

사용자 이분그래프모형을 이용한 온라인 커뮤니티 토론 네트워크의 균집성과 극성 분석[☆]

Cluster and Polarity Analysis of Online Discussion Communities Using User Bipartite Graph Model

김 성 환¹ 탁 해 성¹ 조 환 규^{1*}
Sung-Hwan Kim Haesung Tak Hwan-Gue Cho

요 약

온라인 커뮤니티에서는 많은 수의 참여자들이 시공간적인 제약을 받지 않고 서로간의 다양한 의견을 댓글로 교환한다. 온라인 공간은 시공간적인 제약으로부터 자유롭기 때문에 신속하고 자유로운 의사소통을 가능하게 하지만, 동시에 불필요한 언쟁과 갈등을 쉽게 유발시킬 수 있다는 문제점이 있다. 토론 과정에서 형성되는 참여자 간의 네트워크는 참여자들 간의 대립 양상을 파악하고 앞으로 일어날 분쟁을 예측하여 방지하기 위한 중요한 단서가 된다. 본 논문에서는 온라인 커뮤니티에서의 댓글 교환으로 나타나는 사용자 토론 네트워크상에서 관찰되는 집단의 극성을 분석하기 위한 이분그래프 기반의 정량적 지표를 제안한다. 제안 기법은 댓글 교환 정보를 이용하여 사용자 상호작용 네트워크 그래프를 구성하고, 구성된 그래프 상에서 최대신장트리를 구한 후 버텍스 컬러링을 통하여 사용자를 두 부분집합으로 분할한다. 분할된 사용자 집합 간의 댓글 교환 비율을 이용하여 극성 지표를 계산함으로써 주어진 토론의 참가자들이 양분화된 정도를 정량적으로 측정한다. 실험을 통해 제안 기법이 진영의 양분화를 탐지하는데 효과적임을 보이고 동시에 온라인 커뮤니티에서 발생하는 개별 토론의 참여자들이 두 진영으로 양분되어 논쟁을 벌이는 것을 확인하였다.

☞ 주제어 : 인터넷 커뮤니티, 온라인 토론, 댓글 트리, 이분 그래프, 소셜 네트워크

ABSTRACT

In online communities, a large number of participants can exchange their opinion using replies without time and space restrictions. While the online space provides quick and free communication, it also easily triggers unnecessary quarrels and conflicts. The network established on the discussion participants is an important cue to analyze the confrontation and predict serious disputes. In this paper, we present a quantitative measure for polarity observed on the discussion network built from reply exchanges in online communities. The proposed method uses the comment exchange information to establish the user interaction network graph, computes its maximum spanning tree, and then performs vertex coloring to assign two colors to each node in order to divide the discussion participants into two subsets. Using the proportion of the comment exchanges across the partitioned user subsets, we compute the polarity measure, and quantify how discussion participants are bipolarized. Using experimental results, we demonstrate the effectiveness of our method for detecting polarization and show participants of a specific discussion subject tend to be divided into two camps when they debate.

☞ keyword : Internet Community, Online Discussion, Reply Tree, Bipartite Graph, Social Network

1. 연구동기

의사소통의 시공간적 제약이 없는 온라인상에서는 사

용자들이 다양한 주제에 대하여 상호간의 의견과 감정을 주고받는다. 온라인의 신속한 의사소통 기능은 다양한 사용자들이 동시에 자유로이 의견을 교환할 수 있도록 도움을 주기도 하지만, 때로는 공격적인 언사와 불필요한 언쟁을 통해 온라인상의 분쟁을 유발시키기도 한다.

사용자들 간의 의견이 대립 양상을 보이는 현상에 대한 특성을 파악하는 일은 앞으로 발생할 수 있는 분쟁을 예측하기 위한 중요한 단서가 된다. 특히 온라인 커뮤니티에서는 팔로잉 등을 통해 명시적으로 나타나는 소셜 네트워크 서비스와는 달리 사용자간의 관계가 댓글 교환을 통해 간접적으로 관찰되며, 이를 통해 형성된 네트워

¹ Department of Electrical and Computer Engineering, Pusan National University, Busan, 46241, Korea.

* Corresponding author (hgcho@pusan.ac.kr)

[Received 03 August 2018, Accepted 17 September 2018]

☆ 이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No. NRF-2017RID1A1A02018504)

☆ 본 논문은 2018년도 한국인터넷정보학회 춘계학술발표대회 우수 논문 추천에 따라 확장 및 수정된 논문임.

크에 온라인상의 대립 관계에 대한 정보가 담겨 있다. 그러나 이러한 정보로부터의 대립관계를 추출하고 이를 정량적으로 분석하는 연구는 전반적으로 미진한 편이다.

본 논문에서는 온라인 커뮤니티에서 일어나는 토론 과정에서 형성되는 사용자 상호작용 네트워크에서 사용자들이 어떻게 진영을 이루어 대립하는지를 정량적으로 분석하기 위한 극성 지표를 제안한다. 제안 지표를 이용하여 각각의 발제글에 대하여 토론 참가자들이 양분화되어 서로 대립하는 양상을 보임을 실험을 통해 보인다.

2. 관련 연구

기존에 온라인상의 의사소통과 관련된 연구는 주로 소셜 네트워크 서비스 상에서의 정보 전파[1]나, 군집화[2]를 중심으로 이루어져 왔다. 그 중 사용자 극성 분석과 관련된 연구는 주로 사용자들간의 관계가 팔로잉과 같이 명시적으로 드러나는 소셜 네트워크 서비스를 기반으로 이루어졌다. 대표적으로 트위터 상에서 의견 전파 모형을 활용하여 사용자 군집화를 수행하고 이를 통하여 군집이 양분화되는 것을 보인 연구가 있다[3]. 소셜 네트워크 서비스가 아니더라도 토론의 진영 정보를 시스템적으로 구현하여 토론 참가자로 하여금 자신의 입장을 명확하게 입력하여 우호 및 적대관계가 명시적으로 나타나도록 하는 시스템[4]이나, 온라인 뉴스 사이트에서 언어적 특징과 대략적인 구조 정보를 이용[5]하여 분석을 수행한 사례도 있다. 그러나 일반적인 온라인 커뮤니티에서는 사용자들이 의견교환을 하면서 토론이 이루어지고, 이에 따라 토론 참가자들의 진영이 자연스럽게 형성되어 대립과 갈등이 전개된다. 온라인 커뮤니티 상에서의 토론에 관한 분석 연구가 있기는 하지만, 주로 댓글의 구조적인 특성을 분석[6]하거나 논란을 일으키는 사용자를 탐지하거나 [7], 사용자의 활동을 촉발하는 원인을 분석[8]하는 것에 중점을 두고 있으며, 토론으로 인하여 발생하는 대립 구도에 관한 정량적인 분석은 잘 다루지 않고 있다.

3. 댓글 트리와 사용자 상호작용 그래프

인터넷 커뮤니티에서 나타나는 계층적 댓글 구조는 트리로 나타낼 수 있다. 하나의 발제글과 그에 따른 댓글들은 하나의 트리로 나타낼 수 있다. 이 때 루트는 발제글을 의미하며, 루트의 자식 노드들은 발제글에 직접 달린 댓글, 그 이하의 노드들은 다른 댓글에 달린 대댓글을 의미

한다. 그림 1(a)는 사용자 x가 작성한 발제글에 4명의 사용자 a,b,c,d가 댓글을 작성한 것을 나타내는 트리이다.

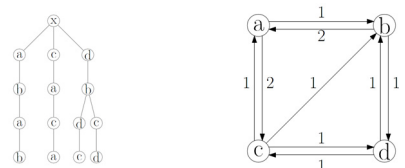
댓글 트리 상에서 누가 누구에게 댓글을 얼마나 작성하였는지에 대한 정보를 이용하여 댓글 트리로부터 사용자 상호작용을 나타내는 그래프를 얻을 수 있다. 노드 r을 루트로 가지는 트리 $T = (V_T, E_T)$ 가 있다고 하자. 노드 u의 작성자를 $\lambda(u)$ 로, u의 부모노드를 $p(u)$ 로 표기할 때, 사용자 상호작용 그래프 $G = (V_G, E_G)$ 는 방향 그래프(directed graph)로서 다음과 같이 구성된다.

$$V_G = \{\lambda(u) | u \in V_T - \{r\}\}$$

$$E_G = \{(\lambda(u), \lambda(v)) | (u, v) \in \hat{E}_T, v = p(u), v \neq r\}$$

$$\text{단, } \hat{E}_T = \{(u, v) \in E_T | v = p(u), v \neq r\}.$$

여기서 루트노드와의 에지를 계산에서 제외하는 이유는 발제글과 그에 직접 달린 댓글의 관계와, 댓글과 그에 대한 대댓글간의 관계는 다소 차이가 있음을 반영하기 위함이다. 이러한 차이는 기본적으로 분량과 작성시간의 차이로부터 기인한다. 댓글은 발제글에 비하여 작성하는데 소요되는 시간이 짧으며, 댓글이 담고 있는 분량은 발제글의 내용에 비하여 상대적으로 매우 적기 때문에, 그러한 댓글에 대하여 공감 또는 반박을 하기 위한 대댓글이 담을 수 있는 내용도 한정될 수밖에 없다. 반면 발제글은 충분한 시간과 분량을 통하여 작성자가 다양한 측면에서의 의견을 개진할 수 있기 때문에 발제글 내용 상의 다양한 세부부분에 공감, 반박을 포함하여 내용 보충, 비슷한 사례 소개, 참고자료 링크 등 다양한 형태의 댓글이 발제글에 직접 달릴 수 있게 된다. 이러한 특성은 기존의 연구에서도 발제글에 직접적으로 달린 댓글의 정량적 특성이 그 이하에 달린 나머지 댓글들에 비하여 뚜렷한 차이가 나는 것을 확인할 수 있다[6]. 따라서 비슷한 특성을 공유하고 있는 관계만을 고려하기 위하여 댓글과 댓글 사이의 연결 관계, 즉 루트노드가 아닌 노드 간의 에지만을 이용하여 사용자 그래프를 구성하였다.



(a) 댓글트리 (a) Reply tree (b) 상호작용그래프 (b) Interaction graph

(그림1) 댓글 트리와 사용자 상호작용 그래프 (Figure 1) A reply tree and its user interaction graph

그래프 G 의 각 에지 $e = (x, y)$ 는 사용자들 간에 댓글을 서로 얼마나 많이 교환했는지를 의미하는 가중치 $w((x, y))$ 를 가지고 있다.

$$w((x, y)) = |\{(u, v) \in \widehat{E}_T | \lambda(u) = x, \lambda(v) = y\}|$$

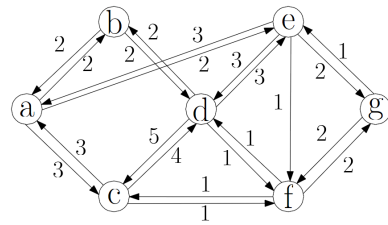
그림 1(b)는 그림 1(a)의 댓글 트리로부터 구한 사용자 상호작용 그래프이다. 이 그래프를 통하여 각 사용자들이 누구에게 댓글을 몇 회 작성하였는지를 확인할 수 있는데, 상호간에 댓글을 작성한 횟수가 많아질수록 두 사용자는 대립관계에 있을 확률이 높다. 왜냐하면 자신의 댓글에 공감하는 댓글이 작성된 경우보다는 반박 댓글이나 감정적인 댓글이 달린 경우 그에 대한 대응을 하기 위해 다시 그 댓글에 대한 대댓글을 작성하게 되고 이러한 과정이 반복되면서 양 사용자간에 많은 양의 댓글이 교환되기 때문이다. 따라서 본 논문에서는 댓글의 상호 교환 횟수가 해당 사용자 쌍의 대립의 정도를 나타낸다는 가정에 기초하여 사용자들이 각각의 발제글에서 대립되는 양상이 양분화 형태를 보일 것이라는 가설을 세우고 이를 보이기 위한 정량적 지표를 제안하고자 한다.

4. 이분그래프 기반의 극성 지표

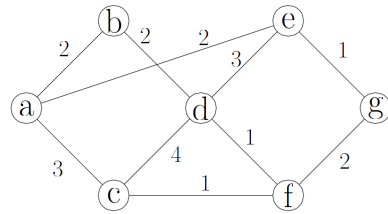
한 발제글 내에서 논쟁을 벌이는 사용자들이 양분화되어 두 진영으로 나뉘고, 댓글 교환이 서로 다른 진영에 속한 사용자 간에만 이루어진다면 상호작용 그래프는 이분 그래프(bipartite graph)가 될 것이며, 2개의 색상으로 노드를 칠할 수 있게 된다. 만약 진영의 양분화가 전혀 일어나지 않고 사용자들 간에 무작위로 논쟁이 발생한다면 상호작용 그래프는 이분그래프와는 먼 형태의 특성을 띠게 될 것이다. 결국 양분화가 얼마나 일어났는지를 보이기 위해서는 주어진 사용자 상호작용 그래프가 이분그래프에 가까운 정도를 정량적으로 나타내는 것이 필요하다.

간단하게 생각할 수 있는 방법으로는 상호작용 그래프의 부분그래프 중 가장 큰 이분그래프인 것을 찾는 것이다. 이 이분 그래프가 상호작용 그래프와 가까울수록 양분화 현상이 뚜렷하게 일어났다고 할 수 있다. 예를 들어 상호작용 그래프에서 이분그래프로 만들기 위해 제거해야 하는 노드의 최소 개수 또는 제거해야 하는 에지들을 최소화하는 문제를 생각할 수 있다. 그러나 이러한 문제들은 주로 NP-Hard문제이므로[9,10] 직접적으로 적용하기에는 어려움이 따른다.

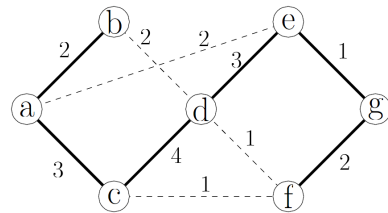
본 논문에서는 이를 위해 그림 2와 같은 과정을 통한 휴리스틱을 사용한다. 그림 2(a)와 같은 그래프가 주어졌을 때, 각 노드간을 연결하는 에지들 중 가중치가 작은 값을 선택하여 무방향화 한다. 즉, 주어진 상호작용 그래프에서 두 노드 x 와 y 를 연결하는 에지의 가중치는 에지의 방향에 따라 $w((x, y))$ 와 $w((y, x))$ 인데, 에지의 방향을 없애면서 두 값 중 작은 값을 새로운 에지의 가중치로 하



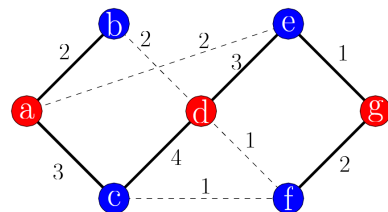
(a) Interaction graph



(b) Making undirected and filtering edges



(c) Maximum spanning tree



(d) 2-Coloring

(그림 2) 상호작용그래프로부터 진영을 계산하는 과정 (Figure 2) Computing Camps from a Graph.

고 만약 가중치가 0인 경우 에지를 제거한다. 무방향화 결과는 그림 2(b)에 나타나있다. 이 작업을 수행하는 이유는 두 사용자 중 한 명이 일반적으로 댓글을 다는 경우에는 이를 쌍방 간의 논쟁이나 대립이라고 볼 수 없기 때문에 양쪽 모두 댓글을 서로 교환한 횟수만 고려하기 위함이다. 이렇게 만들어진 그래프에서 최대신장트리를 계산한 후 최대신장트리 상에서 2-컬러링을 수행한다. 트리를 2-컬러링하는 것은 단순히 깊이우선탐색을 하면서 인접한 노드 간에 교차로 색상을 부여하면 되므로 선형시간에 수행할 수 있다. 그림 2(c)의 굵은 선이 그림 2(b)의 그래프에 대한 최대신장트리이고, 이 트리 상에서 2-컬러링을 수행한 결과가 그림 2(d)에 나타나있다. 2-컬러링을 위해 최대신장트리를 이용하는 이유는 높은 가중치의 에지를 사이에 두고 있는 노드일수록 서로 다른 색상이 부여될 수 있도록 하기 위함이다.

노드에 대한 2-컬러링이 수행되고 난 후, 각 노드의 색상은 그 노드에 해당하는 사용자가 속한 진영이라고 볼 수 있다. 만약 진영이 양분화 될수록 서로 다른 색상의 노드 간에 에지 가중치가 같은 색상의 노드 간 에지 가중치에 비하여 뚜렷하게 높을 것이며, 반대로 양분화 현상이 나타나지 않을수록 에지 가중치와 그 에지 양단의 노드 색상과는 관계가 없어질 것이다. 그래프의 노드 집합 V_G 를 노드의 색상에 따라 부분집합 V_G^+ 와 V_G^- 로 분할한 후, V_G^+ 와 V_G^- 간을 연결하는 에지들의 집합을 E_{inter} 로 정의한다. 또한 서로 다른 색상의 노드들을 잇는 에지 가중치 합을 W_{inter} , 전체 에지 가중치 합을 W_G 라 한다.

$$W_{\text{inter}} = \sum_{e \in E_{\text{inter}}} w(e), \quad W_G = \sum_{e \in E_G} w(e)$$

토론 참가자들이 얼마나 양분화되었는지를 나타내는 극성지표는 전체 에지 가중치 합 중 서로 다른 색상 간의 에지 가중치 합의 비율로 다음과 같이 정의한다.

$$\varphi(G) = W_{\text{inter}} / W_G$$

다시 말해, 어떤 사용자 상호작용 그래프 G 의 극성지표 $\varphi(G)$ 는 해당 그래프 상에서 사용자들이 두 집단으로 나뉘어 서로 다른 집단에 대해서만 댓글을 작성하는 경우 1이 되며, 같은 집단 간의 댓글의 비율이 많아질수록 그 값이 점차 감소하게 된다.

5. 그래프 형태에 따른 극성지표 보정

이전 장에서 제안한 극성 지표를 그대로 양분화 현상을 정량화하여 표현하는데 사용할 수 있지만, 서로 다른 크기와 형태의 그래프의 양분화 정도를 비교하기 위한 용도로는 다소 한계점이 있다. 그래프가 가진 노드의 수나 연결의 형태에 따라 극성지표의 값의 범위가 다소 차이가 발생할 수 있기 때문이다. 예를 들어 노드가 2개인 연결그래프(connected graph)인 경우에는 항상 $\varphi(G) = 1$ 이 된다. 그렇다고 해서 노드 2개인 이 그래프가 노드 수가 100개이면서 극성지표 값이 0.998인 그래프에 비하여 양분화가 더 뚜렷하게 이루어졌다고는 하기 어렵다. 또한 같은 수의 노드를 가진 그래프라고 하더라도 100개의 노드가 하나의 컴포넌트로 연결된 그래프와 노드가 2개씩 연결되어 컴포넌트가 50개인 그래프는 비교 기준이 달라야 할 것이다. 따라서 그래프 크기와 형태의 차이로부터 기인하는 극성지표값 범위를 보정하는 것이 필요하다.

우선 주어진 상호작용 그래프가 연결그래프라고 가정하면, 극성지표값을 보정하기 위한 기준으로서 같은 수의 사용자들이 균등하게 댓글을 주고받았을 때, 즉 완전그래프(complete graph)의 형태를 가지는 경우에 비하여 얼마나 양분화가 되었는지를 이용할 수 있다. 다시 말해 주어진 상호작용 그래프의 극성지표의 보정값은 같은 크기의 완전그래프의 극성지표에 대한 비율로 나타낼 수 있다.

$$\varphi^*(G) = \varphi(G) / \varphi(K_{|G|})$$

$$\text{단, } \varphi(K_n) = \frac{(n/2)^2}{n(n-1)/2} = \frac{n^2}{2n(n-1)}.$$

완전그래프의 노드 수가 홀수인 경우 노드가 정확히 반으로 나누어지지 않으므로 $\varphi(K_n)$ 의 분자 부분이 $k = \lfloor n/2 \rfloor$ 일 때 $k(k+1)$ 이 되어야 하므로 위의 식과는 다소 차이가 있으나 노드 수가 충분히 많다면 큰 차이는 없으므로 노드 수의 홀수 짝수 여부와 관계없이 위의 식을 그대로 이용하도록 한다.

주어진 상호작용 그래프가 연결그래프가 아니라 크기가 $\{c_i\}$ 인 여러 개의 작은 컴포넌트(component)로 이루어져 있다면 각각의 컴포넌트들이 완전그래프인 그래프의 극성지표를 분모로 하여 보정하도록 한다.

$$\varphi\left(\bigcup_i K_{c_i}\right) = \left(\sum_i c_i^2\right) / \left(2 \sum_i c_i(c_i - 1)\right)$$

6. 실험 및 결과

6.1 실험 데이터

실험을 위하여 인터넷 커뮤니티 PGR21의 자유게시판에서 2015년부터 2017년까지 3년간 작성된 발제글 17,517개 및 발제글에 달린 댓글 총 975,512개를 수집하였다. 이 사이트를 선택한 이유는 우선 계층적 구조의 댓글 시스템을 지원하며, 발제글 및 댓글 작성에 있어서 타 커뮤니티에 비하여 다소 엄격한 규정이 있어 사용자 간의 토론을 시스템적으로 유도하고 있다는 점에서 본 실험에 적합하다고 판단하였기 때문이다. 수집한 데이터의 기본적인 특성은 표 1과 같다.

(표 1) 실험 데이터 특성

(Table 1) Data characteristics

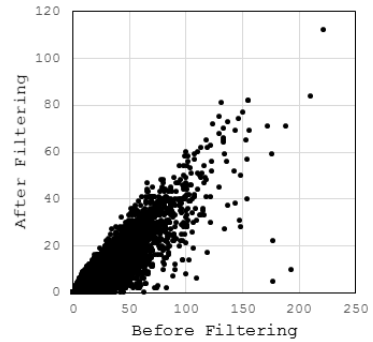
항목	평균	최대
발제글 당 댓글 수	55.69	1362
발제글 당 참여자	35.23	1350
댓글 최대 깊이	4.77	131
Depth-1 댓글 수	29.33	1354

6.2 사용자 상호작용 그래프 특성

각각의 발제글에 대하여 사용자 상호작용 그래프를 구성하였다. 에지 필터링 과정을 거친 후 고립된 노드를 모두 제거했을 때 5,260개의 발제글에 대한 그래프가 빈 그래프(empty graph)가 되는 것을 확인하였다. 이들 발제글에 달린 댓글 수의 평균은 약 19.68개이며, 나머지 12,256개 발제글의 평균인 71.14개에 비하여 매우 적은 양으로 모든 노드가 제거되는 발제글에서는 사용자들 상호간에 충분한 의견교환을 하지 않았음을 확인할 수 있다.

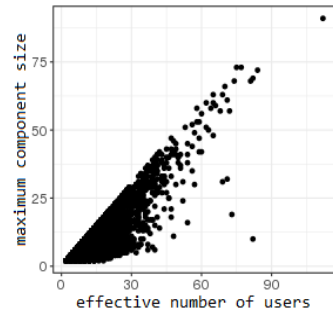
나머지 발제글에서는 필터링 이후에 평균 7.60개의 노드가 남는 것으로 관찰되었다. 즉, 이들 발제글에서 상호간에 댓글을 주고받은 “유효사용자”의 수가 평균 7.60명임을 의미한다. 유효사용자수의 최대값은 112이었다. 그림 3은 필터링 전후의 유효사용자수 변화를 나타낸다.

필터링 이후에 전체 그래프가 연결그래프가 되는 발제글은 5,784개가 있었으며, 가장 많은 경우에는 24개의 컴포넌트가 발생하기도 하였다. 한 발제글의 사용자 상호작용 그래프에서 연결된 컴포넌트의 개수는 평균적으로 2.12개였다. 각 발제글에서 가장 큰 연결된 컴포넌트의 크기는 평균 5.00이었고, 최대 91개의 노드가 서로 연결되는 경우도 있었다.

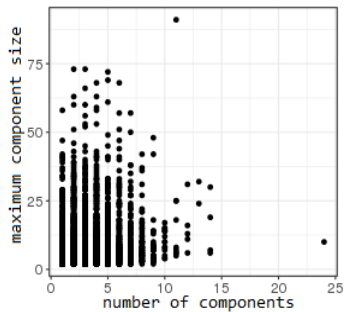


(그림 3) 에지 필터링 전후의 유효사용자 수.

(Figure 3) Number of effective users before and after edge filtering.



(a)

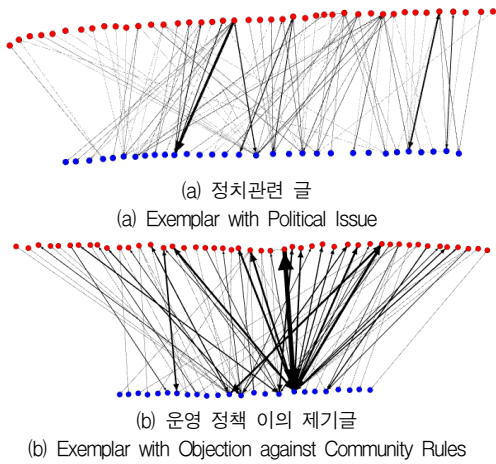


(b)

(그림 4) 사용자 상호작용 그래프의 컴포넌트 특성.

(Figure 4) Component characteristics of User Interaction Graphs.

상호작용 그래프 상에서 컴포넌트가 많을수록 각각 별개의 독립된 토론이 다수 발생하고 있다는 것을 의미하며, 큰 컴포넌트가 있을수록 다수의 사용자가 같은 주제에 대하여 의견을 활발하게 교환하고 있다는 것이라 해석할 수 있다. 그림 4(a)는 유효사용자의 수에 따른 최대



(그림 5) 컬러링 결과로부터 관찰되는 진영 양분화
(Figure 5) Bipolarization observed from coloring

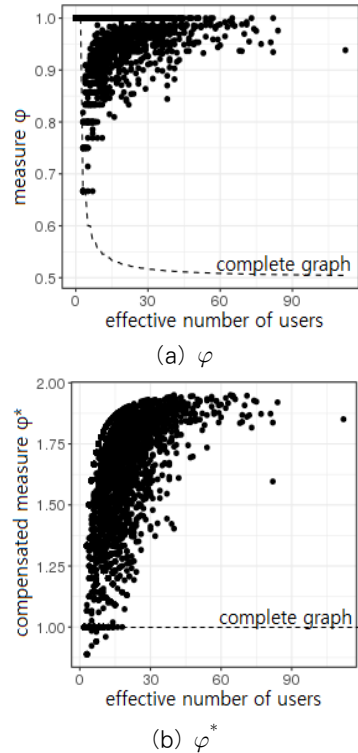
컴포넌트의 크기를 나타낸다. 토론에 참여하는 사용자의 수가 증가할수록 최대 컴포넌트의 크기가 증가하고, 이는 사용자가 모일수록 동일한 주제를 두고 의견을 교환하는 경향이 강해진다는 것을 의미한다고 할 수 있다. 한편 그림 4(b)는 컴포넌트의 개수와 최대 컴포넌트의 크기를 나타낸 것으로 두 특성 사이에는 큰 상관관계가 없다는 것을 확인할 수 있다.

6.3 개별 발제글 그래프 컬러링 결과

각각의 발제글로부터 계산된 상호작용 그래프 상에서 최대신장트리를 이용하여 2-컬러링을 수행한 결과 중 일부 예시가 그림 4에 나타나 있다. 그림 5(a)는 정치관련 발제글에서 해당 발제글의 작성자가 던진 주제에 대해 찬성하는 사용자와 반대하는 사용자 간에 첨예한 대립이 있는 것을 시각적으로 나타내준다. 그림 5(b)는 사이트의 운영정책과 관련한 이의제기글에 다수의 사용자가 소수의 운영자편과 대립을 하고 있는 관계를 보여준다.

6.4 극성 지표 분포 및 특성

각 발제글의 상호작용 그래프로부터 계산할 수 있는 극성지표와 5장에서 보았던 극성지표의 분포가 그림 6(a)와 그림 6(b)에 각각 나타나 있다. 전반적으로 유효사용자의 수가 증가할수록 두 극성지표의 값이 증가하는 경향이 있다는 것을 확인할 수 있다. 그림 6(a)에서 점선은 노드 수(유효사용자 수)에 따른 완전그래프의 극성지표 변



(그림 6) 유효사용자 수에 따른 극성 지표 분포 (점선은 완전그래프의 극성지표)

(Figure 6) Polarity measure values w.r.t the number of effective users (Dashed lines are polarity value for complete graphs)

화를 나타낸 것으로 노드 수가 증가할수록 지표의 값은 0.5에 급격하게 수렴하는 것을 확인할 수 있다. 이는 사용자 간의 댓글 교환이 무작위로 일어나는 사건이 아니라 사용자들이 모일수록 두 진영으로 양분화되는 경향이 강해지고, 서로 다른 진영의 논리를 반박하기 위한 댓글을 많이 작성하게 된다는 것을 알 수 있다.

그림 6(a)에서 확인할 수 있듯이 유효사용자의 수가 적은 발제글의 상호작용 그래프에서 극성지표값이 1.0인 경우가 많이 발생하는데, 이는 발제글 내에서의 토론이 하나로 뭉쳐질 만큼의 충분한 댓글 교환이 이루어지지 않는 상황에서 유효사용자(노드)들이 크기가 2정도인 작은 컴포넌트들로 이루어져 자연스럽게 극성지표값이 최대치인 1.0으로 계산되기 때문이다. 그러나 5장에서 살펴봤듯이 이러한 경우에는 극성지표값이 높더라도 사용자들이 뚜렷하게 양분화되었다고 보기에는 어려움이 있다.

그림 6(b)에서는 보완된 극성지표값을 통하여 이러한 문제점을 해결한 것을 확인할 수 있다. 보완된 극성지표값에서는 완전그래프의 경우 유효사용자의 수와 관계없이 1.0의 값을 지니게 되며, 완전그래프에 비하여 양극화가 뚜렷하게 일어날수록 더욱 높은 값을 가지게 된다.

유효사용자 수가 적은 경우 상호작용 그래프의 극성지표가 완전그래프의 극성지표보다 낮은 경우가 종종 발생하는데 (그림 6(b)의 좌측하단), 이는 최대신장트리를 이용한 2-컬러링 방법이 주어진 그래프에서 최대의 양분화 분할을 정확하게 찾지는 못하기 때문인 것으로 파악된다. 그러나 이 경우는 주로 완전그래프에서 적은 수의 에지가 결손되는 경우 컬러링이 다소 어긋나면서 발생하는 경향이 있으므로 양분화가 뚜렷해지는 상황에서는 크게 문제가 되지 않을 것으로 판단된다.

7. 결론 및 추후연구

본 논문에서는 온라인 커뮤니티에서 발생하는 토론의 대립 양상을 표현하는 정량적 지표를 제안하였다. 주어진 발제글에서 일어나는 토론은 참가자들로 구성된 네트워크를 형성하고, 이 네트워크에 많은 참가자들이 속할수록 이들이 두 진영으로 양분화되어 상대방 진영의 논리에 반박하기 위한 댓글을 작성하는 경향을 확인하였다.

추후 연구로는 우선 본 논문의 기본적인 가정에 대한 실험적인 입증을 수행하고자 한다. 다시말해 실제로 댓글과 대댓글의 관계가 반복될수록 해당 사용자들이 대립관계에 있다는 사실을 댓글 내용을 통하여 직접 분석을 수행함으로써 댓글 교환 횟수에 따라 해당 대화를 나누는 사용자간의 관계가 대립관계일 확률이 어떠한 추세를 나타내는지 밝히고자 한다.

또한 본 논문에서 제시한 네트워크 구성 방법과 정량지표를 이용하여 논쟁을 주로 일으키는 주제나, 논쟁을 유발하는 사용자와 같이 실질적으로 온라인 논쟁을 예측하고 해결하기 위한 적용 방안에 대한 연구를 수행할 예정이다.

본 논문에서는 한 발제글 내에서의 토론 양상에 대하여만 다루었으나 이후 연구에서는 여러 개의 발제글에 걸친 토론에서 참가자들의 대립 양상이 어떻게 변하는지와 같은 시간적 특성을 분석함으로써 온라인 토론을 동적인 측면에서의 모형화할 수 있는 다양한 기법들을 다루고자 한다.

참고문헌(Reference)

- [1] N. Zhao, and X. Liu, "Information Propagation in Social Networks with Overlapping Community Structure," *KSIIT Transactions on Internet and Information Systems*, Vol 11, No 12, pp. 5927-5942, 2017.
[http:// dx.doi.org/10.3837/tiis.2017.12.013](http://dx.doi.org/10.3837/tiis.2017.12.013)
- [2] C.-G. Han, and M.-H. Jo, "An Enhanced Community Detection Algorithm Using Modularity in Large Networks," *Journal of Internet Computing and Services*, Vol 13, No 3, pp.75-82, 2012.
<http:// dx.doi.org/10.7472/jksii.2012.13.3.75>
- [3] A. J. Morales, J. Borondo, J. C. Losada, and R. M. Benito, "Measuring Political Polarization: Twitter Shows the Two Sides of Venezuela," *Chaos*, Vol 25, No 033114, 2015. <http://dx.doi.org/10.1063/1.4913758>
- [4] S. Rosenthal, and K. McKeown, "I Couldn't Agree More: The Role of Conversational Structure in Agreement and Disagreement Detection in Online Discussions," in *Proc. SIGDIAL*, pp. 168-177, 2015.
<http://dx.doi.org/10.18653/v1/W15-4625>
- [5] K. Beelen, E. Kanoulas, and B. van de Velde, "Detecting Controversies in Online News Media," in *Proc. ACM SIGIR*, pp. 1069-1072, 2017. <http://dx.doi.org/10.1145/3077136.3080723>
- [6] V. Gomez, V. Kaltenbrunner and V. Lopez, "Statistical Analysis of the Social Network and Discussion Threads in Slashdot," in *Proc. WWW*, pp.645-654, 2008.
<http://dx.doi.org/10.1145/1367497.1367585>
- [7] T. C. Li, J. Gharibshah, E. E. Papalexakis, and M. Faloutsos, "Trollspot: Detecting Misbehavior in Commenting Platforms," in *Proc. ASONAM*, pp.171-175, 2017.
<http://dx.doi.org/10.1145/3110025.3110057>
- [8] M. Lee, I. Choi, and S. Yang, "When Do People Post a Comment to a News Story on the Internet?" *KSIIT Transactions on Internet and Information Systems*, Vol 6, No 1, pp.434-445, 2015.
<http://dx.doi.org/10.3837/tiis.2015.01.027>
- [9] M. Yannakakis, "Node- and Edge-deletion NP-complete problems," in *Proc. STOC*, pp.253-264, 1978.
<http:// dx.doi.org/10.1145/800133.804355>

- [10] R. Peeters, "The Maximum Edge Biclique Problem is NP-complete," *Discrete Applied Mathematics*, Vol 131, No 3, pp.651-654, 2003.
[http://dx.doi.org/10.1016/S0166-218X\(03\)00333-0](http://dx.doi.org/10.1016/S0166-218X(03)00333-0)

● 저 자 소 개 ●



김 성 환(Sung-Hwan Kim)

2011년 부산대학교 정보컴퓨터공학과(공학사)
2013년 부산대학교 대학원 컴퓨터공학과(공학석사)
2018년 부산대학교 대학원 전기전자컴퓨터공학과(공학박사)
2018년~현재 부산대학교 IT기반융합산업창의인력양성사업단 연수연구원
관심분야 : 데이터마이닝, 알고리즘
E-mail : sunghwan@pusan.ac.kr



탁 해 성(Haesung Tak)

2013년 부산대학교 정보컴퓨터공학부(공학사)
2015년 부산대학교 전기전자컴퓨터공학과(공학석사)
2015년~현재 부산대학교 전기전자컴퓨터공학과 박사과정
관심분야 : 데이터마이닝, 데이터시각화
E-mail : tok33@pusan.ac.kr



조 환 규(Hwan-Gue Cho)

1984년 서울대학교 계산통계학과(이학사)
1986년 KAIST 대학원 전산학과(공학석사)
1990년 KAIST 대학원 전산학과(공학박사)
1990년~현재 부산대학 전기컴퓨터공학부 교수
관심분야 : 알고리즘, 그래프이론, 생물정보학
E-mail : hgcho@pusan.ac.kr