

영화 리뷰 감성분석을 위한 텍스트 마이닝 기반 감성 분류기 구축*

김유영

연세대학교 문헌정보학과
(yuyoungkim@yonsei.ac.kr)

송 민

연세대학교 문헌정보학과
(min.song@yonsei.ac.kr)

누구나 본인이 사용한 제품이나, 이용한 서비스에 대한 후기를 자유롭게 인터넷에 작성할 수 있고, 이러한 데이터의 양은 점점 더 많아지고 있다. 감성분석은 사용자가 생성한 온라인 텍스트 속에 내포된 감성 및 감정을 식별하기 위해 사용된다. 본 연구는 다양한 데이터 도메인 중 영화 리뷰를 분석 대상으로 한다. 영화 리뷰를 이용한 기존 연구에서는 종종 리뷰 평점을 관객의 감성으로 동일시하여 감성분석에 이용한다. 그러나 리뷰 내용과 평점의 실제적 극성 정도가 항상 일치하는 것은 아니기 때문에 연구의 정확성에 한계가 발생할 수 있다. 이에 본 연구에서는 기계학습 기반의 감성 분류기를 구축하고, 이를 통해 리뷰의 감성점수를 산출하여 리뷰에서 나타나는 감성의 수치화를 목표로 한다. 나아가 산출된 감성점수를 이용하여 리뷰와 영화 흥행 간의 연관성을 살펴보았다. 감성분석 모델은 지지벡터 분류기와 신경망을 이용해 구축되었고, 총 1만 건의 영화 리뷰를 학습용 데이터로 하였다. 감성분석은 총 175편의 영화에 대한 1,258,538개의 리뷰에 적용하였다. 리뷰의 평점과 흥행, 그리고 감성점수와 흥행과의 연관성은 상관분석을 통해 살펴보았고, t -검정으로 두 지표의 평균차를 비교하여 감성점수의 활용성을 검증하였다. 연구 결과, 본 연구에서 제시하는 모델 구축 방법은 나이브 베이즈 분류기로 구축한 모델보다 높은 정확성을 보였다. 상관분석 결과로는, 영화의 주간 평균 평점과 관객 수 간의 유의미한 양의 상관관계가 나타났고, 감성점수와 관객 수 간의 상관분석에서도 유사한 결과가 도출되었다. 이에 두 지표 간의 평균을 이용한 t -검정을 수행하고, 이를 바탕으로 산출한 감성점수를 리뷰 평점의 역할을 할 수 있는 지표로써 활용 가능성을 검증하였다. 나아가 검증된 결론을 근거로, 트위터에서 영화를 언급한 트윗을 수집하여 감성분석을 적용한 결과를 살펴봄으로써 감성분석 모델의 활용 방안을 모색하였다. 전체적 실험 및 검증의 과정을 통해 본 연구는 감성분석 연구에 있어 개선된 감성 분류 방법을 제시할 수 있음을 보였고, 이러한 점에서 연구의 의의가 있다.

주제어 : 감성 분석, 감성 분류, 영화 리뷰 분석

논문접수일 : 2016년 7월 28일 논문수정일 : 2016년 9월 6일 게재확정일 : 2016년 9월 7일
원고유형 : 일반논문 교신저자 : 송 민

1. 서론

누구나 제품 및 서비스를 이용하고 그 평가를 인터넷에 자유롭게 남길 수 있으며, 회사 및 조

직은 이러한 고객의 리뷰, 즉 피드백을 이용하여 이익을 창출하기 위해 리뷰 데이터를 다방면으로 분석하는 시대이다. 오피니언 마이닝 혹은 감성분석은 텍스트 마이닝 기법 중 하나로, 어떠한

* 이 논문은 2015년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2015S1A3A2046711).
이 논문은 2015학년도 연세대학교 미래선도연구사업의 부분적인 지원에 의하여 작성된 것임(2015-22-0119).

사용자가 생성한 온라인 텍스트 속에 담긴 감성(sentiment), 정서(affect), 주관(subjectivity), 또는 감정(emotion)을 식별하기 위해 사용된다(Chen and Zimbra, 2010). 즉, 감성분석을 통해 이용자가 제품 또는 서비스로부터 어떠한 느낌을 받았는지 파악하기 위한 것이다. 감성분석 연구의 관련 기법은 2000년대 이후로 활발히 연구되고 있는데, 특히 이러한 감성분석 연구가 급증한 주된 이유는 소셜 미디어의 확산에 있다(Liu 2012; Appel et al. 2015).

감성분석은 포럼, 블로그, 마이크로 블로그, 소셜 네트워크 사이트 등에서 발생하는 텍스트에 적용이 가능하다. 이들은 주로 웹에서의 구전 혹은 온라인 구전(e-word of mouth, eWOM)으로 일컬어지는 것으로, 마케팅 분야에서 특히 중요한 데이터로써 활용되는 추세이다(Chevalier and Mavzlin, 2006; Chen and Xie, 2008; Duan et al., 2008; Ghose and Ipeiritos, 2011; Cui et al., 2012; Pagano and Maalej, 2013). 그 중에서도 활발히 사용되는 데이터 도메인은 제품, 서비스, 영화 등에 대한 리뷰이다. 본 연구에서는 이러한 주요 분석 도메인 중 하나인 영화 리뷰를 감성분석의 대상으로 활용하였다.

영화 리뷰는 감성분석 연구에서 자주 활용되는, 특정 대상(각 영화)에 대한 의견 분석이 가능한 데이터이다. 왜냐하면 리뷰는 영화를 먼저 본 사람들의 평가로써 아직 영화를 보지 않은 사람들에게 큰 영향을 미칠 수 있기 때문이다(Neelamegham and Chintagunta, 1999). 또한 영화 리뷰 데이터는 이용자가 직접 리뷰와 함께 평점(star rating)을 남길 수 있다는 특징이 있다. 따라서 평점과 감성을 동일시하여 관객의 영화에 대한 감성을 파악하는 등의 평점 활용이 가능하다. 그러나 평점이 언제나 감정을 대변하지는 않

며, 영화 리뷰와 평점 사이의 실제적 극성의 정도가 불일치하는 경우가 종종 발생한다는 점에서 연구의 정확성에 영향을 미친다는 한계가 존재한다(오연주, 채수환, 2015).

이에 본 연구는 영화 리뷰를 대상으로 텍스트 마이닝을 이용한 감성분석을 시도하였다. 특히, 기계학습의 한 방법인 지도학습(supervised learning)을 이용한 감성 분류기를 구축하여 리뷰의 감성을 파악하였다. 이때 감성은 하나의 리뷰가 내포하는 긍정 또는 부정적 성향을 감성점수(sentiment score)로써 표현한다. 그리고 감성분석의 결과로 도출된 감성점수와 관객의 직접적 평가인 리뷰 평점이 모두 영화의 흥행으로 대변되는 관객 수와 연관성을 보이는지에 대한 통계적 분석을 수행함으로써 영화의 흥행과 리뷰에서 나타나는 감성에 대한 상관관계를 보고자 한다. 마지막으로, 이를 바탕으로 향후 영화 리뷰 감성분석의 활용을 모색해 보았다. 이러한 전체적인 연구 과정을 통해 지도학습을 기반으로 하는 문헌 분류기를 복합적으로 이용하여 감성분석의 성능을 향상시킬 수 있는 방법론을 제안하고, 활용성을 밝히는 것이 본 연구의 목표이다. 이를 통해 향후 제품 또는 서비스 리뷰와 같은 다른 도메인의 데이터에도 이와 같은 방법론을 적용하여 감성분석을 시도할 수 있다는 점에서 본 연구의 의의가 있다.

본 연구는 다음과 같이 구성되어 있다. 제 2장에서는 감성분석 연구 방법에 대한 소개와 함께 영화 데이터를 이용한 국내의 감성분석 연구에 대하여 소개한다. 제 3장에서는 연구에서 활용한 데이터 및 모델 구축 과정 등을 포함한 연구 설계에 대해 설명한다. 이어지는 제 4장에서는 연구 결과를 기술하며, 이에 따른 결과 분석을 제시한다. 마지막으로 제 5장에서는 결론 및 한계

에 대하여 논의하고자 한다.

2. 관련연구

2.1 감성분석 연구

감성분석 연구는 컴퓨터 과학 분야에서 자연어 처리 중 하나의 주제로써 연구되기 시작하여 현재는 경영관리 분야의 학계 및 산업에서도 활발히 연구되고 있다. 이러한 연구 분야의 확장은 비즈니스 측면에서 사람들이 서비스 혹은 상품에 대해 어떠한 생각을 갖는지에 대해 이해함으로써 그들의 이윤을 극대화 시키고자 하는 동기가 있기 때문이다(Appel et al., 2015).

감성분석에서 중요히 여겨지는 언어의 구문론적, 형식적 부분은 대부분 촘스키의 1950-60년대 연구에 기반을 두고 있다(Appel et al., 2015). 그러나 실질적으로 감성분석 연구가 활발히 이루어지고 많은 발전을 보인 것은 2000년대 이후이다. Pang과 Lee(2008)는 감성분석에 접근하는 다양한 연구 방법들과 이론에 대해 정리했고, Liu(2012)의 정리를 통해 발전된 감성분석 방법론과 적용 방향에 대한 파악이 가능해졌다.

감성분석은 다양한 기계학습을 기반으로 이루어지고 있으며, 이때 기계학습은 크게 지도학습과 비지도학습(unsupervised learning) 기법에 바탕을 두는 것으로 나누어 볼 수 있다. 본 연구에서는 지도학습을 기반으로 하는 감성 분류기 구축을 목표로 하기 때문에 이러한 방법론과 관련된 연구들을 중심으로 파악해보면 다음과 같다.

Appel et al.(2015)의 연구에 의하면, 지도학습 방법은 감성분석에 관한 연구들 중 가장 활발히

사용되는 방법론이다. 학습 분류기의 유형으로는 나이브 베이즈 분류기(Naive Bayes classifier), 결정트리 분류기(Decision Tree), kNN 분류기(k-Nearest Neighbors), 신경망 분류기(Neural Network), 지지벡터 분류기(Support Vector Machine), 최대 엔트로피 모델(Maximum Entropy) 등이 있다. Pang et al.(2002)의 연구는 지도학습 방법을 기반으로 한 감성분석의 대표적인 연구로, 영화 리뷰 데이터를 학습 데이터로 이용한 연구이다. 리뷰 평점이 5점 만점의 4 또는 5점일 때를 긍정으로, 1~3점까지를 부정으로 나누어 학습하는데, 이때 학습 분류기를 나이브 베이즈, 최대 엔트로피, 지지벡터기 세 가지를 이용하여 결과를 비교하였다. 이 실험을 바탕으로 가장 성능이 높은 것은 지지벡터 분류기임을 발견하였다. 이 연구에서 자질을 개별 어휘 뭉치(bag of individual words)인 유니그램(uni-gram)으로 했다면, Mullen과 Collier(2004)는 유니그램을 원형복원(lemmatization)한 것을 추가적인 자질로 사용하였다. 분류 모델은 지지벡터기를 사용하였는데, 결과적으로 유니그램의 원형복원 자질을 추가했을 때 성능이 향상되었음을 알 수 있다.

이후에도 다양한 연구에서 지지벡터기를 이용한 감성분석을 시도하였다(König and Brill, 2006; Kennedy and Inkpen, 2006; Annett and Kondrak, 2008). 그런데 Ferguson et al.(2009)은 지지벡터기보다 다항 나이브 베이즈 분류기(Multinomial Naive Bayes)가 더 좋은 성능을 보였다는 점을 밝히며 금융 관련 블로그 글의 감성 분석을 시도하였다. 이 연구에서는 감성의 분류를 긍정·부정의 이진(binary) 뿐만 아니라 긍정·부정, 중립의 삼진(ternary)으로도 실험하였는데, 결과적으로는 긍정과 부정으로 감성분석을 할 때가 더 정확도

가 높았다. Pak과 Paroubek(2010)의 연구에서도 지지벡터기보다 다항 나이브 베이즈의 성능이 더 높았기 때문에 이 분류기를 사용했다. 학습 데이터로써 직접 사람이 감성을 분류한 트윗 데이터를 사용하였고, 확률 언어 모델인 n-gram과 품사 태깅(part-of-speech tagging, POS tagging)을 자질로서 이용했을 때 좋은 결과가 나왔음을 밝혔다.

위와 같이 분류기를 이용한 감성분석에서 나아가, 최근에는 딥 러닝(deep learning) 혹은 딥 뉴럴 네트워크(deep neural network)에 기반을 두는 감성분석 연구도 증가하는 추세이다(Glorot et al., 2011; Yanagimoto et al., 2013; Hu et al., 2015). 딥 뉴럴 네트워크를 기반으로 감성분석을 수행할 경우, 기존의 감성분석 방법에서 사용한 분류기에 입력(input)으로 사용되는 자질들의 차원 수가 많아졌을 때의 성능이 저하되는 것을 보완할 수 있다. 딥 러닝에서 사용하는 모델로는 Hu et al.(2015)의 연구에서 적용한 계층적 딥 뉴럴 네트워크 모델(hierarchical deep neural network), Yanagimoto et al.(2013)의 연구에서 적용한 제한된 볼츠만 기계(restricted boltzmann machine, RBM) 등이 있다. 이 두 연구에서는 모두 딥 러닝을 이용한 감성분석을 통해 기존에 사용되던 학습 분류기보다 더 나은 성능을 보였다.

2.2 영화 리뷰를 이용한 감성분석

앞서 언급한 바와 같이, 감성분석 연구는 컴퓨터 과학 분야의 한 주제로 시작되었다고 할 수 있다. 그러나 지금은 감성분석 연구에서 활용되는 데이터의 특성상 경영 관리 분야에서도 주목하는 연구 분야이다. 본 절에서는 이러한 점을 고려하여, 감성분석 연구에서 영화 리뷰 데이터

를 이용할 때 두 학문적 분야에서는 주로 어떠한 연구를 하는가에 대해 살펴보았다.

우선 컴퓨터과학 분야에서는 영화 리뷰에 있는 평점을 적극 활용하고 있다. 북미권에서의 연구는 대부분 영화 리뷰 데이터를 미국 최대 규모의 영화 데이터베이스인 Internet Movie Database (IMDb, www.imdb.com) 또는 영화 리뷰 전문 사이트 Rotten Tomatoes (www.rottentomatoes.com)에서 수집하는데, 이 두 사이트 모두 관객이 직접 리뷰와 평점을 남길 수 있는 구조를 가진다. 따라서 영화 리뷰를 이용한다면 자동적으로 평점에 대한 정보를 함께 활용할 수 있어 수동적 방법으로 긍정 또는 부정에 대한 분류를 하지 않아도 되기 때문에 지도학습을 이용한 리뷰 데이터 감성분석 모델 개발에서 주로 활용된다(Pang et al., 2002; Airoidi et al., 2004; Melville et al., 2009). 또한 이러한 감성분석 모델 성능 향상을 넘어 평점 추론을 통한 영화 추천이 가능한 모델 연구도 다양하게 등장하고 있다(Jakob et al., 2009; Sarvabhotla et al., 2010; Seroussi et al., 2011).

컴퓨터 과학 분야에서 감성분석 모델 개발 및 평점 추론 등을 중점적으로 연구한다면, 경영 분야에서는 리뷰의 감성을 이용하는 어플리케이션 측면으로 연구가 확장된다. 영화의 상업적 측면에서 이러한 영화 리뷰의 감성 분석을 수익 예측에 활용하는 것이다. Dellarocas et al.(2005)의 연구를 통해 알 수 있듯이 수익 예측 연구의 초기에는 영화의 평점과 평점의 양을 이용하여 영화의 수입을 예측하는 연구가 주를 이루었다. 특히, 개봉일을 포함한 개봉 첫 주의 영화 수입 예측에 있어서 평점이 중요한 요인이 됨을 밝혔다. 이는 텍스트 마이닝을 이용한 연구라고 볼 수는 없으나, 관객의 영화 평가가 수익 예측으로 연결된다

는 점을 시사한다. Krauss et al.(2008)의 연구에서는 영화 데이터베이스(IMDb)에서 개별 영화를 토픽으로 삼아 이에 대해 커뮤니케이션의 강도를 측정된 것을 하나의 지표로 만들고, 실질적으로 텍스트 내에서 나타나는 긍정적인 정도를 지표화 하여 선정된 25개의 영화가 각각 갖는 값을 통해 영화의 흥행과 아카데미상(Academy awards)의 후보를 예측하는 연구를 수행하기도 하였다. 이 역시 리뷰의 감성 전체를 파악하는 연구는 아니었지만, 긍정정도 지표(positivity index)를 산출했으며 이를 통해 영화의 흥행을 예측할 수 있다는 유의미한 결과를 보여주었다. Asur과 Huberman(2010)은 트윗(tweet)에서 등장한 영화 리뷰의 감성을 긍정, 부정, 중립으로 나누었고, 이 역시 영화의 흥행에 영향을 미치는 요소로 사용될 수 있음을 입증하였다.

국외에서 이렇게 영화 리뷰를 감성분석 하여 흥행 예측 변수로서 활용될 수 있다는 연구가 많이 등장하였으나, 국내 연구는 사실상 미진한 상황이다. 국내에서는 주로 포털의 평점과 리뷰의 양을 이용하는 흥행 예측에 관한 연구가 이루어졌다(김연형, 홍정한, 2011; 박승현 외., 2011; 박승현, 송현주, 2012). 최근에는 조승연 외(2014)의 연구에서 1~10점까지의 평점 중 8점 이상을 받은 긍정적 감성을 가진 리뷰만을 이용하여, 긍정이라는 감정을 나타내는 리뷰에서의 관객이 언급하는 속성을 고려한 흥행성과를 예측하는 모델을 개발하였다. 그리고 허민희 외(2013)는 리뷰에서 자주 관찰되는 구절을 추출한 후, 구절들 간의 공통된 문법적 구조를 파악하고, 이에 맞는 구절을 전체 데이터에서 추출하여 각 구절에 있던 평점들의 평균 및 분산으로 긍정 또는 부정 정도를 산출하는 새로운 방법론을 제시하였다. 그리고 이 방법론을 이용하여 관객 수 예

측에 활용하기도 하였다.

본 연구는 이와 같이 온라인 영화 리뷰를 이용한 감성분석 연구가 부족하고, 평점의 흥행에 대한 영향력 역시 지속적 검증이 필요하다는 점(박승현 외., 2011; 김상호, 한진만, 2014)을 바탕으로 영화 리뷰의 평점과 흥행의 상관관계에 대한 재검증을 하고자 한다. 또한, 지도학습 기반의 영화 리뷰 감성분석으로 영화 리뷰에서 드러나는 감성과 흥행의 연관성에 대해 알아보려 한다.

3. 연구 설계 및 방법

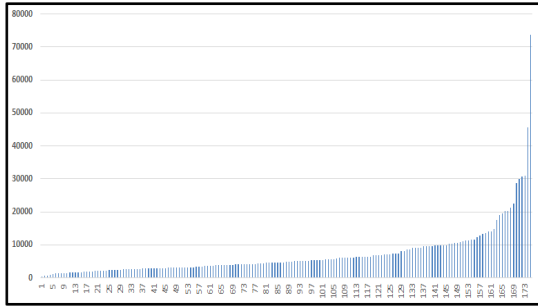
3.1 데이터 컬렉션

본 연구에서는 감성분석 모델 구축을 위해 지도학습 방법을 이용하기 때문에 학습용 데이터와 실험 데이터가 필요하다. 이에 국내 최대 포털 네이버에서 운영 중인 사이트 ‘네이버 영화(movie.naver.com)’에서 데이터를 수집하였다.

자바(Java) 기반의 크롤러를 이용하여 수집한 데이터는 총 175편의 영화에 대한 영화 리뷰와 평점이며, 이때 각 영화의 개봉 후 3주간의 데이터만을 수집 대상으로 하였다. 기간을 3주간으로 한정된 것은 박승철과 송현주(2012)의 연구에서 영화 리뷰가 가장 활발하게 나타나는 기간이 개봉 후 2 ~ 3주 동안이라고 밝힌 것에 기인한다. 영화는 한국영화진흥위원회에서 발표한 2013~2015년의 박스오피스 자료를 기준으로 각 해의 상업영화 중 국내영화 1위부터 30위, 해외영화 1위부터 30위까지에 랭크된 영화를 대상으로 중복된 영화를 제외한 총 175편으로 선정하였다.

이렇게 선정된 영화의 리뷰로 총 1,258,538개

가 수집되었다. 이 리뷰를 실험용 데이터와 모델 구축을 위한 데이터로 사용하였다.

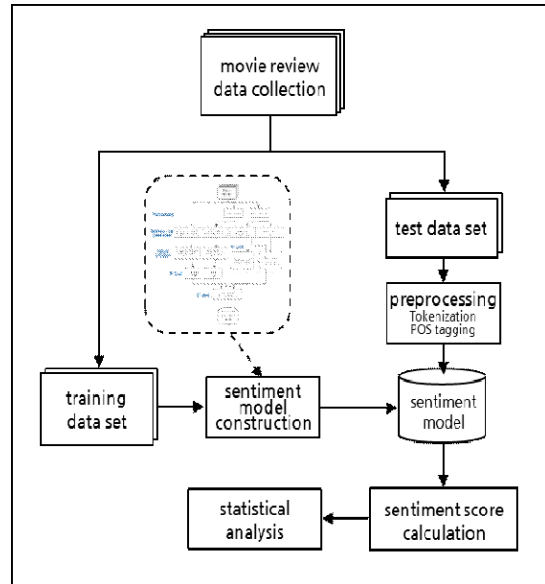


<Figure 1> Data collection

<Figure 1>은 각 영화에 대한 리뷰 수이다. 그림의 X축은 각 영화를, Y축은 리뷰 수를 의미한다. 총 175편의 영화 중 145편의 영화에는 리뷰의 수가 1만 개 이하로 나타났다. 그러나 그림에서도 알 수 있듯이, 몇 편의 영화는 극단적으로 많은 개수의 리뷰를 보유하고 있었다. 영화 <변호인(2013)>이 가장 많은 73,815개의 리뷰를 보유하고 있는 것으로 나타났고, 영화 <개구쟁이 스머프 2 (2013)>가 279개로 가장 적은 리뷰 수를 보였다. 리뷰 수의 중앙값은 4768이다.

3.2 연구 설계

연구의 순서는 <Figure 2>와 같다. 앞서 언급한 바와 같은 데이터 수집 단계를 마치고, 수집된 데이터 중 모델 구축을 위한 10,000개의 데이터를 이용하여 감성분석 모델을 구축하였다. 감성분석 모델은 기계학습 기반으로 이루어진다. 10,000개의 데이터에서 9,000개는 학습용 데이터로, 나머지 1,000개의 데이터는 모델 성능평가용으로 이용하였다.



<Figure 2> Research process

실험용 영화 리뷰가 되는 약 120만 개의 리뷰는 우선 전처리(preprocessing) 과정을 거치게 되는데, 이때 전처리는 문장 내의 공백을 구분자로 하여 문장을 분할하는 토큰화(tokenization)와 각 단어에 해당되는 품사를 결정하는 품사 태깅으로 이루어진다.

전처리 후의 리뷰는 구축된 감성분석 모델을 통해 감성점수를 산출(sentiment score calculation)하였다. 다음으로 이렇게 구해진 리뷰의 감성점수와 관객의 평점이 각각 흥행으로 대변되는 관객 수와 어떠한 상관관계를 보이는지 통계적으로 분석(statistical analysis)하였다.

구체적으로, 통계적 방법 중 상관분석을 수행하며, 유의 수준은 $p = .05$ 를 기준으로 하여 이보다 작은 p-value에 대해서는 유의한 것으로 판단하였다. 두 지표의 평균 차 비교를 위해 독립표본 t-검정도 수행하였다.

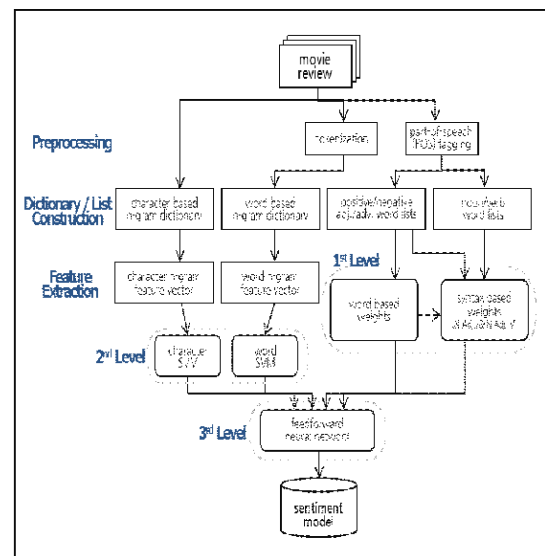
3.3 모델 구축 과정

감성분석 모델은 전처리, 사전 및 어휘 리스트 구축, 자질 추출, 그리고 3개 단계의 기계학습을 이용한 분류 과정을 거쳐서 최종적으로 구축된다. 기계학습의 3개 단계 분류는 Amplayo와 Occidental(2015)의 연구에서 활용한 방법론인 Multi-level Classifier를 응용하였다. 본 연구에서의 전체 구축 과정은 짧은 텍스트에 적합하도록 알고리즘을 작성하였다.

모델 생성의 첫 번째 과정은 전처리이다. 텍스트 전처리 단계는 토큰화와 품사 태깅으로 나뉘며, 이 두 과정에서 한글 형태소 분석기가 필요하여 소셜 미디어에서 생성되는 짧은 텍스트 처리에 적합한 한글 형태소 분석기인 Twitter-Korean-text (<http://github.com/twitter/twitter-korean-text>)를 이용하였다.

다음 과정은 사전과 어휘 리스트 구축이다. 이 과정에서 구축되는 사전은 말뭉치(bag of words)와 같이 단어가 단순히 나열된 것이고, 단어 리스트는 단어의 출현 횟수를 기반으로 내림차순으로 정렬 되어있는 것을 의미한다. 전처리를 통해 품사 태깅을 거친 텍스트는 긍정 형용사(positive adj.), 긍정 부사(positive adv.), 부정 형용사(negative adj.), 부정 부사(negative adv.)의 단어 리스트를 각각 구축하도록 하였고, 명사(noun)와 동사(verb)의 단어 리스트 역시 구축하였다. 긍정과 부정을 나눌 때에는 리뷰 데이터의 평점 정보를 활용하여, 긍정으로 분류될 수 있는 점수와 부정으로 분류될 수 있는 점수를 정해주었다. 토큰화를 거친 텍스트는 이 과정에서 단어 기반의 n-gram 사전을 구축하도록 하였다. 사전 구축 과정에서는 출현 빈도가 일정 수치 이하인 단어를 제거하도록 한다. 형태소 분석기를 이용

하게 되면 철자 오류에 민감하다는 단점이 있기 때문에(신준수, 김학수, 2010) 이를 고려하여 토큰화와 품사 태깅 없이 처음의 입력 데이터를 바탕으로 글자(character) 기반의 n-gram 사전을 구축하는 과정도 사전 구축 단계에 포함하였다. 글자 기반 n-gram 사전 역시 출현 빈도가 일정 수치 이하인 단어를 제거하도록 한다. 이렇게 사전과 어휘 리스트를 구축하는 과정에서 2개의 사전과 6개의 단어 리스트가 구축된다.



<Figure 3> Sentiment model construction process

구축된 글자 기반의 n-gram 사전과 단어 기반의 n-gram 사전은 각각 학습 문헌과의 비교과정을 통해 자질 벡터(feature vector)를 형성한다. 자질 벡터에는 많은 단어가 들어가 있기 때문에 차원 수가 높다. 따라서 차원을 축소하기 위해 각각의 자질 벡터를 감성 분류 연구에서 높은 성능을 보인 지지벡터 분류기(Pang et al., 2002; 신준수, 김학수, 2010) 감성 분류에 대한 학습을 하

여 다음 단계인 신경망 학습의 입력 데이터를 생성한다. 이 분류 과정은 <Figure 3>의 2nd Level로, 감성분석 모델의 3개의 분류 과정 중 2번째 과정에 속한다(Amplayo and Occidental, 2015).

생성된 단어 리스트는 <Figure 3>의 1st Level로 표현된 자연어 처리 과정에서 활용된다. 긍정 형용사, 긍정 부사, 부정 형용사, 부정 부사의 리스트는 각각 어휘 기반과 구문론 기반으로 단어에 가중치가 부여된다. 어휘 기반 가중치는 4가지 단어 리스트에서 출현 빈도가 높은 순으로 정렬된 것에 이를 바탕으로 가중치를 부여해 4개의 결과 파일을 생성한다. 이 때 가중치를 주기 위해 1이 아닌 정수로 2ⁿ을 활용했다. 이 가중치는 리스트에 포함된 전체 단어의 수를 10으로 나눈 것을 첫 집합에 속하는 단어의 개수로 정하고, $n = [(첫 집합의 단어 개수)/10]$ 으로 정하여 부여하는 것이다. 예를 들어, 리스트에 500개의 단어가 포함되어 있을 때, 첫 집합에는 50개의 단어가 속하게 되고, 첫 집합에는 2⁵이 가중치로 50개의 단어에 부여된다. 그리고 두 번째 집합에는 2⁴, 세 번째 집합에는 2³이 가중치로 부여된다. 이렇게 가중치를 부여하는 이유는 단어의 출현빈도를 기준으로 했을 때, 가장 높은 순위의 단어들에 가중치를 뚜렷하게 부여하기 위함이다. 다음으로는 구문론 기반 가중치로, 여기에서는 4가지 리스트에 명사와 동사의 리스트까지 함께 이용하여 (명사-형용사) 또는 (명사-형용사-동사)가 문장에서 나타날 때 역거리 공식(inverse distance)에 어휘 기반 가중치를 곱하는 방법으로 가중치를 부여한다. 문장에서 명사, 형용사, 동사가 출현하는 순서는 고려하지 않으며, 역거리를 이용하기 때문에 단어들이 가까울수록 가중치가 크게 반영되게 된다. 이와 같은 과정을 통해 1st Level에서는 최종적으로 2개의 결과 파일이 생

성된다.

이렇게 두 레벨에서 총 8개의 입력 파일이 신경망 분류기(neural network)를 이용한 분류 과정에서 사용된다. 이 과정은 3rd Level 부분으로, 본 연구에서는 인공 신경망(artificial neural network) 중 하나인 피드포워드 신경망(feedforward neural network, FNN)을 이용하며, 이를 학습시키기 위해 역전파(backpropagation) 알고리즘을 이용하였다. 역전파 알고리즘은 학습 데이터의 가중치들 각 노드(node)에 입력하고 분류 시 발생하는 오류를 이용하여 노드 간의 엣지(edge)의 가중치를 조정하는 것이다(정영미, 2012).

신경망 분류기의 과정을 거쳐 최종적으로 구축된 감성분석 모델은 리뷰에 대한 감성점수를 1과 10사이의 값으로 산출한다. 본 연구에서는 이 스케일을 0과 1사이의 값으로 조정하였다. 이 때 0에 가까울수록 부정적 감성, 1에 가까울수록 긍정적 감성으로 간주하여, 최종적으로는 0과 1의 이진 형태의 값을 통계 분석 과정에서 사용하였다.

4. 연구결과

4.1 감성분석 모델 구축

모델 구축 과정 중 임의로 임계치(threshold)를 설정해주어야 하는 부분에서는 다음과 같은 수로 지정하였다.

우선, 사전 구축 과정에서 글자(character) 기반의 n-gram 사전을 구축할 때에는 n을 4~8로 설정하고, 단어 기반의 n-gram 사전 구축 과정에서는 n을 1~3으로 설정하여 이에 해당하는 글자 및 단어가 사전의 단어로 포함될 수 있도록 하였

다. 그리고 어휘 리스트 구축 과정에서 긍정은 평점으로 7점 이상을, 부정은 4점 이하로 하여 이러한 평점에 속하는 리뷰들에서 어휘를 추출할 수 있도록 하였다.

<Table 1> Examples of word lists

Positive Adj.		Positive Adv.	
좋	시원한	특히	두근두근
재밌	멋있었	훨씬	꼬옥
괜찮	신선한	정말로	마니마니
즐겁	훌륭한	의외로	한껏
Negative Adj.		Negative Adv.	
재미없	딱히	그나마	아무리
아깝	낮은	강	쓸데없이
지루하고	거지같	그저	대놓고
뻘한	부족한	간간히	하나같이
Noun		Verb	
연기	유쾌	봤	즐길
재미	웃음	볼	나오는
코믹	짱	웃고	살렸
가족	스트레스	갔	추천함

이러한 기준을 바탕으로 품사 태깅을 거쳐 추출된 단어 리스트는 총 6개로, 긍정 형용사, 긍정 부사, 부정 형용사, 부정 부사, 그리고 명사와 동사 리스트가 있다. 이때 명사와 동사 리스트는 각각 최대 1,000개의 단어를 추출하도록 하였고, 나머지 리스트에 대해서는 최대 500개의 단어가 포함되도록 하였다. <Table 1>은 추출된 단어들의 예시이며, 이와 같이 실제 리뷰에는 띄어쓰기나 철자의 오류가 상당수 존재했고, 신조어와 같은 다양한 어휘들이 사용되었다. 그러나 글자 기반 n-gram 사전과 같은 다양한 자질을 이용했기 때문에, 다양한 스타일의 텍스트에 대한 감성분석이 가능했다.

다음은 모델의 성능평가로, 본 연구에서는 학습에서 사용된 리뷰와 중복되지 않는 1,000개의

리뷰를 이용하여 성능을 평가하였다. 이 1,000개의 리뷰에 있는 평점이 6~10점 사이이면 1(긍정), 1~5점 사이이면 0(부정)으로 분류하고, 감성분석 모델을 통해 얻어진 감성점수가 1(긍정) 또는 0(부정)으로 표현될 때 평점의 분류와 일치하는 정도를 통해 어느 정도 정확하게 감성을 판별하는지 평가하였다.

<Table 2> Comparison between two sentiment classifiers

Naive Bayes classifier		Proposed sentiment classifier	
Precision	95.2	Precision	95.5
Recall	81.0	Recall	88.2
F1-measure	87.5	F1-measure	91.7
Accuracy	79.9	Accuracy	86

모델의 성능은 <Table 2>에서 나타나는 바와 같다. 학습 분류기 중 하나인 나이브 베이즈 분류기로 구축한 모델은 정확성이 79.9%를 나타낸 반면, 본 연구에서 제안하는 분류 모델은 86%의 정확성을 보여 비교적 좋은 성능을 보임을 알 수 있다.

4.2 상관분석

영화 리뷰의 평점과 텍스트 분석을 통해 나타난 감성점수가 흥행과 어떠한 상관관계를 보이는지 알기위해 각각 통계 분석한 결과는 다음과 같다.

우선, 영화의 리뷰 평점과 흥행으로 대변되는 관객 수와의 상관관계를 보았다. 데이터의 기간이 개봉일 포함 총 21일임을 고려하여, 주간으로 나누어 각 주간의 평점과 관객 수간의 관계를 분석하도록 하였다.

<Table 3> Pearson coefficient of correlations between star rating and number of viewer (overall movies)

		star rating		
		week 1	week 2	week 3
number of viewer	week 1	.052		
	week 2	.262***	.304***	
	week 3	.361***	.399***	.413***

*** $p < .001$

분석 결과 첫 주에는 평점과 관객 수 간의 유의한 상관관계가 나타나지 않았으나, 첫 주의 평점과 둘째 주의 관객 수($r = .262, p < .001$), 둘째 주의 평점과 둘째 주의 관객 수($r = .304, p < .001$), 그리고 셋째 주의 평점과 셋째 주의 관객 수($r = .413, p < .001$) 간의 유의적인 양의 상관관계가 나타났다. 즉, 전체 175편의 영화로 볼 때, 평점과 관객 수 사이에는 시간이 지남에 따라 상관관계수가 커지면서 점차 뚜렷한 상관관계가 나타남을 알 수 있다.

전체 175편의 영화에 대한 평점과 관객의 상관분석 결과는 <Table 3>과 같다. 그런데 3주간의 집계된 관객 수를 비교해보았을 때 175편의 전체 영화에서 관객 수의 편차가 2,218,187로, 그 편차가 상당히 크게 나타난다. 이에 관객 수를 기준으로 흥행한 작품과 흥행에 실패한 비흥행 작품으로 나누어 각 분류에 해당하는 상관분석을 추가적으로 수행하였다. 흥행작과 비흥행작의 구분 기준은 선행 연구를 참고하여 100만 명 이상 관객 수를 흥행 성공의 최소 기준점으로 삼았다(이경제, 장우진, 2006; 김미현 외, 2010; 김연형, 홍정환, 2011).

위 기준으로 보았을 때, 본 연구에서 수집한 175편의 영화 중 총 31편의 영화가 비흥행작으로 분류되었다. 31편의 비흥행 영화들의 3주간

총 관객 수의 평균은 805,749로 나타났다. 나머지 144편의 영화들은 흥행 기준점에 의하면 흥행에는 성공한 영화이나, 3주간 총 관객 수가 가장 많은 영화는 1,500만 명 이상의 관객 수를 동원했다는 점으로 미루어 볼 때, 흥행 영화들 간에도 차이가 크다는 점을 감안하여 관객 수 기준 상위 30위까지의 영화만을 흥행 영화로 분류하여 분석하도록 하였다. 이렇게 선정된 흥행 영화 30편의 3주간 총 관객 수의 평균은 6,674,359으로, 평균적으로 500만 명 이상의 관객 수를 기록하고 있다.

<Table 4> Pearson coefficient of correlations between star rating and number of viewer (successful movies)

		star rating		
		week 1	week 2	week 3
number of viewer	week 1	-.141		
	week 2	.176	.265	
	week 3	.573**	.549**	.502**

** $p < .01$

<Table 4>와 <Table 5>는 흥행 영화의 평점과 관객 수의 상관관계수, 그리고 비흥행 영화의 평점과 관객 수의 상관관계수이다. 두 표를 비교해 보면, 평점과 관객 수의 상관관계는 영화를 흥행과 비흥행 여부로 나누어서 볼 때 그 차이가 확실히 드러난다.

<Table 4>와 같이, 흥행 영화는 3주차에만 평점과 관객 수 간의 유의적인 양의 상관관계가 뚜렷하게 나타났다($r = .502, p < .01$). 이에 비해, 비흥행 영화에서는 1주차의 평점과 관객 수에서 유의적인 음의 상관관계가 강하게 나타났다($r = -.571, p < .01$).

(Table 5) Pearson coefficient of correlations between star rating and number of viewer (unsuccessful movies)

		star rating		
		week 1	week 2	week 3
number of viewer	week 1	-.571**		
	week 2	.219	.129	
	week 3	.472**	.391*	.378*

* $p < .05$ ** $p < .01$

감성점수와 영화의 흥행 간의 상관관계는 다음과 같이 나타났다. 우선, 전체 영화의 주간 감성점수와 관객 수의 상관분석에서는 셋째 주의 감성점수와 둘째, 셋째 주의 관객 수 간의 유의적인 양의 상관관계가 나타났다($r = .224, p < .01$). 아래의 <Table 6>은 상관계수를 보여주며, 이러한 결과는 평점과 관객 수의 상관분석 결과와도 유사하다.

(Table 6) Pearson coefficient of correlations between sentiment score and number of viewer (overall movies)

		sentiment score		
		week 1	week 2	week 3
number of viewer	week 1	-.044		
	week 2	.137	.175*	
	week 3	.183*	.224**	.224**

* $p < .05$ ** $p < .01$

영화를 흥행과 비흥행 여부로 나누었을 때에도 평점에서의 결과와 유사하게 나타났다. <Table 7>과 <Table 8>에서 나타나는 바와 같이 흥행 영화에서는 3주차의 감성점수와 관객 수는 유의적인 양의 상관관계가 있었고($r = .373, p < .05$), 비흥행 영화에서는 1주차의 감성점수와 관객 수에서만 유의적인 음의 상관관계가 있는 것

으로 나타났다($r = -.438, p < .05$). 그 외의 2, 3주차의 감성점수와 관객 수 간의 상관관계에서는 유의적인 관계가 나타나지 않았다.

(Table 7) Pearson coefficient of correlations between sentiment score and number of viewer (successful movies)

		star rating		
		week 1	week 2	week 3
number of viewer	week 1	-.125		
	week 2	.218	.231	
	week 3	.420*	.409*	.373*

* $p < .05$

(Table 8) Pearson coefficient of correlations between sentiment score and number of viewer (unsuccessful movies)

		sentiment score		
		week 1	week 2	week 3
number of viewer	week 1	-.438*		
	week 2	.126	-.016	
	week 3	.318	.173	.239

* $p < .05$

4.3 감성분석 모델 활용 방안 제안

감성점수와 평점은 영화 리뷰를 주간으로 나누어 분석할 때와 영화를 흥행 및 비흥행으로 분류하여 분석할 때 모두 비슷한 패턴을 보임을 알 수 있었다. 이는 비록 감성점수가 평점만큼 관객 수와의 강한 상관을 보이는 것은 아니지만, 평점의 역할을 대신 할 수 있는 가능성을 보인다는 것으로 해석 가능하다. 이와 같은 해석이 가능함을 뒷받침하기 위해 두 지표 간의 t -검정을 수행하여 통계적으로 살펴본 결과를 다음의 <Table 9>로 정리하였다.

〈Table 9〉 An independent sample t-test result

indices		N	average	standard deviation	t
star rating groups	0	81	7.7594	.71500	-7.872***
	1	94	8.5080	.50682	

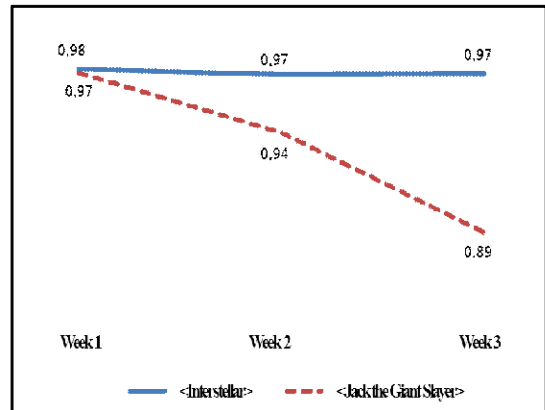
*** $p < .001$

검정대상의 0과 1은 감성점수의 평균인 8.1243보다 낮은 영화를 0으로, 높은 영화를 1로 하여 총 2개의 그룹을 지은 것이다. 그룹 0의 영화는 총 81편, 1의 영화는 총 94편이었고, 각 그룹의 평균은 7.7594와 8.5080으로 나타났다. 두 그룹에 각각 해당하는 영화들의 평점을 이용하여 t -검정을 한 결과, $t(175) = -7.872$, $p < .001$ 로 99.9% 유의수준에서 유의한 것으로 나타났다. 즉, 이러한 결과는 감성 점수의 평균을 기준으로 하여 영화들의 평점 간 평균을 비교하는 것이 유의미함을 의미한다.

t -검정까지의 결과를 통해 감성 점수가 평점의 역할을 할 수 있고, 이러한 비교들이 유의미했음을 알 수 있다. 이에, 구축한 감성분석 모델을 평점이 부여되지 않은 데이터인 트윗(tweet)에 적용하여 감성점수를 활용해 분석하였다.

수집한 트윗은 영화 <인터스텔라 (2014)>와 <잭 더 자이언트 킬러 (2013)>에 대한 것이다. <인터스텔라>는 2014년 11월 6일에 개봉해, 본 연구에서 지정한 3주의 기간 동안 7백만 명 이상의 관객 수를 동원하여 흥행 영화로 분류되었고, 최종적으로는 천만 관객을 동원한 것으로 집계된 영화이다. <잭 더 자이언트 킬러>는 2013년 2월 28일에 개봉하여, 3주 동안 100만 명 이하의 관객을 동원해 비흥행 영화로 분류되었다. 그리고 두 영화에 대해 각각 23,479개, 634개의 한글로 작성된 트윗이 수집되었다.

구축한 모델을 이용하여 감성점수를 산출하고, 각 영화의 주간 감성점수를 비교해본 결과는 다음의 <Figure 4>와 같다.



〈Figure 4〉 Sentiment scores of two movies

트위터 상에서 <인터스텔라>는 3주간 큰 변동 없이 트위터 상에서 호평을 받은 것으로 나타났다. 그러나 비흥행 영화였던 <잭 더 자이언트 킬러>는 첫 주에는 인터스텔라와 같이 긍정적으로 언급되었지만, 시간이 지날수록 감성점수가 하락세를 보였고, 이를 통해 부정적 평가가 이어졌음을 유추해볼 수 있다.

5. 결론

본 연구는 텍스트 마이닝 기반의 감성분석 모델을 구축하고, 이를 이용한 리뷰 내용의 감성분석을 목표로 하였다. 1만 건의 리뷰를 이용함으로써 평점만을 관객의 감성으로 판단하여 사용했을 때 발생하는 선행연구들이 가지는 한계를 극복하고자 하였고, 리뷰의 평점과 모델을 이용

하여 산출된 감성점수가 각각 영화의 흥행과 어떠한 관계가 있는지를 상관분석으로 살펴보았다. 나아가, 리뷰 평점과 감성점수의 평균 차이 비교를 통해 다른 텍스트 데이터로의 감성분석 방법 적용 가능성에 대해 살펴보았다.

연구 결과, 첫째로 평점과 관객 수의 상관관계를 3주간으로 각각 나누어 보았을 때, 시간이 지남에 따라 상관관계수가 커지면서 뚜렷하게 상관관계를 보였다. 관객 수를 기준으로 하여 영화를 흥행 영화와 비흥행 영화 두 집단으로 나누어 추가적으로 상관분석을 한 결과, 흥행 영화에서는 2, 3주차의 평점과 3주차의 관객 수 간의 양의 상관관계가 나타난 반면, 비흥행 영화에서는 1주차의 평점과 관객 수 간의 음의 상관관계가 나타났다. 다음으로, 감성분석 모델을 통해 산출한 감성점수로 관객 수 간의 상관분석을 한 결과는 평점과 관객 수의 상관분석 결과와 비슷했다. 이 점을 바탕으로 하여 감성점수가 리뷰 평점의 역할을 수행할 수 있을 것으로 판단하였고, 이러한 결론을 뒷받침하기 위해 t -검정을 수행하여 유의미한 결과를 도출하였다. 마지막으로, 본 연구에서 제안하는 감성분석 모델을 텍스트 내 감성은 드러나지만 감성 판별이 되지 않은 텍스트 데이터인 트윗에 실험적으로 적용하여 영화별로 도출되는 결과가 어떠한 차이를 보이는지 확인하였다.

이러한 연구과정을 통해 얻어진 실험 결과를 종합해보면, 본 연구에서는 지도학습을 이용한 문헌 분류 방법 중 나이브 베이즈 분류기나 지지 벡터기만을 단순히 감성 분류에 적용한 것이 아닌, 3단계로 구성되는 기계학습 방법을 적용했고, 그 결과 나이브 베이즈 분류기로 구축한 모델 성능에 비해 약 6% 정도의 문헌 분류 정확성 향상을 보였다. 이를 통해 텍스트 마이닝 기법

중 하나인 감성분석에 개선된 문헌 분류 방법을 이용하는 것을 제안할 수 있다는 점에서 의의가 있다. 뿐만 아니라, 기존 연구들에서 밝히지 않은 영화 리뷰의 감성과 흥행성 간의 상관관계를 통계적으로 밝히며, 영화를 흥행 작품과 비흥행 작품으로 구분하여 리뷰의 감성점수와 흥행의 상관관계에 차이가 있음을 새롭게 밝힘으로써 영화의 흥행성에 관한 연구에도 하나의 연구 방향을 제시했다고 할 수 있다.

이러한 연구의 의의가 있음에도 불구하고, 본 연구에서는 영화의 평점과 관객 수, 그리고 감성점수와 관객 수의 상관관계만을 비교 분석해 보았다는 점에서 한계가 있다. 이미 기존의 영화의 흥행과 관련된 연구에서는 스크린 수 등의 특징적인 몇 가지 변수가 영화의 흥행에 상당한 영향을 미친다는 점을 밝힌 바가 있다(박승현 외., 2011; 김상호, 한진만, 2014; Wyatt, 1991). 그리고 실제로 영화의 관객 수, 즉 흥행은 영화를 본 관객의 리뷰뿐만 아니라 스크린 수, 개봉 시점, 배우 및 감독 등의 다양한 변수의 영향을 받는다. 다만, 본 연구에서는 텍스트 마이닝 기반 감성분석 연구에 초점을 맞추므로써 이와 같은 다양한 변수를 고려한 연구가 이루어지지 못하였다. 따라서 평점과 감성점수가 흥행과 연관이 있음을 밝히는 것을 넘어, 다양한 변수들과 함께 고려했을 때 실제로 흥행에 영향을 미치는 관계로 나타나는지에 대한 추후 연구가 이루어진다면 본 연구의 한계점을 보완하며 더욱 완성도 높은 영화 흥행 관련 연구로 이어질 수 있을 것이다.

참고문헌(References)

- Airoidi, E., X. Bai and R. Padman, "Markov blankets and meta-heuristics search: sentiment extraction from unstructured texts," *Proceedings of International Workshop on Knowledge Discovery on the Web*, (2004), 167~187.
- Amplayo, R. K. and J. Occidental, "Multi-level classifier for the detection of insults in social media," *Proceedings of 15th Philippine Computing Science Congress*, (2015).
- Annett, M. and G. Kondrak, "A comparison of sentiment analysis techniques: Polarizing movie blogs," *Advances in Artificial Intelligence*, (2008), 25~35.
- Appel, O., F. Chiclana and J. Carter, "Main concepts, state of the art and future research questions in sentiment analysis," *Acta Polytechnica Hungarica*, Vol.12, No.3(2015), 87~108.
- Asur, S. and B. A. Huberman, "Predicting the future with social media. In Web Intelligence and Intelligent Agent Technology (WI-IAT)," *Proceedings of 2010 IEEE/WIC/ACM International Conference*, (2010), 492~299.
- Chen, Y. and J. Xie, 'Online consumer review: Word-of-mouth as a new element of marketing communication mix," *Management Science*, Vol.54, No.3(2008), 477~491.
- Chen, H. and D. Zimbra, "AI and opinion mining," *IEEE Intelligent Systems*, Vol.25, No.3(2010), 74~80.
- Chevalier, J. A. and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research*, Vol.43, No.3(2006), 345~354.
- Cui, G., H. K. Lui and X. Guo, "The effect of online consumer reviews on new product sales," *International Journal of Electronic Commerce*, Vol.17, No.1(2012), 39~58.
- Dellarocas, C., N. Awad and M. Zhang, "Using online ratings as a proxy of word-of-mouth in motion picture revenue forecasting," *Smith School of Business, University of Maryland*, 2005.
- Duan, W. and A. B. Whinston, "The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry," *Journal of Retailing*, Vol.84, No.2(2008), 233~242.
- Ferguson, P., N. O'Hare, M. Davy, A. Bermingham, P. Sheridan, C. Gurrin and A. F. Smeaton, "Exploring the use of paragraph-level annotations for sentiment analysis of financial blogs," *Proceedings of WOMAS 2009-Workshop on Opinion Mining and Sentiment Analysis*, (2009).
- Ghose, Al, and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *Knowledge and Data Engineering, IEEE Transactions on*, Vol.23, No.10(2011), 1498~1512.
- Glorot, X., A. Bordes and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," *Proceedings of the 28th International Conference on Machine Learning*, (2011), 513~520.
- Heo, M. H., P. S. Kang and S. Cho, "Predicting Box-office with Opinion mining reviews," *Proceedings of the Korean Operations and Management Science Society Conference*, (2013), 487~500.

- Hu, Z., W. Ding and X. Zheng, "Review sentiment analysis based on deep learning," *Proceedings of e-Business Engineering (ICEBE) 2015 IEEE 12th International Conference*, (2015), 87~94.
- Jakob, N., S. H. Weber, M. C. Müller and I. Gurevych, "Beyond the stars: exploiting free-text user reviews to improve the accuracy of movie recommendations," *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, (2009), 57~64.
- Jo, S. Y., H.-K. Kim, B. Kim and H.-W. Kim, "Predicting Movie Revenue by Online Review Mining : Using the Opening Week Online Review," *Information Systems Review*, Vol.16, No.3(2014), 113~134.
- Jung, Y., *Research in Information Retrieval*, revised edition, Yonsei University Press, 2012.
- Kennedy, A. and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters," *Computational intelligence*, Vol.22, No.2(2006), 110~125.
- Kim, M. H., S. E. Kim and Y. J. Choi, "The Determinants of Box-office Performance of Korean Films and Implications for Policies," *Film Studies*, No.46(2010), 31~56.
- Kim, Y. H. and J. H. Hong, "A Study for the Development of Motion Picture Box-office Prediction Model," *Communications for Statistical Applications and Methods*, Vol.18, No.6(2011), 859~869.
- Kim, S. H. and J. M. Han, "An Analysis of Motion Picture Box Office Performance : Focusing on Korean Movies Released in 2012," *Social Science Studies*, Vol.53, No.1 (2014), 191~214.
- König, A. C. and E. Brill, "Reducing the human overhead in text categorization," *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2006), 598~603.
- Krauss, J., S. Nann, D. Simon, P. A. Gloor and K. Fischbach, "Predicting Movie Success and Academy Awards through Sentiment and Social Network Analysis," In *ECIS*, (2008), 2026~2037.
- Lee, K. J. and W. J. Jang, "Predicting Financial Success of a Movie Using Bayesian Choice Model," *Proceedings of the Korean Operations and Management Science Society Conference*, (2006), 1428~1433.
- Liu, B., "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, Vol.5, No.1(2012), 1~167.
- Melville, P., W. Gryc and R. D. Lawrence, "Sentiment analysis of blogs by combining lexical knowledge with text classification," *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2009), 1275~1284.
- Mullen, T. and N. Collier, "Sentiment analysis using support vector machines with diverse information sources," *Proceedings of Empirical Methods in Natural Language Processing*, (2004), 412~418.
- Neelamegham, R. and P. Chintagunta, "A Bayesian model to forecast new product performance in domestic and international markets," *Marketing Science*, Vol.18, No.2(1999), 115~136.
- Oh, Y.-J. and S.-H. Chae, "Movie Rating Inference by Construction of Movie Sentiment Sentence

- using Movie comments and ratings,” *Journal of Internet Computing and Services*, Vol.16, No.2(2015), 41~28.
- Pagano, D. and W. Maalej, “User feedback in the appstore: An empirical study,” *Proceedings of Requirements Engineering Conference, 2013 21st IEEE International*, (2013), 125~134.
- Pak, A. and P. Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining,” *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, (2010), 1320~1326.
- Pang, B., L. Lee and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, (2002), 79~86.
- Pang, B. and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, Vol.2, No.1-2 (2008), 1~135.
- Park, S. H. and H.-J. Song, “Word of Mouth and Box Office Performance: WOM’s Impact on Weekly Box Office Revenues,” *Korean Journal of Journalism and Communication Studies*, Vol.56, No.4(2012), 210~235.
- Park, S. H., H.-J. Song and W.-K. Jung, “The Determinants of Motion Picture Box Office Performance : Evidence from Korean Movies Released in 2009-2010,” *Journal of Communication Science*, Vol.11, No.4(2011), 231~258.
- Sarvabhotla, K., P. Pingali and V. Varma, “Supervised learning approaches for rating customer reviews,” *Journal of Intelligent Systems*, Vol.19, No.1(2010), 79-94.
- Seroussi, Y., F. Bohnert and I. Zukerman, “Personalised rating prediction for new users using latent factor models,” *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, (2011), 47~56.
- Shin, J. and H. Kim, “A Robust Pattern-based Feature Extraction Method for Sentiment Categorization of Korean Customer Reviews,” *Journal of KIISE : Software and Applications*, Vol.37, No.12(2010), 946~950.
- Wyatt, J., “High concept, product differentiation, and the contemporary US film industry,” *Current Research in Film: Audiences, Economics and Law*, Vol.5(1991), 86~105.
- Yanagimoto, H., M. Shimada and A. Yoshimura, “Document similarity estimation for sentiment analysis using neural network”. *Proceedings of Computer and Information Science (ICIS), 2013 IEEE/ACIS 12th International Conference*, (2013), 105~110.

Abstract

A Study on Analyzing Sentiments on Movie Reviews by Multi-Level Sentiment Classifier*

Yuyoung Kim**·Min Song***

Sentiment analysis is used for identifying emotions or sentiments embedded in the user generated data such as customer reviews from blogs, social network services, and so on. Various research fields such as computer science and business management can take advantage of this feature to analyze customer-generated opinions.

In previous studies, the star rating of a review is regarded as the same as sentiment embedded in the text. However, it does not always correspond to the sentiment polarity. Due to this supposition, previous studies have some limitations in their accuracy. To solve this issue, the present study uses a supervised sentiment classification model to measure a more accurate sentiment polarity. This study aims to propose an advanced sentiment classifier and to discover the correlation between movie reviews and box-office success.

The advanced sentiment classifier is based on two supervised machine learning techniques, the Support Vector Machines (SVM) and Feedforward Neural Network (FNN). The sentiment scores of the movie reviews are measured by the sentiment classifier and are analyzed by statistical correlations between movie reviews and box-office success. Movie reviews are collected along with a star-rate. The dataset used in this study consists of 1,258,538 reviews from 175 films gathered from Naver Movie website (movie.naver.com).

The results show that the proposed sentiment classifier outperforms Naive Bayes (NB) classifier as its accuracy is about 6% higher than NB. Furthermore, the results indicate that there are positive

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2015S1A3A2046711).

This work was supported (in part) by the Yonsei University Future-leading Research Initiative of 2015 (2015-22-0119).

** Master's Course, Graduate School of Library and Information Science, Yonsei University

*** Corresponding Author: Min Song

Department of Library and Information Science, Yonsei University

50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea

Tel: +82-2-2123-2416, Fax: +82-2-393-8348, E-mail: min.song@yonsei.ac.kr

correlations between the star-rate and the number of audiences, which can be regarded as the box-office success of a movie. The study also shows that there is the mild, positive correlation between the sentiment scores estimated by the classifier and the number of audiences. To verify the applicability of the sentiment scores, an independent sample t-test was conducted. For this, the movies were divided into two groups using the average of sentiment scores. The two groups are significantly different in terms of the star-rated scores.

Key Words : sentiment analysis, sentiment classification, movie review analysis, text mining

Received : July 28, 2016 Revised : September 6, 2016 Accepted : September 7, 2016

Publication Type : Regular Paper Corresponding Author : Min Song

저 자 소 개



김 유 영

연세대학교 대학원 문헌정보학 석사과정에 재학중이다. 관심분야는 정보학 분야로, 현재 텍스트 마이닝을 중점으로 공부하고 있다. 텍스트 마이닝 중 social media mining을 주로 연구중이다.



송 민

연세대학교에서 도서관학 학사를 받고, Indiana University 문헌정보학 석사, 그리고 Drexel University의 Information Science and Technology에서 박사 학위를 마쳤다. 2012년 연세대학교 문헌정보학과에 임용되어 현재 정교수로 재직중이다. 주 연구분야는 text mining, biomedical literature mining, social media mining, entitymetrics, big data이며, SCI(SSCI) 등재 국제학술지에 총 65건, 학술진흥재단 등재지에 총 18건의 논문을 게재하였고, 국제학술대회에서 70편의 논문을 발표하는 등 활발한 학술활동을 펼치고 있다.