

소셜 네트워크에서 관계 랭킹 모델

A Model for Ranking Semantic Associations in a Social Network

오선주(Sunju Oh)*

초 록

실생활에서 소셜 네트워크 서비스의 사용은 활성화되고 있으나 이를 비즈니스 차원에서 활용하기 위한 이론적이며 실증적인 연구가 부족한 상황이다. 기존의 다양한 데이터로부터 소셜 네트워크를 구축하고, 구축된 소셜 네트워크에서 잠재적 관계를 도출하거나 찾는 등의 유용한 활용 방법에 대한 연구가 요구된다. 본 연구는 소셜 네트워크에서 잠재되어 있는 관계를 인식하여 유용한 관계를 찾기 위한 방안으로서 소셜 네트워크에서 구성원간 관계를 검색하기 위한 랭킹 방법을 제안한다. 본 연구에서는 온톨로지를 기반으로 개체간 의미적 관계를 유추하여 확장하고 이를 바탕으로 다양한 랭킹 기준을 융통성 있게 조합하여 검색하고자 하는 관계를 효율적으로 찾기 위한 랭킹 모델을 제시하였다. 또한 제안한 연구 방법이 유의미한 것을 보이기 위하여 기업과 대학 간 사회적 네트워크에서 임의의 관계를 검색하고 강도를 측정하는 데 연구 모델을 적용하여 보았다. 본 연구에서 제안하는 시맨틱 웹기반 소셜 네트워크에서 임의의 관계를 검색하여 랭킹하는 방법은 빅데이터 시대에 유용한 관계 정보를 편리하게 검색할 수 있는 효과적인 방법으로 활용이 기대된다.

ABSTRACT

Much Interest has focused on social network services such as Facebook and Twitter. Previous research conducted on social network often emphasized the architecture of the social network that is the existence of path between any objects on network and the centrality of the object in the network. However, studies on the semantic association in the network are rare. Studies on searching semantic associations between entities are necessary for future business enhancements.

In this research, the ontology based social network analysis is performed. A new method to search and rank relation sequences that consist of several relations between entities is proposed. In addition, several heuristics to measure the strength of the relation sequences are proposed.

To evaluate the proposed method, an experiment was performed. A group of social relationships among the university and organizations are constructed. Some social connections are searched using the proposed ranking method. The proposed method is expected to be used to search the association among entities in ontology based knowledge base.

키워드 : 관계 검색, 랭킹 기준, 소셜 네트워크

Association Retrieval, Ranking Criteria, Social Network

연구는 경상대학교 경영경제연구소 지원에 의해 수행되었음.

* Gyeongsang National University, Division of Management Information System, Assistant Prof.

(E-mail : ohsunju@gnu.ac.kr)

2013년 04월 25일 접수, 2013년 07월 05일 심사완료 후 2013년 07월 25일 게재확정.

1. 서 론

21세기는 네트워크 시대라 일컬어지며 소셜 네트워크의 사용이 활성화되고 있다. 기업의 경쟁력은 기업이 보유한 자원의 가치에 의해 결정되며, 자원과 자원 간 잠재되어 있는 관계를 발굴하고 활용하는 것은 기업의 수익 증대를 위해서 중요하게 인식되고 있다. 이에 따라, 데이터마이닝, 패턴인식, 통계적 기법 등 의미 있는 정보를 발굴하여 사용하기 위한 관련 기술들이 발전하였으며 이러한 기술들을 적용하여 타 정보 주체와의 관계를 효과적으로 찾아 내어 활용하기 위한 많은 노력이 기울여져 왔다[1~5]. 그 예로 Travers and Milgram[13]의 소셜 네트워크 관련 실험에 의하면 소셜 네트워크 링크의 여섯 단계만 거치면 네트워크 상의 임의의 사람과 연결 가능하다는 연구 결과가 있다. 또한, Mark[8]는 매일 접하는 강인한 인적 관계 보다 가끔 연락을 주고 받는 비교적 약한 연관 관계의 영향력과 중요성(strength of weak ties)을 강조하였다.

기존 연구에서 시맨틱 웹 기반의 관계망에서 온톨로지 기반 관계 추론을 통하여 보다 많은 관계를 도출, 확장시키고 잠재되어 있는 관계까지 관계 망의 범위로 포괄할 수 있었다[6, 7]. 그러나 기존 연구를 통해 제시된 방법을 적용하여 도출, 확장된 관계들이 실제 소셜 네트워크 구조에서의 구성원간 관계의 특성과 의미를 잘 고려하지 못하는 문제점이 있다. 즉, 현실의 사회적 네트워크에서의 특정 구성원간 거리는 시맨틱 웹 기반 소셜 네트워크 이론에 근거한 거리와는 의미상 차이가 있는데, 이는 구성원간의 친근감, 정서적 유대감 등이 실제 대인 관계에 내재되어 있

으나 소셜 네트워크에서 상세하고 개별적으로 표현되기 어렵기 때문이다. 그러므로 명문화되지 않는 관계들을 좀 더 다양한 방법으로 발굴하여 표현할 필요성이 있으며 이를 바탕으로 관계 검색을 하고자 하는 사용자의 의도를 심도 있게 고려할 수 있는 랭킹 방법을 개발하는 것이 요구된다.

본 연구에서는 시맨틱 웹 기반 소셜 네트워크에서 구성 엔터티간 관계를 검색하는 랭킹 모델을 제시하고 임의의 구성원간의 관계를 찾아 검색 결과의 순위를 결정하는 랭킹 기준을 제시한다.

본 연구의 이론적, 실무적 기여 점은 다음과 같다. 첫째, 사용자의 검색 의도에 맞게 선택 적용할 수 있는 다양한 랭킹 기준을 제안하고 실 사례에 적용하였다. 즉, 연결성 검색(Connectivity retrieval)과 정보화 검색(Informative retrieval) 등의 두 가지 검색 방식에 따라 관계 경로 상의 노드의 중심성, 관계의 회귀성 등 다양한 랭킹 기준을 제시하며 랭킹 기준을 융통성 있게 선택하여 반영할 수 있도록 하였다.

둘째, 사용자 선호 컨텍스트를 검색 랭킹에 반영하였다. 즉, 교육, 취업, 산학 등 관계 경로가 위치하는 특정 컨텍스트에 대한 사용자의 선호도를 가중치로 반영하여 사용자가 의도하는 컨텍스트에서 적합한 경로를 파악할 수 있도록 하였다.

셋째, 관계 확장 방법을 적용함으로써 현실 세계와 유사한 소셜 네트워크를 표현하고 검색 대상으로 포함하고자 하였다. 이를 위하여 우선, 규칙 추론 기능을 적용하여 기존의 관계를 확장하고 이를 네트워크의 관계 검색 대상에 반영하였다.

본 연구의 구성은 다음과 같다. 제 2장은 소셜 네트워크에서 관계 검색과 랭킹에 관련된 기존 연구들을 살펴본다. 제 3장에서는 소셜 네트워크를 구성하는 개별 인스턴스간의 관계를 찾는 연구 모델을 제시한다. 제 4장에서는 제시한 연구 모델을 평가하는 실험을 수행한다. 마지막 제 5장은 연구의 결론을 맺는다.

2. 관련 연구

웹에서의 정보 검색은 키워드 기반의 자원 검색이 주류를 이루어왔으나 점차 의미적인 측면을 고려한 시맨틱 검색에 대한 연구가 활발히 진행되어 왔다. 그 중에서도 기존의 자원 중심 검색 방식에서 더 나아가 자원 자체보다 자원 사이의 관계의 검색에 대한 관심이 높아지고 있다. 특히, 시맨틱 웹과 소셜 네트워크를 기반으로 한 네트워크에서 표현되지 않은 잠재적인 관계를 발굴하여 활용하기 위한 다양한 연구들이 이루어져 왔다[1~5, 8~11, 14].

Anyanwu et al.[4]는 엔터티들을 서로 연결시켜 주는 의미적 관계(Semantic association) 개념을 기반으로 시맨틱 웹에서의 엔터티간 관계를 찾고 랭킹하는 방법을 제시한다 [3, 4, 5]. 랭킹 기준으로 정보량(information content), 비예측성(unpredictability), 일치성(the extent of match), 사회성(popularity of nodes)을 제안하고 사용자의 의도에 기반한 융통성을 강조하였다. 그들은 일상적인 검색보다 비일상적인 검색 경로가 정보로서 가치가 높은 경우가 많다고 주장하고 비일상적이고 예측하기 어려운 검색 경로의 랭킹 순위

를 높이는 정책을 기술하였다.

Alemen-Meza et al.[1]은 사용자의 컨텍스트 정보를 반영한 랭킹 알고리즘을 제시하였다. 이들은 주요 랭킹 기준으로 시맨틱 메트릭(Metric)과 통계적 메트릭을 기술하였다 [1, 2]. 시맨틱 메트릭으로는 사용자 컨텍스트, subsumption, 신뢰 등을 제시하였고 통계적 메트릭으로는 인스턴스간 관계 전체에 대한 특정 관계의 비율, 다른 엔터티들과 맺고 있는 관계의 정도 즉, 사회성(popularity), 관계의 길이 등을 제시하였다. 이들 메트릭을 이용하여 사용자의 컨텍스트에 맞게 가중치를 부여하는 융통성 있는 관계 측정 모델을 제시하였으며 테스트베드 SWETO[11] 온톨로지에 적용하였다.

한편, Miki et al.[9]는 학술 커뮤니티에서 사회적 관계를 찾아내는 방법을 제시하였다. 웹 페이지에서의 저자 상호 참조 관계와 같이 직접 링크만을 고려하여 네트워크를 생성하였다. 그리고 일반적인 사회적 관계 네트워크를 생성하고 두 가지 네트워크를 통합하였다.

Thusar et al.[12]은 일상적으로 빈번하게 발생하는 일상적 관계(common associations)와 일반적으로 자주 발생하지 않으며 정보로서 가치가 높은 정보화 관계(informative associations) 등 두 종류의 의미 관계(semantic association)를 정의하였다. 또한, 의미 관계를 RDF 형식의 관계 경로로 표현하고 선호하는 검색 방식에 따라 검색 결과가 랭킹되도록 하였다. 검색 랭킹 기준으로 관계 빈도에 기반을 둔 정보량, 다양한 도메인의 경유 가능성, 특정 키워드 혹은 관계와의 매칭 정확도, 관계 경로상의 노드의 사회적 인지도 등을 정의하였다.

Zhang et al.[14]는 시맨틱 링크 네트워크에서 링크 타입을 추론적 관계와 통계적 관계로 구분하고 각각의 관계의 강도를 측정하는 방법을 제시하였다[14]. 추론 관계는 추론 관계 그래프에서 조건부 확률을 이용하여 관계 강도를 계산하였으며 통계적 관계 측정을 위해서는 시맨틱 링크 네트워크에서 발생 빈도가 높은 관계를 찾아내기 위한 알고리즘을 제시하였다.

이와 같이 시맨틱 웹 환경에서 웹 자원간 관계의 추출 및 검색에 관하여 활발한 연구가 이루어져 왔다. 기존 연구에서 시맨틱 웹 스키마 수준에서의 추론을 통하여 관계를 확장하고 확장된 관계 정보를 대상으로 검색을 하였다[16, 17]. 그런데 근래의 소셜 네트워크 연구에서는 구성원간의 연결성 평가와 효율적 연계 경로의 발견에 대한 많은 관심이 모아지고 있으며 따라서, 소셜 네트워크의 개별 인스턴스 수준에서의 관계 경로를 발견하기 위한 효율적인 관계 검색 기법이 요구된다. 본 연구에서는 인스턴스 수준에서의 관계 경로를 랭킹하는 기준과 방법을 제시한다.

3. 소셜 관계 경로 랭킹 모델

본 연구에서는 온톨로지 기반 소셜 네트워크에서 임의의 소셜 링크를 찾는 방법을 제시한다. 이를 위하여 우선, 소셜 네트워크를 구성하는 관계 혹은 규칙을 바탕으로 온톨로지 기반 추론을 적용하여 다수의 관계들을 검색 대상으로 포함시킬 수 있도록 관계를 확장한다. 또한, 검색하고자 하는 관계 경로가 여러 도메인 컨텍스트에 속하는 경우 각

각의 컨텍스트를 반영하는 최단 경로를 적용한다. 관계 경로 검색을 효율화하기 위해서는 실세계의 풍부한 관계 데이터 정보를 온톨로지에 표현하는 것이 필요하다. 수집한 관계 정보를 바탕으로 온톨로지를 구성하는 주요 개념의 인스턴스간 연결성 유무를 발견하고, 효율적인 관계 경로를 검색하기 위한 랭킹 모델을 정의한다. 본 연구에서 제안하는 랭킹 모델은 랭킹 기준과 랭킹 함수로 구성된다. 랭킹 기준으로 관계 매칭의 정확도와 구체성, 관계의 거리, 사회적 인지도, 관계의 빈도, 컨텍스트 이질성, 컨텍스트 선호도 등을 제안한다. 랭킹 함수는 다양한 랭킹 기준의 조합으로 구성되는데, 사용자의 요구 사항을 반영하여 융통성 있게 랭킹 기준을 선택, 조합하여 구성할 수 있다.

관계 경로(relation sequence : rs) : 네트워크 상의 임의의 두 노드 a, b 를 연결하는 인접한 관계들의 집합이다. 관계 경로 $rs = [r_1, r_2, \dots, r_n]$ 는 두 인스턴스 a, b 간 n 개의 서로 연결된 관계들로 구성된다. 관계 경로 rs 를 구성하는 각각의 관계 r_i 는 $r_i[1] = r_{i+1}[0]$, $r_1[0] = a$, $r_n[1] = b$, $1 \leq i \leq n$ 를 만족한다. 즉, $r_1[0] = a$, $r_1[1] = r_2[0]$, $r_2[1] = r_3[0]$, \dots , $r_n[1] = b$ 등 a 에서 b 까지의 서로 연결된 관계들이 관계 경로를 구성한다.

3.1 랭킹 기준

본 장에서는 소셜 네트워크에서 두 구성원간 임의의 관계 경로를 검색하기 위한 랭킹 기준을 제시한다. 제시하는 랭킹 기준은 관계 매칭 이론에 근거한 기준과 사회적 네트워크

의 구조적 특성을 반영하기 위한 기준들로 구성된다.

매칭 이론(Matching theory)에 근거한 기준은 관계를 구성하는 각 엔터티 혹은 인스턴스의 일치하는 정도에 비례하는 반면 관계 거리에 반비례하는 특성이 있다[14]. 매칭 이론에 근거한 기준으로는 기존 연구를 바탕으로 인스턴스 간 관계 매칭의 정확도와 구체성과 관계의 거리 등을 정의하였다. 이들은 사용자가 제시한 키워드와의 일치성이 높고 보다 구체적이며 강력한 관계를 찾기 위한 기준이다.

한편, 사회적 네트워크의 구조적 특성(Social network characteristics)이란 사회적 네트워크 특성을 반영하기 위한 기준으로 관계를 구성하는 엔터티 혹은 인스턴스의 사회적, 중심성, 관계 빈도 등으로 관계의 특성을 표현한다[15]. 관련 랭킹 기준으로는 검색 대상 인스턴스의 사회적 인지도, 관계 빈도, 컨텍스트 이질성, 컨텍스트 선호도 등을 정의하였다. 이들은 네트워크의 경로를 구성하는 구성원의 영향력이 높은 경로를 찾거나 혹은 이질적인 컨텍스트를 경유하는 드물게 발생하는 경로를 찾기 위해 사용자의 컨텍스트에 대한 선호도를 반영하기 위한 기준이다.

관계 매칭의 정확도와 구체성(specificity : sp)

관계 매칭의 정확도와 구체성은 관계를 구성하는 두 인스턴스와 인스턴스 간 경로가 사용자가 제시한 키워드와 매칭되는 정확도와 구체성을 의미한다. 또한, 동일한 인스턴스 간 관계라 하더라도 인스턴스에 대한 일반적인 정보보다 구체화된 정보가 제시되는 것이 정보량 측면에서 효율적인 측면이 있다. 구체성

기준은 매칭의 정확도를 반영하기 위한 기준으로 동일한 인스턴스 간 관계라도 보다 많은 정보를 포함한 인스턴스 간 관계가 정보로서 중요성이 높다고 간주한다. 즉, 온톨로지 계층 구조상 하위에 위치한 엔터티는 상위에 위치한 엔터티 보다 구체적이며 내포된 정보량이 많다. 그러므로 하위에 위치하는 엔터티 간 관계인 경우 구체성이 높다. 예를 들면 인간 홍길동과의 관계보다는 A 대학 정보학과 재학생인 홍길동과의 관계가 구체성이 높다.

$$sp(i) = |Hi|/|H|$$

|Hi| : 계층 구조상 i 노드의 위치

|H| : 클래스 계층구조 전체의 깊이

관계의 거리(distance)

소셜 네트워크에서 노드 간 관계 거리는 검색 대상이 되는 두 노드 사이를 연결하는 링크 수를 의미한다. 가중치가 부여된 네트워크 일 경우, 사용자가 부여한 가중치를 관계 거리에 반영한다.

$$distance(rs) = |c|,$$

c : number of components on the path rs

사회적 인지도(popularity)

인스턴스의 사회적 인지도는 임의의 인스턴스가 다른 인스턴스들과 가지는 관계의 수, 강도 등으로 정의된다. 다른 인스턴스와 다양한 관계를 유지하는 인스턴스는 지속적으로 활발한 사회 활동을 하며 사회적 영향력이 크다. 그러므로 관계 경로를 구성하는 노드들에 대한 사회적 인지도는 사회적 파급 능력 등을 전체적으로 고려하여 관계 검색의 랭킹

기준으로 사용한다.

$$\text{popularity}(rs) = \frac{\sum_{i=1}^n \text{degree of node } i \text{ on the path } rs}{\text{number of node on the path } rs}$$

관계의 빈도(frequency)

관계의 빈도란 관계가 성립하는 빈도를 나타낸다. 두 인스턴스를 연결하는 관계 경로가 여러 가지 경로가 있는 경우는 관계 빈도가 높으나 목적지에 도달하기 위한 유일한 경로만 있는 경우는 관계 빈도가 낮다. 관계 빈도가 높은 경로는 관계 경로를 통해 전달되는 정보의 양이 많으므로 연결성 검색에서 순위가 높다.

$$\text{frequency}(r) = \frac{\text{occurrence of property } r}{\text{number of all properties}}$$

$$\text{frequency}(rs) = \max_{i=1, n} \text{frequency}(r_i)$$

컨텍스트 이질성(refraction : RF)

컨텍스트 이질성은 찾고자 하는 관계 경로가 단일 컨텍스트에 포함되는 관계들로만 구성된 관계 경로인지 혹은 서로 다른 여러 컨텍스트에 속한 관계의 집합인지 여부를 나타낸다. 정보화 검색시 이질적인 컨텍스트를 경유하며 다양한 컨텍스트에 속하는 경로는 빈번하게 발생하지 않으며 정보로서의 가치가 높다고 간주된다.

$$\begin{aligned} \text{Context}(r) &: \text{context for relation } r, \\ \text{refraction}(r_i, r_{i+1}) & \\ &= 0, \text{ if } \text{Context}(r_i) = \text{Context}(r_{i+1}), \\ &= 1, \text{ otherwise} \end{aligned}$$

$$\begin{aligned} \text{RF}(rs) &= \sum_{i=1}^{n-1} \text{refraction}(r_i, r_{i+1}) \text{ for } n \geq 2 \\ &= 0, \text{ otherwise} \end{aligned}$$

컨텍스트 선호도(Context Weight : CW)

특정 컨텍스트에 대한 사용자의 선호도를 고려하기 위하여 해당 컨텍스트를 경유하는 경로에 가중치를 부여하기 위한 기준이다.

$$\text{CW}(rs) = \frac{\sum_{i=1}^n \text{CW}(r_i)}{\text{number of node on the path } rs}$$

$$\begin{aligned} \text{CW}(r_i) &= 1, \text{ if relation } r_i \in \text{preferred context,} \\ &= 0, \text{ otherwise.} \end{aligned}$$

3.2 랭킹 함수

랭킹 함수는 랭킹 기준을 바탕으로 각각의 관계 경로에 랭킹 값을 계산하는 수식이다. 계산된 함수값을 관계 검색 랭킹 결과에 반영한다. 본 연구에서는 두 가지 종류의 관계 경로 검색에 랭킹 함수를 적용한다. 연결성(Connectivity) 경로 검색과 정보화(Informative) 경로 검색이다. 연결성 검색은 다수의 연결 경로 중 강력하고 효율적인 경로를 찾기 위한 검색 방법이며 정보화 검색은 두 인스턴스간 경로 중 자주 발생하지 않으나 중요도가 높은 경로를 발견하기 위한 방법이다. 그러므로 두 가지 서로 다른 검색 방식에서 적용되는 랭킹 기준도 상이하다(<Table 1> 참조). 즉, 연결성 검색에서는 관계 빈도, 인스턴스의 사회적 인지도, 매칭 정확도와 구체성, 컨텍스트 선호도 등의 기준이 랭킹과 정비례의 관계인 반면, 관계 거리, 컨텍스트 이질성 기준은 반비례한다. 한편, 정보화 검색에서는 인스턴스의 매칭 정확도, 컨텍스트 이질성, 컨텍스트

$$\text{Rank for connectivity}(rs) = \frac{\text{popularity}(rs)}{\text{distance}(rs)} \times \text{frequency}(rs) \times (1 + \text{CW}(rs)) \times 1/(1+\text{RF}(rs)) \quad (1)$$

$$\text{Rank for informative}(rs) = \frac{(1 - \text{popularity}(rs))}{\text{distance}(rs)} \times (1 - \text{frequency}(rs)) \times (1 + \text{CW}(rs)) \times (\text{RF}(rs) + 1) \quad (2)$$

$$\text{SRank for connectivity}(rs) = \frac{\text{popularity}(rs)}{\text{distance}(rs)} \times \text{frequency}(rs) \times (1 + \text{CW}(rs)) \times 1/(1+\text{RF}(rs)) \times \text{sp}(rs) \quad (3)$$

$$\text{SRank for informative}(rs) = \frac{(1 - \text{popularity}(rs))}{\text{distance}(rs)} \times (1 - \text{frequency}(rs)) \times (1 + \text{CW}(rs)) \times (\text{RF}(rs) + 1) \times \text{sp}(rs) \quad (4)$$

선호도 등이 정비례 관계이며 관계의 거리, 관계의 빈도, 사회적 인지도 등이 반비례 관계이다.

즉, 사용자의 검색 목적에 따라 사회적 네트워크의 랭킹 기준은 상이하게 적용된다. 관계의 밀도가 높은 다중 경로 망에서의 검색에서 목적 노드를 연결하는 강력한 관계를 찾기 위해서는 관계 빈도, 인지도 등을 고려하며, 반면 일반적으로 거의 발생하지 않는 연관 관계를 찾기 위해서는 관계 강도보다는 목적 노드까지의 도달 가능성을 점검하고 다양한 관계 경로 중 컨텍스트 이질성이 높은 경로를 높은 우선순위에 둔다.

소셜 네트워크에서 원하는 관계를 찾기 위해서는 <Table 1>에 정의된 랭킹 기준을 융통성 있게 자유롭게 조합하여 랭킹 함수를 구성하고 적용한다. 즉, 랭킹 함수는 위에서

정의한 랭킹 기준을 종합적으로 반영하여 다음과 같이 정의한다.

본 연구에서는 관계의 거리, 사회적 인지도, 관계의 빈도, 컨텍스트 이질성을 고려한 (1), (2) 함수를 적용한다. (1)은 노드 간 연결성을 탐색하기 위한 랭킹 함수이며 (2)는 관계 경로를 통해 전달되는 정보량 크기를 기준으로 하는 랭킹 함수이다. 그 외에 구체성 기준을 반영한 (3), (4) 함수를 다양하게 구성할 수 있다. 연구 모델의 검증을 위해 (1), (2) 랭킹 함수를 적용하여 실험 및 분석을 수행하였다.

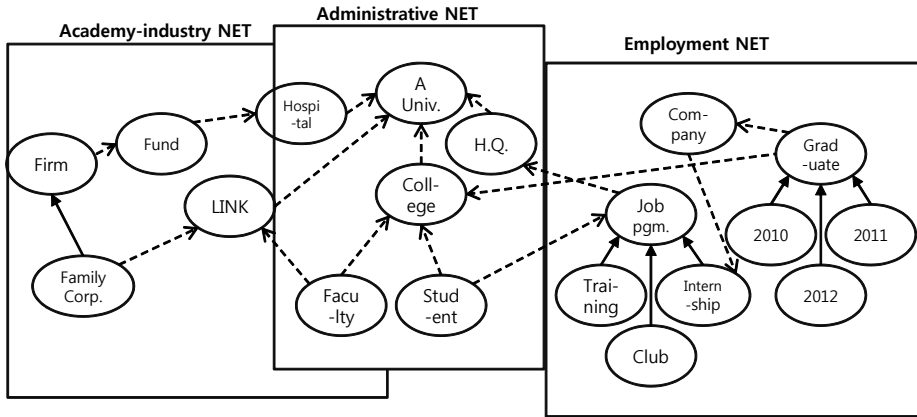
4. 실험 및 분석

기업과 대학간의 소셜 네트워크에서 사용 목적에 적합한 관계 경로를 찾기 위하여 검색 랭킹 모델을 대학의 소셜 네트워크 구축에 적용하였다. A 대학교 경영대학과 협력 기업 간의 산학 네트워크 구축 현황을 조사하고 협력 관계를 증진시키기 위해 2010년부터 2012년까지의 산학 협력 현황, 공동 프로젝트 현황, 졸업생 취업 연계 현황, 기업 임직원 교육 현황 등의 다양한 컨텍스트 데이터를 반영하여 기업과 학계의 협력 관계를 조사하였다. 조사 결과로 700여 개 가족 기업

<Table 1> Ranking Criteria

	Ranking criteria	Connectivity retrieval*	Informative retrieval*
Matching theory	specificity	+	+
	distance	-	-
Social network characteristic	popularity	+	-
	frequency	+	-
	refraction	-	+
	context weight	+	+

* : +Direct proportion, - Inverse proportion.



<Figure 1> Social Network

중 교육 참여, 공동 연구, 산학 협력 등 비교적 상관도가 높다고 판단되는 기업에 대한 관련 데이터를 수집하였다. 이를 바탕으로 산학협력, 행정, 졸업생 취업 컨텍스트에서 <Figure 1>과 같은 소셜 네트워크를 구성하였다.

또한, 업무 관련성을 기반으로 아래와 같은 규칙을 정의하였다. 조사된 네트워크에 아래와 같이 정의된 규칙을 적용하여 기존의 관계를 확장하였으며 기존 네트워크에서 새로운 관계 경로가 도출되었다.

규칙 1 : 기업 ^ 공동 프로젝트의 추진 ^ 기업의 규모 → 가족 기업

규칙 2 : 재학생 ^ 경영대학 ^ 산학 협력 ^ 협력 프로그램 → 취업 준비

또한 산학, 행정, 취업 컨텍스트에서 각각의 인스턴스간 관계를 고려하여 종합적인 산학 통합 네트워크를 구성하였다(<Figure 1> 참조). 기존 단위 망인 산학(Academy-industry), 행정(Administrative), 취업(Employment) 망

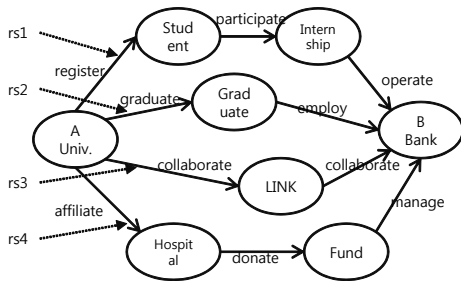
을 비교 통합하여 통합 망을 구성함으로써 기존에 없었던 관계 경로가 새로이 생성되거나 기존의 경로가 단축되는 효과가 발생했다.

산학 통합 네트워크에서 대학과 기업간의 관계를 검색하기 위해 연구 모델을 적용하였다. 즉, 관계 경로 검색 방식으로 경로의 연결성을 검색하는 연결성 검색과 특이한 경로를 검색하는 정보화 검색 등, 두 가지 검색 방식에 제 3장의 연구 랭킹 모델 (1)과 (2)를 적용하였다. 관계 검색에서 목적 노드에 이르기까지 경유하는 인스턴스의 사회적 중심성과 관계 빈도, 사용자의 선호도를 고려하여 순위를 매기고 검색 결과를 비교하였다. 경유 인스턴스의 사회적 중심성을 반영한 실험에서 A 대학과 가족 기업 B 은행을 연결하는 다수의 경로 중 주요 경로인 경로 1, 경로 2, 경로 3, 경로 4는 다음과 같다(<Figure 2> 참조).

경로 1 (rs1) : A 대학 - 등록 - 재학생 - 참가하다 - 인턴프로그램 - 참여하다 - B 은행

경로 2 (rs2) : A 대학 - 졸업하다 - 졸업생

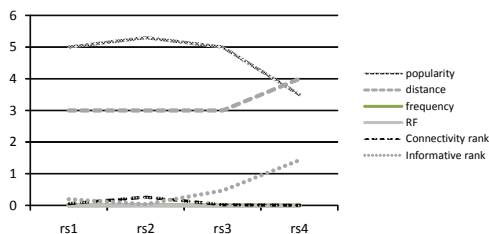
- 취업하다 - B 은행
- 경로 3 (rs3) : A 대학 - 관련되다 - 산학협력 (LINK) - 협력하다 - B 은행
- 경로 4 (rs4) : A 대학 - 부속기관 - A 대학 병원 - 기반하다 - 기금 - 기탁하다 - B 은행



〈Figure 2〉 Paths between A Univ. and B Bank

경로 1은 중심성이 높은 엔터티 재학생을 경유하는 경로이다. 경로 4는 희귀한 경로인 A 대학 - A 대학병원 - 기금 - B 은행 인스턴스를 경유하는 경로이다. 경로 4는 비교적 사회적 인지도가 낮게 나타났으며 서로 다른 컨텍스트를 경유하는 드물게 발생하는 경로이다. 이들 경로를 비교한 결과, 경로 2에서의 인스턴스의 사회적 인지도가 가장 높게 나타났다(〈Figure 3〉 참조).

제 3장에서 제안한 랭킹 함수 (1), (2)를 적



〈Figure 3〉 Rank value of rs1, rs2, rs3, and rs4

용한 결과, 연결성 검색에서 각 경로의 랭킹 값은 〈Table 2〉와 같이 나타났으며 랭킹 순위는 경로 2, 경로 1, 경로 3, 경로 4의 순서로 나타났다.

〈Table 2〉 Ranking

	Connectivity retrieval rank	Ranking	Informative retrieval rank	Ranking
rs1	0.003	2	40.000	3
rs2	0.013	1	8.386	4
rs3	0.002	3	46.667	2
rs4	0.000	4	142.857	1

또한, 정보화 검색에서는 경로 4, 경로 3, 경로 1, 경로 2의 순으로 나타났다(〈Table 2〉 참조). 관계 빈도가 높은 경로 1과 중심성이 높은 경로 2가 연결성 검색에서 상위에 랭크 되었으며 정보화 검색에서는 드물게 발생하는 경로 3과 경로 4가 상위 랭크에 올랐다.

이와 같은 랭킹 결과가 나오게 된 이유를 분석해 보면, 경로 1은 인스턴스의 수가 많은 재학생 노드와 인턴십 노드를 포함하고 있다. 따라서, 경로 1의 관계의 빈도가 상대적으로 높고 사회적 인지도 측면에서도 노드의 중심성이 높아 랭킹 순위를 높이는 역할을 한다. 경로 2는 A 대학, 졸업생과 취업사인 B 은행으로 구성되는데, 졸업생 노드의 관계의 빈도와 중심성이 높아 랭킹 순위를 높이는 역할을 한다. 또한 경로 2의 관계 거리가 3이므로 관계 거리가 4인 경로 1보다 랭킹 순위를 더욱 높이는 효과가 있다. 따라서 경로 2가 랭킹 순위가 가장 높은 결과가 나오는데, 이는 A 대학과 B 은행을 연결하는 경로를 검색할 때 경로 2가 가장 적합한 결과로 인식되도록 한다. 또한, 경로 3은

중심성이 높은 산학 협력단, LINK 노드를 포함하고 있으나 B 은행에 이르는 관계 빈도가 낮아 정보화 검색에서 상위권 순위를 나타내었다. 경로 4는 병원과 기금 등 관계 빈도가 높지 않은 경로로서 연결성 검색 순위에서는 낮은 순위를 나타나게 한다.

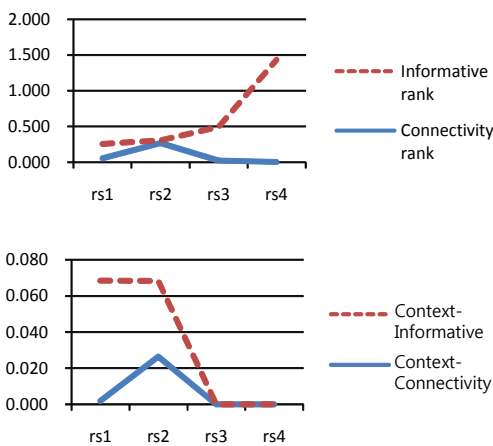
또한, 사용자 컨텍스트 선호도가 랭킹 순위에 미치는 영향을 분석해 보았다. <Table 2>의 실험 결과에 사용자 선호 컨텍스트의 가중치를 부여하여 원하는 컨텍스트를 경유하는 경로를 찾아 보았다. 예를 들면, 사용자가 특

별히 취업에 관심이 있다면 여러 가지 컨텍스트 중에서 취업 컨텍스트에 가중치를 부여하여 경로를 찾는 방식을 적용할 수 있다.

또한, 가중치를 고려하는 방법과 고려하지 않는 두 가지 방법을 비교함으로써 컨텍스트 가중치 부여를 통한 개인의 컨텍스트에 대한 선호도를 랭킹에 반영할 수 있다. 실험에서 A 대학교와 B 은행 간의 관계 경로를 검색 시 검색 사용자가 취업 컨텍스트에 대한 관심이 높다고 가정하고 취업 컨텍스트의 선호도를 높게 반영하였다. 또한, 사용자의 검색 의도에 맞는 검색 결과를 찾을 수 있도록 컨텍스트 선호도 랭킹 기준을 적용하였다. 즉, 선호하는 컨텍스트인 취업 망에 속하는 관계에 가중치를 높게 부여하여 선호하는 컨텍스트를 경유하는 경로의 랭킹 순위가 조절되었다(<Table 3> <Figure 4> 참조). 결과적으로, 정보화 검색에서 취업 컨텍스트를 경유하는 경로 1의 순위가 높아졌다.

<Table 3> Weighed Ranking

	Connectivity retrieval rank	Ranking	Informative retrieval rank	Ranking
rs1	0.004	2	60.000	2
rs2	0.022	1	13.976	4
rs3	0.002	3	46.667	3
rs4	0.000	4	142.857	1



<Figure 4> Ranking(context weight considered(upper) and ignored(lower))

5 결론

여러 가지 형태의 사회적 네트워크에서 네트워크를 구성하는 임의의 두 인스턴스간의 의미 있는 관계가 있는지 찾아보고 관련성이 어느 정도인지 랭킹을 매기는 작업은 이론적 연구와 함께 실무적 적용이 필요한 분야이다. 본 연구에서는 소셜 네트워크에서 구성원간 관계를 효과적으로 찾기 위한 랭킹 기준과 랭킹 함수를 제시하였다. 매칭 이론과 사회적 네트워크 특성을 고려하여 랭킹 기준과 방법을 제시하였다. 또한, 제안한 랭킹 방법을 검증하기 위해 교육 및 산학 협력 도메인에 적

용하여 대학과 기업 간의 임의의 관계를 찾아보았다. 적용 결과, 연결성 검색에서 인스턴스들간 인지도 등 랭킹 기준이 유리하게 작용한 경로가 높은 순위에 올랐으며 정보화 검색에서는 드물게 발생하는 관계를 포함하는 경로가 높은 순위에 올랐다.

소셜 네트워크에서 구성원간의 관계의 특성은 일반적인 네트워크 그래프에서의 노드간 관계와는 차이가 있다. 즉, 사회 구성원간의 친밀성, 사용자 선호도 등 관련 특성들은 네트워크의 구조만으로 표현하기 어렵다. 이와 같은 차이를 인식하고 소셜 네트워크에서 반영할 수 있는 방안으로서 본 모델을 제안하였다. 특히, 사회적 네트워크에서 임의의 구성원 사이의 관계 경로가 성립하는지 찾고 어떤 경로로 서로 효과적으로 연결되는지 파악하여 활용하는 작업은 이론적 연구로서의 의의와 함께 실무적인 측면에서 매우 유용하다 할 수 있다. 또한, 사용자가 선호하는 특정 도메인 컨텍스트에서 찾기를 원하는 관계 경로를 검색하는 등 개인적인 측면에서 보다 효과적으로 활용될 수 있을 것이다.

본 연구의 실험 및 분석에 사용한 대학, 기업간 소셜 네트워크는 실제 규모 면에서는 방대함에도 불구하고 실험에 적용된 소셜 네트워크 경로 사례 데이터가 양적 질적으로 부족하게 보일 수 있다. 그러나 실험에서 랭킹 예로 사용된 A 대학과 B 은행 간의 경로는 다양한 경로 중 관계 거리가 짧고 랭킹 상위에 오르는 주요 경로를 선별하여 적용하였으며, 실제 경로 상에 위치한 노드들은 다수의 인스턴스를 포함하며 관계를 구성하고 있어 데이터의 규모 면에서 실험의 신뢰성과 타당성을 저해하지 않는다고 볼 수 있다. 그

러나 향후 연구에서 여러 가지 형태의 관계 경로에 대하여 모델을 적용하는 일반화가 필요하다. 또한, 연구 모델을 적용한 사례를 다양하게 개발하고 여러 가지 분야에 본 연구 방법을 적용함으로써 연구 모델의 일반성을 검증하고 랭킹 기준과 랭킹 함수를 개선하는 것이 요구된다.

References

- [1] Aleman-Meza, B., Halaschek-Wiener, C., Budak Arpinar, I., Ramakrishnan, C., and Sheth, A., "Ranking complex relationships on the semantic web," In Proc. of WWW2005, 2005.
- [2] Aleman-Meza, B., Nagajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A., and Arpinar, I., "Semantic analytics on social networks : Experiences in addressing the problem of conflict of interest detection," In Proc. of WWW2006, 2006.
- [3] Anyanwu, K. and Sheth, A., " ρ -Queries : Enabling querying for semantic associations on the Semantic Web," In Proc. of WWW2003, pp. 690-699, 2003.
- [4] Anyanwu, K. and Sheth, A., "The ρ operator : Discovering and Ranking associations on the semantic web," SIGMOD, 2002.
- [5] Anyanwu, K., Maduko, A., and Sheth, A., "SemRank : Ranking Complex relationship search results on the Semantic

- Web,” In Proc. of WWW, 2005.
- [6] Barnaghi, P. M. and Kareen, S. A., “Relation robustness evaluation for the semantic associations,” *The electronic Journal of Knowledge Management* Vol. 5, No. 3, pp. 265-272, 2007.
- [7] Faloutsos, C., McCurley, K. S., and Tomkins, A., “Fast Discovery of Connection Subgraphs,” In Proc. of the Tenth ACM SIGKDD Conference, 2004.
- [8] Lee, W., “A framework for discovering meaningful associations in the annotated life sciences web,” Doctoral dissertation, University of Maryland, 2008.
- [9] Mark, S. G., “The Strength of Weak Ties,” *American Journal of Sociology*, Vol. 78, No. 6, pp. 1360-1380, 1973.
- [10] Miki, T., Nomura, S., and Ishida, T., “Semantic Web link analysis to discover social relationships in academic communities,” In Proc. of SAINT’05, 2005.
- [11] Oh, S., Ahn, J., and Park, J., “Ontology Selection Ranking Model based on Semantic Similarity Approach,” *Society for e-Business Studies*, Vol. 14, No. 2, pp. 95-116, 2009.
- [12] Sicilia, M., Kop, C., and Sartori, F., “Ontology, Conceptualization, and Epistemology for Information Systems, Software Engineering and Service Science,” 4th International Workshop ONTOSE2010, 2010.
- [13] Thushar, A. K. and Thilagam, P. S., “An approach for discovering the relevant semantic associations in a social network,” In Proc. of ADCOM2008, pp. 214-220, 2008.
- [14] Travers, T. and Milgram, S., “An Experimental Study of the Small World Problem,” *Sociometry*, Vol. 32, No. 4, p. 425, 1969.
- [15] Turney, P. D., “Measuring Semantic Similarity by Latent Relational Analysis,” In Proc of the Nineteenth International Joint on Artificial Intelligence(IJCAI-05), pp. 1136-1141, Edinburgh, Scotland, 2005.
- [16] Wasserman, S. and Faust, K., “*Social Network Analysis*,” Cambridge, 1994.
- [17] Zhang, J., Wang, H., and Sun, Y., “Discovering associations among semantic links, International conference on Web information systems and mining,” p. 204, 2009.
- [18] Zhuge, H. and Zheng P., “Ranking Semantic-linked Network,” In Proc. of WWW2003, 2003.

저 자 소개



오선주

1986년

1993년

2009년

2009년

2012년~현재

관심분야

(E-mail : ohsunju7@gnu.ac.kr)

서울대학교 계산통계학과 (학사)

서울대학교 계산통계학과 (이학 석사)

서울대학교 경영학과 MIS전공 (경영학 박사)

서울대학교 컴퓨터공학과 연구교수

경상대학교 경영정보학과 조교수

온톨로지, e-비즈니스, 정보기술 수용, 모바일 비즈니스