

논문 2012-49-9-10

퍼지 연관 곱을 이용한 SNS 사용자의 계층적 시각화 방법

(Hierarchy Visualization method of SNS User using Fuzzy Relational)

박 선*, 권 장 우**, 정 민 아***, 이 연 우****, 이 성 로*****

(Sun Park, JangWoo Kwon, Min-A Jeong, Yeonwoo Lee, and Seong Ro Lee)

요 약

시각화는 소셜 네트워크의 분석을 위해서 소셜 네트워크의 사용자 관계를 새로운 시각에서 이해할 수 있도록 하는 중요한 역할을 담당하고 있다. 기존의 대부분의 시각화 방법은 복잡한 다차원 그래프를 통하여 소셜 네트워크상의 사용자의 관계를 집중적으로 표현하고 있다. 그러나 이러한 방법은 개인 사용자 중심으로 사회관계의 중요도를 직관적으로 파악하기 힘들다. 또한 대부분의 시각화 방법은 네트워크상의 노드들 간의 상호작용 정보에 의해서만 사용자 관계를 나타내기 때문에 사용자의 메시지가 사용자 상호관계에 반영되는 것이 미흡한 실정이다. 제안방법은 퍼지 연관 곱을 이용하여 사용자의 내부관계를 계산하며, 노드들 간의 상호작용 정보를 이용하여서 사용자의 네트워크상의 외부 접근 정보를 계산한다. 계산된 내부관계정보와 외부 접근 정보를 이용하여서 사용자중심의 계층적 시각화방법을 제안한다. 이러한 문제를 해결하기 위해서 본 논문은 퍼지 연관 곱에 의한 사용자들의 내부관계와 네트워크상의 외부 접근 정보를 이용하여서 사용자를 중심으로 계층적으로 시각화하는 새로운 방법을 제안한다.

Abstract

Visualizations have played an important role in understanding new insights of users of social network for social network analysis. Most of the previous works of visualization focus on representing user's relationship on social network by a complex multi dimension graph. However, this method is difficult to identify the important of relationship to focus on personal user intuitively. Besides, the user's messages to reflect the interrelation between users is insufficient, since most of visualization methods represent the user relationship using information of interaction between nodes on networks. In order to resolve above problem, this paper proposes a new visualization method to visualize user based hierarchy that uses internal relationship of users by fuzzy relational product and external access information of network.

Keywords : 소셜 네트워크 서비스, 소셜 네트워크 분석, 시각화, 퍼지 연관 곱, 계층

* 정회원, 목포대학교 정보산업연구소

(Institute Research of Information Science and Engineering, Mokpo National University)

** 정회원, 인하대학교 컴퓨터정보공학부

(Department of Computer & Information Engineering, INHA University)

*** 정회원, 목포대학교 컴퓨터공학과

(Department of Computer Engineering, Mokpo National University)

**** 정회원, 목포대학교 정보통신공학과

(Department of Information Communication Engineering, Mokpo National University)

***** 정회원-교신전자, 목포대학교 정보전자공학과

(Department of Information Electronic Engineering, Mokpo National University)

※ 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구소 지원사업으로 수행된 연구임(2010-0028295), 본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음(NIPA-2012-H0301-12-2005)

접수일자: 2012년4월3일, 수정완료일: 2012년5월21일

I. 서 론

소셜 네트워크 서비스(SNS, social network service)는 사람들 사이의 사회적 관계나 사회적 네트워크를 반영하거나 구성하는 것을 목적으로 하는 사이트, 플랫폼 등의 온라인 서비스이다. 정보통신과 스마트 단말기의 발전으로 현실의 사회적 네트워크 관계들이 온라인상의 사회적 관계형성으로 신속히 이동하고 있다. 또한 현실에서는 제한적이었던 사회적 관계들이 온라인상에 형성되면서 시간 및 지역, 계층, 국가 등의 시공간의 제약을 극복하며 다양한 계층의 목적에 맞추어 형성되고 있다. 이렇게 온라인상에 형성되는 다양한 사회적 네트워크의 정보들은 온라인이나 오프라인상의 상업 활동의 추천정보나 기타 다양한 분야의 분석정보로 사용될 수 있는 유용한 정보이다. 이러한 이유 때문에 소셜 네트워크 서비스 분석에 대해서 현재 많은 연구들이 진행되고 있다. 특히 사회적 관계 중요한 공동체나 중심적인 역할을 수행하는 사용자를 네트워크상에서 그래프로 표현하는 시각화방법이 소셜 네트워크의 중요한 분석방법으로 많은 선호를 받고 있다.

현재 소셜 네트워크 시각화를 위한 연구의 대표적인 접근방법으로는 노드 링크(NL, node-link) 접근방법^[1], 행렬 그래프(MAT, matrix graph) 접근방법^[2], 노드 링크와 행렬 그래프의 혼합 접근방법(hybrid of NL and MAT)^[3~4]이 주로 연구되고 있으며, 이외에도 다양한 기법들을 기존방법들에 적용하여 성능을 향상시키는 확장 접근방법들이 있다^[5~9].

노드 링크를 이용한 접근방법들의 경우 소셜 네트워크상의 모든 노들을 시각화 하는데 유용한 방법이나, 그래프로 표현되는 노드들이 서로 겹치고 에지들이 서로 엇갈려서 노드의 양이 많아질수록 네트워크 구조를 읽기 어려운 문제를 가지고 있다^[1, 3, 4]. 행렬 그래프 접근방법은 노드 링크 기반방법에 비해서 쉽게 노드와 노드의 경로를 파악할 수 있다. 그러나 희소(sparse)행렬로 표현되는 노드는 공간낭비가 발생하며 노드 링크 기반방법과 마찬가지로 경로탐색이 어려운 문제를 가지고 있다^[2~3]. 노드 링크와 행렬 그래프의 혼합 접근방법은 노드 링크 방법의 경로를 행렬 그래프 접근방법의 노드에 연결하여서 노드 링크접근방법과 행렬 그래프 접근방법의 문제점을 해결하였다. 그러나 이 방법 역시 행렬 그래프를 기반으로 하고 있기 때문에 행렬 상에 나

타나는 노드의 관계를 이해하기 어렵다^[3~4]. 확장접근방법들은 복잡한 다차원 그래프를 기반으로 네트워크상의 노들들 간의 관계 표현에만 관심을 가지고 사용자 노들 간의 상호작용을 시각화하고 있기 때문에 사용자를 중심으로 한 사회적 관계의 중요도를 직관적으로 파악하기 힘들다. 또한 이 방법들이 대부분 네트워크상의 노드간의 접근 양에 의해서만 사용자 관계를 나타내기 때문에 사용자 메시지의 내용이 상호관계에 반영되는 것이 미흡한 실정이다.

본 논문은 이전 방법들의 문제점을 해결하기 위해서 소셜 네트워크의 사용자를 중심으로 시각화하는 새로운 방법을 제안한다. 제안방법은 퍼지 연관 곱을 이용하여 사용자의 내부관계를 계산하며, 노들들 간의 상호작용 정보를 이용하여서 사용자의 네트워크상의 외부 접근 정보를 계산한다. 계산된 내부관계정보와 외부 접근 정보를 이용하여서 사용자중심의 계층적 시각화방법을 제안한다.

제안방법은 다음과 같은 장점을 갖는다. 첫째, 사용자의 네트워크상의 내부와 외부의 관련정보를 반영함으로써 사용자들 간의 관계를 시각화에 더 자세히 표현할 수 있다. 즉, 사용자가 작성한 메시지를 네트워크의 내부관계에 반영함으로써 소셜 네트워크상에 쟁점이 되는 사안들을 사용자관계에 반영할 수 있으며, 사용자들 간의 노드의 접근 관계인 외부정보를 이용하여서 노드간의 경로를 쉽게 파악할 수 있다. 둘째, 사용자 중심으로 사용자관계를 계층적으로 표현하기 때문에 소셜 네트워크상에서 중요한 공동체나 중심적인 역할을 수행하는 사용자를 쉽게 찾을 수 있다. 마지막으로 사용자의 관계를 계층적으로 표시하기 때문에 이전관련연구들의 시각화 방법들의 읽기와 이해하기 어려운 문제를 해결할 수 있다.

본 논문의 구성은 다음과 같다. II장에서는 관련연구로 이전에 연구된 시각화방법들과 퍼지 연관 곱에 대하여 알아보고, III장에서는 퍼지 연관 곱에 의한 사용자 내부관계와 사용자의 외부 접근 정보를 이용한 시각화 방법에 대하여 알아본다. IV장에서는 제안방법을 이용한 사례 연구를 설명한다. 마지막 V장에서는 결론 및 향후 연구에 대하여 기술한다.

II. 관련 연구

2.1 시각화 방법

소셜 네트워크의 시각화 방법으로는 노드 링크 접근 방법, 행렬 그래프 접근방법, 노드 링크와 행렬 그래프의 혼합 접근방법, 기타 확장 접근방법이 있다. 노드 링크를 이용한 접근방법의 경우 네트워크의 전체 구조를 시각화하는데 유용한 방법이다. 이 때문에 많은 소셜 네트워크 분석 소프트웨어 들이 노드 링크를 이용하여 소셜 네트워크를 탐색하거나 시각화하고 있다^[10]. Ghoniem와 저자들은 노드 링크 방법에 시각화되는 노드들이 서로 겹치고 에지들이 서로 엇갈리는 문제를 지적하였다^[11]. 즉, 노드 링크를 이용하여 표현하는 사용자 간의 관계가 많아질수록 사용자의 관계를 파악할 수 없는 문제가 있다. 이러한 문제를 해결하기 위해서는 샘플링, 필터링, 군집과 같은 후처리를 통하여 필요한 부분만을 정제하여 해결할 수 있는 다양한 방법들이 제안되었으나, 역시 시각화되는 결과가 이해하기 힘들거나 비용이 많이 드는 문제를 가지고 있다^[1, 3~4, 11].

행렬 그래프 접근방법은 노드 링크 기반방법의 시각화 결과에 대한 가독성이 어려운 문제를 해결하기 위해서 제안된 방법이다. 행렬 그래프 기반방법은 노드 링크 기반방법에 비해서 쉽게 노드를 파악할 수 있으며, 노드들 간의 경로도 역시 쉽게 파악할 수 있다. 그러나 행렬 형태로 표현되는 노드들이 행렬 상에 최소행렬로 표시되어 많은 공간을 사용하기 때문에 공간에 대한 낭비가 발생하며 노드 링크 기반방법과 마찬가지로 경로 탐색이 어려운 문제를 가지고 있다. 또한 행렬 그래프로 표현되는 노드들이 이해하기 힘든 문제를 가지고 있다^[2~3, 11].

노드 링크와 행렬 그래프의 혼합 접근방법으로는 대표적으로 Henry와 저자들이 제안한 MatLink^[3]와 MatTrix^[4]가 있다. MatLink 방법은 노드 링크 방법과 행렬 그래프 방법의 경로를 해결하기 위해서 제안된 방법이다. 이 방법은 노드 링크 방법의 에지를 행렬 그래프 방법에 겹쳐서 행렬 그래프 방법의 성능을 향상시키고 있다. 그러나 이 방법 역시 행렬에 연결된 링크들이 복잡하게 구성되어 있기 때문에 이해하기 어려운 단점을 가지고 있다^[1, 3~4]. MatTrix 방법은 MatLink방법의 이해하기 어려운 문제점을 해결하기 위해서 제안된 방법이다. 이 방법은 행렬 그래프를 더 작은 여러 개의 행

렬 그래프로 분해해서 링크를 연결한다. 그러나 이 방법 역시 행렬 그래프를 기반으로 하고 있기 때문에 행렬 상에 나타나는 노드의 관계를 이해하기 어렵다^[4].

확장접근 방법으로, Heer와 Boyd는 vizster를 노드 링크와 비슷한 구조를 갖고 있으면서 노드 및 네트워크를 탐색 및 검색할 수 있는 기능을 제공하는 vizster라는 시각화 시스템을 제안하였다^[5]. Brandes와 Wagner는 그래프 이론 개념을 이용하여 소셜 네트워크의 구조를 설명하고, 기술하고 이해할 수 있는 visone 소프트웨어를 제안하였다^[6]. Smith와 저자들은 엑셀 2007에 네트워크 탐색을 위한 공개 소스인 NodeXL을 이용하여 네트워크를 탐색 및 검색할 수 있으며 네트워크를 개괄적으로 표현하여 분석할 수 있는 도구를 구현하였다^[7]. Ahn와 저자들은 NodeXL과 다른 시간대의 장소를 탐색하게 하는 TempoVis를 이용하여 시간 변화에 따라서 동적으로 소셜 네트워크를 시각화하는 방법을 제안하였다^[8]. Shen와 저자들은 온톨로지 그래프와 의미 그래프를 정의하여서 대량의 이기종의 네트워크를 시각적으로 분석할 수 있는 방법을 제안하였다^[9].

2.2 퍼지 연관 곱

퍼지 연관 곱(fuzzy relational product)은 크리스프 함의 연산자 (crisp implication operator)를 확장하여 퍼지에 적용한 것으로서, 크리스프 함의 연산자는 $\{0,1\} \times \{0,1\} \rightarrow \{0,1\}$ 로 정의되는데 반해, 퍼지 함의 연산자는 $[0,1] \times [0,1] \rightarrow [0,1]$ 로서 단위 구간의 다치 논리로 확장된 것이다. 퍼지 함의 연산자의 종류는 무수히 많으며 본 논문에서는 대표적인 Kleene-Diense 퍼지함의 연산자를 사용하며 예는 다음 식(1)과 같다^[12~13].

$$\begin{aligned} a \rightarrow b &= (1-a) \vee b = \max(1-a, b), \\ a &= 0 \sim 1, b = 0 \sim 1 \end{aligned} \quad (1)$$

집합이론에서 " $A \subseteq B$ "는 " $\forall x, x \in A \rightarrow x \in B$ "와 같고 " $A \in \mathcal{S}(B)$ "와도 같다. 여기서 $\mathcal{S}(B)$ 는 B 의 멱집합(power set)이다. 따라서 퍼지 집합에서의 " $A \in B$ 인 정도"는 $A \in \mathcal{S}(B)$ 인 정도이므로 $\mu_{\mathcal{S}(B)}A$ 로서 나타낼 수 있으며 다음과 같이 정의된다.

(정의1) 퍼지 함의 연산자 \rightarrow 와 크리스프 전체집합 U 의 퍼지 집합 B 가 주어진 상태에서 B 의 퍼지 멱집합의 멤버십 함수 $\mu_{\mathcal{S}B}$ 는 다음과 같이 주어진다.

$$\mu_{\phi B} A = \bigwedge_{x \in U} (\mu_A x \rightarrow \mu_B x) \quad \blacklozenge$$

(정의2) U_1, U_2, U_3 는 유한한 전체 집합이라 하고, R 은 U_1 에서 U_2 로의 퍼지관계이고, S 는 U_2 에서 U_3 로의 퍼지 관계이다. 즉, R 은 $U_1 U_2$ 의 퍼지 부분집합이고 S 는 $U_2 U_3$ 의 퍼지 부분집합이다. 퍼지 관계 곱은 a 는 U_1 이고 c 는 U_3 일 때, a 가 c 에 관련되어 있는 정도를 나타낼 사용되는 퍼지연산이다. U_1 에서 U_2 로의 퍼지관계인 삼각 논리곱 \triangleleft , 는 다음과 같이 정의된다.

$$(R \triangle S)_{ik} = \frac{1}{N_j} \sum_j (R_{ij} \rightarrow S_{jk}) \quad (2)$$

이것을 퍼지 관계곱 (fuzzy relational products) 이라 한다. \blacklozenge

(정의3) 퍼지 함의 연산자는 주어진 문제의 범주에 따라 달라진다. $a \in U_1$ 에 대한 후위집합 (afterset) aR 은 a 와 연관된 $y \in U_2$ 로 구성된 U_2 의 퍼지 부분집합이며 그 멤버십 함수는 $\mu_{aR}(y) = \mu_R(a, y)$ 로 주어진다. $c \in U_3$ 에 대한 전위집합 (foreset) Sc 는 c 에 연관된 $y \in U_2$ 로 구성된 U_2 의 퍼지 부분집합이며 그 멤버십 함수는 $Sc(y) = S(y, c)$ 로 주어진다. aR 이 Sc 의 부분집합인 평균정도는 $y \in aR$ 의 멤버십 정도가 $y \in Sc$ 의 멤버십 정도를 함의하는 평균정도로써 다음과 같이 정의된다.

$$\pi_m(aR \subseteq Sc) = \frac{1}{N_{U_2}} \sum_{y \in U_2} (\mu_{aR}(y) \rightarrow \mu_{Sc}(y)) \quad (3)$$

여기서 π_m 은 평균 정도를 나타내는 함수이다.

위의 평균 정도는 $R \triangleleft S$ 에 의해서 a 가 c 에 관련되는 정도를 나타낼 수 있다^[10~12].

III. 제안방법

본 논문에서 방법은 다음 그림1과 같이 전처리, 사용자 내외부 관계계산, 사용자 관계 계층적 시각화로 구성된다. 전처리 단계에서는 소셜 네트워크로부터 자료를 수집하여 내부정보와 외부정보의 관계 계산을 위해서 각 수집 자료를 벡터 행렬로 변환한다. 사용자 내외부 관계 계산 단계에서는 전처리된 벡터행렬을 이용하여 사용자 내부 관계 및 노드 간 접근 정보를 계산한다. 사용자 관계 계층적 시각화단계에서는 사용자 내부 관계와 노드 접근 정보를 합산하고, 사용자의 중요도를

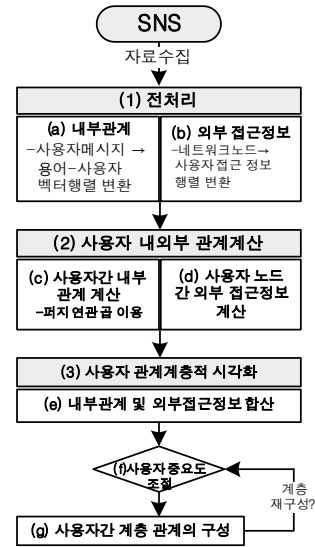


그림 1. 퍼지 연관 곱과 외부접근정보를 이용한 시각화 방법

Fig. 1. Visualization method using fuzzy relational product and external access information.

조절하여 시각화한다. 만약 시각화에 나타난 사용자관계를 조절할 필요가 있으면 중요도를 조절하여 재 시각화 한다.

3.1 전처리

그림1(1)의 전처리 단계에서는 그림1(a)의 내부관계와 그림1(b)의 외부 접근정보로 구분된다. 내부관계는 사용자가 소셜 네트워크에 게재한 메시지들로 사용자들의 메시지들 간의 관계정보를 이용한다. 외부 접근정보는 소셜 네트워크상에서 노드들 간에 접근되는 정보로 네트워크상의 노드들 간의 전송되는 메시지의 양과 사용자간 메시지가 참조되는 횟수를 이용한다.

내부관계에 대한 전처리는 사용자들의 메시지에서 다음과 같이 용어-사용자 빈도 행렬을 구성한다. 한글 메시지 경우 한글 형태소 분석 도구^[14]를 사용하여 명사만을 추출하여 빈도 행렬을 구성하며, 영문 메시지 경우 불용어 제거, 어근 추출한 후에 빈도 행렬을 구성한다. 불용어는 Rijsbergen의 불용어 목록^[15, 16]에 정의되어있는 불용어를 어휘 분석하여 무의미한 용어를 제거한다. 어근 추출은 Porter의 어근 추출 알고리즘^[15, 16]을 이용하여 영어의 파생어들 중에서 가장 중심이 되는 용어인 어근으로 변환한다. 용어-사용자 빈도 행렬의 $M_{ij} = [t_{1j}, t_{2j}, \dots, t_{nj}]^T$ 는 사용자 메시지인 j 번째 열에 포함된 용어의 출현 빈도를 나타낸다. 여기서 용어빈도 t_{ij} 는 j

표 1. 외부 접근정보를 표시하는 기호
Table 1. Representing symbols of external access information.

기호	의미
<i>nm</i> (number of message)	한사용자가 다른 사용자에게 보내는 메시지의 개수
<i>tm</i> (total of message)	사용자들 간 참조되는 메시지의 총 개수
<i>nrm</i> (number of resend message)	사용자가 재전송하는 메시지의 개수
<i>ru</i> (re-reference user)	메시지를 재 참조하는 사용자의 개수
<i>d_i</i> (day)	<i>i</i> 번째 일의 하루 동안 받은 메시지의 총 개수
<i>td</i> (total day)	메시지를 주고받는 총 일 수
<i>nt</i> (number of term)	두 사용자가 사용한 메시지에 포함된 명사의 개수

번째 사용자 메시지에 출현한 *i*번째 용어의 출현 빈도이다.

외부 접근정보에 대한 전처리는 사용자간 메시지 참조 기간, 참조 횟수, 참조 량에 대하여 표 1과 같이 정보를 수집하여 전처리한다.

3.2 사용자 내외부 관계계산

그림 1(2)의 사용자 내외부 관계계산은 그림 1(c)의 사용자간 내부관계 계산과 그림 1(d)의 사용자 노드 간 외부 접근정보계산으로 구분된다. 사용자의 내부 관계계산은 소셜 네트워크에서 전송되는 사용자의 메시지 내부의 정보를 얼마나 반영되는지를 나타낸다. 사용자 노드 간 외부 접근정보계산은 소셜 네트워크상에서 참조되는 사용자의 메시지의 양이 소셜 네트워크상에서 얼마나 반영되었는가를 나타낸다.

3.2.1 사용자 내부 관계 계산

그림 1(c)의 사용자간 내부관계 계산은, 전처리된 용어-사용자 빈도행렬에 식(1)의 Kleene-Diense 퍼지 합의 연산자를 기반으로 한 식(4)의 퍼지 연관 곱을 계산하여서, 사용자와 사용자간의 포함관계를 기반으로 사용자 내부 관계를 계산한다. 일반적으로 식(3)의 퍼지 연관 곱을 적용하여 분류들 간의 퍼지 연관 관계, $C_i \rightarrow C_j$ 를 유도할 수 있다. 그러나 C_i 에 멤버십 값($\mu_{C_i}(x)$)이 작은 원소 x 가 많으면, $C_i \subseteq C_j$ 의 포함여부와 관계없이 항상 1에 가까운 값이 나오는 문제점이 있다. 따라서 다음과 같이 정의하여 두 분류 퍼지 집합의 합의 관계, $\mu_{m,\beta}(C_i \subseteq C_j)$ 를 계산한다.

$$\mu_{m,\beta}(C_i \subseteq C_j) = (R^T \Delta_\beta R)_{ij} = \frac{1}{|C_{i\beta}|} \sum_{K_{i\beta} \in C_{i\beta}} (R_{ik}^T \rightarrow R_{kj}) \tag{4}$$

여기서, K_k 는 k 번째 용어이고, C_i, C_j 는 i 번째와 j 번째 분류이며, $C_{i\beta}$ 는 C_i 의 β -제약, $\{x | \mu_{C_i}(x) \geq \beta\}$ 이고 $|C_{i\beta}|$ 는 $C_{i\beta}$ 의 원소의 갯수이다. R 은 $m \times n$ 행렬로서 R_{ij} 는 $\mu_{C_i}(K_j)$, 즉, $K_j \in C_i$ 인 정도이다. R^T 는 행렬 R 의 전치행렬로서 $R_{ij} = R^T_{ji}$ 이다.

3.2.2 외부 접근정보 계산

그림1(b)의 외부 접근정보는 소셜 네트워크상에서 참조되는 메시지를 이용하기 때문에 방향성이 고려되어야 한다. 즉, 사용자 a 의 입장에서는 사용자 b 에게 보내는 메시지의 비율이 증가할수록 a 에서 b 로의 외부 접근관계가 증가한다. 또한 최근에 참조하는 메시지일수록 외부 접근관계가 증가한다. 이외에도 메시지 참조는 메시지를 게재하는 사용자나 참조하는 사용자나 모두 참조 행위를 인지할 있는데 비해, 재 참조는 재전송하는 b 사용자만이 누구에게 보내는지 인지할 수 고, a 사용자는 다른 사용자의 재전송 메시지참조 유무는 알 수 없다. 이 때문에 참조메시지에 비해서 재 참조되는 메시지는 외부관계에 중요도가 적게 반영된다. 본 논문에서는 이러한 사용자의 노드상의 접근정보를 반영하여서 식(5)와 같이 사용자간 노드 외부 접근정보 ai (access information)가 계산된다.

$$ai(a \rightarrow b) = \left(\frac{nm \times nt}{tm} \right) \times \sum_{i=1}^{td} \left(\frac{d_i}{td - (1-i)} \right) + \left(\frac{nrm}{tm \times ru} \right) \tag{5}$$

여기서 nm 은 사용자 a 가 사용자 b 에게 보내는(즉, b 가 참조하는 a 의) 메시지 개수, nt 는 두 사용자가 참조한 모든 메시지에 포함된 명사 용어의 개수, tm 은 소셜 네트워크상에서 모든 사용자가 참조하는 메시지의 개수이다. td 는 메시지가 참조되는 총일자의 개수이며, 일자별 참조되는 메시지의 개수 d 는 최근 날짜를 기준으로 내림차순으로 정렬한다. nrm 은 사용자가 재전송하는 메시지의 개수, ru 는 실제 메시지를 재 참조하는 사용자의 개수이다.

3.3 사용자 관계계층적 시각화

사용자 관계 계층적 시각화단계는 다음과 같다. 첫째, 계산된 내부 관계와 외부접근정보를 식(6)을 이용하여 정규화한다. 둘째, 정규화된 내외부 관계를 식(7)을 이용하여 합산한다. 셋째, 식(8)을 이용하여서 내외부관계의 합산 행렬에서 평균이상의 요소 값을 가지는 사용자들만 관계그래프로 표현한다. 표현된 관계그래프의 상관 값의 노드간의 합을 이용하여 사용자의 계층관계도를 구성한다. 즉, 사용자의 노드는 노드에 접속되는 에지의 합이 큰 쪽이 계층도의 상위노드가 되며, 노드와 노드는 한번만 연결할 수 있다. 마지막으로 합산 행렬의 평균값을 기준으로 사용자 계층관계의 중요도를 조정한다. 즉, 평균값 이상이나 이하의 가중치를 부여해서 시각화된 계층적 관계를 조정할 수 있다.

사용자의 내부관계와 외부 접근정보를 시각화에 반영하기 위해서 내외부 관계를 합산해야 한다. 그러나 내부관계와 외부 접근정보가 나타내는 원소 값의 정량적 비율이 다르기 때문에 이를 시각화에 반영하면 내부관계나 외부관계한쪽으로 치우쳐서 시각화되는 문제가 발생할 수 있다. 이러한 문제를 해결하기 위해서 내외부관계의 요소 값이 전체 사용자들 중에서 차지하는 비율로 정규화(normalization)하여서 원소 값의 불균형 문제를 해결한다. 다음 식(6)은 내부 상관관계 및 노드 외부관계의 원소를 정규화 하는 식이다.

$$nor(a \rightarrow b) = \frac{u_{ab}}{\sum_{a=1}^l \sum_{b=1}^l u_{ab}} \quad (6)$$

여기서 $nor()$ 는 관계의 원소 값을 정규화 시키는 함수이며, $a \rightarrow b$ 는 a사용자를 참조하는 b사용자를 나타내며, l은 사용자의 총인원수, u_{ab} 는 a사용자를 참조하는 b사용자관계의 원소 값을 나타낸다.

내외부관계의 원소 값을 정규화 한 후에 이들을 시각화에 반영하기 위해서는 내부관계와 외부 접근정보의 원소 값을 다음 식(7)과 같이 합산해야 한다.

$$sie(a \rightarrow b) = nor(c_{a,b}) + nor(er(a \rightarrow b)) \quad (7)$$

여기서 $sie()$ 는 내부관계와 외부관계 원소를 합산(sum of internal and external relation element)을 시키는 함수이며, $nor()$ 는 정규화 함수, ca,b 는 a와 b사용자에 대한 상관관계 원소, $er(a \rightarrow b)$ 는 a와 b사용자에 대한 외부

관계의 원소를 나타낸다.

다음 식(8)은 사용자의 관계를 계층적으로 시각화시 기준점을 제시하는 식이다. 본 논문에서는 내외부관계의 합산 행렬에서 평균이상의 요소 값을 이용한다.

$$coverage(u_i) = \frac{\sum_{j=1}^l u_{ij}}{l} \quad (8)$$

여기서 u_i 는 사용자를 나타내며, l은 사용자의 수, u_{ij} 는 사용자와 사용자의 상관관계를 나타낸다.

IV. 사례 연구

현재 소셜 네트워크의 시각화를 위한 표준화된 평가 방법이나 표준 평가 자료들이 존재하지 않기 때문에 시각화방법에 대한 성능평가는 어려운 부분 중의 하나이다. 이 때문에 대부분의 이전 소셜 네트워크의 시각화 방법에 대한 관련연구에서는 성능평가 대신에 사례연구를 이용하여 제안방법을 설명하고 있다. 이는 대부분의 시각화 방법들이 소셜 네트워크의 분석을 위해서 도구의 하나로 사용되기 때문에 분석을 위한 제안방법과 이들의 특성을 설명하고 있다. 본 논문도 사례연구를 통하여 본 논문에서 제시한 방법과 특성을 설명한다.

다음은 사례연구로서 소셜 네트워크 사용자의 관계를 계층적으로 시각화하는 방법을 설명한다. 이 예는 사용자들 간의 관계시각화단계를 단순히 하여서 쉽게 이해할 수 있도록 사용자의 내부 관계만을 이용하여 설명한다. 다음 표 2는 5명의 사용자의 메시지에 포함된 7개의 용어를 출현의 빈도로 나타낸 용어-사용자 빈도 행렬이다. 표 3은 표2를 0과 1사이의 퍼지 멤버십으로 계산한 결과이다. 표 4는 표 3에 식(4)를 이용하여 퍼지 연관 곱을 계산한 사용자-사용자 관계 행렬이다. 표5는 표 4의 사용자 관계를 가지는 사용자들의 평균을 식(8)을 이용하여 계산한 결과이다.

표 2. 용어-사용자 빈도행렬

Table 2. Term-user frequency matrix.

	$u1$	$u2$	$u3$	$u4$	$u5$
$t1$	0	1	0	0	0
$t2$	1	1	0	0	0
$t3$	2	3	0	1	2
$t4$	1	0	5	3	1
$t5$	0	0	2	1	1
$t6$	0	0	1	2	1
$t7$	0	0	0	0	1

표 3. 표2의 퍼지 멤버십

Table 3. Fuzzy membership of Table 2.

	$u1$	$u2$	$u3$	$u4$	$u5$
$t1$	0.00	0.33	0.00	0.00	0.00
$t2$	0.50	0.33	0.00	0.00	0.00
$t3$	1.00	1.00	0.00	0.33	1.00
$t4$	0.50	0.00	1.00	1.00	0.50
$t5$	0.00	0.00	0.40	0.33	0.50
$t6$	0.00	0.00	0.20	0.67	0.50
$t7$	0.00	0.00	0.00	0.00	0.50

표 4. 표3의 퍼지 연관 곱 계산결과

Table 4. Results of fuzzy relational product of Table 3.

	$u1$	$u2$	$u3$	$u4$	$u5$
$u1$	0.29	0.24	0.23	0.33	0.33
$u2$	0.71	0.71	0.71	0.71	0.33
$u3$	0.29	0.24	0.23	0.33	0.43
$u4$	0.29	0.24	0.23	0.33	0.43
$u5$	0.29	0.24	0.23	0.33	0.43

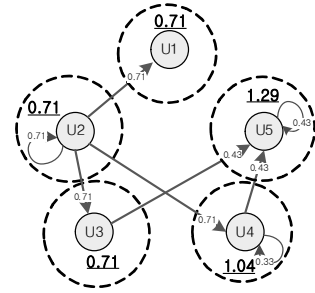
표 5. 표4의 사용자의 평균

Table 5. Average of user in Table 4

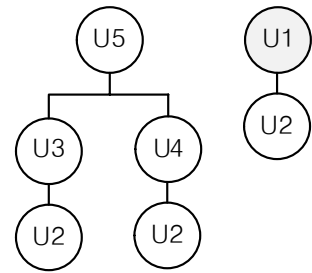
구분	$u1$	$u2$	$u3$	$u4$	$u5$
평균	0.371	0.333	0.325	0.409	0.485

그림 2는 표 4의 퍼지 연관 곱의 계산 결과와 표 5의 내부관계의 사용자 평균을 이용하여서 도식화한 사용자 관계 그래프와 사용자 관계 계층도이다. 그림 2(a) 및 그림 2(b)는 사용자 u1을 중심으로 그래프 및 계층도를 나타낸 것이다. 즉, 그림 2(a)는 사용자 u1의 평균(0.371)이상의 값을 가지는 사용자 노드만을 연결하여 그래프로 도식화한 것이다. 그림 2(b)는 본 논문에서 제안한 방법으로 그림 2(a)의 에지의 합을 기준으로 노드의 상하계층을 결정하여 도식화한 것이다. 그림 2(b)와 그림 2(d)의 사용자 u3를 중심으로 관계 계층도를 나타낸 것이다. 그림 3(b)에서는 u1과 u5가 최상위 계층으로 있으며, u1은 u2를 자식 노드로 갖으며, u3와 u4는 형제관계로 u5의 자식노드이다. u2는 모든 자식 노드의 단말 노드이다. 즉, u1을 중심으로 보았을 때, u1의 메시지는 주로 u2가 참조하는 것을 알 수 있으며, 다른 계층의 메시지 역시 u2가 주로 참조한다는 것을 알 수 있다. 그림 2(d)에서 사용자 u4는 모든 사용자의 최상의 노드이고, u1, u2, u5는 u4의 자식노드로 서로 형제관계를 가지고 있다. 또한 u3는 u5의 자식 노드로 단말노드로 구성되고 있다. 즉, u3는 주로 u5의 메시지를 참조하나 모든 노드의 메시지 역시 참조하는 것을 알 수 있다.

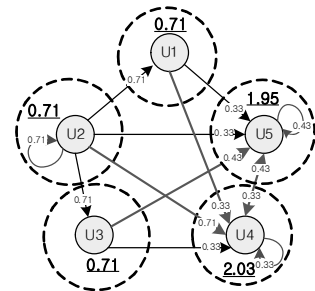
그림 2(b)와 그림 2(d)에서 보이는 것과 같이 사용자의 기준 관점에 따라서 소셜 네트워크의 사용자 상관관계 계층도가 달라지는 것을 알 수 있다. 또한 만약 사용



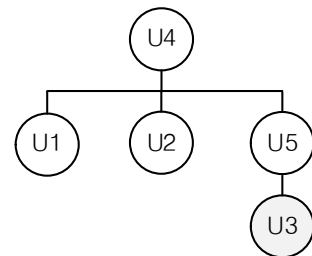
(a) 사용자 u1의 관계 그래프
(a) Relationship diagram of user u1



(b) 사용자 u1의 관계 계층도
(b) Hierarchy diagram of user u1



(c) 사용자 u3의 관계 그래프
(c) Relationship diagram of user u3



(d) 사용자 u3의 관계 계층도
(d) Hierarchy diagram of user u3

그림 2. 사용자 관계를 이용한 관계 그래프와 계층도
Fig. 2. Relationship diagram and hierarchy diagram using user relationship.

자 관계 계층도를 특정 사용자를 기준으로 재구성할 필요가 있으면 그 사용자의 퍼지 연관 곱의 평균값을 기

준으로 재구성 할 수 있다. 즉, 그림 2에서 보는 것과 같이 평균값 이하로 기준 값을 설정하면 계층도가 확대 되며, 평균값 이상으로 기준 값을 설정하면 계층도가 축소되는 특성을 지닌다.

V. 결 론

본 논문은 이전 소셜 네트워크 시각화 방법의 문제점을 해결하기 위해서 새로운 시각의 시각화방법을 제안하였다. 제안된 방법은 소셜 네트워크상의 사용자간의 내부관계와 노드의 외부 접근 정보를 반영하여 사용자간의 관계를 사용자 중심으로 계층적 시각화하였다. 제안방법은 다음과 같은 장점을 갖는다. 첫째, 사회적 관계를 계층적으로 시각화하기 때문에 직관적으로 쉽게 중요도를 파악 및 분석할 수 있다. 둘째, 사용자가 작성한 메시지를 네트워크의 내부관계에 반영함으로써 소셜 네트워크상에 쟁점이 되는 사안들을 사용자관계 계층에 반영할 수 있다. 셋째, 네트워크 노드의 관계를 외부 접근정보에 반영함으로써 노드간의 경로를 쉽게 파악할 수 있다. 마지막으로 사용자 중심으로 사용자관계를 계층적으로 표현하기 때문에 소셜 네트워크상에서 중요한 공동체나 중심적인 역할을 수행하는 사용자를 쉽게 찾을 수 있다. 마지막으로 사용자 관계의 계층적 시각화는 이전 관련 연구들의 시각화결과의 읽기와 이해하기 어려운 문제를 해결할 수 있다.

참 고 문 헌

[1] M. Ghoniem, J. D. Fekete, P. Castagliola, "On the readability of graphs using node-link and matrix based representations, a controlled experiment and statistical analysis, Information Visualization, Vol. V, No. 2, pp.114-143. 2005.

[2] S. Wasserman, K. Faust, "Social Network Analysis", Cambridge University Press, Combridge, 1994.

[3] N. Henry, J.-D. Fekete, "MatLink: Enhanced Matrix Visualization for Analyzing Social Networks", LNCS 4663, Part II, pp. 288-302, 2007.

[4] N. Henry, J.-D. Fekete, M. J. McGuffin, "NodeTrix: a Hybrid Visualization of Social Networks", IEEE Transactions on Visualization and Computer Graphics, Vol. 13 Issue:6,

pp.1302-1309, 2007.

[5] J. Heer, D. Boyd, "Vizster: Visualizing Online Social Networks", IEEE Symposium on Information Visualization 2005, pp.32-39, 2005.

[6] U. Brandes, D. Wagner, "Visone: Analysis and Visualization of Social Networks", Graph Drawing Software, Springer, pp.321-340, 2004.

[7] M. A. Smith, B. Shneiderman, N. Milic-Frayling, E. M. Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer, E. Gleave, "Analyzing (Social Media) Networks with NodeXL", In proceedings of the fourth international conference on communities and technologies, C&T'09, pp.255-264, 2009.

[8] J.-W. Ahn, M. Taieb-Maimon, A. Sapan, C. Plaisant, B. Shneiderman, "Temporal Visualization of Social Network Dynamics: Prototypes for Nation of Neighbors", LNCS 6589, Springer, pp.309-316, 2011.

[9] Z. Shen, K.-L. Ma, T. Eliassi-Rad, "Visual Analysis of Large heterogeneous Social Networks by Semantic and Structural Abstraction", IEEE Transactions on Visualization and Computer Graphics, Special Issue on Visual Analytics, pp.1427-1439, 2006.

[10] International Network for Social Network Analysis, "http://www.insna.org", 2012.

[11] M. Ghoniem, J.-D. Fekete, P. Castagliola, "On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis", Information Visualization, Vol. 4, pp.114-135, 2005.

[12] W. Bandler and L. Kohout. "Semantics of Implication Operators and Fuzzy Relational Products," International Journal of Man-Machine Studies. Vol. 12, pp.89-116, 1980.

[13] B. G. Choi, J. H. Lee, S. Park. "Dynamic Construction of Category Hierarchy Using Fuzzy Relational Products," IDEAL 2003, pp.296-302, 2003.

[14] 한경남, 남경완, "한국어 정보 처리 입문 : 컴퓨터가 우리말을 이해하려면", 커뮤니케이션북스, 2007.

[15] W. B. Frakes, B. Y. Ricardo, "Information Retrieval : Data Structure & Algorithms", Prentice-Hall, 1992.

[16] B. Y. Ricardo, R.-N. Berthier, "Modern Information Retrieval the concepts and technology behind search second edition", Addison Wesley, 2011.

— 저 자 소 개 —



박 선(정회원)
 1996년 전주대학교 전자계산학과 학사
 2001년 한남대학교 정보통신학과 석사
 2007년 인하대학교 컴퓨터정보공학과 박사
 2008년~2009년 호남대학교 컴퓨터공학과 전임강사.
 2010년 전북대학교 전기전자정보인력양성사업단 박사후과정.
 2011년~현재 목포대학교 정보산업연구소 연구교수.
 <주관심분야 : 정보검색, 데이터마이닝, 데이터베이스, 해양생물 IT정보융합>



정 민 아(정회원)
 1994년 2월 전남대학교 전산통계학과 석사
 2002년 2월 전남대학교 전산통계학과 박사
 2005년 3월~현재 목포대학교 컴퓨터공학과 부교수
 <주관심분야 : 데이터베이스/데이터마이닝, 생체인식시스템, 무선통신응용분야(RFID, USN, 텔레메틱스), 임베디드시스템>



이 연 우(정회원)
 1994년 2월 고려대학교 전자공학과 석사
 2000년 2월 고려대학교 전자공학과 박사
 2000년 10월~2003년 12월 영국 Edinburgh 대학교 Research Fellow
 2004년 1월~2005년 8월 삼성종합기술원
 2005년 9월~현재 국립목포대학교 정보공학부 정보통신공학전공, 부교수
 <주관심분야 : 해상무선통신, e-Navigation, Cognitive Radio, 4G 이동통신>



권 장 우(정회원)
 JangWoo Kwon received the B.S degree in electronic Eng. from INHA University in 1990, the M.E. and Ph.D. degree in electronic engineering from INHA University in 1992 and 1996, respectively. In 1992 he was a visiting Researcher at Department of Biomedical Engineering of Tokyo University, Tokyo, Japan. From 1996 to 1998 he was a deputy director of Korea Industrial Property Office (KIPO) where his responsibility was to examine patents. From 1998 to 2009 he was an Associate Professor of Department of Computer Engineering at Tongmyoung University, Pusan, Korea. He had been a Dean of Research Institute for Information Eng. Tech. at Tongmyoung University from 2002 to 2006. From 2010 to 2012 he was an Associate Professor of Department of Computer Eng. at Kyungwon University, Kyeong-gi Province, Korea. Since 2006, he has been a Director of Human Resource Development Division of National IT industry promotion agency of Korea. Currently, his research area is in sensor networks and human computer interaction using biomedical signals. For the last 20 years he has been working in biomedical signal analysis and its recognition using artificial intelligence.



이 성 로(정회원)-교신저자
 1987년 고려대학교 전자공학과 졸업
 1990년 한국과학기술원 전기및 전자공학과 석사
 1996년 한국과학기술원 전기및 전자공학과 박사
 1997년 9월~현재 목포대학교 공과대학 정보전자공학과 교수
 <주관심분야 : 디지털통신시스템, 이동 및 위성통신시스템, USN/텔레메틱스응용분야, 임베디드시스템>