

블로그 연결망의 성향 판정 방안 (An Approach for Determining Propensities of Blog Networks)

윤석호[†] 박선주^{**} 김상욱^{***}
(Seok-Ho Yoon) (Sunju Park) (Sang-Wook Kim)

요약 블로그(blog)는 사용자가 글을 작성하여 게시할 수 있는 일종의 개인 웹사이트이며, 블로그는 다른 블로그와 관계를 맺을 수 있다. 본 논문에서는 블로그와 블로그 간의 관계들로 이루어진 연결망을 블로그 연결망(blog network)이라고 정의한다. 블로그 연결망은 그 내부에 게시된 글의 목적에 따라 정보 중시 성향과 친분 중시 성향을 가질 수 있다. 블로그 연결망이 가지고 있는 각 성향의 정도는 블로그 연결망의 비즈니스 정책 결정에 중요한 요소로서 사용된다. 본 논문에서는 주어진 블로그 연결망의 두 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. 본 논문에서는 제안하는 방안의 우수성을 검증하기 위하여 실제 블로그 데이터를 이용한 다양한 실험을 수행한다. 실험 결과에 의하면, 판정된 관계 성향의 정도는 두 성향에 대하여 모두 약 93%의 높은 정확도를 보였다. 또한, 실제 블로그 연결망의 각 성향의 정도를 판정함으로써 제안하는 방안이 정보 중시 성향과 친분 중시 성향의 연결망을 올바르게 판정할 수 있는 지를 보였다.

키워드 : 사회연결망분석, 데이터 마이닝, 블로그, 성향 판정

Abstract A blog is a personal website where its owner publishes his/her articles for others. A blog can have relationships with other blogs. In this paper, we define a network that is composed of blogs connected together with such relationships as a blog network. Blog networks can have two different propensities characterized by the articles published in the blogs: information-valued propensity and friendship-valued propensity. The degree of each propensity of a blog network plays an important role in deciding business policies for blog networks. In this paper, we address the problem of determining the degrees of two propensities of a given blog network. First, we determine the degree of the propensity of every relationship, a basic unit of a blog network, by using classification that is one of data mining functionalities. Then, by utilizing the result thus obtained, we compute the degrees of two propensities of the whole blog network. Also, we propose a method to solve the problem that the degree of propensities depends on the size of blog networks. To verify the superiority of the proposed approach, we perform extensive experiments using a huge volume of real-world blog data. The results show that our approach provides high accuracy of around 93% in determining the degrees of both propensities of relationships between arbitrary two blogs. We also verify the applicability of the proposed approach by showing that it determines the degrees of the information-valued and friendship-valued propensities correctly in real-world blog networks.

Key words : blog networks, social network analysis, propensity determination

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

† 학생회원 : 한양대학교 전자컴퓨터통신공학과
bogely@agape.hanyang.ac.kr
** 종신회원 : 연세대학교 경영학부 교수
boxenju@yonsei.ac.kr
*** 종신회원 : 한양대학교 정보통신학부 교수
wook@hanyang.ac.kr

논문접수 : 2007년 7월 18일
심사완료 : 2009년 1월 14일

Copyright©2009 한국정보과학회: 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

1. 서론

사회연결망(social network)은 한 사회 내에 존재하는 구성원들 간의 관계 정보를 이용하여 그 사회 전체의 관계 구조를 연결망으로 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[3,4], Granovetter[5], Milgram[6], Nowak[7], Watts[8,9] 등의 연구가 이러한 형태의 전형적인 연구이다.

인터넷의 발달로 인하여 온라인상에서도 이러한 사회연결망이 나타났다. 온라인 사회연결망(online social network)의 가장 큰 특징은 기존의 사회연결망과는 달리 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[10,11].

블로그는 사용자가 자신의 글을 온라인상에 저장할 수 있는 일종의 개인 웹사이트이며, 블로그 세계(blog world)는 이러한 블로그로 이루어진 대표적인 온라인 사회연결망이다[12-15]. 각 블로그 사용자는 타인의 블로그와 관계를 맺을 수 있으며, 이러한 관계들에 의하여 온라인 사회연결망이 형성된다[16]. 본 논문에서는 블로그와 블로그들 간의 관계들로 이루어진 사회연결망을 블로그 연결망(blog network)이라고 정의한다.

블로그 연결망은 그 내부에 존재하는 글의 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.

이러한 블로그 연결망의 성향은 비즈니스에서 다양하게 활용될 수 있다. 예를 들어 블로그 연결망에 정보를 목적으로 블로그 연결망을 이용하는 사용자가 많은 경우, 블로그 사용자들에 의해서 정보가 파급되는 현상을 이용한 바이러스성 마케팅(viral marketing)기법이 적용된다. 또한, 블로그 연결망에 친분을 목적으로 블로그 연결망을 이용하는 사용자가 많은 경우, 블로그 사용자들이 가지고 있는 친분 관계를 이용하여 기업의 상품 또는 서비스 정보 등을 전파하는 입소문 마케팅기법

(word of mouth marketing)을 적용한다[17-19].

따라서, 비즈니스의 대상이 되는 블로그 연결망의 두 성향 정도를 파악할 수 있다면, 블로그 연결망에 효과적이고 적절한 비즈니스 기법을 적용할 수 있을 것이다.

이와 연관된 기존 연구로는 사회연결망의 위상 구조적 특성을 이용해서 유사한 특징의 구성원들로 이루어진 집단을 찾는 연구[20], 사회연결망에서 임의의 사용자 그룹이 주어졌을 때 그 그룹을 형성하는데 관여하는 관계들과 그 관여 정도를 파악하는 연구[21], 그리고 뉴스 그룹을 사용하는 사람들을 특정한 의견에 대하여 찬성하는 사람들과 반대하는 사람들로 분류하는 연구[22] 등이 있었다. 이러한 연구들은 모두 사회연결망을 대상으로 이질적인 특징을 가지는 구성원 집단을 구분하는 점에서 공통점이 있으나, 본 논문에서 해결하고자 하는 성향 판정에 적용하기에는 부적합하다. 따라서 본 논문에서는 블로그 연결망의 성향 정도를 판정할 수 있는 다음과 같은 새로운 방안을 제안한다.

먼저, 블로그 연결망의 기본 단위인 블로그들 간의 관계에 대해서 성향의 정도를 판정하고, 그 결과를 이용해서 블로그 연결망이 가지고 있는 성향의 정도를 판정한다. 두 블로그들 간 관계의 성향 정도는 해당 블로그 사용자들이 스스로 판정할 수 있다. 따라서 블로그 사용자에게 자신이 직접적으로 참여하는 관계에 대해서 설문을 통하여 성향 정도를 질의하고, 그 결과를 이용해서 블로그 연결망이 가지고 있는 전체 성향 정도를 판정한다. 그러나 블로그 연결망에 대한 성향 판정이 요구 될 때마다 매번 블로그 사용자에게 성향의 정도에 관하여 질의 할 수는 없다. 본 논문에서는 성향 판정을 자동화하기 위해서 데이터 마이닝의 분류 기법(classification)[23]을 이용한다.

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.

본 논문에서는 제시한 성향 판정 방안의 정확도와 적용 가능성을 검증한다. 먼저, 블로그들 간의 관계에 대한 성향 판정의 정확도를 검증한다. 제안하는 방안은 약 91%에서 94%까지의 높은 정확도를 보였다. 또한, 특정 성향이 강한 블로그 연결망과 무작위로 추출한 블로그 연결망이 가지는 성향 정도를 서로 비교함으로써 제안된 성향 판정 방안의 적용 가능성을 검증한다. 그 결과, 특정 성향이 강한 블로그 연결망이 무작위로 추출된 블로그 연결망보다 특정 성향의 정도가 높은 것으로 나타났다.

본 논문의 구성은 다음과 같다. 제2장에서는 본 논문에서 해결하고자 하는 문제를 정의한다. 제3장에서는 관련 연구로서 본 논문과 연관된 기존 연구들을 소개한다. 제4장에서는 본 논문에서 제안하는 블로그 연결망의 성향 판정 방안에 대하여 상세히 설명한다. 제5장에서 다양한 실험을 통해서 본 논문에서 제안하는 방안에 대한 정확도를 검증한다. 마지막으로, 제6장에서는 본 논문에 대한 결론을 내리며, 향후 연구 방향을 제시한다.

2. 문제 정의

본 장에서는 블로그 연결망에 대하여 소개를 하고, 본 논문에서 해결하고자 하는 문제를 정의한다.

블로그는 사용자의 관심사에 따라 자유롭게 글을 올릴 수 있는 일종의 개인 웹사이트이다[12-15]. 블로그를 통해서 사용자가 할 수 있는 행동은 크게 자신의 블로그에 할 수 있는 행동과 타인의 블로그에 할 수 있는 행동으로 구분할 수 있다. 첫째로 사용자가 자신의 블로그에 할 수 있는 행동으로는 자신의 게시글을 보는 것, 자신의 블로그에 게시글을 작성하는 것, 자신의 블로그에 타인의 게시글에 대해서 링크를 만들거나 자신의 블로그로 타인의 게시글을 복사해서 가져오는 스크랩이 있다. 둘째, 사용자가 타인의 블로그에 할 수 있는 행동으로는 타인의 게시글을 보는 것, 타인의 게시글에 대해서 의견을 쓰는 댓글이 있다. 블로그 사용자는 관심이 있는 타인의 블로그와 자신의 블로그 사이에 관계를 맺을 수 있다. 이러한 관계로 인해서 블로그 세계에서는 온라인 사회연결망이 형성된다[16]. 본 논문에서는 블로그로 구성되어 있는 온라인 사회연결망을 블로그 연결망(blog network)이라 정의한다.

그림 1은 블로그 연결망을 도식화한 예이다. A~H로 표현된 둥근 사각형은 각 블로그를 나타내며, a~g로 표현된 작은 사각형은 블로그에 저장되어 있는 게시글을 의미한다. 실선 화살표는 타인의 블로그에 있는 게시글

에 대해서 자신의 블로그에 링크를 만들었거나 스크랩을 한 것을 의미한다. 예를 들어, 블로그 D의 게시글 c는 블로그 A의 게시글 c에 대하여 링크를 만들거나 스크랩을 한 경우로 게시글의 내용이 같다. 블로그와 블로그 사이에 있는 실선은 블로그와 블로그 사이의 관계가 존재함을 의미한다. 예를 들어, 블로그 A와 블로그 D는 서로 관계를 맺었음을 알 수 있다.

블로그 연결망은 그 내부에 존재하는 글의 목적에 따라 구분되는 서로 다른 두 가지 성향을 가질 수 있다. 어떤 블로그 연결망에서는 사용자들이 주식, 여행 등과 관련된 정보성 글을 작성하고 수집한다. 본 논문에서는 정보의 전파와 수집을 목적으로 글을 주고받는 성향을 정보 중시 성향이라고 정의한다. 반면, 어떤 블로그 연결망에서는 사용자들이 안부, 소식 등과 같은 사용자간의 관계를 유지하고 발전시키기 위한 글을 작성한다. 본 논문에서는 이와 같이 사용자간에 친분을 목적으로 글을 주고받는 성향을 친분 중시 성향이라고 정의한다.

본 논문에서 정의한 두 성향 중 정보 중시 성향이 강한 블로그 연결망의 대표적인 예는 마이스페이스이다[24]. 마이스페이스는 사용자가 필요로 하는 다양한 정보를 저장하고 검색할 수 있도록 정보 검색 및 정보 관리 기능을 제공하고 있다. 친분 중시 성향이 강한 블로그 연결망의 대표적인 예는 한국의 싸이월드이다[12]. 싸이월드는 사용자가 자신과 친분이 있는 사용자와 친분이 없는 사용자를 차별화 할 수 있는 기능과 사용자들 사이에 주고 받을 수 있는 사이버 선물을 제공하고 있다.

기업들은 블로그 연결망에 적용 가능한 비즈니스 기법을 선정하기 위해서 해당 블로그 연결망이 가지고 있는 다양한 요소들을 고려한다. 예를 들어, 블로그 연결망의 규모, 블로그 사용자들의 연령 분포 등을 고려한다. 대표적으로 고려되는 요소는 블로그 연결망의 성향 정도이다.

블로그 연결망의 성향 정도를 고려한 비즈니스의 예

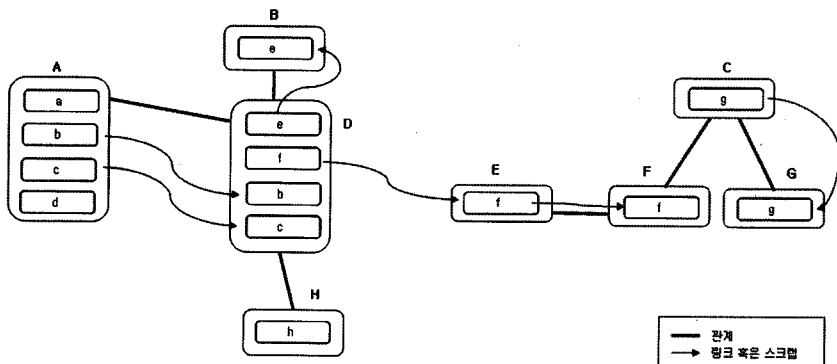


그림 1 블로그 연결망의 예

는 다음과 같다. 먼저, 블로그 연결망에 정보를 목적으로 블로그 연결망을 이용하는 사용자가 많은 경우, 블로그 사용자들에 의해서 정보가 파급되는 현상을 이용한 바이러싱 마케팅(viral marketing)기법이 적용된다. 바이러싱 마케팅은 기존의 마케팅 기법 중 하나인데, 블로그 연결망에서의 바이러싱 마케팅은 블로그 사용자가 타인의 블로그와 글을 만드는 링크를 통하여 정보가 빠르게 전파되는 현상을 이용한다. 즉, 기업의 상품과 서비스에 대한 홍보 자료를 정보성 글로 생성하여 블로그 연결망 전체에 빠르게 홍보 한다.

둘째로, 블로그 연결망에 친분을 목적으로 블로그 연결망을 이용하는 사용자가 많은 경우, 블로그 사용자들이 가지고 있는 친분 관계를 이용하여 기업의 상품 또는 서비스 정보 등을 전파하는 입소문 마케팅기법을 적용한다. 입소문 마케팅(word of mouth marketing)기법 또한 기존의 마케팅 기법 중 하나이다. 이러한 기법은 친분을 목적으로 블로그 연결망을 이용하는 사용자들이 많은 블로그 연결망에서 블로그 사이트 또는 블로그 사이트 내에서 제공하는 서비스 가입 시에 사용자에게 친구의 가입을 권유하는 방법으로 많은 수의 사용자들을 확보하는데 이용된다. 또한 친분 관계가 많은 사용자들을 집중적으로 홍보하여 적은 홍보 비용으로 좋은 결과를 얻고 있다.

따라서, 두 성향을 이용한 비즈니스 기법을 블로그 연결망에 효과적이고 적절하게 적용하기 위해서는 비즈니스의 대상이 되는 블로그 연결망의 두 성향 정도를 파악해야 한다.

그림 2는 블로그 연결망을 그래프 G로 표현한 예이다. 정점은 각 블로그를 의미하며, 실선은 블로그들 간의 관계를 의미한다. G_1 은 G의 부분 연결망을 나타낸다. 주어진 부분 연결망 G_1 에 대한 성향 판정을 통하여 부분 연결망 G_1 의 정보 중시 성향 정도와 친분 중시 성향 정도를 파악할 수 있으며, 이 결과를 다양한 비즈니스 정책 결정에 활용할 수 있다.

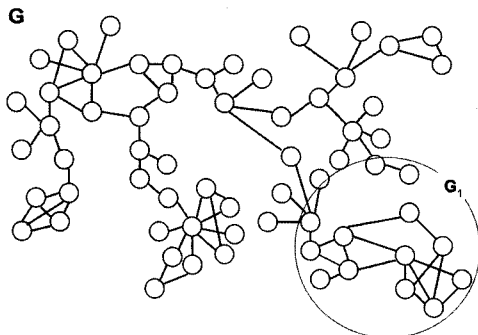


그림 2 성향 판정을 위한 블로그 연결망의 예

3. 관련 연구

본 장에서는 성향 판정과 관련이 있는 기존의 연구 결과들을 요약하고, 이러한 결과들을 본 연구에 적용 가능한지의 여부를 논의한다.

참고문헌 [22]에서는 뉴스 그룹에서 게시된 특정 의견에 대해 찬성하는 그룹과 반대하는 그룹으로 저자들을 분류하는 기법을 제시하였다. 이 기법은 임의의 글을 인용한 사람은 이 글의 의견에 반대되는 의견을 개진하는 경향이 강하다라는 이론을 기반으로 한다. 예를 들어, a라는 글이 뉴스 그룹에 게시되었을 때, 이 글을 인용해서 b라는 글을 쓴 저자는 a라는 글을 쓴 저자와 상반된 의견을 가지고 있다고 해석한다. 또한, b라는 글을 인용해서 c라는 글을 쓴 저자는 b의 저자와 의견이 상반된다고 해석한다. a와 c의 두 저자는 모두 b의 저자와 상반된 의견을 가지므로 a와 c는 같은 의견으로 간주한다. 이러한 기법을 통하여 뉴스 그룹에 글을 게시한 저자들을 특정 글에 대해서 찬성하는 저자들과 반대하는 저자들로 구분할 수 있다.

이 연구는 인용이라는 관계를 통해서 형성된 사회 연결망을 분석하여 저자들을 분류하고 있다. 블로그 연결망의 성향 판정 문제는 블로그 연결망의 성향 정도를 판정하는 문제이므로 일종의 분류 문제로 해석되어 질 수 있다는 점에서 이 문제와 유사한 부분이 있다. 그러나 성향 판정 문제는 블로그 연결망을 정보 중시 성향과 친분 중시 성향들 중 하나로 판정하는 것이 아니라 두 성향 모두에 대해서 성향 정도를 판정하는 문제라는 점에서 차이가 있다. 또한, 이 연구는 저자를 분류하는 문제를 이론에 근거해서 해결하고 있다. 이는 구체적인 데이터가 존재하는 블로그 연결망에 적용하기에는 부적합하다.

참고문헌 [20]에서는 사회연결망에 존재하는 클러스터들을 찾아내는 기법을 제시하였다. 클러스터란 유사한 특징의 구성원들로 이루어진 집단을 의미한다[23]. 이 연구는 클러스터와 클러스터 사이에는 약한 연결만이 존재한다는 사회연결망의 특징을 이용해서 클러스터를 파악한다[5]. 클러스터는 블로그 연결망에서 하나의 성향이 강한 부분 블로그 연결망과 대응되는 개념이다. 그러나 이 논문에서 제시한 기법은 사회연결망의 위상 구조적 특성만을 가지고 클러스터를 찾아내는 방식을 사용한다. 블로그 연결망의 성향은 내부에 게재된 글에 의해 나타나므로 위상 구조적 특성만으로 블로그 연결망의 성향을 판정하는 것은 좋은 해결 방안이 될 수 없다.

참고문헌 [21]에서는 다양한 관계가 존재하는 사회 연결망과 임의의 그룹이 주어졌을 때, 임의의 그룹을 형성하는데 관여하는 관계들을 도출하고, 도출된 각 관계가

해당 그룹을 형성하는데 얼마나 관여했는지 파악하는 기법을 제시하였다. 이 연구에서는 사회연결망 내 구성원 간에는 다양한 형태의 관계와 그 관계의 정도가 존재한다고 가정한다. 이것은 본 논문에서 판정하고자 하는 각 성향과 정도에 대응되는 개념으로 볼 수 있다. 따라서 이 연구에서 필요로 하는 다양한 형태의 관계와 그 관계의 정도를 본 논문에서 제시하는 방안을 이용하여 제공할 수 있다.

위에서 살펴본 기존의 연구에서 해결하고자 한 문제들은 블로그 연결망의 성향 판정 문제와는 차이점이 있기 때문에 기존의 기법들을 성향 판정 문제에 적용하기에는 무리가 있다. 따라서 본 논문에서는 블로그 연결망을 성향 판정 할 수 있는 새로운 방안을 제시하고자 한다.

4. 블로그 연결망의 성향 판정 방안

4.1 개관

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. 블로그 사용자는 자신이 연결망 내에서 직접 주고받는 글에 대해서만 글의 목적을 파악 할 수 있으나, 연결망 내의 자신과 직접 관련되지 않는 글들에 대해서는 목적을 파악 할 수 없다. 따라서 본 논문에서는 먼저 블로그 연결망의 기본 단위원 하나의 관계의 성향을 그 관계에 참여하는 두 블로그 사용자들에게 판정하게 하고, 이런 방식으로 판정된 관계들의 성향을 이용해서 전체 블로그 연결망의 성향을 판정하는 방식을 제안한다.

블로그 연결망의 성향 판정이 요구될 때마다 매번 구성원인 블로그 사용자에게 성향 판정을 요구할 수는 없다. 본 논문에서는 두 블로그 간 관계의 성향 판정을 위하여 데이터 마이닝의 분류 기법을 사용한다[23]. 즉, 블로그 사용자들에 의해서 실제 성향 정도가 제시된 관계 집합을 학습 데이터로 이용하여 학습을 수행하고, 학습된 결과를 이용하여 새로 주어진 블로그 연결망의 성향을 자동적으로 판정하는 방식을 채택한다.

4.2 관계 성향 판정

4.2.1 설문 조사

본 논문에서는 두 블로그 간 관계 성향을 판정하기 위해서 설문 조사를 실시한다. 사용자는 자신이 직접적으로 관여하는 관계에 대한 성향 정도에 대해서 알 수 있다. 그러나 성향의 정도에 대해서 각 사용자마다 주관적으로 판단하므로 객관적으로 성향 정도를 판정할 수 있는 기준이 필요하다. 따라서 본 설문 조사에서는 각 관계에 직접적으로 관여하는 두 블로그 사용자에게 각

성향이 강하다고 판단되는 글의 종류를 제시하고, 제시된 글을 사용자와 직접적으로 관계가 있는 다른 블로그에서 얼마나 자주 보았는지 질의한다. 그 후에 설문 조사 결과를 계량화함으로써 각 관계의 실제 성향 정도를 판정한다. 설문은 정보 중시 성향에 대해서 물어보는 문항들과 친분 중시 성향에 대해서 물어보는 문항들로 구성된다. 정보 중시 성향에 대한 문항과 친분 중시 성향에 대한 문항은 각각 세 개씩이다.

4.2.2 클래스 레이블 부여 절차

분류 모델을 만들기 위해서는 먼저 판정하고자 하는 대상을 정해야하는데, 이를 클래스 레이블(class label)이라고 한다[23]. 설문 조사를 통해서 얻고자 하는 결과가 관계 성향 정도이므로 본 연구에서의 클래스 레이블은 각 관계의 성향 정도이다. 본 절에서는 설문 조사를 통해서 얻은 결과를 이용해서 클래스 레이블을 결정하는 절차에 대해서 논의한다.

특정 종류의 글을 얼마나 보았는지에 대한 응답을 이용하여 성향 정도를 판정하기 위해서는 각 보기에 성향 점수가 부여되어야 한다. 점수 설정은 도메인 전문가에 의해서 응용 분야에 따라서 다르게 설정할 수 있다. 표 1은 본 논문에서 설정한 점수를 나타낸다.

표 2는 성향의 정도를 점수로 변환하는 공식을 나타낸다. 이 공식은 성향에 따른 문항들에 대하여 사용자가 선택한 보기의 점수(W_i)를 합한 값이 총 성향 점수가 됨을 보여준다. 본 논문에서 실시한 설문 조사에서는 각 성향에 대한 문항이 3개씩 존재했으므로 3개의 문항에 대한 점수의 합으로 각 관계의 성향 정도를 판정할 수 있다. 예를 들어, 어떤 블로그 사용자가 친분 중시 성향에 대한 세 문항에 대해서 모두 “하루에 한번 이상”이라고 답하고, 정보 중시 성향에 대한 세 문항에 대해서 모두 “한 달에 한번 이상”이라고 답한 경우, 친분 성향 점수는 $90(3 \times 30)$ 점이 되고 정보 중시 성향 점수는

표 1 보기에 대한 성향 점수

“하루에 한번 이상” : 30	“일주일에 한번 이상” : 15
“한 달에 한번 이상” : 5	“총 한 두 번 이상” : 1
“없다” : 0	

표 2 설문 조사를 이용한 성향 점수의 계산

정보 중시 성향 점수	친분 중시 성향 점수
$PE_I = \sum_i^{N(Q_I)} W_i$	$PE_R = \sum_i^{N(Q_R)} W_i$
i: 문항 번호, W_i : i번 문항의 점수 PE_R : 관계의 친분 중시 성향 점수, PE_I : 관계의 정보 중시 성향 점수 $N(Q_R)$: 친분 중시 성향에 대한 문항 수, $N(Q_I)$: 정보 중시 성향에 대한 문항 수	

15(3×5)점이 된다.

그러나 하나의 관계에 대해서 두 명의 블로그 사용자에게 설문 조사를 실시하였으므로 그 관계 맺고 있는 두 명의 설문 결과를 합하고 평균을 내어서 관계 성향 점수를 산출한다. 또한 분류 모델을 만들기 위해서 산출된 성향 점수를 클래스 레이블로 변환한다. 예를 들어, 클래스 레이블을 상, 중, 하로 정할 수 있다. 그러나 응용의 성격에 따라서 좀 더 세분화된 클래스 레이블을 이용할 수 있다.

4.2.3 분류를 위한 관계 속성의 생성

본 논문에서는 분류 모델을 만들기 위해서 가장 보편적인 방법인 의사결정트리(decision trees)를 사용한다. 이외에도 신경망, 베이시안 분류, 지네틱 알고리즘 등과 같은 다른 분류 기법을 이용하는 것도 가능하다.

의사결정트리를 생성할 때 의사 결정에 기준이 되는 속성의 선택은 대단히 중요하다[23]. 본 논문에서 판정하고자 하는 것이 관계 성향의 정도이기 때문에 관계의 특성을 설명할 수 있는 속성이 필요하다. 그러나 블로그 연결망에는 관계에 대한 데이터는 없고 블로그들에 대한 데이터만 존재한다. 따라서 본 연구에서는 하나의 관계에 속한 두 블로그의 데이터를 이용해서 관계에 대한 속성을 생성한다. 생성된 전체 관계 속성의 수는 30개다.

4.2.4 분류 모델의 생성

관계 성향의 정도는 정보 중시 성향과 친분 중시 성향 모두에 대해서 판정되어야 하기 때문에 각 성향에 대해서 의사결정트리를 생성한다. 그림 3은 친분 중시 성향을 판정하기 위하여 생성된 의사결정트리의 일부를 나타낸다. 맨 위에 있는 속성이 덧글의 수이므로 덧글의 수가 친분 중시 성향을 판정하는데 가장 큰 기준이 된

다는 것을 알 수 있다. 두 번째 중요한 속성은 주고받은 선물의 수이다. 블로그 연결망에서 선물은 아는 사람들 사이에서 많이 이용된다. 따라서 선물을 주고받은 횟수가 친분 중시 성향을 판정하는데 중요한 기준이 된다.

그림 4는 정보 중시 성향을 판정하기 위한 의사결정트리의 일부를 나타낸다. 맨 위에 있는 속성이 덧글의 수이므로 덧글의 수가 정보 중시 성향을 판정하는데 가장 큰 기준이 된다는 것을 알 수 있다. 두 번째 중요한 속성은 스크랩 수이다. 정보 수집을 목적으로 하는 사람은 타인의 블로그에서 관심 있는 글을 복사해오는 경우가 많다. 따라서 스크랩 수는 정보 중시 성향을 판정하는데 중요한 기준이 된다.

블로그 사용자에게 의해서 성향이 판정된 관계들을 가지고 친분 중시 성향과 정보 중시 성향의 정도를 판정하기 위한 두 가지 의사결정트리들이 생성된다. 이후, 이와 같이 생성된 의사결정트리들을 통해서 성향이 판정되지 않은 관계가 주어지면, 이 관계에 대한 각 성향의 정도를 판정 할 수 있다.

4.3 연결망 성향 판정

본 절에서는 제4.2절에서 제시한 기법으로 구한 관계들의 성향 정도를 이용하여 주어진 연결망의 전체 성향 정도를 판정하는 기법에 대하여 언급한다.

판정된 각 관계 성향 정도는 점수화 할 수 있다. 예를 들어, 클래스 레이블을 상, 중, 하가 아닌 100, 50, 0으로 정할 수 있다. 관계 성향 정도가 점수화되면 그 점수의 합을 통해서 주어진 블로그 연결망의 성향 정도를 계산할 수 있다. 그러나 연결망 성향 정도를 관계 성향 점수의 합으로 계산하는 경우, 연결망의 규모에 따라서 성향의 정도가 좌우되는 문제가 발생할 수 있다. 그림 5

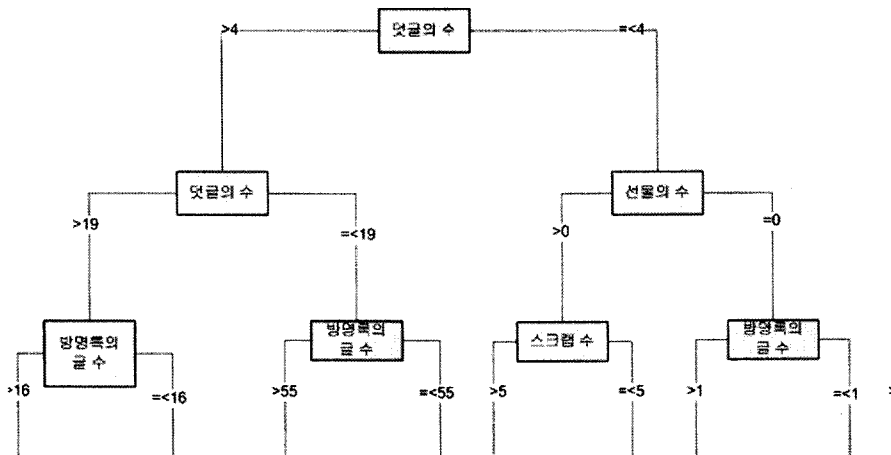


그림 3 친분 중시 성향 판정을 위한 의사결정트리의 예

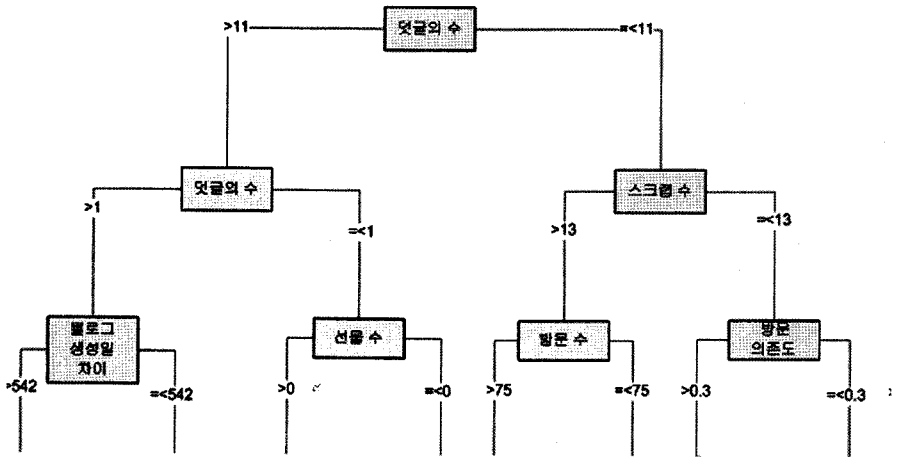
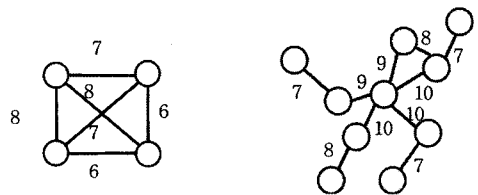


그림 4 정보 중시 성향 판정을 위한 의사결정트리의 예

는 블로그 연결망 A와 블로그 연결망 B에서 나타나는 규모 문제의 예를 나타낸다. 그래프의 각 정점들은 블로그를 의미하며 간선은 블로그 간 관계를 의미한다. 또한, 각 간선에 표시된 숫자는 관계 성향 점수를 의미한다. 이 때, 블로그 연결망 A는 정점과 간선의 수가 적지만 정점들 간에 간선의 수는 많아서 블로그 연결망 B보다 밀집도가 높다. 그에 비해 블로그 연결망 B는 정점과 간선의 수가 많지만 정점들 간에 간선의 수는 적어서 밀집도가 블로그 연결망 B에 비해 낮다. 만약 블로그 연결망의 성향 판정을 단순히 관계 성향 점수의 합만으로 할 경우 블로그 연결망 A가 블로그 연결망 B보다 성향의 정도가 낮은 것으로 판정된다. 따라서, 본 논문에서는 이와 같이 블로그 연결망의 규모에 따라서 연결망의 성향 정도가 좌우되는 문제를 해결하기 위해서 연결망의 규모를 결정하는 블로그의 수와 관계의 수로 연결망 성향 정도의 평균값을 계산하는 두 가지 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. 예를 들어, 그림 5에서 블로그 연결망 A는 정점의 수가 4개이고 간선들의 성향 점수의 합이 42이므로 블로그의 수를 통해 평균을 계산하면 성향 점수는 10.5로 판정된다. 블로그 연결망 B는 정점의 수가 10개 간선들의 성향 점수의 합이 85이므로 성향 점수는 8.5로 판정된다. 블로그의 수를 이용한 평균 기법을 사용할 경우 블로그 연결망 A가 블로그 연결망 B보다 성향 점수가 높



블로그연결망 A(성향 점수 42) 블로그연결망 B(성향 점수 85)
그림 5 관계 성향 점수의 합으로 계산한 연결망의 성향 점수

게 판정된다. 이는 블로그 연결망 A가 가지고 있는 하나의 정점에 연결되어 있는 간선의 수가 블로그 연결망 B의 경우보다 많기 때문이다.

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. 예를 들어, 그림 5에서 블로그 연결망 A는 간선의 수가 6개이고 간선들의 성향 점수의 합이 42이므로 관계의 수를 통해 평균을 계산하면 성향 점수는 7로 판정된다. 블로그 연결망 B는 간선의 수가 10개이고, 간선들의 성향 점수의 합이 85이므로 성향 점수는 8.5로 판정된다. 이는 블로그 연결망 B가 가지고 있는 간선의 성향 점수가 평균적으로 블로그 연결망 A의 경우보다 높기 때문이다.

표 3은 블로그 연결망의 성향 판정을 위한 기법들을 나타낸다. 두 기법 중 어느 기법을 사용할 것인가는 응용 분야의 특성을 고려하여 결정한다. 예를 들어, 블로그

표 3 블로그 연결망의 전체 성향 판정

블로그의 수를 통한 평균 기법	관계의 수를 통한 평균 기법
$PN_R = \frac{\sum PE_R}{N_v}$ $PN_I = \frac{\sum PE_I}{N_v}$	$PN_R = \frac{\sum PE_R}{N_e}$ $PN_I = \frac{\sum PE_I}{N_e}$
N_v : 블로그의 수 N_e : 관계의 수 PN_R : 연결망의 친분 중시 성향 점수 PN_I : 연결망의 정보 중시 성향 점수 PE_R : 관계의 친분 중시 성향 점수 PE_I : 관계의 정보 중시 성향 점수	

그의 수를 통한 평균 기법은 사이버 아이템을 이용한 연결망 활성화 전략에 사용될 수 있다. 사이버 아이템은 주로 블로그들 간의 관계가 많으면서 관계의 친분 중시 성향이 높은 블로그 연결망에서 활성화된다. 따라서 사이버 아이템을 이용한 연결망 활성화 전략은 블로그의 수를 통한 평균 기법에 의해서 성향 점수가 높게 판정된 연결망을 대상으로 활성화 전략을 적용할 수 있다. 관계의 수를 통한 평균 기법은 정보성 콘텐츠를 이용한 연결망 활성화 전략에 사용될 수 있다. 정보성 콘텐츠를 이용한 연결망 활성화 전략은 주로 정보성 콘텐츠가 많은 사람에게 노출되어야 한다. 따라서 정보성 콘텐츠를 이용한 연결망 활성화 전략은 관계의 수를 이용한 평균 기법에 의해서 성향 점수가 높게 판정된 블로그 연결망에 적용하기에 적합하다.

5. 성능 평가

본 장에서는 제안한 성향 판정 방안에 대한 타당성을 실험을 통해서 검증하고자 한다. 제 5.1절에서는 실험을 위한 환경을 기술하고, 제 5.2절에서는 실험 결과를 분석한다.

5.1 실험 환경

본 논문에서는 성향 판정 방안에 대한 타당성을 검증 을 위해서 다음과 같은 실험 환경을 구축했다. 실험용 PC의 하드웨어 환경은 3.0G의 CPU를 가진 컴퓨터를 사용하였으며, 소프트웨어 환경은 Windows XP Professional과 MS-SQL 2000을 사용하였다.

설문 조사의 대상자는 블로그 연결망 내에 존재하는 모든 관계들 중에서 조사 기간 중에 글을 주고받은 적이 있는 35,000 관계를 무작위로 선택했다. 설문 조사는 블로그에서 직접 이루어졌다. 설문 조사에 대해서 실제 블로그 데이터의 유사성을 비교하는 전처리 과정을 수행했다. 그 결과 유효한 관계의 수는 1408개였다.

본 실험에서 사용한 데이터는 2006년에 수개월간 수집된 블로그 데이터를 사용하였으며, 데이터 수집 기간

동안 자신 또는 타인의 블로그에 한 번도 방문하지 않은 블로그 사용자는 실험에서 제외하였다.

의사결정트리(C5.0 알고리즘)를 이용하였으며 입력 데이터로 선물의 수, 스크랩 수, 댓글의 수, 방명록에 남긴 글의 수, 블로그 생성일의 차이 등 총 30개의 관계 속성을 사용하였다. 의사결정트리에 대한 검증은 전체 관계 정보들을 10개의 집합으로 나누고, 그 중에 9개를 학습 데이터로 1개를 테스트 데이터로 사용하여 각 단계마다 학습 데이터와 테스트 데이터를 바꾸어 10번에 걸쳐 정확도를 측정한다. 최종 정확도는 각 단계마다 측정된 정확도의 평균이다. 또한, 연결망 성향 판정에 대한 실험을 위하여 실제 블로그 세계에서 각 성향이 강하다고 판단되는 블로그 연결망들을 추출하였다.

블로그 연결망의 성향 판정 방안에 대한 검증은 관계 성향 판정과 연결망 성향 판정에 대해서 이루어진다. 관계 성향 판정에 대한 정확도는 다음과 같이 두 가지 방식으로 측정한다. 첫째는 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. 예를 들어 성향 정도를 총 5단계로 나눈다면 최대 오차는 4가 된다.

$$\text{오차} = |\text{예측단계} - \text{실제단계}|$$

주어진 연결망의 전체 성향 판정은 실제 성향 정도를 파악할 수 없으므로 직접적으로 정확도를 측정하기 어렵다. 따라서 본 논문에서는 한 쪽 성향이 강하다고 추측되는 각각의 블로그 연결망과 무작위로 선택된 블로그 연결망에 대해서 성향을 판정하고, 그 결과를 상호 비교함으로써 본 논문에서 제안하는 방안의 적용 가능성을 보이고자 한다.

본 논문에서는 블로그에서 정보 중시 성향이 강한 블로그 연결망을 추출하기 위해 다음과 같은 방식을 사용한다. 먼저, 다른 사용자가 스크랩을 많이 한 글을 작성한 블로그를 선택한다. 선택된 블로그와 관계를 맺고 있는 블로그들 중에서 선택된 블로그가 가지고 있는 글을 스크랩한 적이 있는 블로그들을 추출한다. 추출된 블로그들 중에서 다른 블로그에 방문한 수가 많은 순으로 100개의 블로그들을 선택한다. 친분 중시 성향이 강한 연결망은 다른 블로그들과 상호 관계를 많이 맺고 있는 블로그를 선택한다. 선택한 이 블로그를 중심으로 상호 관계를 맺고 있는 블로그들을 추출한다. 추출된 블로그

들 중에서 방문수가 많은 순으로 100개의 블로그들을 선택한다. 무작위 연결망은 무작위로 선택한 블로그들 중심으로 이 블로그와 관계를 맺고 있는 100개의 블로그들을 추출한다.

그림 6은 실험에 사용하기 위해 블로그 연결망을 추출하는 방법을 그림으로 나타낸 것이다. 각 화살표는 관계를 맺은 방향을 의미하며 화살표가 한쪽으로 되어 있는 것은 이웃을 의미하고, 양쪽으로 되어 있는 것은 상호 이웃을 의미한다. 정보 중시 성향이 강한 연결망과 무작위로 추출한 블로그 연결망 그리고 친분 중시 성향이 강한 연결망과 무작위로 추출한 블로그 연결망의 성향 판정 점수를 비교하여 제안된 블로그 연결망의 성향 판정 기법의 적용 가능성을 검증한다. 실험에서는 친분 중시 성향이 강한 연결망을 친분 중시 연결망, 정보 중시 성향이 강한 연결망을 정보 중시 연결망, 무작위로 블로그를 추출한 블로그 연결망을 무작위 연결망이라고 한다.

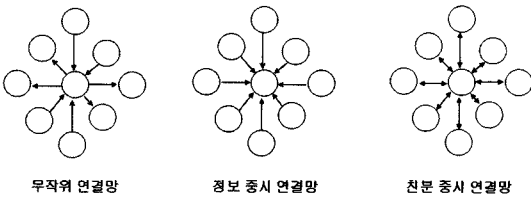


그림 6 연결망 성향 판정의 실험을 위한 연결망 추출

5.2 결과 분석

본 장에서는 블로그 연결망의 성향 판정 방법에 대한 정확도 측정 결과를 분석한다.

5.2.1 관계 성향 판정의 정확도 검증

5.2.1.1 적중률

적중률은 분류 모델에 의해 얻어진 성향 정도와 설문지를 통해서 얻은 실제 성향 정도 간에 일치 여부를 계산한 것이다. 표 4는 블로그 간 관계의 적중률을 측정된 결과를 나타낸다. 단계 수는 성향 정도를 몇 가지 종류로 나누었는가를 의미한다. 예를 들어서 단계 수 3은 성향 정도를 상, 중, 하 세 단계로 나누는 경우이다. 단계 수 9는 성향 정도를 아홉 단계로 나누는 경우이다. 표에서 볼 수 있듯이 친분 중시 성향과 정보 중시 성향 모두 단계 수와 상관없이 약 93%의 높은 적중률을 보인다. 단계의 수가 늘어남에 따라 적중률이 약간 감소하는데 이는, 단계의 수가 늘어남에 따라 관계가 가질 수 있는 클래스 레이블의 종류가 많아지기 때문이다. 그러나 제안한 기법은 큰 영향 없이 높은 정확도를 보인다. 따라서 본 논문에서 제안하는 방안의 정확도가 높다는 것이 입증된다.

표 4 관계 성향 판정에 대한 적중률

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

5.2.1.2 오차

오차는 분류 모델에 의해 얻어진 성향 정도와 설문지를 통해서 얻은 실제 성향 정도 사이의 차이이다. 표 5는 블로그 간 관계의 오차를 측정된 결과이다. 결과에서 나타난 바와 같이 단계수가 증가함에 따라 오차도 증가한다. 단계수가 9일 때의 오차는 약 0.2로 나타났으며, 이는 실제 성향 정도와 분류 모델이 제시한 성향 정도의 차이가 ±0.2인 것을 의미한다. 즉, 실제 성향 정도가 50이라면 분류 모델의 예측은 대략 49.8~50.2 사이가 된다. 단계의 수가 증가함에 따라 오차도 조금씩 증가하는데 이는 단계수가 3일 때는 최대오차가 2이나 단계수가 9일 때는 최대오차가 8이기 때문이다. 본 실험을 통해서 제안한 방안의 오차가 적다는 것을 알 수 있다.

표 5 관계 성향 판정에 대한 오차

단계수 \ 모델 종류	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

5.2.2 연결망 성향 판정의 타당성 검증

표 6은 연결망 성향 판정의 적용 가능성을 검증하기 위해 친분 중시 연결망과 무작위 연결망을 비교한 결과를 나타낸다. 친분 중시 연결망은 두 가지 평균 기법으로 얻어진 성향 정도가 무작위 연결망과 비교하여 모두 높게 나타났다. 값의 차이는 블로그 수를 통한 평균의 경우 1.19이고, 관계의 수를 통한 평균의 경우 0.27로서 친분 중시 연결망이 무작위 연결망보다 각 평균 기법에서 37%와 20% 높은 값을 가지는 것을 의미한다. 따라서 친분 중시 성향에 대해서 블로그 수를 통한 평균과 관계 수를 통한 평균 방안 모두 사용 가능하다.

표 7은 정보 중시 연결망과 무작위 연결망을 비교한 결과를 나타낸다. 정보 중시 연결망은 관계 수를 통한 평균 성향 점수가 무작위 연결망과 비교하여 0.29(21%) 높게 나타났다. 그러나 블로그 수를 통한 평균 성향 점

표 6 친분 중시 성향 판정 결과에 대한 비교

연결망 종류 \ 평균 방안	블로그 수를 통한 평균	관계 수를 통한 평균
무작위 연결망	3.18	1.34
친분 중시 연결망	4.37	1.61

표 7 정보 중시 성향 판정 결과에 대한 비교

연결망 종류 \ 평균 방안	블로그 수를 통한 평균	관계 수를 통한 평균
무작위 연결망	3.19	1.38
정보 중시 연결망	2.50	1.67

수는 무작위 연결망보다 오히려 0.69(27%) 낮게 나타났다. 이는 정보 중시 성향은 정보의 수집과 전달을 목적으로 하기 때문에 동일한 정보를 가지고 있는 블로그들에 대해서 관계를 맺을 필요가 없다. 즉, 정보를 많이 가지고 있는 블로그를 중심으로 성형(star)으로 형성되어 있기 때문에 블로그 연결망에 관계가 무작위 연결망보다도 많지 않다. 따라서 관계의 수를 통한 평균 기법을 이용하는 경우, 성향 정도가 낮아지는 경향이 있다. 블로그 연결망에서 정보 중시 성향을 판정할 시에는 관계 수를 통한 평균 방안을 사용할 것을 추천한다.

6. 결론 및 향후 과제

블로그 연결망에는 블로그들의 행동 양식을 결정하는데 중요한 요소가 되는 두 가지 성향인 정보 중시 성향과 친분 중시 성향이 존재한다. 블로그 연결망을 효과적으로 활성화하기 위해서는 블로그 연결망의 성향 정도를 판정하는 것이 매우 중요하다. 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 설문을 통하여 그 정도를 판정하고, 그 결과를 데이터 마이닝의 분류 기법을 이용하여 자동화하였다. 분류 기법을 통해서 판정된 관계 성향 정도를 이용하여 전체 블로그 연결망의 성향 정도를 판정하였다.

전체 블로그 연결망의 성향 판정은 기본적으로 블로그 연결망 내에 모든 관계가 가지는 성향 정도를 더함으로써 가능하다. 그러나 이런 단순한 방식은 블로그 연결망의 규모에 따라서 성향 정도가 좌우되는 문제를 유발시킨다. 따라서 본 논문에서는 이러한 문제를 해결하기 위하여 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. 주어진 연결망의 전체 성향 판정은 실제 성향 정도를 파악할 수 없으므로 직접적으로 정확도를 측정하기 어렵다. 따라서 본 논문

에서는 한 쪽 성향이 강하다고 추측되는 각각의 블로그 연결망과 무작위로 선택된 블로그 연결망에 대해서 성향을 판정하고, 그 결과를 상호 비교함으로써 본 논문에서 제안하는 방안의 적용 가능성을 보였다. 그 결과, 주어진 연결망의 성향 판정 결과는 친분 중시 연결망이 무작위 연결망보다 관계의 수를 통한 평균 기법에서 0.27(20%) 높았고 블로그 수를 통한 평균 기법에서 1.19(37%) 높았다. 정보 중시 연결망은 무작위 연결망보다 관계의 수를 통한 평균 기법은 0.29(21%) 높았으나, 블로그 수를 통한 평균 기법은 0.69(27%) 낮았다. 연결망을 성향 판정함에 있어서 친분 중시 성향 판정에 대해서는 블로그 수를 통한 평균과 관계 수를 통한 평균 기법 모두 사용 가능하며, 정보 중시 성향 판정에 대해서는 관계 수를 통한 평균 방안만이 사용 가능함을 보였다.

본 연구에서는 주어진 블로그 연결망의 성향 정도를 판정하는 방안에 대하여 논의하였다. 향후 과제로서 제안된 성향 판정 방안을 기반으로 전체 블로그 연결망에서 특정 성향이 강한 부분 블로그 연결망을 추출하는 방안에 대해서 연구하고자 한다.

참고 문헌

- [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] R. Albert, H. Jeong, and A. Barabasi, "Diameter of the World Wide Web," *Nature*, Vol.401, pp. 130-131, 1999.
- [4] H. Jeong et al., "The Large-scale Organization of Metabolic Networks," *Nature*, Vol.407, pp. 651-654, 2000.
- [5] M. Granovetter, "The Strength of Weak Ties," *American Journal of Sociology*, Vol.78, No.6, pp. 1360-1380, 1973.
- [6] S. Milgram, "The Small World Problem," *Physiology Today*, Vol.2, pp. 60-67, 1967.
- [7] A. Nowak, *Virus Dynamics: Mathematical Principles of Immunology and Virology*, Oxford University Press, 2000.
- [8] D. Watts and S. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Nature*, Vol.393, pp. 440-442, 1998.
- [9] D. Watts, *Small Worlds : The Dynamics of Networks Between Order and Randomness*, Princeton, New Jersey: Princeton University Press, 1999.
- [10] L. Adamic, O. Buyukkokten, and E. Adar, "A Social Network Caught in the Web," *First Monday*, Vol.8, No.6, pp 1-22, 2003.

[11] 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.

[12] (주)SK Communications, <http://www.cyworld.com>.

[13] (주)엠파스, <http://blog.empas.com>.

[14] (주)아이세이브존, <http://www.isavezone.com>.

[15] (주)NHN, <http://blog.naver.com>.

[16] 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.

[17] S. Gardner, *Buzz Marketing With Blogs For Dummies*, John Wiley & Sons Inc, 2005.

[18] R. Scoble and S. Israel, *Naked Conversations*, John Wiley & Sons Inc, 2006.

[19] J. Wright, *Blog Marketing*, McGraw-Hill, 2005.

[20] M. Girvan and M. Newman, "Community Structure in Social and Biological Networks," In *Proc. National Academic Science*, Vol.99, No.12, pp. 7821-7826, 2002.

[21] D. Cai et al., "Mining Hidden Community in Heterogeneous Social Networks," In *Proc. Int'l Workshop on Link Discovery: Issues, Approaches and Applications*, pp. 58-65, 2005.

[22] R. Agrawal et al., "Mining Newsgroups Using Networks Arising From Social Behavior," In *Proc. Int'l. Conf. on World Wide Web*, pp. 529-535, 2003.

[23] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Academic Press, 2001.

[24] News Corporation Co., Ltd. <http://www.myspace.com>.



김 상 옥

1989년 2월 서울대학교 컴퓨터공학과(학사). 1991년 2월 한국과학기술원 전산학과(석사). 1994년 2월 한국과학기술원 전산학과(박사). 1991년 7월~1991년 8월 미국 Stanford University, Computer Science Department, 방문 연구원. 1994년 3월~1995년 2월 KAIST 정보전자연구소 전문 연구원. 1999년 8월~2000년 8월 미국 IBM T.J. Watson Research Center, Post-Doc. 1995년 3월~2003년 2월 강원대학교 정보통신공학과 부교수. 2003년 3월~현재 한양대학교 정보통신대학 정보통신학부 교수. 2009년 1월~현재 미국 Carnegie Mellon University, Visiting Scholar. 관심분야는 데이터베이스 시스템, 저장 시스템, 트랜잭션 관리, 데이터 마이닝, 멀티미디어 정보 검색, 공간 데이터베이스/GIS, 주기억장치 데이터베이스, 이동객체 데이터베이스/텔레매틱스, 사회 연결망 분석, 웹 데이터 분석



윤 석 호

2005년 성결대학교 컴퓨터공학과 졸업(학사). 2007년 한양대학교 정보통신대학원 졸업(공학석사). 2007년~현재 한양대학교 대학원 전자통신컴퓨터공학과 박사과정 재학중. 관심분야는 사회연결망분석, 인터넷 포탈 데이터 분석, e-비지니스, 데이터 마이닝



박 선 주

1989년 서울대학교 컴퓨터공학과(학사)
 1991년 서울대학교 컴퓨터공학과(석사)
 1999년 U of Michigan, Ann Arbor, CSE(박사). 1999년~2005년 Rutgers University, MSIS Department(조교수)
 2005년~현재 연세대학교 경영학과(부교수). 관심분야는 에이전트 시스템, 옥션, 온라인 사회 연결망, 네트워크 가격정책