

# 매개 중심성 분석을 이용한 콘텐츠 추천 방법

Yun-feng Pei, 손종수, Qing Wang, 송태성, 정인정  
고려대학교 컴퓨터정보학과

e-mail : {beayunfeng, mis026, wangqing, song87, chung}@korea.ac.kr

## Contents Recommendation Method Using Betweenness Centrality Analysis

Yun-feng Pei, Jong-soo Sohn, Qing Wang, Tae-sung Song, In-jeong Chung  
Dept. of Computer and Information Science, Korea University

### 요 약

최근 웹 2.0의 발전 및 사용자들의 커뮤니티에 대한 수요가 확대되면서 소셜네트워크 서비스(Social network service)들이 큰 인기를 얻고 있으며 SNS를 통해 많은 콘텐츠가 유통되고 있다. 그러나 사용자들에게 필요한 정보를 찾는 데 있어서 어려운 판별 과정이 요구된다. 이에 소셜네트워크 서비스 사용자들간에 공유된 정보들을 추천하기 위한 연구가 요구되고 있다. 본 논문에서는 소셜네트워크 분석 기법(Social network analysis) 중 하나인 매개 중심성(Betweenness centrality) 분석을 활용하여 양질의 콘텐츠를 추천하는 방법을 제안한다. 본 논문에서 제안한 방법은 소셜네트워크 서비스인 트위터(Twitter)의 사용자들이 배포한 타임라인(Timeline) 콘텐츠와 사용자의 프로필(Profile)을 수집하여 매개 정도 중심성을 분석한다. 그리고 분석된 매개 정도 중심성의 값을 이용하여 콘텐츠를 추천한다. 본 논문에서 제안한 방법을 검증하기 위하여 수집된 데이터 중에서 10개 키워드에 관련된 콘텐츠를 추출한 후, 이에 대한 정확도를 분석하였다. 본 논문에서 제안한 추천 방법은 기존 소셜네트워크 서비스에서 제공하는 타임라인 콘텐츠보다 약 41% 높은 정확도를 보인다. 본 논문에서 제안한 방법을 활용하면 사용자에게 양질의 콘텐츠를 추천할 수 있다.

### 1. 서 론

최근 웹 2.0의 발전 및 온라인 사용자들이 커뮤니티에 대한 수요가 확대되면서 소셜네트워크 서비스들이 큰 인기를 얻고 있다. 웹 2.0 기술을 활용한 소셜네트워크 서비스는 사용자의 지식 및 경험에 대한 콘텐츠를 작성하고 배포할 수 있는 기능을 제공한다[1].

인터넷이 발달하고 소셜네트워크 서비스들이 발전함에 따라 사용자들이 배포한 콘텐츠의 양이 폭발적으로 증가하고 있다. 이에 사용자들이 필요한 정보를 찾는 데 있어서 어려운 판별 과정이 요구된다[2].

본 논문에서는 기존 소셜네트워크 서비스에서 제공되는 타임라인 형식의 한계를 극복하기 위해 매개 중심성 분석을 이용한 콘텐츠 추천 방법을 제안한다. 본 논문에서 제안한 방법을 검증하기 위해 우리는 질의어에 관련된 10개 키워드의 정확도를 각각 분석하였다. 실험을 통해 본 논문에서 제안한 방법은 기존 소셜네트워크 서비스에서 제공한 타임라인 콘텐츠보다 높은 정확도를 얻게 됨을 확인하였다. 따라서 본 논문에서 제안한 방법을 활용하면 사용자에게 양질의 콘텐츠를 추천할 수 있다.

### 2. 배경 기술

소셜네트워크 서비스는 웹 2.0 기술을 기반으로 사용자들이 직접 참여하고, 사용자들로 하여금 프로필을 작성하며, 사용자들 간의 상호작용이 가능하게끔 하는 웹 기반의 서비스이다[1]. 또한 소셜네트워크 서비스는 사용자

가 검색 기능을 제공하여, 사용자들이 배포한 관련 콘텐츠들을 타임라인 형식으로 보여준다[3].

소셜네트워크 분석은 소셜네트워크에서 존재하는 서로 독립적인 노드(Node)들 간의 연결 관계를 이용하여, 노드들의 중심성(Centrality) 혹은 밀도(Density) 등 사회적 관계를 분석하는 방법이다. 소셜네트워크 분석 방법 중 하나인 매개 중심성 분석은 소셜네트워크를 구성하는 한 노드와 다른 한 노드를 연결시키는 특정 노드의 중개 정도를 측정하는 지표이다[4].

### 3. 매개 중심성 분석을 이용한 콘텐츠 추천

#### 3.1 매개 중심성 분석

본 논문에서는 소셜네트워크 분석 기법 중 하나인 매개 중심성 분석[4]을 이용한다. 특정 노드의 매개 중심성은 그 노드들을 제외한 다른 모든 노드들의 최단경로 개수와 해당 최단경로에서 특정 노드가 존재하는 개수의 비율을 나타낸다. 본 절에서 보이는 매개 중심성 분석은 질의어가 포함된 콘텐츠를 배포한 사용자가 소셜네트워크에서 차지하는 매개 중심성을 분석하는데 사용된다.

매개 중심성을 계산하는 과정에서 특정 사용자를 제외한 다른 노드의 쌍들의 최단경로 개수와 해당 최단경로에서 특정 사용자가 존재하는 개수의 비율은 (식 1)으로 표현된다.

$$b_{ij}(P_k) = \frac{g_{ij}(P_k)}{g_{ij}} \quad (\text{식 1})$$

(식 1)에서  $P_k$ 는 사용자이고,  $g_{ij}$ 는 노드  $i$ 와 노드  $j$ 를 연결하고 있는 최단경로 개수이며,  $g_{ij}(P_k)$ 는 노드  $i$ 와 노드  $j$ 사이의 최단경로 중에서 노드  $P_k$ 를 포함되는 최단경로의 개수이다. 그러므로  $b_{ij}(P_k)$ 는 특정 사용자  $P_k$ 를 제외한 노드  $i$ 와 노드  $j$ 사이의 최단 거리의 수와 해당 최단 경로에서 사용자가  $P_k$ 가 존재하는 최단경로 개수의 비율이다.

그리고  $P_k$ 라는 사용자가 소셜네트워크에서 차지하는 매개 중심성은 (식 2)에서 보이는 바와 같이  $C_B(P_k)$ 로 표현된다.

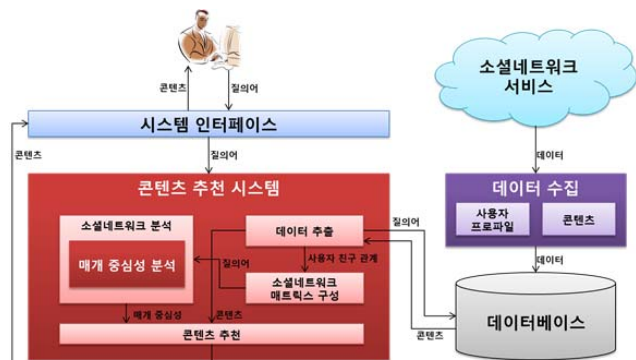
$$C_B(P_k) = \sum_{i \neq j \neq k}^n b_{ij}(P_k) \quad (\text{식 } 2)$$

위 (식 2)은 (식 1)를 이용하여 사용자  $P_k$ 를 제외한 전체 노드  $n$ 중에서 노드  $i$ 와 노드  $j$ 사이의 최단경로 개수와 해당 최단경로에서 사용자가 존재하는 개수 비율의 합이다. 그러므로 (식 2)를 통해 특정 사용자  $P_k$ 가 전체 소셜네트워크에서 중계자 역할을 얼마나 잘하는지를 측정할 수 있다.

본 논문에서 제안한 방법은 소셜네트워크에서 사용자들이 차지하는 매개 중심성을 분석한다. 그리고 본 논문은 계산된 사용자의 매개 중심성을 이용하여 기존 타임라인 콘텐츠들을 상위 순서로 다시 정렬한 후 사용자에게 추천한다. 이에 본 논문에서 제안한 방법을 활용하면 사용자들의 위상이 반영된 신뢰성이 있는 콘텐츠를 사용자에게 추천할 수 있다.

### 3.2 시스템 아키텍처

본 논문에서 제안한 매개 중심성 분석을 이용한 콘텐츠 추천 방법은 다음 (그림 1)과 같은 아키텍처를 통해 구현되었다. (그림 1)에서 시스템은 크게 데이터 수집 부분과 콘텐츠 추천 부분으로 구성된다. 데이터 수집 부분에서 데이터 수집 모듈은 소셜네트워크 서비스에서 제공된 데이터를 수집한 후 데이터베이스에 저장한다. 콘텐츠 추천 부분에서는 데이터 추출 모듈, 소셜네트워크 매트릭스 구성 모듈, 소셜네트워크 분석 모듈, 콘텐츠 추천 모듈로 구성된다. 이와 같은 모듈로 구성된 콘텐츠 추천 시스템은 질의어에 관련된 콘텐츠를 사용자에게 추천한다.



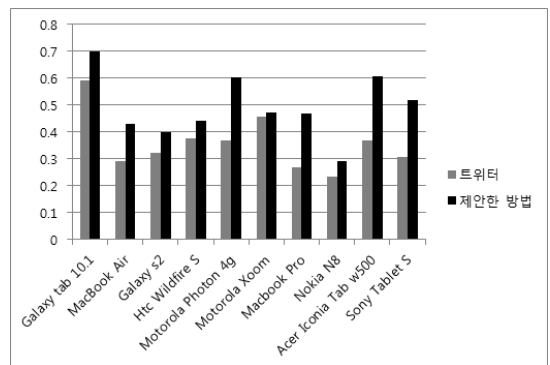
(그림 1) 시스템 아키텍처

### 4. 실험

본 논문에서 제안한 추천 방법을 검증하기 위하여 트위터에서 제공한 API를 통해 데이터를 수집하였다. 데이터 수집 시스템의 개발환경은 이클립스(Eclipse)과 마이크로소프트(Microsoft) SQL 서버를 이용하였으며 개발언어는 PHP 5를 이용하였다.

본 실험을 위하여 우리는 총 38784개의 노드를 트위터 API를 통해 수집한 후 데이터베이스에 저장하였다. 그리고 저장된 데이터 중에서 10개 키워드에 관련된 콘텐츠를 타임라인 형식으로 추출하였다. 추출한 각 키워드에 관련된 콘텐츠 개수는 25개이며, 10개 키워드에 대하여 총 250개 콘텐츠들이 추출되었다.

실험의 평가를 위해 정보 검색 분야의 대표적인 평가 도구인 정확도(Precision)를 사용하여 제안한 방법의 성능을 평가하였다.



(그림 2) 정확도

(그림 2)에서 보이는 바와 같이 본 논문에서 제안한 방법이 기존 트위터 검색 결과 보다 10개 키워드에 대하여 평균적으로 약 41% 높은 정확도를 보임을 알 수 있다.

### 5. 결론

본 논문에서는 매개 중심성 분석을 이용한 콘텐츠 추천 방법을 제안하였다. 실험을 통해 제안한 방법은 기존 소셜네트워크 서비스에서 제공한 타임라인 콘텐츠보다 높은 정확도를 가진다. 따라서 본 논문에서 제안한 방법을 활용하면 신뢰성이 있는 콘텐츠를 사용자에게 추천할 수 있다.

### 참고문헌

[1] Danah M. Boyd, Nicole B. Ellison, "Social Network Sites: Definition, History, and Scholarship", Journal of Computer-Mediated Communication, Vol.13, October 2007, pp.210-230  
 [2] 이성운, 이태휘, 김형주. "사용자의 활동과 영향력을 이용한 트위터의 URL 추천 시스템", 『한국정보과학회 논문지』 제17권 제8호, 2011, pp.447-456  
 [3] www.twitter.com  
 [4] Linton C. Freeman, "A Set of Measures of Centrality Based on Betweenness", Sociometry, Vol.40, No.1, Mar 1977, pp.35-41