사용자
Term Similarity	두 사용자간 용어 유사도
Structural Similarity	두 사용자간 관계 구조 유사도
Intimacy	두 사용자간 친밀도
Multi-dimensional interaction	두 사용자간 다차원 상호관계

(표 3) 친밀도, 유사도 수식 기호

분 류	기호	의 미
공통	x, y, z	사용자 x, y, z
친밀도	ms()	사용자간 mention 점수
	rs()	사용자간 retweet 점수
	inti()	사용자간 친밀도
	n	사용자간 주고받은 mention, retweet 의 총 개수
	sm	한 사용자가 보낸 mention 수
	ed	마지막으로 Mentnio, Retweet이 일어난 날부터 해당 Mention, Retweet이 일어난 날 수
	td	mention, retweet을 주고받은 총 일 수
용어 유사도	n	두 사용자가 사용한 명사 개수
	첨자 _i	x, y 사용자의 i번째 명사 개수
관계 유사도	ssim()	사용자간 Following 관계구조 유사도

작성된 twitter4j 2.1.7*을 활용하였다.

3.2 친밀도, 유사도(용어, 관계구조) 산정

본 절에서는 트위터로 부터 수집된 정보를 가공하여, 사용자간 친밀도, 용어 사용유사도, 그리고 관계유사도를 도출하는 방법론을 소개한다. (표 3)은 이와 같은 정보를 도출하기 위해 사용된

기호들을 설명한다.

친밀도는 사용자간 주고받은 Mention, Retweet 을 통해 산정되며, Mention 점수는 방향성을 고려할 때 사용자 x의 입장에서 y 에게 보낸 메시지의 비율이 증가할수록 x에서 y로의 Mention점수가 커진다. 또한, 시간을 고려 할 때 최근에 받은 Mention 일수록 큰 값을 갖게 되며 (수식 1)과 같이 산출 된다.

$$ms(x \rightarrow y) = \frac{sm}{n} \times \sum_{i=1}^n \left(\frac{ed_i}{td} \right)$$

(수식 1) mention 점수

Mention 점수의 경우 주고받는 사용자 모두가 Mention행위를 인지 할 수 있는데 비해, Retweet은 Retweet 하는 사용자만 Retweet 행위를 인지하지 때문에 Retweet 점수를 산정할 때는 x라는 사용자가 y라는 사용자의 글을 n번 Retweet 해왔다면 Retweet 개수와 시기만 Retweet 점수에 반영되며 (수식 2)와 같이 산정된다.

$$rs(x \rightarrow y) = \sum_{i=1}^n \left(\frac{ed_i}{td} \right)$$

(수식 2) retweet 점수

위 점수들을 합산하여 친밀도를 산정하며, 이를 수식으로 표현하면 (수식 3)과 같다.

$$inti(x \rightarrow y) = ms(x \rightarrow y) + rs(x \rightarrow y)$$

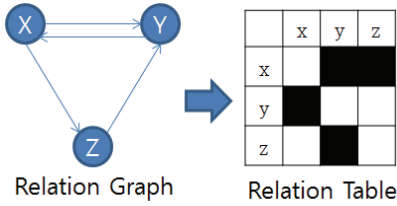
(수식 3) 친밀도 산정

용어 유사도 측정 방법은 사용자가 특정기간 내에 남긴 모든 Tweet에서 명사를 추출하고 그 빈도를 기록한 뒤, 용어와 빈도를 이용하여 사용자별 용어 백터를 구성한다. 두 사용자간 용어 사용 유사도는 해당 두 벡터(x, y)의 Cosine 유사도로 산정하며 이를 식으로 표현하면 (수식 4)와 같다.[6, 8]

* Korean Morpheme Analyzer V0.9.1 by therocks :

<http://therocks.tistory.com/14>

** Twitter4J : <http://twitter4j.org>



(그림 2) 관계구조 표현

$$Cosine(x, y) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{(\sum_{i=1}^n x_i^2) \times (\sum_{i=1}^n y_i^2)}}$$

(수식 4) 단어의 Cosine 유사도

관계구조 유사도 측정은 사용자들의 Following 관계구조를 분석하여 산정한다.

Following 관계는 (그림 2)과 같이 표현 가능하며, 이를 이용하여 사용자 별 Following 벡터 추출 후, XOR 연산결과의 차수를 [0, 1]사이의 값을 갖도록 (n-1)*2로 나누어 관계구조의 비유사도를 도출한다. 끝으로, ‘1 - 관계구조 비유사도’를 통해 관계구조 유사도를 산출한다. 예를 들어, 그림 2에서 벡터 y, z를 표현 하면 $\vec{y}_{yz} = (0, 1, 1, 1)$, $\vec{z}_{yz} = (1, 0, 0, 1)$ 이 되며 (수식 5)와 같이 산정한다.

$$ssim(y, z) = 1 - \frac{|\vec{y}_{yz} \oplus \vec{z}_{yz}|}{2(n-1)}$$

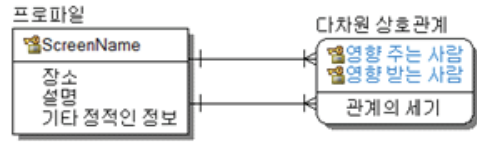
(수식 5) 관계구조유사도

3.3 각 요소 정규화 및 합산

본 절에서는 3.2에서 소개된 지표를 이용하여 사용자 간의 다차원 상호관계 계산방식을 설명한다. 먼저, 사용자별로 계산된 친밀도, 용어 사용 유사도, 관계 구조 유사도를 각각 [0, 1]로 정규화한다.

정규화는 각 요소의 최대값으로 나누어 주었으며 친밀도는 연락 횟수에 따른 변동 값이 작아지도록 자연로그를 취하였다.

정규화 된 3가지 값을 더하여 사용자별로 다차



(그림 3) 통합 프로필 모델

원 상호관계의 세기를 표현하며 이를 수식으로 나타내면 (수식 6)과 같다.

$$multi-dimension(x \rightarrow y) = \frac{\ln(inti(x \rightarrow y))}{\ln(\max(inti))} + \frac{\cos(x, y)}{\max(\cos)} + \frac{ssim(x, y)}{\max(ssim)}$$

(수식 6) 다차원 상호관계

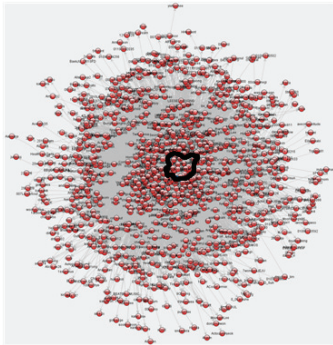
이 후, 이와 같이 산정된 ‘다차원 상호관계’를 추천 서비스와 같은 외부 서비스에 제공하기 위하여 통합 프로필을 구축한다. 구축방법은 사용자 별로 다른 사용자와의 ‘다차원 상호관계’ 세기를 표현 하는 것이며, 통합된 프로필의 모습은 (그림 3)과 같다.

4. 실험 및 실험결과

본 실험에서는 트위터 사용자 중 20대 후반~40대 서울 지역 남녀 17명을 무작위로 정하여 각 사용자의 2단계 Following 관계에 있는 사용자 정보를 수집하였다. 해당 사용자 중 2010년 8월 1일 ~10월 31일(92일) 간 Tweet 개수 상위 1000명을 선정하였다. 외국인 사용자, 광고성 사용자는 제외 했으며, 위 조건을 모두 만족하는 사용자의 Tweet에서 Mention, Retweet을 추출하였다. 이와 같은 방법을 통해 구축된 실험 데이터는 1000명의 Following 관계 23870개, Mention 38238개, Retweet 10342개, Tweet 565541개를 포함한다.

4.1 트위터 기본 정보를 이용한 사용자 성향 분석

Following(Following 받은 수) 상위 5명의 직업군은 방송인(oisso-이외수, keumkangkyung-김제동),

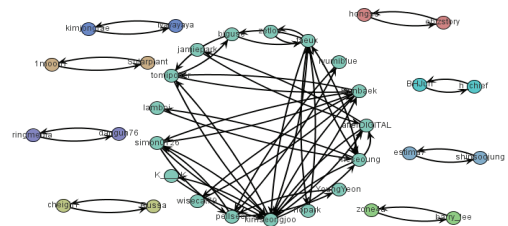


(그림 4) Following 상위 2%

뉴스(dogsul-독설닷컴), IT 종사자(chanjin, kimseongjoo)이었다. 많은 Following을 받았다는 것은 대중의 관심을 받는다는 것을 의미하며[7], 이를 반영하듯 상위 2%에서 대중의 관심을 받는 공인과 뉴스를 합치면 전체의 40%로 가장 큰 비중을 차지한다. 연령은 30~40대가, 성별은 남성이 70%를 넘어 가장 큰 비중을 차지하는 것을 보인다. 또한, ‘사용자간 Following 관계 그래프’ 내에서 상위 2%의 관계가 24%(5810개)를 차지하여 (그림 4)와 같이 그래프 중앙에 위치하여 높은 Centrality를 보인다.

받은 Mention은 보낸 Mention에 대한 답장일 수 있으므로[7], 사용자간 주고받은 전체 Mention 중 받은 Mention의 비율이 높은 사용자를 추출 하였다. 받은 Mention비율 상위 5명은 IT종사자(hmpyo), 학생(metrosexual), 방송인(moviejhp-박중훈, doc0102-김창렬, kolleen22-박칼린)으로 학생인 metrosexual을 제외하면 주로 방송인과 IT 종사자였다. Mention 상위 2% 분석 결과, 방송업 부터 의료, 연구원, 기타(주부) 까지 다양한 직업군을 포함하며, 주 연령대는 30~40대가 성별은 남성이 각각 전체의 약 65%의 비중을 차지했다.

Retweet이 많이 된 상위 5명의 사용자는 대학교수(hiconcep), 뉴스(dogsul-독설닷컴, wiktiree-위키투리, estima7), IT 종사자(taeuk)였다. 상위 2%의 직업군은 IT종사자부터 종교인, 군인까지 다양하게 나타난다.



(그림 5) 친밀도 사용자 관계도($t > 3.0$)

뉴스, CEO, 교수의 비율이 높은 것을 통해 사용자들은 신뢰할 수 있는 사용자의 Tweet 으로부터 Retweet 한다는 경향을 알 수 있다[7]. 연령대는 30~40대, 성별은 남성이 70%를 차지하며 상위에 위치한다.

트위터의 기본 정보인 Following 관계, Mention, Retweet으로 사용자 분석결과 공통적으로 가장 큰 활동량을 보인 사용자의 직업은 방송인, IT종사자였으며 각 요소별 상위 2%를 비교한 결과 모든 요소에서 상위에 나타난 사용자는 2명(dogsul-뉴스, kimjuha-방송인)이었다. Retweet이 많이 된 사용자와 Follower를 많이 가진 사용자는 9명으로 50%에 가까운 사용자가 일치하며 높은 관계성을 보여주었으나, Mention 정보와 다른 두 가지 지표에서는 3명만 겹치며 전체적으로 서로 적은 관계성을 보여 주었다.

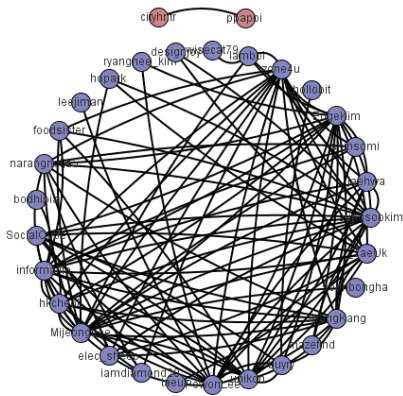
4.2 친밀도, 관계구조 유사도, 용어 유사도 분석

본 절에서는 실험 데이터로 부터 친밀도, 관계구조 유사도, 용어 유사도를 산정하여 사용자간 관계를 분석한 결과를 제시한다.

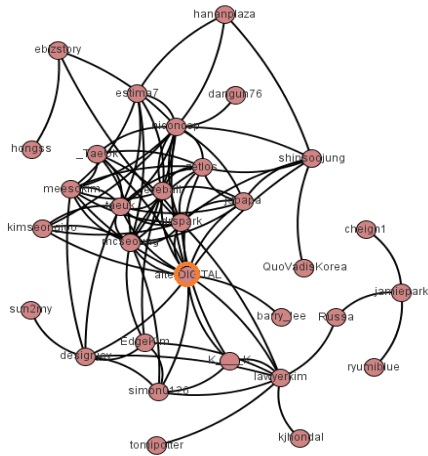
4.2.1 친밀도 실험

친밀도 상위 33명의 친밀도 관계 58개를 (그림 6)과 같이 나타내었다. 그림 5에서 나타나는 바와 같이 33명이 9개의 그룹으로 나뉘지며 2명으로 이루어진 그룹 8개와 17명으로 이루어진 그룹 1개로 나타난다.

(그림 6)에서, 2명으로 이루어진 8개 그룹에서



(그림 7) 관계구조 유사도 ($t > 0.93$)



(그림 8) 용어 유사도($t > 0.78$)

‘zone4u’와 ‘barry_lee’로 이루어진 그룹만 30대/남성/IT 종사자로 비슷한 성향을 지닌 것으로 나타난다. 17명으로 이루어진 그룹에서는 ‘kimseongjoo’가 7명과 친밀한 관계를 가지며 ‘afterDIGITAL’, ‘mcseoung’, ‘taeuk’, ‘Hanback’, ‘simon0126’처럼 비슷한 직업을 사용자인 경우도 있으나 예외가 많고 전반적인 성향을 표현하기에 다소 무리가 있다.

4.2.2 관계구조 유사도 실험

관계 구조 유사도는 Following 관계를 통해 산출된 Closeness Centrality가 높은 100명의 사용자들에 대한 요소를 분석하였다. 해당 사용자 100명은 실제 트위터 상에서도 Following수 249~99825(평균 7073.91명), Follower수 255~96383(10062.67명)으로 비교적 많은 관계를 맺고 있는 사용자들 이었다. 해당 사용자 중 관계구조 유사도를 기준으로 0.93 이상의 값을 가진 상위 31명 간 115개의 관계를 표현하면 (그림 6)과 같다. 31명이 2 개의 그룹으로 나누어지며 2명으로 이루어진 그룹이 1개, 29명으로 이루어진 그룹이 1개로 나타난다.

(그림 7)에서, 2명으로 된 그룹의 ‘cityhnr’와 ‘ppappi’는 모두 IT 종사자였으며, 29명으로 구성된 그룹에서는 ‘MijeongLee’ 라는 사용자가 19명의 가장 많은 사용자와 비슷한 관계구조를 가진

것으로 나타났으며, 해당 사용자와 관계구조 유사성이 높은 상위 5명과 비슷한 직업군임을 보여주었다.

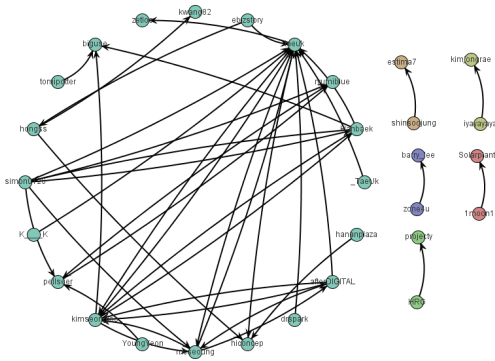
4.2.3 용어 사용 유사도 실험

용어 사용 유사도를 기준으로 서로 0.78 이상의 값을 가진 상위 31명 간 91개의 관계를 나타내면 (그림 7)과 같이 모두 연결되어 있으며, 이 중 ‘afterDIGITAL’ 이라는 사용자가 16명으로, 가장 많은 사용자와 비슷한 용어 유사도를 보여준다. 상위 5명과 비교해 볼 경우 ‘afterDIGITAL’은 IT 칼럼리스트, 교수, IT종사자, 변호사 등의 다양한 직업을 가진 사용자와 높은 용어 유사도를 보여주었다.

4.3 다차원 상호관계 분석

본 논문에서 제안한 다차원 상호관계를 기준으로 2.1 이상의 값을 가진 상위 30명 간 41개의 관계를 표현하면 (그림 8)과 같이 6개의 그룹이 나타나며, 2명인 그룹 5개, 20명으로 이루어진 그룹 1개로 나뉜다.

2명으로 이루어진 그룹들을 살펴보면 각각 ‘estima7’은 ‘shinsoojung’에게 ‘kimjongrae’는 ‘iyayaya’



(그림 9) 다차원 상호관계도(t)2.1)

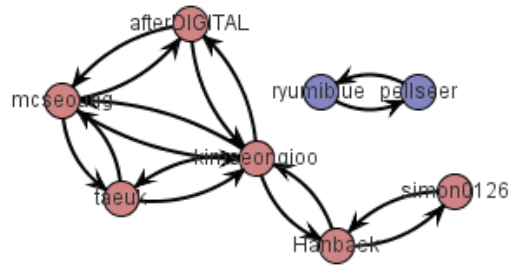
에게 ‘barry_lee’는 ‘zone4u’에게 ‘Solarplant’는 ‘Imoon1’에게 그리고 ‘projecty’는 ‘HRG’에게 영향을 주는 사용자로 나타났다. ‘estima7’, ‘shinsoojung’, ‘barry_lee’, ‘zone4u’는 IT업에 종사하며 ‘kimjongrae’, ‘iyayaya’는 마케팅 분야, ‘Solarplant’, ‘Imoon1’은 같은 기업의 CEO와 직원으로 나타났다. 또한 ‘projecty’와 ‘HRG’는 IT 칼럼리스트와 UX개발자로 비슷한 분야의 직업을 가졌으며 ‘Microsoft’사와 연관된 사용자들로 프로필상 비슷한 점들이 발견되었다.

20명으로 이루어진 그룹에서는 ‘taeuk’이라는 사용자가 가장 많은 사용자와 높은 영향력을 보이며, 해당 사용자는 IT 업계 연구소장으로 본인의 다른 계정인 ‘_TaeUK’과 타 IT 연구소장을 포함한 6명의 IT 관련업종 종사자와 1명의 종교인(Hanbaek)에게 영향을 주며, 의료 IT, 의사(교수) 종사자 인 2명의 사용자에게 영향을 받는 것으로 나타났다.

방향성을 고려하여 서로 영향을 2.1 이상 주고 받는 사용자들을 분석한 결과 (그림 9)처럼 나타나며, 6명으로 이루어진 그룹에서 ‘Hanbaek’을 제외한 전부가 IT 종사자이며 ‘ryumibluе’, ‘pellseer’는 각각 마케팅 관련 종사자와 학생으로 나타났다.

5. 결론 및 고찰

트위터에서 얻을 수 있는 기본적인 정보를, 본



(그림 12) 방향성 다차원 상호관계(t)2.1)

연구에서 제안한 방법을 통해 분석한 결과 IT종사자의 활동이 두드러졌으며 이는 아직 트위터가 서비스의 특성상 IT종사자들 중심으로 사용자가 확장되고 있다고 해석할 수 있다. 또한, Retweet과 Following 받는 수가 상위 20명인 사용자가 50% 께치는 결과를 나타내 차미영 교수의 연구결과[7]와 상이하였으며 이는 무작위 1,000명에 대한 데이터만 분석한 결과로서, 모집단의 성격의 차이에 기인한 것이라 추측할 수 있다.

대체로 30~40대 남자 사용자이기 때문에 그룹을 특징지을 요소로 성별, 연령을 사용하기에는 어려움이 있어 직업과의 연관성을 찾으려고 시도 하였지만 사용자의 직업군에서 비슷한 성향을 보여주기에는 어려움이 있었다. 이는 트위터를 사용할 때 사회적 이유보다는 개인적으로 사용하는 경우가 많아 용어나 관계구조에서 직업적 특성이 드러나지 않았기 때문으로 보인다.

따라서, SNS 상에서 기본적으로 얻을 수 있는 (표 1)과 같은 정보나 친밀도, 관계구조, 용어 사용 유사도 중 한 가지 값으로는 나타나지 않던 같은 직업군 같은 사용자간 관계가 다차원 적인 분석으로 좀 더 드러나는 것을 보여주었다. 또한, 이렇게 산출된 사용자간 영향관계는 사용자의 프로필에 저장되어 해당 사용자의 정보 부재 시 보조적인 수단으로 사용될 수 있을 것이며 추천과 같은 외부서비스에서 정보 부족으로 생기는 문제를 일부 해결 할 수 있을 것으로 기대된다.

Acknowledgement

본 연구는 지식경제부/방송통신위원회/한국산업기술평가원의 산업원천기술개발사업, “차세대 IPTV 인프라 기반 Social TV 서비스 인에이블러 개발” 연구 과제의 지원을 받아 수행되었습니다.

This work was supported by IT R&D Program of MKE/KEIT/KCC, [Development of Social TV Service Enabler based on Next Generation IPTV Infrastructure]

참 고 문 헌

- [1] Danah M. Boyd/Nicole B. Ellison, 'Social Network Sites: Definition, History, and Scholarship', Michigan State University, 2007
- [2] Tamas Nepusz and Fulop Bazso, 'Measuring tie-strength in virtual social networks', <http://www.insna.org>, 2007
- [3] Yuan-yuan Yang, Su-fen Dong, Gui-fa Teng, Wei Yao, 'Research of Nodes Intimacy in SNS Networks', <http://ieeexplore.ieee.org>, 2010
- [4] 오정운, 윤성웅, 이상훈, Social Networks 사용자간의 친밀도 산정, '08 한국컴퓨터종합학술대회 논문집 Vol35, No 1(C), 2008
- [5] Ido Guy, Michal Jacovi, Adam Perer, Inbal Ronen, Erel Uziel, 'Same Places, Same Things, Same People? - Mining User Similarity on Social Media', <http://www.acm.org>, 2010
- [6] 정경용, 이정현, 개인화 추천 시스템의 예측 정확도 향상을 위한 사용자 유사도 가중치에 대한 비교 평가, '05년 11월 전자공학회 논문지 제42권 CI편 제6호, 2005
- [7] Meeyoung Cha, Hamed Haddadiy, Fabrício Benevenutoz, Krishna P. Gummadi, 'Measuring User Influence in Twitter: The Million Follower Fallacy', International AAAI Conference on Weblogs and Social Media (ICWSM), May 2010
- [8] Senator Jeong, Soo Kyoung Lee, Hong-Gee Kim, 'Knowledge Structure of Korean Medical Informatics: A Social Network Analysis of Articles in Journal and Proceedings', Health Inform Res. 16(1):52-59, March 2010
- [9] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, Bobby Bhattacharjee, 'Measurement and Analysis of Online Social Networks', the 5th ACM/USENIX Internet Measurement Conference (IMC'07), 2007
- [10] 김민수, 유홍진, 채진석, 사용자의 인적정보를 이용한 소셜 네트워크 활성화 방안, '09 한국컴퓨터종합학술대회 논문집 Vol.36, No.1 (D), 2009
- [11] Christine Hine, 'VIRTUAL ETHNOGRAPHY' <http://socantcafe.org/uploads//2009/10/hine-2000-virtual-ethno.pdf>, 2000
- [12] Nicole B. Ellison, Charles Steinfield, Cliff Lampe, 'The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites' Journal of Computer-Mediated Communication, 23 AUG 2007
- [13] Michael J. Brzozowski, Tad Hogg, Gabor Szabo, Friends and Foes: Ideological social networking, <http://www.acm.org>, 2008

● 저 자 소 개 ●

이 경 민



2010년 명지대학교 컴퓨터소프트웨어 학과 졸업(학사)
2010년~현재 서울대학교 협동과정 의료정보학과 석사과정
관심분야 : Biomedical text mining, Data mining etc.
E-mail : sephiroce@snu.ac.kr

남 공 현



2004년 전북대학교 컴퓨터공학과 졸업(학사)
2006년 과학기술연합대학원 대학교 컴퓨터소프트웨어 및 공학과 졸업(석사)
2006년~현재 서울대학교 치의학과 박사과정
관심분야 : 시맨틱 웹, 온톨로지, 대용량 데이터 처리
E-mail : ngh@snu.ac.kr

김 응 희



2007년 선문대학교 컴퓨터정보학과 졸업(이학학사)
2009년 서울대학교 대학원 치의과학과 졸업(공학석사)
2011년~현재 서울대학교 협동과정 의료정보학과 박사과정
관심분야 : Formal concept analysis, association rule mining, data mining
E-mail : eungheekim@snu.ac.kr

이 강 용



2004년 아주대학교 정보 및 컴퓨터 공학과 졸업(학사)
2006년 University of Science & Technology 광대역네트워크공학과 졸업 (석사)
2006년~현재 한국전자통신연구원 방송통신융합연구부 연구원
2010년~현재 KAIST 산업 및 시스템공학과 박사과정
관심분야 : 네트워크 최적화, 통신망 성능 분석, 시뮬레이션
E-mail : kanglee@etri.re.kr

김 흥 기



1985년 고려대학교 심리학과 졸업(학사)
1993년 조지아대학교 대학원 전산학(인공지능)학과 졸업(석사)
1996년 조지아대학교 대학원 철학(AI, 인지과학)학과 졸업(박사)
2005년~현재 서울대학교 의료정보학과 부교수
관심분야 : Medical Informatics, Semantic Web, Ontology Engineering, Knowledge Representation.
E-mail : hgkim@snu.ac.kr