

소셜 네트워크의 태그와 시간 정보를 반영한 추천 알고리즘[☆]

A recommendation algorithm which reflects tag and time information of social network

조 현1 홍 종 현2 최 준 연3* 김 성 희4
Jo Hyeon Hong Jong-hyun Choeh Joon Yeon Kim Soung Hie

요 약

최근 다수의 소셜 네트워크가 빠르게 확산되었다. 그 중에서도 소셜 북마킹 시스템은 가장 널리 사용되는 것 중 하나이다. 소셜 북마킹 시스템은 사용자가 온라인 자원에 태그를 부여해서 공유하고 관리할 수 있는 환경을 제공한다. 소셜 북마킹 시스템에서는 품질향상을 위해 태그와 시간 정보를 반영하여 개인에 특화된 추천을 할 수 있다. 본 논문에서는 가중치와 유사도 측정 과정에서 태그와 시간을 반영한 추천 시스템을 제안하였다. 또한 제안 방법론을 실제 데이터에 적용하였고, 실험결과 태그와 시간 정보를 함께 반영하였을 때 추천 성능이 향상됨을 확인하였다.

주제어 : 추천시스템, 협업 북마킹, 소셜 네트워크

ABSTRACT

In recent years, the number of social network system has grown rapidly. Among them, social bookmarking system(SBS) is one of the most popular systems. SBS provides network platform which users can share and manage various types of online resources by using tags. In SBS, it can be possible to reflect tag and time in order to enhance the quality of personalized recommendation. In this paper, we proposed recommender system which reflect tag and time of weight generation and similarity calculation. Also we adapted proposed method to real dataset and the result of experiment showed that the our method offers better performance when such information is integrated..

keyword : Recommender System, Social Bookmarking System, Social Network

1. 서 론

최근 인터넷 전자상거래 시스템이 지속적으로 진화하고 활용됨에 따라 기업이나 조직과 같은 전문 정보 제공자뿐만 아니라 개인 사용자들에 의해서도 정보가 활발하게 생성되고 공유되고 있다[1]. 특히 웹의 발전으로 인하여 인터넷 이용자들은 네트워크 서비스를 통해서 스스로 콘텐츠를 생산하여 공유하고, 집단적으로 협업을 수

행한다[2]. 이러한 서비스들 중에서도 태그를 활용하여 사이트나 블로그, 사진과 같은 콘텐츠들을 관리하는 소셜 네트워크 형태의 북마킹 시스템이 지속적으로 증가하고 있다. 특히 소셜 북마킹 시스템은 전통적인 분류학인 Taxonomy의 단점을 보완해주는 유용한 분류체계 방법으로 알려져 왔다[3]. 소셜 북마킹 시스템을 통해서 소셜 네트워크 이용자들은 웹 리소스를 설명하고 식별하기 위해 협업적으로 태그를 붙여 관리할 수 있다. 이를 통해 이용자들은 사회적 관심사를 공유하고, 유용한 정보에 접근한다[4]. 본 연구에서는 소셜 북마킹 서비스의 거래 데이터로부터 이용자들의 태깅 빈도와 시점을 고려하여 추천 시스템을 고안하였다. 기존의 연구들은 인터넷 전자상거래 이용자들의 일반적인 구매행위나 클릭 빈도 등을 토대로 추천 시스템을 연구하였지만, 본 연구에서는 소셜 네트워크 거래 데이터로부터 얻을 수 있는 태그 정보를 반영하여 특화된 시스템을 고안하였다. 태그 정보 기반의 추천 방법과 시점 정보 기반의 추천 방법 그리고 태그와 시점을 함께 반영한 추천 방법을 고안하여 실제 소셜

1 Department of Management Information Systems, Dong-A University, Busan, 602-760, Korea

2 Department of supercomputing strategy, KISTI, Daejeon, 305-806, Korea

3 Department of Digital Contents, Sejong University, Seoul, 143-747, Korea

4 KSIM, KAIST Business School, Seoul, 130-722, Korea

* Corresponding author (zoom@sejong.ac.kr)

☆ 본 논문은 동아대학교 교내연구비 지원에 의하여 연구되었음
[Received 16 October 2012, Reviewed 22 October 2012(R2 23 January 2013), Accepted 8 March 2013]

북마킹 거래 데이터에 적용하여 실험하였다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구와 관련된 선행 연구들을 정리하고 3장에서는 추천 시스템 개발 과정을 소개하며 4장에서는 실제 소셜 북마킹 거래 데이터에 적용한 실험 결과를 도출하였고, 5장에서 본 연구의 결론에 대해 논의하였다.

2. 선행 연구

본 연구의 연구 대상이 소셜 네트워크 서비스의 한 유형인 소셜 북마킹 시스템이며, 연구 방법이 추천 시스템이므로 소셜 네트워크와 추천 시스템 분야의 선행 연구들을 살펴보고자 한다.

2.1. 소셜 네트워크

소셜 네트워크는 사회과학 분야를 비롯한 다양한 영역에서 사회 주체들의 네트워크 관계를 규명하는데 이용되는 이론적 구조를 의미한다. 소셜 네트워크는 노드와 링크로 구성이 되는데, 노드는 일반적으로 사회적 관계의 주체가 되는 개인이나 조직 등을 의미하고, 링크는 그들 간의 관계를 이론적으로 표현한다.

소셜 네트워크 분야의 연구는 2000년 초반 웹 사용자들의 데이터 수집이 가능해지면서 활발하게 진행되었다 [5-8]. 이들은 인터넷 사용자들의 행위 패턴 및 웹 사이트 간의 네트워크 현상을 연구하였는데, 분석에 따르면 네트워크상에 존재하는 임의의 노드는 다른 노드에 있어서 상대적으로 소수의 링크만을 거치고도 도달할 수 있다. 또한 이들에 따르면 네트워크 행위를 하는 노드들의 분포는 일반적인 사회현상에 발견되는 멱함수 분포를 따른다고 한다. 특히 Albert and Barabasi가 제안한 연구 모형에 따르면 네트워크는 국지적인 흐름에 따라 진화를 하는데, 특히 시간의 흐름에 따라 새로운 노드가 생성이 되고, 오래된 링크들이 새롭게 부활되는 등의 현상이 나타난다[9]. 이들은 각 요소들의 상대적인 빈도를 통해 대상 네트워크가 통태일 분포인지 지수 분포인지를 판별할 수 있다고 주장하였다. Newman은 협업적 네트워크의 경우 멱함수 분포를 따르며, 노드 선택의 근거는 선호적 연결에 의해 결정된다고 주장하였다[10-12]. 추천 분야에서도 소셜 네트워크에 관한 연구가 진행되어 왔다. McDonald은 소셜 네트워크를 활용하여 추천 시스템을 연구하였는데, 거시적인 네트워크 집합체를 통한 추천은 일반 개인에게는 최적의 추천을 선사하지 못한다는 결론을 제시하

였다[13]. Palau et al.은 추천 시스템 상에서의 협업 기능을 분석하고 표현하기 위해 소셜 네트워크를 사용하였고 [14], Lam은 SNACK이라는 방법론을 제시하여 추천을 위한 자동화된 협업 시스템을 검증하였다[15]. 이시화 등은 콘텐츠 추천을 위한 태그 기반의 소셜 네트워크 구축을 연구하였으며[16], 엄태영 등은 태그 네트워크를 이용하여 개인화 북마크 추천 시스템을 제안하였고[17], 조현 등은 온라인 태깅 시스템의 사용자들 행위 패턴 정보를 바탕으로 태그 네트워크를 구축하고 정보를 추출하여 제시한 방법론을 검증하였다[18]. 이렇듯 소셜 네트워크를 활용하거나 소셜 네트워크에 여러 방법론을 적용한 다양한 연구들이 진행되어왔다.

2.2. 추천 시스템

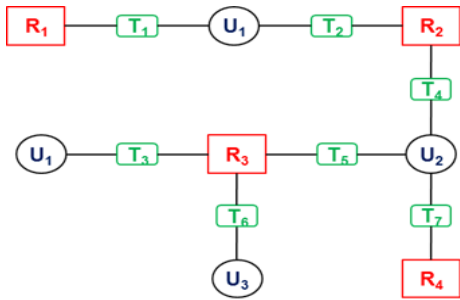
데이터 정보 분야에서 대표적인 추천 시스템 방법으로 협업적 필터링(Collaborative Filtering: CF)을 들 수 있다. 협업적 필터링은 구매자나 사용자의 행위 이력을 바탕으로 선호도가 유사한 이웃군을 추출하여 추천 목록을 제시해주는 방법론이다. 전통적인 협업적 필터링은 사용자-아이템 행렬과 수치화된 가중치로 구성이 된다[19]. 일반적으로 선호도 지표는 사용자, 아이템, 선호도 정보로 구성이 되는데, 사용자가 명시적으로 특정 아이템에 수치를 부여하여 아이템에 대한 선호 성향을 측정하게 된다. 협업적 필터링에 사용되는 선호도 지표들은 인터넷 전자상거래에서의 클릭 행위, 구매 이력 혹은 암묵적으로 웹 방문 기록을 통해서 얻어진다[20]. 이러한 경우 특정 사용자가 대상 아이템(웹 페이지, 상품)을 ‘보거나’, ‘구매했다’는 식의 방식으로 기록에 남는다.

협업적 필터링에 대한 연구 분야에서는 개인화된 추천 시스템 고안을 위해서 정형화된 프로세스를 고안하거나, 정황적 정보를 활용한 연구들이 꾸준히 수행되어왔다[21-23]. Adomavicius et al.은 장소와 시간 같은 정황적 정보를 포괄하는 다차원적 추천을 위해, 축약 기반의 접근을 제안하였으며[24], Chen은 정황 인지가 가능한 협업적 시스템을 제안하여 특정 환경 속에서 미래 선호도를 예측하여 검증하였다[25]. 소셜 네트워크 및 협업적 공유 시스템이 보급되면서 사회적 관계나, 태그 부여 빈도, 태깅 시점 등과 같은 추가적인 정황 정보들이 가용해졌다. 특히 태그 정보의 경우 다양한 연구들을 통해서 그 중요성이 입증되었다. 이동균과 권준희의 연구에서는 태그의 중요성을 입증하기 위해 동일한 태그를 공유하는 웹 문서들 간의 유사도를 측정하고, 새로운 클러스터링 알고

리즘을 제안하였다[26]. Kim et al.은 태그를 활용한 추천 시스템을 제공하기 위하여 협업적 여과 기법을 도입하였다. 또한 이들은 협업적 태깅을 도입하여 데이터 희박성(sparseness)이나 cold-start 문제를 해결하였다[27]. 태그 네트워크를 이용한 개인화 북마크 추천 시스템을 제안하여 북마크 사이트에서 제시하는 순위에 비해 순위 정보 성능이 더 우수함을 입증한 연구가 있었으며[17], 사용자, 객체(object), 태그의 세 개체로 이루어진 관계를 활용하여 개인화 추천 시스템을 개발한 연구가 있었다[28]. 사용자의 태깅 정보를 수집하여 태그를 클러스터링하고 선호도가 유사한 집단을 구성한 연구가 있었고[16], 기존의 태그 추천 방법론들을 비교·분석하고 유형별로 분류한 연구가 있었다[29].

3. 연구 방법

소셜 네트워크의 한 유형인 소셜 북마킹 시스템은 사용자(user), 리소스(resource), 태그(tag)의 세 가지 객체로 이루어지는 구조를 지닌다[28]. 한 명의 사용자는 하나 이상의 태그를 사용하여 리소스를 공유하는데, 그들 간의 일차적인 관계도를 도식화 하면 (그림 1)과 같다.



(그림 1) 3 객체 구조
(Figure 1) Three object structure

본 연구에서는 이러한 구조를 통해 얻을 수 있는 사용자들의 태깅 행위정보를 활용하여 선호도와 유사도를 측정하기 위한 산정 방법을 제안하였다. 구체적인 협업적 필터링은 총 세 단계로 구성된다. 첫 단계인 특정 리소스에 대한 가중치 생성 부분에서는 소셜 북마킹 사용자들의 태그 부여 빈도 및 시점을 기반으로 가중치를 생성하였고, 두 번째 단계에서는 가중치 벡터를 바탕으로 사용자들 간의 유사도를 측정하였다. 마지막 단계에서는 사용자들의 가중치 정보와 이웃군 정보를 바탕으로 선호도

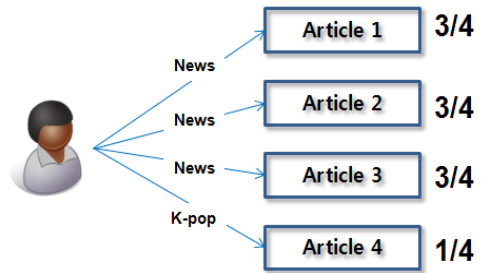
측정을 수행하였다.

3.1 가중치 생성

가중치 생성 단계에서는 소셜 북마킹 시스템으로부터 고유하게 얻을 수 있는 태그 빈도 정보와 시간 정보에 근거하여 가중치를 산정하였다.

3.1.1 태그 기반 가중치

사용자가 특정 태그를 다른 태그에 비해 많이 사용하면, 그 태그와 관련된 리소스에 더 많은 가중치를 부여해야 할 것이다. 예를 들어 어떤 사용자가 'News'라는 태그를 세 번 사용하고, 'K-pop'이라는 태그를 한번 사용하였다면 'News'에 더 관심을 가진다고 생각할 수 있다.



(그림 2) 태그 빈도 기반의 가중치
(Figure 2) Weight based on tag frequency

따라서 본 연구에서는 특정 사용자 u 가 리소스 r 에 갖는 태그 기반의 가중치 태그 벡터 $TaW(u,r)$ 를 다음과 같이 정의 한다.

$$TaW(u,r) = \sum_{t_a \in tag(u,r)} w_{u,t_a} \quad (1)$$

t_a 는 특정 사용자가 부여한 개별 태그를 의미하고, w_{u,t_a} 는 사용자 u 가 t_a 라는 태그를 활용한 상대지표로써, 전체 태깅 빈도 수 대비 특정 태그의 사용비율을 의미한다.

3.1.2 시간 기반 가중치

사용자들은 다양한 시점에 태그를 부여한다. 인간의 관심사는 시간에 따라 변화할 수 있는데, 일반적으로 과거에 태그를 부여한 리소스보다는 최근에 태그를 부여한 리소스에 더 관심을 가질 것이다. 따라서 본 연구에서는

태그의 부여 시점을 활용하여 가중치를 산정하였다. 관심사의 변화나 망각의 정도를 연구함에 있어서 지수함수가 널리 이용되어 왔다[30]. 특정 사용자 u 가 리소스 r 에 갖는 시점 기반의 가중치 타임 벡터 $TiW(u,r)$ 를 다음과 같이 정의 한다.

$$TiW(u,r) = exp\left(-ln2 \times \frac{t}{hs_t}\right) \quad (2)$$

$$t = \begin{cases} 0, & t_i = t_l \\ \frac{t_l - t_i + 1}{t_l - t_f + 1}, & t_i \neq t_l \end{cases}$$

t_l 은 특정 사용자가 태그를 부여한 가장 마지막 날짜를 의미하고, t_i 는 특정 태그를 부여한 시점을 의미하며 t_f 는 태그를 부여한 첫 번째 날짜를 뜻한다. hs_t 는 특정 사용자가 태그를 부여한 전체 타임 스패를 1로 환산하였을 때 정확히 중간이 되는 지점에서의 시점 지표를 의미한다. 태그의 부여 시점이 가장 최근 날짜와 동일하면 t 값은 0이 되어 가중치는 1의 값으로 예측되고, 태그의 부여 시점이 타임 스패의 정확히 중간에 위치하면 가중치는 1/2의 값으로 떨어진다. 결국 최근에 태그를 부여한 리소스일수록 높은 값을 지니며, 과거에 태깅한 리소스는 낮은 값으로 산출된다.

3.1.3 종합 가중치

앞선 태그 기반 가중치와 시점 기반의 가중치를 함께 반영하기 위하여 종합 가중치(ToW: Total Weight)를 고안하였다. 종합가중치는 선형조합으로 구성되었고, λ 의 크기에 따라서 태그와 시간의 반영 비율이 상대적으로 변화한다.

$$ToW(u,r) = \lambda \cdot TiW(u,r) + (1-\lambda) \cdot TaW(u,r) \quad (3)$$

3.2 사용자 유사도 측정

사용자들 간의 유사도를 측정할 때는 일반적으로 두 벡터의 거리를 계산하는데 사용하는 cosine similarity를 활용하였다. 태그 가중치 벡터와 타임 가중치 벡터 그리고 종합가중치 벡터를 활용하여 다음과 같이 유사도를 측정하였다.

$$Sim_i(u,v) = \frac{\sum w_i(u,r) \cdot w_i(v,r)}{\sqrt{\sum (w_i(u,r))^2} \cdot \sqrt{\sum (w_i(v,r))^2}} \quad (4)$$

i 는 태그, 타임, 종합 가중치를 의미하고, $w_i(u,r)$ 과 $w_i(v,r)$ 은 각각 사용자 u 와 v 의 리소스 r 에 대한 가중치 벡터를 의미한다.

3.3 선호도 예측

추천 목록을 도출하기 위해서는 마지막으로 사용자가 리소스를 얼마나 선호할지에 대한 수치를 측정하여야 한다. 앞에서 측정한 가중치 벡터와 유사도 행렬을 활용하여 다음과 같이 선호도 계측을 실시하였다.

$$P(u,r) = \sum_{v \in KNN(u)} (w_i(v,r)) \cdot Sim_i(u,v) \quad (5)$$

사용자 u 가 리소스 r 에 대한 선호 예측치는 유사도가 가장 높은 K 명의 이웃들($KNN(u)$)과의 유사도와 그 이웃들의 리소스 r 에 대한 가중치 벡터를 함께 반영하여 산출한다.

4. 실험 결과

4.1 데이터 및 실험 절차

본 연구에서는 미국의 대표적인 논문 서지 관리 및 공유 서비스인 CiteULike와 미국의 소셜 북마킹 사이트인 Del.icio.us의 거래 기록 자료를 수집하여 활용하였다. CiteULike은 공개된 데이터셋[31]을 제공하고 있으며, delicious 자료는 Web crawling 기법으로 수집하였다.

태깅 행위를 20번 이상으로 한 사용자들만을 추출하여 그 중에서도 일부를 무작위로 채택하였다. 각 사이트 별 데이터 규모는 표 1과 같다. 데이터는 과거 80%의 기록을 훈련 집단(Training Set)으로 분류하고, 나머지는 20%의 기록을 실험 집단(Testing Set)으로 분류하였다. 각 사용자별로 추천 목록은 10개를 추출한 후 실험결과를 측정하였다. 이웃군의 규모는 3명에서부터 30명까지 3명씩 증가시켜 실험을 반복하였다.

(표 1) 실험 데이터 요약
(Table 1) Experiment Data Summary

	CiteULike	Del.icio.us
사용자 수	50,056	43,028
리소스 수	20,506	9,383

4.2 평가 지표

성능 평가를 하기 위해서 추천 시스템 성능 평가에 널리 활용되는 Precision, Recall 그리고 F-ratio를 도입하였다[32].

$$Precision = \frac{|test \cap top K|}{|top K|} \quad (6)$$

$$Recall = \frac{|test \cap top K|}{|test|} \quad (7)$$

top K는 실험 대상에서 도출된 K개의 추천 목록을 의미하며, test는 각 사용자들이 실제로 사용한 리소스 집합을 의미한다. 추가적으로 Precision과 Recall의 조화평균을 측정해 이들의 지표를 함께 반영한 F-ratio를 측정하였다.

$$F = \frac{2 \cdot Recall \cdot Precision}{(Recall + Precision)} \quad (8)$$

4.3 실험 결과

실험 및 성과 측정에 앞서, 3.1.3에서 기술한 바와 같이 태그 기반 가중치와 시점 기반 가중치간의 최적 비율을 탐색하기 위한 사전 분석을 수행하였다.

(표 2) 유사도 측정 방법 별 λ 수치
(Table 2) λ value according to similarity measurements

Site	I	TaVS	TiVS	ToVS
CiteULike	P	0.2	0	0.2
	R	0.2	0.3	0.2
	F	0.2	0.3	0.2
Delicious	P	1	1	0.5
	R	0.4	0.3	0.8
	F	0.4	0.5	0.5

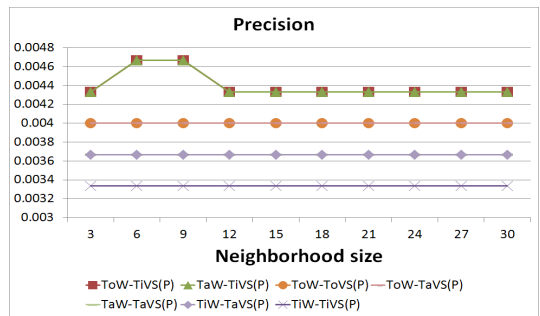
각 유사도 측정의 유형별, 평가 지표별로 태그와 시간의 상대 비율을 0에서 부터 1까지 0.1씩 변화시켜 가면서 가장 우수한 성능을 보이는 λ값을 관측한 결과 아래와 같은 결과를 얻었다. 본 실험에서 사용한 CiteULike 및 Delicious 별로 λ값을 (표 2)에 정리하였다.

TaVS(Tag Vector Similarity)는 태그 기반의 가중치를 토대로 유사도를 측정된 방법을 의미하고, TiVS(Time Vector Similarity)와 ToVS(Total Vector Similarity)는 각각 시간 벡터와 종합 벡터를 활용하여 유사도를 측정된 방

법을 지칭한다. P, R, F는 각각 Precision, Recall, F-ratio를 의미한다.

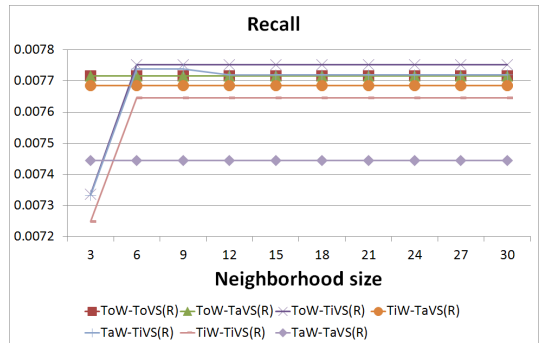
4.3.1 CiteULike 결과

(그림 3)은 CiteULike 사용자들의 이웃군의 규모를 변화시켰을 때 나타나는 Precision 지표의 결과 값을 담고 있다. Precision 지표 수치가 가장 높은 추천 방법(ToW-TiVS, TaW-TiVS)과 가장 낮은 추천 방법 (TiW-TiVS)을 비교했을 때, 이웃군의 규모별로 성능 수치가 0.10%에서 0.13% 정도의 차이를 보였다.



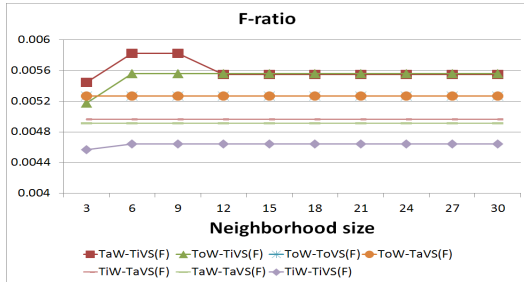
(그림 3) 추천 방법별 Precision 결과
(Figure 3) Precision result by recommendation methods

Recall 지표 수치의 경우 다양한 이웃군 규모 전반에 걸쳐 ToW-TiVS 방법이 가장 높게 나타난 반면, TaW-TaVS 방법이 가장 낮게 기록됐다. 가중치와 유사도를 측정함에 있어서 태그 빈도와 시점을 함께 반영함으로써 성능을 향상 시킬 수 있음을 알 수 있다.



(그림 4) 추천 방법별 Recall 결과
(Figure 4) Recall result by recommendation methods

(그림 5)에서는 이웃군 규모에 따라 제안된 방법들의 F-ratio를 보여준다.

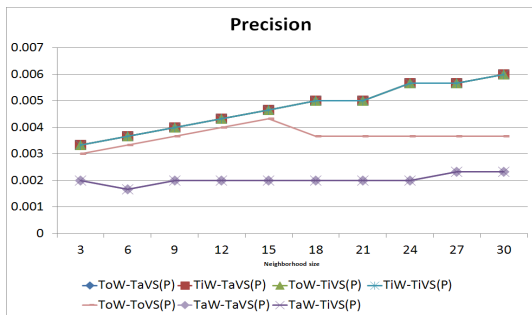


(그림 5) 추천 방법별 F-ratio 결과
(Figure 5) F-ratio result by recommendation methods

전반적으로 TaW-TiVS 방법이 가장 우수한 성능을 보였으며, TiW-TiVS 방법이 가장 낮은 수치를 기록했다. 지표 수치는 0.09%에서 0.12% 정도의 차이를 보였으며, 이웃군 규모가 증가함에 따라 F-ratio 수치도 조금 상승하는 것을 관찰 할 수 있었다.

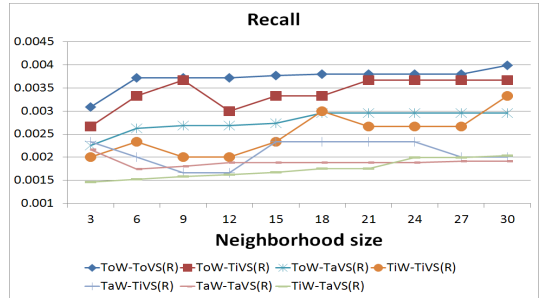
4.3.2 Delicious 결과

(그림 6)은 Delicious 사용자들의 이웃군의 규모를 변화시켰을 때 나타는 Precision 지표의 결과 값을 담고 있다.



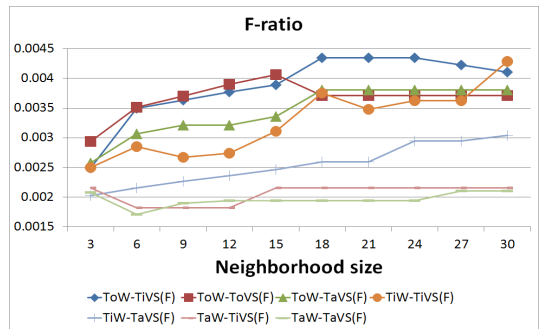
(그림 6) 추천 방법별 Precision 결과
(Figure 6) Precision result by recommendation methods

Precision 지표 수치의 경우, ToW-TaVS, ToW-TiVS, TiW-TaVS, TiW-TiVS 방법이 동일하게 가장 좋은 성능을 보였다. 이들의 성능과 가장 나쁜 성능의 방법 (TaW-TiVS, TaW-TaVS)을 비교했을 때, 이웃군의 규모별로 성능 수치가 0.03%에서 0.33% 정도의 차이를 보였다.



(그림 7) 추천 방법별 Recall 결과
(Figure 7) Recall result by recommendation methods

Recall 지표 수치의 경우 다양한 이웃군 규모 전반에 걸쳐 ToW-ToVS 방법이 가장 높게 나타난 반면, TiW-TaVS 방법이 가장 낮게 기록됐다.



(그림 8) 추천 방법별 F-ratio 결과
(Figure 8) F-ratio result by recommendation methods

(그림 8)에서는 이웃군 규모에 따라 제안된 방법들의 F-ratio를 보여준다. 이웃군의 규모에 따라서 이웃이 15명 이하일 때 까지는 ToW-ToVS 방법이 가장 우수한 성능을 보였으나, 15명을 초과하는 경우부터는 ToW-TiVS 방법이 가장 우수한 성능을 나타냈다. 각 이웃군 규모별로 성능이 가장 좋은 경우와 나쁜 경우를 비교했을 때 성능의 차이는 0.12%에서 0.30% 정도까지 나타났다.

4.3.3 실험 결과 정리

소셜 네트워크의 협업 분류 서비스의 실제 거래 기록 데이터를 대상으로 세 가지의 성능 지표로 검증 해 본 결과, 태그 빈도나 태그 시점을 단독으로 반영하는 것 보다 이들의 정보를 함께 감안할 때 성능이 향상되는 것을 알 수 있었다.

(표 3)과 (표 5)는 각 성능 지표별로 가장 우수한 방법에서부터 순차적으로 정리한 결과이며, (표 4)와 (표 6)은 이때 각각의 방법별 성능지표 결과이다.

태깅 정보와 시간 정보는 이웃군의 규모 및 소셜 네트워크의 사용 환경에 따라서 상대적인 중요도가 변화함을 알 수 있다. 즉, 대중의 사회적 관심사가 개인의 개별 관심사 및 시간에 따라 변화하듯이, 본 실험에서 사용된 논문 서지 관리 사이트 및 소셜 북마킹 사이트의 사용자 집단도 시간의 흐름에 따라 관심사 및 북마킹 패턴이 변화함을 확인 하였다.

(표 3) 추천 방법별 성능 순위 (CiteULike)
(Table 3) Performance rank by recommendation method (CiteULike)

	Precision	Recall	F-ratio
1	ToW-TiVS	ToW-ToVS	TaW-TiVS
2	TaW-TiVS	ToW-TaVS	ToW-TiVS
3	ToW-ToVS	ToW-TiVS	ToW-ToVS
4	ToW-TaVS	TiW-TaVS	ToW-TaVS
5	TaW-TaVS	TaW-TiVS	TiW-TaVS
6	TiW-TaVS	TiW-TiVS	TaW-TaVS
7	TiW-TiVS	TaW-TaVS	TiW-TiVS

(표 4) 추천 방법별 성능 지표 (CiteULike)
(Table 4) Performance value by recommendation method

	Precision	Recall	F-ratio
1	0.004400	0.007716	0.005595
2	0.004400	0.007716	0.005521
3	0.004400	0.007710	0.005269
4	0.004400	0.007686	0.005269
5	0.003666	0.007684	0.004965
6	0.003666	0.007607	0.004913
7	0.003333	0.007445	0.004635

(표 5) 추천 방법별 성능 순위 (Delicious)
(Table 5) Performance rank by recommendation method (Delicious)

	Precision	Recall	F-ratio
1	ToW-TaVS	ToW-ToVS	ToW-TiVS
2	TiW-TaVS	ToW-TiVS	ToW-ToVS
3	ToW-TiVS	ToW-TaVS	ToW-TaVS
4	TiW-TiVS	TiW-TiVS	TiW-TiVS
5	ToW-ToVS	TaW-TiVS	TiW-TaVS
6	TaW-TaVS	TaW-TaVS	TaW-TiVS
7	TaW-TiVS	TW-TaVS	TaW-TaVS

(표 6) 추천 방법별 성능 지표 (Delicious)
(Table 6) Performance value by recommendation method (Delicious)

	Precision	Recall	F-ratio
1	0.004733	0.003719	0.003864
2	0.004733	0.003400	0.003668
3	0.004733	0.002778	0.003445
4	0.004733	0.002500	0.003263
5	0.003667	0.002100	0.002541
6	0.002033	0.001894	0.002053
7	0.002033	0.001739	0.001958

5. 결 론

본 연구에서는 최근 활발하게 활용되고 있는 소셜 네트워크의 한 유형인 소셜 북마킹 시스템을 대상으로 추천 시스템 방법론을 연구하였다. 소셜 북마킹 시스템의 고유한 성격을 분석하여 태그 빈도와 태깅 시점 정보를 추출하였고, 이를 활용하여 사용자들의 선호도 가중치 및 유사도 측정 방법을 제안하였다. 선호도 측정 방법에는 태그 정보를 반영하는 태그 벡터와 태깅 시점을 반영한 타임 벡터, 태그와 타임을 함께 반영한 토탈 벡터를 제시하였다. 토탈 벡터는 태그 빈도와 태깅 시점을 선형으로 조합하여 단계별로 비율을 변화시키며 구성하였다. 선호도 벡터를 기초로 유사도 측정을 수행하였으며, 특정 사용자와 가장 유사한 집단을 3명부터 30명까지 3명 간격으로 변화시키며 실험을 수행하였다. 성능 평가는 추천 시스템의 평가에 널리 활용되고 있는 Precision, Recall, F-ratio를 도입하여 수행하였다.

실험결과, 각각의 유사도 측정 방법 및 평가 지표 별로 λ 값이 0에서 1까지 다양하게 기록되었다. 즉, 태그 빈도 정보가 태깅 시점 정보보다 상대적으로 의미가 큰 것으로 나타났다. 각 평가 지표별로 성능을 비교한 결과, 태그 빈도나 태깅 시점 어느 한 정보만을 단독으로 반영한 방법보다는 이 두 정보를 함께 반영했을 때 성능이 향상되는 것으로 나타났다. 이웃군의 규모별로는 12명까지는 성능에 차이가 나타났으나 그 이후부터는 모든 평가 지표에서 성능이 동일하게 나타났다. 즉, 소셜 북마킹 서비스 제공자가 사용자 제고를 위해 추천 시스템을 운영한다면, 특정 임계치 수준의 이웃만을 반영하면 될 것이다.

본 연구에서는 고안한 추천 방법을 특정 사이트 자료에만 적용하였다는 한계가 있지만 여러 소셜 북마킹 시스템에도 확대 적용을 하여 시사점을 도출하고 개선한다

던 소셜 네트워크 서비스 발전에 유용한 지침을 제공할 수 있으리라 본다.

참 고 문 헌(Reference)

- [1] Choi, J. Y., Rosen, J., Maini, S., Pierce, M. E. and Fox, G. C., "Collective collaborative tagging system", Grid computing environment workshop, 2008, pp. 1-7.
- [2] Yu, H. L. and Song, I. G., "Social network evolution according web service type change", Journal of Korean Society for Internet Information, Vol 11, No. 3, 2010, pp. 52-62.
- [3] Golder, S. A. and Huberman, B. A., "Usage patterns of collaborative tagging system", Journal of Information Science, Vol. 32, No. 2, 2006, pp. 198-208.
- [4] Jo, H., Cheoh, J. Y. and Kim, S. H., "A Study About User Pattern of Social Bookmarking System", Journal of Korean Society for Internet Information, Vol. 12 No. 5, 2011a, pp. 29-37.
- [5] Albert, R., Jeong, H. and Barabási, A.-L., "Diameter of the World-Wide Web", Nature, Vol. 401, No. 1999, pp. 130-131.
- [6] Adamic, L. A., "The Small World Web", 99 Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries, Vol. 1696, pp. 443-452.
- [7] Dodds, P. S., Muhamad, R. and WatTiVS, D. J., "An experimental study of search in global social networks", Science, Vol. 301, No. 2003, pp. 827-829.
- [8] Crossley, N., "Review article: the new social physics and the science of small world networks", The Sociological Review, Vol. 53, No. 2005, pp. 351-359.
- [9] Albert, R. and Barabási, A.-L., "Topology of Evolving Networks: Local Events and Universality", Physical Review Letters, Vol. 85, No. 24, 2000, pp. 5234-5237.
- [10] Newman, M. E. J., "The structure and function of complex networks", SIAM Review, Vol. 45, No. 2003, pp. 167-256.
- [11] Newman, M. E. J., "Fast algorithm for detecting community structure in networks", Physical Review E, Vol. 69, No. 6, 2004, pp. 1-5.
- [12] Newman, M. E. J., "Power laws, pareto distributions and zipf's law", Contemporary Physics, Vol. 46, No. 5, 2005, pp. 323-351.
- [13] McDonald, D. W., "Recommending collaboration with social networks: a comparative evaluation", CHI '03 Proceedings of the SIGCHI conference on Human factors in computing systems, 2003, pp. 593-600.
- [14] Palau, J., Montaner, M., López, B. and Rosa, J. L. d. l., "Collaboration analysis in recommender systems using social networks", Cooperative Information Agents VIII, Vol. 3191, No. 2004, pp. 137-151.
- [15] Lam, C., "SNACK: incorporating social network information in automated collaborative filtering", The 5th ACM Conference on Electronic Commerce, 2004, pp. 254-255.
- [16] Lee, S., Park, S., Lee, M. and Hwang, D., "A Study on Construction of Tag-based Social Network for Content Recommendation", Proceedings of Korean Multimedia Society, Vol. 12 No. 1, 2009, pp. 152-155.
- [17] Eom, T. Y., Kim, W. and Park, S., "Personalized Bookmark Recommendation System Using Tag Network", The Journal of Society for e-Business Studies, Vol. 15, No. 4, 2010, pp. 181-195.
- [18] Jo, H., Cheoh, J. Y. and Kim, S. H., "A Study on Information Retrieval of On-line Tagging System", Journal of Korean Institute of Information Technology, Vol. 9 No. 10, 2011b, pp. 215-221.
- [19] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J., "Analysis of recommendation algorithms for E-commerce", The Second ACM Conference on Electronic Commerce, 2000, pp. 158-167.
- [20] Breese, J. S., Heckerman, D. and Kadie, C., "Empirical analysis of predictive algorithms for collaborative filtering", The 14th Conference on Uncertainty in Artificial Intelligence, 1998, pp. 43-52.

- [21] Kim, K., "A Hybrid Collaborative Filtering Algorithm for Personalized Recommendation and its Application to the Internet Electronic Commerce", *The Journal of Internet Electronic Commerce Research*, Vol. 8, No. 4, 2008, pp. 1-20.
- [22] Kim, Y., -s. and Kang, J., -g., "Development of a Process for Recommender System Applications and a Framework of an Analysis of Personalizability", *The Journal of Internet Electronic Commerce Research*, Vol. 9, No. 3, 2009, pp. 213-241.
- [23] Kim, J., Ahn, B. and Jung D., "A Recommender System using Mixed Filtering for Health Products", *The Journal of Internet Electronic Commerce Research*, Vol 12, No. 2, 2012, pp. 109-124.
- [24] Adomavicius, G., R. Sankaranarayanan, S. Sen and Tuzhilin, A., "Incorporating contextual information in recommender systems using a multidimensional approach", *ACM Transactions on Information Systems* Vol. 23, 2005, pp. 103-145.
- [25] Chen, A., "Context-Aware Collaborative Filtering System: Predicting the User's Preferences in Ubiquitous Computing Environment", *Lecture Notes in Computer Science*, Vol. 3479, No. 2005, pp. 244-253.
- [26] Lee, D. and Kwon, J., "Recommendation of Subscribed Tag Using Folksonomy Mashup", *Korean Institute of Information Technology Proceeding*, 2009, pp. 544-547.
- [27] Kim, H. N., Ji, A. T., Ha, I. and Jo, G. S., "Collaborative filtering based on collaborative tagging for enhancing the quality of recommendation", *Electronic Commerce Research and Applications*, Vol. 9, No. 1, 2010, pp. 73-83.
- [28] Shang, M.-S., Lu, L., Zhang, Y.-C. and Zhou, T., "Empirical analysis of web-based user-object bipartite networks", *Europhysics Letters*, Vol. 90, 2010, 48006.
- [29] Kim, H., Lee, K. and Kim, H., "Tag Recommendation Algorithms in Tagging System", *Journal of Information Science Society: Computing reality and letter*, Vol. 16권, No. 9, 2010, pp. 927-932.
- [30] Aggarwal, C. C., Han, J., J. Wang and Yu, P. S. "A framework for projected clustering of high dimensional data streams", *Proceedings of the Thirtieth international conference on Very large data bases*, Toronto, Canada. 2004.
- [31] <http://www.citeulike.org/faq/data.adp>
- [32] Herlocker, J. L., Konstan, J. A., Terveen, L. G. and Riedl, J. T., "Evaluating collaborative filtering recommender systems", *ACM Transactions on Information Systems*, Vol. 22, No. 1, 2004, pp. 5-53.

◎ 저 자 소개 ◎

조 현



2004년 KAIST 경영공학과 졸업(학사)
2006년 KAIST 경영공학과 졸업(석사)
2012년 KAIST 경영공학과 졸업(박사)
2013년~현재 동아대학교 경영정보학과 교수
관심분야 : Behavioral MIS, Social network, Intelligent information system
E-mail : sineoriz@gmail.com

홍 증 현



2011년 한양대학교 정보기술경영학과 졸업(학사)
2013년 KAIST 경영공학과 졸업(석사)
2013년~현재 KISTI 국가슈퍼컴퓨팅연구소 슈퍼컴퓨팅전략실 연구원
관심분야 : data analytics
E-mail : hyunsit@gmail.com

최 준 연



1996년 KAIST 전기 및 전자공학과 졸업(학사)
1998년 KAIST 경영공학과 졸업(석사)
2007년 KAIST 경영공학과 졸업(박사)
2008년~현재 세종대학교 디지털콘텐츠학과 교수
관심분야 : 데이터마이닝, 추천시스템.
E-mail : zoon@sejong.ac.kr

김 성 희



1973년 서울대학교 섬유공업 졸업(학사)
1978년 University of Missouri-Columbia 산업공학 졸업(석사)
1983년 Stanford university 경영정보학과 졸업(박사)
1983~현재 한국과학기술원 교수
관심분야 : 경영정보, 의사결정 지원 시스템, 전자상거래, 모바일 정책, 그린 IT
E-mail : seekim@business.kaist.ac.kr