

블로그 연결망의 성향 판정 방안

윤석호*, 김상욱*, 박선주**
 *한양대학교 전자컴퓨터통신공학과
 **연세대학교 경영학과
 e-mail: bogely@agape.hanyang.ac.kr

A Decision Method for Propensities of Blog Networks

Seok-Ho Yoon*, Sang-Wook Kim*, Sunju Park**

*Department of Electronics and Computer Engineering, Hanyang University

**Department of Business, Yonsei University

요 약

본 논문에서는 주어진 블로그 연결망의 정보 중시 성향과 친분 중시 성향의 정도를 판정하는 방안에 관하여 논의한다. 먼저, 데이터 마이닝 기법의 하나인 분류(classification)를 이용하여 블로그 연결망의 기본 단위인 관계 성향의 정도를 판정하고, 그 결과를 이용하여 주어진 연결망의 전체 성향의 정도를 판정한다. 또한, 블로그 연결망의 규모에 따라 성향의 정도가 좌우되는 문제를 해결하기 위한 기법을 제안한다. 실제 블로그 데이터를 이용한 실험을 통하여 제안하는 방안의 우수성을 검증한다.

1. 서론

사회연결망(social network)은 한 사회 내에 존재하는 구성원들 간의 관계 정보를 이용하여 그 사회 전체의 관계 구조를 그래프의 형태로 표현한 것이다. 이러한 관계 구조를 연구하여 그 사회가 가지고 있는 고유의 특징을 도출하는 것을 사회연결망 분석(social network analysis)이라고 한다[1]. 과거에는 사회연결망 내의 구성원 사이의 관계 여부를 나타내는 데이터를 확보하는데 큰 어려움이 있었다. 따라서 기존 연구의 초점은 구성원들 간의 관계에 대한 구체적인 데이터 없이 사회연결망이 과연 어떠한 구조와 특징을 가지는가를 모델링하는 데에 있었다[2]. Barabasi[3], Granovetter[4], Milgram[5] 등의 연구가 이러한 형태의 전형적인 연구이다.

인터넷의 발달로 인하여 온라인상에서도 사회연결망이 나타났다. 온라인 사회연결망의 가장 큰 특징은 구성원들 간의 관계를 설명할 수 있는 구체적인 정보와 그 관계에 의하여 주고받는 정보를 데이터베이스 내에 저장하고 있다는 것이다. 이러한 특징으로 인하여 최근 들어 온라인 사회연결망을 분석하고자 하는 연구가 이루어지고 있다 [6].

블로그는 사용자가 자신의 글을 온라인상에 저장할 수 있는 일종의 개인 웹사이트이며, 블로그 세계(blog world)는 이러한 블로그로 이루어진 대표적인 온라인 사회연결망이다. 각 블로그 사용자는 타인의 블로그와 관계를 맺을 수 있으며, 이러한 관계들에 의하여 온라인 사회연결망이 형성된다[7]. 본 논문에서는 블로그들 간의 관계들로 이루어진 사회연결망을 블로그 연결망(blog network)이라고 정의한다.

블로그 연결망은 글을 주고받는 목적에 따라 구분되는 정보 중시 성향과 친분 중시 성향을 가질 수 있다. 정보 중시 성향은 유용한 정보의 전파와 수집을 목적으로 글을 주고받는 성향으로 정의되며, 친분 중시 성향은 사용자 상호간의 친분을 목적으로 글을 주고받는 성향으로 정의된다.

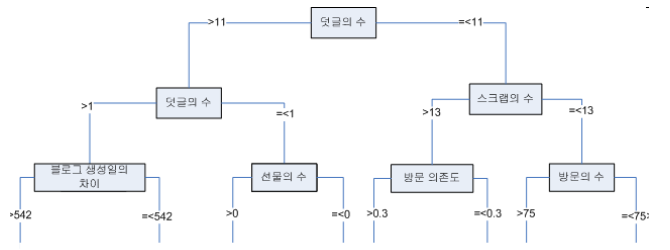
블로그 세계를 운영하는 운영자 또는 기업의 목표는 블로그들의 활동이 많아져서 전체적으로 블로그 세계가

활성화 되는 것이다[8]. 블로그 연결망이 가지고 있는 각 성향 정도는 해당 블로그 연결망에 대한 활성화 정책을 세우기 위한 중요한 판단 근거가 된다. 본 논문에서는 주어진 블로그 연결망에 대하여 두 성향의 정도를 판정하는 방안에 관하여 논의하고자 한다.

2. 블로그 연결망의 성향 판정

블로그 연결망이 가지고 있는 두 가지 성향은 글을 주고받는 목적에 따라서 구분된다. 따라서 연결망의 구성원인 블로그 사용자들이 연결망 내에서 주고받은 글들의 목적을 알려준다면, 그 결과를 이용해서 블로그 연결망에 대한 성향의 정도를 판정을 할 수 있다. 따라서 본 논문에서는 먼저 블로그 연결망의 기본 단위인 하나의 관계의 성향을 그 관계에 참여하는 두 블로그 사용자들에게 설문문을 통하여 판정하게 하고, 이런 방식으로 판정된 관계들의 성향을 이용해서 전체 블로그 연결망의 성향을 판정하는 방식을 제안한다.

그러나 블로그 연결망의 성향 판정이 요구될 때마다 매번 구성원인 블로그 사용자에게 성향 판정을 요구할 수는 없다. 따라서 본 논문에서는 두 블로그 간 관계의 성향 판정을 위하여 데이터 마이닝의 분류 기법을 사용한다[9]. (그림 1)은 데이터 마이닝의 분류 기법을 이용해서 생성한 의사결정트리의 한 예이다. 이를 통해 주어진 관계를 자동적으로 판정할 수 있다.



(그림 1) 분류 기법으로 생성한 의사결정트리의 예.

전체 블로그 연결망의 성향 판정은 기본적으로 블로그 연결망 내에 모든 관계가 가지는 성향 정도를 더함으로써

가능하다. 그러나 이런 단순한 방식은 블로그 연결망의 규모에 따라서 성향의 정도가 좌우되는 문제를 유발시킨다. 본 논문에서는 이러한 문제를 해결하기 위하여 연결망의 규모를 결정하는 블로그 수와 관계의 수를 이용하여 평균 성향의 정도를 구하는 기법을 제시한다.

첫 번째 평균 기법은 블로그 연결망 내에 있는 관계 성향 점수의 합을 블로그들의 수로 나눈다. 블로그의 수를 통한 평균 방안은 하나의 블로그가 가지고 있는 평균 성향의 정도를 의미한다. 이러한 평균 기법은 블로그 연결망에 정점이 적고 간선이 많을수록 높게 판정된다.

두 번째 평균 기법은 관계의 수로 전체 관계 성향 점수의 합을 나눈다. 관계의 수를 통한 평균 방안은 블로그 연결망의 관계의 평균 성향 점수를 의미한다. 이러한 평균 방안은 블로그 연결망 내에 성향 점수가 높은 관계들이 많을수록 높게 판정된다. <표 1>은 평균 기법을 수식으로 표현한 것이다.

<표 1> 블로그 연결망의 전체 성향 판정.

성향의 종류 기법 종류	정보 중시 성향	친분 중시 성향
블로그의 수를 통한 평균 기법	$PN_I = \frac{\sum PE_I}{N_v}$	$PN_F = \frac{\sum PE_F}{N_v}$
관계의 수를 통한 평균 기법	$PN_I = \frac{\sum PE_I}{N_e}$	$PN_F = \frac{\sum PE_F}{N_e}$
기호 설명	N_v : 블로그의 수 N_e : 관계의 수 PN_F : 연결망의 친분 중시 성향 점수 PN_I : 연결망의 정보 중시 성향 점수 PE_F : 관계의 친분 중시 성향 점수 PE_I : 관계의 정보 중시 성향 점수	

3. 실험

관계 성향 판정에 대한 정확도는 다음과 같이 두 가지 방식으로 측정한다. 첫째 의사결정트리에 의해 얻어진 성향 정도와 설문지를 통해서 얻은 실제 성향 정도 간에 일치 여부인 적중률을 측정한다. 둘째, 의사결정트리에 의해 얻어진 성향 정도와 설문지를 통해서 얻은 실제 성향 정도간의 오차를 예측단계와 실제단계의 차이로 측정한다. 적중률 외에도 오차를 측정하는 이유는 성향 정도를 상, 중, 하로 나타 낼 때 예측이 맞지 않는 경우에도 한 단계 다른 경우와 두 단계 다른 경우는 구분되어야 하기 때문이다.

3.1. 적중률

<표 2> 관계 성향 판정에 대한 적중률.

모델 종류 \ 단계수	3	5	7	9
친분 중시 성향	94%	93%	93%	91%
정보 중시 성향	93%	93%	93%	92%

<표 2>는 블로그 간 관계의 적중률을 측정한 결과를 나타낸다. 단계 수는 성향 정도를 몇 가지 종류로 나누었는가를 의미한다. 예를 들어, 단계 수 3은 성향 정도를 상, 중, 하 세 단계로 나눈 경우이다. 단계 수 9는 성향 정도를 아홉 단계로 나눈 경우이다. 표에서 볼 수 있듯이 친분 중시 성향과 정보 중시 성향 모두 단계 수와 상관없이 약 93%의 높은 적중률을 보인다.

3.2 오차

<표 3> 관계 성향 판정에 대한 오차.

모델 종류 \ 단계수	3	5	7	9
친분 중시 성향	0.07	0.14	0.18	0.23
정보 중시 성향	0.08	0.13	0.21	0.22

<표 3>은 블로그 간 관계의 오차를 측정된 결과이다. 결과에서 나타난 바와 같이 단계수가 증가함에 따라 오차도 증가한다. 단계수가 9일 때의 오차는 약 0.2로 나타났다. 이는 5번 성향 판정을 했을 때 5번 중에 4번 정확하게 예측하고, 1번은 틀리나 1단계 정도의 오차가 발생한다는 것을 의미한다. 위에서 실시한 두 가지 실험을 통해서 본 논문이 제안한 방안은 정확도와 오차 측면에서 모두 우수하다는 것을 검증하였다.

4. 결론

본 논문에서는 임의의 블로그 연결망의 성향을 판정할 수 있는 방안에 관하여 논의하였다. 블로그 연결망의 성향 판정 방안은 주어진 연결망 전체의 성향 판정 문제를 연결망이 가지고 있는 관계 성향 판정 문제로 분할하였다. 관계 성향 판정은 직접 관계에 참여하는 두 블로그 사용자에게 설문지를 통하여 그 정도를 판정하고, 그 결과를 데이터 마이닝의 분류 기법을 이용하여 자동화하였다. 또한, 블로그 연결망의 규모에 따라서 성향 정도가 좌우되는 문제를 해결하기 위해서 연결망의 규모를 결정하는 블로그 수와 관계의 수를 이용하여 평균 성향의 정도를 구하는 기법을 제시하였다. 본 논문에서 제안한 방안은 관계 성향 판정에 대해서 두 성향 모두 적중률은 약 91-94%이며 오차는 0.07-0.23으로 높은 정확도를 보였다.

감사의 글

본 연구는 NHN(주)의 지원을 받았습니다. 그러나, 본 논문에서 제시된 의견이나 결론, 또는 권고 등은 온전히 저자(들)의 것이며, 반드시 지원회사의 입장을 대변하는 것은 아닙니다.

참고문헌

- [1] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications, Cambridge University Press, 1994.
- [2] S. Redner, "How Popular Is Your Paper?," European Physics Journal B, Vol. 4, No. 2, pp. 131-134, 1998.
- [3] H. Jeong et al., "The Large-scale Organization of Metabolic Networks," Nature, Vol. 407, pp. 651-654, 2000.
- [4] M. Granovetter, "The Strength of Weak Ties," American Journal of Sociology, Vol. 78, No. 6, pp. 1360-1380, 1973.
- [5] S. Milgram, "The Small World Problem," Physiology Today, Vol. 2, pp. 60-67, 1967.
- [6] R. Kumar, J. Novak, and A. Tomkins, "Structure and Evolution of Online Social Networks," In Proc. Int'l. Conf. on Knowledge Discovery and Data, pp. 611-617, 2006.
- [7] A. Chin and M. Chignell, "A Social Hypertext Model for Finding Community in Blogs," In Proc. Int'l. Conf. on Hypertext and Hypermedia, pp. 11-22, 2006.
- [8] (주)위고넷 이브랜딩 전략연구소, 커뮤니티를 승부하는 브랜드 전략, 이디자인, 2004.
- [9] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Academic Press, 2001.