대립 관계에 있는 이슈에서의 바이어스 탐지

권아롱, 촐몽 바야르, 이경순
전북대학교 컴퓨터공학과
e-mail : lifecorrect@naver.com, bayar_277@yahoo.com, selfsolee@chonbuk.ac.kr

Bias Detection on Opposition Issue

A-Rong Kwon, Bayar Tsolmon, Kyung-Soon Lee
Dept. of Computer Science and Engineering, Chonbuk National University

요 약

사람들은 기업이나 제품에 대해 자신의 생각이 긍정적인지 부정적인지 표현하고자 한다. 트위터 사용자들은 트윗을 통해 자신의 생각을 표현한다. 본 논문에서는 트위터 데이터를 대상으로 대립관계에 있는 이슈에서의 바이어스 탐지 방법을 제안한다. 비지도학습 방법을 이용하여 트윗 패턴을 통해 세부자질을 추출하며, 세부자질에 대한 감정에 따른 확률 테이블을 구축하여 바이어스 탐지를 수행한다. 제안 방법의 유효성을 검증하기 위해 4 개의 대립 이슈에 대해 평가를 하였으며, 제안 모델이 기존의 모델보다 우수한 성능을 보였다.

1. 서론

웹상에서 인맥 관계를 강화시키고 또 새로운 인맥을 쌓으며 폭넓은 인적 네트워크를 형성할 수 있도록 해주는 서비스를 소셜 네트워크 서비스(Social Network Service; SNS)라 한다. 그 중에서도 트위터(twitter)는 한글이든 영문이든, 공백과 기호를 포함해 한 번에 최대 140 글자로 작성할 수 있는 소셜 네트워크 서비스이다. 또한 트위터 상에서 사용자들이 작성하는 메시지를 트윗(tweet)이라고 한다. 트윗은 사용자가 자신의 상태를 표현하거나 지인들과 대화하고, 다양한 정보를 공유하는 등의 목적으로 사용되고 있다. 또한 트위터 사용자들은 임의의 토픽(topic) 대해 긍정적인지 부정적인지 자신의 감정을 표현한다.

오피니언 마이닝(opinion mining)에 관한 연구로 Park[1]은 트위터 데이터 셋을 자동으로 구축하는 방법을 소개하고, 구축된 데이터 셋을 이용하여 긍정과 부정을 분류하는 분류기를 제안하였다. Popescu[2]는 이벤트 및 그 이벤트에 대한 설명을 추출하기 위해 학습을 통한 분류기와 긍정, 부정, 중립으로 구성 되는 감정 사전을 이용하였다.

사용자의 감정을 요약하고자 하는 연구로 Hu[3]는 상품 리뷰에서 상품에 대한 사용자들의 감정을 요약하는 방법을 제안하였다. Zhao[4]는 트위터를 요약하기 위해 2 개 이상의 단어로 된 언어단위를 이용하였다. Qiu[5]는 구문 관계(syntactic relation)를 이용하여 세부자질(target)을 추출하였다.

사용자의 감정을 분석하여 바이어스 탐지를 하는 연구로 Jiang[6]는 트윗에 대한 감정 분류를 하기 위해 세부자질과 구문론적 특성을을 이용하였다.

또한 Somasundaran[7]는 세부자질에 대한 감정 확률을 이용한 방법을 제안하였다. 토픽에 대한 세부자질을 추출하기 위해 감정 단어와 구문론적 특성을 이용하였으며, 추출된 세부자질에 대한 확률 테이블을 구축하기 위해 토픽에 대해 긍정 또는 부정으로 분류가 된 데이터를 이용한다. 또한 대립되는 두 토픽에서, 임의의 한 토픽에 대해 부정적으로 생각하는 것은 대립관계의 토픽에 대해 긍정적으로 생각하는 것으로 볼 수 있다는 특성을 이용한다.

본 논문에서는 이러한 특성을 이용하였으나, [7]과는 다르게 토픽과 동사 사이에 나오는 단어들을 tf·idf 방법을 통해 세부자질을 추출하였으며, 사용자들이 제품에 대한 감정을 피드백(feedback) 하지 않은 트윗 데이터를 이용한 비지도학습(unsupervised learning) 방법을 통해 세부자질에 대한 확률 테이블을 구축하였다. 또한 토픽에 대한 직접적인 감정 단어를 추출하여 바이어스 탐지하는데 이용하였다.

본 논문의 구성은 다음과 같다. 2 장에서는 바이어스 탐지를 위한 학습 과정을 소개하고, 3 장에서는 학습 과정에서 추출한 자질들을 이용하여, 대립 관계에 있는 이슈에서의 바이어스 탐지 방법을 제안한다. 4 장에서는 실험 및 분석을, 5 장에서는 결론 및 향후 연구에 대해 논하겠다.

2. 바이어스 탐지를 위한 학습 과정

사람들은 토픽에 대해 직접적으로 감정 단어를 이용하여 자신의 생각을 표현하기도 하며, 토픽과 관계있는 세부자질을 통해 표현하기도 한다. 다음은 트위터 사용자가 토픽에 대해 감정을 표현한 예이다.

*"삼성전자 **갤럭시탭** **추천**합니다. **무게**가 **가볍고** **두께**도 8.6mm 로 얇아 **편리한** 것이 특징이죠. **휴대성**도 **좋네요**!!"*

‘갤럭시탭’은 토픽이며, ‘추천’은 토픽에 대한 감정 단어, ‘무게’, ‘두께’, ‘휴대성’ 은 세부자질, ‘가볍고’, ‘편리한’, ‘좋네요’ 는 세부자질에 대한 감정단어이다. 토픽에 대해 직접적으로 감정 단어를 이용하여 생각을 표현할 수도 있으며, 세부자질에 대해 감정 단어를 언급함으로써 토픽에 대한 생각을 표현했음을 알 수 있다.

본 논문에서는 이러한 특징을 이용하여 바이어스 탐지를 하기 위해 세부자질 목록과 세부자질에 대한 확률 테이블을 구축하여 바이어스 탐지를 수행한다.

2.1 패턴 기반 세부자질 추출

세부자질을 추출하기 위해 패턴 방식을 이용하였으며 다음과 같다.

- **패턴 :** <topic> <target> <동사>
- **예제 :** *“니콘 렌즈 가격 어때요”*

토픽과 동사 사이에 있으면서, 동사 앞에 거리가 2 이내의 단어들이 세부자질 후보로 추출됨을 의미한다. 거리는 형태소 분석 결과를 기준으로 한다. 예제에서 ‘어때요’ 라는 동사 앞에 ‘렌즈’, ‘가격’ 이라는 단어가 거리 2 이내에 있으므로 이 두 단어는 세부자질 후보로 추출된다. 이렇게 추출된 세부자질 후보들은 빈도수에 의해 순위화되고, 그 다음에 tf·idf 방법에 의해 재순위화한다. tf·idf 방법에 의해 추출된 세부자질 추출 결과는 <표 1>과 같다.

<표 1> 세부자질 추출 결과

순위	갤럭시탭 vs. 아이패드	니콘 vs. 캐논	윈도우 vs. 맥 OS	삼성 vs. 애플
1	아이폰	렌즈	***설치***	제품
2	***사용***	카메라	*7*	아이폰
3	어플	***가방***	***사용***	***기업***
4	***출시***	DSLR	컴퓨터	***가격***
5	기능	***사용***	***모바일***	***판매***
6	***가격***	***가격***	***포맷***	***사용***
7	***삼성***	***D***	안드로이드	***우리나라***
8	전화	콘	***지원***	***서비스***
9	***요금***	***중고***	파일	***돈***
10	화면	55	프로그램	***회사***

“갤럭시탭 vs. 아이패드”는 대립되는 두 이슈인 ‘갤럭시탭’ 또는 ‘아이패드’ 가 언급된 트윗 집합을 의미한다. 두 토픽이 동시에 언급된 트윗들은 분석하기 어려워 학습과정에서는 다루지 않지만, 바이어스 탐지 과정에서는 다룬다. ‘사용’, ‘출시’, ‘가격’, ‘삼성’ 등과 같은 세부자질들은 제품의 사양정보에 없지만 토픽에 적합한 세부자질로 볼 수 있다. 이러한 세부자질들은 제품에 대한 정보를 이용하여 세부자질들을 수동적으로 구축하였을 경우 추출하지 못하는 세부자질들이다. ‘D’, ‘7’ 은 제품에 대한 모델명을 의미한다. <표 1>과 같이 세부자질 추출 결과 모두 적합함을 알 수 있다.

2.2 세부자질에 대한 확률 테이블 구축

본 논문에서는 바이어스 탐지를 하기 위해 세부자질에 대한 확률 테이블을 이용한다. 트위터 사용자들은 토픽의 세부자질에 대해 좋거나 나쁜 감정을 구체화하기 때문이다. 예를 들면 다음과 같다.

*“삼성전자 **갤럭시탭** 추천합니다. 무게가 가볍고 **두께**도 8.6mm 로 얇아 편리한 것이 특징이죠. **휴대성**도 좋네요!!”*

‘갤럭시탭’은 토픽이며, ‘무게’, ‘두께’, ‘휴대성’ 과 같이 세부자질에 감정이 표현되었을 경우, 각 세부자질에 대한 확률을 이용하여 바이어스 탐지를 할 수 있다.

이러한 특징을 이용하기 위해 세부자질에 대한 확률 테이블을 구축할 필요가 있다. 세부자질 뒤에 감정 단어가 있을 경우와 없을 경우에 대한 두 가지의 확률 테이블을 구축한다.

(1) 토픽에 대한 긍정 또는 부정 분류

확률 테이블을 구축하기 위해 토픽에 대해 긍정의 트윗인지, 부정의 트윗인지 알 수 없는 학습 데이터를 이용하며 비지도 학습 방법을 사용한다. 그러므로 해당 트윗이 임의의 토픽에 대해 긍정의 트윗인지, 부정의 트윗인지 자동적으로 미리 분류할 방법이 필요하다. 트윗을 자동적으로 긍정 또는 부정으로 분류하기 위해 패턴 방법을 사용하였다. 패턴은 크게 2 단계 방법이 있으며, 모든 학습 집합 트윗에 동시에 적용된다.

첫번째, 토픽 다음에 감정 단어가 나올 경우이다. 트위터를 이용하는 사용자들은 토픽을 언급하고, 그 다음에 감정 단어를 언급함으로써 해당 토픽에 대한 의견을 표현하는 것을 관찰하였다. 이러한 특징을 패턴으로 만들어 토픽에 대한 긍정 또는 부정을 자동 분류 하였다. 패턴은 다음과 같다.

- **패턴 :** <topic> <sentiment>
- **예제 :** *“**갤럭시탭**은 무게와 휴대성이 **좋습니다**”*

토픽 뒤에 감정 단어가 거리 10 이내에 언급되면 감정 단어에 의해 긍정 부정으로 분류한다. 예제에서 ‘갤럭시탭’ 토픽 다음에 ‘좋습니다’ 라는 긍정 단어가 나왔으므로, ‘갤럭시탭’에 긍정인 트윗으로 분류한다.

두번째, 이전의 패턴 방법인 토픽 다음에 나오는 감정 단어로 분류하는 방법은 토픽 주변의 문맥 정보를 고려하지 않았다. 문맥 정보를 고려한 패턴은 다음과 같다.

- **토픽이 비교 대상인 패턴 :**
  <topic> <‘보다’|‘비해’> <sentiment>

- **의문문 패턴 :**
  <topic> <sentimentt> <?>
  <topic> <'왜'> <sentiment>
- **감정 단어가 감정 사전에 없는 패턴 :**
  <topic> <sentiment> <'않'|'못'|'없' >
  <topic> <'안'| '못'> <sentiment>

토픽이 비교 대상인 패턴은 토픽 뒤의 거리 1 인 단어가 '보다' 또는 '비해'가 나오는 패턴이며, 해당 감정 단어를 반대로 적용한다. 의문문 패턴은 감정 단어 뒤의 거리 1 인 단어가 '?'또는 감정 단어 앞의 거리 1 인 단어가 '왜'가 나오는 패턴이며, 해당 감정 단어를 무시한다. 감정 단어가 감정 사전에 없는 패턴은 감정 단어 뒤의 거리 1 인 단어가 '않', '못', '없' 또는 앞의 거리 1 인 단어가 '안', '못' 이 나오는 패턴이며, 해당 감정 단어를 반대로 적용한다.

(2) 패턴 기반 세부자질에 대한 확률 추출

앞에서 설명한 긍정 또는 부정으로 자동 분류한 트윗 데이터를 이용하여, 바이어스 탐지에 이용될 세부자질에 대한 확률 테이블을 구축한다. 세부자질 뒤에 감정 단어의 유무에 따라 두 가지의 확률 테이블을 구축한다.

먼저 세부자질 뒤에 감정 단어가 있을 경우의 확률 테이블을 구축하기 위해 다음과 같은 패턴을 이용하였다.

- **패턴 :** <topic> <sentiment> <target> <sentiment>

토픽이 나오고 일정 거리 내에 감정 단어가 나왔을 경우, 토픽에 대해 긍정 또는 부정 트윗으로 자동 분류를 할 수 있다. 이렇게 분류를 할 수 있는 트윗에서만 세부자질에 대한 감정에 따른 확률을 추출한다.

확률을 추출하는 수식은 다음과 같다.

$$P(topic_j \mid target^{\pm}) = P(topic_j^{+} \mid target^{\pm}) + P(topic_i^{-} \mid target^{\pm}) \quad (1)$$

조건부 확률 $P(topic_j^{+} \mid target^{\pm})$에서 $topic_j^{+}$ 은 토픽 j 에 대한 긍정 트윗을 의미하며, 조건부 확률 $P(topic_i^{-} \mid target^{\pm})$에서 $topic_i^{-}$ 은 토픽 j 와 대립 관계인 토픽 i 의 부정 트윗을 의미한다. $target^{\pm}$은 세부자질 뒤에 감정 단어가 있을 경우를 의미한다. $P(topic_j^{+} \mid target^{\pm})$ 확률은 토픽 j 에 대해 긍정인 트윗이면서 세부자질 뒤에 감정 단어가 있는 트윗의 개수를 세부자질에 뒤에 감정 단어가 있는 트윗 개수로 나눈 값이다.

다음으로, 트위터 사용자들은 세부자질에 대해 감정이 구체화되지 않아도 세부자질의 언급만으로 토픽에 대한 자신의 의견을 표현한다. 세부자질 뒤에 감정 단어가 없을 경우의 확률 테이블을 구축하기 위해 다음과 같은 패턴을 이용하였다.

- **패턴 :** <topic> <sentiment> <target>

확률을 추출하는 수식은 다음과 같다.

$$P(topic_j \mid target) = P(topic_j^{+} \mid target) + P(topic_i^{-} \mid target) \quad (2)$$

조건부 확률 $P(topic_j^{+} \mid target)$에서 $topic_j^{+}$ 은 토픽 j 에 대한 긍정 트윗을 의미하며, 조건부 확률 $P(topic_i^{-} \mid target)$에서 $topic_i^{-}$ 은 토픽 j 와 대립 관계인 토픽 i 의 부정 트윗을 의미한다. target 은 세부자질 뒤에 감정 단어가 없을 경우를 의미한다.

3. 바이어스 탐지 기법

트위터가 대립관계에 있는 두 토픽이 언급되어 있다면, 어느 한쪽 입장에 치중되어서 의견을 나타내고 있는지를 탐지한다. 또한, 트위터가 하나의 토픽에 대해서만 언급되어 있다면, 그 토픽에 대한 긍정 또는 부정 의견을 인식한다.

본 논문에서는 임의의 트윗에 대한 바이어스 탐지를 하기 위해 각 토픽에 대한 최종 확률 $P(topic_j)$을 계산한다. $P(topic_j)$값을 구하는 수식은 다음과 같다.

$$P(topic_j) = \alpha \times P(topic_j^{\pm}) + (1-\alpha) \times \{P(topic_j \mid target^{\pm}) + P(topic_j \mid target)\} \quad (3)$$

α 값은 실험에 의한 파라미터 값이다. 본 논문에서는 0.5 로 설정한다.

$P(topic_j^{\pm})$ 은 토픽 다음에 감정 단어가 나왔을 경우에 대한 확률이다. 토픽 뒤에 나오는 감정 단어의 확률을 이용하여 바이어스 탐지한다. 예를 들면 다음과 같다.

*"**니콘**도 **좋고** **캐논**도 **좋지만** 개인적으로 **니콘**이 더 **좋다고** 생각합니다"*

$P(니콘^{\pm})$은 2/3 이며, $P(캐논^{\pm})$은 1/3 이다.

$P(topic_j \mid target^{\pm})$ 은 토픽 다음에 세부자질이 나오고 세부자질에 대한 감정 단어가 나왔을 경우에 대한 확률이다. 학습 과정에서 추출한 세부자질 뒤에 감정 단어가 있는 확률을 이용한다.

$P(topic_j \mid target)$ 은 토픽 다음에 세부자질만 나왔을 경우에 대한 확률이다. 학습 과정에서 추출한 세부자질 뒤에 감정 단어가 없는 확률을 이용한다.

수식 (3)을 이용하여 트윗에 등장하는 각 토픽에 대해 $P(topic_j)$ 값을 계산한 다음, 값이 높은 토픽으로 바이어스 탐지를 수행한다.

제안 시스템에서는 기존 방법[7]과는 다르게 토픽 다음에 감정 단어가 나왔을 경우와 세부자질 뒤에 감정 단어가 없는 확률을 고려하였다.

4. 실험 및 분석

4.1 실험 집합

본 논문에서 제안한 방법을 실험하기 위해 4 가지 이슈에 대해 Twitter API 를 이용하여 수집하였다. 실험 데이터 집합의 트윗 문서 개수와 이슈는 아래의 <표 2>와 같다.

<표 2> 실험 데이터 집합

	$Topic_1$	$Topic_2$	$Topic_1$ & $Topic_2$	총 개수
갤럭시탭 vs. 아이패드	100	100	300	500
캐논 vs. 니콘	100	100	300	500
윈도우 vs. 맥 OS	100	100	300	500
삼성 vs. 애플	100	100	300	500

정답 집합 구축은 대학원생 2 명이 직접 판정하였으며, 판정이 일치하지 않는 트윗은 의견이 일치하도록 하였다.

4.2 비교 실험 방법

- **OpTarget**
  : 세부자질에 대한 확률 테이블을 이용한 방법[7]
- **OpTopic**
  : 감정 단어의 빈도수를 이용한 방법
- **OpTopic & OpTarget^s**
  : 본 논문에서 제안하는 방법

4.3 실험 결과

성능 평가는 정확도(accuracy)로 평가하였으며 수식은 다음과 같다.

$$\text{accuracy} = \frac{\text{올바르게 탐지한 개수}}{\text{시스템이 탐지한 개수}} \quad (4)$$

수식 (4)에서 시스템이 탐지한 개수는 제안하는 시스템이 실험 집합에서 긍정, 부정, 중립 세 가지 중의 하나로 탐지한 트윗 수를 의미하며, 탐지 결과 정답 집합과 비교하여 올바르게 탐지한 개수를 정확도에 이용한다.

비교 실험 결과는 <표 3>과 같다.

<표 3> 비교 실험 결과

	OpTarget	OpTopic	OpTopic & OpTarget^s
갤럭시탭 vs. 아이패드	0.55	0.543	0.634
캐논 vs. 니콘	0.508	0.546	0.628
윈도우 vs. 맥 OS	0.484	0.548	0.611
삼성 vs. 애플	0.466	0.562	0.596
전체 집합	0.502 (-)	0.55 (+9.5%)	0.617 (+22.9%)

실험 결과에서 OpTarget 방법보다 본 논문에서 제안하는 방법이 모든 대립 이슈에서 우수한 성능을 보였으며 22.9% 향상률을 보인 것을 알 수 있다.

5. 결론 및 향후 연구

트위터 사용자들은 임의의 토픽에 대한 감정 표현을 세부자질을 통해 표현한다는 관찰을 통해 세부자질에 대한 확률 테이블을 이용한 바이어스 탐지 방법을 제안하였다. 세부 자질에 대한 확률 테이블을 구축하기 위해서 세부자질 목록을 추출하였고, 비지도 학습에서 각 세부자질들에 대한 확률을 구하기 위해 트윗의 긍정 부정 분류를 토픽 뒤에 나오는 감정 단어와 주변의 어휘를 고려한 패턴 방법을 이용하였다.

비교실험은 감정 단어를 이용한 방법과 세부자질에 대한 확률 테이블을 이용한 방법[7], 그리고 본 논문에서 제안하는 방법을 통해 수행하였다.

향후 연구에서는 토픽에 대한 사건이 일어나면 일련의 사건들은 토픽의 이미지에 영향을 끼칠 수 있다는 관찰을 통해 시간별 핵심사건을 고려한 바이어스 탐지 모델을 결합할 것이다.

참고문헌

[1] A. Park, P. Paroubek, “Twitter as a Corpus for Sentiment AnalysE and Opinion Mining.” In Proceedings of LREC 2010.

[2] A.-M. Popescu, M.Pennacchiotti, Deepa Arun Paranjpe. “Extracting events and event descriptions from Twitter” In Proceedings of WWW 2011.

[3] M.Hu, B.Liu “Mining Opinion Features in Customer Reviews” In Proceedings of AAAI 2004, pages 755-760.

[4] W.Zhao, J.Jiang, J.He, Y.Song, P.Achananuparp, Ee-Peng Lim, X.Li, “Topical Keyphrase Extraction from Twitter” In Proceedings of ACL 2011, pages 379-388.

[5] G.Qui, B.Liu, J.Bu, C.Chen, “Opinion Word Expansion and Target Extraction through Double propagation” In Proceedings of ACL 2011, pages 9-27.

[6] L.Jiang, M.Yu, M.Zhou, X.Liu, T.Zhao, “Target-dependent Twitter Sentiment Classification” In Proceedings of ACL 2011, pages 151-160.

[7] S.Somasundaran, J.Wiebe, “Recognizing Stances in Online Debates” In Proceedings of ACL 2009, pages 226-234.