시간의 흐름에 따른 한국 블로그 공간의 변화 분석

하지운, 배덕호, 김상욱
한양대학교 전자컴퓨터통신공학과
e-mail : oneofus@agape.hanyang.ac.kr

Analyzing a Korean Blogosphere over Time

Jiwoon Ha, Duck-Ho Bae, Sang-Wook Kim
Dept. of Electronics and Computer Engineering, Hanyang University

요 약

블로그 공간은 블로거들의 이동 및 블로거들의 행동에 의해 시간에 따라 지속적으로 변화한다. 이러한 블로그 공간의 변화를 분석함으로써 블로그 공간이 발전해 온 경향을 알 수 있으며, 나아가 향후 발전 과정을 예측할 수 있다. 본 논문에서는 한국 블로그 공간의 구조적 특성이 시간의 흐름에 따라 변화하는 모습을 심도 있게 분석한다.

1. 서론

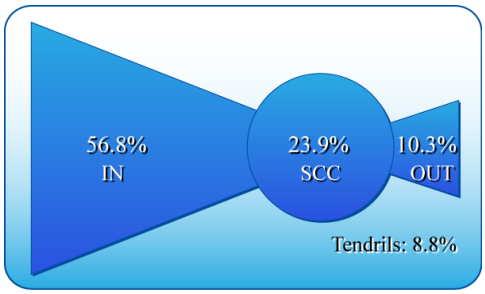
블로그 공간(blogosphere)은 블로그 공간의 사용자인 블로거, 블로그, 포스트, 그리고 작성된 포스트에 대해 블로거들이 행한 액션으로 구성되어있다 [Her05]. 블로그 공간은 시간이 경과함에 따라 지속적으로 변화한다. 블로그 공간에 신규 블로거들이 유입되기도 하며, 기존의 블로거들이 탈퇴하기도 한다. 또한, 블로거들은 자신의 생각을 알리기 위해 새로운 포스트를 작성하거나, 관심이 있는 포스트들에 대하여 액션들을 수행하기도 한다. 이러한 개별 블로거들의 행동 패턴은 전체 블로그 공간의 특성 변화에 반영된다. 이러한 블로그 공간의 변화를 분석함으로써 블로그 공간이 발전해 온 경향을 알 수 있으며, 나아가 향후 발전 과정을 예측할 수 있다.

한국 블로그 공간은 기존의 블로그 공간과는 다른 고유한 특징을 가지고 있다 [Ha11]. 한국 블로그 공간의 고유한 특징의 영향으로, 한국 블로그 공간의 구조적 특성은 웹 공간 또는 기존의 블로그 공간의 구조적 특성과 차이를 보인다 [Ha11]. 본 논문에서는 이러한 한국 블로그 공간의 구조적 특성이 시간의 흐름에 따라 변화하는 모습을 분석하고, 더 나아가 한국 블로그 공간의 블로거들의 행동 패턴을 분석한다

2. 관련 연구

2.1 Bow-tie 구조

참고 문헌 [Bro00]은 웹 연결망을 거시적 관점에서 분석하기 위한 Bow-tie 구조를 제안하였다. Bow-tie 구조는 크게 SCC, IN, OUT, Tendrils 의 4 가지 컴포넌트로 구성된다. 그림 1 은 한국 블로그 공간의 Bow-tie 구조를 나타낸 것으로 각 컴포넌트 내의 숫자는 해당 컴포넌트에 속하는 노드 수의 비율을 의미한다.



(그림 1) 한국 블로그 공간의 Bow-tie 구조 [Ha11].

2.2 Entropy plot

참고 문헌 [Got09][Mcg07][Mcg08]는 entropy plot [Wan02]을 통해 블로그 연결망의 burstiness 를 측정하였다. Entropy plot 은 resolution W 에 대한 특정 현상의 entropy $H(W)$의 분포로 나타낸다. Resolution 이란 전체 구간을 몇 개의 세부 구간으로 나누었는가를 의미하고, Resolution W 에서의 entropy $H(W)$는 식 (1)과 같이 정의된다 [Wan02]. 여기서, $p_{i,w}$ 는 전체 구간에 대한 i 구간에서의 현상의 수의 비율을 의미한다. 즉, $pi,w = ni,w / n$.

$$H(W) = -\sum_i p_{i,w} \log_2(p_{i,w}) \quad \text{..... (1)}$$

Entropy plot 은 $H(W)$를 세로축으로, $\log_2(W)$를 가로축으로 하는 그래프로 정의한다. Entropy plot 의 자료들이 선형 함수의 모습으로 그려지는 경우, 해당 현상은 self-similar 하다고 정의한다 [Wan02]. 특정 현상이 self-similar 하다는 것은 전체 구간에서의 해당 현상의 패턴이 세부 구간에서도 동일하게 나타남을 뜻한다. Self-similarity 는 프랙탈의 대표적 특성이기도 하다.

또한, entropy plot 의 요소들에 대한 선형 함수 회귀선을 그렸을 때, 그 선형 함수의 기울기 f 를 information fractal dimension 이라 한다 [Wan02]. 기울기 f 를 통해 해당 현상의 burstiness 를 알 수 있다. f 가 1 이라면 해당 현상은 uniform 하게 일어남을 의미하고, f 가 0 이라면 해당 현상이 정확히 동일한 시각에 일어났음을 의미한다.

3. 시간의 흐름에 따른 Bow-tie 구조의 변화

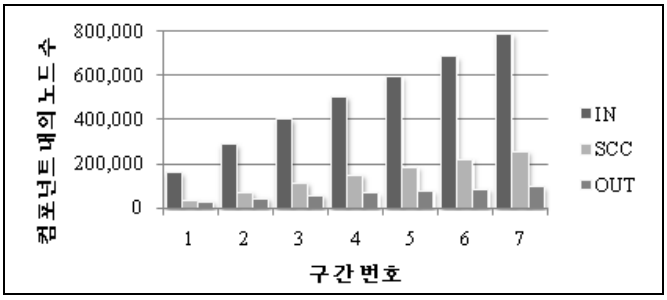
본 연구에서는 분석을 위해 한국의 최대 블로그 서비스로부터 수개월간 수집하여 익명으로 처리한 데이터를 사용하였다. 사용된 데이터는 네이버 블로그의 일부로서 블로거 약 170 만 명, 포스트 약 1 억 개, 스크랩 회수 약 4 천만 회이다. 또한, 시간에 따른 변화를 분석하기 위해 전체 데이터를 7 개의 시간 구간으로 분할하였다.

그림 2(a)는 시간의 흐름에 따른 블로그 연결망의 Bow-tie 컴포넌트의 비율 변화를 나타낸 것이다. Bow-tie 구조의 약 90%에 해당하는 핵심적 요소인 IN, SCC, OUT 을 대상으로 실험을 수행하였다. 가로 축은 각 구간, 세로 축은 Bow-tie 구조 각 컴포넌트 내의 노드 수를 나타낸다. 그림 2(b)는 블로그 연결망의 Bow-tie 컴포넌트의 상대적 비율 변화를 나타낸 것이다. 가로 축은 각 구간, 세로 축은 Bow-tie 구조에서 각 컴포넌트가 차지하는 비율을 나타낸다. 블로그 공간의 성장에 따른 Bow-tie 구조의 변화를

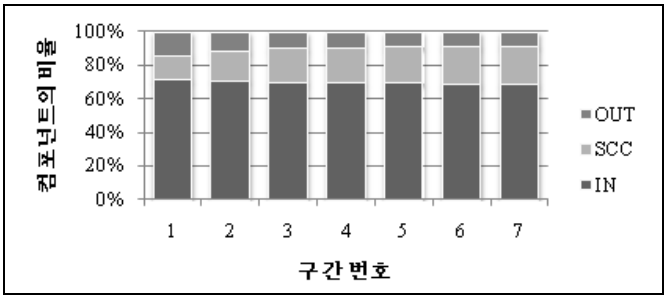
분석하기 위해 각 구간의 데이터를 누적하여 실험하였다.

그림 2(a)에서 보이는 것과 같이, 시간의 흐름에 따라 IN 내의 노드 수는 빠르게 증가하고, SCC 내의 노드 수도 점진적으로 증가하고 있으며, OUT 내의 노드 수는 아주 느리게 증가하고 있음을 알 수 있다. IN 내의 노드들은 주로 다른 블로거의 포스트를 스크랩하는 블로거들로, 이는 새롭게 유입된 블로거들의 특성으로 볼 수 있다. 한국 블로그 서비스 업체들이 제공하는 편리한 블로그 생성 및 스크랩 기능이 신규 블로거들이 쉽게 블로그 공간에 진입할 수 있는 기회를 제공하고 있음을 알 수 있다.

IN 내의 노드들의 수가 가장 빠른 속도로 증가하고 있는 것으로 보이는 것과 달리, Bow-tie 컴포넌트의 비율의 측면에서는 SCC 의 비율이 점점 증가하고 있음을 알 수 있다(그림 2(b)). SCC 에 속하는 노드들은 블로그 공간에서 서로 영향을 주고 받으며 활발하게 활동하는 블로거들을 의미한다. 따라서 블로그 공간이 성장함에 따라 활발하게 활동하는 블로거들의 비율이 높아지고, 일방적으로 스크랩을 하기만 하거나 스크랩을 받기만 하는 블로거의 비율은 상대적으로 줄어드는 것을 알 수 있다. 이는 한국의 블로그 공간이 양적 성장과 더불어 질적 성장도 함께 하고 있음을 보여주는 것이다.



(a) Bow-tie 컴포넌트 내의 노드 수 변화.



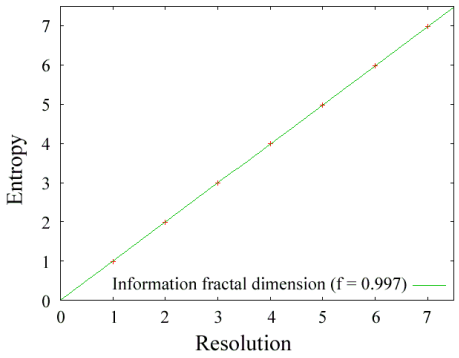
(b) Bow-tie 컴포넌트의 비율 변화.

(그림 2) 시간의 흐름에 따른 블로그 연결망의 Bow-tie 구조 변화.

4. 블로거 포스트 생성 패턴 분석

블로그 공간에서 블로거의 포스트 생성 패턴을 확인하기 위해 포스트 생성 수의 entropy plot 을 구하고 그에 따른 information fractal dimension 을 측정하였다. 실험 결과, 그림 3 과 같이 포스트 생성 수의 entropy plot 은 선형 함수의 형태로 나타났다. 이는 블로그 공간에서의 포스트 생성이 self-similar 한 특성을 가짐을 의미한다. 또한, 포스트 생성 수의 information fractal dimension 값은 기존의 블로그 공간에서의 information fractal dimension 값인 0.88 [Got09], 0.72 [Mcg07]에 비해 0.997 로 1 에 더 가까운 값을 나타냈는데, 이는 한국 블로그 공간에서의 포스트 생성이 다른 블로그 공간에서의 포스트 생성보다 더 uniform 하게 나타남을 의미한다. 포스트의 생성이 uniform 하게 일어났다는 것은 일 별 생성 포스트 수가 서로 유사함을 의미한다.

이는 특별한 사회적 이슈가 존재 할 때 포스트 생성이 급격하게 증가 할 것이라는 직관과는 다른 결과로, 사회적 이슈와 관계 없이 포스트 생성 수가 거의 일정함을 뜻한다. 이로부터, 한국 블로그 공간의 블로거들은 자신이 관심 있는 고유한 주제에 대한 연재 형식으로 포스트를 생성하거나 자신의 일상에 대한 내용에 대한 포스트를 생성하는 것으로 생각할 수 있다.



(그림 3) 포스트 생성 수의 entropy plot.

5. 결론

본 논문에서는 시간의 흐름에 따른 Bow-tie 컴포넌트의 비율 변화 분석을 통해 한국 시간의 흐름에 따른 블로그 공간의 특성 변화를 분석하고, entropy plot 을 통해 블로거들의 행동 패턴에 대한 분석을 수행하였다.

감사의 글

본 연구는 2010 년도 정부(교육과학기술부)의 재원으로 한국연구재단 (No.2008-0061006) 및 지식경제부 및 정보통신산업진흥원의 'IT 융합 고급인력과정 지원사업' (NIPA-2011-C6150-1101-0001) 및 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

[Her05] S. Herring et al., "Conversations in the Blogosphere: An Analysis "From the Bottom Up"," In *Proc. the 38th Annual Hawaii Int'l. Conf. on System Sciences*, pp. 107b-107b, 2005.

[Ha11] J. Ha et al., “Analyzing a Korean Blogosphere: A Social Network Analysis Perspective,” In Proc. the 26th Symposium on Applied Computing, pp. 771-775, 2011.

[Bro00] A. Broder et al., "Graph Structure in the Web," Computer Networks, Vol. 33, No. 1, pp. 309-320, 2000.

[Got09] M. Götz et al., “Modeling Blog Dynamics,” In Proc. Int’l. Conf. on Weblogs and Social Media, ICWSM, 2009.

[Mcg07] M. McGlohon et al., “Finding Patterns in Blog Shapes and Blog Evolution,” In Proc. Int’l. Conf. on Weblogs and Social Media, ICWSM, 2007

[Mcg08] M. McGlohon, L. Akoglu, and C. Faloutsos, “Weighted Graphs and Disconnected Components: Patterns and a Generator,” In Proc. ACM Int’l. Conf. on Knowledge Discovery and Data Mining, ACM SIGKDD, pp. 524-532, 2008.

[Wan02] M. Wang et al., “Data Mining Meets Performance Evaluation: Fast Algorithms for Modeling Bursty Traffic,” In Proc. IEEE Int’l Conf. on Data Mining, ICDE, pp. 507-516, 2002.