준지도학습을 통한 세부감성 어휘 구축

조요한O, 오효정, 이충희, 김현기
한국전자통신연구원
yohan.jo@etri.re.kr, ohj@etri.re.kr, forever@etri.re.kr, hkk@etri.re.kr

Fine-grained Sentiment Lexicon Construction via Semi-supervised Learning

Yohan JoO, Hyo-Jung Oh, Chung-Hee Lee, Hyunki Kim
Electronics and Telecommunications Research Institute

요 약

소셜미디어를 통한 여론분석과 브랜드 모니터링에 대한 요구가 증가하면서, 빅데이터로부터 감성을 분석하는 기술에 대한 필요가 늘고 있다. 이를 위해, 본 논문에서는 단순 긍/부정 감성이 아닌 20종류의 세분화된 감성을 분석하기 위한 감성어휘 구축 알고리즘을 제시한다. 감성어휘 구축을 위해서는 준지도학습을 사용하였으며, 도메인에 특화되지 않은 일반 감성어휘를 구축하도록 학습되었다. 학습된 감성어휘를 인물, 스마트기기, 정책 등 다양한 도메인의 트위터 데이터에 적용하여 세부감성을 분석한 결과, 알고리즘의 특성상 재현율이 낮다는 한계를 가지고 있었으나, 대부분의 감성에 대해 높은 정확도를 지닌 감성어휘를 구축할 수 있었고, 감성을 직간접적으로 나타내는 표현들을 학습할 수 있었다.

주제어: 감성 분석, 오피니언 마이닝, 감성 어휘

1. 서론

트위터, 페이스북, 블로그 등의 소셜미디어 사용량이 증가하면서, 이런 빅데이터로부터 자동으로 오피니언 정보를 분석하려는 시도가 활발히 이루어지고 있다. 가령, 기업은 자사 제품이나 서비스가 소셜미디어 상에서 어떤 평가를 받고 있는지 파악함으로써 마케팅 전략에 참고할 수 있고, 정책 기관에서는 정책에 대한 여론분석을 통해 정책 수정방향 및 홍보방식 등을 결정할 수 있다. 이러한 필요가 대두되면서 빅데이터 기반 오피니언 마이닝을 전문으로 하는 브랜드 모니터링 서비스 사업도 활발해지고 있다.

본 논문에서는, 빅데이터 기반의 브랜드 모니터링 서비스를 위한 감성분석을 위해 감성어휘를 구축하는 기법을 제시한다. 그동안 학계에서 제시되어 온 감성분석 기법들은 각기 장단점을 가지고 있으며, 사용 목적에 따라 적절한 기법을 선택할 필요가 있다. 본 논문에서 초점을 맞추고 있는 브랜드 모니터링 서비스를 위해서는 다음과 같은 조건을 만족시키는 감성분석 기법이 필요하다.

먼저 단순 극성(긍정/부정)감성이 아닌 더 세분화된 세부감성을 제공할 수 있어야 한다. 기존의 감성분석 기법들이 주로 다루는 극성감성은 정보력에 한계가 있다. 가령, 어떤 기업의 제품에 대해 '걱정'하는 여론이 주를 이룬다면 기업에서는 소비자를 안심시킬 수 있는 홍보를 기획하는 반면, '불만' 여론이 주를 이룬다면 소비자들이 제품에 대해서 불만을 갖는 부분을 개선하는 방향으로 전략을 세울 수 있을 것이다. 하지만 '걱정'과 '불만'은 모두 부정적인 여론이므로, 극성감성 분석만 가지고는 이와 같은 구체적인 전략을 세우기가 어렵다.

또한, 감성분석을 통해 텍스트에서 찾아낸 감성에 대해, 사람이 납득할 만한 단서를 제시할 수 있어야 한다. 그 이유는, 브랜드 모니터링 서비스를 이용하는 사용자들은 서비스의 신뢰도 파악을 위해 서비스에서 제공하는 감성분석 결과에 대한 판단 근거를 요구하기 때문이다. 단서를 제공하기 위해서는 어휘(lexicon) 기반의 감성분석이 적합하다. 가령, 지지벡터기계(SVM: Support Vector Machine)[1]처럼 여러 자질들의 조합을 통해 감성을 분류하는 방식으로는 감성 판단의 근거를 사용자들에게 납득시키기가 쉽지 않다. 반면 어휘를 기반으로 하는 경우, 감성 판단의 근거가 되는 단어 혹은 자질을 제시함으로써 사용자들을 이해시키기가 비교적 용이하다.

다음으로, 높은 정확도(precision)가 높은 재현율(recall)보다 우선순위에 있어야 한다. 서비스 사용자들은 감성분석 된 결과량이 많지 않을지언정 틀린 결과를 보고 싶어 하지 않는다. 게다가 재현율이 낮은 단점은 빅데이터 분석 결과를 활용함으로써 보상 가능하다. 일반적으로 어휘 기반의 감성분석 방식은 다른 기계학습 방식에 비해 재현율이 낮다는 단점을 가지고 있는데, 왜냐하면 기본적으로 어휘에 들어가는 자질은 사용자에게 보여지기 위한 것이고, 따라서 정확도가 높은 자질들 위주로 들어가기 때문이다.

마지막으로, 너무 많은 양의 학습데이터 구축을 필요로 해서는 안된다. 브랜드 모니터링 서비스 사용자들이 요구하는 다양한 도메인을 반영하는 어휘가 모두 포함되도록 학습데이터를 구축하기에는 너무 많은 비용과 노력이 수반된다. 따라서 본 논문에서는 준지도학습(semi-supervised learning) 방식을 취함으로써, 태깅된 데이터와 태깅되지 않은 데이터를 모두 사용하여 학습한다.

이후 장들에서는 위의 네 가지 조건을 전제로 하여, 한국어 트윗으로부터 세부감성 분석을 위한 어휘를 구축하는 알고리즘을 제시하고 평가 결과를 제시한다. 2장에서는 관련 연구를 살펴보고, 3장에서는 구체적인 알고리즘을 제시하며, 4장에서는 평가데이터 및 평가척도에 대해 명시하고, 5장에서는 평가 결과를 분석하여, 6장에서 이 논문의 결론을 지으며 마무리 할 것이다.

2. 관련 연구

세부감성을 분석하는 연구는 크게 벡터표현 기반, 매트릭스 기반, 그래프 기반의 연구가 이루어졌다. 먼저 벡터표현 기반의 연구[2]에서는 ExperienceProject.com에 있는 고백에 관한 글들을 대상으로 다섯 종류의 감성을 분석한다. 모든 단어와 감성은 벡터로 표현되고, 문서의 감성은 문서에 들어있는 단어벡터와 감성벡터의 내적으로 정의된다. 이 방식에서는 지도학습을 통해 단어벡터와 감성벡터를 계

산하며, 태깅된 데이터에 존재하지 않는 단어에 대해서는 단어벡터를 계산할 수 없다. 매트릭스 기반의 연구[3]에서는 뉴스 헤드라인에 대하여 네 종류의 감성을 분석한다. 문서와 감성을 매트릭스로 표현하고 매트릭스 분해(matrix factorization)를 통해 감성을 분석하는데, 역시 태깅된 데이터에 존재하지 않는 단어에 대해서는 계산이 불가능하다. 그래프 기반의 연구[4]는 본 논문과 가장 유사한 접근 방식을 취한다. 단어들을 꼭지점으로 나타내고 단어들 간의 관계를 변으로 나타낸 뒤, 감성을 지닌 단어들로부터 감성이 알려지지 않은 단어들로 감성을 확장시킨다. 감성을 모르는 대량의 단어들에 대해서 학습할 수 있다는 장점이 있으나, 원래 논문에서는 긍정/부정 두 종류의 극성감성만을 사용하였다. 그밖에도 일반적으로 널리 사용되는 Structural SVM[5]등을 사용해 세부감성을 분류할 수 있으나, 역시 태깅된 데이터로만 학습할 수 있다는 한계가 있다.

감성	포함 감성 예	감성	포함 감성 예
자신감	자신감	두려움	공포 무서움 걱정 불안
감동	감탄 존경 경의	화남	격노 분노 억울함 짜증
감사	감사	싫어함	미움 증오 혐오 지루함 심심함 싫증
기대감	설렘 희망 기대됨	슬픔	연민 애처로움 쓸쓸함 참담함 비참함 안타까움
좋아함	호감 사랑	실망	체념 아쉬움 그리움 허무 어이없음 황당함 후회
기쁨	즐거움 흥미 행복함 반가움 만족	수치심	창피 무안 부끄러움 민망
안심	위안 안심 안도	곤란	서먹 어색 거북함 곤란 난처
신뢰	신뢰 믿음	미안함	미안함
선의	인사말 축하 응원 격려	부러움	질투 시기
의심	불신	반대	비동의 비납득

표 1 세부감성 및 포함되는 감성의 예

3. 방법

3.1 감성 분류 체계

본 논문에서는 세부감성을 위해, [6]에서 제안하는 감성분류체계를 기반으로 한다. 이 감성분류체계는 기존의 심리학에서 사용되는 감성체계[7]를 소셜 웹 미디어에 적합하도록 수정한 19개의 카테고리로 구성되어 있다. 본 논문에서는 기본적으로 이 분류체계를 사용하되, 서비스에 적절하도록 다음과 같은 수정을 하였다.

- 세부감성이 애매한 경우를 위해 만들어 놓았던 '나쁨(bad)', '좋음(good)' 제거
- 사전적 정의에 따라 '두려움'과 '걱정' 병합
- 소셜미디어에 많은 '의심', '수치심', '곤란' 감성 추가
- 출현 빈도가 높은 '감사'는 '감동'으로부터 분리
- 쓰임새의 특성상 '안심'을 '만족'으로부터 분리하고, '만족'은 '기쁨'에 병합
- '인정'에서 감성이 불분명한 '이해', '납득', '동감'을 제거

이렇게 수정된 세부감성 분류체계가 표 1에 있다.

3.2 감성어휘 구축 모델

본 논문에서 제시하는 모델은 [4]의 모델을 응용한다. 이는 그래프 기반의 알고리즘으로서, 원래 논문에서는 극성 감성어휘를 구축하기 위해 사용되었다. 감성어휘에 들어갈 후보들이 꼭지점이 되고, 후보들 간의 상호정보량(mutual information)이 변이 되며, 극성이 이미 태깅된 꼭지점이 몇 개 있다. 기본적인 아이디어는 태깅된 꼭지점과의 상호정보량이 큰 꼭지점들은 비슷한 강도의 극성을 갖고, 반면 상호정보량이 작은 꼭지점들은 낮은 강도의 극성을 갖도록 그래프 상에서 극성값을 확산시키는 것이다.

구체적인 알고리즘을 설명하기에 앞서, 본 논문에서 특별히 그래프 확산 알고리즘을 선택한 이유는 다음과 같다. 대부분의 지도학습은 태깅된 학습데이터에 나타나는 자질 외에는 학습이 어려운데, 빅데이터에서 모든 중요한 자질을 포함하는 태깅 데이터를 구축하기는 쉽지 않다. 즉, 트위터와 같은 빅데이터를 충분히 수용할 수 있을 정도의 감성어휘를 위한 학습데이터를 구축하는 것은 현실적으로 어렵기 때문에 준지도학습 방식이 적합하다. 또한, 지도학습은 하나의 학습데이터 포인트에 대해 카테고리 수만큼의 true/false가 태깅되어 있어야 하는 것에 반해, 그래프 확산 알고리즘에서는 하나의 데이터 포인트에 대해 false 정보를 태깅할 필요가 없어서 학습데이터 구축이 용이하다. 특히 본 논문에서처럼 20 종류의 감성을 사용하는 경우, 지도학습 방식에서처럼 하나의 트윗에 대해 20개의 true/false를 태깅하는 것보다, 각 감성별로 해당하는 단어를 독립적으로 수집하는 것이 훨씬 효율적이다.

3.2.1 그래프 구축

본 알고리즘에서 그래프 $G=(V,E)$를 구축하고 감성어휘를 구축하는 방법을 구체적으로 설명하도록 한다. 먼저 감성집합 $S=\{s_1,\cdots,s_M\}$이 있고, M은 감성 종류의 개수이다. 각 감성에 대해서는 학습을 위한 기본감성어휘 $L_k^S, 1 \leq k \leq M$가 있다. 임의의 학습데이터 트윗 d는 다음과 같은 자질 집합으로 나타낼 수 있다.

$$d = F_d \cup \{LABEL(s_k) : s_k \text{는 } d\text{에 들어있는 감성}\}$$

F_d는 트윗 d에 있는 형태소나 단어 등 학습을 통해 감성어휘에 들어가게 될 자질들의 집합이다. 어떤 트윗 d가 L_k^S에 속하는 자질을 적어도 하나 이상 포함하는 경우, s_k가 d에 들어있다고 보고, 감성 s_k의 감성값 $LABEL(s_k)$을 하나의 자질로 간주한다. 한편, 준지도학습에서는 태깅이 되어있지 않은 트윗들도 학습에 함께 사용된다. 이런 트윗들에 대해서는 레이블 자질이 존재하지 않는다.

G의 꼭지점 집합 $V = V^F \cup V^L$에 대해서, 꼭지점 $v \in V^F$는 감성어휘에 들어갈 후보 자질을 나타내고 V^L은 $\{LABEL(s_k) : 1 \leq k \leq M\}$을 나타낸다. 임의의 꼭지점 v_i가 나타내는 자질을 f_i라고 할 때, 꼭지점 $v_i, v_j \in V$를 연결하는 변 $e_{ij} \in E$에 대해 PMI h_{ij}를 다음과 같이 정의한다.

$$\begin{aligned} h_{ij} &= PMI(f_i, f_j) \\ &= \log \frac{p(f_i, f_j)}{p(f_i)p(f_j)} \\ &= \log \frac{n(f_i, f_j)/N}{n(f_i)/N \cdot n(f_j)/N} \\ &= \log \frac{n(f_i, f_j)N}{n(f_i)n(f_j)} \end{aligned}$$

여기에서 $n(f_i, f_j)$는 f_i와 f_j가 모두 등장하는 트윗의 개수, $n(f_i)$는 f_i가 등장하는 트윗의 개수, N은 모든 트윗 개수이고, h_{ij}의 범위는 다음과 같다.

$$-\infty \leq h_{ij} \leq \min\{\log(N/n(f_i)), \log(N/n(f_j))\}$$

e_{ij}의 무게값에 대해서는, f_i와 f_j의 PMI 값이 0보다 작을 경우 무게값을 0으로 설정하고, 0 이상일 경우 h_{ij}를 [0,1]로 정규화시킨 값을 무게값으로 사용한다. 즉, e_{ij}의 무게값 w_{ij}는 다음과 같

이 정의된다.

$$w_{ij} = \begin{cases} 0 & \text{if } h_{ij} < 0 \\ h_{ij}/\log(N) & \text{otherwise} \end{cases}$$

그래프 상에서 감성이 확산되는 방식은 다음과 같다. 임의의 꼭지점 $v \in V$는 모든 감성에 대해 감성강도 $I_k, 1 \leq k \leq M$를 가지고 있다. 감성 레이블 꼭지점 $v \in V^L$는 그 레이블이 의미하는 감성 s_k에 대해 $I_k = 1$의 값을 갖는다. 임의의 꼭지점 $v_i \in V$와 인접한 꼭지점 $v_j \in V$에 대해 v_i의 감성강도 I_{ik}는 다음과 같이 계산된다.

$$I_{ik} = \max_j \{I_{jk} \cdot w_{ij}\}$$

각 $v \in V_L$에 대해, v와 연결된 모든 꼭지점들에 대해 I를 계산하고, 그 꼭지점들과 연결된 꼭지점들에 대해서 또 I를 계산을 하는 식으로 확산해 나간다. 초기에는 감성 레이블 꼭지점들에 대해서만 I의 값이 0보다 크지만, 감성이 확산됨에 따라 감성 레이블과 직접적(같은 트윗에 들어있는 자질들), 간접적으로 연결된 자질들도 0보다 큰 I를 갖게 된다. 이것이 준지도학습의 핵심이다. 한편 레이블 꼭지점이 두 개 이상인 감성의 경우, 감성강도의 정의에 따라 그래프의 각 꼭지점은 감성강도가 가장 큰 경로만을 취한다. 따라서 임의의 꼭지점의 I는 증가하는 방향으로만 업데이트 되므로 모든 감성 레이블 꼭지점들에 대하여 순차적으로 업데이트가 완료되면 I는 실제 감성강도 값으로 수렴한다. 또한 그래프에 싸이클이 존재하는 경우에도, w_{ij}은 [0,1]의 값을 가지므로 문제없이 감성강도를 계산할 수 있다.

한편, 모든 감성에 대해 높은 감성강도를 갖는 자질을 처리하기 위해서, 감성강도의 평균값을 빼서 조정하도록 한다. 즉, 조정된 감성강도 I_{ik}'는 다음과 같이 계산된다.

$$I_{ik}' = I_{ik} - \frac{1}{M-1} \sum_{\substack{p \neq k \\ 1 \leq p \leq M}} I_{ip}$$

이렇게 계산된 감성강도에 대해서, 각 감성별로 적절한 역치(threshold)를 선정하여 감성어휘에 포함시킨다.

3.2.2 트윗의 RT 처리

하나의 트윗에 리트윗한 내용이 포함되어 있을 수 있다. 예를 들어, "아후~!! 속 터져~!! RT @asdfjkl: 그리고 어제부터 쇼고객센터 로그인이 안되는데ㅠㅠ"라는 트윗은 아이디가 asdfjkl인 사용자가 쓴 트윗에 자신의 생각을 덧붙여 작성한 것이고, 이 경우 asdfjkl의 트윗이 리트윗되었다고 한다. 이런 경우에 트윗 하나의 범위를 어디까지 할 것인지 결정해야 한다. 인기가 많아서 리트윗이 많이 되는 트윗은 데이터에 자주 중복되어 나타나게 되고 이렇게 빈도가 비정상적으로 높아진 트윗은 자질 간의 상호정보량을 계산하는 데에 혼란을 줄 수 있다. 따라서 본 논문에서는 트윗의 범위를 다음의 세 가지 경우로 구분하여 실험하였다.

- RTLevel: 어떤 트윗이 리트윗을 포함할 경우, 각 리트윗을 별개의 트윗으로 간주한다. 가령, 어떤 트윗에 리트윗이 두 개 포함되어 있을 경우, 세 개의 트윗(직접 작성한 내용 + 두 개의 리트윗)으로 간주한다.
- TweetLevel: 어떤 트윗이 리트윗을 포함하고 있더라도 모두 합쳐서 하나의 트윗으로 간주한다.
- TweetLevel+: TweetLevel와 비슷한 방식이나, 리트윗 내용은 자질 간 상호정보량을 계산하지 않는다. 이는 TweetLevel에서 빈도가 높은 리트윗은 자질 간 상호정보량이 중복 계산되어 결과가 왜곡되는 것을 막기 위함이다.

4. 평가 준비

본 논문에서 제안하는 알고리즘을 통해 구축된 감성어휘를 평가하기 위해서, 구축된 감성어휘를 이용해 트윗의 감성을 분석하는 성능을 평가하기로 한다. 이 장에서는 구체적인 평가 방법, 기본감성어휘를 구축하는 방법, 평가데이터를 구축하는 방법, 선택한 자질들에 대해서 설명하도록 한다.

4.1 평가 방법 및 척도

본 논문에서 제시하는 알고리즘의 목적은 트위터 데이터 같은 빅데이터의 감성을 분석하여 여론을 파악하는 것이다. 이를 위해, 알고리즘을 통해 구축된 감성어휘를 사용하여 각 트윗별로 20종류의 감성을 분석하는 성능을 측정하도록 한다. 하나의 트윗이 다양한 감성을 표현할 수 있기 때문에, 이 실험에서는 하나의 트윗을 하나의 세부감성으로 분류하지 않고 대신 트윗이 가지고 있는 모든 감성을 찾도록 한다. 이를 위해, 구축된 감성어휘에 포함된 자질이 들어있는 트윗은 해당 감성을 지니고 있다고 판단하였다. 한편, 각 감성에 대해서 감성강도가 얼마인 자질들까지 감성어휘에 포함시킬지 그 역치를 정해야 한다. 본 실험에서는 0부터 1 사이를 0.02 간격으로 역치를 선정하여 총 50가지 경우에 대해 모두 평가해보고 성능이 가장 좋은 역치를 선택하였다.

일반적으로 이러한 정보추출의 성능을 평가하기 위한 척도로서 정확도와 재현율, 그리고 둘의 조화평균인 F1 score가 많이 사용된다. 각 감성별로 감성어휘의 역치에 따른 F1 score를 계산한 후 바로 성능 척도로 사용하는 것이 직관적인 방법이기는 하나, 본 논문에서 성능 척도로 사용하기에는 곤란한 점이 있다. 지도학습이 아닌 준지도학습을 사용한다는 점, 어휘 기반으로 감성을 분석한다는 점, 그리고 트윗의 절반 이상은 감성이 없다는 점 때문에, 재현율이 정확도에 비해 매우 낮다. 다시 말해, F1 score가 재현율에 지배적이게 되며, 재현율을 높이기 위해서는 정확도가 크게 희생된다. 이는 본 논문에서 목적을 두고 있는 브랜드 모니터링 서비스에서 정확도를 우선시 하는 것과 상충한다.

따라서 본 실험에서는 기본적으로 사용자들이 납득할 만한 수준의 정확도를 실험적으로 70%로 잡고, 정확도가 70% 이상인 감성어휘는 70% 미만인 감성어휘보다 무조건 우위에 있도록 한다. 정확도가 70% 이상인 감성어휘들 간에는 F1 score를 이용하여 성능을 비교하고, 정확도가 70% 미만인 감성어휘들 간에는 정확도를 기준으로 성능을 비교한다. 즉, 감성어휘 L의 성능 $PERF(L)$은 다음과 같이 정의된다.

$$PERF(L) = \begin{cases} Fscore(L) + 1 & \text{if } Prec(L) \geq 0.70 \\ Prec(L) & \text{otherwise} \end{cases}$$

4.2 기본감성어휘 구축

영어로 된 감성어휘 중에서 널리 사용되는 어휘로는 WordNet-Affect[8], LIWC[9], SentiWordNet[10] 등이 있으나, 한국어를 위해 사용할 만한 감성어휘는 구하기 어려운 실정이다. 세부감성을 이용해 노래 가사의 감성을 분석한 비교적 최근 논문[11]에서는 감성어휘를 수작업으로 구축하여 사용하였다. 따라서 본 논문에서는 간단한 방식을 사용하여 다음 항목들로 구성된 기본감성어휘를 구축하였다.

(ㄱ) 세부감성의 이름 (예: "감동", "슬픔")
(ㄴ) 세부감성에 포함되는 감성의 이름(표 1) (예: "감탄", "존경", "연민", "애처로움")
(ㄷ) 세부감성의 동의어 (예: "감동"의 동의어 "감명")

도메인	키워드	트윗 개수
인물	김태촌, 곽노현, 박근혜, 나경원, 박원순, 이명박, 한명숙, 문재인, 안철수, 나꼼수, 정봉주	1700
스마트 기기	아이폰, 스마트폰, 배터리, 킨들, 블랙베리, 베가, 갤럭시탭, 애플TV, LTE, 모바일OS, 옵티머스, 넥서스폰, 아이팟, 갤럭시폰, 아이패드, 구글폰	2000
정책	민주당, 전자팔찌, 종합편성채널, 셧다운제, 천안함, 반값등록금, 학생인권조례, 새누리당, 실명제, 무상급식, 요일제, 한미FTA	2000

표 2 평가데이터 키워드

(ㄷ)에서 너무 일반적인 동의어는 제외하였다. 예를 들어, "감동"의 동의어 중에 "느낌"이 있는데, "느낌"은 일반적으로 "감동"의 의미보다 더 포괄적으로 사용되므로 제외하였다. 또한 감성 이름 중에 명사형으로밖에 사용될 수 없는 경우는 동사로 사용될 수 있도록 형태를 확장하였다. 예를 들어, "두려움"과 같은 감성은 "두렵다", "두려워" 등도 포함할 수 있도록 형태를 확장하였다. 이렇게 하여, (ㄱ)과 (ㄴ)을 통해 구축된 용어는 140개, (ㄷ)을 통해 형태 확장된 용어는 60개가 되어 총 200개의 용어를 가진 기본감성어휘를 구축하였다.

4.3 학습데이터

학습데이터는 다음과 같은 방식으로 트윗을 샘플링하여 구축하였다. 본 논문에서 전제로 하는 서비스의 특성상 특정 도메인에 특화되지 않은 범용화 된 어휘사전을 구축하는 것이 목표이므로, 트위터 상의 모든 트윗을 대상으로 샘플링하였다. 그러나 트위터 상의 트윗 중 85% 이상은 감성이 없기 때문에[6], 완전히 임의로 샘플링을 할 경우 감성과 잠재적으로 관련 있는 자질을 충분히 확보하기 위해서 필요한 샘플의 크기가 매우 커지게 되고 자연히 학습에 필요한 메모리와 시간도 크게 증가한다. 따라서 감성과 관련 있는 자질을 확보하면서 샘플의 크기를 줄이기 위해서 "기분", "가슴이", "느낌"이 들어간 트윗들을 샘플링하였다. 이는 한 연구[12]에서 "We feel"이라는 텍스트가 들어있는 문장을 분석하여 블로그 상의 감성을 분석한 데에서 아이디어를 얻은 것이다. 이런 방식으로, 2010년 11월부터 2011년 3월 사이에 작성된 트윗을 대상으로 2백만 개의 트윗을 샘플링하여 학습데이터로 사용하였다. 그중 기본감성어휘에 포함된 용어가 들어있는 트윗은 태깅 데이터로, 나머지 트윗들은 태깅되지 않은 데이터로 사용되어 준지도학습을 하게 된다. 2백만 개의 트윗 중에서 감성이 태깅된 데이터는 419,456개이고, 태깅이 되지 않은 데이터는 1,580,544개이다. 학습데이터에서 기본감성어휘를 통해 태깅된 세부감성 분포가 그림 1에 나와 있다. '기쁨'과 '좋아함' 감성이 많이 태깅되었는데, 이는 두 감성의 기본감성어휘 중 "기쁘다", "즐겁다", "사랑", "좋아하다" 등의 표현이 텍스트 상에 많이 나타났기 때문이다.

4.4 평가데이터

평가데이터를 구축하기 위해서, 2011년에 작성된 트윗 중에서 다양한 이슈와 관련된 트윗들을 샘플링하고 여기에 20종류의 세부감성을 태깅하였다. 고려한 이슈는 인물, 스마트기기, 정책 중 하나에 속하며, 그 리스트가 표 2에 나열되어 있다.

태깅은 각 트윗에 대하여 트윗에 들어있는 모든 세부감성을 태깅하였다. 하나의 트윗에 대해 세 명이 상의하여 둘 이상의 동의를 얻은 감성만 태깅하였고, 그런 감성이 없는 트윗은 '중립'으로 태깅하였다. 총 5700개의 트윗 중 63%의 트윗이 '중립'으로 판단되었다. 감성이 존재하는 트윗에 대해서 세부감성의 비율이 그림 2에 있다. '실망', '싫어함', '화남', '슬픔'과 같은 부정적인 감성이 많은 비율을 차지하고 있고, '자신감', '미안함', '곤란' 등의 감성은 그 비율이 매우 낮은 것을 알 수 있다.

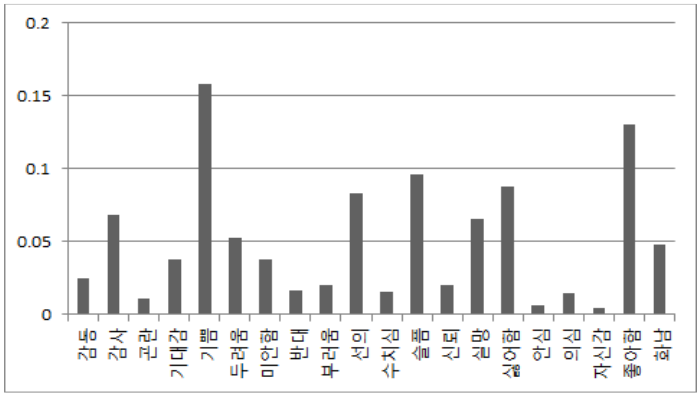

그림 1 학습데이터 내 태깅된 세부감성 분포

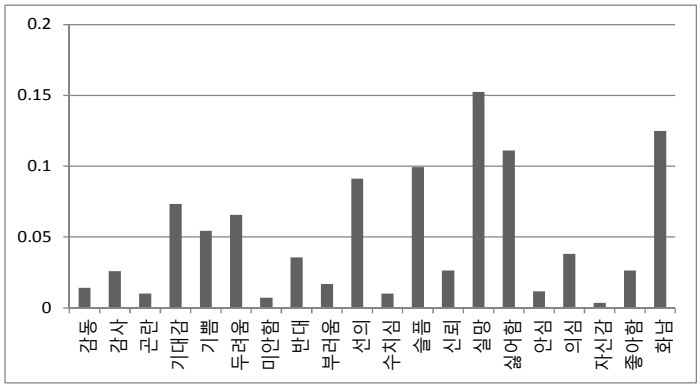

그림 2 평가데이터 세부감성 분포

평가데이터에 포함된 트윗의 예는 다음과 같다.

- 실망: #MLB_ 9호선은 DMB가 안 되는군요... 급행 빼고는 좋은 게 없는 듯... 4칸 짜리 객차로 어쩌려고...
- 기쁨: 수요일은 차량요일제로 전철 이용하는 날이 많습니다. 전철에서 간만에 Podcast 골라 듣고 회사 입구에서 커피 한잔 사들고 오는 맛도 괜찮군요!

4.5 자질

감성어휘에 들어갈 자질로 다음 여섯 개를 사용했다.

- MRP: 명사, 동사, 형용사 형태소 (예: 따분하/pa)
- MRPN: 명사, 동사, 형용사, 부사, 의존명사 형태소의 바이그램 (예: 우울하/pa 기분/pv)
- WD: 단어 (예: 감사합니다!, 디자인이)
- WDN: 단어 바이그램 (예: 뭔가 어색한, 본의 아니게)
- PRD: 명사 형태소 + 용언(동사, 형용사, 명사) 형태소 (예: 걱정/nc 많/pa, 희망/nc 없/pa)
- RM: 문장의 마지막 용언 형태소 (예: 방심하/pv)

형태소와 관련된 자질에 관해서는, 자질을 추출하기 전에 몇 개의 정규표현을 사용해 부정어 처리를 하였다. 예를 들어, "기분/nc 이/jc 좋/pa 지/ec 않/px 다/ef"의 경우 "기분/nc 이/jc 안_좋/pa 다/ef"와 같이 변환하였다. 학습데이터에 대해, 빈도가 낮은 자질은

MRP	MRPN	WD	WDN	PRD	RM
13700	12536	18553	14892	10807	3975

표 3 자질별 빈도

필터링하였고, 최종적으로 남은 자질의 개수가 표 3에 나와 있다.

5. 평가 결과

트윗을 구분하는 세 가지 방법(RTLevel, TweetLevel, TweetLevel+)과 여섯 종류의 자질(MRP, MRPN, WD, WDN, PRD, RM)을 조합하여 세부감성 분석 성능을 측정하였다. 각각의 트윗 구분 방법에 대해, 최대 두 개의 자질을 조합하였고, 감성별로 가장 높은 성능을 보이는 감성어휘를 선택하였다.

5.1 트윗 구분 방법별 성능

먼저 기본감성어휘만을 이용한 방법과 트윗을 구분하는 세 가지 방법 RTLevel, TweetLevel, TweetLevel+에 대해 성능을 측정하였고, 그 결과가 표 4에 나와 있다. 세 방법 중에서 가장 높은 성능을 보인 결과를 볼드체 및 음영으로 표시하였다.

전체적으로 몇 개의 감성을 제외하고는 재현율이 매우 낮은 것을 알 수 있다. 이는 도메인에 특화된 처리를 전혀 하지 않은 상태에서 중립 트윗이 절반 이상을 차지하는 평가데이터에 대해 높은 정확도를 얻으려고 하다 보니 맞닥뜨리게 되는 한계로 보인다.

기본감성어휘가 들어있는지 여부로 감성을 판단한 결과는, 구축된 감성어휘를 사용한 결과에 비해 전반적으로 정확도가 떨어졌다. 주된 이유는 기본감성어휘에 있는 "만족", "사랑" 등의 짧은 표현들이 감성을 부정확하게 판단했기 때문이다. 다만 '부러움'과 '슬픔' 감성에 대해서는 구축된 감성어휘에 비해서 높은 성능을 나타내었다. '부러움'의 경우 "부럽다", "질투", "탐나다"의 표현만으로 높은 정확도와 재현율을 내었고, 이를 통해 트위터 상에서 '부러움' 감성이 이 표현들을 통해 주로 나타남을 알 수 있다. 반면 '슬픔'에 해당하는 기본감성어휘에는 "슬프다", "연민", "애처롭다" 등 약 15 종류의 표현이 들어있다. 이 표현들의 존재를 통해 감성을 정확하게 판단할 수 있었으며, 또한 표현의 다양성으로 인해 재현율도 높았다.

RTLevel, TweetLevel, TweetLevel+의 성능을 비교해보면 눈에 띄는 차이는 보이지 않는다. 다만 '감동'의 경우에는 TweetLevel와 TweetLevel+에서 정확도 50% 이상인 감성어휘를 얻을 수 없었던 반면에 RTLevel에서는 얻을 수 있었다. 결과 어휘를 자세히 분석한 결과, RTLevel에서는 '감동'을 직접적으로 표현하는 RM 자질(예: 감탄스럽/pa, 감동적/nc)이 상위를 차지하고 있었던 반면, RTLevel+는 주로 명사(예: 감동/nc)가 상위를 차지하고 있었다. 이는 트윗에서 직접 작성한 내용과 리트윗을 함께 고려할 때 서로 영향을 주어 좀 더 직접적인 감성표현을 얻지 못하는 것으로 보인다. '미안함' 감정도 비슷하게 RTLevel에서 더 높은 정확도와 재현율을 얻을 수 있었다. 결과 어휘를 분석한 결과, 이번에는 '감동'과 반대로 RTLevel에서는 '미안함'의 이유가 되는 MRPN 자질들(예: "기분/nc 상하/pa", "기분/nc 나쁘/pa")이 상위를 차지한 반면, TweetLevel에서는 '미안함'을 직접적으로 표현하는 RM 자질들(예: 죄송하/pa, 미안하/pa)이 상위를 차지했음을 알 수 있었다. 이는 트윗에서 '미안함'을 표현할 경우, 주로 리트윗과 상관없이 미안함을 표현하기 때문이다.

한편, TweetLevel와 TweetLevel+를 비교하면, TweetLevel이 TweetLevel+에 비해 다소 높은 성능을 내는 것 같으나 차이가 그리 크지는 않았다.

	기본감성어휘			RTLevel			TweetLevel			TweetLevel+		
	Prec	Recl	PERF	Prec	Recl	PERF	Prec	Recl	PERF	Prec	Recl	PERF
감동	0.31	0.14	0.31	**1.00**	**0.03**	**1.07**	0.50	0.03	0.50	0.50	0.07	0.50
감사	0.65	0.53	0.65	**0.74**	**0.53**	**1.62**	**0.74**	**0.53**	**1.62**	**0.76**	**0.53**	**1.62**
곤란	0.50	0.21	0.50	**1.00**	**0.05**	**1.10**	**1.00**	**0.05**	**1.10**	**1.00**	**0.05**	**1.10**
기대감	0.38	0.06	0.38	**0.71**	**0.03**	**1.06**	**0.71**	**0.03**	**1.06**	**0.71**	**0.03**	**1.06**
기쁨	0.22	0.10	0.22	0.22	0.04	0.22	**0.24**	**0.04**	**0.24**	**0.24**	**0.04**	**0.24**
두려움	0.43	0.16	0.43	**0.88**	**0.06**	**1.10**	0.86	0.05	1.09	0.86	0.05	1.09
미안함	0.44	0.57	0.44	**0.86**	**0.43**	**1.57**	0.80	0.29	1.42	0.80	0.29	1.42
반대	0.18	0.16	0.18	**1.00**	**0.01**	**1.03**	**1.00**	**0.01**	**1.03**	**1.00**	**0.01**	**1.03**
부러움	**0.79**	**0.40**	**1.53**	0.78	0.37	1.50	0.78	0.37	1.50	0.78	0.37	1.50
선의	0.52	0.23	0.52	**0.75**	**0.03**	**1.07**	0.83	0.03	1.05	0.83	0.03	1.05
수치심	0.29	0.20	0.29	**1.00**	**0.05**	**1.10**	**1.00**	**0.05**	**1.10**	**1.00**	**0.05**	**1.10**
슬픔	**0.87**	**0.14**	**1.24**	0.88	0.04	1.07	0.88	0.04	1.07	0.88	0.04	1.07
신뢰	0.03	0.02	0.03	**0.33**	**0.02**	**0.33**	**0.33**	**0.02**	**0.33**	**0.33**	**0.02**	**0.33**
실망	0.60	0.07	0.60	0.70	0.05	1.08	**0.71**	**0.06**	**1.10**	**0.71**	**0.06**	**1.10**
싫어함	0.55	0.14	0.55	0.78	0.03	1.06	**0.80**	**0.05**	**1.10**	0.77	0.05	1.09
안심	0.01	0.04	0.01	**1.00**	**0.04**	**1.08**	**1.00**	**0.04**	**1.08**	**1.00**	**0.04**	**1.08**
의심	-	0.00	-	0.11	0.01	0.11	**0.13**	**0.01**	**0.13**	0.07	0.01	0.07
자신감	0.00	0.00	0.00	0.01	0.29	0.01	0.01	0.14	0.01	**0.02**	**0.57**	**0.02**
좋아함	0.22	0.14	0.22	**1.00**	**0.02**	**1.04**	**1.00**	**0.02**	**1.04**	**1.00**	**0.02**	**1.04**
화남	0.54	0.05	0.54	**1.00**	**0.01**	**1.02**	**1.00**	**0.01**	**1.02**	**1.00**	**0.01**	**1.02**

표 4 RTLevel, TweetLevel, TweetLevel+ 성능

	PERF	자질들	예
감동	1.07	PRD+RM	RM:감탄스럽/pa RM:감격하/pa
감사	1.62	MRP+WD	WD:감사감사 MRP:감사합니당/nc
곤란	1.10	PRD+RM	RM:안_어색하/pa PRD0:느낌/nc 어색하/pa
기대감	1.06	MRPN+WD	MRP2:가슴/nc 설레/pv WD:희망찬
두려움	1.10	MRPN+RM	MRP2:불안하/pv 기분/nc RM:섬뜩하/pa
미안함	1.57	MRP+MRPN	MRP2:ㅠㅠ기분/nc 나쁘/pa MRP2:혹시/mag 기분/nc
반대	1.03	MRPN+RM	RM:반대하/pv RM:거랑같/nc
부러움	1.50	MRP+MRPN	MRP:탐나/pv MRP:질투/nc
선의	1.07	RM+WDN	RM:ㅎㅎ화이팅/nc WD2:힘내!!!
수치심	1.10	MRPN+RM	RM:치욕스럽/pa MRP2:부끄러움/nc 느끼/pv
슬픔	1.07	MRPN+RM	RM:우울증/nc MRP2:우울하/pv 날/nc
실망	1.10	MRP+MRPN	MRP:아쉽/pa MRP2:아쉬움/nc 남/pv
싫어함	1.10	MRP+WD	WD:싫어 MRP:지루하/pa WD:불쾌한
안심	1.08	MRPN+PRD	MRP2:한숨/nc 쉬/pv PRD0:스트레스/nc 날리/pv
좋아함	1.04	RM	RM:호감/nc RM:사랑/nc
화남	1.02	RM+WDN	RM:일투성이/nc RM:삭감하/nc

표 5 감성별 어휘 자질

5.2 감성별 어휘 분석

다음으로, 트윗 구분 방법에 관계없이 감성별로 가장 좋은 성능을 얻은 감성어휘들을 분석하였다. *PERF* 값이 1 이상인 감성들에 대해 상위를 차지하는 자질들의 예가 표 5에 나와 있다.

대부분의 감성에서 감성을 직접적으로 표현하는 자질들이 상위를 차지하고 있음을 알 수 있다. 하지만 '미안함'의 경우 "기분이 나빴다면"이나 "혹시 기분이 상하셨다면"과 같은 표현을 학습했음을 알

	통합 전			통합 후			향상 성능		
	Prec	Recl	PERF	Prec	Recl	PERF	Prec	Recl	PERF
감동	1.00	0.03	1.07	1.00	0.07	1.13	-	+0.04	+0.06
미안함	0.86	0.43	1.57	0.70	0.50	1.58	-0.16	+0.07	+0.01
선의	0.75	0.03	1.07	0.83	0.06	1.11	+0.08	+0.02	+0.04
슬픔	0.88	0.04	1.07	0.90	0.05	1.09	+0.03	+0.01	+0.02
싫어함	0.80	0.05	1.10	0.79	0.07	1.12	-0.01	+0.01	+0.02
좋아함	1.00	0.02	1.04	1.00	0.04	1.07	-	+0.02	+0.03
화남	1.00	0.01	1.02	0.80	0.02	1.03	-0.20	+0.00	+0.01

표 6 어휘 통합 전후 성능 비교

수 있고, '안심'의 경우 "한숨을 쉬다" 혹은 "스트레스를 날리다"와 같은 표현을 학습했음을 알 수 있다. 또한 '화남'의 경우에는 "일투성이"나 "임금을 삭감"한다는 표현을 학습하였다. 특정 상품이나 정책에 관한 트윗들을 대상으로 학습을 하였다면 그 도메인에 특화된 간접적인 감성 표현들(예를 들어, "너무 비싸다")이 많이 추출될 수 있겠지만, 본 실험에서는 일반적인 트윗을 사용해 학습하였기 때문에 감성을 직접적으로 나타내는 표현들의 비율이 상대적으로 높다. 그럼에도 불구하고 감성을 직접적으로 표현하는 기본감성어휘만 가지고 위와 같은 간접적인 감성 표현들을 학습해 낸 것을 통하여, 본 알고리즘에서 사용한 준지도학습의 유용성을 알 수 있다.

한편, '곤란'에서는 "안_어색하/pa"와 같은 자질이 상위로 잘못 추출되었다. 이는 학습데이터에 기본감성어휘를 적용할 때, 부정어 처리를 하지 않았기 때문에 발생한 문제이다. 기본감성어휘와 대량의 데이터에 대해서 정교한 부정어 처리를 할 경우에 성능이 향상될 것으로 예상된다.

정확도가 높은 감성어휘들을 통합하여 사용할 경우, 정확도를 크게 희생하지 않으면서 낮은 재현율을 보충할 수 있을 것이라 예상할 수 있다. 따라서 트윗 구분 방법과 자질 조합에 관계없이 감성별로 정확도가 70% 이상인 감성어휘들을 통합하여 성능을 측정하였고, 성능 향상을 보인 감성들이 표 6에 나와 있다. '감동', '미안함', '선의', '슬픔', '싫어함', '좋아함', '화남'에 대해 PERF가 향상되었다. 자질들을 통합함으로써 정확도가 떨어지는 경우가 많으나, 재현율이 올라가면서 전체적인 성능은 향상되었다. 이는 서비스 사용자들이 납득할 만한 정확도 범위 내에서 재현율이 상승되었음을 의미한다. '감동'과 '미안함'은 재현율이 큰 폭으로 상승하였고, '선의', '슬픔', '좋아함'은 정확도의 손실 없이 재현율이 향상되었다.

6. 결론

본 논문에서는 한국어로 된 트윗으로부터 20종류의 세부감성을 분석하기 위한 감성어휘를 구축하는 알고리즘을 제시하고 평가 결과를 제시하였다. 그래프를 기반으로 한 준지도학습을 이용하여, 적은 양의 기본감성어휘로부터 감성을 확장하는 알고리즘을 제시하였다.

도메인에 특화되지 않은 일반적인 트윗과, 직접적인 감성 표현들로 이루어진 소수의 기본감성어휘만 가지고 준지도학습을 통하여 감성어휘를 구축할 수 있었다. 트윗을 구분하는 방법에 따라 성능을 측정한 결과, 하나의 트윗 내에서 직접 작성한 부분과 리트윗 부분을 모두 하나의 트윗으로 간주한 경우와 모든 리트윗을 별개의 트윗으로 간주한 경우 성능 차이가 크지는 않았으나, 몇 감성에 대해서 후자의 경우 성능이 크게 증가하는 것을 확인하였다. 또한 학습된 감성어휘를 분석한 결과, 대부분 직접적인 감성표현이 상위를 차지하였으나 일부 감성들에서는 간접적인 감성표현들이 상위를 차지하는 것을 확인하였다.

본 논문은 세부감성 분석에 대한 기초 연구로서 발전 가능성이 크다. 먼저 본 논문에서 사용한 기본감성어휘는 주로 명사형이고 부정어 처리를 하지 않았기 때문에 감성을 태깅하는 데에 한계가 있다. 기본감성어휘를 좀 더 정교하게 처리함으로써 성능을 향상 시킬 수 있을 것이다. 또한 본 논문에서는 구축된 감성어휘의 일반성을 위해 도메인에 특화된 어떠한 정보나 데이터도 사용하지 않았다. 특정 도메인에 해당되는 학습데이터와 메타정보들을 활용한다면 더 높은 성능을 얻을 수 있을 것이다.

참고문헌

[1] Cortes, C. & Vapnik, V., Support-vector networks. *Machine learning*, 20(3), 273-297, 1995.

[1] Cortes, C. and Vapnik, V. N., Support-Vector Networks, *Machine Learning*, Springer, 1995.

[2] Maas, A. L., Ng, A. Y., and Potts, C., *M*ulti-Dimensional Sentiment Analysis with Learned Representations. 2011.

[3] Kim, S. M., Valitutti, A., and Calvo, R. A., Evaluation of unsupervised emotion models to textual affect recognition. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pp. 62-70., 2010.

[4] Velikovich, L., Blair-Goldensohn, S., Hannan, K. and McDonald, R., The viability of web-derived polarity lexicons. In *the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 777-785., 2010.

[5] Shawe-Taylor, C. A., and Schölkopf, S., The Support Vector Machine, 2000.

[6] 장문수, 심리학적 감정과 소셜 웹 자료를 이용한 감성의 실증적 분류. *한국지능시스템학회 논문지* 제22권 제5호, 2012.10, 563-569, 2012.

[7] Plutchik, R. and Kellerman, H., Emotion: Theory, Research, and Experience. Vol.5, Academic Press, 1990.

[8] Strapparava, C. and Valitutti, A., WordNet-Affect: an affective extension of WordNet. In *Proc. of 4th International Conference on Language Resources and Evaluation*, 2004.

[9] Pennebaker, J. W., Francis, M. E., and Booth, R. J., Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* (2001): 71, 2001.

[10] Esuli, A., and Sebastiani, F., Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, vol. 6, pp. 417-422. 2006.

[11] 윤애선, 임경업, 윤애선, 권혁철, 감정 온톨로지를 활용한 가사 기반의 음악 감정 추출, *한국지능정보시스템학회* 2010년 추계학술대회, pp.333-337, 2010.

[12] Kamvar, S. D. and Harris, J., We feel fine and searching the emotional web. *In Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 117-126. ACM, 2011.