Journal of The Korea Society of Computer and Information
Vol. 19, No. 1, January 2014

www.ksci.re.kr
http://dx.doi.org/10.9708/jksci.2014.19.1.111

트위터 사용자정보의 유사성을 기반으로 한 팔로어 분류시스템

계 용 선*, 윤 영 미**

Follower classification system based on the similarity of Twitter node information

Yong-Sun Kye*, Youngmi Yoon**

요 약

현재 트위터에서 제공되는 친구추천 시스템은 영향력이 높은 사용자를 우선적으로 추천해준다. 하지만 사용자정보의 유사성이 높은 다른 사용자는 추천되지 않는 단점을 가지고 있다. 사용자들은 정보의 유사성이 높은 사용자 추천을 원하기 때문에 이러한 단점을 극복하기 위하여 본 논문에서는 사용자정보의 유사성을 기반으로 팔로어 추천 시스템을 구현하였다. 본 논문에서 사용된 데이터는 SNAP(Stanford Network Analysis Platform)에서 제공하는 데이터로, 팔로어의 수가 10,000명이상인 트위터의 사용자정보와 노드간 연결 데이터로 구성된다. 이 데이터를 트레이닝 데이터로 활용하여 팔로어간의 관계를 분류해줄 수 있는 분류자를 생성하고, 10-Fold Cross Validation을 활용하여, 분류자의 정확도를 판단한다. 두 트위터의 정보가 주어지면 그 들 사이에 친구 관계, 팔로우 관계, 비연결 관계를 추천한다.

▸ Keywords : 소셜 데이터마이닝, 트위터

Abstract

Current friend recommendation system on Twitter primarily recommends the most influential twitter. However, this way of recommendation has drawbacks where it does not recommend the users of which attributes of interests are similar to theirs. Since users want other users of which attributes are similar, this study implements follower recommendation system based on the similarity of twitter node informations. The data in this study is from SNAP(Stanford Network Analysis Platform), and it consists of twitter node information of which number of followers is over 10,000 and twitter link information. We used the SNAP data as a training data, and generated a

•제1저자 : 계용선 •교신저자 : 윤영미
•투고일 : 2013. 11. 30, 심사일 : 2013. 12. 06, 게재확정일 : 2014. 01. 14
* 가천대학교 컴퓨터공학과(Dept. of Computer Engineering, Gachon University)
** 가천대학교 컴퓨터공학과(Dept. of Computer Engineering, Gachon University)
•이 논문은 2013년도 가천대학교 교내연구비 지원에 의한 결과임.(GCU-2013-R364)

classifier which recommends and predicts the relation between followers. We evaluated the classifier by 10-Fold Cross validation. Once two twitter node informations are given, our model can recommend the relationship of the two twitters as one of following such as: FoFo(Follower Follower), FoFr(Follower Friend), NC(Not Connected).

▸ Keywords : Social Media Data, Twitter

I. 서 론

소셜 네트워크 서비스(SNS, Social Networking Service)는 사용자들 간 공통의 관심사나 활동 등을 바탕으로 서로 간의 사회적 관계를 형성하고 이를 반영한 온라인 기반의 서비스를 제공한다[1]. 이러한 소셜 네트워크를 이용한 정보의 확산은 전 세계적으로 매우 빠른 속도로 인기를 얻었고, 대표적인 SNS에는 트위터(Twitter), 페이스북(Facebook), 마이스페이스(Myspace), 싸이월드(Cyworld)등이 있으며, 그중 트위터의 성장은 2006년에 설립하여 1억 명이 넘는 사용자를 확보하는 등 전 세계적으로 SNS 열풍을 주도하고 있다[2].

SNS를 사용하는데 있어 중요한 요소 중 하나는 관계를 맺는 것이다. 관계를 맺음으로써 사람들과 정보의 교환뿐만 아니라 일상적인 담화를 나누며 SNS의 목표인 지인과의 관계를 강화하고자 한다[3]. 트위터에서는 사용자 추천 기능을 지원하지만 영향력이 높은 사람을 추천해주는 경우가 많다. 영향력이란 트위터 내에서의 파급효과가 큰 사용자를 의미하며, 사회적으로 큰 이슈를 일으킬 수 있는 유명인인 경우가 많다[4]. 하지만 사용자들은 주변 사람들이나 정보의 유사성이 높은 사람들과 관계를 원한다[5]. 이러한 문제점을 발견하여 본 논문에서는 효율적인 추천방법을 제시하고자 한다.

본 논문에서는 대표적인 SNS인 트위터에서 보다 효율적으로 관계를 맺고자 사용자정보와 연결 데이터를 활용한 관계 예측을 수행한다. 예측방법으로는 사용자정보의 유사성을 기반으로 한 관계 예측 및 추천을 제시할 수 있는 시스템을 구현하였다. 시스템 구현을 위하여 트위터 사용자정보와 연결 데이터를 활용하였다. 본 논문에서는 관계정의를 그림 1처럼 친구 관계(FoFo, Follower Follower), 팔로우 관계(FoFr, Follower Friend), 비연결 관계(NC, Not Connected) 3가지로 정의한다. 친구 관계는 노드간 상호 팔로우 관계를 맺고 있는 관계이며, 팔로우 관계는 한 노드만 팔로우 관계를 맺고 있는 관계이다. 그리고 비연결 관계는 어떠한 관계를 맺고 있지 않은 노드 상태를 의미한다. 그러므로 트위터에서의 노드정보를 활용하여 사용자에게 맞는 관계 예측이나 추천을 제시해 줄 수 있다.

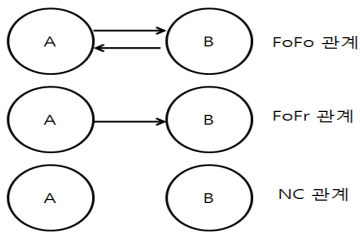

그림 1. 관계 정의
Fig. 1. Definition of Relation

본 논문의 구성은 다음과 같다. 1장 서론에 이어 2장에서는 관련연구로서 소셜네트워크 서비스, 페이스북, 트위터에 대하여 기술하고, 3장에서는 노드정보를 활용한 관계 예측 기법을 제시한다. 4장에서는 실험 결과 및 분석을 보여주고, 5장에서 논문에 대한 결론을 맺는다.

II. 관련 연구

2.1 소셜네트워크 서비스

소셜 네트워크 서비스는 인터넷상에서 공통의 관심사를 지니고 있는 사용자들 간의 관계형성을 지원하고, 이렇게 형성된 지인 관계를 바탕으로 인맥 관리, 정보 및 콘텐츠 공유 등 다양한 활동을 할 수 있도록 지원하는 서비스를 의미한다[6]. 즉, 사람들이 다양한 상호 관계를 구성할 수 있도록 지원하는 인터넷 기반의 서비스이다[7].

대표적으로 트위터, 페이스북 등이 있는데, 뉴스의 전달, 개인적인 사소한 일상의 공유, 정치적 견해의 표현, 마케팅, 기업 조직 간의 비공식 커뮤니케이션 채널, 현재 발생하는 사

건 실황의 추적 등 다양한 용도로 매우 활발하게 사용되고 있다[8].

2.2 페이스북

페이스북은 6억 명 이상의 사용자를 확보한 세계 최대의 소셜네트워크 서비스이다[9]. 페이스북은 사용자들이 친구(Friends) 목록을 공유함으로서 더욱 쉽게 친구를 추천받을 수 있으며, 담벼락(WALL) 기능을 통하여 실시간으로 정보를 공유한다[10].

페이스북은 개인이 사용하는 프로필, 기업이나 단체가 사용하는 페이지, 비공개가 가능한 그룹, 어플리케이션 등으로 그 기능이 크게 나누어진다. 특히 페이스북은 2007년 5월에 'F8 플랫폼'이라는 API를 개발해 세상에 공개함으로써 소프트웨어 개발업체가 페이스북과 연동되는 어플리케이션을 개발할 수 있도록 보장하는 오픈플랫폼 정책을 도입했다. 이러한 개방 정책이 결국 페이스북 급성장의 원동력이 되었다[11].

2.3 트위터

트위터는 140byte 한도 내의 단문으로 된 메시지를 사람들과 주고받을 수 있는 SNS이다. 트위터는 기존의 블로그와 SMS(Short Message Service), 메신저, 커뮤니티의 장점을 흡수한 변형된 SNS로서, 무수한 사람들이 창출하고 쏟아내는 이야기나 정보 중에 자신과 관계가 있고 듣고 싶은 이야기만을 걸려서 커뮤니케이션을 할 수 있는 수용자의 선택성을 강화시킨 미디어라 볼 수 있다[12].

트위터의 의사소통 체계는 관심 있는 사람이나 단체 등의 트위터 계정을 임의로 팔로우(follow) 할 수 있고, 이후에는 해당 트위터에서 업데이트 되는 트윗(tweet)이 자신의 타임라인(timeline)에 올라오게 된다. 자신의 팔로어(follower)와 공유하고 싶은 트윗이 올라올 경우 리트윗(retweet)을 통해 이를 일괄적으로 전파할 수 있다[13].

현재 트위터에서는 친구 추천 기능을 제공한다. 트위터에서 제공하는 친구 추천 알고리즘에는 실제 사회적 관계를 가지고 있는 친구를 추천해주는 것이 아니라 트위터에서의 영향력이 높은 사람을 우선적으로 추천해준다. 영향력은 팔로어가 많은 사람을 뜻하며 영향력이 높은 사람은 현실에서 연예인이거나 유명인이 대부분 포함되어있다. 이를 통한 트위터의 장점은 지구 반대편에 있는 유명인과 친구를 맺을 수 있어서 다른 소셜 네트워크와 가장 큰 차이점을 가지고 있다[14].

소셜네트워크 서비스의 성장세는 단순한 유행에 그치는 것이 아니라 사용자들의 온라인 생활의 일부로 정착되고 있으며, 온라인 콘텐츠의 생산, 소비, 유통 방식을 크게 변화시키면서 온라인 서비스 시장에서의 영향력을 확대해 나아가고 있다. 이처럼 소셜네트워크 서비스가 확장됨에 따라 본 논문에서는 트위터의 관계에 대하여 사용자정보의 유사성을 기반으로 한 관계를 예측 및 추천을 제시해 줄 수 있다.

III. 본 론

3.1 시스템 구성도

본 논문에서 제안하는, 사용자정보의 유사성을 중심으로 한 트위터 팔로어 관계 분류 시스템은 그림 2와 같이 트위터 사용자정보(Twitter_User_information)와 연결 데이터(Twitter_Link)를 이용하여 관계상태를 예측한다. 트위터 사용자정보는 표 1의 속성으로 이루어진 1인 트위터의 정보이다. 연결 데이터 정보는 2인 트위터간의 기존 연결 여부를 제공하는 관계정보이다.

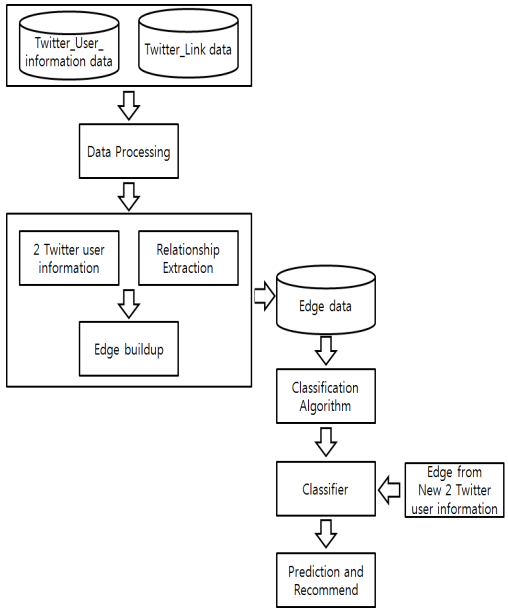

그림 2. 시스템 구성도
Fig. 2. System Diagram

사용자정보의 유사성을 기반으로 한, 데이터 가공 과정은 사용자정보 데이터에서 2명의 사용자정보와, 연결 데이터에서 관계를 추출하여 엣지 데이터를 생성한다. 1개의 엣지 데이터는 2개의 트위터 노드 정보와 관계에 대한 상태 (FoFo,

FoFr, NC)로 이루어져 있다.

엣지 데이터로 이루어진 트레이닝 데이터를 이용하여, 분류 분석 알고리즘을 통하여, 분류자를 생성한다. 관계상태가 알려지지 않은, 새로운 2 트위터의 속성정보만으로 테스트 데이터를 구성한다. 생성된 분류자를 이용하여 테스트데이터의 엣지의 클래스, 즉 관계 상태를 예측 및 추천을 하는 시스템이다.

3.2 데이터 가공 과정

본 논문에서 사용된 데이터는 스탠포드 대학교의 SNAP(Stanford Network Analysis Platform)에서 제공되는 Twitter 데이터이다[15]. 본 데이터는 만 명 이상의 팔로우가 있는 사용자로 총 1,500명으로 구성된 노드 데이터와 사용자간의 연결을 나타내는 관계데이터가 제공된다. 제공되는 데이터는 기본적으로 2009년 6월 1일부터 2009년 12월 31일까지의 기간 동안 수집된 데이터이다.

표 1. 노드 정보
Table 1. Node Information

번호	속성 명	설명
1	profile_bar_color	프로필에서 제공되는 바에 대한 색상
2	text_color	글자에 대한 색상
3	followers_count	팔로어의 수
4	location	트윗(글)을 올린 위치
5	background_color	배경 색상
6	statuses_count	사용자가 트위터에 트윗(글)을 올린 수
7	friends_count	친구의 수
8	favourites_count	다른 사용자의 트윗(글)에 대한 관심 트윗(글)의 수
9	time_zone	표준 시간대
10	gender	성별
11	Class	엣지의 관계에 대한 관계

3.2.1 사용자정보의 유사성을 기반으로 한 데이터 가공 과정

사용자정보의 유사성을 기반으로 한 데이터 가공을 위하여 엣지 데이터를 생성하였다. 하나의 엣지는 2개의 노드 정보와 관계에 대한 상태를 가지고 있다. 한 노드의 정보는 표 1에 표시되어있다. 표 2는 관계에 따라 추출한 정보와 사용자정보와의 결합된 벡터데이터가 어떻게 표현되는지를 나타낸다. 본 실험에서 고려한 엣지의 총 개수는 21,585개이다.

3.3 분류 분석 및 관계 예측

표 2의 엣지 데이터를 Weka에서 제공 하는, 카테고리별 대표적인 알고리즘 Bayes Net, J48, Decision Table, Logistic을 이용하여 분류분석을 수행하였다[16, 17]. 트레이닝 데이터는 아래의 표2와 같이 두 트위터 노드의 속성정보와, 클래스 정보로 이루어져 있다. 클래스 정보는 그 두 트위터의 기존 관계 상태를 의미한다. 테스트 데이터의 클래스 판정을 위하여, 전체 트레이닝 데이터셋에서, 매번 트레이닝 데이터의 10%를 테스트데이터로 활용하는 10-Fold Cross Validation 실험을 수행하였다[18]. 성능평가에 대한 측도와 결과는 4.2장에 기술하였다.

표2. 트레이닝 데이터
Table 2. Training Data

edge	A_profile_bar_color	A_text_color	A_followers_count	A_location	A_background_color	A_statuses_count	A_friends_count	A_favourites_count	A_timezone	A_gender	B_profile_bar_color	B_text_color	B_followers_count	B_location	B_background_color	B_statuses_count	B_friends_count	B_favourites_count	B_timezone	B_gender	Class
e1	pink	black	895829	San Francisco	brown	4209	574	614	Pacific Time	m	gray	green	14062	San Francisco	black	7060	255	449	Tijuana	f	FoFo
e2	orange	black	17118	Ross	blue	2081	8636	4	Pacific Time	f	gray	brown	14454	Boston	white	3635	10592	39	Central Time	m	FoFr
e3	yellow green	black	23128	New York	sky blue	2993	8366	349	Eastern Time	f	violet	black	10523	Btazil	white	1827	510	1	Pacific Time	m	NC
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

새로운 2 트위터의 속성 정보가 주어진다면, 분류자를 이용하여 2 트위터 간의 관계, 즉 클래스를 예측한다. 클래스의 판정이 FoFo(Follower Follower) 라면, 노드 간 상호 팔로우관계를 맺기를 추천한다. 클래스의 판정이 NC(Not Connected) 라면 노드간 팔로우관계를 맺지 않기를 추천한다. FoFo의 관계를 추천 받았다면, 새로운 2 트위터의 사용자정보의 유사성이 높기 때문에 상호 팔로어 관계를 맺는 것이 바람직함을 의미한다. 관계의 예측을 통하여 사용자는 서로 속성정보가 유사한 다른 트위터와 팔로우 관계를 고려 할 수 있다.

IV. 결과

4.1 실험 환경

실험은 Windows 7 Home Premium K(x86) 운영체제와 4.00G RAM, Intel(R) Core(TM) i5 CPU M460(2.53GHz) 사양의 PC에서 수행했다. 실험에서 사용된 분류 알고리즘은 Weka V3.6.5의 구현을 이용했다.

4.2 성능 평가 방법

일반적인 이분 분류분석(Binary Class Classification)이 아니라, 표 3와 같이 FoFo, FoFr, Nc의 세 클래스를 고려하는 복수클래스 분류분석(Multi-Class Classification)을 수행하였다[19].

표 3. 추론클래스와 실질클래스의 엣지의 개수 (3 클래스 기준)
Table 3. Number of Edges in a Predicted Class and a Real Class (Based on 3 Classes)

		실질 분류		
	Class	*FoFo*	*FoFr*	*NC*
추론된 분류	*FoFo*	$num_{FoFoFoFo}$	$num_{FoFoFoFr}$	num_{FoFoNC}
	FoFr	$num_{FoFrFoFo}$	$num_{FoFrFoFr}$	num_{FoFrNC}
	NC	num_{NCFoFo}	num_{NCFoFr}	num_{NCNC}
실질클래스별 엣지의 개수		num_{FoFo}	num_{FoFr}	num_{NC}

각 엣지는 표 2에서 제공되는 엣지 데이터를 의미한다. 아래첨자의 전자는 추론된 클래스를 의미하고 후자는 실질 클래스를 의미한다. 즉 num_{FoFoNC}는 추론된 클래스는 *FoFo*이지만, 실질클래스는 *NC*인 엣지 데이터의 개수를 의미한다. 또한 $num_{totaldass}$는 전체 엣지 데이터 개수를 의미하며, num_{FOFO}는 실질 클래스 *FoFo*에 속한 엣지의 개수를 의미한다.

실험의 성능을 평가하는 방법으로는 Sensitivity, Precision, Accuracy, F-measure을 이용하였다[20]. Sensitivity, Specificity는 분류가 얼마나 정확하게 나뉘었는지 나타내는 방법으로 식(1), (2)와 같이 표현 된다.

$$Sensitivity = \frac{\sum_{i\in\{FoFo,FoFr,NC\}} Sensitivity_i * num_i}{num_{totaldass}} \quad (1)$$

$$Specificity = \frac{\sum_{i\in\{FoFo,FoFr,NC\}} Specificity_i * num_i}{num_{totaldass}} \quad (2)$$

Precision은 정확률로서 분류기가 악성이라고 판별한 자질들 중에서 악성자질로 정확히 판별한 자질의 비율을 나타낸다. Precision은 식(3)로 표현된다.

$$Precision = \frac{\sum_{i\in\{FoFo,FoFr,NC\}} Precision_i * num_i}{num_{totaldass}} \quad (3)$$

Accuracy는 분류자의 정확성을 나타내는 정확도로서 식(4)과 같이 표현된다.

$$Accuracy = \frac{num_{FoFoFoFo} + num_{FoFrFoFr} + num_{NCNC}}{num_{totaldass}} \quad (4)$$

F-measure는 테스트 데이터의 자질이 판별된 클래스에 대한 신뢰성을 나타내며 Recall(=Sensitivity)과 Precision에 동등한 중요도를 부여하여 Recall과 Precision의 합으로 그 값의 두 배에 해당하는 값을 나누어 계산하는 평가방법 중 하나이다. F-measure는 식(5)로 표현된다.

$$F\_measure = \frac{\sum_{i\in\{FoFo,FoFr,NC\}} F\_measure_i * num_i}{num_{totaldass}} \quad (5)$$

각각의 엣지를 기준으로 Sensitivity, Precision, F-measure 다음과 같이 계산된다. 아래 수식은 num_{FoFo}를 기준으로 한 계산식이며 $\overline{FoFo}$는 num_{FoFo}를 제외한 모든 클래스를 의미한다. 식 (6), (7), (8)은 num_{FoFo}를 기준으로 계산한 Sensitivity, Precision, F-measure에 대한 식이다.

$$Sensitivity_{FoFo} = \frac{num_{FoFoFoFo}}{num_{FoFoFoFo} + num_{\overline{FoFo}FoFo}} \quad (6)$$

$$Precision_{FoFo} = \frac{num_{FoFoFoFo}}{num_{FoFoFoFo} num_{FoFo\overline{FoFo}}} \quad (7)$$

$$F\_measure_{FoFo} = 2 \cdot \frac{Precision_{FoFo} \cdot recall_{FoFo}}{Precision_{FoFo} + recall_{FoFo}} \quad (8)$$

4.3 트위터 데이터의 분류 및 예측 분석

관계에 따라 추출한 벡터 데이터를 트레이닝데이터로, 복수 클래스 분류 분석을 수행하여 트위터의 노드정보에 따른 관계가 어떻게 이루어지는지를 예측하는 분류자를 생성하였다.

아래 표 4는 각각의 분류 분석 알고리즘에 대한 분석 결과이다. 4가지의 알고리즘을 통한 분석 결과는 Sensitivity, Precision, Accuracy, F-measure, AUC(Aread Under the Curve)을 통하여 보여 진다. 또한 그림 3은 4개의 알고리즘 중 정확도가 가장 높은 Decision Table의 클래스별 ROC curve(Receiver Operating Characteristic curve)를 나타낸다.

표 4. 분석 결과
Table 4. Classification Results

Algorithm	sensitivity	Precision	accuracy	F-measure	AUC
Bayes Net	0.716	0.735	71.6284	0.725	0.822
J48	0.761	0.713	76.0852	0.71	0.767
Decision Table	0.762	0.717	76.164	0.724	0.841
Logistic	0.746	0.701	74.5796	0.713	0.82

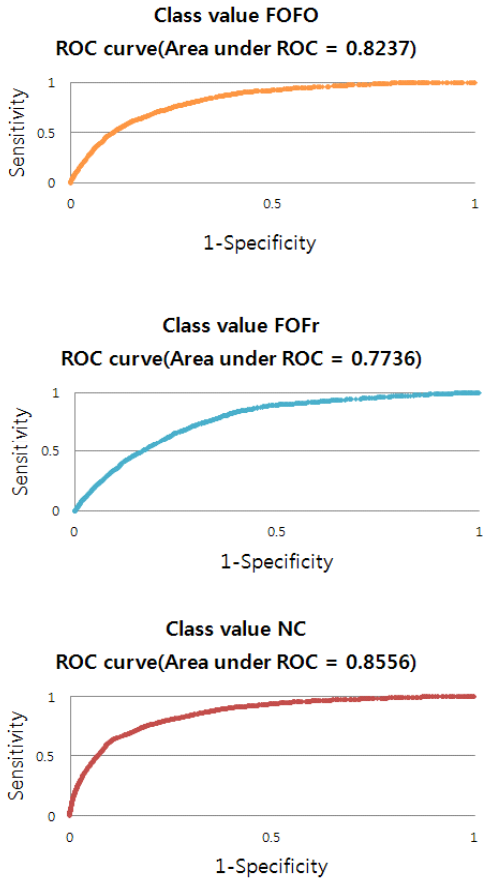

그림 3. Decision Table알고리즘의 클래스별 ROC curve
Fig. 3. ROC curve of Decision Table Algorithm According to the Class.

V. 결 론

현재 까지 트위터는 팔로어 수가 많은 사용자를 중심으로 친구 추천을 해주고 있다. 이것은 지구 반대편의 유명한 사람과 친구를 맺을 수 있는 장점을 가지고 있으나 사용자와 유사성 이 높은 친구를 추천해주지 못하는 단점을 가지고 있다. 본 논문에서는 이러한 단점을 극복하고자 트위터 사용자정보의 유사성을 기반으로 한 팔로어 예측 및 추천을 받을 수 있는 시스템을 구현하였다.

트위터 사용자정보의 유사성 통하여 정보가 비슷한 사용자에게 팔로어 관계를 추천 받을 수 있기 때문에 사용자는 더욱 친밀감을 가지고 트위터를 사용할 수 있게 된다. 또한 현재

트위터에서 진행되고 있는 추천시스템과 병행하여, 본 논문에서 제시한 정보의 유사성이 높은 사용자를 추천해준다면 트위터에서의 친구추천 단점을 극복하고 더욱 강력한 SNS로 발전할 것이다.

향후 연구로는 관계에 영향을 미칠 수 있는 속성을 추가할 예정이며, 네트워크에 구조를 활용한 분류분석을 수행할 예정이다.

참고문헌

〔1〕 Danah Boyd, and Nicole Elison, "Social Network Sites: Definition, History, and Scholarship," Journal of Computer-Mediated Communication, Vol. 13, No. 1, pp. 210-230, 2008.

〔2〕 J. Jansen, Zhang Mimi, Kate Sobel, and Abdur Chowdury, "Twitter power: Tweets as electronic word of mouth," Journal of the American Society for Information Science and Technology archive, Vol. 60, No. 11, pp. 2169-2188, November 2009.

〔3〕 Dong Hee, "Analysis of Online Social Networks: A Cross-National Study," Online Information Review, Vol. 34, No. 3, pp. 473-495, 2010.

〔4〕 R. Hazlewood, K. Makice, and W. Ryan, "Twitter space: A co-developed display using twitter to enhance community awareness," Proceeding PDC '08 Proceedings of the Tenth Anniversary Conference on Participatory Design, pp. 230-233 Indiana University Indianapolis, IN, USA, 2008.

〔5〕 A. Golder and Yardi, "Structural predictors of tie formation in twitter: Transitivity and mutuality," Proceedings of the second IEEE international conference on social computing, Minneapolis, MN, USA, August 2010.

〔6〕 P. Brandtzag, and J. Heim, "Why people use social networking sites," Online Communities and Social Computing, Vol. 5621, pp. 143-152, 2009.

〔7〕 R. Baker, and M. White, "Predicting adolescents' use of social networking sites from an extended theory of planned behaviour perspective," Computers in Human Behavior, Vol. 26, No. 6, pp. 1591-1597, 2010.

〔8〕 Kwak Haewoon, Lee Changhyun , Park Hosung , and Moon Sue, "What is Twitter, a social network or a news media?," Proceedings of the 19th international conference on World wide we, New York, USA, pp. 591-600, 2010.

〔9〕 Hong SamYull, "Comparative Analysis of User Access Factor of Twitter and Facebook," Korean Internet Information Association Autumn Conference Thesis. Vol. 11, No 2, pp. 248-252, 2010.

〔10〕 B. Ellison, C. Steinfield, and C. Lampe, "The benefits of Facebook friends: social capital and college students use of online social network sites," Journal of Computer-Mediated Communication, Vol. 12, No. 4, pp. 1143-1168, 2007.

〔11〕 Kuss Daria, Griffiths Mark, "Online Social Networking and Addiction: A Review of the Psychological Literature," Online Social Networking and Addiction: A Review of the Psychological Literature, Vol. 8, No. 9, pp. 3528-355, 2011.

〔12〕 B. Krishnamurthy, P. Gill, and M. Arlitt. A few chirps about twitter. In Proc. of the 1st workshop on Online social networks. ACM, 2008.

〔13〕 M. Chen,"Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others," Computers in Human Behavior, Vol. 27, No.2, pp. 755-762, 2011.

〔14〕 D. Davidiv, O. Oren Tsur, and A. Rappoport, "Enhanced sentiment learning using twitter hashtags and smileys," Proceedings of the 23rd International Conference on Computational Linguistics, 2010.

〔15〕 SNAP Stanford, http://snap.stanford.edu/.

〔16〕 I. Witten, E. Frank, and M. Hall, "Data Mining : Practical Machine Learning Tools and Techniques, Third Edition", Morgan Kaufmann Publishers, Burlington, MA 01803, USA, ISBN

978-0-12-374856-0, 2011.

〔17〕 M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and H. Ian, "The WEKA Data Mining Software: An Update", ACM SIGKDD Explorations Newsletter, Vol. 11, No. 1, pp. 10-18, 2009.

〔18〕 Hyeoniee Jeong, Youngmi Yoon, "Class prediction of an independent sample using a set of gene modules consisting of gene-pairs which were condition(Tumor, Normal) specific," Journal of The Korea Society of Computer and Information, Vol. 15, No. 12, pp. 197-207, 2010.

〔19〕 YoungmiYoon, Young-HoLee, "Emotion Classification System for Chatting Data," Journal of The Korea Society of Computer and Information, Vol. 14, No. 5, pp. 11-17, 2009.

〔20〕 Liaw, Andy and Wiener, Matthew "Classification and Regression by randomForest," R News Vol. 2 No.3, pp. 18-22, 2002.

저 자 소 개

계 용 선

2014: 가천대학교
컴퓨터공학과 학사 졸업 예정

관심분야: 데이터 마이닝,
소프트웨어공학

Email : yongsun0705@naver.com

윤 영 미

1981: 서울대학교
자연과학대학 학사 졸업

1983: 오하이오 주립대학
수학과 학사 수료

1987: 스탠포드대학교
컴퓨터학과 이학 석사 졸업

2008: 연세대학교
컴퓨터과학과 공학 박사 졸업

현 재: 가천대학교 컴퓨터공학과 교수

관심분야: 데이터베이스 시스템,
데이터 마이닝,
바이오인포매틱스

Email : ymyoon@gachon.ac.kr