

영어 트위터 감성 분석을 위한 SentiWordNet 활용 기법 비교

A Comparative Study on Using SentiWordNet for English Twitter Sentiment Analysis

강인수
In-Su Kang

경성대학교 공과대학 컴퓨터공학부
School of Computer Science & Engineering, College of Engineering, Kyungsoong University

요 약

트위터 감성 분석은 트윗글의 감성을 긍정과 부정으로 분류하는 작업이다. 이 연구에서는 SentiWordNet(SWN) 감성 사전에 기반한 트윗글 감성 분석을 다룬다. SWN은 전체 영어 단어에 대해 단어의 의미별로 긍정, 부정의 감성 강도를 저장해 둔 감성 사전이다. 기존 SWN 기반 감성 분석 연구들은 문서에 출현하는 각 용어의 감성을 SWN으로부터 결정된 다음 이를 바탕으로 문서 전체의 감성을 결정하였는데, 그 방법들이 매우 다양하다. 예를 들어, 한 용어의 감성 결정 시 해당 용어의 SWN 내 의미별 긍정, 부정 감성 강도 차이들의 평균을 계산하거나 긍정과 부정 각각의 감성 강도 평균 혹은 최대값을 구하기도 하며, 문서 전체의 감성을 결정하는 경우에도 문서 내 용어들의 감성 값들에 대해 평균 혹은 최대값을 취하기도 하였다. 또한 SWN 내 형용사, 동사, 명사, 부사의 품사 집합 전체 혹은 특정 부분집합에 대해 위의 감성 결정 작업을 적용하기도 한다. 이처럼 기존 연구에서는 SWN 기반의 다양한 감성 자질 추출 절차가 시도되고 있으나 이들 자질 추출 기법 전반에 대한 성능 비교 연구는 찾기 힘들다. 이 연구에서는 SWN을 트위터 감성 분석에 활용하는 다양한 방법들을 일반화하는 절차들을 소개하고 각 방법들의 성능 비교 및 분석 결과를 제시한다.

키워드 : 감성 분석, 트위터, SentiWordNet, 기계학습, 워드넷

Abstract

Twitter sentiment analysis is to classify a tweet (message) into positive and negative sentiment class. This study deals with SentiWordNet(SWN)-based twitter sentiment analysis. SWN is a sentiment dictionary in which each sense of an English word has a positive and negative sentimental strength. There has been a variety of SWN-based sentiment feature extraction methods which typically first determine the sentiment orientation (SO) of a term in a document and then decide SO of the document from such terms' SO values. For example, for SO of a term, some calculated the maximum or average of sentiment scores of its senses, and others computed the average of the difference of positive and negative sentiment scores. For SO of a document, many researchers employ the maximum or average of terms' SO values. In addition, the above procedure may be applied to the whole set (adjective, adverb, noun, and verb) of parts-of-speech or its subset. This work provides a comparative study on SWN-based sentiment feature extraction schemes with performance evaluation on a well-known twitter dataset.

Key Words : Sentiment Analysis, Twitter, SentiWordNet, WordNet

1. 서론

트위터(Twitter) 감성 분석(Sentiment Analysis)은 하나

접수일자: 2013년 5월 15 일

심사(수정)일자: 2013년 7월 4일

게재확정일자 : 2013년 7월 5일

† Corresponding author

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

의 트윗글(Tweet Message)을 입력으로 받아 그 트윗글을 통해 전달되는 감성의 극성(Polarity)을 긍정(Positive)과 부정(Negative)으로 분류하는 작업이다. 기존 의견 분석(Opinion Mining)이 시도되었던 영화평, 도서평, 상품평 등의 텍스트 유형과 달리, 트윗글은 최대 140자 길이 제약과 함께 이모티콘(Emoticon), 해쉬태그(Hash Tag), 유저태그(Username), 리트윗(RT), URL 등의 독특한 용어/구문적 특성을 갖고 있어, 새로운 복잡도를 갖는 감성 분석 대상으로 인식되고 있다.

영어 트위터 감성 분석을 위한 기존 연구에서는 n-gram 형식의 용어 자질을 사용한 기계학습법[1]으로 시작하여, 품사(POS, Part-of-speech) 자질[2,3], 감성 사전에서 추출된 자질[2,4], 그리고 전술한 이모티콘, 해쉬태그, 유저태그 등의 마이크로블로깅(microblogging) 자질을 활용하는 방법들

[4,5]이 시도되었다. 최근에는 트윗글에 출현하는 용어에 내재된 PERSON, COMPANY, PRODUCT 등의 개념 자질들을 활용하여 감성 분석의 성능 향상이 시도되고 있다[6].

이 중 감성 사전은 형용사, 부사와 같은 감성 용어들의 긍정 및 부정 감성 강도를 저장해 둔 것으로 감성 분석의 핵심 언어 자원 중 하나이다. 현존하는 대부분의 감성 사전들은 형용사 중심의 어휘들에 대해 의미 수준이 아닌 단어 수준의 감성 강도를 수작업으로 부여하여 구축된다. 기존 감성 사전과 달리 SentiWordNet¹⁾(이후 SWN으로 줄여 표기함)[7]은 영어 전체 어휘에 대해 각 단어의 의미(Word Sense)별로 긍정, 부정의 감성 강도 값이 기계학습 기법을 통해 부여되어 있어, 감성 분석기의 범용성, 도메인 확장성, 의미 수준 고급 처리 등의 측면에서 장점이 많다.

기존 SWN 기반 감성 분석 연구들은 대부분 입력 문서 내 각 용어의 감성 자질을 추출한 다음, 이를 바탕으로 문서 전체의 감성 자질을 추출하여 기계학습 분류기에 적용하였다. 용어 감성 자질의 경우, 용어의 여러 의미들이 갖는 감성 강도들의 평균[8]을 긍정과 부정 각각에 대해 구하거나, 긍정과 부정 감성 강도의 차이들의 평균[9]을 구하여 사용하였다. 문서 감성 자질의 경우, 용어 감성 자질들의 평균[8,10,11], 최대값[12], 총합[13] 등을 계산하여 추출하였다. 또한 SWN에서 각 단어의 제일의미(The first sense)만을 사용하거나[9], 형용사, 동사, 명사와 같은 특정 품사 집합[8]에 대해서만 전술한 감성 자질 추출 절차가 적용되기도 하였다.

이처럼 기존 연구에서는 SWN 기반의 다양한 감성 자질 추출 절차가 시도되고 있다. 그러나 이들 자질 추출 기법 전반에 대한 성능 비교 연구는 찾기 힘들다. 이 연구에서는 SWN을 트위터 감성 분석에 활용하는 다양한 방법을 일반화된 절차로 제시하고, 성능 비교 및 분석 결과를 기술한다. 실험에서는 총 720가지 SWN 기반 감성 자질 추출법에 대해 SVM 기계학습을 적용한다.

감성 분석에 SWN을 사용하는 방법으로, 크게 어의중의 성해소(WSD, Word Sense Disambiguation)를 적용하는 방법과 그렇지 않은 방법으로 나눌 수 있다. 그러나, 트윗글과 같은 단문의 비정형 텍스트에 대해 WSD를 적용할 경우, 성능 저하 우려가 있어 기존 연구에서는 후자의 방법이 주로 시도되었다. 이 연구에서도, 트윗글에 WSD를 적용하지 않으면서, 트윗글에 출현한 각 용어의 SWN 내에서의 의미별 감성 분포로부터 해당 용어의 감성 자질들을 추출하여 기계학습 기반의 감성 분석을 수행하는 접근법을 취한다.

논문의 구성은 다음과 같다. 2장에서 배경 지식으로 SentiWordNet에 대한 소개를 다루고, 3장에서 관련 연구를 기술한다. 4장에서는 SWN 기반 감성 자질 추출 기법에 대해서 상술한다. 5장에서는 실험과 결과를 기술하고 6장에서 결론을 맺는다.

2. SentiWordNet

SentiWordNet(SWN)은 감성 분류와 오피니언 마이닝 응용을 위해 WordNet에 의미 수준의 긍정, 부정 감성 강도 값을 부착하여 구축된 어휘 사전으로 2010년 기준으로 전 세계 300여개 이상의 연구 그룹에서 활용되고 있다[7]. SWN은 2006년 첫 버전이 알려진 이후 현재 WordNet 3.0에 대해 작업된 SentiWordNet 3.0이 공개된 상태이다. 이

논문에서는 SentiWordNet 3.0을 사용한다. WordNet 3.0은 명사, 동사, 형용사, 부사의 네 개 품사에 속하는 총 147,278개 영어 단어들을 117,659개 동의어집합(synset, synonym set)들로 수작업 분류해 놓은 어휘 사전이다²⁾. 즉 SWN은 WordNet의 각 동의어집합에 대해 긍정, 부정, 중립 감성의 값을 자동 부착해 둔 것이다. 또한 긍정, 부정, 중립 감성의 총합은 1이 되도록 정규화되어 있다. 표 1은 단어 good에 대한 SWN의 간략화된 예를 보인 것으로 영어 단어 good의 품사별로 각 의미의 긍정(Positive), 부정(Negative), 중립(Objective) 감성 강도 값을 보여준다. 예를 들어 skillful의 의미를 갖는 형용사 good의 경우 긍정, 부정, 중립 감성 값이 각각 0.625, 0, 0.375임을 보인다.

표 1. SentiWordNet 예제
Table 1. An example of SentiWordNet

POS	Sense	Pos	Neg	Obj	Example
Adjective	good#1	0.75	0	0.25	good news
	good#2	0.5	0	0.5	a good(secure) investment
	good#3	0.625	0	0.375	a good(skillful) mechanic
Noun	good#1	0.5	0	0.5	for your own good
	good#2	0	0	1	expensive goods
Adverb	good#1	0.375	0	0.625	The baby walks pretty good

3. 관련 연구

텍스트 감성 분석 방법은 긍정 및 부정 감성 분류가 기록된 학습 데이터 집합에 기계학습을 적용하여 감성 분류기를 학습하는 교사(Supervised) 방식과, 텍스트 출현 용어들의 감성 강도들을 미리 준비된 감성 사전 등으로부터 구한 다음 텍스트 전체의 감성을 결정하는 비교사(Unsupervised) 방식으로 나뉜다[14]. 이 논문의 연구는 두 방식을 결합하여 트윗글 감성 분류를 시도한 것이다. 일반적인 감성 분류의 기존 연구에 대한 상세하고 체계적인 검토를 위해서는 Liu[15]와 Pang[16]의 논문을 참조하기 바라며, 이 절에서는 감성 분석의 기존 연구들을, SWN 기반 대표적 시도들을 중심으로 개괄한다.

Taboada 등은 상품/영화/도서평 등 다양한 의견 텍스트의 감성 분석을 위한 감성 사전 기반 방법을 시도하였다[9]. 그들은 SWN으로부터 어휘의 감성 값을 표현하기 위해 어휘의 각 의미가 갖는 긍정 감성 값과 부정 감성 값의 차이를 이용하였다. 어휘의 제일의미(first sense)만 사용하는 경우 긍정과 부정 감성의 차이를 사용하였고, 어휘의 모든 의미를 사용하는 경우 각 의미의 긍정과 부정 감성 값의 차이들의 평균을 계산하여 활용하였다.

Denecke 등은 의학 분야 웹 블로그 문서의 유형을 affective와 informative로 분류하는 기계학습 자질로서 SWN에서 얻어지는 단어의 감성 정보를 사용하였다[8]. 구체적으로 그들은 블로그 문서에 출현하는 각 단어의 전체 의미에 대해 긍정/부정/중립 감성 강도 각각의 세 가지 평균을 구한 다음, 문서 내 같은 품사(동사, 형용사, 명사)를 갖는 각 단어의 전술한 평균들의 평균을 긍정/부정/중립 감성별로 계산하여 문서의 감성 자질로 이용하였다. Hamouna 등도 상품평의 감성 분석을 위해 이와 유사한 방법을 사용하였다[10].

1) <http://sentiwordnet.isti.cnr.it/>

2) <http://wordnet.princeton.edu/man/wnstats.7WN.html>

Ohana 등은 리뷰 문서 감성 분석을 위한 SVM 자료로서 문서 내 같은 품사(형용사, 부사, 동사)를 갖는 단어의 의미들의 감성 강도의 총합을 긍정 및 부정 감성 각각에 대해 계산하여 추출하였으며, 이후 그러한 긍정과 부정 감성 총합의 비율도 품사별로 계산하여 활용하였다[13].

Dehkharghani 등은 호텔 및 영화 리뷰 문서의 감성 분석을 위해, 문서 내 각 단어의 감성 값을 긍정과 부정 감성의 차이에 기반한 방법으로 계산하고, 여러 단어들의 이러한 감성 값의 평균을 기계학습 분류기의 자질로 활용하였다[11].

Lee 등은 질의 주제와 관련된 블로그 글의 감성 분석을 위한 용어 감성 자질 표현을 위해 SWN을 활용하였다. 그들은 한 단어의 모든 의미가 갖는 긍정, 부정 감성 강도들 중 최대값을 해당 단어의 감성 값으로 사용하였다[12].

4. SentiWordNet 기반 감성 자질 추출

SWN 기반 감성 자질 추출은, 입력으로 주어지는 하나의 트윗글로부터 그 트윗글에 실린 감성을 나타내는 하나 이상의 감성 자질(들)을 SWN으로부터 추출하는 것이다. 그 절차는, 먼저 트윗글을 구성하는 각 용어의 감성 정보를 SWN을 참조하여 만들고, 다음으로 용어들의 감성 정보들로부터 트윗글 전체의 감성 정보를 자질화하는 것이다. 이 절에서는 총 세 단어 t_1, t_2, t_3 로 구성된 하나의 트윗글 $s=[t_1, t_2, t_3]$ 와 이 세 단어의 SWN 예제 데이터(그림 1 참조)를 사용하여, 트윗글로부터 감성 자질을 추출하는 다양한 기법들을 설명한다.

그림 1은 세 단어 t_1, t_2, t_3 의 품사(형용사, 명사, 동사, 부사)별 의미별 긍정(Pos) 및 부정(Neg) 감성 강도를 표시한 것이다. 예를 들어 단어 t_2 의 경우 형용사, 명사, 동사로의 사용이 가능하고, 형용사의 경우 총 세 개 의미가 있으며 그 중 세 번째 의미인 $t_2\#3$ 의 경우 긍정, 부정 감성 강도 값이 각각 0.2, 0.1임을 알 수 있다.

t_1			t_2			t_3		
Adj	Pos	Neg	Adj	Pos	Neg			
$t_1\#1$	0.15	0	$t_2\#1$	0.6	0.3			
$t_1\#2$	0	0.25	$t_2\#2$	0.3	0			
			$t_2\#3$	0.2	0.1			
Noun	Pos	Neg	Noun	Pos	Neg	Noun	Pos	Neg
$t_1\#1$	0.1	0.1	$t_2\#1$	0	0.1	$t_3\#1$	0	0.05
			$t_2\#2$	0.4	0	$t_3\#2$	0.75	0.1
			Verb	Pos	Neg			
			$t_2\#1$	0.2	0			
Adv	Pos	Neg						
$t_1\#1$	0.1	0.25						

그림 1. SentiWordNet 예제 데이터
Fig. 1. An example of SentiWordNet

이 연구에서는 트윗글 감성 자질 추출 절차로 의미 수준에서의 긍정-부정 극성 차 적용 여부와 품사 구분 자질 생성 여부에 따라 다음의 네 가지로 구분하였다. 각 절차는 용어 수준의 감성을 결정하는 함수 F_t 와 문장(트윗글) 수준

의 감성을 결정하는 함수 F_s 를 순차 적용하는 과정을 거친다. SPD, PoS는 각각 Sentiment Polarity Difference, Part-of-Speech의 약자이다.

- SPDn-PoS_n: 의미별 극성차 무시, 품사 무관 감성 자질 추출 절차
- SPDy-PoS_n: 의미별 극성차 기반, 품사 무관 감성 자질 추출 절차
- SPDn-PoS_y: 의미별 극성차 무시, 품사 구분 감성 자질 추출 절차
- SPDy-PoS_y: 의미별 극성차 기반, 품사 구분 감성 자질 추출 절차

그림 2는 의미 수준의 극성 차이를 사용하지 않고 만들어진 감성 자질을 품사 구분 없이 생성하는 절차 SPDn-PoS_n를, 입력 트윗글 $s=[t_1, t_2, t_3]$ 에 대한 적용 예 ($F_t=\max, F_s=\max$ 가 사용된다고 가정)와 함께 보인 것이다. 먼저 $F_t=\max$ 는 s 의 각 용어 t_i 에 대해, SWN 내에서 t_i 가 취할 수 있는 모든 품사의 모든 의미들의 긍정 및 부정 감성 강도들의 최대값을 선택한다. 예를 들어, t_1 에 대한 $F_t=\max$ 적용 결과인 Pos=0.15, Neg=0.25는 SWN 내에서 t_1 의 가능한 모든 품사들(형용사, 명사, 부사)의 의미들의 긍정 및 부정 감성 강도 최대값들에 해당한다(그림 1 참조).

다음으로 $F_s=\max$ 는 트윗글 s 전체의 감성 대표값으로, s 를 구성하는 용어들이 갖는 긍정 및 부정 감성 대표값들의 최대값을 선택한다. 즉 Pos=0.75는 각 용어의 긍정(Pos) 감성 대표값 0.15, 0.6, 0.75의 최대값이다. 이렇게 얻어진 Pos=0.75, Neg=0.3은 기계학습 감성 분류기의 입력 인스턴스가 된다. 또한 Pos=0.75, Neg=0.3의 차를 계산한 Pos-Neg=0.45를 기계학습 분류기의 입력으로 사용할 수도 있다.

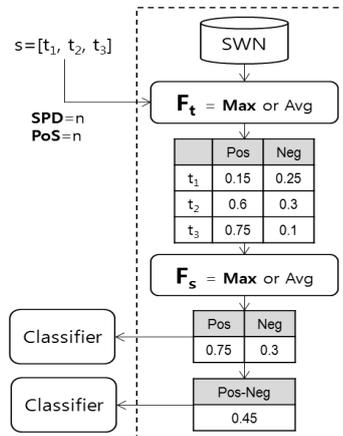


그림 2. SPDn-PoS_n: 의미 극성차 무시 품사 무관 트윗글 감성 자질 추출 절차
Fig. 2. SPDn-PoS_n: a feature-extraction process not using SPD, and ignoring POS

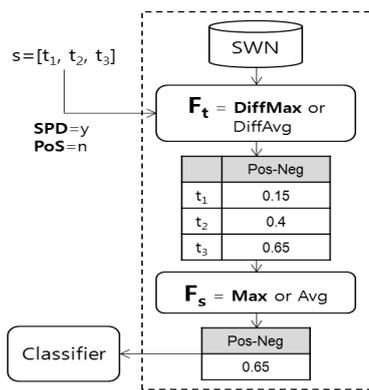


그림 3. SPDy-PoS: 의미 극성차 기반 품사 무관 트윗글 감성 자질 추출 절차
 Fig. 3. SPDy-PoS: a feature-extraction process using SPD, and ignoring POS

그림 3은 의미 수준의 극성 차에 기반한 감성 자질을 품사 구분 없이 생성하는 절차 SPDy-PoS를, 입력 트윗글 $s=[t_1, t_2, t_3]$ 에 대한 적용 예($F_t=diffmax, F_s=max$ 가 사용된다고 가정)와 함께 보인 것이다. 먼저 $F_t=diffmax$ 는 s 의 각 용어 t_i 에 대해, SWN 내에서 t_i 가 취할 수 있는 모든 품사의 각 의미별 긍정 부정 감성 강도 차이의 최대값을 취한다. 예를 들어, t_1 에 대한 $F_t=diffmax$ 적용 결과인 Pos-Neg=0.15는, SWN 내 t_1 의 모든 품사의 의미별 긍정 부정 감성 차이 값들 0.15, -0.25, 0, -0.15 중 최대값 0.15를 취한 것이다. 다음으로 $F_s=max$ 는 앞에서 얻어진 s 의 각 용어 감성 대표값들 0.15, 0.4, 0.65 중 그 최대값 0.65를 트윗글 s 전체의 감성 대표값으로 선택한다. 이렇게 얻어진 0.65는 기계학습 감성 분류기로의 단일 자질을 갖는 입력 인스턴스가 된다.

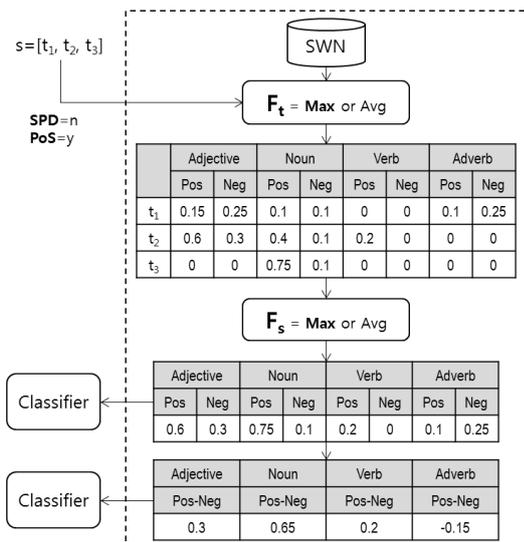


그림 4. SPDn-PoS: 의미 극성차 무시 품사 구분 트윗글 감성 자질 추출 절차
 Fig. 4. SPDn-PoS: a feature-extraction process not using SPD, and differentiating POS

그림 4는 의미 수준의 극성 차이를 사용하지 않고 만들

어진 감성 자질을 품사별로 구별하여 생성하는 절차 SPDn-PoS를, 입력 트윗글 $s=[t_1, t_2, t_3]$ 에 대한 적용 예 ($F_t=max, F_s=max$ 가 사용된다고 가정)와 함께 보인 것이다. 그림 2와의 차이는 긍정 부정 감성 자질들을 품사별로 구분하여 생성한다는 것이다. 예를 들어 $F_t=max$ 는 s 의 용어 t_1 에 대해, t_1 이 취할 수 있는 형용사, 명사, 부사 각 품사의 의미들의 긍정 및 부정 감성 강도들의 최대값을 선택한다. 즉 t_1 이 형용사인 경우 긍정 및 부정 감성 최대값은 각각 0.15, 0.25가 된다. $F_t=max$ 의 적용 과정 또한 품사별 구분을 제외하고는 그림 2와 동일하다.

그림 5은 의미 수준의 극성 차에 기반한 감성 자질을 품사별로 구분하여 생성하는 절차 SPDy-PoS를, 입력 트윗글 $s=[t_1, t_2, t_3]$ 에 대한 적용 예($F_t=diffmax, F_s=max$ 가 사용된다고 가정)와 함께 보인 것이다. 그림 3과의 차이는 긍정 부정의 감성 차이 자질들을 품사별로 구분하여 생성한다는 것이다. 예를 들어 $F_t=diffmax$ 는 s 의 용어 t_i 에 대해, t_i 이 취할 수 있는 형용사, 명사, 부사 각 품사의 의미별 긍정 부정 감성 강도 차이들의 최대값을 선택한다. 즉 t_i 이 형용사인 경우 의미별 긍정 부정 감성 차이값들 0.15, -0.25의 최대값은 0.15가 된다. $F_t=max$ 의 적용 과정 또한 품사별 구분을 제외하고는 그림 3과 동일하다.

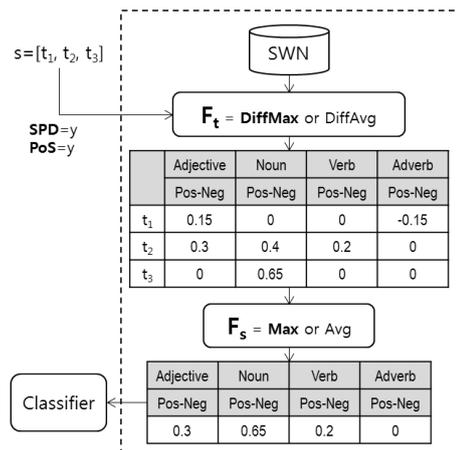


그림 5. SPDy-PoS: 의미 극성차 기반 품사 구분 트윗글 감성 자질 추출 절차
 Fig. 5. SPDy-PoS: a feature-extraction process using SPD, and differentiating POS

5. 실험

5.1 평가 집합, 평가 방법

실험에서는 4장에서 소개된 SWN 기반 감성 자질 추출 기법이 트윗글 감성 분석 성능에 미치는 영향을 평가한다. 이를 위해 가장 잘 알려진 트윗글 감성 분석 데이터 집합[1]인 스탠포드 트위터 데이터셋³⁾(Stanford Twitter Dataset)을 사용한다. 이 데이터셋은 회사명, 인명, 제품명, 영화명 등의 질의로 트위터를 검색하여 얻어진 총 359개 트윗글을 긍정 및 부정 감성을 갖는 트윗글로 수작업 분류한 것이며, 긍정 트윗글 182개, 부정 트윗글 177개로 구성되어 있다.

감성 분류 방법으로는 자질 벡터에 유클리디언 길이 정규화를 적용한 후 SVM⁴⁾ 기계학습을 사용하였다. 평가 척

3) <http://help.sentiment140.com/for-students>

도로는 전술한 데이터셋에 대해 3-fold 교차 검증(Cross validation)을 100회 반복 적용한 평균 성능을 긍정 및 부정 분류 정확률(Precision), 재현율(Recall), F1, 그리고 Accuracy에 대해 적용하여 사용한다.

5.2 순수 SWN 감성 자질 기반 감성 분류

먼저 4장의 SWN 기반 감성 자질들로만 학습한 경우의 감성 분류 성능은 표 2와 같다. 4장의 네 가지 각 방법들은 차례로 총 8, 4, 8, 4개의 방법들로 세분되므로 기본적으로 총 24가지 자질 추출 방법이 존재한다. 표 2는 24가지 각 방법의 성능을 Accuracy 순으로 상위 7개, 하위 3개를 제시한 것으로, 열 2,3은 SWN 전체를 사용한 경우의 성능이고, 열 4,5는 SWN에서 일순위의미(First Sense)만을 사용한 경우의 성능을 보인 것이다. 표에서 PoSy, PoSn는 품사 구분 자질의 추출 유무이고, avg-max, diffavg-avg 등은 F₁, F₂에 적용된 함수들을 차례로 표시한 것이다. PN과 DI는 SPDn-PoSn, SPDn-PoSy에서 긍정과 부정 감성 대표 값을 각각 자질로 추출(PN)하거나 그 차이를 자질로 추출(DI)한 방법을 의미한다.

표 2. 순수 SWN 감성 자질 기반 감성 분류 성능 (전체 품사 사용)

Table 2. Performance of sentiment analysis based on pure SWN features using all POSs

Rank	SWN_All	Acc	SWN_FirstSense	Acc
1	avg-max-PoSy-DI	0.6897	max-max-PoSy-DI	0.7057
2	diffavg-avg-PoSy	0.6886	avg-max-PoSy-DI	0.7057
3	avg-avg-PoSy-DI	0.6864	avg-max-PoSn-DI	0.705
4	avg-max-PoSy-PN	0.6813	avg-max-PoSn-PN	0.7034
5	avg-avg-PoSy-PN	0.6782	avg-avg-PoSy-DI	0.6985
6	max-max-PoSy-DI	0.6771	max-avg-PoSy-DI	0.6979
7	diffmax-avg-PoSy	0.6732	diffmax-avg-PoSy	0.6977
22	diffmax-avg-PoSn	0.5623	diffmax-max-PoSn	0.5849
23	avg-avg-PoSn-DI	0.4797	diffavg-avg-PoSn	0.4784
24	diffavg-avg-PoSn	0.4785	avg-avg-PoSn-DI	0.4765

표 2를 통해 서로 다른 SWN 감성 자질 추출 방법들은 감성 분류 성능에서 20% 이상 차이가 벌어짐을 알 수 있으며, 품사별 감성 자질을 구별하는 것(PoSy)과 긍정과 부정 감성 강도의 차이(diffavg, diffmax, DI)를 자질로 추출하는 것이 효과적임을 보인다. 또한 용어의 전체 의미(SWN_All) 대신 대표의미(SWN_FirstSense)의 감성 수치만 사용하는 것이 상대적으로 높은 감성 분류 성능을 얻을 수 있음을 보인다.

표 2는 SWN 내 형용사, 부사, 동사, 명사의 네 가지 전체 품사를 모두 사용한 경우의 성능이다. 그러나 사전 기반의 감성 분류 연구의 경우 형용사, 부사 중심의 감성 사전을 주로 활용한다는 점을 감안할 때 SWN 기반의 감성 분류를 위해 품사별 자질 영향력의 차이를 살펴볼 필요가 있다. 표 3은 이를 반영한 실험 결과로 형용사, 부사, 동사, 명사의 네 품사의 가능한 조합 15가지에 대해 표 2에서의 실험을 반복한 것이다. 표 2는 모든 품사가 사용된 한 가지 조합(arvn)에 대한 실험이다. 모든 품사 조합이 추가되는 경우 총 720(=24×2×15)가지 자질 추출 방법들에 대한 성능 비교가 되며 표 3은 그 중 상위 10개 방법들을 SWN_All과 SWN_FirstSense에 대해 구분하여 제시한 것이다. 표에서

arvn의 a, r, v, n은 각각 adjective, adverb, verb, noun의 약자이다.

표 3. 순수 SWN 감성 자질 기반 품사 조합별 감성 분류 성능

Table 3. Performance of sentiment analysis based on pure SWN features using different POS combinations

Rank	SWN_All	Acc	SWN_FirstSense	Acc
1	avg-max-PoSn-arv-DI	0.7033	max-max-PoSy-arvn-DI	0.7057
2	avg-max-PoSn-arv-PN	0.6916	avg-max-PoSy-arvn-DI	0.7057
3	avg-max-PoSy-arv-PN	0.6902	avg-max-PoSn-arvn-DI	0.705
4	avg-max-PoSy-arvn-DI	0.6897	avg-max-PoSn-arvn-PN	0.7034
5	avg-max-PoSn-av-DI	0.6887	avg-max-PoSy-avn-DI	0.7023
6	diffavg-avg-PoSy-arvn	0.6886	max-max-PoSy-avn-DI	0.7022
7	avg-max-PoSy-avn-DI	0.6876	max-max-PoSn-av-DI	0.7006
8	avg-avg-PoSy-avn-DI	0.6864	avg-max-PoSy-arv-DI	0.6998
9	avg-avg-PoSy-arvn-DI	0.6864	max-max-PoSy-arv-DI	0.6995
10	diffavg-avg-PoSy-avn	0.685	avg-max-PoSn-av-DI	0.6994

표 3을 통해서도 표 2에서의 같은 분석 결과를 재확인할 수 있다. 특히 품사 조합과 관련하여 감성 분석에서 형용사, 동사의 상대적 필요성과 함께 다양한 품사의 사용이 감성 분석에 도움이 됨을 보인다. 이는 단문의 트윗글이 갖는 적은 수의 용어 집합으로부터 최대한 다양한 감성 자질을 추출하는데, 다양한 품사 집합의 사용이 긍정적 영향을 미쳤기 때문인 것으로 판단된다.

5.3 SWN 감성 자질 기반 감성 분류

트윗글 감성 분류의 기존 연구들은 트윗글 내 용어 자질을 필수 자질 집합으로 사용한다. 이 절에서는 SWN 감성 자질이 기존 연구에서의 용어 자질과 결합된 경우 트윗글의 감성 분류 성능에 미치는 영향을 실험적으로 제시하고 분석한다.

표 4의 BOW는 1-gram 용어 자질만을 사용한 감성 분류 방법이다. 이는 Go 등의 기존 연구[1]를 따라, 트윗글에서 @ 기호로 시작하는 사용자명(예: @angel)을 "USERNAME"으로, http://로 시작하는 웹 주소를 "URL"로 각각 통일시키고 "loooooove"에서와 같이 3회 이상 반복되는 문자들을 "loove"와 같이 2회 반복으로 교체시키는 전처리를 거친 다음, 1-gram 용어 자질들을 추출하고 SVM 학습을 거쳐 감성 분류기를 만든 것이다. 이렇게 얻어진 1-gram 용어 자질의 가중치 표현에서 이진 표현을 사용한 것이 BOW (binary)이고 TF-IDF 가중치 표현을 사용한 것이 BOW (tf-idf)이다. 또한 전술한 것처럼 가중치 벡터는 SVM 학습 전에 유클리디언 길이 정규화를 적용하였다.

표 4. SWN 감성 자질 기반 감성 분류 성능

Table 4. Performance of sentiment analysis

Method	Positive Sentiment Class			Negative Sentiment Class			Total	
	Pre	Rec	F1	Pre	Rec	F1	F1	Acc
BOW (binary)	77.5	83.5	80.2	81.6	75.0	77.9	79.0	79.2
BOW (tf-idf)	80.3	82.8	81.3	81.7	79.0	80.1	80.7	80.8
BOW+SWN	82.9	83.6	83.0	82.9	82.3	82.4	82.7	82.7

4) <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

TF-IDF 가중치 표현은 이진 표현에 비해 감성 분류 성능이 더 좋았다. BOW+SWN은 TF-IDF 가중치 방식의 용어 자질 벡터와 5.2절에 사용된 SWN 감성 자질 벡터를 결합하여 SVM 학습을 통해 감성 분류기로 만든 것이다. 표를 통해 SWN 감성 자질의 추가는 기존 용어 기반 감성 분류 성능을 전체 F1 및 Accuracy 기준으로 약 2% 향상시킴을 알 수 있다. 이는 SWN에 내재된 단어 의미별 긍정 부정 감성 강도로부터 얻어진 단어의 사전(prior) 감성 자질 값들이, 트윗글 학습 데이터로부터 추출되는 용어 기반의 자질들을 보완하는 효과가 있음을 의미한다.

표 5에서는 4장의 SWN 감성 자질 표현법 중 용어 자질과의 결합에서 높은 성능을 보인 상위 10개 방법을 보인 것이다. 그 결과 다양한 SWN 감성 자질 표현법 중 BOW 자질과의 결합에서 전체 감성 분류 성능 향상에 도움이 되는 것들은 대부분 품사별 구분 자질을 사용하지 않으면서 (PoSn), diffmax, DI 등의 절차를 통해 긍정 부정 감성 강도의 차이를 활용하는 방법들이었다.

그러나 흥미롭게도 순수 SWN 자질만을 사용한 경우와 SWN 자질을 BOW 자질과 결합하는 두 경우 각각에 있어 상위 감성 분류 성능을 보인 SWN 감성 자질 표현법들이 서로 일치하지 않는다(표 3, 표 5 비교). 이러한 상충은 어휘(용어)와 감성이라는 서로 다른 유형의 자질 집합들이 결합되는 과정에서 발생한 것으로 판단된다. 이후 기술은 이에 대한 분석을 포함하고 있다.

표 5. 서로 다른 SWN 감성 자질 표현법들에 대한 BOW+SWN 감성 분류 성능

Table 5. Performance of sentiment analysis based on BOW+SWN features

Rank	DB_All	Acc	DB_FirstSense	Acc
1	max-avg-PoSn-arn-DI	0.8274	diffmax-avg-PoSn-arn	0.8173
2	diffmax-avg-PoSn-arn	0.8244	diffmax-avg-PoSn-arnv	0.8172
3	max-avg-PoSn-an-PN	0.8236	max-avg-PoSn-arnv-DI	0.8169
4	diffmax-avg-PoSn-arnv	0.8213	diffmax-avg-PoSn-arv	0.8168
5	max-avg-PoSn-arn-PN	0.8207	max-avg-PoSn-an-DI	0.8165
6	diffmax-avg-PoSn-an	0.8207	max-avg-PoSn-arn-DI	0.8163
7	max-avg-PoSn-avn-DI	0.8206	diffmax-avg-PoSn-ar	0.8163
8	max-avg-PoSn-arnv-DI	0.8202	max-avg-PoSn-arv-DI	0.816
9	avg-avg-PoSn-arnv-DI	0.82	max-avg-PoSn-ar-DI	0.8157
10	max-avg-PoSn-an-DI	0.8193	diffmax-avg-PoSn-an	0.8155

그림 6은 BOW (tf-idf) 감성 분류기의 Accuracy 성능을 기준으로 720가지 각 BOW+SWN 결합 분류기의 성능을 비교한 것으로, 대략 상위 100순위 내의 SWN 감성 자질 표현법들이 BOW+SWN의 결합을 통해 어휘 자질 기반 감성 분류 성능을 향상시켰음을 보여준다. 그림 7은 총 720개 각 SWN 감성 자질 표현법에 대해, BOW+SWN 감성 분류 성능 순위와 해당 SWN 자질 표현법의 감성 자질 수를 도식화한 것으로, BOW 자질과의 결합을 통해 감성 분류에 도움이 되는 대략 100순위 이내 SWN 자질 표현들은 그 자질의 수가 대부분 1개 혹은 2개임을 알 수 있다(상위 100순위 내 평균 1.1개). 그러나 그림 8에 보인 것처럼 순수 SWN 자질만을 사용한 감성 분류의 경우에는 상위 성능의 SWN 자질 표현법들일수록 자질의 수가 상대적으로 많다는 것을 알 수 있다(상위 100순위 내 평균 3.2개).

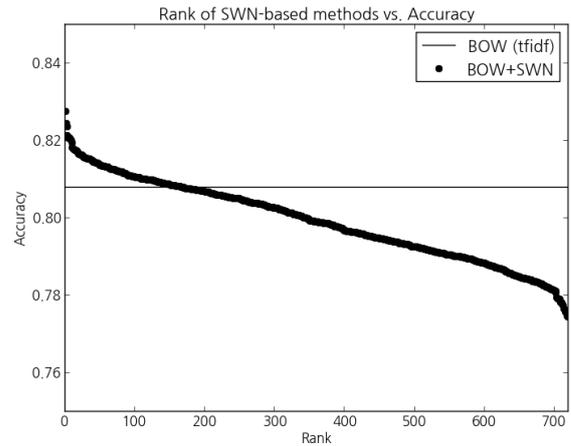


그림 6. BOW 감성 분류 성능 대비 720개 BOW+SWN 감성 분류기 성능 비교

Fig. 6. Relationship between ranks and accuracies for 720 sentiment analysis methods based on BOW+SWN features

그 이유는 다음과 같이 생각된다. 순수 SWN 자질 기반 감성 분류기의 경우 감성 자질을 가능한 한 다양화하는 것이 트윗글 극성 판단에 도움이 되겠지만, BOW+SWN 감성 분류의 경우에는 기본이 되는 어휘 자질 집합에서 결여된 소량의 감성 판단 자질들만을 SWN으로 보완하는 것이 효과적이기 때문일 것이다. 이와 관련하여 향후 BOW와 SWN 자질들의 적절한 결합 방법을 탐구하고 전술한 SWN 자질 표현에서의 자질 개수의 영향을 좀 더 고찰할 필요가 있다.

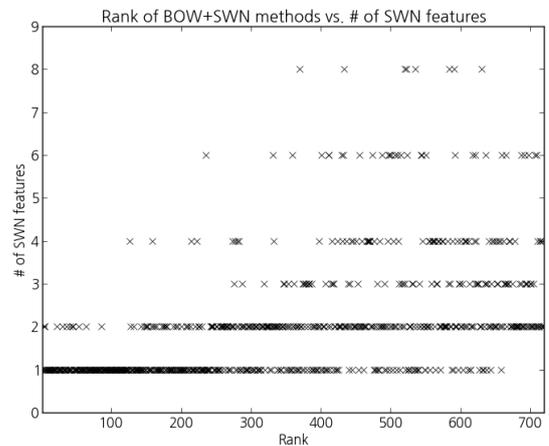


그림 7. BOW+SWN 감성 분류 성능 순위와 SWN 자질 수의 관계

Fig. 7. Relationship between ranks and the number of SWN features for 720 sentiment analysis methods based on BOW+SWN features

References

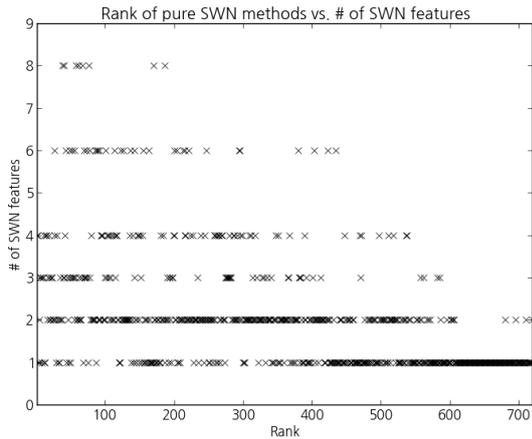


그림 8. 순수 SWN 자질 기반 감성 분류 성능 순위와 SWN 자질 수의 관계

Fig. 8. Relationship between ranks and the number of SWN features of 720 sentiment analysis methods based on pure-SWN features

6. 결론

이 연구는 리뷰 문서의 감성 분석을 위해 기존에 시도되었던 SentiWordNet 기반의 감성 자질 추출 방법들이 트윗글의 감성 분석에 미치는 효과들을 실험적으로 비교하고 분석하였다. 그 결과 SWN 감성 자질 중 긍정과 부정의 감성 강도 차이를 정량화하는 방법이, 순수 SWN 자질 기반의 감성 분류와 SWN 자질이 BOW 어휘자질과 결합된 감성 분류의 두 경우 모두에 있어 트윗글 감성 분석에 효과적이었다. 추가적으로 순수 SWN 자질만을 사용하는 경우는 품사별 자질을 구분하는 방법이 좋았고, BOW+SWN의 결합에 있어서는 품사 구분을 무시하는 방법이 감성 분류에 도움이 되었다.

한편 논문에서 시도된 총 720개 감성 자질 추출 절차들 중 약 20% 정도만 BOW와의 결합을 통해 BOW 자질만 사용한 경우의 감성 분류 성능을 향상시켰으며, 순수 SWN 자질만 사용한 경우에도 그 성능의 범위가 최저 47%에서 최고 70%로 상당한 격차가 발생함을 확인하였다. 이는 SWN 기반 감성 분석에서 적절한 SWN 감성 자질의 선택이 주의 깊게 고려되어야 함을 의미한다.

실험에서는 BOW 어휘 자질과 SWN 감성 자질 같은 상이한 특성의 자질 집합을 결합하는 적절한 방법의 연구 필요성이 제기되었다. 이와 관련하여 향후 Gehler[17] 등의 이미지 처리 분야에서 그 효과가 보고된 MKL(Multiple Kernel Learning), Boosting 방법 등을 시도할 계획이다.

현재 연구에서 다룬 트윗 감성 분석 방법은 영어에 대한 것으로 이후 한국어 트윗글에서 재확인되고 개선될 필요가 있다. 이를 위해 영어 워드넷과 연결된 한국어 워드넷을 통해 한국어 어휘에 대한 영어 SentiWordNet 정보의 활용이 가능할 것이다. 또한 한국어 트윗의 경우 형태소 분석의 어려움에서 발생하는 BOW 자질 표현의 불충분함을 보완하기 위해 문자열 커널을 한국어에 적용한 음절 커널[18] 기반 분류 기법을 동시 고려할 필요가 있을 것이다.

- [1] A. Go, R. Bhayani, L. Huang, "Twitter Sentiment Classification using Distant Supervision," *CS224N Project Report*, Stanford, 2009.
- [2] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data," *Proceedings of the Workshop on Languages in Social Media (LSM)*, 2011.
- [3] A. Pak, P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, 2010.
- [4] L. Barbosa, J. Feng, "Robust Sentiment Detection on Twitter from Biased and Noisy Data," *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, 2010.
- [5] E. Kouloumpis, T. Wilson, J. Moore, "Twitter Sentiment Analysis: The Good the Bad and the OMG!," *Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM)*, 2011.
- [6] H. Saif, Y. He, H. Alani, "Semantic sentiment analysis of twitter," *Proceedings of the 11th International Conference on The Semantic Web (ISWC)*, 2012.
- [7] S. Baccianella, A. Esuli, F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, 2010.
- [8] K. Denecke, W. Nejdl, "How valuable is medical social media data? Content analysis of the medical web," *Information Sciences*, vol. 179, no. 12, pp. 1870-1880, 2009.
- [9] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, "Lexicon-Based Methods for Sentiment Analysis," *Computational Linguistics*, vol. 37, no. 2, pp. 267-307, 2011.
- [10] A. Hamouda, M. Rohaim, "Reviews Classification Using SentiWordNet Lexicon," *The Online Journal on Computer Science and Information Technology (OJCSIT)*, vol. 2, no. 1, pp. 120-123, 2011.
- [11] R. Dehkharghani, B. Yanikoglu, D. Tapucu, Y. Saygin, "Adaptation and Use of Subjectivity Lexicons for Domain Dependent Sentiment Classification," *IEEE 12th International Conference on Data Mining Workshops (ICDMW)*, 2012.
- [12] Y. Lee, S. Na, J. Kim, S. Nam, H. Jung, J. Lee, "KLE at TREC 2008 Blog Track: Blog Post and Feed Retrieval," *Proceedings of The Seventeenth Text REtrieval Conference (TREC)*, 2008.
- [13] B. Ohana, B. Tierney, "Sentiment classification of reviews using SentiWordNet," *Proceedings of the 9th IT&T Conference*, 2009.

- [14] R. Feldman, "Techniques and applications for sentiment analysis," *Communications of the ACM*, vol. 56, no. 4, pp. 82-89, 2013.
- [15] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, Morgan & Claypool Publishers, 2012.
- [16] B. Pang, L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1-135, 2008.
- [17] P. Gehler, S. Nowozin, "On feature combination for multiclass object classification," *Proceedings of IEEE 12th International Conference on Computer Vision (ICCV)*, 2009.
- [18] S. Kim, S. Park, S. Park, S. Lee, K. Kim, "A Syllable Kernel based Sentiment Classification for Movie Reviews," *Journal of Korean Institute of Intelligent Systems*, vol. 20, no. 2, pp. 202-207, 2010.

저 자 소 개



강인수(In-Su Kang)

1995년 : 경북대학교 컴퓨터공학 공학사
1999년 : POSTECH 컴퓨터공학 공학석사
2006년 : POSTECH 컴퓨터공학 공학박사
1995년~1997년 : 포스태이타
2006년~2008년 : 한국과학기술정보연구원
2008년~현재 : 경성대학교 컴퓨터공학부

관심분야 : 자연어처리, 정보검색, 시맨틱 웹
Phone : +82-51-663-5147
E-mail : dbaisk@ks.ac.kr