

# 블로그 월드에서 비명시적 관계를 고려한 정보 파급 모델

(An Information Diffusion Model  
Considering Non-explicit  
Relationships in the Blog World)

권 용 석 <sup>†</sup> 김 상 육 <sup>‡</sup>

(Yong-Suk Kwon) (Sang-Wook Kim)

박 선 주 <sup>\*\*\*</sup> 임 승 환 <sup>†</sup>

(Sunju Park) (Seung-Hwan Lim)

이 재 범 <sup>\*\*\*\*</sup>

(Jae Bum Lee)

**요약** 블로그 월드에서의 정보의 파급에 대한 기존의 연구들은 블로그들 간에 명시적 관계를 설정하고, 정보가 파급되는 요인으로서 이 관계를 통해서 발생하는 입소문 효과만을 고려하였다. 그러나 본 연구자들의 선행 연구 결과에 따르면 블로그 월드에서 실제 발생하는 정보의 파급 현상들은 약 85%는 비명시적 관계를 통해서 발생한 것이다. 따라서 본 논문에서는 명시적 관계, 비명시적 관계를 통한 정보의 파급을 함께 고려하여 이들 관계를 통한 정보의 파급 현

† 본 연구는 NHN(주)의 지원을 받았습니다. 또한, 지식경제부 및 정보통신 연구진흥원의 대학IT연구센터지원사업(IITA-2009-C1090-0902-0040)과 한국과학재단의 2009년도 특정기초연구사업(No. R01-2008-000-20872-0)의 부분적인 지원을 받았습니다.

‡ 이 논문은 제35회 추계학술대회에서 '블로그 월드에서 정보 파급 모델'의 제목으로 발표된 논문을 확장한 것임

† 비 회 원 : 한양대학교 전자컴퓨터통신 공학과  
neo\_zion@hanyang.ac.kr  
firemoon2000@hanmail.net

\*\* 종신회원 : 한양대학교 전자컴퓨터통신 공학과 교수  
wook@hanyang.ac.kr

\*\*\* 종신회원 : 연세대학교 경영대학 교수  
boxenju@yonsei.ac.kr

\*\*\*\* 종신회원 : (주) NHN  
jblee@nhncorp.com

논문접수 : 2008년 12월 23일

심사완료 : 2009년 3월 4일

Copyright©2009 한국정보과학회 : 개인 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 : 컴퓨팅의 실제 및 레터 제15권 제5호(2009.5)

상을 효과적으로 분석할 수 있는 새로운 정보 파급 모델을 제안한다. 또한, 실제 블로그 연결망에서 제안하는 기법과 기존의 기법을 이용하여 정보의 파급 분석의 수행 성능을 비교함으로써 제안하는 기법의 우수성을 검증한다.

**키워드 :** 사회연결망 분석, 블로그, 데이터 마이닝, 정보 파급, 정보 파급 모델

**Abstract** Analyzing information diffusion in a blog world is a very useful research issue, which can be used for predicting information diffusion, abnormally detection, marketing, and revitalizing the blog world. Existing studies on information diffusion in blog networks establish explicit relationship between blogs, and analyze only the word-of-mouth effect through such explicit relationships. However, we observed that more than 85% of all information diffusion in a blog world occurs through non-explicit relationships. In this paper, we propose a new model that considers both explicit and non-explicit relationships between blogs in order to explain all information diffusion phenomena in a blog world. We verify the superiority of our proposed models through extensive experiments of information diffusions at a real blog network.

**Key words :** Social Network Analysis, Blog, Data Mining, Information Diffusion, Information Diffusion Model

## 1. 서 론

온라인 사회연결망을 이용한 대표적인 서비스로서 블로그 서비스(blog service)를 들 수 있다. 블로그(blog)는 블로그의 소유자인 블로거(blogger)가 자신의 생각을 온라인상에 게시할 수 있는 일종의 개인 웹사이트이다 [1-4]. 블로거가 작성한 게시글(post)은 다른 블로거들에게 유용한 정보로서 활용될 수 있다. 따라서 블로그에는 필요한 정보를 쉽게 얻거나 친분 관계를 유지하기 위해서 관심이 있는 다른 블로그들과 관계를 맺을 수 있는 기능이 존재한다. 이는 웹 브라우저의 '바로가기'와 유사한 기능으로서 원하는 블로그로 언제든지 쉽게 방문을 할 수 있도록 도와주는 역할을 한다. 본 논문에서는 이러한 기능과 같이 블로그들 사이에 설정된 관계를 명시적 관계(explicit relationship)라고 정의한다. 또한, 블로그들과 명시적 관계로 구성된 온라인 사회 연결망을 블로그 월드(blog world)라고 정의 한다.

한국의 대표적인 블로그 서비스들[1-4]에서는 블로그 사용자가 다른 사용자의 게시글을 자신의 블로그로 복사해 올 수 있는 기능을 지원하고 있는데, 이를 스크랩 기능이라고 한다. 스크랩과 트랙백 기능이 지원되는 블로그 월드내에서는 이들을 통해서 게시글들의 파급이 빈번하게 발생한다. 블로그 월드에서 발생하는

이러한 정보의 파급을 분석하는 것은 정보 파급의 예측, 이상 정보의 검출, 마케팅에의 응용, 블로그 월드의 활성화 등에 이용될 수 있는 매우 유용한 연구 이슈이다[5,6].

블로그 월드에서 정보 파급 분석에 관한 기존의 연구들은 주로 사회 연결망의 구성원들을 통한 입소문 효과를 통해서 정보의 파급이 일어나는 현상들을 설명하는데 초점을 맞추고 진행되었다[7]. 블로그 사용자들은 자신과 명시적 관계를 맺고 있는 사용자들의 블로그를 방문하여 게시글을 파급할 수 있으므로, 이러한 접근방법은 일면 타당하다고 볼 수 있다. 그러나 참고문헌 [8]의 분석 결과, 블로그 월드 내에 존재하는 게시글의 파급 현상들의 85%가 입소문 효과 이외의 방법을 통해서 파급된 것임이 밝혀졌다. 그 대표적인 예로서, 특정 게시글이 블로그 서비스 업체의 메인 페이지에 소개됨으로써 명시적 관계가 존재하지 않는 블로그들에게 폭발적으로 파급되는 현상을 들 수 있다[8].

명시적 관계들을 통해서 발생하는 정보의 파급은 일반적으로 파급에 참여한 블로그들의 수가 선형적으로 증가하는 경향을 보인다. 그러나 비명시적 관계를 통해서 발생하는 정보의 파급은 파급에 참여한 블로그들의 수가 급격하게 증가하는 경향을 보인다. 그림 1은 이러한 파급 경향을 보인 것으로서 x축은 시간, y축은 정보의 파급력을 누적한 값을 나타낸다. 여기서 정보의 파급력이란 해당 게시글에 의해 직간접적으로 동화된 사용자의 수를 의미한다. 시간이 지남에 따라 비명시적 관계를 통해서 발생한 정보의 파급력이 명시적 관계를 통한 정보의 파급력에 비하여 300배 이상 폭발적으로 증가함을 알 수 있다.

기존의 정보 파급 분석을 위한 연구들[5,9~12]은 명시적 관계를 통한 점진적인 파급 현상만을 대상으로 하고

있으므로 그림 1의 예에서와 같은 정보의 폭발적인 파급 현상은 설명할 수 없다. 그러나 이러한 현상은 블로그 월드내에서 발생하는 정보파급의 대부분을 차지하고 있다. 따라서 본 논문에서는 이러한 현상을 모델링할 수 있는 새로운 정보 파급 모델을 제안한다. 이를 위하여 기존의 정보 파급 모델인 독립 전파 모델(independent cascade model)[5]을 기반으로 하여, 새로운 요소들을 추가한 모델을 제안한다. 여기서 새로운 요소들이란, 블로그 서비스의 메인 웹페이지에 게시글을 노출시키는 공간, 이 공간과 각 사용자들 간의 관계를 의미한다. 또한, 제안하는 모델의 정확도를 높이기 위한 방안에 대하여 논의하고, 이를 반영한 확장 모델을 제안한다. 끝으로, 실험을 통하여 제안하는 기법의 정확도를 검증한다.

## 2. 제안하는 기법

### 2.1 기본 정보 파급 모델

본 절에서는 명시적 관계와 비명시적 관계를 통한 정보의 파급을 분석할 수 있는 기본 정보 파급 모델을 제안한다.

기본 정보 파급 모델에서는 명시적 관계를 통한 정보의 파급 현상을 설명하기 위해서 기존의 정보 파급 모델을 이용한다. 기존의 정보 파급 모델에서는 블로그들 간에 명시적 관계를 설정하고, 동화확률로써 적절한 값을 부여하여 블로그들과 명시적 관계들을 이용하여 블로그 연결망을 생성하였다. 기존의 정보 파급 분석 기법은 이러한 방법으로 생성된 블로그 연결망을 대상으로 독립 전파 모델을 이용하여 정보의 파급을 분석한 것이다.

또한, 기본 정보 파급 모델에서 비명시적 관계를 통한 정보의 파급 현상을 설명하기 위해서 기존의 명시적 관계만을 고려한 파급 모델에 슈퍼노드, 방송 에지, 등록 에지 등의 새로운 개념을 추가하였다. 여기서, 슈퍼노드는 블로그 서비스 제공 업체의 메인 웹페이지를 모델링 한 것으로서, 블로그 연결망 내의 블로그들과 마찬가지로 정보들을 게시하고 있는 공간이다. 그러나 슈퍼노드는 블로그들과 달리 블로그 연결망 내의 모든 블로그들과 양방향의 관계들을 갖고 있다. 이러한 관계들 중, 슈퍼노드로부터 블로그들에게 향하는 관계를 방송 에지라고 부르며, 블로그들로부터 슈퍼노드로 향하는 관계를 등록 에지라고 부른다. 슈퍼노드는 방송 에지를 통해서 블로그들을 동화시킬 수 있고, 등록 에지를 통해서 블로그들에게 동화될 수 있다. 이 때, 슈퍼노드가 블로그에게 동화된다는 것은 블로그 서비스 업체의 메인 페이지 관리자에 의해서 해당 블로그의 정보가 선정되었다는 것을 의미한다. 그림 2는 제안하는 기본 정보 파급 모델을 보인 것이다.

앞으로의 논의 전개를 위해서 필요한 용어 및 기호들을 정리하면 다음과 같다. SN은 슈퍼노드를 의미하고,

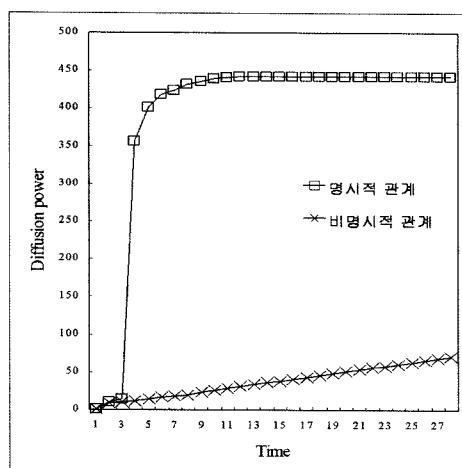


그림 1 정보의 폭발적 파급의 예

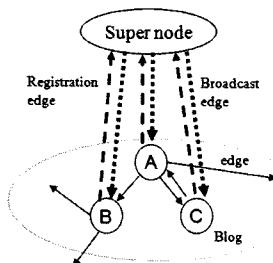


그림 2 기본 정보 파급 모델

$U_i$ 는 식별자가  $i$ 인 사용자를 의미한다.  $P_{i \rightarrow j}$ 는  $U_i$ 의 정보가  $U_j$ 에게 파급될 확률 즉,  $U_i$ 가  $U_j$ 에게 동화될 확률을 의미한다.  $D_{SN}$ 은 SN에 등록된 게시글들의 집합을 의미하고, SN의 노출 가능한 정보의 개수를  $m$ 이라고 할 때,  $D_{SN}$ 에는  $m$ 개의 게시글이 포함될 수 있다.  $D_i$ 는  $U_i$ 가 소유한 게시글들의 집합을 의미하고,  $D_{i,j}$ 는  $U_i$ 의  $j$ 번째 게시글을 의미한다. BE는 방송에지들의 집합을 의미한다.  $BE_i$ 는 SN으로부터  $U_i$ 로 향하는 방송에지를 의미하고,  $BE_i$ 의 동화확률은  $P_{SN \rightarrow i}$ 으로 나타낸다. RE는 등록에지들의 집합을 의미한다. RE는  $U_i$ 로부터 SN으로 향하는 등록에지를 의미하고, RE의 동화확률은  $P_{i \rightarrow SN}$ 으로 나타낸다. 블로그 연결망 내의 전체 블로그들의 수를  $n$ 이라고 할 때, BE와 RE는 각각  $n$ 개의 에지들을 포함한다.

블로그 연결망 내에 존재하는 관계들의 동화확률은 다음과 같은 방법으로 계산한다. 식 (1)은 블로그들간의 명시적 관계들의 동화확률 즉,  $U_i$ 가  $U_j$ 를 동화시킬 확률  $P_{i \rightarrow j}$ 를 계산하는 방법을 나타낸 것이다. 블로그 사용자들은 다른 사용자들이 관심을 가질만한 게시글을 작성할 뿐 아니라, 사생활에 대한 게시글과 같이 다른 사용자들이 관심을 갖지 않는 게시글도 작성할 수 있다. 따라서 본 연구에서는  $U_i$ 의 게시글이  $U_j$ 에게 파급될 확률로서  $D_i$  중에서 다른 사용자들이 관심을 갖고 있는 게시글들과 이 중에서 실제로  $U_j$ 에게 파급된 게시글들의 비율을 부여한다. 이 때,  $D_i$  중에서 다른 사용자들이 관심을 갖고 있는 게시글들을 식별하기 위해서 파급이력을 분석하여 임계값  $th$  이상 파급을 발생한 게시글들만을 식별하는 방법을 이용한다.

$$P_{i \rightarrow j} = \frac{D_j \text{ 중 } D_i \text{로부터 파급된 게시글의 수}}{D_i \text{ 중 파급력이 } th \text{ 이상인 게시글의 수}} \quad (1)$$

식 (2)은 방송에지들의 동화확률을 계산하는 방법을 나타낸 것이다. 슈퍼노드가  $U_i$ 를 동화시킬 확률  $P_{SN \rightarrow i}$ 는  $DSN$  중에서  $U_i$ 에게 실제로 파급된 게시글들의 비율로 계산하였다.

$$P_{SN \rightarrow i} = \frac{D_i \text{ 중 } D_{SN} \text{으로부터 파급된 게시글의 수}}{|D_{SN}|} \quad (2)$$

식 (3)은 등록 에지들의 동화확률을 계산하는 방법을 나타낸 것이다.  $U_i$ 의 게시글이 슈퍼노드를 동화시킬 확률  $P_{i \rightarrow SN}$ 은  $D_i$  중에서 다른 사용자들이 관심을 갖고 있는 게시글들과 이 중에서 실제로 슈퍼노드에 파급된 게시글들의 비율로 계산한다. 식 (1)에서와 마찬가지로  $D_i$  중에서 다른 사용자들이 관심을 갖고 있는 게시글들을 식별하기 위해서 파급이력을 분석하여 임계값  $th$  이상 파급을 발생한 게시글들만을 식별하는 방법을 이용한다.

$$P_{i \rightarrow SN} = \frac{D_{SN} \text{ 중 } D_i \text{로부터 파급된 게시글의 수}}{D_i \text{ 중 파급력이 } th \text{ 이상인 게시글의 수}} \quad (3)$$

## 2.2 확장 정보 파급 모델

본 절에서는 기본 정보 파급 모델의 성능을 개선하기 위한 방안에 대하여 논의하고 이를 반영한 확장 정보 파급 모델을 제안한다.

슈퍼노드에 의해서 노출되는 게시글들에 대하여 블로그 사용자들은 다양한 관심을 보일 수 있다. 특정 게시글에 대해서는 많은 사용자들이 관심을 보이고 해당 게시글을 파급할 수 있는 반면, 다른 게시글에 대해서는 대다수의 사용자들이 관심을 보이지 않을 수 있다. 그림 3은  $DSN$ 에 존재하는 게시글들의 파급력 분포를 보인 것이다. x축은 파급력을 기준으로 한 게시글의 순위, y축은 해당 게시글의 파급력을 나타낸다. 그림 3을 통해서  $DSN$ 에 존재하는 게시글들의 파급력 분포가 역함수를 따르고 있음을 알 수 있다.

기본 정보 파급 모델에서는 슈퍼노드에 의해서 특정 사용자가 동화될 확률을 계산할 때,  $DSN$ 에 존재하는 게시글들의 파급력의 차이를 고려하지 않고,  $DSN$ 의 게시글들이 파급되는 방송에지  $BE_i$ 에 단 하나의 동화확률  $P_{SN \rightarrow i}$ 만을 부여하였다. 이로 인해 파급력이 큰 게시글과 파급력이 작은 게시글이 파급 시에 동일한 동화확률을 갖게 됨으로써 정보 파급의 분석 정확도가 떨어지게 된다.

본 논문에서는 기본 정보 파급 모델의 이러한 문제점을 극복한 확장 정보 파급 모델을 제안한다. 본 연구에서는  $DSN$ 에 존재하는 게시글들 중에서 동일한 사용자가

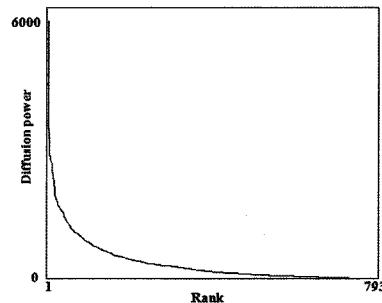


그림 3 DSN의 파급력 분포

작성한 게시글들은 서로 유사한 파급력을 갖는다는 가정한다. 식 (4)는  $U_i$ 가 작성하여  $D_{SN}$ 에 포함된 게시글들이  $U_j$ 에게 파급될 확률  $P_{i \rightarrow SN \rightarrow j}$ 을 계산하는 방법을 나타낸 것이다.  $P_{i \rightarrow SN \rightarrow j}$ 는  $P_{SN \rightarrow j}$ 에  $a_i$ 를 곱하여 계산한다. 여기서,  $a_i$ 는  $U_i$ 가 작성하여  $D_{SN}$ 에 포함된 게시글들의 동화확률을 보정계수이다.

$$P_{i \rightarrow SN \rightarrow j} = a_i * P_{SN \rightarrow j} \quad (4)$$

식 (5)은  $U_i$ 에 대한 동화확률 보정계수  $a_i$ 를 계산하는 방법을 나타낸 것이다.  $a_i$ 는  $U_i$ 가 작성하여  $D_{SN}$ 에 포함된 게시글들의 평균 파급력을  $D_{SN}$ 의 모든 게시글들의 평균 파급력으로 나누어 계산한다.

$$a_i = \frac{D_i \text{ 중 } D_{SN} \text{에 포함되는 정보들의 평균 파급력}}{D_{SN} \text{의 평균 파급력}} \quad (5)$$

확장 정보 파급 모델을 이용하여 정보의 파급 현상을 분석하기 위한 과정은 다음과 같다. 1) 정보 파급 모델을 이용하여 블로그 연결망을 구축한다. 2) 슈퍼노드에 존재하는 게시글들의 작성자에 대한 동화확률 보정계수를 계산하고, 이 값을 반영하여 슈퍼노드에 존재하는 게시글들의 동화확률을 부여한다. 3) 블로그 연결망을 대상으로 독립 전파 모델을 수행하여 정보의 파급을 분석한다.

### 3. 성능 평가

#### 3.1 실험 환경

본 논문에서는 실험을 위해 국내에서 가장 큰 블로그 월드 중 하나인 네이버 블로그[4]에서 2006년 7월부터 수개월간 수집하여 익명으로 처리한 데이터를 사용하였으며, 블로그 간의 명시적 관계로서 블로그 서비스를 시작한 시점부터 생성된 이웃 관계를 사용하였다. 블로그 서비스 업체의 메인 페이지에는 24시간을 주기로 15개의 게시글들이 관리자에 의해 수집되어 사용자들에게 노출되었고, 분석기간 중에는 총 1,200여개의 게시글들이 수집되어 사용자들에게 노출되었다.

본 연구에서 성능 평가의 대상으로 선정한 정보 파급 모델은 블로그들 간의 명시적 관계만을 고려한 기준의 모델을 이용한 기법(Previous), 본 논문에서 제안하는 기본 정보 파급 모델을 이용한 기법(BID), 기본 정보 파급 모델의 성능을 개선한 확장 정보 파급 모델을 이용한 기법(EID) 등 세가지이다.

본 연구에서는 위의 세가지 기법의 성능을 비교하기 위하여 성능 평가의 척도로서 오차값 Error와 오차 개선율 ImprovedErrorRatio를 사용한다. 오차값은 블로그 연결망 내의 특정 사용자로부터 실제로 정보가 파급된 결과와, 지정한 정보 파급 모델을 이용하여 정보의 파급을 시뮬레이션한 결과와의 차이를 계산한 것이다. 또한,

오차 개선율은 두 개의 정보 파급 모델을 이용하여 각각 특정 사용자로부터 정보의 파급을 시뮬레이션 한 결과의 오차값을 비교한 것이다.

식 (6)은 오차값을 계산하는 방법을 나타낸 것이다. M은 블로그 연결망을 생성하는데 사용한 정보 파급 모델을 의미한다.  $DH_i$ 는 파급이력 중에서  $U_i$ 로부터 실제로 정보가 파급된 결과만을 추출한 것이고, Simulation( $M, U_i$ )은 M을 이용하여 생성된 블로그 연결망에서 독립 전파 모델을 이용한 시뮬레이션을 통하여  $U_i$ 로부터의 정보 파급을 수행한 결과이다.  $DH_i$ 와 Simulation( $M, U_i$ )은 시간의 흐름에 따라  $U_i$ 에 의해 직간접적으로 동화된 사용자의 수를 누적하여 기록한 것으로서 시계열 데이터의 특성을 갖는다. 따라서 본 논문에는  $DH_i$ 와 Simulation( $M, U_i$ )의 차이를 측정하기 위해서 시계열 데이터간의 유용한 거리함수인 DTW[13]를 사용하였다.

$$\text{Error}(M, U_i) = DTW(DH_i, \text{Simulation}(M, U_i)) \quad (6)$$

식 (7)은 오차 개선율을 계산하는 방법을 나타낸 것이다. M1, M2는 각각 정보 파급 모델을 의미한다. M1과 M2를 이용하여  $U_i$ 의 정보의 파급을 시뮬레이션한 결과, M1의 M2에 대한 오차 개선율은  $\text{Error}(M1, U_i)$ 를  $\text{Error}(M2, U_i)$ 로 나누어 계산한다. 오차 개선율의 값이 0 이상 1미만인 경우, M1이 M2에 비하여 적은 오차를 유발한다는 것을 의미하고, 1보다 클 경우 M1이 M2에 비하여 많은 오차를 유발한다는 것을 의미한다.

$$\text{ImprovedErrorRatio}(M, M2, U_i) = \frac{\text{Error}(M, U_i)}{\text{Error}(M2, U_i)} \quad (7)$$

#### 3.2 실험 결과

본 논문에서는 두 종류의 실험을 수행하였다. 실험 1에서는 BID와 Previous의 성능을 비교하였고, 실험 2에서는 EID와 Previous의 성능을 비교하였다.

실험 1에서 BID와 Previous의 성능을 비교하기 위하여 BID의 Previous에 대한 오차 개선율을 분석하였다. 그림 4는 실험 1의 결과를 보인 것이다. 점들은 슈퍼노드에 존재하는 게시글들을 나타내고, x축은 해당 게시글의 파급력, y축은 각 게시글의 파급 시뮬레이션을 수행한 결과, BID의 Previous에 대한 오차 개선율을 나타낸다. 실험 결과, BID의 Previous에 대한 오차 개선율의 평균값은 1.65로 나타났다. 이것은 BID가 Previous에 비하여 오차를 65%가량 더 발생한다는 것을 의미한다. 이러한 결과가 나타난 이유는 슈퍼노드에는 파급력이 다양한 게시글들이 포함되어 있는데, BID는 이러한 요인을 고려하지 않았기 때문이다.

실험 2에서는 EID와 Previous의 성능을 비교하기 위하여 EID의 Previous에 대한 오차 개선율을 분석하였다. 그림 5는 실험 2의 결과를 보인 것이다. 실험 1에서

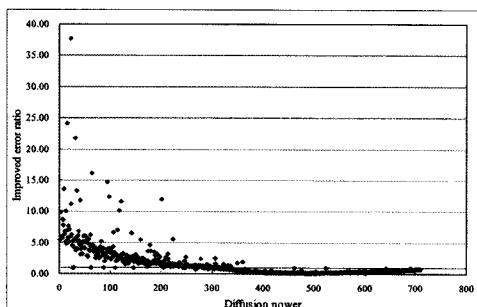


그림 4 오차 개선율(BID vs Previous)

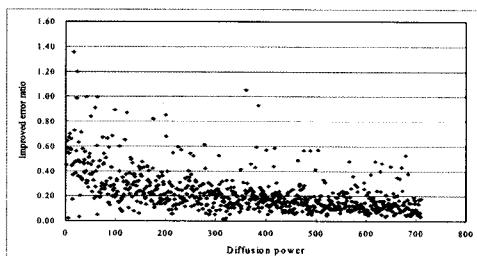


그림 5 오차 개선율(EID vs Previous)

와 마찬가지로 점들은 슈퍼노드에 존재하는 게시글들을 나타내고, x축은 해당 게시글의 파급력, y축은 각 게시글의 파급 시뮬레이션을 수행한 결과, EID의 Previous에 대한 오차 개선율을 나타낸다. 실험 결과, EID의 Previous에 대한 오차 개선율의 평균값은 0.23으로 나타났다. 이것은 EID가 Previous에 비하여 오차를 77%가량 적게 발생한다는 것을 의미한다. EID는 BID와 달리, 슈퍼노드에 존재하는 게시글들의 파급력이 다양한 경우에도 게시글들의 파급 현상을 잘 설명할 수 있는 모델임을 알 수 있다.

#### 4. 결 론

본 논문에서는 명시적 관계, 비명시적 관계를 통한 정보의 파급을 함께 고려하여 이를 관계를 통한 정보의 파급 현상을 분석할 수 있는 기본 정보 파급 모델을 제안하였다. 기본 정보 파급 모델은 기존의 정보 파급 모델에 블로그 서비스 업체의 메인 페이지를 모델링한 슈퍼노드와 방송에지, 등록에지를 추가한 것이다. 또한, 기본 파급 모델을 통해서 생성한 블로그 연결망의 각 관계들에 실질적인 동화확률을 부여하는 방안을 제안하였다.

기본 정보 모델은 슈퍼노드에 존재하는 게시글들의 파급력이 다양한 경우, 이러한 요인을 고려하고 있지 않으므로 분석 성능이 떨어지게 된다. 따라서 본 논문에서는 슈퍼노드에 존재하는 각 게시글들의 다양한 파급력

을 고려하여 정보 파급의 분석 성능을 향상시킨 확장 정보 파급 모델을 제안하였다.

실험을 통하여 제안하는 기법이 기존의 정보 파급 모델에 비하여 우수한 성능을 보임을 검증하였다. 실험 결과, 확장 정보 파급 모델은 게시글들의 파급력이 다양한 환경에서 기존의 정보 파급 모델에 비하여 오차를 77% 가량 적게 발생하는 것으로 나타났다.

#### 참 고 문 헌

- [1] (주)SK커뮤니케이션, <http://www.cyworld.com>.
- [2] (주)엠파스, <http://blog.empas.com>.
- [3] (주)아이세이브존, <http://www.isavezone.com>.
- [4] (주)NHN, 블로그홈, [blog.naver.com](http://blog.naver.com).
- [5] J. Goldenberg, B. Libai, and E. Muller, "Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth," *Marketing Letters*, Vol.12, No.3, pp. 211-223, 2001.
- [6] (주)위고넷 이브랜딩 전략연구소, 커뮤니티를 승부하는 브랜드 전략, 이디자인, 2004.
- [7] J. Brown and P. Reinegen, "Social Ties and Word-of-Mouth Referral Behavior," In *Proc. Journal of Consumer Research*, JCR, Vol.1, No.3, pp. 350-362, 1987.
- [8] 권용석, 김상욱, 박신주, "블로그 월드에서 정보 파급 분석", 한국정보처리학회 춘계학술 발표대회 Vol.15, No.1, pp. 223-226, 2008.
- [9] 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.
- [10] M. Granovetter, "Threshold Models of Collective Behavior," In *Proc. American Journal of Sociology*, Vol.83, No.6, pp. 1420-1443, 1978.
- [11] D. Gruhl et al., "Information Diffusion Through Blogspace," In *Proc. Int'l. Conf. on World Wide Web*, WWW, pp. 491-501, 2004.
- [12] A. Java et al., "Modeling the Spread of Influence on the Blogosphere," In *Proc. Int'l. Conf. on World Wide Web*, WWW, 2006.
- [13] C. Ratanamahatana and E. Keogh, "Making Time-series Classification more Accurate using Learned Constraints," In *Proc. Int'l. Conf. on Society for Industrial and Applied Mathematics*, SIAM, pp. 11-22, 2004.