

감정자질과 커널모형을 이용한 영화평 평점 예측 시스템

허향란^o, 정형일, 서정연*
서강대학교 컴퓨터공학과, *서강대학교 컴퓨터공학과/바이오융합기술협동과정
{xyzljz, hijeong}@gmail.com, jyseo@sogang.ac.kr

A Rating System on Movie Reviews using the Emotion Feature and Kernel Model

Xiang-Lan Xu^o, Hyoung-Il Jeong, Jung-Yun Seo*

Department of Computer Science and Engineering, Sogang University

*Department of Computer Science and Engineering, and Interdisciplinary Program of Integrated Biotechnology, Sogang University

요 약

본 논문에서는 최근 많은 관심을 받고 있는 Opinion Mining으로서 사용자들의 자연어 형태의 영화평 문장을 분석하여 자동으로 평점을 예측하는 시스템을 제안한다. 제안 시스템은 영화평 분석에 적합한 어휘 자질, 감정 자질, 가치 자질 및 기타 자질들을 추출하고, 10점 척도의 영화평의 평점을 10개의 범주로 가정하여, 커널모형인 다중 범주 Support Vector Machine (SVM) 모델을 이용하여 높은 성능으로 영화평의 평점을 범주 분류한다.

주제어: Opinion Mining, 영화평 평점 예측, 감정 자질, Support Vector Machine (SVM)

1. 서론

최근 온라인 커뮤니티나 Social Network Services (SNS) 등이 활성화됨에 따라, 많은 사람들이 특정 상품에 대하여 온라인상에서 많은 의견을 표현하고 있다. Opinion Mining은 감정분석이라 알려져 있으며, 텍스트에서 긍정 혹은 부정의 의견을 추출하는 기술이며[1], 특히 자연어 형태의 영화평으로부터 평점을 예측하는 것도 일종의 Opinion Mining으로서 영화평 속에 담긴 사용자의 정서와 선호도를 파악 하는 것을 목표로 한다.

Opinion Mining의 개념을 이용하여 연구자들은 감정분류에 관한 연구를 많이 해왔으며[2-3], 영화평 분석에 대한 연구도 많이 진행되고 있다[4]. Turney는 영어에 대하여 형용사와 부사에 대한 감정 분류 시스템을 제안하였고[2], Kamps, et al.는 WordNet[5]에서의 어휘관계를 통하여 형용사단어에 대한 감정 분류 방법을 제안하였다[3]. 김상도 외는 문자열 커널의 확장 모델인 음절 커널에 기반 하여 영화평 감성 분류를 하였다[4].

온라인 영화 커뮤니티 등에서의 평점은 평점자가 직접 수동으로 평점을 매기는 방법이 주로 사용되고 있다. 본 논문에서는 자연어 형태의 영화평을 분석하여 그 평의 적합한 평점을 자동으로 예측하는 시스템을 제안한다. 제안하는 시스템은 영화평의 평점을 범주로 가정하고, 커널모형인 다중 범주 Support Vector Machine (SVM)[6]을 통해 평점 범주를 분류한다. 제안 시스템은 평점 범주의 특징을 잘 나타낼 수 있는 어휘 자질과 감정 자질, 가치 자질 등을 추출하여 평점 범주를 분류한다.

본 논문의 구성은 다음과 같다. 2절에서는 제안 시스템의 전체구성과 사용하는 자질을 설명하고, 3절에서는

실험 및 평가를 보인다. 마지막으로 4절에서는 결론 및 향후연구를 서술한다.

2. 영화평 평점 예측 시스템

제안하는 영화평 평점 예측 시스템은 학습 코퍼스의 영화평 문장에서 영화 평점의 특징을 잘 나타내는 자질을 추출하고, 추출된 자질을 이용하여 다중 범주 SVM 모델을 기계학습 한다. 영화평의 평점을 SVM의 범주로 정의하여 1점부터 10점까지 1점 단위의 총 10개의 범주로 분류한다. 평점 범주에서 1점은 최악을, 10점은 최고를 의미한다.

제안 시스템은 어휘 자질, 감정 자질, 가치 자질 등을 사용한다. 각 자질은 다음의 2.1~4절에서 소개한다.

2.1 어휘 자질

어휘 자질은 영화평 문장에 출현한 어절, 음절, 형태소/품사 쌍 등을 의미한다. 이러한 어휘들의 출현 여부를 SVM의 학습 자질벡터로 사용한다. 다음의 표 1은 “어휴~ 박찬욱 이제 끝났네~”라는 영화평 문장에서 각각의 어휘자질을 추출하는 예를 나타내는 표이다.

표 1. 어휘자질의 추출 예

영화평 예	“어휴~ 박찬욱 이제 끝났네~”
어절	“어휴~” “박찬욱” “이제” “끝났네~”
음절	“어” “휴” “~” “박” “찬” “욱” “이” “제” ...
형태소/품사	“어휴/감탄사” “~/기호” “박찬욱/명사” “이제/관형사” ...

본 논문에서는 어휘자질만을 사용하는 영화평 평점 분류시스템을 Baseline 시스템으로 정한다. 각 어휘 자질별 실험 및 평가는 4절에서 보인다.

2.2 명시적 감정 자질

명시적 감정 자질은 평가자의 감정을 명시적으로 나타내는 표현을 의미한다. 제안 시스템에서 사용한 감정 자질은 표 2와 같이, 총 14가지의 감정 분류로 구분된다.

표 2. 명시적 감정 분류

명시적 감정 분류	의미	명시적 감정표현 예
SAD	슬픔	“눈물만 쏟아졌다”, “불행하게도”
HAPPY	기쁨	“감동 받아서”, “기뻐어”
AFFECTION	그리움	“아련하다지만”, “그리웠어”
FEAR	공포	“겁이 나긴 하는데”, “끔찍하더라”
SURPRISE	놀람	“충격적이야”, “깜짝 놀랐어”
REGRET	미안함	“용서해주라”, “죄송해요”
ENVY	부러움	“부러웠던”, “질투난다”
SHAME	부끄러움	“쑥스럽군”, “창피하다”
HATE	싫음	“고통스럽게”, “근란해요”
POKERFACE	무표정	“허무하더군요”, “심심하시겠어요”
UNHAPPY	불쾌	“기분더럽네”, “지저분하네”
ANGRY	화남	“나쁜 놈들이”, “성질을 건드렸다”
SULKY	삐짐	“치사한 놈”, “섭섭해라”
LOVELY	사랑	“사랑한다면”, “애교도 부리고”

다음의 표 3은 영화평 문장에서 명시적 감정 분류를 찾은 예이다.

표 3. 명시적 감정 표현 예

예문	명시적 감정표현	평점
“1점도 주기 힘들다. 살다 살다 이렇게 고통스럽게 본 영화는 처음이다.”	“힘들다”→“HATE” “고통스럽게”→“HATE”	1점
“난 재미없는 영화 볼때 담배 생각이 나는데 1시간 넘으면서 담배 생각이 절실 하더이다.”	“재미없는” →“POKERFACE”	4점
“신기한장르... 전 재밌었어 요!!”	“신기”→“SURPRISE” “재밌었어”→“HAPPY”	10점

제안 시스템은 명시적 감정 분류를 SVM의 학습자질로 사용한다. 그러나 명시적 감정 분류를 나타내는 표현들은 각각의 활용형이 매우 다양하기 때문에, 입력 영화평 문장내의 명시적 감정 표현을 찾는 것은 어렵다. 이러한 문제를 해결하기 위해 온라인 커뮤니티 등을 대상으로 수집한 감정표현들을 이용하여 “감정패턴 사전”과 “구어 및 속어 감정사전”을 구축하였다. 감정패턴 사전은 감정 표현의 형태소 분석결과에서 핵심적인 의미를 갖는 형태소/품사들만을 나열한 형태이며, 933개의 패턴을 포함한다. 구어 및 속어 감정사전은 형태소분석이 불가능한 914개의 감정표현들을 포함한다. 다음의 그림 1에서는 명시적 감정 분류의 추출 과정을 보인다.

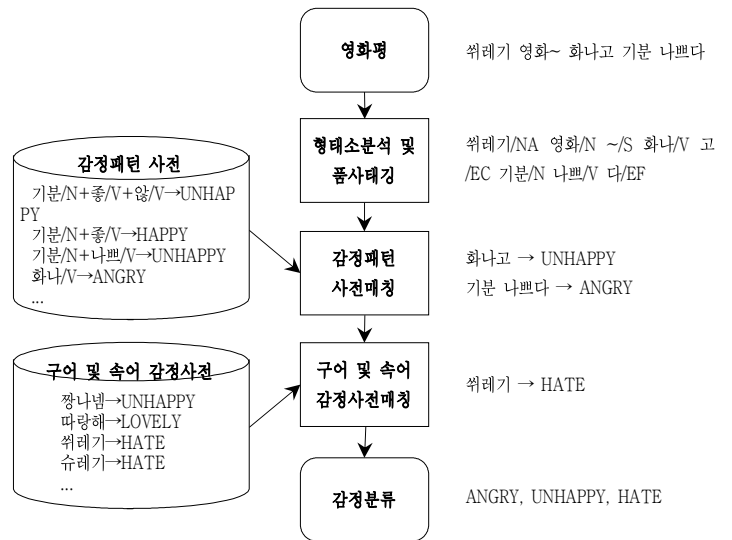


그림 1의 예처럼, 영화평의 형태소 분석 결과를 이용하여 감정패턴사전을 검색하여 “화나고”와 “기분나쁘다”에 대응하는 “ANGRY”와 “UNHAPPY”라는 명시적 감정 분류를 찾고, 구어 및 속어 감정사전을 검색하여 “썩레기”에 대응하는 “HATE”의 명시적 감정 분류를 찾는다. 추출된 명시적 감정 분류 태그를 SVM의 자질로 사용한다.

2.3 목시적 감정 자질

목시적 가치 자질은 명시적 감정 자질처럼 사용자의 감정을 명시적으로 표현하지는 않지만, 영화에 대한 선호도를 관용적인 어구를 통해 표현하는 것을 의미한다. 영화를 선호하는 관용적 표현이 있는 경우는 “LIKE”로, 선호하지 않은 관용적 표현이 있는 경우에는 “DISLIKE”로 나타낸다. 다음의 표 4는 영화평 문장에서 목시적 감정 자질을 추출한 예이다.

표 4. 목시적 감정 표현 예

예문	목시적 감정표현	평점
“보지마라..”	“보지마라”→“DISLIKE”	1점
“최악이다 돈날렸다. 영화로서 갖춰야할 요소중 하나인 흥미가 없다.”	“돈날렸다”→“DISLIKE”	2점
“시간가는 줄 모르고 볼 정도로 몰입도는 최고였으나 스토리는 기대 이하.”	“시간가는 줄 모르고” →“LIKE”	7점
“사람의 이중성 배우들의 명연기 볼 만한 영화 꼭 보세요.”	“볼만한 영화” →“LIKE”, “보세요”→“LIKE”	10점

목시적 감정분류는 명시적 감정분류와 동일한 방법으로 추출하며, 목시적 감정사전 역시 온라인 커뮤니티 등에서 목시적 감정표현을 수집하여 구축하였다. 목시적 감정패턴 사전은 101개의 패턴으로 구성하였으며, 구어 및 속어 감정사전은 19개로 구성하였다.

2.4 기타 자질

본 논문에서 제안한 시스템은 2.1~3절에서 언급한 자질들 이외에 단어의 초성열과 띄어쓰기 횟수를 자질로 사용한다.

평점자가 영화에 대해 비판적인 경우 비속어를 사용하며, 비속어는 초성열로 표현하는 경향이 있다. 따라서 제안시스템은 단어의 초성열을 자질로 사용한다. 또한 영화에 비판적인 평점자는 영화평을 작성할 때 띄어쓰기를 잘 하지 않은 경향이 있으므로 띄어쓰기 횟수가 5회 이하 또는 6회 이상인지 판단하여 자질로 사용한다. 표 5는 비속어 또는 띄어쓰기 횟수에 따른 영화평과 평점을 보여주는 예이다.

3. 실험 및 평가

본 연구에서는 제안 시스템의 기계 학습 및 평점 범주 할당을 위해 선형커널을 이용한 LIBSVM[7]을 사용하였다. 실험을 위해 사용한 데이터는 Naver lab에서 제공한

표 5. 기타 자질의 추출 예

예문	추출 자질	평점
“꽁자로보여준다고해서봤는데 안보는건데... 사비 박찬욱너이런사람이었니?”	초성열: “사비” 띄어쓰기: 3회	1점
“송강호는송강호다 김옥빈짱이다 박찬욱은ㄱㅈㅈ”	초성열: “ㄱㅈㅈ” 띄어쓰기: 2회	1점
“모든 배우들의 열연이 빛나는 영화- 대중적이진 않지만 사람을 매혹하는 무언가가있다 ”	초성열 없음 띄어쓰기: 9회	10점
“보는 내내 입벌리고 보았다. 슬픔,절망,광기를 아는 자라면 10점을 안줄수 없을 것.”	초성열 없음 띄어쓰기: 10회	10점

영화 <박쥐>와 <해운대>의 40자평 데이터 세트이다[8]. 영화의 평은 총 2만 개이며 각 영화평에 1~10점의 평점이 부여되어 있다. 각 평점별 영화평의 개수에 대한 학습 및 평가 데이터의 구성은 다음의 표 6에서 보인다.

표 6 영화평 데이터 평점 구성

평점	전체 데이터	학습 데이터	평가 데이터
1	4,556	3,645	911
2	628	502	126
3	432	346	86
4	696	557	139
5	754	603	151
6	969	775	194
7	844	675	169
8	1,312	1,050	262
9	1,150	920	230
10	8,659	6,927	1,732
합	20,000	16,000	4,000

평점을 부여하는 기준은 평점자 개인에 따라 다르기 때문에, 전문가 집단에 의한 평가를 하더라도 각 전문가의 평점은 서로 일치하지 않기 때문에 일반적으로 사용하는 척도들은 영화평 평점에는 적합하지 않다. 이러한 상황을 고려하여, 본 연구에서는 실험 척도를 아래의 식 (1)과 (2)로 제안하여 평가하였다.

$$\text{평균평점자} = \frac{\text{시스템이 제안한 평점과 실제 평점 차의 합}}{\text{영화평의 수}} \quad (1)$$

$$\text{정확도} = \frac{\text{시스템이 제안한 평점과 실제 평점 차가 1이내인 영화평의 수}}{\text{영화평의 수}}$$

(2)

식 (1)의 평균평점차는 시스템이 제안한 평점과 실제 평점들의 차의 평균을 계산하는 식이고, 식 (2)의 정확도는 시스템이 제안한 평점과 실제 평점의 차가 1 이내인 경우를 정답으로 인정하여, 시스템이 정답을 제시하는 비율을 계산하는 식이다.

본 논문에서는 가장 적합한 어휘 자질의 조합을 찾기 위해, 여러 가지 형태의 어휘 자질들을 비교 실험하였다. 다음의 표 7에서는 대표적인 어휘 자질들의 조합을 사용했을 때의 평균평점차와 정확도를 나타낸다.

표 7. 각 어휘 자질 조합별 실험 결과

어휘 자질 조합	평균평점차	정확도
어절 Unigram	4.32	37.36%
어절 Uni~Bigram	3.68	45.82%
어절 Uni~Trigram	4.42	37.36%
음절 Unigram	3.27	50.04%
음절 Uni~Bigram	2.92	54.12%
음절 Uni~Trigram	3.07	52.16%
형태소/품사 Unigram	2.93	53.93%
형태소/품사 Uni~Bigram	3.34	49.31%
음절Uni~Bigram+형태소/품사Unigram	2.92	54.14%
어절Uni~Bigram+음절Uni~Bigram	3.17	46.26%
어절Uni~Bigram+음절Uni~Trigram	3.68	45.52%
어절Uni~Trigram+형태소/품사Unigram	3.49	47.68%

표 7의 실험결과에서 볼 수 있듯이, 형태소/품사 쌍의 Unigram과 음절의 Unigram과 Bigram을 어휘 자질로 사용했을 때가 가장 높은 성능을 보인다. 일반적으로 의미를 갖는 가장 작은 단위인 형태소를 어휘자질로 사용할 경우가 성능이 가장 높은 것이 자연스런 현상이다. 그러나 이와는 다르게 음절 N-gram의 성능이 높은 이유는 데이터의 특성상 표준어를 사용하지 않는 경우가 많아 형태소분석 및 품사부착의 성능이 좋지 않기 때문이다.

본 논문에서는 형태소/품사 쌍의 Unigram과 음절의 Unigram과 Bigram을 사용하는 시스템을 Baseline 시스템으로 정하였다. 다음의 표 8에서는 명시적 감정 자질, 묵시적 감정 자질 및 기타 자질을 추가한 제안 시스템의 성능을 보인다.

표 8. 각 자질 조합별 실험 결과

각 자질 조합	평균평점차	정확도
Baseline	2.92	54.14%
Baseline+명시적감정	2.22 (-0.70)	63.02% (+ 8.88%)
Baseline+묵시적감정	2.53 (-0.39)	56.98% (+ 2.84%)
Baseline+기타	2.88 (-0.04)	55.46% (+ 1.32%)
Baseline+명시적+묵시적감정	2.04 (-0.88)	64.99% (+10.85%)
Baseline+ALL	1.98 (-0.94)	66.67% (+12.53%)

표 8에서 볼 수 있듯이 어휘 자질, 명시적 감정 자질, 묵시적 감정 자질, 기타 자질을 모두 사용한 제안 시스템이 가장 높은 성능으로 영화 평점을 예측하는 것을 알 수 있다.

3. 결론

본 논문에서 제안한 시스템은 자연어 형태의 영화평 문장에서 어휘 자질, 명시적 감정 자질, 묵시적 감정 자질 등을 추출하여 자동으로 평점을 예측하는 시스템을 구현하였다. 제안한 시스템은 1.98의 평균평점차와 66.67%의 정확도를 보였으며, Baseline시스템에 비하여 0.94의 평균평점차의 감소 및 12.53%의 정확도 향상을 보였다.

온라인 영화평의 특성상 평점자들은 맞춤법에 구애받지 않고 자유롭게 의견을 표현하는 경우가 많다. 따라서 일반적인 형태소 분석 및 품사 부착기로는 분석이 가능하지 않거나 분석 결과에 오류가 포함된 경우가 많아 평점 예측 시스템의 성능 향상이 매우 어렵다. 향후, 형태소 분석이 가능하지 않은 비표준어에 적합한 자질을 연구하여, 성능 향상을 기대하고 있다.

또한, 제안한 영화평 평점 시스템은 한국어 영화평만을 대상으로 하였으나, 차후 한국어 이외의 외국어에도 적용할 수 있도록 연구할 예정이다.

* 본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업(NIPA-2011-(C1090-1131-0008))의 연구결과로 수행되었음.

참고문헌

- [1] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1-2, pp. 1-135, 2008.
- [2] P. D. Turney and M. L. Littman, "Measuring Praise and Criticism: Inference of Semantic Orientation from Association," ACM Transactions on Information Systems, vol. 21, no. 4, pp. 315-346, 2003.
- [3] J. Kamps, M. Marx, R. J. Mokken, and M. D. Rijke, "Using WordNet to Measure Semantic Orientation of Adjectives," In Proceedings of the

- 4th International Conference on Language Resources and Evaluation, pp. 1115-1118, 2004.
- [4] 김상도, 박성배, 박세영, 이상조, 김권양, "음절 커널 기반 영화평 감성 분류," 한국지능시스템학회 논문지, 제20권, 제2호, pp. 202-207, 2010.
- [5] G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, and K. Miller, "WordNet: An Online Lexical Database," International Journal of Lexicograph, vol. 3, no. 4, pp. 235-244, 1995.
- [6] J. Weston and C. Watkins, "Support Vector Machines for Multiclass Pattern Recognition," In Proceedings of the 7th European Symposium on Artificial Neural Networks, pp. 219-224, 1999.
- [7] <http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html>
- [8] <http://lab.naver.com/research/>