

블로그 연결망 활성화를 위한 콘텐츠 파워 유저의 파악 방안

(Determining Contents Power Users for Revitalizing Blog Networks)

임승환[†] 김상욱^{**} 박선주^{***} 이준호^{****}
 (Seung-Hwan Lim) (Sang-Wook Kim) (Sunju Park) (Joon Ho Lee)

요약 블로그 연결망 내에는 다른 사용자들이 블로그 서비스를 활발하게 이용하도록 유도하는 특별한 사용자들이 존재한다. 본 논문에서는 이러한 사용자들 가운데 블로그 내의 콘텐츠가 다른 사용자에게 큰 영향을 끼치는 사용자들을 콘텐츠 파워 유저라고 정의하고 이들을 파악하는 기법에 대하여 논의한다. 우선, 각 사용자가 소유하고 있는 각 게시글들의 콘텐츠 영향력을 계량화하는 방법을 제안하고, 이 값들의 합을 이용하여 해당 사용자의 콘텐츠 영향력을 계산하는 방법을 제안한다. 끝으로, 실제 블로그 연결망에서 제안하는 기법과 기존의 기법들을 이용하여 콘텐츠 파워 유저들을 파악하는 실험을 수행하고, 이 결과를 비교 및 분석한다.

키워드 : 사회연결망 분석, 콘텐츠 파워 유저, 블로그, 데이터 마이닝, 정보파급

Abstract In a blog network, there are special users who induce other users to actively utilize blog services. In this paper, these users whose contents exhibit large influence over other bloggers are defined as 'Content Power Users' (CPUs). It is important to accurately determine who content power users are in a blog network in order to establish business policies that will stimulate usage of blog services. In this paper, we discuss a novel method of determining content power users. First, we propose a system of measuring the influence of content of each post owned by individual users. Then, by adjusting the measured values based on the time of exposure and adding them up, we calculate the power of influence for corresponding users. Finally, by applying the proposed method to actual blog networks and comparing the selected power users to those of a preexisting method, we analyze different methods of determining power users. The experimental results demonstrate that our method of determining power users reflects well dynamic changes in a blog network.

Key words : Social Network Analysis, Content Power User, Blog, Data Mining, Information Diffusion

· 본 연구는 NHN(주)의 지원을 받았습니다. 그러나, 본 논문에서 제시된 의견이나 결론, 또는 권고 등은 온전히 저자(들)의 것이며, 반드시 지원 회사의 입장을 대변하는 것은 아닙니다. 또한 본 연구는 한국과학재단의 2009년도 특장기초연구사업(No. R01-2008-000-20872-0)의 부분적인 지원을 받아 수행되었습니다.

· 본 연구의 기본 아이디어와 이에 대한 선행 실험 결과가 국제 컨퍼런스인 ACM SNA-KDD 2009에 발표된 바 있습니다.

† 학생회원 : 한양대학교 전자컴퓨터 통신공학과
 firemoon2000@hanmail.net
 ** 중신회원 : 한양대학교 전자컴퓨터 통신공학과 교수
 wook@hanyang.ac.kr
 *** 중신회원 : 연세대학교 경영대학 교수
 boxenju@yonsei.ac.kr
 **** 중신회원 : (주) NHN
 joonho@nhn.com
 논문접수 : 2008년 9월 24일
 심사완료 : 2009년 10월 8일

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.
 정보과학회논문지: 데이터베이스 제36권 제6호(2009.12)

1. 서론

최근, 인터넷의 발달로 인해서 개인의 의사 표현과 사교를 위한 장이 오프라인에서 온라인으로 급속히 이동하고 있다. 따라서 온라인상에서 개인들이 서로 관계를 맺고 교류를 갖는 온라인 사회가 출현하였다. 이러한 온라인 사회를 대상으로 하는 서비스의 대표적인 예로서 블로그 서비스(blog service)를 들 수 있다.

블로그(blog)는 사용자가 자신의 글을 온라인상에 게시할 수 있는 일종의 개인 웹사이트이며, 블로그 서비스는 사용자가 블로그를 생성하고 운영할 수 있도록 지원해주는 서비스이다. 블로그 서비스 내의 사용자들은 서로 관계를 맺을 수 있으며, 이를 통해서 사회연결망¹⁾이

1) 사회 연결망(social network)은 임의의 사회 내에 존재하는 구성원들 간의 관계 정보를 이용하여 그 사회를 연결망(network)으로 표현한 것이다[1].

형성된다. 본 논문에서는 이러한 블로그들로 이루어진 사회연결망을 블로그 연결망(blog network)이라고 정의한다.

사회연결망 분석(social network analysis)은 사회연결망을 분석함으로써 그 사회의 특성을 도출하는 연구 분야이다[1]. 기존의 사회연결망 데이터들은 구성원간의 관계 유무에 대한 정보만을 포함하고 있었기 때문에, 대부분의 연구들은 주로 사회연결망의 위상구조적인 특성만을 대상으로 분석을 수행하였다[2-9].

블로그 서비스 제공자(blog service provider)는 사용자들이 서비스를 이용한 주요 기록들을 데이터베이스에 저장하고 있다. 본 논문에서는 데이터베이스를 분석하여 블로그 사용자들 간의 관계의 유무에 대한 정보뿐만 아니라 관계의 정도에 대한 정보를 파악하고자 한다. 이러한 구성원간의 관계의 정도에 대한 정보는 기존의 사회연결망 데이터들에는 존재하지 않는 것으로써, 이를 이용하여 사회연결망의 위상 구조만을 분석하던 기존 연구에 비하여 실제적이고 정밀하게 분석하기 위한 연구를 가능하게 한다[10,11].

블로그 연결망을 대상으로 하는 비즈니스의 성공을 위해서는 블로그 연결망 내의 사용자들이 블로그 서비스를 활발하게 이용하도록 하는 것이 선행되어야 한다[12]. 블로그 연결망 내에는 다른 블로그 사용자들이 블로그 서비스를 활발하게 이용하도록 유도하는 역할을 하고 있는 특별한 블로그 사용자들이 존재한다. 이러한 사용자들을 파악할 수 있다면, 이들을 중심으로 블로그 서비스 활성화를 위한 다양한 정책을 수립할 수 있을 것이다.

본 논문에서는 이와 같이 온라인 사회연결망 내에서 일반 사용자들과 차별되는 고유의 영향력을 가진 사용자들을 파워 유저(power user)라고 정의한다. 기존의 연구에서는 사회연결망의 위상 구조적인 특징만을 고려하여 파워 유저를 파악한 바 있다[13]. 만일, 사회연결망을 분석하는데 있어서 사용자들 간의 관계의 정도를 함께 이용할 수 있다면, 사회연결망에서 실제로 영향력이 큰 사용자들을 올바르게 파악할 수 있을 것이다. 본 논문에서는 블로그 연결망 내에 사용자들간의 관계의 정도를 파악할 수 있는 정보가 저장되어 있다는 점에 착안하여 이러한 관계 정도를 함께 분석하여 실제로 블로그 서비스를 활성화하는데 공헌하는 파워 유저를 파악할 수 있는 새로운 방안을 제시한다.

본 논문에서는 블로그 연결망을 활성화시키는데 실질적인 공헌을 하는 파워 유저로서 콘텐츠 파워 유저

(contents power user)를 제안한다. 콘텐츠 파워 유저는 콘텐츠를 통해서 다른 사용자들에게 큰 영향력을 미치고 있는 사용자이다. 콘텐츠 파워 유저들을 파악하기 위해서 우선, 사용자의 각 게시글들의 영향력을 계량화하는 방안을 제안하고, 이를 토대로 각 사용자의 콘텐츠 영향력을 계량화하는 방안을 제안한다. 게시글의 영향력은 블로그 연결망에 노출된 시간이 지날수록 증가하는 경향이 있는데, 이를 보정하여 사용자 콘텐츠 영향력의 정확도를 증대시키는 방안을 제안한다. 또한, 실제 블로그 연결망 데이터를 사용하여 본 논문에서 제안하는 방법으로 파악한 콘텐츠 파워 유저 집합과 기존의 위상구조를 분석하여 파악한 파워 유저 집합들을 각각 구하고, 이 결과를 비교 및 분석한다. 실험 결과에 의하면, 콘텐츠 파워 유저 집합과 기존의 위상 구조에 의한 파워 유저 집합은 크게 다른 것으로 나타났으며, 콘텐츠 파워 유저는 기존의 위상 구조에 의한 파워 유저들과 달리 동적으로 변화하는 블로그 연결망의 특성을 잘 반영하고 있는 것으로 나타났다.

본 논문의 구성은 다음과 같다. 제2장에서는 블로그 연결망을 소개하고, 본 논문에서 해결하고자 하는 문제를 정리한다. 제3장에서는 파워 유저를 파악하기 위한 기존의 연구를 고찰하고, 문제점을 지적한다. 제4장에서는 본 논문이 제안하는 콘텐츠 파워 유저를 파악하기 위한 기법에 대해서 상세히 설명한다. 제5장에서는 제안하는 기법과 기존의 기법들을 이용해서 실제 블로그 연결망을 분석하고, 그 결과를 비교 분석한다. 제6장에서는 본 논문에 대한 결론을 내리고, 향후 연구 방향을 제시한다.

2. 연구 동기

본 장에서는 블로그 연결망에 대해서 소개하고, 본 논문에서 해결하고자 하는 문제를 정의한다. 또한, 이러한 문제를 해결함으로써 인해서 얻게 되는 기대 효과에 대하여 설명한다.

블로그(blog)는 사용자가 자신의 글을 올릴 수 있는 일종의 개인 웹사이트이다[14-17]. 블로그 연결망 내의 각 사용자가 자신의 블로그에 게시글을 작성하면, 다른 사용자들은 이 게시글에 대하여 (1) 조회하거나, (2) 댓글을 남기거나, (3) 스크랩하거나, (4) 워인글 달기를 할 수 있다. 여기서, 조회하기는 게시글을 읽는 행동이며, 댓글 남기기는 게시글에 대하여 자신의 의견을 남기는 행동이다. 또한, 스크랩 하기는 게시글의 내용을 복사하여 자신의 블로그 내에 게시글로 등록하는 행동이며, 워인글 달기는 게시글과 연관된 새로운 내용을 자신의 블로그 내에 게시글로 작성하는 행동이다. 스크랩하기와 워인글 달기를 통하여 재생산된 게시글도 일반 게시글

2) 기존의 연구에서는 이와 같이 위상 구조적 측면에서 더 나은 사용자들 허브(hub)라 부른다.

과 마찬가지로 조회하기, 댓글 남기기, 스크랩하기, 워인 글 달기 등의 새로운 행동을 유발할 수 있다.

그림 1은 블로그 연결망 내에서 이루어지는 행동들의 설명을 위한 간단한 예를 나타낸 것이다. 굵은 사각형은 사용자 A, B, C, D와 대응되는 블로그를 나타내며, 그 내부의 작은 등근 사각형은 해당 블로그에 존재하는 게시글을 나타낸다. 화살표는 사용자 A의 1번 게시글과 사용자 B, C, D가 취한 행동으로 인하여 재생산된 게시글들간의 관계를 나타낸 것이다. 사용자 A의 1번 게시글에 대하여 다른 사용자들은 조회하기, 댓글 남기기, 스크랩하기, 워인글 달기의 다양한 행동을 취하였다. 특히, 사용자 B는 이 게시글을 자신의 2번 게시글로 스크랩 하였으며, 사용자 C는 이 게시글에 대한 워인글 달기를 함으로써 자신의 3번 게시글로 등록하였다. 또한, 사용자 D는 이 사용자 C의 3번 게시글을 자신의 2번 게시글로 스크랩하였다.

블로그 서비스 제공 기업의 목표는 사용자들의 블로그 서비스 이용을 활성화함으로써 이를 통하여 수익을 극대화 하는 것이다. 이를 위하여 본 논문에서는 다른 사용자들이 활발하게 블로그 서비스를 이용하도록 유도 하는 역할을 하고 있는 특별한 사용자들을 파악하는 방안 에 대하여 논의한다. 본 논문에서는 이러한 사용자들을 블로그 연결망에서의 파워 유저라고 정의한다.

블로그 연결망 전체에 영향을 미치는 소수의 파워 유저를 파악할 수 있다면, 이들만을 대상으로 블로그 연결 망 활성화를 위한 정책을 경제적으로 수행할 수 있다. 예를 들어, 파워 유저들에게 다양한 혜택을 제공하여 이 들이 활발하게 블로그 서비스를 이용하도록 장려함으로써 전체 블로그 연결망의 활성화를 위한 촉매제의 역할을 하게 할 수 있다.

블로그 연결망은 기존의 온라인 연결망과 달리 사용자들간의 관계에 게시글 내의 콘텐츠가 중요한 매개체 의 역할을 담당하고 있다. 본 논문에서는 이 점에 착안

하여 블로그 연결망 내에서 콘텐츠의 질을 분석함으로써 다른 사용자들의 행동에 크게 영향을 미치는 사용자 들을 파악하는 방안에 대하여 논의한다.

3. 관련 연구

사회연결망에서 파워 유저를 파악하기 위한 기존의 연구에서는 주로 사회연결망의 위상 구조적인 특징을 기반으로 하는 방법들을 제시하였다[1,18,19]. 본 장에서는 파워 유저 파악에 관련된 기존의 연구 결과들을 요약하고, 이들을 블로그 연결망에 적용 가능한가의 여부를 논의한다.

파워 유저를 파악하기 위한 기존의 방법으로는 사회 연결망에서 해당 사용자가 중앙에 위치한 정도를 측정 하는 방법들이 제안되었다. 사용자의 중앙성을 측정하는 방법에 따라 연결 중앙성, 인접 중앙성, 사이 중앙성으로 세분화된다[1,20-22].

연결 중앙성(degree centrality)은 다른 사용자와의 이웃 관계의 수를 각 사용자의 파워로 부여하는 방법이다. 여기서 사용자의 파워란 각 사용자가 다른 사용자에 게 영향을 끼치는 정도이다. 이웃 관계를 많이 갖고 있는 사용자가 해당 사회연결망에서 중요한 사용자로 판정 된다. 식 (1)에서 U_i 는 식별자가 i 인 사용자, $degree(U_i)$ 는 U_i 의 이웃 관계의 수, N 은 연결망 내의 전체 사용자 들의 수, $C_D(U_i)$ 는 U_i 의 연결 중앙성을 의미한다. 서로 다른 사회연결망에 포함된 사용자들의 연결 중앙성을 상호 비교하기 위해서 $C_D(U_i)$ 를 최대값으로 정규화하여 식 (2)와 같이 나타낼 수 있다. $C_D(U_i)$ 가 최대값을 갖는 경우는 U_i 가 사회연결망 내의 모든 사용자들과 이웃 관 계를 갖고 있는 경우이며, 이 때, $C_D(U_i)$ 는 $N-1$ 이 된다.

$$C_D(U_i) = degree(U_i) \tag{1}$$

$$C'_D(U_i) = \frac{degree(U_i)}{N-1} \tag{2}$$

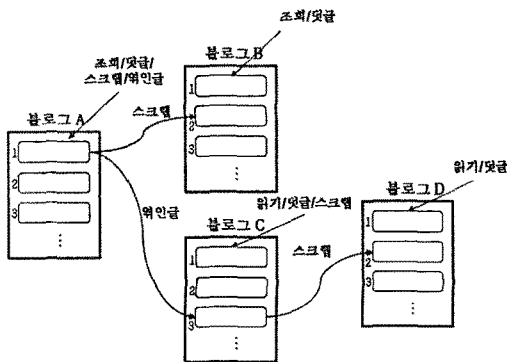


그림 1 블로그 연결망의 예

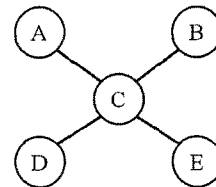


그림 2 연결망의 예

연결 중앙성을 그림 2의 연결망을 이용하여 설명하면 다음과 같다. $C_D(U_A)=1$, $C_D(U_C)=4$ 가 된다. 또한, $C'_D(U_A)=1/4$, $C'_D(U_C)=1$ 이 된다.

인접 중앙성(closeness centrality)은 다른 사용자들과 의 최단 거리들의 합을 각 사용자의 파워로 부여하는

방법이다. 즉, 다른 사용자들과의 최단 거리들의 합이 작은 사용자가 해당 사회연결망에서 중요한 사용자로 판정된다. 식 (3)에서 $distance(U_i, U_j)$ 는 U_i 와 U_j 와의 최단 거리[1], $C_c(U_i)$ 는 U_i 의 인접 중앙성을 의미한다. 서로 다른 사회연결망에 포함된 사용자들의 인접 중앙성을 상호 비교하기 위해서 $C_c(U_i)$ 를 최소값으로 정규화하여 식 (4)와 같이 나타낼 수 있다. $C_c(U_i)$ 가 최소값을 갖는 경우는 U_i 가 사회연결망 내의 모든 사용자들과 이웃 관계를 갖고 있는 경우이며, 이 때, $C_c(U_i)$ 는 $1/(N-1)$ 이 된다. 그림 2에서 $C_c(U_A)=1/(2+1+2+2)=1/7$, $C_c(U_C)=1/(1+1+1+1)=1/4$ 이 된다. 또한, $C'_c(U_A)=4*1/7=4/7$, $C'_c(U_C)=4*1/4=1$ 이 된다.

$$C_c(U_i) = \left[\sum_j^N distance(U_i, U_j) \right]^{-1} \quad (3)$$

$$C'_c(U_i) = \frac{N-1}{\left[\sum_j^N distance(U_i, U_j) \right]} \quad (4)$$

$$= (N-1)C_c(U_i)$$

사이 중앙성(betweenness centrality)은 각 사용자가 다른 사용자들 사이의 최단 경로 상에 위치하는 정도를 사용자의 파워로 부여하는 방법이다. 즉, 다른 사용자들 사이의 최단 경로 상에 위치하는 횟수가 많은 사용자가 해당 연결망에서 중요한 사용자로 판정된다. 식 (5)에 나타난 $C_B(U_i)$ 는 U_i 의 사이 중앙성을 의미한다. U_j 와 U_k 사이에 $distance(U_i, U_j)$ 가 최소값을 갖는 경로들은 여러개가 존재할 수 있다. 여기서, g_{ik} 는 이러한 U_j 와 U_k 의 최단 경로들의 개수, g_{ijk} 는 U_i 를 포함하는 U_j 와 U_k 의 최단 경로들의 개수를 의미한다. 서로 다른 연결망에 포함된 사용자들의 사이 중앙성을 비교하기 위해서 $C_B(U_i)$ 를 $C_B(U_i)$ 의 가능한 최대값으로 정규화하여 식 (6)과 같이 나타낼 수 있다. $C_B(U_i)$ 이 최대값을 갖는 경우는 U_i 가 다른 모든 사용자들 사이의 최단 경로 상에 위치하고 있는 경우이며, 이 때, $C_B(U_i)$ 는 $(N-1)(N-2)/2$ 가 된다. 그림 2에서 $C_B(U_A)=0$, $C_B(U_C)=1$ 이 된다.

$$C_B(U_i) = \sum_j^N \sum_k^N \frac{g_{jik}}{g_{jk}} \quad (5)$$

$$C'_B(U_i) = \frac{C_B(U_i)}{[(N-1)(N-2)/2]} \quad (6)$$

기존의 연구에서는 위에서 언급한 연결 중앙성, 인접 중앙성, 사이 중앙성 등을 기준으로 파워 유저를 선정해 왔다[1,18,19]. 그러나 이와 같이 사회연결망의 위상 구조적 특징만을 고려하는 기준을 사용하는 경우, 실제 연결망에서 영향력을 크게 발휘하는 파워 유저를 올바르게 선정할 수 없다. 예를 들어, 블로그 연결망에서 많은 이웃 관계를 갖고 있는 사용자가 반드시 다른 사용자들

에게 더 큰 영향력을 행사하고 있는 것은 아니다. 오히려 적은 이웃 관계를 갖고 있는 사용자가 다른 사용자들에게 큰 영향력을 행사할 수도 있다. 블로그 연결망 내에는 관계 정보와 더불어 사용자들 간의 관계의 정도를 파악할 수 있는 정보가 저장되어 있다. 본 논문에서는 이점에 착안하여 관계 정도를 함께 분석하여 실제로 블로그 서비스를 활성화하는데 공헌하는 실질적인 파워 유저를 파악할 수 있는 새로운 방안을 제시한다.

4. 제안하는 기법

본 장에서는 블로그 연결망에서 콘텐츠로 인한 영향력이 큰 사용자들을 파악하기 위한 새로운 방안을 제시한다. 각 사용자가 소유하고 있는 게시글들의 콘텐츠 영향력의 합 즉, 각 사용자 콘텐츠 영향력 = \sum (게시글들의 콘텐츠 영향력)을 이용하여 해당 사용자의 콘텐츠 영향력을 계산하는 방법을 제시하고, 이를 이용하여 콘텐츠 영향력이 큰 파워 유저들을 결정하는 방법을 제안한다. 이와 같이, 콘텐츠 영향력이 큰 사용자들을 본 논문에서는 콘텐츠 파워 유저라 정의한다.

4.1 용어 정리

본 절에서는 표 1을 통하여 앞으로의 논의 전개를 위해서 필요한 용어 및 기호들을 정리한다. U_i 는 식별자가 i 인 사용자를 의미한다. D_i 는 U_i 가 소유한 게시글들의 집합을 의미하고, D_{ij} 는 U_i 의 j 번째 게시글을 의미한다. 게시글 D_{ij} 가 사용자들에게 미치는 콘텐츠 영향력을 이 게시글의 콘텐츠 파워(document contents power)라고 정의하며, DocumentContentsPower(D_{ij})로 표기한다. 또한, 사용자 U_i 가 다른 사용자들에게 미치는 콘텐츠 영향력을 사용자의 콘텐츠 파워(user contents power)라고 정의하고, UserContentsPower(U_i)로 표기한다. 사용자가 블로그 서비스를 이용하여 취할 수 있는 행동으로는 게시글 작성(write), 조회하기(read), 댓글 남기기(comment), 스크랩 하기(scrap), 엮인글 달기(link) 등이 있으며, 이러한 행동을 각각 W, R, C, S, L로 표기

표 1 용어 정의

U_i	식별자가 i 인 사용자
$D_i = \{D_{i1}, D_{i2}, \dots\}$	사용자 U_i 가 소유한 게시글들의 집합
D_{ij}	U_i 의 블로그에 등록된 j 번째 게시글
DocumentContentsPower(D_{ij})	게시글 D_{ij} 의 콘텐츠 파워
UserContentsPower(U_i)	사용자 U_i 의 콘텐츠 파워
$AT = \{A_1, A_2, \dots\}$	블로그 서비스를 이용한 사용자 행동의 타입
A_k	Action of type k
W_{Ak}	Weight of A_k

한다. 게시글의 콘텐츠 영향력을 계량화할 때, 각각의 행동에 다른 의미를 부여하기 위하여 서로 다른 가중치를 할당할 수 있다. 행동 W, R, C, S, L을 위한 이러한 가중치를 w_w, w_r, w_c, w_s, w_l 로 표기한다.

4.2 게시글의 콘텐츠 파워

본 절에서는 게시글의 콘텐츠 파워를 계량화하는 방안에 대하여 논의한다. 특정 게시글에 대하여 다른 사용자들이 행동을 보인다는 것은 이 게시글로 인하여 영향을 받았음을 의미한다. 이와 같은 사실에 착안하여 본 논문에서는 게시글의 콘텐츠 파워를 계량화하기 위해서 각 게시글에 대한 사용자들의 행동의 가중치와 빈도를 곱하는 방법을 사용한다.

각 게시글은 처음으로 작성된 블로그 내에서 다른 사용자에게 직접적으로 영향을 미칠 수 있고, 스크랩 하거나 워인글 달기를 통해서 전파된 다른 블로그 내에서 다른 사용자에게 간접적으로 영향을 미칠 수 있다. 본 논문에서는 전자를 게시글의 직접적인 영향, 후자를 게시글의 간접적인 영향이라고 부르며, 이를 계량화한 값을 각각 해당 게시글의 직접 콘텐츠 파워(direct contents power), 간접 콘텐츠 파워(indirect contents power)라고 정의한다. 각 게시글의 콘텐츠 파워를 해당 콘텐츠의 직접 콘텐츠 파워와 간접 콘텐츠 파워의 합으로 계산한다. 여기서, 직접 콘텐츠 파워와 간접 콘텐츠 파워의 반영 비율은 응용에 따라 각각에 대한 가중치 w_D 와 w_I 를 부여하여 조절 할 수 있다.

표 2는 게시글의 콘텐츠 파워의 계량화 방법을 정리한 것이다.

표 2 게시글의 콘텐츠 파워 계산

DocumentContentsPower(D_i)
= w_D *DirectContentsPower(D_i) + w_I *IndirectContentsPower(D_i)
DirectContentsPower(D_i) = $\sum_{A_k \in AT} W_{A_k} * \text{Count}(D_i, A_k)$
Count(D_i, A_k) = D_i 내에서 A_k 의 발생 빈도수
IndirectContentsPower(D_i) = $\sum_j \text{DirectContentsPower}(D_j)$
여기서, D_j 은 D_i 로부터 직접적으로 재생산된 게시글을 의미

4.3 파급 이력

게시글의 콘텐츠 파워를 효과적으로 계산하기 위해서는 게시글들 간의 파급 관계를 효과적으로 검색할 수 있는 구조가 필요하다. 이를 위해서 본 논문에서는 그림 3에 나타난 각 원본 게시글에 대한 파급 이력 구조(diffusion history structure)를 사용한다. 파급 이력의 첫 번째 행은 파급 이력에 포함된 게시글들의 원본이 되는 게시글을 의미한다. (UID, D#)는 각 게시글 $D_{UID,D\#}$ 의 식별자를 나타내며, TYPE은 해당 게시글이

어떠한 행동을 통하여 등록되었는지를 나타낸다. 또한, (PUID, PD#)은 게시글 $D_{UID,D\#}$ 의 파급의 출처가 되는 직전 게시글 $D_{PUID,PD\#}$ 의 식별자를 나타내며, TIME은 해당 게시글이 등록된 시각을 의미한다. 그림 3은 그림 1의 첫 게시글을 원본으로 하는 게시글들의 파급 이력을 나타낸 것이다. U_A 가 t_1 에 $D_{A,1}$ 을 새로 작성하였고, 이 게시글을 $D_{A,1}$ 에 대하여 U_C 가 t_2 에 워인글 달기를 하여 $D_{C,3}$ 를 등록하였다. 또한, U_B 는 게시글 $D_{A,1}$ 에 대하여 t_3 에 스크랩을 하였으며, U_C 에 의하여 등록된 $D_{C,3}$ 를 U_D 가 t_4 에 스크랩하였다. 이와 같이, 각 원본 게시글의 파급 이력을 미리 구축하여 두면 게시글들 간의 파급 관계를 쉽게 확인할 수 있으므로 게시글의 간접 콘텐츠 파워의 효율적인 계산이 가능해진다.

그림 4는 그림 3의 파급 이력에 나타난 게시글들 간의 파급 관계를 이용하여 게시글의 콘텐츠 파워를 계산하는 과정의 예를 보인 것이다. 여기서는 직접 콘텐츠 파워와 간접 콘텐츠 파워의 가중치를 각각 동일하게 1로 설정한 경우를 대상으로 하였다. 각 게시글 안의 값들은 <직접 콘텐츠 파워, 간접 콘텐츠 파워, 전체 콘텐츠 파워>를 의미한다. 그림 4에서 보는 바와 같이 게시글의 파급 이력은 트리의 형태를 갖는다. $D_{D,2}$ 는 이후에 파급된 기록이 없으므로 간접 콘텐츠 파워가 0이며, 전체 콘텐츠 파워는 직접 콘텐츠 파워인 5가 된다. 따라서 $D_{C,3}$ 의 간접 콘텐츠 파워는 5가 되고, 이 값에 직접 콘텐츠 파워인 3을 더하여 $D_{C,3}$ 의 전체 콘텐츠 파워는 8이 된다. 또한, 게시글의 원본이 되는 $D_{A,1}$ 의 간접 콘텐츠

Diffusion History #3					
UID	D#	TYPE	PUID	PD#	TIME
A	1	W	-	-	t_1
C	3	L	A	1	t_2
B	2	S	A	1	t_3
D	2	S	C	3	t_4

그림 3 파급 이력의 예

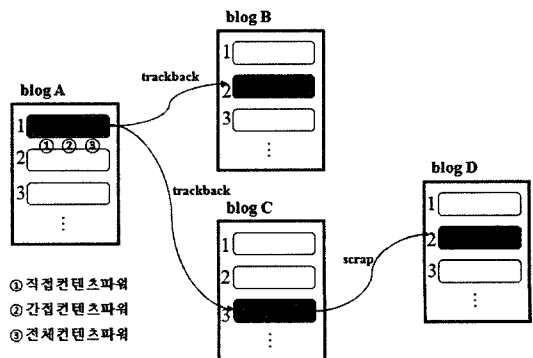


그림 4 파급 이력을 이용한 게시글의 콘텐츠 파워 계산

파워는 D_{C_3} 의 전체 콘텐츠 파워인 8과 D_{B_2} 의 전체 콘텐츠 파워인 4를 합한 12가 되고, 여기에 직접 콘텐츠 파워인 4를 더한 16이 된다.

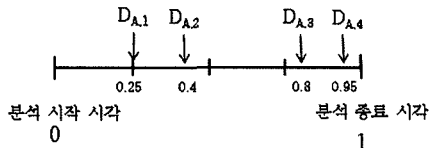
4.4 사용자 콘텐츠 파워

사용자의 콘텐츠 파워는 사용자 블로그에 등록되어 있는 모든 게시글들의 콘텐츠 파워의 합을 이용하여 계산한다. 오래전에 등록된 게시글은 다른 사용자들에게 오랜 시간 동안 노출되어 많은 행동들을 유발할 수 있기 때문에 최근에 등록된 게시글에 비해서 대체로 높은 콘텐츠 파워를 갖게 된다. 이로 인해 노출 시간이 큰 게시글들이 사용자의 콘텐츠 파워에서 큰 비중을 차지하게 되는데, 이는 사용자의 콘텐츠 파워로서 사용자가 과거에 보인 영향력 보다, 분석 시점인 현재에 보이고 있는 영향력을 계산하고자 하는 일반적인 분석 목적에 부합되지 않는다. 따라서 본 논문에서는 이러한 문제를 해결하기 위해서 각 게시글의 콘텐츠 파워에 등록 이후의 노출 시간을 반영하는 방법을 사용한다. 즉, 전체 블로그 데이터의 분석 기간을 1로 간주하고, 해당 게시글이 등록된 시각을 참조하여 해당 게시글의 상대적 노출 시간의 역수를 계산한다. 이 값을 게시글의 콘텐츠 파워에 곱함으로써 노출 시간의 차이로 인한 왜곡을 보정한다.

식 (7)은 게시글의 콘텐츠 파워를 이용하여 사용자의 콘텐츠 파워를 계산하는 방법을 나타낸 것이다. 여기서, $IET_{D_{i,j}}$ 는 게시글 $D_{i,j}$ 의 상대적 노출 시간의 역수(Inverse of the Exposed Time)를 의미한다.

$$User\ Contents\ Power(U_i) = \sum_j IET_{D_{i,j}} * Document\ Contents\ Power(D_{i,j}) \tag{7}$$

그림 5는 식 (7)을 이용하여 $UserContentsPower(U_A)$ 를 계산하는 과정을 나타낸 것이다. 이 예에서는 가장 오래 전에 등록된 게시글 $D_{A,1}$ 의 IET는 0.25, 가장 최근에 등록된 게시글 $D_{A,4}$ 의 IET는 0.95이고, $UserContentsPower(U_A) = 50*0.25 + 35*0.4 + 35*0.8 + 10*0.95 = 64$ 가 된다.



Document	Document Contents Power	IET
$D_{A,1}$	50	0.25
$D_{A,2}$	35	0.4
$D_{A,3}$	35	0.8
$D_{A,4}$	10	0.95

그림 5 사용자 콘텐츠 파워 계산의 예

블로그 연결망 내에서 n 명의 콘텐츠 파워 유저들을 식별하기 위해서는 앞서 제안한 기법을 이용하여 각 사용자의 콘텐츠 파워를 계산하고, 전체 사용자들 중 상위 n 명의 사용자들을 파워 유저로 선정한다.

5. 블로그 연결망 분석

본 장에서는 제안하는 기법과 기존의 기법들을 이용하여 실제 블로그 연결망에서 파워 유저를 선정한 결과를 제시한다. 먼저, 제 5.1절에서는 성능 평가를 위한 실험 환경을 설명하고, 제 5.2절에서는 도출된 실험 결과를 분석함으로써 제안된 기법의 유용성을 규명한다.

5.1 실험 환경

본 논문에서는 파워 유저 분석을 위하여 국내 최대의 블로그 서비스인 네이버 블로그[17]의 2006년 4월 17일부터 2006년 7월 17일까지의 약 30기가 바이트의 대용량 데이터를 사용하였다. 이 데이터는 임의의 정수로 표현되는 사용자 정보와 사용자간의 교류 정보들로 구성되어 있다. 네이버 블로그 연결망의 정점은 블로그 서비스의 사용자로 설정하였고, 간선은 블로그 간의 이웃 관계 여부가 아니라 교류의 빈번한 정도가 일정 이상인 경우의 관계로 설정하였다. 이는 블로그 간에 이웃 관계가 설정되어 있어도 빈번한 교류를 갖지 않는 경우가 많다는 사실에 근거한 것이다. 실험의 수행을 위하여 네이버 내부의 서버를 사용하였다.

본 실험에서 성능 평가의 대상으로 선정된 파워 유저 선정 기법은 연결 중앙성을 이용하는 기법(DEG), 사용자가 직접적으로 생산한 게시글들만을 대상으로 하는 콘텐츠 파워를 이용하는 기법인 $ContentsPower_Direct(CP_D)$, 사용자가 스크랩하기 혹은 위인글 달기 등의 행동을 통하여 간접적으로 생산한 게시글들만을 대상으로 하는 콘텐츠 파워를 이용하는 기법인 $ContentsPower_Indirect(CP_I)$, 직접적으로 생산한 게시글들과 간접적으로 생산한 게시글들을 모두 대상으로 하는 콘텐츠 파워를 이용하는 기법인 $ContentsPower_Total(CP_T)$ 을 사용하였다. 즉, 기법 (CP_D) 는 표 2에서 $W_D=1, W_I=0$, 기법 (CP_I) 는 $W_D=0, W_I=1$, 기법 (CP_T) 는 $W_D=1, W_I=1$ 로 설정한 것을 의미한다. 파워 유저를 식별하는 기존의 기법 중에서 인접 중앙성과 사이 중앙성을 이용하는 기법은 정점이 수천만에 이르는 네이버 블로그 연결망을 대상으로 계산하기 위해서는 비현실적인 수행 시간을 필요로 하므로, 본 실험에서의 평가 대상에서 제외하였다.

기법 (CP_D) , 기법 (CP_I) , 기법 (CP_T) 에서 게시글의 콘텐츠 파워를 계산하기 위해서 사용된 사용자 행동의 가중치는 게시글 작성 1, 조회하기 1, 댓글 남기기 3, 스크랩 하기 7, 위인글 달기 7로 설정하였다. 행동의 가

표 3 각 기법에 따른 파라미터 정리

기법	WD:WI	사용자 행동	가중치
CP_D	1:0	게시글 작성	1
		조회하기	1
CP_I	0:1	댓글 남기기	3
		스크랩 하기	7
CP_T	1:1	워인글 달기	7

중치 값들은 응용의 목적에 따라서 다양하게 설정할 수 있다. 예를 들어, 분석가가 조회하기, 댓글 남기기를 많이 유발하는 사용자에게 높은 파워를 부여하기를 원하는 경우, 조회하기, 댓글 남기기 행동의 가중치를 스크랩 하기, 워인글 달기에 비하여 높게 책정하여 분석을 수행한다. 본 실험에서는 스크랩 하기, 워인글 달기 행동을 유발하는 사용자들이 조회하기, 댓글 남기기를 유발하는 사용자들에 비해서 블로그 연결망 활성화에 도움이 된다고 판단하여 스크랩 하기, 워인글 달기 행동의 가중치를 다른 행동들에 비해서 높게 책정하였다. 실험에서 사용된 각 기법에 따른 사용자 행동의 가중치 및 WD:WI의 비율을 정리하면 표 3과 같다.

본 실험에서 게시글 작성 행동은 게시글의 파급 시에 단 한번만 발생하므로, 게시글의 특성을 반영하고 있지 않은 것으로 판단되어 콘텐츠 파워의 계산 시에 제외하였다. 또한, 네이버 블로그 서비스에서는 게시글의 조회 행동에 대한 정보를 저장하고 있지 않기 때문에, 조회하기 행동은 계산 시에 반영하지 않았다. 또한, 기법(CP_T)에서 직접 콘텐츠 파워와 간접 콘텐츠 파워의 반영 비율은 각각 동일하게 1로 설정하였다. 본 실험에서는 각 기법이 선정한 파워 유저 집합의 시간의 흐름에 따른 변화를 살펴보기 위해서 전체 분석 구간을 10일 간격으로 총 9개의 구간으로 나누어 실험을 수행하였다.

5.2 실험 결과 및 분석

본 논문에서는 세가지 실험을 수행하였다. 실험 1에서는 각 기법을 이용하여 선정한 파워 유저 집합의 결과들을 제시하고, 결과의 의미를 해석하였다. 실험 2에서는 시간의 흐름에 따라 각 기법을 이용하여 선정한 파워 유저 집합의 생존률 변화를 비교하였다. 실험 3에서는 시간의 흐름에 따라 각 기법을 이용하여 선정한 파워 유저 집합이 변화하는 정도를 비교하였다.

파워유저는 분석의 목적에 따라서 다양하게 정의될 수 있으며, 이들을 파악하기 위해서는 각 정의를 만족하는 다양한 기법들이 필요하다. 따라서 본 논문에서는 각 기법들 간의 정확도를 비교하는 것은 의미가 없다고 판단하여, 다음의 사항들을 실험을 통하여 논의 하고자 한다. (1) 블로그와 관련된 비즈니스에서 다른 사용자들이

많은 행동을 일으키도록 유발하는 사용자를 파악하는 것은 명백하게 중요하며, 콘텐츠 파워유저의 정의는 이러한 요구에 부응하기에 적절한 개념이다. (2) 파급 이력 구조를 이용한 제안하는 기법은 이러한 콘텐츠 파워 유저를 효과적으로 파악하기에 적합하다. (3) 사용자의 콘텐츠 파워는 블로그 연결망 내에서 발생하는 사용자들의 행동을 성공적으로 계량화한 것이다.

실험 1에서는 각 기법이 선정한 30명의 파워 유저 집합의 결과를 비교하였다. 그림 6은 실험 1의 결과를 보인 것이다. 블로그 사용자의 아이디는 개인 정보 보호를 위해서 일련번호로 대체하여 나타내었다. 선정한 파워 유저들을 각 기법이 이용하는 파워를 기준으로 하여 정렬하였다. 회색 바탕의 사용자는 위상 구조만을 이용하는 DEG를 통해서 선정한 파워 유저들 중 다른 기법을 통해서 선정한 파워 유저 집합들에 적어도 한번 이상 포함된 사용자를 의미한다.

실험 결과, DEG에 의하여 선정된 파워 유저 중에서 사용자 307819, 사용자 227694, 사용자 191454, 사용자 1050536 등 넷 외에는 다른 기법에 의하여 선정된 파워 유저 집합에 전혀 포함되지 않은 것으로 나타났다. 이러한 결과는 사용자간의 관계의 정도에 관한 정보를 분석하여 파워 유저를 선정하는 기법의 결과와 위상 구조만을 이용하여 파워 유저를 선정하는 기법의 결과가 매우 다르다는 것을 의미한다. 이는 블로그 연결망 활성화를 위해서 다른 사용자들에게 많은 행동을 유발하는 사용자들을 파악하기 위해서 위상 구조만을 이용하는 기법을 사용하는 것은 적절하지 않다는 것을 의미한다.

Rank	DEG	Contents Power		
		CP_D	CP_I	CP_T
1	21608	307819	205649	
2	1212576	227694	521114	
3	1223870	360763	1066638	521114
4	307819	957182	609211	360763
5	173326	240513	38291	957182
6	571678	976756	228000	205649
7	373712	1007110	156438	976756
8	360430	275612	579543	240513
9	720669	1329246	1283112	1007110
10	1030582	1254859	29155	1153439
11	221718	1153439	481166	275612
12	227694	1050536	330592	1329246
13	1269927	954535	1139729	1254859
14	465355	1214458	1191046	1050536
15	191454	84535	591032	1214458
16	1138926	1120363	660655	954535
17	1263255	1041565	1336501	579543
18	1244929	782455	1223256	999317
19	108545	208456	1384606	84535
20	127904	999317	938203	1120363
21	1050536	493135	975704	782455
22	351335	131454	51090	1041565
23	985608	140607	1009253	493135
24	629130	417359	33801	140607
25	608703	371450	1380719	191454
26	709632	1362815	1354168	208456
27	561782	337868	745863	84463
28	433300	1049339	269898	417359
29	639	1341008	1032290	371450
30	336602	84463	513037	966173

그림 6 각 기법에 따른 파워 유저 집합의 결과

실험 2에서는 각 기법을 이용하여 선정된 파워 유저 집합의 시간의 흐름에 따르는 생존률의 변화를 비교하였다. 파워 유저 집합의 생존률이란 분석구간 1에서 선정된 파워 유저들이 이후의 각 분석 구간에서 선정된 파워 유저 집합에 포함되는 비율을 의미한다. 식 (8)은 생존률의 계산식을 나타낸다. 여기서 PU_t 는 t 구간에서 선정된 파워유저 집합을 의미하고, $PU_t \cap PU_1$ 는 PU_t 와 PU_1 에 동시에 포함되는 사용자들을 의미한다.

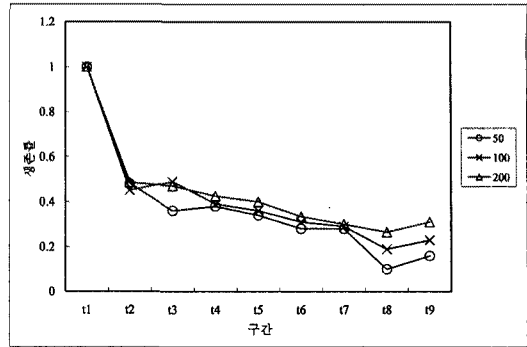
$$\text{생존률}(PU_t) = |PU_t \cap PU_1| / |PU_t| \quad (8)$$

이러한 비교를 통해서 특정 시점에서 선정된 파워 유저가 이후에도 파워 유저로 다시 선정되는 정도를 예측할 수 있으며, 선정된 시점부터 향후 어느 시점까지를 파워 유저로서 인정할 것인지를 결정하는데 도움을 줄 수 있다.

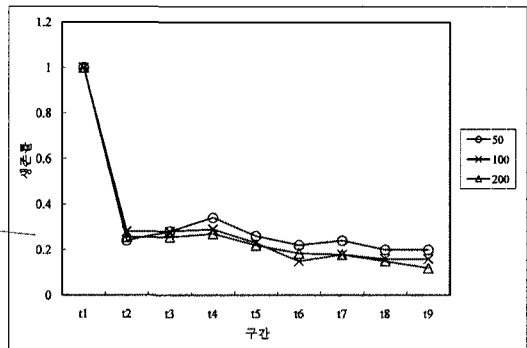
그림 7과 그림 8은 실험 2의 결과를 나타낸다. 선정된 파워 유저 집합의 크기는 50명, 100명, 200명으로 각각 다르게 설정하였다. 가로축은 분석 구간 1과 비교 대상이 되는 분석 구간을 나타내고, 세로축은 생존률을 나타낸다.

먼저, 그림 7은 위상 구조만을 이용한 DEG의 실험 결과를 보인 것이다. 파워 유저 집합의 크기에 관계없이 시간의 흐름에 따라 생존률은 서서히 감소하였다. 50명의 파워 유저 집합이 구간 t9에서 66%로서 가장 낮은 생존률을 보였고, 100명과 200명의 파워 유저 집합들은 구간 t9에서 75%의 유사한 생존률을 보였다.

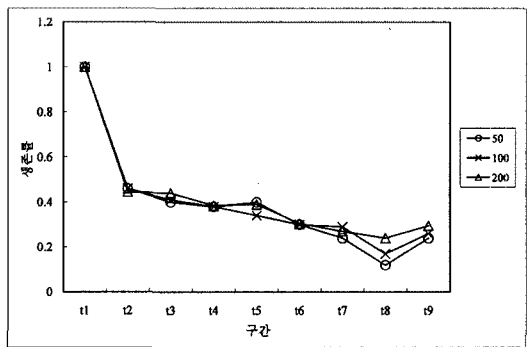
그림 8은 게시글의 컨텐츠파워를 이용한 기법 CP_D, 기법 CP_I, 기법 CP_T의 실험 결과를 보인 것이다. 파워 유저 집합의 크기에 관계없이 시간의 흐름에 따라 파워 유저의 생존률은 급격히 감소하는 것으로 나타났다. 50명의 파워 유저 집합이 구간 t9에서 16%~24%의 가장 낮은 생존률을 보였고, 100명과 200명의 파워 유저 집합은 구간 t9에서 16%~26%와 12%~31%의 유사한 생존률을 보였다. t9에서 기법 DEG에 의한 파워 유저의 생존률이 75%임을 고려하면 기법 CP_D, 기법 CP_I, 기법 CP_T에 의한 파워 유저 집합은 매우 급격



기법(CP_D)



기법(CP_1)



기법(CP_T)

그림 8 댓글의 개수를 이용한 기법의 생존률 변화

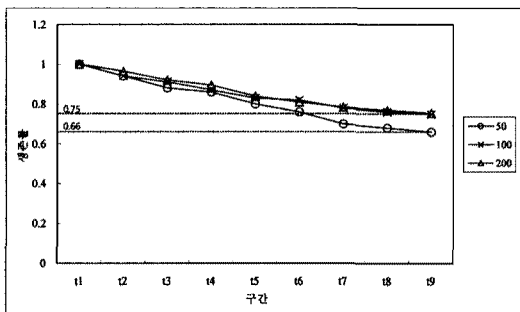


그림 7 기법(DEG)의 파워 유저 생존률 변화

히 생존률이 떨어진다는 것을 알 수 있다. 이 결과를 통해서 컨텐츠 파워를 이용하는 기법은 위상 구조를 이용한 기법과 달리, 사용자 수에 관계 없이 생존률이 급격하게 떨어진다는 것을 알 수 있다. 따라서 컨텐츠 파워 유저를 대상으로 비즈니스 정책을 수행할 때에는 위상 구조를 이용하여 선정된 파워유저를 이용하는 경우와 달리, 파워유저 집합에 지속적으로 의미있는 파워유저들을 포함시키기 위해서, 파워유저 집합의 크기와는 상관 없이 빈번하게 파워유저 선정 작업을 수행해야 함을 알 수 있다.

실험 3에서는 시간의 흐름에 따라 각 기법을 이용하여 선정된 파워 유저 집합의 유사도를 비교하였다. 파워 유저 집합의 유사도는 이전 분석구간에서의 파워 유저 집합이 현재의 분석구간에서 선정된 파워 유저 집합에 포함되는 정도를 이용하며, 이 값이 클수록 분석 구간간의 파워 유저 집합의 변화가 적다는 것을 의미한다. 식 (9)는 유사도의 계산식을 나타낸다. 여기서 PU_t 는 t 구간에서 선정된 파워유저 집합을 의미하고, $PU_t \cap PU_{t-1}$ 은 PU_t 과 PU_{t-1} 에 동시에 포함되는 사용자들을 의미한다.

$$\text{유사도}(PU_t) = |PU_t \cap PU_{t-1}| / |PU_t| \quad (9)$$

이러한 분석을 통해서 각 기법에 따라 선정된 파워 유저들의 파워 유저 집합으로의 진입과 퇴출의 정도를 파악할 수 있으며, 파워 유저 집합의 성향을 파악을 위한 연구에 도움을 줄 수 있다.

그림 9와 그림 10은 실험 3의 결과를 나타낸다. 선정된 파워 유저 집합의 크기는 50명, 100명, 200명으로 각각 다르게 설정하였다. 가로축은 비교의 대상이 되는 분석구간을 나타내고 세로축은 비교 분석구간의 파워 유저 집합의 유사도를 나타낸다.

먼저, 그림 9는 사용자의 위상적인 특성만을 이용한 DEG의 실험 결과를 보인 것이다. 파워 유저 집합의 크기에 관계없이 시간의 흐름에 관계없이 변화의 정도가 거의 없는 것으로 나타났다. 이러한 결과가 나타난 이유는 블로그 사용자들간의 서로이웃 관계의 수정이 거의 일어나지 않기 때문이다. 50명의 파워 유저 집합이 비교구간 t8-t9에서 96%로서 가장 낮은 유사도를 보였고, 100명과 200명의 파워유저 집합들은 비교구간 t8-t9에서 98%의 유사한 생존율을 보였다.

그림 10은 콘텐츠 파워를 이용한 기법 CP_D, 기법 CP_I, 기법 CP_T의 실험 결과를 보인 것이다. 기법(CP_D), 기법(CP_T)에 대해서 파워 유저 집합의 크기에 관계없이 시간의 흐름에 따라 유사도는 평균 약 0.5의 분포를 보였으며, 기법(CP_I)는 평균 약 0.4의 분포를 보였다. 실험 결과, 파워 유저를 선정하는 기법에 상관없이 시간의 흐름에 따른 파워 유저 집합이 변화하는

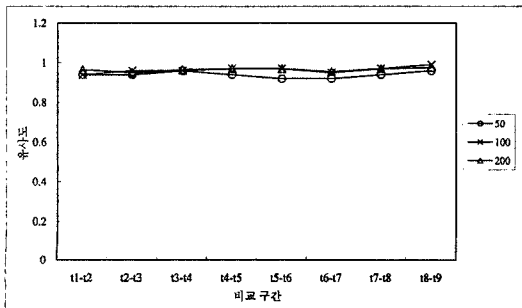
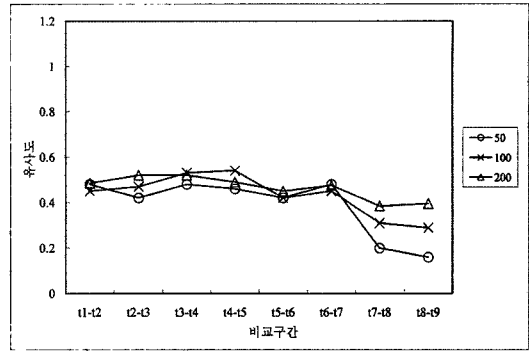
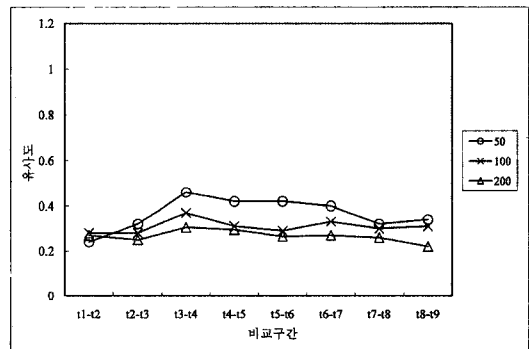


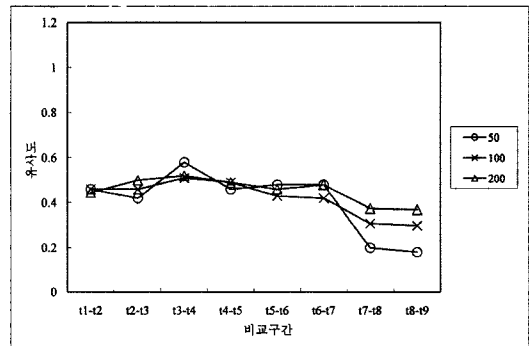
그림 9 기법(DEG)의 파워 유저 집합의 변화



기법(CP_D)



기법(CP_I)



기법(CP_T)

그림 10 콘텐츠 파워를 이용한 기법의 파워 유저의 변화율

정도는 유사하며, DEG에 비해 CP를 이용하는 기법은 변화의 정도가 큰 것을 알 수 있다. 기법(CP_I)는 기법(CP_D), 기법(CP_T)에 비해서 사용자가 적을수록 높은 유사도를 보이는데, 이는 재생산 행동을 많이 유발하는 사용자들 중에서 상위에 위치한 사용자들은 거의 변하지 않는다는 사실에 기인한 것이다.

실험 결과, 콘텐츠 파워를 이용하는 기법은 위상 구조를 이용한 기법에 비하여 각 구간에서 선정된 파워유저 집합들간의 유사도가 크게 낮은 것으로 나타났다. 이 결

과를 통해서 실험 2의 결과와 마찬가지로, 콘텐츠 파워 유저를 대상으로 비즈니스 정책을 수행할 때에는 위상 구조를 이용하여 선정된 파워유저를 이용하는 경우와 달리, 파워유저 집합에 지속적으로 의미있는 파워유저들을 포함시키기 위해서, 파워유저 집합의 크기와는 상관 없이 빈번하게 파워유저 선정 작업을 수행해야 함을 알 수 있다.

6. 결론

본 논문에서는 블로그 연결망에서 파워 유저를 파악하는 방안에 대하여 논의하였다. 기존의 연구에서는 사회연결망의 관계 정도에 대한 정보의 부재로 인해서 위상 구조적인 특징만을 고려하여 파워 유저를 파악하였다. 그러나 이러한 방식은 응용의 요구 사항을 원천적으로 배제하므로 대부분의 응용에 적합하지 않다. 본 논문에서는 블로그 연결망 내에 사용자들간의 관계의 정도를 파악할 수 있는 정보가 저장되어 있다는 점에 착안하여 이러한 관계 정도를 함께 분석함으로써 블로그 서비스를 활성화하는데 실제로 공헌하는 파워 유저를 파악하는 새로운 방안을 제시하였다.

본 논문의 주요 공헌은 다음과 같다.

- 위상 구조적인 특징만을 이용한 기존의 파워 유저의 개념과는 달리 블로그 연결망 내의 사용자의 실질적인 영향력을 반영하는 콘텐츠 파워 유저라는 새로운 개념을 제안하였다.
- 블로그 연결망에서 게시글들의 실질적인 영향력을 분석함으로써 각 사용자의 콘텐츠 파워를 계량화하는 방안을 제안하였다.
- 사용자 콘텐츠 파워의 정확도를 향상시키기 위하여 게시글들의 영향력을 노출 시간에 따라 보정하는 방안을 제안하였다.
- 제안하는 기법과 기존의 기법들을 이용하여 실제 블로그 연결망에서 파워 유저들을 파악하는 실험을 수행하고, 이 결과를 비교 및 분석하였다.

본 논문에서 제안하는 콘텐츠 파워유저를 이용하여 블로그 연결망의 활성화에 이바지하는 사용자들을 선정할 수 있다. 이러한 파워유저들을 블로그 서비스 상에서 공개함으로써, 파워 유저들의 활동을 촉진시킬 수 있으며, 파워 유저가 아닌 일반 사용자들을 자극함으로써 파워 유저로 변화하는 촉매제 역할을 할 수 있다. 또한, 타겟 마케팅 측면에서는 특정 주제의 상품 및 서비스를 홍보할 때, 파워유저들만을 대상으로 마케팅함으로써 최소의 비용으로 최대의 광고효과를 기대할 수 있다.

본 논문에서는 제안한 새로운 파워 유저의 개념은 블로그 연결망에서 콘텐츠를 통하여 다른 사용자들에게 많은 행동을 유발하는 사용자이다. 그러나 이러한 정의

는 이 파워 유저의 영향을 받는 사용자의 수를 전혀 고려하지 않은 것이다. 어떤 응용 분야에서는 가급적 다수의 사용자들에게 영향을 미칠 수 있는 사용자를 파워 유저라고 선정하기를 원할 수 있다. 이러한 점을 고려하여 현재 콘텐츠를 통하여 많은 사용자의 행동을 유발하는 사용자들을 파악할 수 있는 방안에 대하여 연구를 진행하고 있다.

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] X. Song et al., "Mining in Social Networks Information Flow Modeling based on Diffusion Rate for Prediction and Ranking," In *Proc. Int'l. Conf. on World Wide Web*, pp.191-200, 2007.
- [11] J. Iribarren and E. Moro, "Information Diffusion Epidemics in Social Networks," *Arxiv*, 0706.0641, 2007.
- [12] (주)위고넷 이브랜딩 전략연구소, *커뮤니티를 승부하는 브랜드 전략*, 이디자인, 2004.
- [13] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the Spread of Influence through a Social Network," In *Proc. ACM Int'l. Conf. on Knowledge Discovery and Data Mining*, ACM SIGKDD, pp. 137-146, 2003.
- [14] (주)SK Communications, <http://www.cyworld.com>.
- [15] (주)엠포스, <http://blog.empas.com>.
- [16] (주)아이세이브존, <http://www.isavezone.com>.
- [17] (주)NHN, 블로그플, blog.naver.com.
- [18] J. Brown and P. Reinegen, "Social Ties and

Word-of-Mouth Referral Behavior," *Journal of Consumer Research*, vol.1, no.3, pp.350-362, 1987.

- [19] J. Coleman, H. Menzel, and E. Katz, *Medical Innovations: A Diffusion Study*, Bobbs Merrill, 1966.
- [20] A. Bavelas, "Communication patterns in task-oriented groups," *Journal of the Acoustical Society of America*, 22, pp. 271-282, 1950.
- [21] C. Proctor and C. Loomis, *Analysis of sociometric data*, Research Methods in Social Relations, pp. 561-586, 1951.
- [22] A. Shimbel, "Structural parameters of communication networks," *Bulletin of Mathematical Biophysics*, 15, pp.501-507, 1953.



임 승 환

2003년 한양대학교 전자컴퓨터공학부 졸업(학사). 2005년 한양대학교 정보통신대학원 졸업(공학석사). 2005년~현재 한양대학교 대학원 전자통신컴퓨터공학과 박사과정 재학중. 관심분야는 데이터베이스, 데이터 마이닝, 사회 연결망 분석

김 상 욱

한국정보과학회논문지 : 데이터베이스
제 36 권 제 1 호 참조



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, 육선, 온라인 사회연결망, 네트워크 가격정책



이 준 호

1987년 서울대학교 전자계산기공학과(학사). 1989년 한국과학기술원 전산학과(석사). 1993년 한국과학기술원 전산학사(박사). 1993년~1994년 한국과학기술원 인공지능연구센터, 연구원. 1994년~1995년 코넬대학교 전산학과, 방문연구원. 1994년~1997년 연구개발정보센터, 선임연구원. 2003년~2005년 매사추세츠대학교 전산학과, 방문교수. 1997년~2008년 숭실대학교 IT대학 컴퓨터학부, 부교수. 현재 NHN주식회사 COO. 2000년~현재 편집위원, Information Retrieval. 1996년~1998, 2002, 2003 학술위원 ACM SIGIR 컨퍼런스. 1996년~1999, 2003 학술위원 IRAL 워크숍. 1997년 학술위원, RIAO 컨퍼런스. 2001년~2003년 편집위원, 한국정보처리학회 논문지. 1996년~2000년 편집위원, 한국정보과학회 데이터베이스 연구회. 2000년~2002년 운영위원, 한국정보과학회 데이터베이스 연구회. 관심분야는 정보검색