creative commons

COMMONS DEED

저작자표시 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

조선대학교
CHOSUN UNIVERSITY

[UCI]I804:24011-200000267158

2019년 2월
석사학위 논문

드라마 리뷰 속성별 감성분류 방법

조선대학교 산업기술융합대학원

소프트웨어융합공학과

한 두 진

조선대학교
CHOSUN UNIVERSITY

2019년 2월 석사학위논문

드라마 리뷰 속성별 감성분류 방법

한 두 진

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

드라마 리뷰 속성별 감성분류 방법

Method of Emotion Classification by Attribute
for Drama Review

2019년 2월 25일

조선대학교 산업기술융합대학원

소프트웨어융합공학과

한 두 진

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

드라마 리뷰 속성별 감성분류 방법

지도교수 반 성 범

이 논문을 공학석사학위신청 논문으로 제출함

2018년 11월

조선대학교 산업기술융합대학원

소프트웨어융합공학과

한 두 진

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

한두진의 석사학위논문을 인준함

위원장 조선대학교 교수 김 판 구 (인)

위 원 조선대학교 교수 반 성 범 (인)

위 원 조선대학교 교수 신 주 현 (인)

2018년 11월

조선대학교 산업기술융합대학원

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

목 차

ABSTRACT

I. 서론 ····· 1
A. 연구 배경 및 목적 ····· 1
B. 연구 내용 및 구성 ····· 3

II. 관련연구 ····· 4
A. 감성분석 ····· 4
1. 사전 기반의 감성분석 ····· 6
2. 속성 단위의 극성 분석 ····· 7
B. Word2Vec ····· 9

III. 드라마 리뷰 속성별 감성분류 방법 ····· 11
A. 시스템 구성도 ····· 11
B. 속성 사전 구축 ····· 13
1. 데이터 전처리 및 속성단어 유사어 추출 ····· 13
2. 속성사전 구축 ····· 17
C. 속성별 감성분석 ····· 18
1. 속성별 문장 분류 ····· 18
2. 속성별 감성 어휘 추출 ····· 20
3. 속성별 문장 감성분류 ····· 21

IV. 실험 및 평가 ····· 24
A. 실험 결과 및 비교 실험 ····· 25

V. 결론 및 제언 ····· 29

참고문헌 ····· 31

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

표 목 차

[표 2-1] 감성분석 3단계 ····· 4
[표 3-1] 리뷰 데이터 토크나이징 ····· 14
[표 3-2] 한국어 불용어 리스트 ····· 14
[표 3-3] 불용어 제거 예시 ····· 15
[표 3-4] 속성단어 유사어 추출 ····· 16
[표 3-5] 속성사전 구축 ····· 17
[표 3-6] 리뷰 데이터 일부 ····· 18
[표 3-7] 속성별 문장 분류 ····· 19
[표 3-8] 감성 어휘 추출 예시 ····· 20
[표 3-9] KNU 긍/부정 사전 예시 ····· 21
[표 3-10] 속성별 감성에 따른 긍정/부정 판별 ····· 22
[표 4-1] 실험 환경 ····· 24
[표 4-2] 긍/부정 결과 ····· 25
[표 4-3] 속성별 감성분류 결과 예시 ····· 26
[표 4-4] 기존 감성분석과 속성별 감성분류의 비교실험 결과 ····· 26
[표 4-5] 속성별 드라마 인기 순위 예시 ····· 27

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

그 림 목 차

[그림 1-1] 시청률이 낮은 드라마 예시 ····· 1
[그림 2-1] 상품 고유 속성 ····· 7
[그림 2-2] CBOW 모델과 Skip-gram 모델 구조 ····· 9
[그림 2-3] Word Embedding 과정 ····· 10
[그림 3-1] 속성별 감성분류를 위한 시스템 구성도 ····· 11
[그림 3-2] 시청자 리뷰 ····· 12
[그림 3-3] Word2Vec 코드 ····· 15
[그림 3-4] 속성 문장 추출 코드 ····· 18
[그림 3-5] easySenti 패키지를 이용한 감성분석 코드 ····· 22
[그림 3-6] 속성별 긍/부정 결과 그래프 ····· 23

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

ABSTRACT

Method of Emotion Classification by Attribute for Drama Review

DooJin Han
Advisor : Prof. Pan, Sung Bum
Department of Software Convergence Engineering
Graduate School of Industry Technology Convergence, Chosun University

Unlike the past when dramas were mainly limited to terrestrial channels, numerous dramas are being broadcast through Internet web dramas as well as cable. People's opinions have become important for those who choose to watch dramas because it can be fun even if they have low ratings. Accordingly, people's opinions can be helpful because they can understand how they feel and watch dramas without a separate investigation process. However, the time and expense of reading and selecting all of the review data, which was the opinion of people online, led to the need for data analysis techniques to automatically analyze and meter the information needed in the review data. And those objectives are bringing about the development of emotional analysis. Existing emotional analysis analyzes the review written by view

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

ers to find out whether it is positive or negative, but it can also mean that there are sentences that are unrelated to the drama and that one wants to know what one likes or dislikes. Therefore, in this thesis, one can select a property keyword related to a drama and provide information considering one's preferences to those who want to watch the drama through emotional classification method by nature. Similar terms, which refer to attributes within a sentence, also use the Word2Vec model to avoid missing sentences related to the attribute. This paper also presents the results for verifying efficiency through comparative experiments with the proposed classification of sensitivities by attributes and with existing emotional analysis, and also shows examples of the popularity ranking of dramas by nature based on positive/negative results.

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

Ⅰ. 서론

A. 연구 배경 및 목적

드라마는 주로 지상파에 한정되었던 과거와는 달리 현재 케이블 방송사뿐만 아니라 인터넷 웹 드라마를 통하여 수많은 드라마가 방영되고 있다. 지상파 방송사만 해도 일주일에 대략 40편의 드라마를 편성 한 해 평균 130편이라는 드라마가 제작, 방영되고 있다[1]. 수많은 드라마 중 어떤 드라마를 시청할지 재미있는 드라마가 무엇인지 고민하는 사람들에게 시청률이 높다고 드라마를 선택하기에는 어려움이 있다. 과거 시청률이 높으면 인기 있는 드라마였었지만 현재 인터넷과 스마트폰의 발달로 인하여 드라마를 시청할 방법이 다양해지면서 시청률이 낮더라도 사람들에게 재미있고 인기 있는 드라마일 수 있다. 예를 들어 [그림2-1]을 확인하면 방영 중인 드라마 '뷰티인사이드' 시청률은 3.6%로 낮은 시청률이지만 국내 드라마 일간 검색어를 확인하면 2위로 높은 순위를 가지고 있는 것을 확인할 수 있다.

[그림1-1] 시청률이 낮은 드라마 예시

그러므로 시청률은 인기 있는 드라마의 절대적인 지표가 될 수 없기 때문에 드라마에 대한 사람들의 의견을 중요시 하게 되었다. 사람들 드라마에 대한 의견을 SNS와 온라

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

인 게시판, 블로그, 커뮤니티에 적극적으로 참여하고 있으며 이렇게 축적된 리뷰 데이터는 시청자가 느끼는 감정을 직접적으로, 별도의 조사 과정 없이 파악할 수 있기 때문에 드라마를 선택하려는 사람들에게 도움을 줄 수 있다[2]. 따라서 시청자들은 시청률이 높거나 드라마 인기투표 결과를 통하여 드라마를 선택하기 보다는 온라인상에 남겨진 드라마 후기를 통하여 정보를 얻어 선택할 수 있음을 알 수 있다. 그렇지만 온라인상의 수많은 리뷰 데이터를 한 사람이 모두 읽고 드라마를 선택하기에는 막대한 시간과 비용이 소요되기 때문에 수많은 리뷰 데이터에서 필요한 정보를 자동으로 분석하고 계측화하기 위한 데이터 분석 기법이 필요하게 되었다. 그리고 그런 목적들이 감성분석의 발전을 가져다주고 있다. 물론 시청자의 감성을 측정하기 위해 감성분석을 이용하지 않고도 '별점'과 같은 단순한 수치를 이용할 수도 있지만 리뷰와 별점이 일치하지 않을 수도 있으며 의도적으로 별점을 높게 주게 된다면 신뢰성이 떨어지는 리뷰가 될 수 있다. 신뢰성의 문제를 해결하기 위해 실제 시청자가 작성한 리뷰 자체를 분석해 긍정 또는 부정 여부를 파악해야한다. 하지만 단순히 드라마 리뷰를 감성분석을 이용하여 분석하게 되면 리뷰를 작성한 사람이 모든 문장에 주제와 관련 된 문장을 남기지 않을 수도 있으며, 감성분석은 컴퓨터가 분석하기 때문에 기계적으로 긍정과 부정 단어가 몇 차례 나왔는지 단순히 횟수를 세기 때문에 리뷰 전체의 의미를 잘못 해석할 수 있다. 또한 드라마 제작 관계자 시점에서는 드라마에 대한 평가가 단순히 긍정인지 부정인지 아는 것에서 나아가 무엇을 좋아하고 무엇을 싫어하는지 알고 싶어 할 수 있다. 또는 드라마 패션과 관련된 사람들에게는 배우의 어떤 의상이 이슈가 되는지 메이크업은 어떻게 했는지 이슈 되는 머리스타일은 무엇인지 알고 싶어 한다. 이러한 한계점을 극복하고 사용자의 취향별 의미를 해석하기 위한 연구로써 드라마 리뷰에 대한 속성별 분석이 필요하다. 예를 들어 '배우'가 누가 출현하는지 중요시하는 사람이 있는 반면 극중 배우의 '패션'에 대한 정보가 필요한 사람이 있을 수 있으며, 또는 단지 내용을 중요시하는 사람들에게는 드라마 '스토리' 일수도 있다.

이를 바탕으로 본 논문에서는 드라마 리뷰 데이터에서 속성을 기반으로 한 감성분류를 하고자 한다. 속성사전 구축을 위해 속성과 유사한 속성어를 추출하여 감성분류를 진행한다. 감성분류를 통하여 드라마에서 무엇을 얼마나 중요시하는지 즉 사람들에게 필요한 정보를 반영할 수 있는 드라마 리뷰 속성별 감성분류 방법을 제안한다.

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

B. 연구 내용 및 구성

본 연구는 드라마 리뷰 감성분류를 위하여 실시간으로 시청자 리뷰가 활발히 진행되고 있는 2018년 tvn에서 방영중인 '뷰티인사이드' 시청자 리뷰를 수집하였다. 수집한 데이터는 감성분류에 적합하도록 전처리과정을 통하여 감정과는 불필요한 용어들을 제거한 후 Word2Vec 모델을 이용하여 속성사전을 구축하였다. 구축된 속성사전을 활용하여 문장을 분류 하였고, 사전에 구축된 긍/부정 사전을 이용하여 속성별 감성분류를 진행 하였다. 본 연구는 서론을 포함하여 총4장으로 구성되며 내용은 다음과 같다.

1장 서론에서는 본 논문 주제의 선정 배경과 목적을 기술하고, 2장 관련 연구에서는 본 논문의 리뷰 분석 기법인 감성분석에 대해 알아보고 속성 사전구축을 위한 Word2Vec 속성어 추출에 대해 기술한다. 3장에서는 'Word2Vec'모델을 통한 속성어를 추출하여 속성사전을 구축하는 방법과 속성사전을 이용한 속성문장을 감성 분류하는 방법을 제시 한다. 4장에서는 본 논문의 결론과 향후 연구 방향에 대해 기술하고 마무리한다.

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

Ⅱ. 관련연구

A. 감성분석

감성분석은 텍스트마이닝 분석의 한 분야로 오피니언 마이닝 또는 감성분류로 불리며 텍스트에서 사람들의 의견과 태도, 성향 등을 분석하는 기술이다[3]. 감성분석은 문서가 긍정, 부정, 중립 중 어떤 의견이나 생각을 갖고 있는지 판별하는 하나의 과정이라 볼 수 있으며 각 문서 최소 단위인 어휘의 감성 극성에 기반을 두어 분석이 이루어진다[4]. 그러므로 단어의 감성극성을 정확히 반영한 감성사전을 사용하는 것이 중요하다.

[표 2-1] 감성분석 3단계

데이터수집	SNS, 리뷰게시판, 블로그에서 리뷰 데이터를 수집
주관성 탐지	사용자의 주관이 드러난 부분만을 분류
극성 탐지	추출한 감성 데이터를 ‘긍정’과 ‘부정’의 두 감성으로 분류

[표 2-1]은 감성분석 3단계에 대한 정보이며 첫 번째 온라인상에 있는 방대한 양의 데이터는 감성 분석의 핵심 요소다[5]. 각종 커뮤니티, 블로그, 영화나 드라마 리뷰 게시판, 쇼핑몰 상품 평가란 등 공개적인 데이터를 수집하는 ‘데이터 수집’ 단계이다. 데이터를 수집하기 위해서는 사용자가 원하는 정보를 정확하게 찾기 위한 검색 엔진을 활용하며 사용자의 질의어가 포함된 모든 문서, 리뷰, 코멘트, 별점이 포함한 관련 데이터를 수집한다. 두 번째 ‘주관성 탐지’는 수집된 데이터를 감성 분석에 사용될 텍스트 요소만을 분류하는 작업이 필요하다. 일반적으로 온라인상에서 수집된 텍스트는 문장 내에서 주관성이 없거나 ‘감성’과는 관련이 없다고 판단되는 부분을 제외시키며 또한 지나친 정보 수집에서 따르는 문제를 피하기 위해 개인 정보 즉 텍스트 저자의 이름, 성별 같은 정보를 걸러내는 과정을 거친다. 세 번째 ‘극성 탐지’는 주어진 데이터가 ‘긍정’인지, ‘부정’인지를 판단하는 ‘극성 분석’ 작업이 이뤄진다. 컴퓨터는 텍스트 안에 있는 긍정적 단어와 부정적인 단어를 탐지하여 이를 정량화 한 뒤 통계적 기법을 적용

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

한다. 예를 들어, 문서에서 각 단어가 나타나는 '빈도'나, 긍정이나 부정에 따라 가중치나 점수를 부여한 뒤, 각 단어가 나타내는 점수의 총합이나 평균을 구해 전체 텍스트가 과연 긍정적인지 혹은 부정적인지 알아내는 것이다.

기존의 감성분석 연구에서는 극성분류에 제품의 특징을 이용하여 불필요한 단어들의 극성 분류 횟수를 줄이고, 특징마다 사용되는 감성 단어의 극성을 분류하는 방법을 제안하였다[6]. 소셜 미디어에서 특정 제품이나 서비스에 대해 부정 표현의 빈도가 급격히 상승한 경우 위기 상황을 감지하고 대응하기 위한 위기관리 모델을 제안하였다[7].

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

1. 사전 기반의 감성분석

사전기반의 감성분석은 데이터 문장에서 감정 단어를 추출하고 이를 감성사전 이용하여 어휘의 극성을 판별하는 기법이다[8]. 사전은 긍정적인 의미의 단어와 부정적의미의 단어를 모아놓은 대표적인 사전을 의미한다. 정확한 감성 분석을 위해서는 신뢰도가 높은 감성사전이 구축되어야한다. 신뢰도가 높은 감성사전을 구축하려면 사람이 직접 단어를 선별하는 과정이 필요하다. 하지만 언어의 사용 과정에서 발생하는 다양한 의미의 변이와 동태적 활용, 동음이의어가 분석에 있어 어려운 작업으로 인식 되고 있다[9]. 사전기반 감성분석 단계로는 첫 번째 분석할 문장을 형태소 단위로 분리하여 감성어휘를 추출한다. 두 번째 단계로는 공개된 감성 사전과 감성어휘와 매칭 하여 단어의 감성점수를 추출 할 수 있다. 끝으로 단어의 감성 점수를 토대로 문장의 감성점수를 계산 할 수 있는 방법이다. 한국어 감성사전의 대표적인 예는 KOSAC와 KNU 한국어 감성사전이 있다.

KOSAC(한국어감성분석코퍼스) 감성사전은 서울대학교에서 개발한 한글의 말뭉치사전이며 세종 구문분석 코퍼스로부터 선별한 332개의 기사, 7,744개 문장을 주석대상으로 17582개의 감정 표현이 수록돼 있다[10]. 극성분류 실험을 위해 SVM 모델을 사용하였다.

KNU 한국어 감성사전은 군산대학교에서 구축한 한국어 감성사전으로 국립국어원에서 제공하는 '표준국어대사전(국립국어원 2018)'을 구성하는 모든 단어와 그 단어에 해당하는 뜻풀이를 수집한 후 정제하여 만들었으며 그 외에도 축약어, 이모티콘 등 표준국어 대사전에서 추출되지 않은 새로운 감성어를 추출하였다. KNU 감성사전은 영화, 드라마, 음악, 스마트폰, 컴퓨터 등 특정 도메인에서 사용되는 긍부정어보다는 인간의 보편적인 기본 감정 표현을 나타내는 긍부정어로 구성된다[11].

언어의 사용 과정에서 문서의 특성에 맞지 않는 범용감성사전 사용 시 감성분석의 정확성이 낮아지는 문제가 발생하기 때문에 문서의 주제에 맞는 맞춤형감성사전을 구축하고 정확성을 높이려는 연구가 수행되고 있다[12]. 기존 연구로는 주제지향적인 감성사전을 구축하는 방법을 제안하고 뉴스 분석을 통하여 주가지수의 상승을 예측하는 연구를 하였다[13].

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

2. 속성 단위의 극성 분석

감성분석에서 문서 전체의 긍/부정 여부에만 집중하게 되면 여러 문제가 생길 수 있다. 예를 들어 한 사람이 상품 리뷰에 대해 장점을 나열 하였지만 마지막 문장에 '그러나 이 상품을 사고 싶지는 않다'라고 남겼다면 이 리뷰는 부정적인 리뷰지만 컴퓨터는 긍정인 문장이 많기 때문에 긍정 리뷰로 해석할 것이다. 또한 단순히 상품이 긍정인지 부정인지 아는 것에서 나아가 '무엇'이 좋고 '무엇'이 싫은지 알고 싶어 한다[14]. 이러한 문제를 해결하기 위해서는 문서 전체의 긍/부정 여부가 아니라 문장단위로 텍스트를 분석하는 속성별 극성 분석이 필요하다[14]. '속성'이란 [그림2-1]와 같이 상품 고유 속성상품의 고유 속성 '기본사양', '디스플레이', '시스템'과 같은 여러 특징을 의미한다.

기본사양	
운영체제	안드로이드 8.0 오레오
TTA-VoLTE	지원
유심타입	나노유심
디스플레이	
화면크기(인치)	14.7cm(5.8")
ppi	570ppi
패널종류	SUPER AMOLED
HDR규격	
시스템	
AP종류	엑시노스9 9810
코어클럭	2.8GHz+1.7GHz
시스템램	램:4GB
저장장치	UFS2.1

[그림2-1] 상품 고유 속성

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

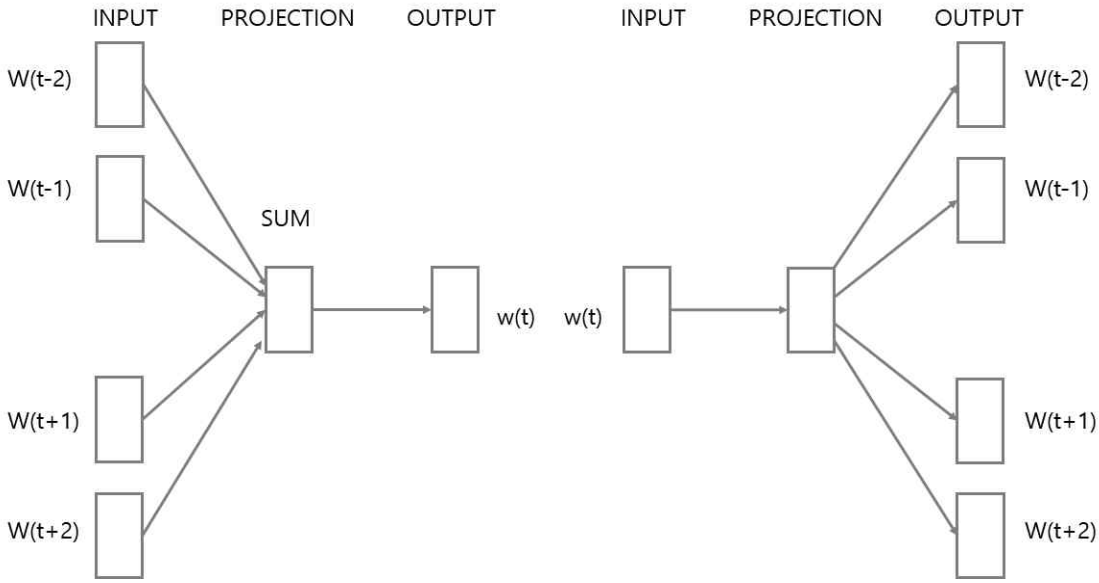
속성 단위의 극성 분석을 하기 위해서는 우선 대상이 어떤 속성으로 구성되어 있는지, '속성명'을 추출한 다음, 각각의 속성에 따른 감성어를 찾은 뒤 그 감성어가 긍정에 해당하는지, 부정에 해당하는지 분류하는 과정이 필요하다. 이제 사람들은 단순히 '좋아요'가 몇 개인지 아는 것에서 나아가 '무엇을' 좋아하고 있는지도 알고 싶어 한다. 예를 들어 '이 스마트폰은 디자인이 좋긴 하지만 카메라가 별로다'라는 텍스트의 경우, 문서 자체의 긍/부정 여부를 따지는 것이 아니라 문장 단위로 들어가 사용자가 구체적으로 스마트폰의 어떤 점을 좋아하고 싫어하는지에 대한 분석을 할 수 있다. 관련 논문으로는 오피니언 마이닝은 개인, 이슈, 토픽, 서비스, 조직, 사건, 제품, 그리고 이들의 여러 속성에 대한 사람들의 태도, 평가, 의견, 감정 등을 분석하는 일련의 과정을 의미한다 하였다[15]. 그러나 실제 일상에서는 문서의 주제에 대한 감정만이 존재하는 것이 아니라 다양한 세부의견들이 포함되어 있으므로 이를 세분화하는 오피니언 타겟이 필요하다고 하였다. 이러한 타겟을 속성으로 하여 다양한 의견 세분화가 가능한데 그 예로 두 가지 종류의 스마트 폰에 대한 의견을 비교할 때 전체적인 긍정과 부정뿐만 아니라 디자인, 배터리 수명, 카메라, 사이즈 등의 세부 요소에 대한 감성분석이 가능함을 보여주었다. 또한 소비자 개인이 작성한 전체 상품평에 대한 감정의 극성도 중요하지만 상품의 속성 또는 특성에 대한 극성을 판단하는 것이 중요하다 하였다[16].

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

B. Word2Vec

Word2Vec은 2013년 구글에서 개발한 방법론이다. 단어(Word)를 벡터(Vector)로 바꿔주는 방법이며 이를 임베딩(Embedding)이라고 한다[17]. Word2Vec의 효과는 단어를 벡터화 할 때 단어의 문맥적 의미를 보존하게 된다.



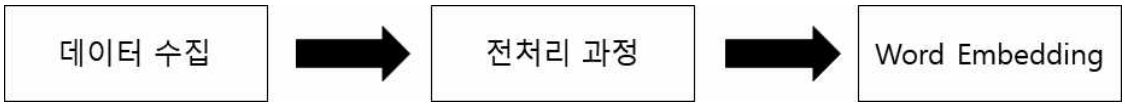
[그림 2-2] CBOW 모델과 Skip-gram 모델 구조

Word2Vec는 단어를 벡터로 바꾸는 방법론으로 크게 CMOW(Continuous Bag of Words)와 Skip-Gram 두 가지 방식이 있다. [그림 2-2]는 CBOW 모델과 Skip-gram 모델 구조이다. 전자는 주변에 있는 단어들을 가지고 중심에 있는 단어를 맞추는 방식이고, 후자는 중심에 있는 단어로 주변 단어를 예측하는 방법이다. 예를 들어 “오늘도 나는____에 간다.” 라는 문장이 있다면 빈칸에 들어갈 수 있는 단어는 다양하다. ‘집’ 일 수도 있고 ‘학교’ 또는 ‘회사’일 수도 있다. 이렇게 주변 단어를 가지고 중심에 있는 단어를 맞춤으로써 단어 벡터들을 만들어 내는 방법이 CBOW이다. 그러나 “____외나무다리____”앞 뒤로 어떤 단어가 올지 즉 대상 단어로 주변 단어를 예측하는 방법은 Skip-Gram이다[18]. Word2Vec은 비슷한 위치에 등장하는 단어들을 그 의미도 유사할 것

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

이라는 전제가 깔려 있다.



[그림 2-3] Word Embedding 과정

[그림 2-3] 온라인상 리뷰를 수집하여 속성과 유사한 단어를 추출하기 위한 과정이다. 유사어를 추출하기 전에 조사나 너무 긴 단어의 제거 등을 위해 토크나이징을 실시한다. 대표적인 한국어 형태소분석으로는 KoNLPy가 있다. 토크나이징한 결과물을 Python gensim 패키지를 활용해 Word Embedding하여 속성과 유사한 단어를 찾는다.

[19]에서는 Word2Vec은 학습하는 문서에 의존적으로 키워드를 분석 할 수 있지만 키워드 분석에 활용하는 사례는 거의 진행하지 않았기 때문에 키워드의 특성 및 성향을 쉽게 파악할 수 있다는 이유로 Word2Vec을 감성 분석에 활용하였다.

[20]에서는 연구자가 구축한 감성사전을 Word2Vec 모델을 구축하기 위해 Word2Vec 모델 학습 단계에 결함 Word2vec 모델 학습 단계에 결합하여 감성사전 기반 Word2Vec 모델을 구축하고 이를 통해 분석하고자 하는 문서의 자질을 구축하여 감성분석을 시행하였다.

Collection @ chosun

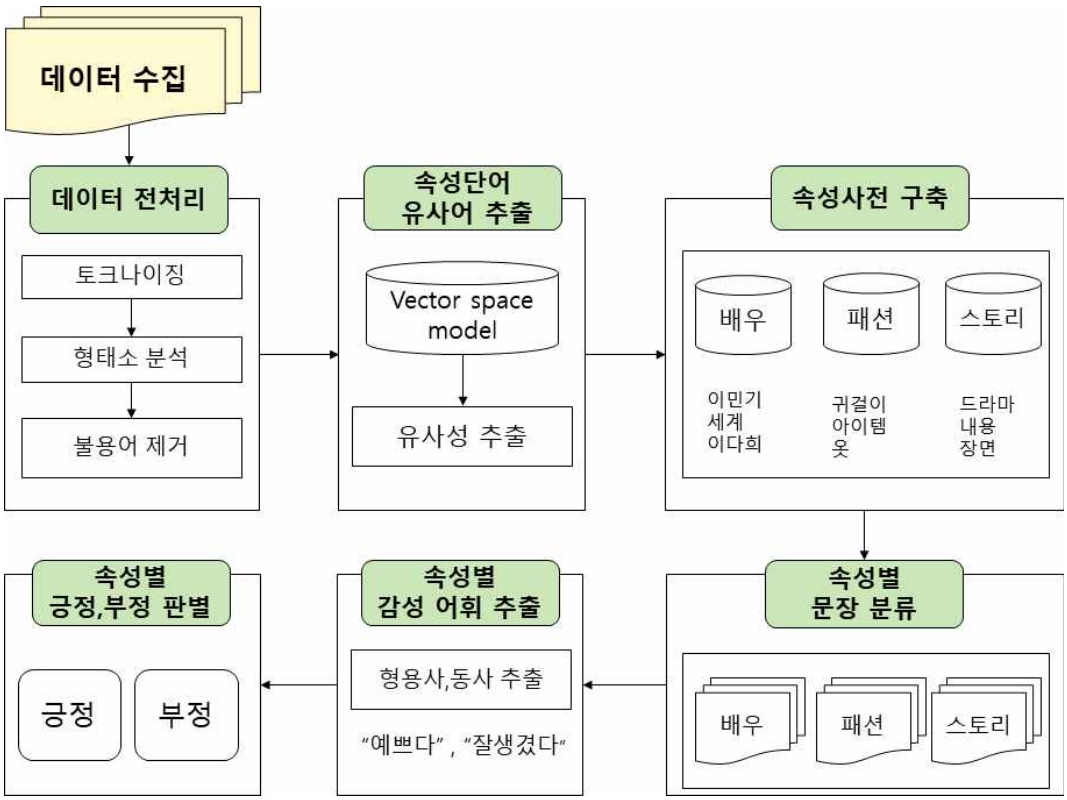
조선대학교
CHOSUN UNIVERSITY

Ⅲ. 드라마 리뷰 속성별 감성분류 방법

본 논문에서 제안하는 감성분류 방법은 기존의 드라마 댓글이나 리뷰를 수집 후 리뷰 전체를 감성 분석하는 방법이 아닌 속성에 따른 긍정/부정 분류를 통해 기존의 감성 분석보다 세밀한 결과를 도출하여 시청자가 구체적으로 어떤 속성을 좋아하고 싫어하는지에 대한 분석 방법을 제안하고자 한다.

A. 시스템 구성도

[그림 3-1]은 속성별 감성분류 방법의 시스템 구성도이다. 드라마 리뷰 속성별 감성분류 방법은 크게 데이터 수집 및 전처리 단계와 속성사전 구축 단계, 그리고 속성별 긍/부정 판별 단계로 구성된다.



[그림 3-1] 속성별 감성분류를 위한 시스템 구성도

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

실험을 위한 데이터는 JTBC 드라마 '뷰티인사이드' 시청 소감 게시판과 NAVER에서 제공하는 실시간 TALK에서 2018.10.01.부터 2018.10.10.까지 17467개의 데이터를 수집한다. 수집된 데이터는 특수한 문장 부호, 띄어쓰기, 숫자 등의 문제가 있으므로 전처리 단계에서 이들을 정제할 필요가 있으며 Word2Vec 학습을 통해 속성과 유사한 단어를 추출한다. 이후 속성사전을 구축하여 속성별 문장의 분류를 통하여 긍/부정을 판별한다.

기존의 감성분석 방법은 리뷰 전체의 감성 단어를 긍/부정 분류하였으나 드라마 리뷰 특성상 개인의 취향에 즉 속성에 따라서 느끼는 감정이 다르기 때문에 본 연구에서는 이러한 점을 보안하기 위해 속성을 기반으로 데이터를 긍/부정 분류하는 실험을 진행 하였다. [그림 3-2]는 실시간 드라마 시청자 리뷰 게시판이다.

No	제목
공지	[이벤트] 바뀐 그림 찾기 이벤트 당첨자 발표
99	이래서 영화, 드라마를 사랑할 수 밖에 없습니다.
98	매번 앤딩이 강렬해요 다음회가 기대되요
97	한세계의 대반전
96	편집 좀 자연스럽게 부탁드립니다..
95	ost!!공개만 기다려요ㅠㅠ
94	4화 이다희씨 립스틱
93	안재현ost
92	Ost
91	말 못한다는 아이가 어찌 소리를 냈을까요?

[그림 3-2] 시청자 리뷰

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

B. 속성사전 구축

속성사전을 구축하기에 앞서서 속성 키워드를 선별하는 기준은 연구자의 판단 기준과 사용 용도에 따라 다를 수 있으며 본 논문에서는 사람마다 드라마를 시청하는 이유가 다르기 때문에 여러 가지 속성이 있지만 가장 큰 이유 3가지로 '배우','스토리','패션'으로 선정하였다. 선정하게 되는 근거로 사람들은 드라마 시청률 상승에 영향을 미치는 요인 즉 드라마에 어떤 스타 배우가 나왔는지에 따라 사람들은 드라마를 시청하기 때문에 '배우'를 속성으로 정하였다. 배우에 대한 중요성은 특히 초반 시청률에서 스타 배우가 참여한 드라마가 그렇지 않은 드라마보다 통계적으로 유의하게 높은 시청률을 나타낸다[21]. 또한 국내 드라마가 시청률 측면에서 성공을 거두기 위해 갖추어야 할 핵심 요인에 관한 질문에서 전문가들은 스토리 구성, 인적 자원 구성의 중요성, 편성 요인 순서로 중요하다 답하였다.[22] 그렇기 때문에 사람들은 스토리가 탄탄한 드라마는 초반에 시청률이 저조하더라도 많은 관심을 주기 때문에 '스토리'를 선정하게 되었다. 드라마가 패션은 사람들의 미의 개념의 변화는 다양해지고 개성이 강한 미적 아름다움을 추구하기 때문에 미디어의 발전은 유행에 따른 패션, 헤어스타일, 외목관심, 메이크업 등의 분야 전반에 영향을 끼치고 있다. 때문에 드라마 속 배우들의 패션스타일은 소비자에게 모방심리를 가지게 하기 때문에 패션과 외모에 대한 관심을 겨냥한 마케팅 활동은 여러 분야에서 확인되고 있다.[23] 즉 배우의 머리부터 발끝까지의 패션은 사람들에게 정보가 되고 있으며 수 많은 쇼핑몰, 블로그, 어플리케이션, 뉴스기사로도 활용 되고 있다. 그렇기 때문에 이 같은 사람들을 위한 '패션'을 속성 키워드로 정하게 되었다. 본 논문에서는 속성 '배우', '패션', '스토리'로 속성사전을 구축한다.

1. 데이터 전처리 및 속성단어 유사어 추출

수집한 데이터가 저장된 txt 파일로 저장된 리뷰데이터를 속성과 유사한 단어를 추출하는 전처리 과정으로 토크나이징 기법을 이용하고, 형태소분석을 수행하기 위해 Python KoNlpy라이브러리를 이용하였다

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

[표 3-1] 리뷰 데이터 토크나이징

배우	리뷰 원문
원본 데이터	서현진씨 연기밖에 안 보이네요 ^^ 월화요정 서현진님... 다시 월화요일이 즐거워지기 시작했어요 서현진씨 작품 하나도 빼먹지 않고 챙겨보고 있어요
토크 나이징 적용 후 데이터	서현진, 씨, 연기, 밖에, 안, 보이네요, ^^ 월화, 요정,서현진, 님,... ,다시, 월화, 요일, 이, 즐거워지기, 시작, 했어요 서현진, 씨, 작품, 하나, 도, 빼먹지, 않고, 챙겨, 보고, 있어요

토크나이징은 문서나 문장을 분석에 용이하도록 하나의 단어 단위로 세분화시키는 작업이며, 수집 후 토크나이징을 적용한 데이터는 [표 3-1]과 같이 저장하였다. 이후 형태소 분석과 속성과 유사한 단어를 판별하기 위해 [표 3-2] 한국어 불용어 리스트 중 고빈도 불용어 제거를 진행으로 전처리를 수행한다.

[표 3-2] 한국어 불용어 리스트

형태	비율	형태	비율
이	0.01828	나오	0.000725
있	0.011699	가지	0.00072
하	0.009774	씨	0.00071
것	0.009733	시키	0.000708
들	0.006898	만들	0.000704
그	0.005327	지금	0.0007
되	0.003613	생각하	0.000695
수	0.003474	그러	0.000692
이	0.003361	속	0.000685
…	…	…	…

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

[표 3-3] 불용어 제거 예시

배우	리뷰
토크나이징 적용 데이터	서현진, 씨, 연기, 밖에, 안, 보이네요, ^^ 월화, 요정, 서현진, 님,... ,다시, 월화, 요일, 이, 즐거워지기, 시작, 했어요 서현진, 씨, 작품, 하나, 도, 빼먹지, 않고, 챙겨, 보고, 있어요
불용어 제거 완료 데이터	서현진,연기, 월화,요정,서현진,월화,요일,시작 서현진,작품

[표 3-3]과 같이 불용어 제거가 완료된 리뷰 데이터들은 파이썬 gensim 패키지를 활용해 Word2Vec 방법론을 적용한다. [그림 3-3]에서 gensim 패키지가 제공하는 기능 중 'most_similar' 함수는 두 벡터 사이의 코사인 유사도를 구해주며 그 값이 클수록 비슷한 단어임을 의미한다.

```
train_docs = [(tokenize(row[0])) for row in train_data]

from gensim.models import word2vec

wv_model_ko = word2vec.Word2Vec(train_docs)  # 단어의 벡터표현 학습
wv_model_ko.init_sims(replace=True)
wv_model_ko.save('ko_word2vec.model')  # 나중 사용을 위해 정리

# '배우'와 유사한 단어 검색
print(wv_model_ko.most_similar(tokenize('배우'),topn=20))
```

[그림 3-3] Word2Vec 코드

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

[표 3-4] 속성단어 유사어 추출

속성	유사어
배우	이민기, 세계, 이다희, 안재현, 연기, 캐릭터 …
패션	귀걸이, 아이템, 옷, 원피스, 셔츠, 가방 …
스토리	드라마 , 내용 , 장면, 뷰티, 사이드, 영화 …

유사도 계산으로 [표 3-4]와 같이 배우'라는 단어와 가장 비슷한 즉 코사인 유사도가 큰 상위 20개 단어목록을 저장 하였다. 상위 랭크에 저장 될수록 속성과 유사도가 높은 단어로 추측이 가능하다.

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

2. 속성사전 구축

추출된 속성 유사어는 아래 랭크로 내려갈수록 사용자가 의미하는 속성과 다를 수 있다. 리뷰에서 속성과 유사율이 높은 명사로 되어 있는 유사 단어 상위 20개를 추출하여 [그림 3-5]와 같이 속성 사전을 구축한다.

[표 3-5] 속성사전 구축

배우	패션	스토리
이민기	귀걸이	드라마
세계	아이템	내용
이다희	옷	장면
:	:	:

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

C. 속성별 감성분석

1. 속성별 문장 분류

구축된 속성사전을 이용하여 [표 3-6]같은 리뷰 데이터들을 속성과 속성 유사 단어가 속해 있는 문장을 클러스터링 한다.

[표 3-6] 리뷰 데이터 일부

영화를 참 좋아했었는데, 드라마로 나오는군요.흥행 기대합니다. 2018 가을 모두를 아름답게할 뷰티 인사이드를 응원합니다! 왠지 기대가 되는 드라마네요 서현진씨가 갑자기 명MC 전현무로 바뀐다는 그 드라마 그 장면은 이민기씨와 서현진씨가 쇼핑하러갈때 바로 그 장면입니다. 한세계역 서현진씨와 서도재역 이민기씨 케미 너무 기대되요 첫방날짜 빨리 왔으면 좋겠어요 세계 얼른와 뷰티인사이드 기다리고 있어요 꿀잼 예상합니다 서현진차기작을 기다렸는데 이제 얼마 안남네요 서현진 이민기 케미 기대합니다 또 오해영 감독님 또 웰메이드 드라마 하나 만들어 주세요 웃음과 감동을 주는 드라마 민기오빠 나오는 드라마라니 이번생은 처음이라에 이어 인생드라마 가나요 너무너무 기대돼요 관리자님 제발제방 대본이벤트나 포스터 이벤트 해주세요 뼈를 갈아 열심히 참여할게요 내가 사랑하는 여자가 남자로 바뀐다면 이민기가 서현진씨와 옷가게에서 옷을 사준다. 서현진씨는 원피스를 입으러 탈의실로 들어가고서현진씨는 이민기를 유혹한다. 이민기는 탈의실로 따라들어가는데 이게 웬일 서현진씨가 전현무로 변해버린다. 이민기는 한혜진씨로 바뀌는거 아이가어쨌든 10월1일 첫방송 많이 참여해주시기 바랍니다.

```
setwd("E:/Rtest")

docs <- readLines("beauty51.txt")

docs[grep("배우", docs)]
write.table(sets, "E:/Rtest/irisssss.txt",sep = ",",row.names = FALSE,quote = FALSE,append =  TRUE,na ="NA")
```

[그림 3-4] 속성 문장 추출 코드

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

'R'에는 grep 함수를 이용하면 어떤 데이터 프레임에서 원하는 문자열이 포함된 행을 뽑을 수 있으며 [그림 3-4]와 같이 grep함수를 이용하여 속성단어와 속성 유사어가 들어가 있는 문장을 추출하여 'irisssss.txt' 파일명으로 저장하여 속성별로 분류하는 실험을 진행 하였다. [표 3-7]와 같이 속성별 문장이 분류됨을 확인 할 수 있다.

[표 3-7] 속성별 문장 분류

속성	댓글/리뷰 원문
배우	서현진 배우님 너무 아름다우시고 친절하시고 이민기 배우님도 너무 다정하신 분! 뷰티인사이드 첫방 정말로 기대하고 있어요. 첫화부터 대박 터지길 꼭 본방사수하고 주위에도 홍보할게요! 정말 좋아하는 배우분들이 주연을 맡아서 영화로도 재미있게 본 뷰티인사이드 드라마 버젼이라니 너무 기대되고 설렙니당 이민기 배우님과 안재현배우님 전에 출연하셨던 작품을 너무 재밌게 봐서 이번 드라마두 너무 기대가되네요 이번드라마 기대하겠습니다. 화이팅하세용
패션	드라마뷰티인사이드 시상식 드레스인데 웨딩드레스 빰치게 이쁜 드라마뷰티인사이드패션 병원씬에서 등장한 노랑 가디건 과 반지 드라마 보다 보면 꼭 빼놓을 수 없는 게 드라마 속 여배우 패션이죠 얼마 전, 뷰티인사이드 서현진 패션은 포스팅을 했는데 개인적으로 더 마음에 드는 취향 아무래도 제가 관심을 갖고 있는 남주의 패션에 관심이 많은 편이랍니다
스토리	신사의 품격, 태양의 후예, 상속자들, 도깨비 등 지금까지 김은숙작가님과 함께한 임메아리작가님 히스토리를 보며 기대하고 있습니다. 대박나시길 기원합니다. 스토리 전개대사는 하나 없고 일단 저는 영화 뷰티인사이드를 정말 재미있게 봐서 드라마도 무척 기대를 했습니다. 영화와 다른 캐릭터 설정이 잘못된다고 생각하지 않습니다. 오히려 그런 요소가 더욱 재미있는 스토리를 만들어낼 수 있을 것이라고 기대했습니다. 그리고 기대하며 1화와 2화를 보았습니다.

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

2. 속성별 감성 어휘 추출

분류된 속성별 문장들을 감성분류하기 위해 감성 어휘를 추출한다. 감성 어휘는 형태소 분석을 통하여 추출할 수 있으며 형태소 분석기 Konlpy 패키지로 진행 한다. 사람들은 주로 감성을 드러내는 품사인 '형용사' 예쁘다, 좋다, 아름답다 와 감성을 드러내기도 하는 동사인 '즐기다, 신나다, 느끼다'를 모두 감성 단어로 설정하였다. [표 3-8]은 감성 어휘 추출 예시를 표로 작성 하였다.

[표 3-8] 감성 어휘 추출 예시

속성	분류된 감성어휘
배우	아름다우시고, 친절하시고, 다정하신, 기대하고, 있어요
패션	이쁜, 취향이다, 관심이 많다
스토리	기대하고, 재미있는

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

3. 속성별 문장 감성분류

감성분류는 기본적으로 텍스트가 어떤 극성 값을 가지고 있는지 판별하는 과정이다. 또한 감성사전을 이용하며 문서 최소 단위인 어휘의 감성 극성에 기반을 두어 이루어진다. 본 논문에서 속성별로 분류된 문장을 분석하여 감성의 극성을 판단 한다. 속성별로 분류된 속성문장들을 공개된 긍/부정 사전인 KNU 사전을 사용하여 감성분류를 수행한다. R프로그램을 통해 easySenti 패키지를 이용하여 속성별 감성분류를 수행한다. [표 3-9]는 공개된 KNU 사전의 긍정과 부정사전의 일부를 나타낸다.

[표 3-9] KNU 긍/부정 사전 예시

Positive	Negative
기대된다	서글픔
깜찍하다	서럽다
깨끗이 하다	성가시게
깨끗이 함	쇠약한
또렷하다	수준이 낮다
매우 훌륭한	초라한
멋진	취약하다
명쾌한	크게 우는
명쾌함	추하다
미소를 보냄을	편하지 못한
믿다	편하지 아니하고
감사	포악하다
보송보송	핀잔
부드럽게	하소연하는
:	:

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

본 연구에서는 [그림 3-5] 코드를 이용하여 실험을 진행 하였다. 감정경계 기본값 0을 사용하여 0초과시 '긍정', 0미만 '부정'의 결과가 나온다. 즉 easySenti 패지지는 긍/부정 사전에 들어있는 어휘와 속성 리뷰를 비교 분석하여 긍정어휘 개수와 부정어휘 개수를 판별하며 차액 값 즉 결과의 (긍정단어 개수 - 부정단어 개수)를 하여 긍정단어가 개수가 많다면 긍정 부정단어 개수가 많다면 부정을 나타낸다. [표 3-10]은 속성별 단어 개수와 결과를 확인하는 결과며 [그림 3-6]은 배우, 패션, 스토리 의 감성분류 결과 긍정단어와 부정단어의 개수를 구성한 그래프이다.

```
library(devtools)

library(easySenti)


setwd("E:/Rtest")
positive <- readLines("positive.txt")
negative <- readLines("negative.txt")
docs <- readLines("fox.txt")


easySenti(docs, positive, negative)
```

[그림 3-5] easySenti 패키지를 이용한 감성분석 코드

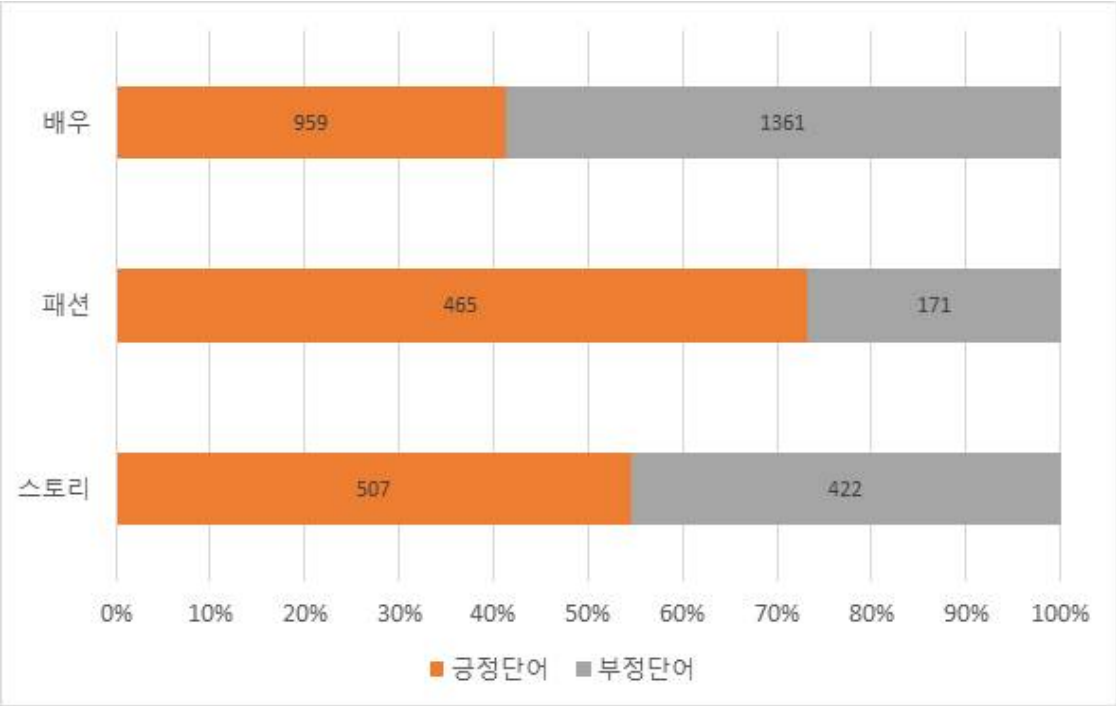
[표 3-10] 속성별 감성에 따른 긍정/부정 판별

속성	긍정단어	부정단어	결과
배우	959	1361	부정
패션	465	171	긍정
스토리	707	422	긍정

‘

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY



[그림 3-6] 속성별 긍/부정 결과그래프

‘배우’ 의 경우 긍정단어 959개 부정단어 1361개로 부정단어가 ‘부정’의 결과 값을 확인 할 수 있다. ‘패션’ 과 ‘스토리’는 부정단어의 비율보다 긍정단어의 비율이 많아 긍정의 결과를 보여 준다.

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

Ⅳ.실험 및 평가

본 장에서는 제안하는 드라마 리뷰에 대한 속성별 감성분류 방법에 대한 평가를 위해 총 17,467개의 NAVER에서 제공하는 드라마 '뷰티인사이드' 시청자 게시판과 블로그 및 실시간 talk 리뷰를 이용하여 실험을 진행하였고 제안하는 방법과 기존 감성분석 방법을 비교 평가를 진행한다. 실험환경은 [표 4-1]과 같다.

[표 4-1] 실험 환경

Division		Contents
HW	CPU	Intel(R) core(TM) i5-7600K CPU @ 3.80GHz
	GPU	GeForce GTX 750
	RAM	8.00GB
	HDD	560GB
SW	OS	Windows10
	Program Language	Python, R

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

A. 실험 결과 및 비교 실험

데이터의 주제인 드라마의 속성을 활용한 감성분류를 위해 제안하는 방법과 기존 감성사전을 이용한 감성분석의 결과 값을 비교 실험을 진행한다. 먼저 실험은 2단계를 거쳐 진행한다. 첫 번째로 속성 단어가 포함된 하나의 리뷰의 긍정인지 부정인지 판단하여 결과를 확인한다. 두 번째 전처리만이 진행된 전체 리뷰 데이터를 감성사전을 이용하여 감성을 분류한 다음 긍정 단어와 부정 단어 개수를 도출하여 긍/부정을 판별하여 속성별 감성분석 결과와 비교 분석한다.

첫 번째 긍/부정 사전을 이용한 감성분류는 리뷰에서 긍정단어와 부정 단어가 몇 차례 나왔는지 분류 되어 차액 값을 통해 극성 값을 알 수 있다. 속성이 포함된 문장을 'x'라 하고 'c'는 긍정인지 부정인지 판단하는 'class'라 지칭한다. 그러므로 문장 'x'가 긍정이면 +1의 값을 가지고 부정이면 -1의 값을 가진다. 수식 (1)은 긍정과 부정을 구하는 계산식을 나타낸다.

$$\begin{aligned} &p(x_1,x_2,x_3,\ldots x_n|c) = p(x_1|c)+p(x_2|c)+p(x_3|c)\ldots p(x_n|c) \\ &p(x_1,x_2,x_3,\ldots x_n|c) > 0 \text{ 긍정} \\ &p(x_1,x_2,x_3,\ldots x_n|c) < 0 \text{ 부정} \end{aligned} \tag{1}$$

[표 4-2] 긍/부정 결과

Review	배우	패션	스토리	합	POS/NEG
rev1	긍정(+1)	x	긍정(+1)	+2	POS
rev2	부정(-1)	x	부정	-2	NEG
rev3	긍정(+1)	패션(+1)	긍정(+1)	+3	POS
⋮	⋮	⋮	⋮	⋮	⋮

긍정개수 = +1, 부정개수 = -1을 이용하여 하나의 리뷰씩 속성별 감성분류를 진행 하였다. [표 4-2]와 같이 감성 값을 확인 할 수 있다.

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

[표 4-3] 속성별 감성분류 결과 예시

속성	리뷰
배우	서현진 배우님 너무 아름다우시고 친절하시고 이민기 배우님도 너무 다정하신 분! 뷰티인사이드 첫방 정말로 기대하고 있어요. 첫화부터 대박 터지길 꼭 본방사수하고 주위에도 홍보할게요! 류화영 보기 거북합니다. 류화영배우..안나오면 안되요? 그렇게 여배우가 없나요..
패션	드라마뷰티인사이드 시상식 드레스인데 웨딩드레스 빰치게 이쁜 드라마뷰티인사이드패션 병원씬에서 등장한 노랑 가디건 과 반지 드라마 보다 보면 꼭 빼놓을 수 없는 게 드라마 속 여배우 패션이죠 얼마 전, 뷰티인사이드 서현진 패션은 포스팅을 했는데 개인적으로 더 마음에 드는 취향
스토리	신사의 품격, 태양의 후예, 상속자들, 도깨비 등 지금까지 김은숙작가님과 함께한 임메아리작가님 히스토리를 보며 기대하고 있습니다. 대박나시길 기원합니다. 일단 저는 영화 뷰티인사이드를 정말 재미있게 봐서 드라마도 무척 기대를 했습니다.

[표 4-3]은 속성별 감성분류 결과 예시이며 '배우'의 리뷰를 확인하면 부정적인 감성어휘가 많이 포함되어 있으며 '패션'과 '스토리'는 긍정의 감성어휘가 많이 포함 되고 있음을 확인 할 수 있다. 두 번째 드라마 전체 리뷰를 감성분석 하여 긍정단어 개수와 부정단어 개수를 확인한다. 그리고 제안하는 속성별 감성분류와 결과를 비교하여 수치를 확인한다.

[표 4-4] 기존 감성분석과 속성별 감성분류의 비교실험 결과

Review	기존 감성분석		제안하는 속성별 감성분석							
			배우		패션		스토리		TOT	
	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS	NEG
전체 리뷰	3359	2159	959	1361	465	171	707	422	2131	1954
Ratio(%)	60.8	39.1	41.3	58.7	73.1	26.9	62.6	37.4	52.2	47.8
POS/NEG	POS		NEG		POS		POS		POS	

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

[표 4-3]는 속성별 감성분류 결과와 기존 감성분석 비교 실험에 대한 결과를 나타나낸다. 비교 실험은 항목별 감정어휘에서 긍정단어와 부정단어가 차지하는 비율을 수치화 하여 비교한다. 기존 감성분석으로 전체 리뷰를 분석결과 '긍정'인 결과를 확인할 수 있다. 그러나 기존 감성분석은 드라마 리뷰는 '긍정'인 드라마지만, 제안하는 속성별 감성분류 결과를 확인하면 드라마를 선택함에 있어서 '배우'에 대한 속성을 중요시하는 시청자에게는 잘못된 정보일 수 있다. 리뷰 데이터의 모든 감정 어휘를 분석하기 때문에 이와 같은 결과가 도출 될 수 있다. 속성을 기반으로 필요한 문장만을 감성분석하게 된다면 불필요한 감정 어휘를 포함하지 않으며 각 속성에 대한 결과를 제공하기 때문에 취향을 고려하는 분석이 가능하다는 것을 알 수 있었다.

[표 4-5] 속성별 드라마 인기 순위 예시

No.	배우	패션	스토리
1			
2			
3			나의 아저씨

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

사람들의 취향을 고려한 속성별 감성분류 결과인 긍/부정 결과 값을 통하여 사람들의 취향에 맞는 드라마를 선택할 수 있도록 속성별 드라마 인기순위를 제시할 수 있다. [표 4-5]은 속성별로 드라마들의 인기 순위를 나열한 예시이다. 본 논문에서 '뷰티인사이드' 리뷰를 분석 데이터로 사용하여 나온 결과로 '배우'는 부정적인 의견이지만 '패션'에서는 높은 긍정의 값을 갖기 때문에 '패션'의 높은 순위를 가지고 있지만 '스토리'에 관해서는 다른 드라마 비해 낮은 순위를 가질 수 있다. 그러므로 사람들에게 속성별 인기순위를 통하여 개인의 취향에 맞는 드라마를 선택할 수 있도록 한다.

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

V. 결론 및 제언

지상파, 케이블, 인터넷 웹 드라마를 통하여 수많은 드라마들이 반영되고 많은 사람들에게 사랑받고 있다. 수많은 드라마 중 어떤 드라마를 시청할지, 재미있는 드라마가 무엇인지 고민하는 사람들에게 시청률이 높다고 드라마를 선택하기에는 어려움이 있다. 과거 시청률이 높으면 인기 있는 드라마였었지만 현재 인터넷과 스마트폰의 발달로 인하여 드라마를 시청할 수 있는 방법이 다양해지면서 시청률이 낮더라도 재미있고 사람들에게 인기 있는 드라마일 수 있다. 그러므로 사람들은 드라마에 대한 사람들의 의견을 중요시 하게 되었으며 드라마 시청자 리뷰 또는 블로그, 실시간 Talk를 통해 정보를 얻게 된다. 하지만 한 사람이 모든 리뷰를 읽고 분석하기에는 많은 어려움이 있기 때문에 리뷰의 감정을 분석하는 감성분석 방법이 필요하게 되었다.. 감성분석은 텍스트 데이터에서 긍정, 부정, 중립 중 어떤 견해를 갖는지 판별하는 방법이다. 기존 감성분석은 드라마 전체 리뷰의 감성어휘를 추출하여 이 드라마가 긍정인지 부정인지 중립인지 판단하기만 했지만 모든 리뷰가 드라마와 관련도니 문장만이 있지 않을뿐더러 무엇이 좋은지 무엇이 나쁜지도 알 수 없다. 그러므로 본 논문에서는 시청자들이 드라마를 시청한 후 남긴 리뷰 데이터에서 감성분석을 진행하였고 리뷰 데이터에서 의미 없는 문장을 배제하고 효율적인 감성분석을 위해 속성별 감성분류를 진행 하였다. 속성이란 드라마에서 중요시 되는 배우, 스토리, 패션, OST 와 같은 드라마의 중요 요소다. 본 논문에서는 사람마다 드라마를 선택하는 속성이 다르기 때문에 크게 세 가지를 정하여 시청률에 크게 영향을 끼치는 배우, 스토리와 드라마의 패션은 사람들에게 모방심리를 크게 작용하기 때문에 선정하게 되었다. 그리고 드라마에서 '배우'라는 단어에는 배우의 실제 이름과, 극중 이름 또한 의미하기 때문에 속성을 의미하는 유사한 단어 또한 Word2Vec을 사용하여 추출한다. 추출된 단어를 이용하여 속성사전을 구축하여 속성별로 문장을 분류 할 수 있다. 분류된 속성별 문장들을 감성 사전으로 긍/부정 판별하는 방법을 제안하였다.

본 연구에 제안한 속성별 감성분류 결과와 기존 감성분석 결과의 비교 실험을 통하여 드라마를 시청하려하는 사람에게 기존 감성분석 방법은 잘못된 정보를 제공할 수 있음을 확인 하였다. 기존의 감성분석 방법은 모든 리뷰의 감성어휘를 통하여 긍정인지 부정인지 결과를 알 수 있지만 속성별 감성분류는 속성별로 긍정인지 부정인지 결

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

과를 확인 할 수 있기 때문에 원하는 정보를 제공 할 수 있다. 그러므로 드라마 '뷰티인사이드' 리뷰를 대다수의 사람들이 중요시하는 키워드 3가지를 주제로 정하여 각각 감성분석을 실험하였다. 실험 결과 '배우'의 경우 부정의 의견이 나왔지만 '패션'과 '스토리'는 긍정의 의견이 나왔다. 이와 같이 무엇이 좋은지 싫은지 알 수 있지만 전체 리뷰를 감성분석 하게 되면 단지 이 드라마는 긍정이라는 결과만 확인 할 수 있다. 즉 '뷰티인사이드'는 '배우'를 중요시하는 사람들에게 '배우'도 좋다는 결과를 내포할 수 있기 때문에 세분화된 속성별 감성분류 방법의 필요성을 검증하였다.

본 연구의 한계는 다음과 같다. 긍정사전과 부정사전 이용하여 긍정단어 개수와 부정단어 개수의 차를 통하여 이진분류를 했기 때문에 결과가 도출된다. 그러나 모든 감성어휘가 같은 감정값을 가질 수 없기 때문에 이를 보안하기 위해 기존 공개사전에서 제공하는 수치화된 감정어휘와 사용자 평가점수가 아주 높은 리뷰의 고빈도 단어를 긍정 표현으로, 사용자 평가점수가 아주 낮은 리뷰에서 고빈도 단어를 부정 표현으로 감정어휘를 추가하여 사전을 구축하고 모든 어휘를 점수화 하여 어휘마다 차별을 둘 수 있다.

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

참고문헌

[1] 강명현, "국내 드라마의 속성에 따른 방영기간의 차이 연구." 지역과 커뮤니케이션 Vol.16.3 pp.3-30, 2012.

[2] 박승수, "온라인 리뷰와 머신러닝을 활용한 드라마 시청률 예측 모델 연구." 연세대학교 정보대학원 학위논문, 2017

[3] 남민지, "SNS 해시태그를 이용한 사용자 감정 분류 방법에 관한 연구" 조선대학교 산업기술융합대학원 석사학위논문, 2015

[4] 이상훈, 최정, and 김종우. "영역별 맞춤형 감성사전 구축을 통한 영화리뷰 감성분석." 지능정보연구 Vol.22.2 pp.97-113, 2016

[5] 김정호, "영역별 감성사전 구축을 위한 그래프 기반 방법" 한국항공대학교 대학원 박사학위논문, 2015

[6] 서정열, "오피니언 마이닝의 제품 특징을 이용한 극성 분류" 서울과학기술대학교 산업대학원 석사학위논문, 2014

[7] 차승준, "소셜미디어 오피니언 마이닝에 기반한 기업의 위기관리에 관한 연구", 한국정보과학회 학술발표논문집, Vol39.1C, pp.142-144, 2012

[8] 장은지, "소셜미디어에서의 인터랙션정보기반 감성분석기법 연구", 서울여자대학교 일반대학원 석사학위논문, 2018

[9] 안정국, and 김희웅. "집단지성을 이용한 한글 감성어 사전 구축." 지능정보연구 Vol.21.2, pp.49-67, 2015

[10] 김문형, et al. "KOSAC (Korean Sentiment Analysis Corpus) 한국 감정 및 의견 분석 코퍼스" 한국정보과학회 학술발표논문집, pp.650-652

[11] KNU 한국어 감성사전. http://dilab.kunsan.ac.kr/knusl.html

[12] 이상훈, 최정, and 김종우. "영역별 맞춤형 감성사전 구축을 통한 영화리뷰 감성분석." 지능정보연구 Vol.22.2, pp.97-113, 2016

[13] 유은지, "오피니언 마이닝의 정확도 향상을 위한 주제지향 감성사전 구축 및 활용: 주가예측 적용 사례", 국민대학교 비즈니스IT전문대학원 석사학위논문, 2014

[14] 신수정, "글에서 감정을 읽다 감성 분석의 이해", IDG Tech Report

[15] Liu, Bing. "Sentiment analysis and opinion mining." Synthesis lectures on human language technologies Vol.5.1, pp.1-167, 2012

Collection @ chosun

조선대학교
CHOSUN UNIVERSITY

[16] 장재영, “온라인 쇼핑몰의 상품평 자동분류를 위한 감성분석 알고리즘” 한국전자거래학회지, Vol.14.4, pp.19-33, 2009

[17] Word2Vec 문장분류. https://ratsgo.github.io/natural%20language%20processing/2017/03/08/word2vec/

[18] 어균선, 이건창, “Word2vec을 이용한 오피니언 마이닝 성과분석 연구”, 한국콘텐츠학회 종합학술대회 논문집, pp.7-8, 2018

[19] 장환석, 장은영, 정광용, “Word2Vec를 이용한 감성어 분석 방법”, 한국정보과학회 학술발표논문집, pp.661-663, 2017

[20] 명은진, “감성사전 기반 Word2vec 자질을 이용한 감성 분류 시스템”, 이화여자대학교 대학원:빅데이터분석학협동과정 석사학위논문, 2018

[21] “스타의 역량이 방송프로그램 성공에 미치는 영향”, 한국콘텐츠진흥원, 2014

[22] 권호영, “드라마 성공 요인 분석”, 한국콘텐츠진흥원 KOCCA 연구보고서 09-50, 2009

[23] 박수연, “드라마 속에 나타난 남성 패션스타일 분석”, 한국엔터테인먼트산업학회 논문지, Vol.6.4, pp.164-169, 2012

Collection @ chosun