

감정 표현구 단위 분류기와 문장 단위 분류기의 결합을 통한 주관적 문장 분류의 성능 향상

강 인 호[†]

요 약

주관적 문장이란 주관적인 내용을 포함한 문장으로써 저자의 제품이나 사건에 대한 생각을 알 수 있다. 주관적 내용임을 나타내는 주관적인 표현은 문장 전반적으로 골고루 나타날 수도 있지만 일부 한정된 영역에서만 발견될 수도 있다. 따라서 보다 정확한 분류를 위해서는, 문장 전체를 고려하는 정보 외에 사실이나 감정을 표현하는 주관적 혹은 객관적 표현구 정보의 활용이 필요하다. 본 연구에서는 문장 전체를 이용한 분류 결과와 감정 표현구를 이용한 분류 결과를 결합하여 주/객관적 문장 분류기의 성능을 향상시키는 방법을 제안한다. 한 문장은 여러 개의 표현구를 가질 수 있어 복수개의 표현구 단위 결과를 얻게 되며 기계 학습을 응용하여 문장 단위 결과와 결합한다. 실험을 통한 결과, 표현구 단위 결과물 중 최대값을 가지는 두 가지 결과와 문장 전체를 이용한 결과를 합침으로써 2.5% 성능 향상된 79.7%의 정확률을 얻을 수 있었다.

키워드 : 주관적 문장, 객관적 문장, 주관적 표현구, 감정 표현구, 감정 분석

Combining Sentimental Expression-level and Sentence-level Classifiers to Improve Subjective Sentence Classification

In-Ho Kang[†]

ABSTRACT

Subjective sentences express opinions, emotions, evaluations and other subjective ideas relevant to products or events. These expressions sometimes can be seen in only part of a sentence, thus extracting features from a full-sentence can degrade the performance of subjective-sentence-classification. This paper presents a method for improving the performance of a subjectivity classifier by combining two classifiers generated from the different representations of an input sentence. One representation is a sentimental phrase that represents an automatically identified subjective expression or objective expression and the other representation is a full-sentence. Each representation is used to extract modified n-grams that are composed of a word and its contextual words' polarity information. The best performance, 79.7% accuracy, 2.5% improvement, was obtained when the phrase-level classifier and the sentence-level classifier were merged.

Key Words : Subjective Sentence, Objective Sentence, Subjective Expression, Semantic Phrase, Sentiment Analysis

1. 서 론

문서나 문장을 대상으로 주관적 내용인지 혹은 객관적 내용인지 분류하거나, 주관적 내용에 대해서 긍정 혹은 부정적인 내용인지를 분류하는 감정 분석(Sentiment Analysis)에 대한 요구가 계속 증가하고 있다. 이는 제품에 대한 의견을 자동으로 분류하여 제품의 장단점을 파악하는데 용이함을 제공하기 때문이다[1][2][3]. 그리고 질의 응답 시스템이나 정보 추출 시스템이 추출한 구절(Passage)이나 문장을 대상으로, 사용자의 의견이 반영된 주관적인 내용인지, 아니

면 사용자의 의견이 배제된 객관적인 내용인지를 분류하여 주관적 사실을 제외할 수 있다[4][5]. 또한 주관적 문장 분류기를 통해서 문서의 내용을 분류하거나, 저자의 의견이 나타난 부분을 파악하는데 활용할 수 있다[3]. 본 연구에서는 주어진 문장이 주관적 문장인지 객관적인 문장인지를 판별하는 분류기 제작을 목적으로 한다.

주관적 문장은 저자의 의견, 평가, 추측, 감정 등의 개인 상태(Private state)와 관련된 표현을 포함한다[6]. 이러한 표현은 문장 전체에 골고루 나타나는 경우도 있지만, 극히 일부분에만 나타나는 경우도 많다. 이러한 경우 문장 전체의 정보를 균일하게 고려하는 형태의 분류기는 좋은 결과를 얻지 못할 수 있다. 본 연구에서는 저자의 주관적인 표현을 인식하고 활용하기 위해 구 단위 분류기를 이용한다. 입력

[†] 정 회 원 : CMU/LTI 연구소 박사후과정
논문접수 : 2007년 6월 4일, 심사완료 : 2007년 7월 27일

문장을 객관적 혹은 주관적 표현 단위 형태인 표현구로 자동 분리하여, 각 표현구가 주관적인 표현을 포함한 주관적인 구인지를 판별한다. 주관적인 표현구로 판별될 경우, 이를 포함한 문장 전체를 주관적인 문장으로 분류한다. 표현구 단위 분류기 제작을 위해, 문장보다도 짧은 표현구에서도 분류에 유용한 자질을 충분히 추출할 수 있는 변형된 n-그램(n-gram)을 활용한다. 변형된 n-그램은 단어의 어휘 정보 외에 단어의 원형과 정보 검색기 어간 추출기(stemmer)를 이용한 어간 결과물, 그리고 단어가 가지는 주관적 표현 여부 혹은 긍/부정 선호 정보로 구성된다.

여러 개의 시스템 결과를 결합하여 각 시스템이 가지고 있는 단점을 보완하고 장점을 극대화하여 보다 강건하면서도 높은 성능의 시스템을 개발할 수 있다[7]. 본 연구에서 제안하는 문장 단위 분류기는 문장 전체에 나타나는 정보를 활용하여 멀리 떨어진 단어들의 공기 정보나 단어 활용의 비를 이용할 수 있고, 표현구 단위 정보를 활용하는 구 단위 분류기는 주/객관적 문장을 분류하는 기준이 되는 저자의 표현과 관련된 부분만을 집중적으로 파악할 수 있다는 장점이 있다. 그러나 표현구 단위 분류기는 특정한 경우 표현구가 없는 문장이 있을 수도 있고 인식 오류로 인해 제대로 적용하지 못하는 경우도 있다. 따라서 두 분류기의 단점을 극복하고 장점을 결합하기 위해 시스템 결합이 필요하다. 그러나 하나의 문장에서 얻을 수 있는 구 단위 결과의 수는 가변적이기 때문에 가변적인 수의 분류 결과를 결합하여 성능을 향상시키는 방법이 필요하다. 본 연구에서는 기존의 결합 방법을 살펴보고 주/객관적 문장 분류에 유용한 결합 방법을 보인다.

본 논문의 구성은 다음과 같다. 주관적 객관적 문장 분류에 대한 기존 연구를 2절에서 살펴본 뒤, 저자의 표현구를 포함하고 있는 MPQA 코퍼스에 대해서 3절에서 살펴본다. 4절에서는 학습 말뭉치에서 문장 단위 분류기와 구 단위 분류기 생성을 위해 자질을 추출하는 방법과 결합하는 방법을 제안하고 5절에서는 실험을 통한 제안하는 방법의 유용성을 보인다.

2. 관련 연구

2.1 단어 리스트 기반 분류

감정 표현이나 객관적 사실 표현에 자주 사용되는 단어를 활용하면 주관적/객관적 문장을 분류할 수 있다. 예를 들어 'station'과 'percent'는 뚜렷한 의미적인 관련성이 없는 단어이지만, 이 두 단어가 나타나는 문장은 객관적인 문장일 가능성이 높다. 분류를 위한 단어들은 주관적 혹은 객관적 문서를 통해서 얻을 수 있다. 예를 들어 신문의 사설이나 투고문에 많이 나타나는 단어는 주관적 문장을 구별하는 좋은 단어들로 사용할 수 있으며, 사실을 전달하는 뉴스 기사나 개념을 설명하는 백과사전 같은 문서에 많이 나타나는 단어는 객관적인 문장을 구별하는 좋은 단어들로 사용할 수 있다[8][9]. 초기에는 형용사, 부사, 동사의 단어를 추출 및 활

용하는 연구에서 점차 이들 단어와 같이 나타나는 명사를 활용하는 형태로 발전해왔다. 이러한 단어들은 WordNet과 같은 사전의 동의어 정보를 활용하여 자동으로 단어 리스트를 늘릴 수 있다[10]. Yu(2003)는 단어와 품사 정보 그리고 긍/부정 정보를 결합하는 Naïve Bayes 분류기를 생성하여 높은 재현율과 정확률을 가지는 주관적 문장 분류기 생성 방법을 보였다[4]. 또한 단어 n-그램 정보를 결합하여 객관적 문장 분류기의 성능 또한 향상시킬 수 있음을 보였다.

2.2 패턴 기반 분류

정보 추출 연구 분야에서는 "delegation from <np>"와 같은 구문 패턴을 이용하여 문서에 나타난 개체나 개체간의 관계 정보를 추출한다. Riloff(2003)는 이러한 구문 패턴을 주관적인 문장을 구분하는 데에 사용하였다[11]. <표 1>은 사용된 구문 패턴의 예를 보이며, 이렇게 사람이 작성한 구문 패턴은 bootstrapping 방식을 이용하여 자동으로 확장되고 간략한 구문 해석기를 이용하여 문장에서의 출현여부를 판별한다. 구문 패턴의 활용으로 높은 정확률(90.2%)을 보였으나 낮은 재현율(40.1%)을 보였다. 한편 구문 패턴을 이용하는 대신 구문 패턴과 공기하는 단어를 추출하여 분류기의 재현율을 높이는데 사용하여 좋은 성능을 얻는 방법이 있었다[12]. 또한 패턴을 이용할 경우 비지도 학습으로도 좋은 성능을 얻을 수 있음을 보였다[13].

<표 1> 주관적 객관적 표현 패턴 예제

Subjective Patterns	Objective Patterns
<subj> believes	Plans to produce <dobj>
<subj> was convinced	<sub> took effect
to express <dobj>	occurred on <np>

2.3 시스템 결합

사용하는 정보와 기계 학습 방법의 차이에 따라서 각각의 시스템은 고유한 장단점을 가진다. 단점을 보완하고 장점을 극대화하기 위해 여러 시스템의 결과를 결합하여 메타 분류기라고도 불린다. 결합에 사용되는 시스템이 전반적으로 유사한 성능을 나타낼 경우에는, 시스템들의 과반수 결과를 선택하는 투표(Voting) 방식이나 결과 중 최대값을 출력한 시스템의 결과를 선택하는 최대치 선정 방식과 같은 비지도 학습의 결합 방법이 가능하다. 반면 학습 데이터를 통해서 시스템의 출력값과 정확률의 관계를 분석하여 각 시스템의 장단점을 파악하고, 시스템 출력들간의 관계를 분석하여 결합하는 방법이 가능하다. 이런 방법으로는 각 시스템의 출력값 분포 정보를 입력 노드로 사용하여 결정트리로 결합하는 메타 결정트리(Meta decision-tree)와, 선형회귀법(Linear-regression)을 활용하여 결합하는 다중응답회귀(Multiple-response regression), 그리고 트리구조(tree structure)를 활용하여 회귀분석(regression)을 수행하여 결합하는 다중응답모델트리(Multiple-response model tree) 방법이 있다[7].

3. MPQA 코퍼스

본 연구에서는 분류기를 구축하고 평가하기 위해 MPQA (Multi-Perspective Question Answering) 코퍼스를 사용한다 [6]. MPQA 코퍼스는 뉴스 기사 문서를 대상으로 객관적인 표현과 주관적인 내용의 표현 등의 정보가 사람에게 의해 부착되어 있다. 여기서 객관적인 표현은 *objective speech event*로, 주관적인 표현은 개인상태(Private State)를 직접적으로 나타내는 *direct-subjective element*와 간접적으로 나타내는 *expressive-subjective element*로 나뉘어서 부착되어 있다. 각각의 표현은 강도를 나타내는 *intensity*가 부착되어 있다. 주관적인 문장은 강도 높은 (extreme, high, medium) *expressive-subjective element*나 *direct-subjective element*를 가진 문장으로 정의된다.

다음 두 문장은 MPQA 코퍼스에 부착되어 있는 여러 자질 정보들을 보여준다.

- (1) The news [*o₁* leaked] from the reoccupied Palestinian cities [*s₁* confirms] that Israeli troops have committed [*s₂* dozens] of [*s₃* premeditated massacres] against the civilian Palestinian population.
o₁: objective speech event
s₁: expressive-subj. medium intensity
s₂: expressive-subj. low intensity
s₃: expressive-subj. extreme intensity
- (2) [*s₁* However], [*o₁* according to] broad statistics, Taiwan-born voters [*s₂* favoring independence] account for 65 percent of the island's 23 million population.
s₁: expressive-subj. low intensity
s₂: direct-subj. high intensity
o₁: objective speech event

예를 들어 예제 문장 (1)은 1개의 객관적인 표현과 3개의 주관적인 표현을 가진다. 두 문장은 각각 강도 높은 *expressive-subjective element*와 *direct-subjective element*

를 가지고 있기 때문에 주관적인 문장이다.

MPQA 코퍼스를 이용하여 많은 주/객관적 문장 분류기 및 감정 분류기의 성능을 비교하고 있으며, Riloff(2003)가 제일 좋은 성능을 보였다[12]. 그들은 주관적 명사 (subjective noun), 어간(stems), 화용자질(discourse features)을 활용하였다. 여기서 주관적 명사 정보는 MetaBoot나 Basilisk와 같은 부트스트래핑(bootstrapping) 알고리즘을 이용해서 생성 확장된 단어들과, 대명사, 수사, 숫자, 조동사(Modal), 부사 등의 출현 여부를 사용한다. 그리고 화용정보는 문서에서 입력 문장 앞 뒤로 나타나는 주관적 객관적 단어의 비례 이용하여 문장의 주위 문맥을 고려하는 자질과 문서가 속한 도메인 정보 그리고 문장 길이를 고려한다. 본 연구에서 제안하는 방법은 사용자가 부착한 주관적/객관적 표현구를 사용한다는 차이점을 가진다.

4. 주/객관적 문장 분류

4.1 문장의 전처리

주어진 입력 문장은 심벌의 제거와 단어의 원형 복원, Porter Stemmer를 이용한 어간추출(Stemming), 그리고 Brill 품사 태거를 이용하여 단어 별 정보를 추출한다[14]. 단어의 품사와 원형 정보를 이용하여 각 단어가 가지는 긍/부정 선호 정보와 주관적 표현 선호 정보를 추출한다. 이러한 감정 선호 정보 (Semantic Preference) 리스트는 Wilson(2005)에 의해 작성되었으며 약 8,000 단어를 포함한다[2]. 긍/부정은 긍정(positive), 부정(negative), 그리고 중립(neutral)으로 나뉘며, 주관적 표현은 강한 주관적 표현(SSJ)과 약한 주관적 표현(WSJ)으로 나뉜다. <표 2>는 예제 문장 (1)과 (2)에 나타난 일부 단어의 선호 정보를 나타낸다. 여기서 POS는 선호 정보를 매칭할 때 사용하는 품사의 조건이다. 'anypos'는 어떠한 품사로 문장에 나타나도 상관없다는 뜻이다. SSJ와 WSJ는 각각 Strong Subjective와 Weak Subjective를 의미하며 negative는 부정, positive는 긍정을 뜻한다. 입력 문장의 각 단어는 Subjectivity와 Prior-Polarity를 결합한 값을 감정 선호 정보로 가진다.

(그림 1)은 전처리를 통해서 획득한 단어 단위 정보를 나

The_(DT:the:N:The) news_(NN:news:N:new) <o> leaked_(VBN:leak:wsj;ng:leak) </o> from_(IN:from:N:from) the_(DT:the:N:the) reoccupied_(VBN:reoccupy:N:reoccupi) Palestinian_(JJ:palestinian:N:Palestinian) cities_(NNS:city:N:citi) <s> confirms_(VBZ:confirm:N:confirm) </s> that_(IN:that:N:that) Israeli_(JJ:israeli:N:Israeli) troops_(NNS:troop:N:troop) have_(VBP:have:N:have) committed_(VBN:commit:N:commit) <o> dozens_(NNS:dozen:N:dozen) </o> of_(IN:of:N:of) <s> premeditated_(JJ:premeditate:ssj;ng:premedit) massacres_(NNS:massacre:ssj;ng:massacr) </s> against_(IN:against:wsj;ng:against) the_(DT:the:N:the) civilian_(JJ:civilian:N:civilian) Palestinian_(JJ:palestinian:N:Palestinian) population_{(NN:population:N:popul) *(::N:.)}

(그림 1) 전처리 결과 예제

〈표 2〉 단어별 감정 선호 정보

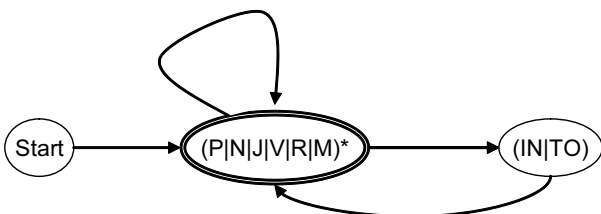
Word-form	POS	Subjectivity	Prior-Polarity
premeditated	adj	SSJ	negative
massacres	noun	SSJ	negative
However	anypos	WSJ	negative
favor	verb	SSJ	positive
independence	noun	WSJ	positive

타낸다. 여기서 <s>, </s> 그리고 <o>, </o>는 각각 주관적인 표현과 객관적인 표현의 범위를 나타낸다. 표현의 범위는 MPQA 코퍼스에 표기된 표현구의 범위와 동일하지만 주관적 표현인지 객관적인 표현인지는 주관적인 문장 구별 기준과 동일하게 표현의 강도가 강한 표현을 주관적인 표현으로 그렇지 않은 표현구는 객관적인 표현구로 결정한다. 예를 들어 단어 ‘massacres’는 품사 NNS, 원형 massacre, 감정 선호 정보 ssj:ng, stem ‘massacr’을 단어 정보로 가진다. 그리고 표현어구 “dozens”는 subjective expression이지만 intensity가 low이기 때문에 objective expression으로 변환한다.

4.2 표현구 단위 분할

주관적, 객관적인 표현은 일반적으로 구 단위로 나타난다. 입력 문장에서 이러한 표현을 자동으로 파악하기 위해 유사 표현구를 사용한다. 유사 표현구는 (그림 2)에 나타난 입력 문장의 품사열 정규 패턴을 이용하여 인식된다. (그림 2)에서 P*, N*, J*, V*, R*, M*는 각각 고유명사(proper noun), 명사(noun), 형용사(adjective), 동사(verb), 부사(adverb)와 조동사(auxiliary verb)를 IN, TO는 ‘in’, ‘to’, ‘of’를 포함하는 전치사를 나타낸다¹⁾. 여기서 인식되는 구는 일반적으로 명사구, 동사구, 형용사구와 부사구에 해당한다. 여기서 관사(determiner)는 구 인식에 중요한 품사이지만 주/객관적 분류에는 유용하지 않기 때문에 제외한다.

인식된 유사 표현구를 대상으로 학습 데이터에서 주관적인 표현으로 태깅된 일부의 표현을 포함하는 유사구는 주관적 표현구로 분류하고 다른 유사구는 객관적 표현구로 분류한다. 예를 들어 예제 문장 (1)에서 세 개의 표현구가 인식되며, 이 중 사람이 주관적인 표현으로 태깅한 ‘confirms’와 ‘premeditated massacres’를 포함하는 유사 표현구는 주관적 유사구 표현으로 태깅된다(그림 3). 그러나 그 외의 표현



(그림 2) 유사 표현구 인식을 위한 상태 전이도

1) Brill POS Tagger가 출력하는 Penn-Treebank 품사체계를 이용한다.

Objective:	news leaked
Subjective:	reoccupied Palestinian cities confirms that Israeli troops have committed dozens of premeditated massacres
Objective:	civilian Palestinian population

(그림 3) 예제 문장 (1)에서 인식한 유사 표현구

구는 사람이 주관적인 표현이라고 한 표현을 포함하지 않기 때문에 객관적 유사 표현구로 태깅된다.

4.3 자질 추출

유사구 주위에 나타나는 문맥적인 표현과 내부에 사용된 표현을 활용하여 주/객관적 표현구를 자동 분류한다. 자료 희소성(data sparseness) 문제를 해결하기 위해 어휘 n-그램 대신 단어의 감정 선호 정보를 결합하는 변형된 n-그램을 사용한다. <표 3>은 변형된 n-그램을 정의한 것으로써, $rootForm(w_i)$, $stem(w_i)$, $semPref(w_i)$ 은 각각 단어 w_i 의 원형, stem, 그리고 감정 선호 정보를 나타낸다.

고유 명사인 경우에는 단어의 어휘를 사용하지 않고 품사를 대신 사용한다. 이는 분류기가 특정 영역에 과잉 맞춤(over-fit)되는 것을 막기 위함이다. 문장 단위 분류기와 표현구 단위 분류기는 자질을 추출하는 문장에서의 영역만 다르고 동일한 자질 템플릿을 사용한다. 유사 표현구에 대한 자질 추출은 유사 표현구 바로 앞의 단어와 표현구를 뒤따르는 두 단어를 자질 추출 대상 영역으로 이용하여 주위 문맥과 표현구 인식의 오류를 보완한다. 예를 들어 w_3, w_4, w_5 가 인식된 유사 표현구일 경우, w_3 에서 w_6 까지의 어휘를 대상으로 <표 3>에 정의된 자질 템플릿에 기반해서 자질을 추출한다. 이때 자질 템플릿 정의에 의해 w_2 와 w_7 의 정보가 포함된 자질을 자연스럽게 포함한다. <표 4>는 예제 (1)의 단어 ‘massacre’에 대해 추출된 자질을 나타낸다.

〈표 3〉 자질 추출 템플릿

ID	Definition
1	$rootForm(w_i) + semPref(w_i)$
2	$stem(w_i)$
3	$semPref(w_i) + stem(w_{i+1})$
4	$stem(w_i) + stem(w_{i+1})$
5	$stem(w_{i-1}) + semPref(w_i)$
6	$stem(w_{i-1}) + stem(w_i)$
7	$stem(w_{i-1}) + semPref(w_i) + semPref(w_{i+1})$
8	$semPref(w_{i-1}) + stem(w_i) + semPref(w_{i+1})$
9	$semPref(w_{i-1}) + semPref(w_i) + stem(w_{i+1})$

〈표 4〉 ‘massacre’에 대해 추출된 자질들

ID	Extracted Feature
1	massacre + ssj:ng
2	massacre
3	ssj:ng + against
4	massacre + against
5	premedit + ssj:ng
6	premedit + massacre
7	premedit + ssj:ng + wsj:ng
8	ssj:ng + massacre + wsj:ng
9	ssj:ng + ssj:ng + against

추출된 자질은 출현 횟수에 관계 없이 동일한 가중치를 가진다. 이는 유사 표현구나 문장이 매우 작아서 빈도나 비를 고려하기 힘들며 주/객관적 표현 분류는 표현의 출현 여부이지 과반수의 값에 의해 결정되는 것이 아니기 때문이다. 유사 표현구 중 하나라도 주관적인 표현구로 분류 될 경우 그 문장은 주관적 문장으로 분류된다.

4.4 자질 선택

전치사 ‘of’, ‘to’나 관사 ‘a’, ‘an’과 같은 단어는 구문 해석 시 구 인식에는 유용할지 모르나 주/객관적 분류에는 큰 도움이 되지 않는다. 학습 데이터에서 추출한 자질을 대상으로 카이제곱(chi-square) 값을 이용하여 유용성을 평가하여 유용한 자질만을 사용한다. 각 자질에 대해서 주관적 표현으로 나타나서 사용되는 경우와 객관적 표현으로 나타나서 사용되는 경우를 계산하여 두 경우의 차가 큰 자질을 사용한다는 의미이다. <표 5>는 자질 선택을 통해 높은 카이제곱(chi-square) 값을 가지는 상위 10개의 자질을 보인다. 문장 단위 자질에서 ‘.’이 높은 값을 가지는 것으로 나왔는데 이는 MPQA 테스트 문장의 59%가 주관적 문장이기 때문이다. 자질 선택을 위해 사용하는 카이제곱(chi-square)의 임계값은 학습 데이터를 이용한 5-fold 교차타당성(cross evaluation)을 통해서 가장 좋은 결과를 보인 값으로 결정한다.

선별된 자질들은 SVM model로 기계 학습을 수행하며 이를 위해 SVMlight 툴킷을 이용한다[15].

〈표 5〉 chi-square로 선별된 자질의 예

Phrase-Level	Sentence-Level
ssj:ng	ssj:ng
N + N + ,	N + N + .
the + N + N	ssj:ps
the + N	N + .
N + ,	ssj:ng + N + the
ssj:ps	will + N + N
N + , + N	.
NNP + N + N	N + emiss + N
NNP	N + “
ssj:ng + N + the	N + N + by

4.5 분류기 생성

학습 데이터를 대상으로 문장 단위 분류기와 유사 표현구

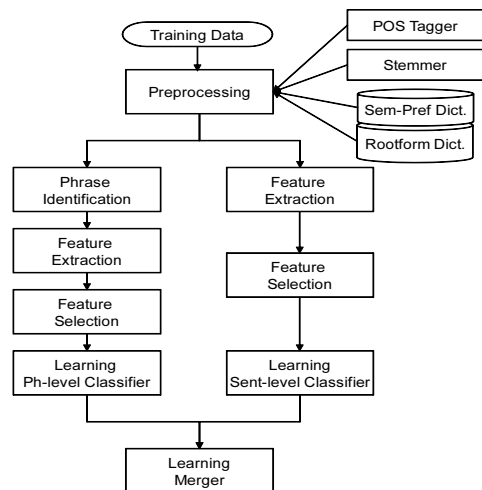
단위 분류기를 각각 생성한다. 두 분류기는 동일한 전처리 과정을 거치며, 사용하는 자질 추출 템플릿과 자질 선택 알고리즘 및 기계 학습 알고리즘도 동일하다. 그러나 문장 단위 분류기는 문장 전체를 대상으로 자질을 추출하고 문장이 주관적 문장인지 객관적 문장인지를 판별하는 반면, 유사 표현구 단위 분류기는 문장에서 추출한 유사 표현구를 대상으로 자질을 추출하고 표현구가 주관적 표현구인지 객관적 표현구인지를 판별한다.

4.6 분류 결과 결합

문장 단위의 분류 결과와 유사 표현구 단위의 결과를 결합하는 메타 분류기를 생성하기 위해 SVM 기계 학습을 활용한다. 기존의 메타 분류기와의 차이점은 결합 할 결과의 개수가 일정하지 않다는 것이다. 문장에서 얻을 수 있는 유사 표현구 개수가 정해져 있지 않기 때문이다. 이러한 복수개의 값을 나타내는 방법으로 평균, 분산, 최고값, 최저값 등의 통계값을 생각할 수 있다. 여기서 기계 학습의 결과로 나오는 값이 주관적인 문장으로 판별될 값이라고 할 경우, 주관적인 문장은 주관적으로 태깅되는 표현구를 가지는 문장이므로 평균이나, 분산 그리고 최저값은 의미가 없다. 본 연구에서는 주관적인 표현구로 판단될 가능성이 높은 2개의 결과 값과 문장 단위 분류기의 결과 값을 결합하는 형태로 두 개의 상이한 분류기의 결과를 합친다. 이때 사용하는 결과 값은 SVM 기계학습을 통해서 얻어낸 값을 그대로 사용한다. 세개의 값은 3차원 벡터값으로 표현되며 이때 정답으로 가져야 할 분류 결과와 함께 SVM 기계 학습을 이용해서 메타분류기를 생성한다.

$$V = (ph_{best}, ph_{snd}, snt_{res})$$

여기서 ph_{best} 와 ph_{snd} 는 유사 표현구 단위 결과값 중에서 상위 2개의 값을 나타내며 snt_{res} 는 문장 단위 분류기의 결과 값을 나타낸다. 이는 SVM 기계 학습을 통해 각 결과값의 최적의 결합 방식을 찾아내는 메타 분류기를 만들기 위함이다.



(그림 4) 문장 분류기 전체 흐름도

(그림 4)는 본 연구에서 제안하는 구 단위 분류기와 문장 단위 분류기의 제작 및 결합의 전체적인 흐름을 보여준다.

5. 실험

본 연구에서 제안하는 분류기의 유용성을 평가하기 위해 기존 최고의 성능을 보이고 있는 Riloff (2003)가 사용한 방법과 동일한 테스트 문장과 테스트 방식을 사용한다. 테스트 문장은 MPQA 코퍼스에서 추출한 총 2,197 문장으로 이중 1,296 (59%) 문장이 주관적 문장이며, 25-fold 교차타당성(cross validation)을 수행한다. 평가를 위해서 주관적 문장에 대한 정확률(SubjP), 재현율(SubjR) 그리고 F-measure(SubjF) 값과 객관적 문장에 대한 정확률(ObjP), 재현율(ObjR) 그리고 F-measure (ObjF) 값 그리고 모든 문장에 대한 정확도(Acc)를 측정한다. 기준 시스템(Baseline System)으로 유니그램(unigram)에서 5-그램(5-gram)까지의 어휘 n-그램을 SVM 모델을 이용하여 결합해서 사용하는 시스템과 문장에 나타난 단어 별 감정 선호 정보만을 이용하는 시스템(Sem-Pref)을 사용한다. <표 6>은 문장 분류 결과를 나타낸다.

<표 6> 객관적/주관적 문장 분류 결과

	SubjP	SubjR	SubjF	ObjP	ObjR	ObjF	Acc
Unigram	76.5	77.3	76.6	67.1	66.3	66.2	72.6
+ Bigram	77.1	75.8	76.2	66.2	67.8	66.5	72.4
+ Trigram	76.3	76.4	76.1	66.4	66.3	65.8	72.1
+ 4-gram	76.1	77.1	76.3	66.8	65.5	65.5	72.1
+ 5-gram	74.9	77.4	75.9	66.4	63.0	64.0	71.3
Sem-Pref	78.9	72.4	75.2	64.3	72.2	67.5	72.1
Phrase-Level	83.0	77.5	79.9	70.2	77.1	73.1	77.2
Sentence-Level	80.6	80.5	80.4	71.9	72.4	71.9	77.1
Riloff (2003)	81	77	79	70	74	72	76

어휘 n-그램을 5-gram까지 사용하여도 성능의 변화가 적으며 오히려 감소하는 것을 알 수 있다. 반면 감정 선호 정보만을 사용한 결과가 어휘를 사용한 결과와 흡사한 성능을 보인다. 이를 통해 주관적 객관적 문장 분류에 감정 선호 정보가 매우 유용함을 알 수 있다. 그리고 유사 표현기반 분류기의 성과와 문장 단위 분류기는 흡사한 정확률을 보였다. 주관적인 문장 분류에서는 유사 표현기반 분류기가 보다 높은 정확률을 보였으며, 문장 단위 분류기가 높은 재현율을 보였다. 이는 본 연구에서 제안한 유사 표현기반 인식 결과가 정확률은 좋으나 재현율이 부족함을 뜻한다.

유사 표현기반 단위 분류기의 높은 정확률과 문장 단위 분류기의 좋은 재현율을 결합하기 위해 두 분류기의 결과를 결합하여 사용한다. <표 7>은 결합을 통한 분류기의 성능의 변화를 보인다.

<표 7> 결합을 통한 문장 분류기의 결과

	SubjP	SubjR	SubjF	ObjP	ObjR	ObjF	Acc
VOTE	67.3	92.1	77.6	75.4	35.6	47.9	69.0
SelBest	72.8	88.3	79.6	75.3	52.5	61.4	73.6
Both Features	78.9	84.7	81.6	75.7	67.8	71.2	77.8
MSVM-1	81.3	84.5	82.6	76.3	72.0	73.7	79.4
MSVM-2	81.5	84.9	83.0	76.9	72.3	74.1	79.7
MSVM-3	80.4	85.2	82.6	76.7	70.1	72.8	79.0

<표 7>에서 *VOTE*와 *SelBest*는 각각 투표표를 통한 과반수 분류 결과를 선택하는 방법과 최대치 결과값을 가지는 분류 결과를 선택하는 방법을 뜻한다. 이 과정에서 유사 표현기반 단위 결과 모두를 동등하게 간주하여 문장 단위 분류기의 결과와 결합한다. 또한 *BothFeatures*는 문장 단위 분류기에서 추출한 자질 정보와 유사 표현기반 단위에서 추출한 자질 정보를 동시에 기계 학습에 사용한 경우를 나타낸다. 즉 문장에서 추출 가능한 모든 자질을 대상으로 문장인지 표현기반인지 출처 표시를 덧붙여 기계 학습을 수행하였다. *MSVM-1*은 유사 표현기반이 주관적 표현으로 분류될 가능성을 나타내는 결과 출력값 중 최대값만을 결합에 사용한 것이고, *MSVM-2* 그리고 *MSVM-3*은 각각 최대값 2개, 3개를 사용했을 때의 결과를 나타낸다. <표 7>을 통해 두 분류기의 결과 값을 기계 학습을 통해 결합하는 방법이 자질을 결합하는 것보다 높은 성능 향상을 얻을 수 있다. 이는 유사 표현기반 단위로 분류기를 학습하는 것이 유용하며, 기계 학습을 통한 결합 방법이 문장 단위 분류기와 유사 표현기반 단위 분류기의 장단점을 잘 결합해준다는 것을 뜻한다. 즉, 유사 표현기반으로 인식하지 못하는 주관적 문장에 대한 재현율을 시스템 결합을 통해 향상시켰다.

<표 8>은 분류기 학습 시 사용한 일부 정보를 제외했을 때의 *MSVM-2*의 결과를 나타낸 것으로 자질 추출에 사용한 정보와 모듈의 중요성을 보인다. 고유 명사를 어휘 정보로 그대로 사용한 결과(w/ Proper Noun)와 어간 추출기를 사용하지 않은 경우(w/o Stemmer) 그리고 단어 별 감정 선호 정보를 사용하지 않았을 때(w/o Sem-Pref)와 감정 선호 정보 중 주관적/객관적 표현 정보를 사용하지 않은 경우(w/o Subjective) 긍정/부정 정보를 사용하지 않았을 때의 결과(w/o Prior-Polarity)를 각각 보여준다.

고유 명사 배제는 큰 영향이 없지만, 어간 추출기 사용을 통해서 객관적 문장 분류의 재현율이 많이 향상되었다는 것

<표 8> 정보 별 문장 분류기 결과

	SubjP	SubjR	SubjF	ObjP	ObjR	ObjF	Acc
w/ Proper Noun	81.6	84.8	82.7	77.1	72.1	73.6	79.3
w/o Stemmer	79.8	85.1	82.3	76.1	69.3	72.2	78.6
w/o Sem-Pref	79.5	82.6	80.9	73.8	69.4	71.3	77.4
w/o Subjective	81.2	81.8	81.4	73.7	72.9	73.0	78.2
w/o Prior-Polarity	80.5	82.6	81.3	73.9	71.1	71.9	77.8

〈표 9〉 정답 표현구를 활용한 문장 분류 결과

	SubjP	SubjR	SubjF	ObjP	ObjR	ObjF	Acc
Phrase-Level	76.1	96.2	84.8	92.0	56.8	70.0	80.1
MSVM-2	84.2	86.8	85.4	80.7	77.0	78.5	82.8

을 알 수 있다. 그리고 단어 별 감정 선호 정보가 제일 큰 영향을 미치는 것을 알 수 있으며 긍/부정 정보를 통해 객관적 문장 분류에 좀더 좋은 성능을 보임을 알 수 있다. 여기서 단어 별 감정 정보를 이용하지 않은 경우는 기존 시스템 성능과 유사하게, 문장 단위 분류기가 71.2%와 구 단위 분류기가 72.5%의 정확도(Acc) 성능을 보였으나 이를 합침으로써 77.4%까지 성능 향상이 가능했다. 이를 통해 본 연구에서 제안하는 결합 방법이 유용함을 알 수 있다.

〈표 9〉는 자동으로 인식하는 표현구 대신 사용자가 태깅한 정답 표현구를 그대로 사용해서 분류기를 학습하고 테스트에 사용했을 경우의 표현구 단위 분류기와 MSVM-2의 성능을 보인다.

결과를 통해 정확률(SubjP, ObjP)의 많은 향상을 볼 수 있다. 그러나 재현율 특히 객관적 문장의 재현율이 좋지 않은 성능을 보인다. 이는 추출한 주관적 표현구 수에 비해서 객관적 표현구 수가 부족하였기 때문이다. 주관적인 표현구를 잘 인식하기 위해서는 그에 대응하는 객관적인 표현구에 대한 정보 수집 또한 필요하다. 본 연구에서는 표현 어구를 품사열 패턴으로 정의하여 주관적 감정 표현과 형태적으로 유사한 많은 양의 객관적 표현 추출을 가능하게 했으며, 이를 통해 좋은 성능을 얻을 수 있게 했음을 알 수 있다.

6. 결 론

문장은 문서에 비해 추출할 수 있는 정보의 수가 한정적이기 때문에 의미 있는 단어의 출현 빈도나 비를 얻어내기가 힘들다. 또한 주관적인 문장을 판별하는 주관적인 표현은 문장 전체에서 골고루 나타나는 것이 아니라 문장의 일부분에서 나타나기도 한다. 따라서 문장 전체를 주관적인 문장이나 객관적인 문장으로 간주하고 자질을 추출할 경우 잘못된 학습을 유도할 수 있다. 본 연구에서는 이러한 문제를 해결하기 위해 저자의 사실 표현이나 주관적인 내용 표현에 사용되는 어구에 대응하는 유사 표현구 단위 분류기를 제안했다. 품사열 패턴을 이용하여 유사 표현구를 자동으로 인식하여 저자의 감정 표현이 부착된 학습 말뭉치에서 충분한 주관적 객관적 표현구를 추출하고 학습하여 주관적/객관적 문장을 분류한다. 문장보다도 짧은 유사 표현구에서 충분한 자질을 획득하기 위해, 기존의 어휘 n-그램 대신 단어 별 감정 선호 정보를 활용한 변형된 n-그램을 추출하여 자질 정보로 사용했다. 실험을 통하여 변형된 n-그램을 활용할 경우 표현구 단위 분류가 좋은 정확률을 그리고 문장 단위 분류기가 좋은 재현율을 보였다. 이 두 분류기의 장점을 결합하기 위해 표현구 단위의 분류 결과 중 상위 2개의 결

과와 문장 단위 분류기의 결과를 기계 학습을 통하여 결합하는 방법이 가장 좋은 성능을 얻을 수 있음을 알았다.

정답 표현구를 사용했을 경우의 실험을 통해 보다 정확한 유사 표현구 인식의 방식이 필요함을 알 수 있었다. 아울러 유사 표현구의 인식을 통해 문장에 나타난 의견, 평가, 감정에 대한 표현을 자동 인식할 수 있는 방법의 연구가 필요하다.

참 고 문 헌

- [1] Kushal Dave, Steve Lawrence and David M. Pennock, "Mining the peanut gallery: opinion extraction and semantic classification of product reviews," Proceedings of the 12th international conference on World Wide Web, 519-528, 2003.
- [2] Theresa Wilson, Janyce Wiebe and Paul Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," Proceedings of Human Language Technology Conference and North American Chapter of the Association for Computational Linguistics, 2005.
- [3] Bo Pang and Lillian Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," Annual Meeting of the Association for Computational Linguistics (ACL), 2004.
- [4] Hong Yu and Vasileios Hatzivassiloglou, "Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences," Proceedings of the Conference on Empirical Methods in Natural Language Processing vol. 10, 129-136, 2003.
- [5] Ellen Riloff, Janyce Wiebe and William Phillips, "Exploiting Subjectivity Classification to Improve Information Extraction," Proceedings of the 20th National Conference on Artificial Intelligence, 2005.
- [6] Theresa Wilson and Janyce Wiebe, "Annotating Opinions in the World Press," Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue, 2003.
- [7] Saso Dzeroski and Bernard Zenko, "Is Combining Classifiers Better than Selecting the Best One?" Machine Learning, 54(3), 255-273, 2004.
- [8] Soo-Min Kim and Eduard Hovy, "Automatic Detection of Opinion Bearing Words and Sentences," Proceedings of the Second International Joint Conference on Natural Language Processing, 2005.
- [9] Bo Pang, Lillian Lee and Shivakumar Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2002.
- [10] George A. Miller, "WordNet: a lexical database for English," Communications of the ACM, 38(11), 39-41, 1995.

- [11] Ellen Riloff and Janyce Wiebe, "Learning Extraction Patterns for Subjective Expressions," Proceedings of the Conference on Empirical Methods in Natural Language Processing, 105-112, 2003.
- [12] Ellen Riloff, Janyce Wiebe and Theresa Wilson, "Learning Subjective Nouns Using Extraction Pattern Bootstrapping," Proceedings of the Seventh Conference on Natural Language Learning, 25-32, 2003.
- [13] Janyce Wiebe and Ellen Riloff, "Creating Subjective and Objective Sentence Classifiers from Unannotated Texts," Proceedings of the Sixth International Conference on Intelligent Text Processing and Computational Linguistics, 2005.
- [14] Eric Brill, "Transformation-based Error-driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging," Computational Linguistics, 1995.
- [15] Thorsten Joachims, "Making large-Scale SVM Learning Practical," Advances in Kernel Methods - Support Vector Learning, MIT Press, 169-184, 1999.



강 인 호

e-mail : ihkang97@cs.cmu.edu

1997년 경북대학교 컴퓨터공학과(학사)

1999년 한국과학기술원 전산학과

(공학석사)

2004년 한국과학기술원 전산학과

(공학박사)

2004년~2005년 삼성종합기술원 전문연구원

2006년~현 재 CMU/LTI 연구소 박사후과정

관심분야: 정보 검색 및 마이닝, 자연언어처리 등