

# 트위터에서 이슈가 되고있는 뉴스 기사에 대한 소셜 사용자 네트워크 기반 정치 성향 분류

김준길, 이경순  
전북대학교 컴퓨터공학부/영상정보신기술연구센터  
e-mail : jgkim@jbnu.ac.kr, selfsolee@jbnu.ac.kr

## Political Bias Classification Based on Social User Networks on Issuable Political News Article in Twitter

Jun-Gil Kim, Kyung-Soon Lee  
Dept of Computer Engineering, Chonbuk National University

### 요 약

트위터에서 정치 성향을 가지거나 관심이 있는 트위터 사용자는 관심있는 정치 인물이나 단체에 대한 뉴스 기사에 대해 자신의 의견을 남기거나 그대로 인용하게 된다. 또한, 자신의 의견과 공감하거나 비공감하는 트윗에 대해서 리트윗을 하거나 추가적인 자신의 의견을 언급하기도 한다. 본 논문에서는 이슈가 되고있는 정치 뉴스 기사에 대해 관심 있는 트위터 사용자들을 찾아 트위터 사용자들 간의 트윗 문서들 사이에서의 관계 정보를 가지는 사용자 네트워크에서의 트위터 사용자들의 성향을 분류해주는 방법을 제안한다. 제안한 방법의 유효성을 검증하기 위해 트위터에서 이슈가 된 정치 뉴스 기사들과 각 뉴스 이슈를 언급한 트위터 데이터에서 트윗 문서 내용 유사도 기반 분류 방법과의 비교 실험 하였다. 실험 결과에서 사용자간의 관계 정보를 이용한 성향 분류 방법이 유효함을 보였다.

### 1. 서론

트위터(Twitter)[1]에서 정치적 성향을 가지거나 관심이 있는 트위터 사용자는 자신의 트위터에 정치 인물이나 단체에 대한 언급을 하거나 관련된 뉴스 기사의 정보를 인용하거나 추가적으로 자신의 의견을 포함하는 트윗(tweet)을 작성한다. 또한, 다른 사용자들의 관심있는 트윗을 그대로 리트윗(Retweet)하거나 자신의 의견을 추가하기도 한다.

정치적으로 이슈가 되고 있는 뉴스 기사에서는 한 쪽으로 치우치거나 다양한 뉴스에 대한 의견들이 나타난다. 이러한 의견들은 진보, 보수 혹은 찬성, 반대와 같이 양측으로 의견을 분류하여 보기가 어렵다. 게다가 트위터에서는 140 자로 제한됨에 따라 트위터 사용자의 성향에 따른 트윗에 대한 내용이 어느 쪽으로 치우쳤는지 판단하기가 어렵다. 더군다나 리트윗을 하면서 자신의 의견을 추가하는 트윗은 더욱 더 판단하기가 어렵다. 따라서, 트위터에 기반하는 뉴스 기사에 대한 트위터 사용자의 성향에 대한 그룹을 자동으로 분류해주는 시스템이 필요하다.

트위터 서비스에 기반한 연구에서 [2]과 [3]는 리트윗 정보가 리트윗한 대상에 대한 내용에 공감에 대한 표현이거나 다른 사람들의 생각을 검증하기 위하는 행동이라고 분석하였다. [4]에서 트위터에서는 정보를 전달하는데 있어 많은 트윗 문서가 나타난다고 분석하였으며, 이슈자질을 포함하는 트윗의 내용 중 85%가 해당하는 뉴스 기사에서의 글이었다고 하였다. [5]

에서는 트위터 데이터에서 시간적인 분석에 감성 자질 및 리트윗을 이용하여 사회적인 핵심 사건을 추출하는 방법을 제안하였다.

뉴스 기사와 댓글 사이에서의 정보를 이용하여 뉴스 기사 혹은 댓글자의 정치적 성향을 분류하는 연구가 활발하게 이루어지고 있다. [6]는 뉴스 기사에 나타난 댓글자들의 정치적 성향에 따른 감정패턴을 이용하여 뉴스 기사의 성향을 분류해주는 방법을 제안하였다. [7]는 논쟁을 초래할 수 있는 이슈에 대한 뉴스 기사들의 성향 분류를 위한 논쟁자 관계 기반 방법을 제안하였다.

본 논문에서는 트위터 서비스에서 이슈가 되고 있는 뉴스 기사에서 트위터 사용자들의 정치적 성향에 대한 그룹을 분류해주는 방법을 제안한다. 트위터에서 뉴스 URL 을 포함하는 트윗과 리트윗에서 많이 인용된 뉴스 URL 을 선택하여 이슈가 되고 있는 뉴스 기사의 데이터를 추출하고 선택된 뉴스 URL 이나 뉴스 제목과 같은 뉴스 정보를 포함하고 있는 트위터 데이터를 수집한다. 트위터 사용자의 성향을 분류하기 위하여 뉴스 이슈를 언급한 트윗 문서들 사이의 관계 정보를 분석하여 사용자 소셜 네트워크를 생성하고 트위터 사용자 사이의 각 사용자의 트위터 히스토리(twitter history)에서 나타난 트윗 문서들 사이의 관계 정보를 이용하여 트위터 사용자의 성향을 분류하는 방법을 제안한다. 실험을 통하여 제안 방법의 유효성을 검증한다.

## 2. 트위터에서의 소셜 사용자 네트워크 구축

본 논문에서는 트위터 사용자 간에 나타난 트윗 문서 정보와 관계를 이용하여 소셜 사용자 네트워크를 구축하였다. 2.1에서는 먼저 트윗 문서에서의 사용자 관계 정보를 추출하기 위한 방법을 소개하였으며 2.2에서는 이러한 네트워크에서 사용자의 성향을 분류하는 방법에 대해 설명한다.

### 2.1 뉴스기사에 대한 트윗과 리트윗에 기반한 사용자 네트워크 구축

트위터 사용자들 간의 관계에 대한 네트워크를 구성하기 위하여 먼저 트윗 문서에서의 특성에 대해 분석하도록 한다.

뉴스 기사에 대한 트윗 문서는 기사에서의 이슈에 대한 자신의 의견을 표현하는 트윗 문서와 다른 사용자의 트윗에 동의하거나 다른 사용자에게 알리고 싶을 경우에 상대 트윗을 그대로 가져오게 되는 리트윗 문서, 그리고 리트윗 문서에 긍정적인 혹은 부정적인 자신의 의견을 추가하는 의견 리트윗 문서로 구분할 수 있다.

본 논문에서는 의견을 포함하는 트윗 문서를 가지는 것에 따라 사용자를 구분하였다. 먼저, 사용자가 뉴스 기사에 대한 트윗 문서를 가지고 있을 경우 그 사용자는 트윗 사용자로 구분을 하였다. 리트윗에 자신의 의견을 추가한 경우 의견 리트윗 사용자로 구분을 하였고 그 이외에 의견이 포함되지 않은 리트윗 문서만 가지는 사용자를 리트윗 사용자로 구분하였다.

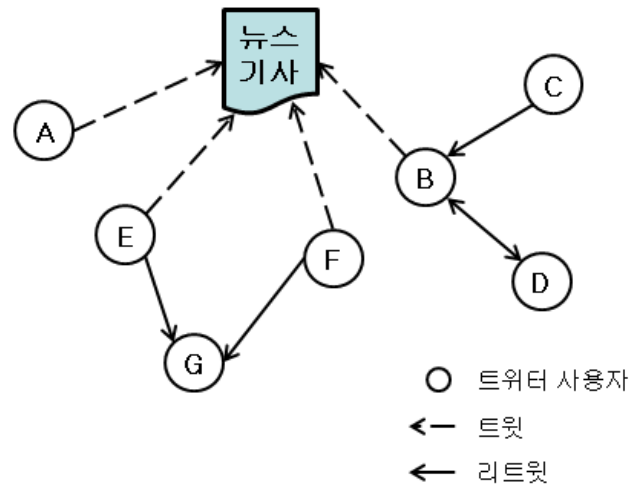
뉴스 기사와 트위터 사용자간의 네트워크를 구축하기 위해 각각의 사용자들을 노드(Node)로 두고, 사용자들이 가지는 트윗과 리트윗 문서에서의 정보를 토대로 에지(Edge)를 생성해 줄 수 있다.

### 2.2 사용자 네트워크에서의 사용자 관계 관찰

트위터 사용자간의 관계를 토대로 뉴스 기사와 트위터 사용자간의 네트워크를 그림(1)과 같이 생성할 수 있게 된다. 트윗 문서에 따른 뉴스 기사와 트위터 사용자 간의 관계는 다음과 같이 나타난다.

사용자 간의 네트워크에서의 사용자간의 관계 정보인 에지는 크게 리트윗 관계, 의견 리트윗 관계, 상호 리트윗 관계, 공동 리트윗 관계로 4 가지로 나누어 볼 수 있다.

- 리트윗 관계란 사용자 C가 사용자 B의 트윗 문서를 그대로 리트윗 하였을 때 C가 B의 의견에 동의한다고 볼 수 있으므로 두 사용자의 성향이 동일하다고 가정할 수 있다.
- 의견 리트윗 관계란 사용자 B가 사용자 C의 트윗 문서를 의견을 추가하면서 리트윗 한 경우에 추가한 의견에 따라 B와 C의 성향이 각각 달라지게 된다. 두 트위터 사용자에 대해 의견이 같고 다를 수 있음을 확인 할 필요가 있다.
- 상호 리트윗 관계란 사용자 B와 사용자 D 같이 서로 리트윗 관계에 있을 경우 서로의 의견에 동의한다고 볼 수 있어 두 사용자는 매우 밀접한 관



그림(1) 뉴스 기사와 관련된 트윗 문서에 대한 사용자 소셜 네트워크

계를 가지게 된다.

- 공동 리트윗 관계란 사용자 E와 사용자 F가 동일한 사용자 G와 리트윗 관계에 있을 경우 E와 F는 간접적으로 유사한 관계가 있음을 알 수 있다.

사용자 간의 리트윗 관계 정보는 트윗 문서 어휘에 대한 분석 없이도 두 사용자의 관계를 알 수 있는 매우 중요한 정보라고 할 수 있다.

## 3. 트위터 사용자 정치 성향 분류 방법

소셜 사용자 네트워크 사이에서의 관계를 이용하여 사용자의 성향을 분석하기 위해 날짜별로 이슈가 되고있는 정치 뉴스 기사를 선택하고 해당 기사에 관심을 보이는 트위터 사용자들의 트윗을 수집하고 각각의 사용자들의 히스토리를 수집하였다. 수집한 데이터를 이용하여 사용자 네트워크에 기반한 사용자 정치 성향을 분류하는 방법을 소개한다.

### 3.1 이슈가 되고있는 정치 뉴스 기사 선택

본 논문에서는 이슈가 되고 있는 정치적 뉴스 기사를 선택하기 위해서 매일 트위터에 뉴스 URL이 많이 나타나는 뉴스 기사를 이슈가 되는 뉴스 기사로 보고 해당되는 뉴스 기사를 선택하였다. 선택한 뉴스 URL을 가지는 트위터 데이터를 수집하기 위해 뉴스 URL이나 뉴스 제목이 포함된 트위터 데이터를 수집하였다.

### 3.2 트위터 히스토리 정보를 이용한 사용자 네트워크 관계 확장

뉴스 기사에서의 사용자 관계만으로는 모든 사용자의 성향을 분류하기가 어렵다. 본 논문에서는 사용자간의 관계 정보를 확장하기 위하여 사용자 각각의 히스토리(history)에서의 리트윗 정보와 리트윗 대상을 이용하여 두 사용자간의 관계를 확장하였다. 여기 트위터 사용자 히스토리란, 트위터 사용자가 이전에 작

<표 1> 뉴스 이슈와 트위터 히스토리에서의 사용자 관계별 네트워크 확장 분포

이슈	트위터 사용자 수	뉴스 기사에서의 관계 분포				트위터 히스토리에서 확장된 관계 분포	
		리트윗 관계	의견 리트윗 관계	상호 리트윗 관계	공동 리트윗 관계	상호 리트윗 관계	공동 리트윗 관계
1	428	440	7	2	26,631	312	87,678
2	1,150	1,179	45	2	37,254	1,350	629,776
3	303	1,016	2	12	18,884	142	15,968
4	573	2,243	24	32	93,469	689	66,106
5	752	675	88	0	9,871	976	268,769

성하였던 모든 트윗과 리트윗 데이터를 말한다. 확장된 관계 정보는 <표 1>과 같다.

추출된 트위터 데이터에서의 사용자들의 히스토리에서의 최근 트위터 문서를 각각 1,000 개씩 수집하였다. 정치와 관련된 인물과 어휘에 대한 간단한 리스트를 생성하여 리스트에 있는 단어가 한 개 이상 나타나는 트위터 문서들을 정치 관련 트위터 문서들로 하였다.

### 3.3 트위터 사용자 네트워크에 기반한 정치 성향 분류 방법

2 장에서 분석한 트위터 사용자들의 관계 정보를 이용하여 트위터 사용자의 성향을 분류하였다. 그림(2)에 나타난 사용자 성향 분류 방법은 다음과 같다.

1) 트윗 사용자와 뉴스 의견 리트윗 사용자들을 각각의 초기 그룹으로 생성한다.

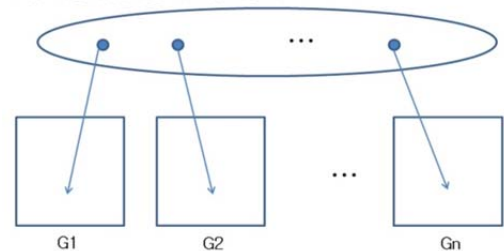
2) 리트윗 트위터 사용자들을 각 그룹마다 리트윗 관계가 가장 많은 그룹으로 분류한다.

3) 그룹간의 상호 리트윗관계와 공동 리트윗관계 정보를 반영한 유사도를 이용하여 계산하여 두 그룹을 묶어준다. 두 그룹 사이에서의 트위터 사용자가 상호 리트윗관계일 경우 유사도를 1로 설정하고 아닐 경우 두 트위터 사용자들의 트위터 히스토리에서 나타는 상호 리트윗관계의 비율을 이용하여 코사인 유사도 이용하여 유사도를 계산한다. 두 그룹 사이에서 발생할 수 있는 모든 유사도 값을 계산하여 평균값으로 두 그룹의 유사도를 계산한다.

4) 모든 그룹 사이에서의 유사도 중 가장 높은 유사도를 선택하여 해당되는 두 그룹을 묶어준다. 이러한 방법을 반복적으로 하여 최종적으로 두 그룹이 될 때까지 3)과 4)를 반복한다.

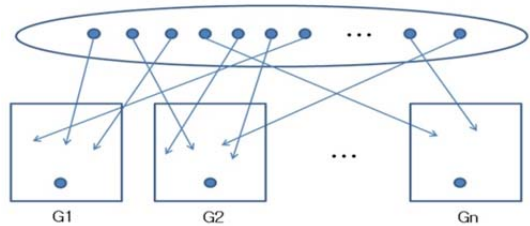
#### 1) 초기 그룹 생성

트윗 사용자 및 의견 리트윗 사용자

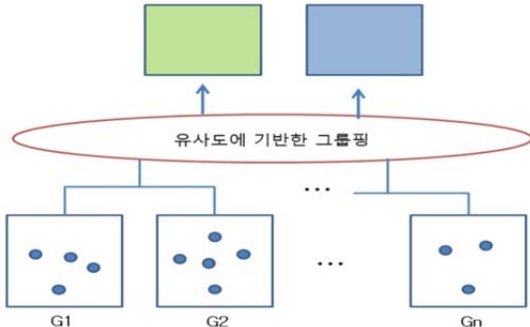


#### 2) 리트윗 사용자 분류

리트윗 사용자



#### 3) 반복적 유사한 그룹 그룹핑



#### 4) 최종적으로 두 그룹으로 분류

그림(2) 트위터 사용자 성향 분류 방법

<표 2> 날짜별 이슈가 되고 있는 정치 뉴스 기사

이슈	날짜	뉴스 기사 제목	뉴스 URL 개수	기사와 관련된 트윗 개수
1	2012. 9. 5	“양경숙 30억 송금계좌, 실제 주인은 양경숙	487	490
2	2012. 9. 7	[단독]"안철수협박, 이슈안되게" 새누리 긴급문자	951	1,581
3	2012. 9. 8	박근혜 “보도보고 알았다. 어떻게 친구사이에...”	1,305	1,319
4	2012. 9. 9	공격 당한 박근혜 오르고 안철수 0.2%P 하락	2,133	3,316
5	2012. 9. 10	MBC 노조 "김재철 해임시키지 않으면 파업재개" 공식선언	681	889

**<표 3> 관계 기반 분류 및 문서 내용 기반 분류 방법에서의 트위터 사용자 그룹 분류 결과에서의 정확도**

이슈	정답 개수	트윗 내용 유사도 기반 집단 평균 연결 클러스터링	소셜 사용자 네트워크 기반 사용자 분류 방법
1	136	52.2% (71/136)	<b>97.0% (132/136)</b>
2	179	54.7% (98/179)	<b>88.2% (158/179)</b>
3	126	57.9% (73/126)	<b>92.0% (116/126)</b>
4	129	<b>92.2% (119/129)</b>	70.5% (91/129)
5	160	92.5% (148/160)	<b>93.7% (150/160)</b>
전체	730	69.7% (509/730)	<b>88.6% (647/730)</b>

#### 4. 실험 및 평가

##### 4.1 실험 방법

본 논문에서 제안한 방법의 유효성을 검증하기 위해 선택한 5 개의 이슈를 포함하는 각각의 트위터 데이터에서 실험을 하였다. 선택된 뉴스 이슈에 대한 정보는 <표 2>와 같다. 제안한 사용자 네트워크 기반 성향 분류 과정에서 그룹 사이에서의 유사도가 0.1 이하가 되거나 그룹이 두 개가 될 때까지 반복하였다. 그룹이 세 그룹 이상 분류되었을 때 가장 많은 트위터 사용자를 가진 두 그룹을 선택하고 나머지 그룹들은 기타 그룹으로 구분하였다.

실험에서는 소셜 사용자 네트워크 관계에 기반하는 정보와 트윗 문서에 나타난 정보만을 이용하는 두 가지 실험 방법에 대해서 비교하였다.

- 소셜 사용자 네트워크 관계 기반 그룹 분류
- 트윗 문서 내용 유사도 기반 그룹 분류

트윗 문서 내용에 기반한 실험에서는 트윗 내용 유사도에 기반하는 집단 평균 연결 클러스터링을 이용하여 실험을 하였다. 트윗 문서는 어휘별로 tf-idf 가중치를 이용하여 벡터로 표현하였다. 트위터 사용자 성향 평가는 각각의 이슈마다 무작위로 120 명의 사용자를 선택하여 두 명의 평가자가 트위터 사용자의 성향을 평가하였다. 두 명의 평가자가 사용자를 동일한 성향으로 평가하였을 경우, 해당되는 성향으로 설정하였고 성향을 다르게 평가하였을 경우 해당 사용자를 기타 그룹으로 설정하였다. 실험에서 분류된 두 그룹에서 가장 많은 트위터 사용자를 가진 그룹을 성향이 많이 나타난 쪽으로 성향을 분류하였고 다른 그룹은 반대 성향을 가진 그룹으로 분류하였다. 성능 평가는 정확도를 이용하여 비교하였다.

##### 4.2 실험 결과

뉴스 이슈 5 개에 대한 트위터 사용자 성향 평가 실험 결과는 <표 3>과 같다. 5 개의 이슈에 대한 실험 결과에서 제안한 소셜 네트워크 기반 사용자 분류 방법의 정확도가 88.6%으로 트윗 내용 유사도 기반 집단 평균 연결 클러스터링 결과인 69.7%에 비해 높은 성능을 보였다.

이슈 3의 정답 데이터는 각각 55, 66 명으로 비율이 비슷하다. 이슈 3의 실험결과는 소셜 사용자 네트워크 기반 사용자 분류 방법에 대한 성능이 92.0%로 트윗 내용 유사도 기반 집단 평균 연결 클러스터링 방

법에 대한 성능 57.9%에 비해 매우 높은 성능을 나타내었다. 이것은 동일한 성향을 가진 트위터 사용자들 이더라도 어휘적으로 다양한 표현과 의견으로 인해 트윗 내용만을 분석하더라도 유사하게 나타나지 않는다고 볼 수 있다. 이러한 결과는 트위터 사용자 간의 관계 정보를 이용하는 것만으로도 사용자의 성향을 분류하는 방법이 의미 있음을 보여준다.

#### 5. 결론

본 논문에서는 트위터에서 이슈가 되고 있는 정치 뉴스 기사에서 트위터 사용자 사이에서의 소셜 네트워크 관계를 이용하여 사용자에 대한 성향을 분류해주는 방법을 제안하였다. 소셜 사용자 네트워크 기반 사용자 성향 분류 실험 결과 88.6%의 높은 성능을 보였다.

향후 연구에서는 추가적으로 뉴스 이슈에 대한 트위터 데이터와 트위터 사용자 히스토리에서의 어휘적 분석을 통하여 좀 더 명확한 사용자 성향 분류 방법에 대한 연구를 할 것이다.

#### 참고문헌

- [1] <http://www.twitter.com>
- [2] Z. Yang, J. Guo, K. Cai, J. Tang, J. Li, L. Zhang, and Z. Su, "Understanding retweeting behaviors in social networks" In Proceedings of CIKM, 2010.
- [3] D. Boyd, S. Golder, and G. Lotan "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter" In Proceedings of HICSS, 2010.
- [4] H. Kwak, C. Lee, H. Park and S. Moon, "What is Twitter, a social network or a news media?", In Proceedings of WWW, 2010.
- [5] 바야르, 이경순, "트위터 문서에서 시간 및 리트윗 분석을 통한 핵심 사건 추출", 정보처리학회논문지 B 한국어정보처리(정보검색분야), 2012.
- [6] S. Park, M. Ko, J. Kim, Y. Liu, J. Song, "The Politics of Comments: Predicting Political Orientation of News Stories with Commenters' Sentiment Patterns", In Proceedings of CSCW, 2011.
- [7] S. Park, M. K. Lee, J. Song, "Contrasting Opposing Views of News Articles on Contentious Issues", In Proceedings of HLT, 2011.