

텍스트 구조적 요소에 따른 감성 분석 성능의 차이: 셰익스피어 『햄릿(Hamlet)』을 통한 어휘 사전과 RNN 기반 비교 *

Structural Factors in Sentiment Analysis: Lexicon vs. RNN on Shakespeare's Hamlet

김민 (Kim, Min)**

 0009-0006-6938-3152

목차

1. 서론
2. 연구방법
 - 2.1 감성사전 기반
 - 2.2 RNN 기반
3. 결과분석
 - 3.1 감성사전기반 정확도 우위의 예문
 - 3.2 RNN 기반 정확도 우위의 예문
4. 결론

초록

해당 연구는 문학 텍스트 셰익스피어의 『햄릿(Hamlet)』 한글 번역본을 바탕으로 감성사전 기반과 인공신경망 RNN 기반의 두가지 감성분석(sentiment analysis)을 수행한다. 두 결과를 바탕으로 특정 텍스트 복잡성에 따른 결과 도출의 적합성 및 오류의 정도를 비교하며, 전산기법의 적용이 어려워지는 구조와 사전기반 감성 분석과 RNN 기반 각각이 더 정확한 결과를 도출하는 문장에 특정 패턴의 여부를 관찰한다. 어휘기반과 구문 및 문법구조를 아우르는 총체적 기반으로 문장을 분석하는 각각의 방식을 비교대조한 결과를 통해 역으로 자연어 패턴이 이해, 처리 되는 주요 방식을 추론할 수 있으며 텍스트 복잡성을 인지하는데 있어서 전산적 분류결과와 직관적 관찰의 차이를 분석할 수 있다.

주제어: 감성분석, RNN, 사전기반, 텍스트 패턴, 셰익스피어 코퍼스

*본 논문은 2023 년 <문화와융합> 가을 학술대회 발표문을 토대로, 내용 전반에 걸쳐 보완 및 확장한 연구입니다.

**The author is a PhD candidate at Yonsei University, specializing in the philosophy of language, NLP, and AI-based language models. The research draws on both literary theory and computational techniques, bridging the humanities and engineering to investigate how meaning emerges across linguistic systems—human and artificial. Through the lens of digital humanities, the work aims to reframe and expand our understanding of language in the age of intelligent systems. Correspondence: text.research.yonsei@outlook.com

Abstract

This study conducts sentiment analysis on the Korean-translated text of Shakespeare's Hamlet using two approaches: a lexicon-based method and a Recurrent Neural Network (RNN)-based method. By comparing the analytical outcomes of both approaches, the study evaluates their respective effectiveness in handling texts of varying linguistic complexity. Specifically, it investigates which types of lexical, syntactic or grammatical patterns lead to greater classification errors or reveal strengths in either model. And the research aims to uncover systematic tendencies where semantic-focused (lexicon-based) and syntax-aware (RNN-based) approaches diverge, thereby offering insights into the mechanisms of natural language comprehension in computational systems. Finally, by suggesting literary text as a crucial for evaluating and refining NLP methodologies, this work helps to comprehend the gap between intuitive human language interpretation and machine-based language processing.

Keywords: Sentiment Analysis, RNN, Lexicon-Based Approach, Text Pattern, Shakespearean Corpus.

1. 서론

문학 텍스트 기반의 텍스트 마이닝 연구는 국내외에서 활발히 진행되고 있다. 특히 셰익스피어 비극 작품을 대상으로 한 텍스트 마이닝 연구는 국내외에서 활발히 이루어지고 있으며, 인물 네트워크 분석, 주제 모델링, 스타일 분석, 패턴 탐지, 단어 빈도 기반의 상호작용 분석 등 다양한 접근법이 활용되고 있다. 이 중 주요 등장인물 간의 감정 분석은 주된 주제 중 하나다. 예를 들어, “대화 코퍼스를 통한 셰익스피어 비극 작품과 주요 남녀 등장인물 간의 감정 분석”(장세은 외, 2019), “딥러닝을 이용한 셰익스피어 작품의 감정 분석”(서혜진 외, 2019), “셰익스피어 Othello 등장인물 대사에 나타난 감성 분석”(홍유림, 2020) 등의 연구는 주로 인공지능 및 텍스트 마이닝 기법을 활용하여 문학 텍스트의 감성 구조를 정밀하게 분석하고, 이를 해석하는 데 집중해 왔다. 또한, “Measuring Emotional Temperatures in Shakespeare's Drama”(Culpeper et al., 2018)과 “Revolutionizing Sentiment Analysis in Literary Text ‘The Immortals of Meluha’ through a Hybrid CNN-RNN Architecture and Advanced Feature Techniques”(Rengarjan et al., 2023) 등의 연구는 감성 분석을 보다 정교하게 수행하기 위한 모델링 기법을 제안하였다.

그러나 이러한 연구들은 대체로 기술을 문학 연구의 보조 도구로 활용하여 텍스트의 의미 해석을 정밀화 하는데 초점을 맞추고 있으며, 디지털 인문학의 연구 목적이 텍스트 마이닝을 통한 문학 독해 및 원거리 읽기(distant reading)의 타당성 확보에 한정되는 경향이 있다. 이러한 접근은 문학 텍스트가 가지는 방대한 잠재성에 비해 디지털 인문학이 여전히 문학의 해석이라는 동일한 범주 안에 머물러 있음을 의미하며, 전산적 문학 연구에 대한 새로운 시각의 모색이 필요함을 시사한다.

이에 본 연구는 기존 연구와 차별화하여 문학 텍스트를 단순히 분석 대상이 아니라, 언어 데이터 객체로 활용하는 방향을 모색한다. 즉, 문학적 의미 해석을 넘어, 언어 이해 및 자연어 처리(NLP) 기술의 성능을 평가하고 분석하는 유효한 자료로서 문학 텍스트를 적극적으로 활용하는 것을 목표로 한다. 셰익스피어의 희곡 텍스트를 연구의 코퍼스로 활용하는 이유는 희곡 장르가 다른 문학 장르와 달리 서술체(narrative prose)보다는 인물 간의 대화(dialogue)로 이루어져 있어, 구조적으로 발화자의 담화(distributed discourse)를 분석하는 데 적합하기 때문이다. 특히 희곡의 대사는 실제 음성 언어와 유사한 문장 구조를 지니며, 대화의 흐름 속에서 감정 표현이 직접적으로 드러난다. 따라서 감성 분석을 수행할 때 서술적 배경 요소나 내레이션 없이 등장인물의 발화만을 기반으로 감정적 패턴을 분석할 수 있다는 장점이 있다. 작품 선정에 있어서는 감성 분석이 보다 효과적으로 이루어질 수 있도록 어휘 선택의 감정적 명암(emotional contrast)이 명확히 대비되는 작품을 고려하였다. 셰익스피어의 비극(Tragedy) 장르는 희극(Comedy)에 비해

감정의 극적인 변화를 직접적으로 보여주는 구조를 가지며, 주요 등장인물의 대사에서 강한 감정적 대비가 나타나는 특징이 있다. 그중에서도 『햄릿』은 다른 비극보다도 감정의 대비가 가장 선명하게 드러나는 작품으로(Pearce and Duffy, 2010; Steenbergh, 2011), 사색적인 독백과 감정적인 대사가 혼재되어 있어 감성 분석을 수행하는 데 있어 중요한 실험적 가치를 지닌다. 햄릿은 "To be, or not to be"(살 것인가, 죽을 것인가)와 같은 독백을 통해 실존적 고뇌를 드러내는 한편, 극 중에서 분노, 절망, 광기, 조롱 등 감정적 색채가 급격히 변화하는 구간이 많다. 맥베스나 오셀로와 같은 다른 비극에서도 강한 감정이 나타나지만, 이들은 특정 감정(욕망, 질투)이 서사의 주요 원동력이 되는 반면, 햄릿은 감정의 폭이 넓고 변화가 급격한 특징이 있다. 또한, 햄릿의 대사는 셰익스피어의 다른 비극과 비교했을 때 문장 길이가 상대적으로 길고, 삽입절 및 종속절을 포함하는 복잡문이 많아 문장 구조적으로 복잡하다. 이로 인해 감성 분석에서 단순 어휘 기반의 분석과 문장 구조 기반의 분석 결과가 어떻게 차이를 보이는지 관찰하기에 적절한 연구 대상이 된다. 따라서 본 연구는 김재남 역의 『햄릿』 한국어 번역본(셰익스피어 저, 서문당, 2014)¹을 기반으로 감성 분석(감성사전/인공신경망 RNN)을 수행하고, 어휘(lexicon) 기반 분석과 문장 구문(syntax) 및 문법 구조(grammatical structure)를 포함하는 RNN 기반 분석 결과를 직접 비교·대조한다. 이를 통해 텍스트 복잡성이 자연어 처리 기술의 성능에 미치는 영향과 분석 오류를 유발하는 특정 패턴을 탐색하여, 인간의 언어 인지와 전산적 자연어 처리 사이의 간극을 이해하는 데 기여하고자 한다.

2. 연구방법

김재남 역의 『햄릿』 한국어 번역본 텍스트 파일에 대하여 Visual studio상에서 정규식(regex)을 통해 일차적으로 삭제되어야 하는 부분을 직접 확인하며 단계별로 제거하는 전처리 작업을 했다: 1)텍스트 앞부분 [인물이름] 모두 찾기 및 공백으로 대체 (삭제): $W[^W]*W$ Ws*, 동일 방법으로 2)()안 지시문 모두 삭제: $W([^W]*W)W$ s*, 동일 방법으로 2)()안 지시문 모두 삭제: $W([^\wedge W]*W)W$ s*, 3) 육안으로 보이는 구두점 제거: [,!---"<>], 4)숫자 제거: [Wd], 5) 해당 line이 길 경우(if length>45) 마침표를 기점으로 문장을 분리, 6)5번 단계 진행을 위해 마침표(period)를 이후 제거, 7)분석 대상으로 지나치게 짧거나 긴 문장 제거(length<13, length>62).

[장] 1장
 ((엘시노어 궁전 앞 망대. 프랜시스코가 보초를 서고 있다. 그 앞으로 버나아도 등장))
 [버나아도] 게 누구얏?
 [프랜시스] 넌 누구야? 정지. 이름을 대라
 [버나아도] 국왕 만세!
 [프랜시스] 버나아도 장교님이시오?
 [버나아도] 그렇다. |
 [프랜시스] 꼭 시간에 대오셨군요.
 [버나아도] 자아 교대다. 가서 자게나.
 [프랜시스] 고마워. 원 육살할 추위도 뻘속가지 얼었네.
 [버나아도] 이상 없었나?

게 누구얏
 넌 누구야 정지 이름을 대라
 국왕 만세
 버나아도 장교님이시오
 그렇다
 꼭 시간에 대오셨군요
 자아 교대다 가서 자게나
 고마워 원 육살할 추위도 뻘속가지 얼었네
 이상 없었나

그림 1. 전처리 전(좌) 후(우) 텍스트 파일

또한, 텍스트 복잡성(Text Complexity)의 경우, 서론에서 제시한 바와 양적 측면에 해당하는 문장의 길이(+ 단어의 개수도 포함할지 여부)가 함께 출력될 수 있도록 하였으며, 질적 측면에 있어서는 기본적으로 문학 텍스트인 만큼 지나치게 단순한 문장은 빈번하지 않지만 연구자가 직접 문장의 난도를 분석하는 방식을 택한다. 웹 기반 텍스트 복잡성 지수 측정 프로그램으로 Lexile(<https://lexile.com>) 혹은 ETS 개발 프로그램(Textevaluator(<https://textevaluator.ets.org>))이 있지만 이는 영어 텍스트 분석에 사용 가능한 것이며, 한글 대상 프로그램이 개발되기도 했지만(www.korean-lab.org) 이는 개별 문장 단 위보다 전체 텍스트 집단을 산출하는 방법으로 본 연구에서는 사용하지 않았다.

2.1 감성사전 기반(Dictionary-based)

타겟 텍스트가 일차적으로 만들어지고 나면(916개 문장) 이를 대상으로 감성사전(KnuSentiLex)² 기반 감성분석³을 진행한다. KoNLPy의 Okt(Open Korean Text) 와 Mecab 을 이용해서 토큰화와 형태소 분석을 진행하며 파이썬 re 모듈 내 re.sub method를 통해 (re.sub('[^WsWdWw]', '', sentence)) 분석에 방해가 되는 문자 외의 표기를 없애는 정제작업을 다시한번 거친다. mecab.morphs method를 통해 문장을 형태소 단위로 쪼개고 나면 반복(iteration)을 거치면서 형태소 단위 단어에 감성사전에 해당하는 단어가 있는 경우 그것이 긍정 단어에 해당할 때 마다 +1, 부정 단어에 해당할 때마다 -1 로 문장 별로 값을 단순 산출한다. 즉, 어휘의 긍/부정 정도에 대한 가중치는 따로 반영되지 않았기 때문에 주관성의 오류를 범하지 않는 가장 단순한 방법임과 동시에 실제적 섬세함이 떨어진다는 단점이 있다. 산출된 값이 각 문장 옆에 출력되어 저장될 수 있도록 코드를 구성하고 나면 0 (중립)을 기점으로 긍정(+) 과 부정(-)의 정도를 해석할 수 있다. 결과값의 범위는 -4~+6으로 관찰된다.

2.2 RNN 기반

다음으로 인경신경망 RNN 기반⁴으로 감성분석을 진행하였다. 감성점수가 표기(라벨링)된 네이

버 영화 리뷰 코퍼스(NSMC: Naver Sentiment Movie Corpus)데이터 로 훈련을 거친 모델을 활용하여 분석하고자 하는 타겟 텍스트의 각 문장에 대한 예측 값을 최종적으로 ~% 확률로 긍정/부정과 같이 추출할 수 있다. 한국어 기반 훈련데이터로 활용할 수 있는 감성점수 태깅 자료는 AIHub출처 감성 대화 말뭉치, 국립국어원의 한국어감성분석(KOSAC: Korean Sentiment Analysis), 네이버 영화리뷰 코퍼스(NSMC: Naver Sentiment Movie Corpus) 등으로 작품 텍스트 분석을 위한 훈련 데이터로서 문장의 어휘, 구조적 난도 측면에서 적합하지 않을 수 있다. 그러나 문학 데이터에 대한 별도의 라벨링 작업과 데이터 구축은 시간과 비용 면에서 한계가 있다. 무엇보다 본 연구는 문학 텍스트 자체의 감성 분석이 목적이 아니라, 이를 일반 담화 코퍼스의 예시로 삼아 격식 있는 문어체를 사용하는 화자의 언어를 표본으로 분석의 타당성과 오류 정도를 측정하는 데 목적이 있다. 이에 따라, 일상적이며 담화적 특성을 지닌 문어체인 리뷰 데이터를 훈련 자료로 사용하는 것이 연구 목적에 더 부합한다고 판단하였다.

다음의 과정은 구체적 기술을 위해 자연어처리 코드에 관한 기본 설명을 포함하고 있다. 감성 사전 기반과 동일한 방식으로 한국어 자연어처리를 위한 konlpy 패키지가 설치된 상태에서 형태소 분석 모듈 Okt 와 Mecab를, 그리고 필요한 pandas, numpy, tensorflow, re 등의 추가 라이브러리를 미리 불러온다. 앞서 언급한 훈련데이터를 pandas frame으로 불러온 상태에서 리뷰가 담긴 ‘document’ 열에 정규식을 이용해 필요한 문자만 남길 수 있도록 전처리를 진행한다. 또한, 빈 문 자열의 경우 Nan 값으로 대체되게 하여 Dataframe에서 삭제시킨다. 모델 성능(performance)을 확인하기 위해 필요한 테스트 데이터(NSMC중 split 된 부분)에 대해서도 같은 방식으로 데이터 내에 중복되는 행이 있는 경우 삭제하고 ‘document’ 열에 대한 기호, Nan값 제거를 진행한다. 다음으로 전처리의 연장선에서 Tokenizer는 훈련데이터에서 ‘document’ 열의 텍스트를 작은 단위로 쪼개어 빈도 상위5000개의 단어에 대해 빈도정보를 반영한다(더 보편적인 단어일수록 낮은 정수 값을 부여받는다.). 정수를 부여하여 텍스트가 일련의 정수(sequences of integers)로 표현되게끔 한다. 상위 5000개로 개수를 한정하는 이유는 dimensionality를 통제함으로써 과정의 효율성을 유지하기 위함이며 (text_to_sequences())는 원데이터(raw data)가 머신러닝 모델로 처리될 수 있는 형식이 되게 하기 위함이다. 다음으로 Padding and Truncating Sequences 과정을 통해 정수로 변환된 텍스트의 길이를 맞춰주는 작업을 진행한다. 정수의 개수가 부족한 경우 비어있는 자리에 0을 채워주고, 오히려 길이가 넘는 경우 뒷부분을 잘라준다. 또한 label열의 값이 숫자가 아닌 경우 NumPy array 변환이 필요하지만 이미 주어진 데이터의 label은 0(부정), 1(긍정)으로 변환된 상태이다. 테스트데이터에 대해서도 같은 전처리 과정이 진행되지만 중요한 점은 훈련데이터에 fit된 Tokenizer를 그대로 사용함으로써 word mapping의 일관성을 유지해야 한다. 다음으로, model architecture 설정에 관해서는 sequential model에 임베딩 층(embedding layer)을 추가해주며 이때 이전에 정수로 인코딩된 단어가 벡터로 변환된다. 벡터 표상은 단어의 의미와 문맥을 포착하게 되며 단어를 벡터로 맵핑(mapping) 함으로써, 임베딩 층은 신경망이 단어들 사이의 유사성과 긴밀한 관계를 인지하고 이해할 수 있게끔 한다. 이 부분이 단순 단어사전 기반 감성분석을 진행하는 것과 가장 큰 차이가 생기는 이유이다. 모델에 양방향 LSTM(Long-Short Term Memory / RNN의 종류)층을 64단위로 쌓아주게 되면 해당 층의 결과가 다음 층으로 전달되며, 추가로 32단위의 LSTM을 쌓아⁵ 비교적 모델을 단순화하되 주요패턴 정보를 더 파악하여 결과도출로 이어질 수 있도록 구성하며 이는 계열 의존성에 대한 정보를 포착하는 과정이다. 임베딩 dimension(=128) 과 LSTM 단위(unit) 설정은 모델의 성능 효율성과 overfitting 가능성을 줄일 수 있는 보편적 구성으로 선택을 하였다. 마지막으로 Dense 층의 추

가에 대한 언급을 할 수 있다. Dense 층의 각 뉴런은 이전 층에 존재하는 모든 뉴런과 연결되어 있으며 한개의 단위로 존재하여 이분지 분류 (binary classification task)에 적합하다. 시그모이드 함수를 활성화함수로 사용하여 긍정 혹은 부 정으로(1 or 0)에 분류될 가능성을 측정하도록 한다. 추가로, 가장 대중적인 adam optimizer를, 손실함수는 이분적 분류에 적합한 binary_crossentropy 로 환경설정을 한다. Earlystopping 설정을 통해 더 이상 loss 가 향상되지 않으면 훈련이 중지되도록 한다(patience=3). batch_size=64, epoch=2를 최종 조건으로 Test Accuracy: 0.7738122344017029 결과를 얻은 모델로 햄릿 데이터 대상으로 분석을 진행한다.

3. 결과분석

다양한 문장 대상의 감성사전 기반과 RNN 기반 결과를 비교/분석한다. 감성사전 기반과 RNN 기반 상위 등을 기준으로 결과 정확도를 비교분석하기 이전에 감성사전기반 도출 결과의 값이 >=4의 큰 값으로 얻어진 문장의 경우 RNN 기반 감성분석 결과도 일관적으로 도출되는지를 관찰해보고자 한다.

표 1. 사전기반 도출 결과의 절대값이 큰 문장에 대한 성능평가 예시

sentence	sent/dic	RNN
(1)정다운 말씀이라고 허어 아 이렇게 험란한 세상 에 정다운 소리가 있음직하나 그래 그게 진짜 같이 들리었나	6	65.04% 확률로부정리뷰.
(2)마음 편히 숨을 거둔 사람처럼 미사를 드리고 명복을 빌어서는 오히려 상태의 예법을 모독하는 짓이오	5	90.64% 확률로긍정리뷰.
(3)두사람의 역량을 잘 아니까 그렇게는 생각하지 않는다	5	58.07% 확률로긍정리뷰.
(4)일순에 천리를 달리는 사랑의 날개보다 더 빨리 원수를 갚으러 날아가겠으니	4	57.62% 확률로긍정리뷰.
(5)에이 꼴사나운 바보 늙은이	-4	98.57% 확률로부정리뷰.
(6)아무리 그대가 얼음같이 정절하고 눈같이 순결해도 세상 구설은 면치 못할꺼라는	4	51.38% 확률로긍정리뷰.
(7)두 형제의 얼굴을 그린 이 초상화와 저것을 이 미간에 서린 기품 이 어른이 바로 당신의 남편이었습니다	4	74.17% 확률로긍정리뷰.
(8)햄릿의 죄를 들어내서 다스리지 못하는 이유는 전혀 왕비 때문 그 밖에 또 하나의 이유	-4	64.69% 확률로긍정리뷰.

감성분석 기반 도출 값이 +6으로 오히려 가장 큰 (1)과 -4 값의 (8)에 대해서는 상반되는 결과가 도출되었으며, 그 외의 문장에 대해서는 상응하는 값이 도출되었다. 그러나, 상응하는 결과의 경우에도, 감성 분석 기반 도출값의 절대값이 크다고 해서, 즉 더욱 명시적인 결과를 산출하더라도 RNN기반에서 큰 값으로 긍/부정을 산출하지 않는다. 이 경우 일관성은 대략적으로 보장이 되지만 분석 결과의 정밀도는 확보되지 않고 있음을 알 수 있다. (1)과 (8)의 경우 상반된 결과 값 중 각각 RNN, 감성사전이 적합한 분석을 도출하고 있으며 (2)와 (6)의 경우는 감성사전과 RNN 기반 모두 틀린 결과값을 도출하고 있다. (2)의 예문은 ‘모독하는 짓이오’의 구문 결말로 명시적인 부정적 의미를 표현하고 있음에도 선행하는 ‘편히’ ‘명복’ ‘빌다’ ‘예법’ 등과 같은 긍정적 단어들로 인해 높은 점수가 산출된 것으로 추정된다. 이는 문장단위처리가 동반되지 않은 결과로 해석될

수 있는데 구문 구조 단위를 분석 근거에 포함하는 RNN의 경우에도 90.64% 긍정으로 높은 오류수치가 산정되는 것으로 보아 의미 함축도가 명시적인 키워드를 많이 포함할수록 RNN 기반의 경우에도 어휘(lexicon) 의존도가 높아지는 것은 아닐지 하는 질문을 낳는다. 해당 내용과 관련한 논의는 다음의 연구들에서도 참조하여 볼 수 있다. Kauf et al. (2024)은 인공 신경망 기반 언어 모델과 인간의 언어 처리 뇌 영역 간 반응 유사성을 분석한 연구에서, 문장의 구문 구조보다 는 어휘-의미적 요소가 더 큰 기여를 한다고 보았다. 이 연구는 ANN 기반 모델이 문맥을 고려함에도 불구하고, 특정 단어의 의미적 명시성이 판단에 주요한 영향을 미친다는 점을 뒷받침한다. 한편, Wilcox et al. (2018)은 RNN 언어 모델이 문법적 의존성을 학습하는 능력을 실험적으로 평가하면서, 복잡한 구문 구조에서는 인간과 다른 방식으로 반응하거나, 제약 조건을 온전히 반영하지 못하는 경향을 보인다고 밝혔다. 이러한 연구들은 RNN 기반 감성 분석 모델이 구문 정보를 포함하고 있더라도, 명시적인 감성 어휘의 빈도나 강조 정도에 따라 결과 해석에 있어 어휘 중심적인 편향이 여전히 발생할 수 있음을 시사한다. (6)의 경우도 동일한 이유로 추정이 가능하지만 이 경우는 RNN에서 상대적으로 정확한 값을 도출하는 점이 다르다. (2)의 ‘오히려’ 보다 (6)의 ‘아무리 ~’ 가 제시하는 구문 구조 범위 내 의미 힌트가 강하였음을 추론할 수 있다. ‘아무리~해도’와 같은 표현은 단일 어휘가 아니라 구문적 틀(syntactic frame)로 작용하며, 문장 전체의 감성 귀결을 강하게 암시하는 구조적 신호를 제공한다. 이 구문은 조건과 역설의 구조를 내포하고 있어, 이 후에 이어지는 부정적 혹은 반전된 감성 표현을 예고하는 역할을 한다. 이에 반해, ‘오히려’는 담화 전환 또는 대조 강조의 기능을 수행하지만, 문장의 전반적인 감성 극성 자체를 독립적으로 예고하거나 결정짓는 구조로 기능하지 않는다. 즉, ‘오히려’는 감성의 방향성을 명시적으로 제시하기보다는 이전 문맥과의 비교나 논리적 대비에 초점을 두기에, RNN이 문맥을 학습하더라도 그 자체로는 강한 감성 힌트를 제공하지 못할 수 있다. 이러한 언어학적 분석은 Construction Grammar(Goldberg, 1995)에서 제시하는 바와 같이, 특정 구문이 고유한 의미를 내포하는 의미 단위(constructional meaning)로 작용할 수 있다는 이론적 기반과 연결된다.

다음으로, 문장의 길이가 타겟 텍스트 내에서 상대적으로 가장 길고 짧은 경우에 속하는 예문에 대해 분석 정확도를 먼저 살펴보고자 한다.⁶

표 2. 가장 길고 짧은 경우에 속하는 예문에 대한 정확도 우위
(d:사전기반 우위/ r:RNN 우위, o:모두 정확도가 높은 경우/ x:모두 정확도가 낮은 경우)

sentence	length	accuracy	sentence	length	accuracy
말을 해 썩 말해 보라	13	o	어떤 보배를 가졌던가요	13	D
어이하여 견히지 않을까	13	o	저기 저 구름 낙타같지	13	d
선왕 장례식에 참례하러	13	o	어머니는 어찌된 노릇이요	13	d
내 빛은 주지도 말 것	13	o	이라 끝끝 마차야 각자	13	d
그러면 잘 있거라 햄릿	13	o	내 아버지는 어디 있소	13	d
있을 법도 한 일인걸요	13	o	이 끝에 입이 있었겠군	13	d
밖에 나다니지 못하게 해	13	o	그때까지 마나를 신중히	13	d
전혀 손도 댈줄 모릅니다	13	o	피를 보고 기절한 것이다	13	d
그 징글맞은 손가락으로	13	o	하늘은 그대를 용서할걸세	13	d
나는 옠피리아를 사랑했다	13	o	얼굴은 보지 못하였겠군	13	r
장검과 단검이라 하옵니다	13	o	좀 더 비싸게 처신해라	13	r
음모다 예잇 문을 잠가라	13	o	하여간 그렇게 해 보지	13	r
그걸 난폭하게 다뤘으니	13	x	넌 제 어미도 몰라보냐	13	r

어디 자리를 옮겨 볼까	13	x	이 무슨 잔인한 것이냐	13	r
오피리어를 안 만날테요	13	x	소리도 들리지 않으시고	13	r
자 레어티즈 말해보아라	13	x	너의 머리가 짜낸 환상야	13	r
그래 말을 걸어보았나	13	d	그 미쳤으니까 보냈지요	13	r
그리고 햄릿 왕자 애긴테	13	d	레어티스 너는 왜 나에게 그런 태도를 취하느냐 천하대장 군 허큘리스 실컷 뽑내봐도 개는 개고 고양이는 고양이다	60	r
저 햄릿 왕자님 얘데요	13	d	이렇게 아다운 아다운 아다운아 애그 몰래 바퀴 장단 에 잘도 맞네 주인집 딸을 훔쳐간게 바로 고 고약한 청지기였대요	64	r
아버님 분부대로 하겠어요	13	d	왕자님 스스로도 자기가 이상해졌단 말씀을 솔직히 고백하오나 어째서 그렇게 되셨다는 이유는 일체 언급치 않았읍니다	63	r
그것이 바로 상사병이라는 거로구나 그 병이 스스로 몸을 망치고 급기야는 무슨 무모한 짓도 저지르고 마는 거야	60	o	햄릿 만일 말이 입김에서 생기고 입김이 목숨에서 생긴거 라면 네 말을 누 설할 입김도 목숨도 내겐 없으니 안심하여라	62	r
말하면 네 얼은 빠지고 네 짧은 피가 얼어 붙을 것이며 두 눈알은 유성같이 튀어 나고 그 머리카락은 곧두설 것이다	63	x	레어티즈가 프랑스에 가 있는 동안 나도 줄 곧 연습은 해왔으니까 허나 어 찌지 가슴이 설 레어 이상야릇한 기분이로군	61	r
그 날도 내가 마음놓고 낮잠을 자고있는데 너의 숙부 그녀 석이 몰래 가까이 와 병에 든 독약을 내 귀 에 부었다	60	d	독 사 같이 빈틈없는 두 동 창 녀석들 이것들이 나를 영국으로 데리고 갑네 하고 어떤 함정으로 몰아놓겠다는 수작이지요	61	d
오 기특한 지고 이 기쁨을 축하하기 위하여 축연을 베풀고 축포를 하늘에 터트려 라 그리하면 하늘도 나의 축배에	63	d	너의 아버지는 정원에서 낮 잠을 자다가 독사에게 물려 죽었다고 거짓 사 인을 꾸며 온 백성을 속이고 있지만 여봐라	60	d

길이(length)=13(lowest)의 경우, ≥ 60 예문의 표본 개수에 차이가 있으나 상대적으로 짧은 예문에서 정확도 결과가 (감성분석 사전 / RNN 기반 모두 정확하게 도출) 높은 경우가 많은 것을 확인할 수 있다. 짧은 구문일수록 한정된 어휘개수와 영역 내에서 점수가 정확히 산정되며 이는 구문 구조, 문맥 단위가 반영되는 RNN의 경우에도 분석이 용이한 것으로 추정하게끔 한다.

길이가 가장 짧고 긴 예문에서 하나의 방법에서만 정확하게 결과가 분석된 경우 감성분석, RNN에 있어서 특정 톨의 우월 경향성이 발견되지는 않았다. 이제 각각이 우월하게 도출되는 예문 특성(하나의 결과만 정확하게 출력됨; 명시적 우위 발생)에 대해 분석의 정확도가 높은 문장과 반대로 정확도가 낮은 문장에 대하여 유의미한 규칙을 발견할 수 있는지 여부를 관찰한다.

3.1 감성사전 기반 정확도 우위의 예문 (표 3 참조)

감성분석 결과가 어떤 쪽에 치우치지 않는 중립적 예문에 대해서는 감성사전 기반이 정확도 우위에 있음을 확인할 수 있다. 대표적으로 ‘(6)소신을 프랑스로 돌아가게 하여 주옵소서’와 ‘(11)아 벌써 날이 새는구나’와 같은 예문에서 RNN기반 감성분석은 90.33% 확률의 긍정으로 도출하고 있으며 특히 길이가 짧고 감성적으로 중립적인 문장에 한해서는, 문맥과 구문 구조를 감성 반영하는 것이 오히려 결과의 오류 편향, 특히 긍정적인 방향으로의 과잉 일반화를 유발할 수 있다는 점이 확인된다. 이러한 결과는, 문맥 정보가 감성적으로 애매한 문장의 의미를 보완하는 데 도움을 준다는 일반적인 기대와는 상반되는 현상이다. 일반적으로 문맥은 의미의 공백을 메워주는 도구로 작용하지만, 중립적이고 단서가 적은 짧은 문장에서는 오히려 그 문맥 정보가 의도하지 않은 방향으로 감성 판단을 유도할 수 있다. 이는 모델이 불확실성을 해소하려는 과정에서, 데이터상 빈도적으로 더 자주 등장하고 정서적으로 더 명확한 긍정적 표현을 우선시하도록 학습되어 있기 때문이다. 실제로 He와 Garcia(2009)는 감성 분석을 포함한 분류 작업에서 데이터의 클래스 불균형 (positive > negative > neutral)이 존재할 경우, 신경망 기반 모델이 중립적이거나 부정적인 표현을 과소 인식하고, 모호하거나 불확실한 경우 빈도수가 많은 긍정 클래스로 분류하는 경향을 보인다고 보고하였다. 이와 유사하게, Dixon et al. (2018)은 감성 분류 모델이 감정적으로 불확실하거나 애매한 입력에 대해, 학습 과정에서 더 많이 노출된 감성 클래스(주로 긍정)나 표현 강도가 뚜렷한 감성 범주에 편향되는 경향이 있다고 지적한다. 이는 단지 모델의 구조적 한계가 아니라, 실제 데이터 구성과 표현 빈도에 따라 학습된 감성 분포 자체가 판단의 편향을 구조화하고 있음을 시사한다.

이러한 현상은 나아가 애매한 감정 표현을 해석하는 데 있어 문맥보다는 특정 어휘의 선택 자체가 더 중요한 단서로 작용할 수 있음을 시사한다. 이때 모델은 상대적으로 더 명시적이거나 학습 빈도가 높은 감성 단어의 영향력에 크게 좌우된다. 이는 애매한 감정이 문장 구조보다 어휘 조합에 의해 형성되는 경향이 크다는 점과도 연결되며, 어휘 중심적 접근이 감정 해석의 정확도를 높이는데 더 실질적인 역할을 할 수 있음을 시사한다.

실제로 Swinney (1979)는 인간 언어 처리 과정에서 모호한 단어의 모든 의미가 문맥과 무관하게 일시적으로 활성화되는 현상을 관찰하였으며, 이는 문맥보다 어휘 자체의 의미 활성화가 초기 해석에 있어 핵심적인 역할을 수행함을 보여준다. 또한 Word Sense Disambiguation for Lexicon-Based Sentiment Analysis (2017)는 감성 분석에서 의미 중의성 해결이 중요한 과제임을 지적하며, 문맥 정보보다는 단어 수준에서의 의미 명확화가 결과에 직접적인 영향을 미친다고 강조한 바 있다. 따라서 감정적으로 모호한 문장을 해석할 때, 문맥은 때로는 의미의 불확실성을 증폭시키는 요인으로 작용할 수 있으며, 이에 비해 핵심 어휘의 조합은 감성 추론을 위한 보다 결정적인 실마리가 될 수 있다. 또한, (2)의 ‘창백’ ‘허깨비’ (4)의 ‘불안’ ‘대관절’ (5)의 ‘무서운’ ‘당한’ ‘죄인’ (7)의 ‘기특’ ‘축하’ ‘축연’ ‘베폴다’ ‘축포’ ‘축배’ (8)의 ‘업화’ ‘지옥’ (9)의 ‘독사’ ‘물리다’ ‘죽다’ ‘거짓’ ‘사인’ ‘속이다’ (10)의 ‘독사’ ‘죽이다’ (12)의 ‘실성’ (13) ‘반갑다’ (14) ‘금지옥엽’ ‘귀하다’ ‘애지중지’부터 (15)~(23) 등의 의미편향이 명시적인 어휘들이 분명히 있음에도 RNN기반 결과가 굳이 반대로 도출되는 점은 정확도와 별개로 어휘 외의 요소들이 RNN 기반 감성분석에 상당히 반영되고 있음을 알 수 있다. 문학적 표현이나 구어체를 포함하는 적절한 훈련 데이터가 충분히 확보되지 않은 상황에서는, 감성 분석 과정에서 특정 표현 유형에 대한 패턴 분석이 불가능해지며 이에 따른 오류 발생 가능성이 존재한다고 추정할 수 있다. 그러나 동일한 조건하에서

도 RNN 기반 모델이 상대적으로 안정적인 예측을 보이는 사례와 비교해 볼 때, 이러한 오류 양상을 전적으로 데이터 부족에 기인한 것으로 단정짓기는 어려울 것으로 보인다.

추가로 문장의 길이와 감성사전 / RNN 기반의 정확도 우위의 상관관계는 유의미하게 발견되지 않는다. 다만, 3-2)RNN 기반 정확도 우위에 속하는 부정감성의 예문에 비해서 3-1)감성사전 기반 정확도 우위의 부정감성 예문의 경우 같은 부정성을 함축하고 있음에도 문장 말미의 구조가 긍정+부정을 포함하거나 의문문을 포괄하는 등 의미+구조 조합이 분명하지 않거나 긍정적 어휘가 쓰이게 될 경우: (2)‘허깨비는 아니지요’ (4)‘엄하게 할까요’ (8)‘또 가까워 왔다’ (9)‘속이고 있지만 여봐라’ (10)‘왕관을 쓰고 있다’ (12)‘분명하오이다’ (16)‘내가 계승할 왕위를’ RNN 기반 분석상의 오류가 빈번히 발생하는 경우를 볼 수 있다. 실제로, 이와 같은 분석적 관찰은 기존 연구들에 의해 뒷받침된다. Maheswaranathan & Sussillo (2020)는 RNN이 문맥 정보를 처리하는 방식에 대한 분석을 통해, 해당 모델이 문장의 시작과 끝 구간을 상이하게 처리하며, 특히 문장 말미에 등장하는 긍정적 어휘가 전체 감성 판단에 과도한 영향을 미치는 경향이 있음을 밝혀냈다. 이러한 결과는 RNN이 문장의 전반적인 의미 흐름보다는 마지막에 위치한 단어의 정서적 극성에 지나치게 의존할 수 있음을 시사하며, 감성 판단의 오류 원인을 구조적으로 설명하는 데 기여한다. 유사한 맥락에서, Barnes et al. (2019)은 부정 표현의 범위(scope)를 명확히 식별하지 못할 경우, 감성 분석 모델이 감성 판단에 오류를 범할 수 있음을 지적하였다. 이들은 특히 부정 표현이 문장 후반부에 위치하거나 긍정 어휘와 결합되어 나타날 경우, RNN 기반 모델이 이러한 복합 구조를 효과적으로 해석하지 못하는 문제점을 강조한다. 이는 부정문에 있어 문맥정보가 문장 전체에 걸쳐 정확히 반영되고 있지 않을 수 있음을 시사한다.

표 3. 감성사전 기반 정확도 우위의 예문

sentence	감성사전	RNN	length
(1)바로 저 자리에 왔을 때 마셀리스와 나 둘뿐이었는데 종이 한점을 땀 뺐을 무렵	0	70.45% 확률로 긍정 리뷰.	43
(2)아니 호레이쇼님 떨고 있잖아요 안색도 창백하고 어때세요 허깨비는 아니지요	-1	61.40% 확률로 긍정 리뷰.	41
(3)시간도 꼭 같은 자정에 무장을 갖추고 이 망대 곁을 지나 갔었어요	0	71.49% 확률로 긍정 리뷰.	37
(4)그렇지 않아도 불안해 죽겠는데 대관절 왜 이렇게 밤마다 파수는 세우며 날이면 날마다 경비를 엄하게 할까요	-1	81.08% 확률로 긍정 리뷰.	59
(5)무서운 호출이나 당한 죄인처럼	-2	71.25% 확률로 긍정 리뷰.	17
(6)소신을 프랑스로 돌아가게 하여 주읍소서	0	81.75% 확률로 부정 리뷰.	21
(7)오 기특한 지고 이 기쁨을 축하하기 위하여 축연을 베풀고 축포를 하늘에 터트려라 그리하면 하늘도 나의 축배에	2	62.14% 확률로 부정 리뷰.	63
(8)이글 이글 타오르는 저 지옥의 업화에 내가 내 몸을 태울 시간이 또 가까워 왔다	-1	56.33% 확률로 긍정 리뷰.	43
(9)너의 아버지는 정원에서 낮잠을 자다가 독사에게 물려 죽었다고 거짓 사인을 꾸며 온 백성을 속이고 있지만 여봐라	-2	52.16% 확률로 긍정 리뷰.	60
(10)네 아버를 죽인 그 독사는 지금 머리에 왕관을 쓰고있다	-2	57.26% 확률로 긍정 리뷰.	32

(11)아 벌써 날이 새는구나	0	90.33% 확률로 긍정 리뷰.	13
(12)감히 말씀드리옵니만 실성한 것이 분명하오이다	-1	55.12% 확률로 긍정 리뷰.	26
(13)대감 반가운 소식이 있습니다	1	51.51% 확률로 부정 리뷰.	15
(14)금지옥엽 귀한 딸을 애지중지하였도다	2	53.02% 확률로 부정 리뷰.	20
(15)매우 점잖게 맞아 주시옵니다	1	76.83% 확률로 부정 리뷰.	15
(16)이렇게 된 마당엔 난들 더 참을 수 없지 않는가 선왕을 살해하고 왕비를 간통한 자가 내가 계승할 왕위를	-1	68.10% 확률로 긍정 리뷰.	58
(17)오죽하면 내가 원수를 갚으려 했겠오만 왕자님 말씀 들으니 이제 마음이 풀리오	1	54.74% 확률로 부정 리뷰.	42
(18)그대의 수완이 한결 드높이드러나 칠야의 별같이 빛나게 될 발판이 되어 드리리다	2	59.08% 확률로 부정 리뷰.	44
(19)암만해도 양심에 걸리는데	-1	50.79% 확률로 긍정 리뷰.	14
(20)피를 보고 기절한 것이다	-1	66.98% 확률로 긍정 리뷰.	13
(21)아 햄릿 저 술 술 술에 독이	-1	56.39% 확률로 긍정 리뷰.	18
(22)왕자님 손에 쥐인 그 칼 뺏겨진 그 칼끝에 독이 독이 칠해 있습니다	-3	61.73% 확률로 긍정 리뷰.	39
(23)이 천하에 둘도 없는 살인 강간자 옛다 이 독약마저 마셔라 네놈의 진주라는 것이 이것이나	-2	63.35% 확률로 긍정 리뷰.	50

3.2 RNN 기반 정확도 우위의 예문

감성사전 기반 감성분석이 우세하게 나타난 예문의 긍/부정 비율에 큰 차이가 없었다면, RNN의 경우 다소 편향되어 나타나는 점을 발견할 수 있다. 아래 표를 참고하면 4개의 예문을 제외하고 RNN이 정확하게 예측하고 있는 예문은 모두 부정예문에 해당한다.

표 4. RNN 기반 정확도 우위의 예문

sentence	sent	RNN	length
(1)넌 누구야 정지 이름을 대라	1	84.38% 확률로 부정 리뷰.	16
(2)쥐새끼 한마리 얼씬하지 않았오	0	63.22% 확률로 부정 리뷰.	17
(3)호레이쇼님은 우리더러 헛것을 보았다는거네	1	50.04% 확률로 부정 리뷰.	23
(4)대체 뭣이나 너는 승하하신 상감의 그 늠름하신 무장을차리고 무엄하게도 이 야삼정에 나타나다니 썩 말해라	0	75.06% 확률로 부정 리뷰.	58
(5)가버렸어 대꾸하기 싫은 모양이군	0	89.43% 확률로 부정 리뷰.	17
(6)혹 이게 나라에 무슨 변괴라도 일어날 흥조나 아닐까	1	60.86% 확률로 부정 리뷰.	29
(7)그렇다면 어째서 그렇게 유별나게 보이는가	2	61.99% 확률로 부정 리뷰.	22
(8)여보게 날 친구로 대해 주게 뭣하러 윗텐버그에서 여기까지	1	67.91% 확률로 부정 리뷰.	31
(9)이런일 당할바에야 차라리 난 천당에서 원수를 만나고 싶어	0	64.42% 확률로 부정 리뷰.	32
(10)두 사람은 놀라 사시나무 떨듯 떨기만 하고 아무 말도 건네 보지 못하였다는군요	2	91.27% 확률로 부정 리뷰.	44

(11)네 처지를 분간 못하는 모양이더만 너는 내 딸이며 미혼의 몸이란걸 알아두어야 해	0	74.51% 확률로 부정 리뷰.	44
(12)정다운 말썸이라고 하여 아 이렇게 험란한 세상에 정다운 소리가 있음직하나 그래 그게 진짜 같이 들리었나	6	65.04% 확률로 부정 리뷰.	58
(13)말을 너무 달리면 맥이 풀어지듯 말도 해두면 실속이 없는 법이나라	3	62.24% 확률로 부정 리뷰.	36
(14)이나라 어딘가가 썩어든 정조조	0	69.37% 확률로 부정 리뷰.	16
(15)어디까지 갈 작정이시오 말을 하시라 여기 아무도 없으니 말을 하시라	2	55.49% 확률로 부정 리뷰.	38
(16)그렇게 정숙하게 보이던 왕비를 꼬여 창피 막심한 난음의	0	52.58% 확률로 부정 리뷰.	31
(17)그럴 때 가령 팔짱을 끼거나 고개를 흔들며 우리는 알고 있어 하면서 내 비밀을 아는 채하지 말란말야	0	63.84% 확률로 부정 리뷰.	56
(18)진정하라 바보 같은 두더지	0	71.90% 확률로 부정 리뷰.	15
(19)아버님 큰일 났어요 무서워요	0	56.26% 확률로 부정 리뷰.	16
(20)그리고서야 겨우 손목을 놔주시더니 어깨 넘어로 곧장 제 얼굴을 보시면서 문밖으로 나가셨어요	0	89.31% 확률로 부정 리뷰.	50
(21)저런 그래서 실성하신 거로군 이건 내 잘못이야	1	87.93% 확률로 부정 리뷰.	25
(22)근자에 나는 모든 기쁨을 잃었어	1	50.71% 확률로 부정 리뷰.	18
(23)이러한 인간이건만 대체 나에겐 뭐란 말인가 쓰레기 떼미 쓰레기로밖에는 보이지 않아	0	95.93% 확률로 부정 리뷰.	45
(24)인간의 꼴이 보기 싫단 말이야	1	70.72% 확률로 부정 리뷰.	17
(25)인간의 꼴이 보기 싫으시다면 극단 배우들이 얼마나 박대를 받겠어요	1	59.59% 확률로 부정 리뷰.	37
(26)역사극 전원극 전원희극 역사전원극 비극역사극 희비역사전원극 닥치는 대로 척척해 냅니다	-1	66.22% 확률로 긍정 리뷰.	47
(27)살인의 죄는 입이 없어도 스스로 실토하기 마련이라니까	0	76.40% 확률로 부정 리뷰.	30
(28)그까진것 믿지 않았어야 할 걸	0	76.20% 확률로 부정 리뷰.	16
(29)심신이 쇠잔하여 길지 못할 이내몸 아차 이 몸 죽거들랑 바라건대 우리왕비 훌륭한 배필맞아 백년해로 다시하오	-2	63.75% 확률로 긍정 리뷰.	59
(30)이래서 세상만사 여의치 않지	0	78.01% 확률로 부정 리뷰.	16
(31)정말 싫으실테지 자 피리를 불어줍세	0	75.63% 확률로 부정 리뷰.	20
(32)내전에 들킨 후 몹시 심기가 좋지 못하신 듯	1	70.25% 확률로 부정 리뷰.	25
(33)그렇게 놀라신 다음에는 그 뒤는 어떻게 되시었지	1	57.53% 확률로 부정 리뷰.	27
(34)지금의 어머님의 얼굴결의 어머니가 되더라도 복종하겠노라 여쭙어라	0	61.04% 확률로 긍정 리뷰.	35
(35)제발 다부지게 힐책하십시오	0	78.71% 확률로 부정 리뷰.	14
(36)햄릿 너는 아버님께 왜 그렇게 불손하냐	1	79.64% 확률로 부정 리뷰.	22
(37)어머니는 나의 아버님께 왜 그렇게 불손하셨어요	1	62.31% 확률로 부정 리뷰.	26
(38)그런 쓸데없는 입은 놀리지 않는 법이야	0	75.49% 확률로 부정 리뷰.	21
(39)넌 제 어미도 몰라보냐	0	82.96% 확률로 부정 리뷰.	13
(40)아니 네가 정 그런 말대답을 한다면 너를 따져 물을 수 있는 사람을 부를테다	1	56.33% 확률로 부정 리뷰.	42
(41)그 마음속을 거울에 환히 비춰 보일테니 그때까진 못 나가십니다	1	59.33% 확률로 부정 리뷰.	35
(42)이 무슨 잔인한 짓이냐	0	74.35% 확률로 부정 리뷰.	13
(43)아버님 그렇게 소자를 보지 마십시오	1	73.41% 확률로 부정 리뷰.	20

(44)있지도 않은 것을 있는 듯이 만들어내는 것이 역시 바로 실성한 증거야	2	58.28% 확률로 부정 리뷰.	39
(45)그러나 정신 차리지 않으시면 더 나쁜게 뒤를 따를 것입니다	0	78.23% 확률로 부정 리뷰.	33
(46)누군지 필시 자살한 사람인가 보군	1	79.62% 확률로 부정 리뷰.	18
(47)교회의 예법으로서는 최선을 다하였오	0	58.87% 확률로 긍정 리뷰.	19
(48)칙명이 관례를 굽혔기 망정이지 그렇지 않았던들 교회 묘비 사용도 어림 없오	1	84.32% 확률로 부정 리뷰.	42
(49)이 이상은 아무래도 안된다는 말씀이오	1	65.96% 확률로 부정 리뷰.	21
(50)그 위조하신 친서의 내용은	0	54.28% 확률로 부정 리뷰.	15
(51)마침 내 주머니에 간직하고 있던 선왕의 옥쇄로 찍었지	1	56.71% 확률로 긍정 리뷰.	29
(52)그게 바로 간신의 말로라는 거야	0	75.50% 확률로 부정 리뷰.	18
(53)레이티즈에겐 이만저만 미안한 것이 아니야	0	54.09% 확률로 부정 리뷰.	23
(54)내가 그의 부친을 죽이다니 그리고 오펜리어의 장지에서 다투었으니 내가 너무 흥분해서 지나친 실수를 했지	0	83.72% 확률로 부정 리뷰.	57
(55)그 신사 이야기는 대관절 무슨 까닭에	1	70.23% 확률로 부정 리뷰.	21
(56)이번 승부 불리하실 것 같소이다	0	53.91% 확률로 부정 리뷰.	17
(57)레이티즈가 프랑스에 가 있는 동안 나도 줄곧 연습은 해왔으니까 허나 어쩐지 가슴이 설레어 이상야릇한 기분이로군	-1	70.39% 확률로 긍정 리뷰.	61
(58)아니야 부녀자나 마음에 걸 그런 하찮은 불안이야	0	63.54% 확률로 부정 리뷰.	26
(59)햄릿 이 진주는 햄릿의 것이다	0	58.81% 확률로 긍정 리뷰.	17
(60)승부를 먼저 끝내겠소이다	-1	65.03% 확률로 긍정 리뷰.	13

즉, 부정 예문을 부정으로 정확하게 판단하는데 있어 RNN기반 감성분석이 더욱 특화된 성능을 구현하는 것으로 보이며 감성사전 기반과 다르게 구조와 문맥정보를 통합적으로 활용하는 RNN의 분석 요인 특성을 고려하면 부정감성 예문의 경우 단순 어휘 조합보다 구조와 문맥정보 조합을 통해 의미를 구성하고 있는 것을 역으로 추론할 수 있다. 또한, 해당 분석에서 80% 이상 부정 감성으로 분류된 예문들이 다수 확인되었다는 점은, 단순히 추상적인 경향이 아닌, 상당히 섬세한 정서 인식 결과로 해석할 수 있다. 특히 극단적인 부정 감정을 담고 있는 텍스트의 경우, 감성 사전 기반 분석에서는 오히려 0(중립)으로 판단되는 사례가 반복적으로 나타났다. 3-1)의 경우로 감성사전이 정확히 분석했던 극단 부정의 예문(ex:(22), (23))과 3-2)의 RNN우위 극단 부정 예문의 비교해 볼 수 있다. (1-22) (1-23)의 경우 ‘독’, ‘칼’, ‘독약’등의 명시적인 어휘가 부정적 의미를 중심에서 형성하고 있다면 (2-10) ‘두 사람은 놀라 사시나무 떨듯 떨기만 하고 아무 말도 건네 보지 못하였다는군요 / 그리고서야 겨우 손목을 놔주시더니 어깨 너머로 곧장 제 얼굴을 보시면서 문밖으로 나가셨어요’에서와 같이 어휘가 아닌 문맥 정보를 주로 부정적 의미가 형성되고 있는 것을 확인할 수 있다. 이를 통해, 인간 언어에서 의미는 때로 어휘 중심, 또는 문맥과 구조 중심이라는 서로 다른 방식에 치우쳐 구성되고 전달될 수 있으며, 극단적인 부정 정서를 담은 문장일지라도 반드시 강한 부정 어휘가 사용되어야 하는 것은 아님을 시사한다. 3-1)의 예문과 3-2)의 예문들의 평균길이에 있어 전자는 30.19 후자는 28.67로 유의미한 차이가 발견되지는 않았으며 이는 구조와 문맥 힌트가 반드시 긴 문장을 통해 발생하고 있지는 않음을 제시한다. 그러나 특히 짧고 의미가 중립적인 예문에서 감성 사전 기반의 어휘 위주 분석이 상대적으로 우위를 보이고, 길고 중립적인 문장에서는 RNN 기반 분석이 더 높은 정확도를 보이는 양상이 뚜렷하게 관찰되었다는 점은 중요하다. 이는, 애매한 의미를 담은 단문에서는 어휘의 정서적 극성이 상대적으로 강한 영향을 미치는 반면, 장문의 경우에는 의미의 흐름, 전환 구조, 구문적 기대 등

문맥 기반 정보가 해석의 핵심 단서로 작용한다는 점을 시사한다. 이러한 경향은 전산언어처리 뿐 아닌 인지심리학 분야에 서도 유사하게 제기된 바 있으며, 대표적으로 garden-path sentence 연구들은 문장의 구조적 복잡 성으로 인해 독자가 초기 해석 단계에서 혼동을 겪는 현상을 설명 하며, 문맥과 구문 구조의 통합적 처리가 특히 긴 문장에서 더 중요한 역할을 한다는 점을 보여 준다.

4. 결론

감성사전과 RNN 딥러닝기법을 포괄하는 머신러닝 기반 접근은 인간의 주관성을 배제한 기계적 객관적 평가를 가능케 한다.⁷ 감성사전 기반과 RNN을 이용한 감성분석 모두 어느 정도 타당한 결과를 도출하고 있다. 그러나 그 값의 일관성에 있어서는 더 의미(감성)함축이 높지만 값이 낮게 산출되거나 두 가지 분석 톨의 결과가 상반되게 나오는 등 신뢰도가 떨어지는 경우가 빈번히 발견된다. 뚜렷한 긍/부정 키워드가 있음에도 오류를 범한 감성사전 분석의 경우는 우선 제외하고 RNN의 경우 특정 문장 구조와 그 표현이 가지는 복합성에 대한 학습이 아직 미비할 경우(문장이 단순함에도 이중부정, 부정적 대상에 대한 부정표현 등) 단서 키워드가 있음에도 오류를 범하는 것을 보면 역으로 RNN이 어휘 외에 다른 요소에 상당히 의존하고 있지 않음을 증명할 수 있는 요소가 되기도 하다. 물론 구문이 단순하고 명시적 키워드를 포함한 경우 RNN의 산출 정확도 수치도 높아지는 부분은 어휘의 중추적 역할을 강조하기도 한다. 짧고 중립적인 문장에서는 특히 어휘 함축의 단순한 분석(감성사전기반)이 정확한 결과 값을 도출하며 반대로 길고 중립적인(어렵고 난해한 문장)의 경우 RNN 기반 분석이 정확한 분석을 제공한다. RNN은 특히 부정문을 부정으로 판별하는데 있어 정확도와 신뢰도가 높으며 이는 부정문의 경우 긍정문보다 구조 문맥적 힌트를 제시하는 경향이 더 큰 점을 시사한다. 또한, 이 경우 문장 말미가 의미와 구조 측면에서 분명 해지지 않는 경우 명시적 어휘가 있음에도 부정감성 문장으로 분별하는데 있어 RNN이 실패하는 경우는, 문장 범주 내에서도 특정 부분이 구조적으로 언어 이해에 더욱 중요한 작용을 하고 있을 수 있음을 함께 보여준다. 추후 두 가지 분석방법에 있어 모두 정확도가 높거나 혹은 모두 분석에 실패한 경우의 예문들을 함께 분석 정리함으로써 인간의 언어 체계에서 인식하는 복잡한 구문과 전산처리상의 복잡성에 있어서 어떠한 차이가 있는지 관찰할 수 있다. 또한, 전산적 결과를 통해 인간의 언어 이해 메커니즘⁸을 역으로 추론해 볼 수 있다. 이후 방대한 감성사전 구축과 복잡하고 난해한 구문을 포괄할 수 있을 훈련데이터 구축과정이 동반되면 관련하여 보다 정교한 주제를 설정하여 분석관찰을 진행할 수 있다.

¹ 김재남 역본은 전통적 문체와 현대 한국어 구문이 혼합된 특징을 지니며, 문장 구조적 복잡성과 감성 표현이 균형 있게 나타나는 판본으로, 감성 분석 및 구조 비교 분석 모두에 적합하다고 판단되어 본 연구에 사용되었다.

² KnuSentiLex 는 군산대학교 소프트웨어융합공학과 Data Intelligence Lab 에서 구축한 한국어 감성 사전으로, 단어에 긍정, 부정, 중립의 감성 극성과 감성 강도 값을 부여하여 감성 분석에 활용할 수 있도록 만들어졌다. 이 사전은 국립국어원 표준국어대사전의 뜻풀이 분석, 김은영(2004)의 긍/부정어 목록, SentiWordNet 및 SenticNet-5.0 에서 번역된 긍/부정어, 그리고 최근 온라인에서 사용되는 축약어 및 이모티콘 목록 등을 기반으로 구축되었다. 감성사전은 다음의 GitHub 페이지를 통해 열람할 수 있다: <https://github.com/park1200656/KnuSentiLex>

³ 전통적인 감성사전 기반 분석(lexicon-based sentiment analysis)은 감성 극성이 사전에 정의된 단어 리스트(감성사전, sentiment lexicon)를 기반으로, 문서나 문장에서 감성적 경향을 추론하는 방식이다. 이

접근법은 주로 각 단어에 부여된 감성 점수를 합산하거나 평균하여 전체 문장의 감성 상태를 추정한다 (Taboada et al., 2011). 문 맥에 독립적인 단어 수준의 판단에 기반하기 때문에 간단하고 빠르며 지도 학습이 필요 없다는 장점이 있으나, 부정어 처리, 다의어, 문맥 의존성 등의 한계가 있다.

⁴ 순환 신경망(Recurrent Neural Networks, RNN)은 텍스트를 단어의 시퀀스로 변환하며, 단어 간의 의존성과 텍스트의 구조적 관계를 파악하도록 설계되어 있다. RNN의 가장 널리 사용되는 구조는 LSTM(Long Short-Term Memory)이며, 이는 트리 구조, 단어 간 관계, 문서의 주제와 같은 요소들을 반영하는 데 적합하다. RNN은 시간축에서의 패턴을 포착하는 반면, 합성곱 신경망(CNN)은 공간 축에서의 패턴을 포착한다. “Recurrent neural networks (RNN) convert text into a sequence of words. It is designed to get the dependency between the word and the structure of the text. The most popular architecture of RNN is long short-term memory (LSTM) in tree structure, word relation and document topic. RNNs capture the pattern in the time dimension, while convolutional neural networks (CNN) capture the pattern in the space dimension.” (Low, KeiKhosroKiani, Asl, 2024)

⁵ RNN은 뛰어난 성능을 얻기 위해 장단기 기억(Long Short Term Memory, LSTM)과 함께 사용되어야 한다. LSTM을 네트워크에 추가하는 것은 네트워크 내부에 입력의 시작 문맥을 기억할 수 있는 메모리 유닛을 추가하는 것과 같다. “RNNs need to be used with a Long Short Term Memory (LSTM) to achieve great results. Adding the LSTM to the network is like adding a memory unit inside the network that can remember context from the very beginning of the input.” (Tarwani, Kanchan M., et al, 2017) 순환신경망(Recurrent Neural Networks, RNNs)은 시퀀스 데이터를 처리하는 데 적합하지만, 기본 구조만으로는 긴 시퀀스 내 장기 의존성(long-term dependency)을 효과적으로 학습하기 어렵다는 한계가 있다. 이는 주로 기울기 소실 (vanishing gradient) 문제 때문인데, 시간이 길어질수록 초반 입력의 정보가 후반 출력에 제대로 반영되지 못하게 된다. 이를 해결하기 위해 제안된 LSTM(Long Short-Term Memory)은 입력 게이트, 출력 게이트, 망각 게이트(forget gate) 등을 통해 중요 정보를 장기간 유지하거나 불필요한 정보를 제거할 수 있는 구조를 갖는다. 결과적으로, LSTM은 입력 시점에서 등장한 의미적 또는 문맥적 요소를 이후 시점까지 기억할 수 있게 하여, 문장 전체의 흐름을 반영한 보다 정교한 예측을 가능하게 한다.

⁶ 실제 absolute syntactic complexity의 평가 요소는 다섯 차원에서 분석된다; length of production unit, amount of subordination, amount of coordination, degree of phrasal sophistication, and overall sentence complexity. (Lu, 2010, Ortega, 2003) Length of production unit(Mean length of sentence, Mean length of clause, Mean length of T-unit)는 문장의 길이가 syntactic[text]complexity의 중요기준임을 반영한다.

⁷ 기계 학습 접근법을 활용하면 수작업 분석으로는 즉시 드러나지 않는 대규모 데이터셋 내의 패턴을 식별할 수 있다. 또한 이러한 접근은 분석 과정에서 인간의 주관성을 배제함으로써 편향을 줄이는 데에도 도움이 된다. “The use of machine learning approaches can help to identify patterns within large datasets that may not be immediately apparent through manual analysis. This approach can also help reduce bias by removing human subjectivity from the process of analysis”. (Low, KeiKhosroKiani, Asl, 2024)

⁸ 인간의 언어 이해 메커니즘에도 패턴과 규칙이 존재한다. “Language cognition research initially revealed some patterns of human language understanding, much more is needed to truly analyze the mechanism of language understanding in the human brain.” (Wang, Ding, et al., 2023) 그런데 인간의 뇌에서 인지하는 총체적, 직관적 패턴과 그것에 바탕하여 혹은 모사하여 만들어진 전산적 패턴 사이의 일관성 및 일치성이 또한 주요한 질문이 될 것이다. 이러한 맥락에서 최근 연구들은 대규모 언어모델이 인간의 뇌 반응과 상당히 높은 정렬도를 보인다는 점을 입증하고 있다(Schrimpf et al., 2021; Toneva & Wehbe, 2019). 전산 모델이 인간의 문법 처리 오류나 예측 불확실성과 유사한 반응을 보인다는 연구(Wilcox et al., 2018)도 이러한 논의를 보완한다.

참고문헌

Alharbi, Najla M., et al (2021). “Evaluation of Sentiment Analysis via Word Embedding and RNN Variants for Amazon Online Reviews.” *Mathematical Problems in Engineering*. 1–10.

Barnes, Jeremy, Erik Velldal, and Lilja Øvrelid (2021). “Improving Sentiment Analysis with Multi-Task Learning of Negation.” *Natural Language Engineering*. 249–269. <https://doi.org/10.1017/S1351324920000510>

-
- Christianson, Kiel, et al (2001). "Thematic Roles Assigned along the Garden Path Linger." *Cognitive Psychology*. 368–407. <https://doi.org/10.1006/cogp.2001.0752>
- Culpeper, Findlay, et al. (2018). "Measuring Emotional Temperatures in Shakespeare's Drama." *English Text Construction*. 10–37.
- Dixon, Lucas, et al (2018). "Measuring and Mitigating Unintended Bias in Text Classification." *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. <https://doi.org/10.1145/3278721.3278729>
- Frazier, Lyn; Rayner, Keith (1982). "Making and Correcting Errors during Sentence Comprehension: Eye Movements in the Analysis of Structurally Ambiguous Sentences." *Cognitive Psychology*. 178–210. [https://doi.org/10.1016/0010-0285\(82\)90008-1](https://doi.org/10.1016/0010-0285(82)90008-1)
- Goldberg, Adele E (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.
- Hong, Yurim (2020). "Sentiment Analysis of Shakespeare's Othello." <영어영문학>. 163–185.
- He, Haibo; Garcia, Eduardo A (2009). "Learning from Imbalanced Data." *IEEE Transactions on Knowledge and Data Engineering*. 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>.
- Jang, See-eun, et al (2019). "Sentiment Analysis of Shakespeare's Tragedy Plays and Their Major Male and Female Characters through Dialogue Corpora." <언어과학>. 115–147.
- Kauf, Carina, et al (2024). "Lexical-Semantic Content, Not Syntactic Structure, Is the Main Contributor to ANN-Brain Similarity of fMRI Responses in the Language Network." *Neurobiology of Language*. 7–42. https://doi.org/10.1162/nol_a_00116.
- Low, Keikhosrokiani, et al (2024). "Decoding Violence Against Women: Analysing Harassment in Middle Eastern Literature with Machine Learning and Sentiment Analysis." *Humanities and Social Sciences Communications*. 1–18.
- Lu, X (2010). "Automatic Analysis of Syntactic Complexity in Second Language Writing." *International Journal of Corpus Linguistics*. 474–496.
- Maheswaranathan, Niru, and David Sussillo (2020). "Reverse Engineering Recurrent Networks for Sentiment Classification Reveals Line Attractor Dynamics." *arXiv preprint arXiv:2004.08013*. <https://doi.org/10.48550/arXiv.1906.10720>.
- Ortega, Lourdes (2003). "Syntactic Complexity Measures and Their Relationship to L2 Proficiency: A Research Synthesis of College-Level L2 Writing." *Applied Linguistics*. 492–518.
- Pamungkas, Endang Wahyu; Putri, Divi Galih Prasety. "Word Sense Disambiguation for Lexicon-Based Sentiment Analysis." *Proceedings of the 9th International Conference on Machine Learning and Computing (ICMLC '17)*. 442–446. <https://doi.org/10.1145/3055635.3056578>.

-
- Pearce, Brian; Duffy, Kevin(2010). "Hamlet: Rational and Emotional Units of Meaning in Four Soliloquies." *Shakespeare in Southern Africa*. 21–28.
- Rengarjan, Manikandan, et al (2023). "Revolutionizing Sentiment Analysis in Literary Text 'The Immorals of Meluha' through a Hybrid CNN-RNN Architecture and Advanced Feature Techniques." *Journal of Theoretical and Applied Information Technology*. 7866–7879.
- Seo, Hyejin; Lee, Jonghyun; Shin, Jeonga (2019). "Sentiment Analysis of Shakespeare's Plays Using a Deep Learning Technique." <한국영어학회 영어학>. 817–836.
- Swinney, David A (1979). "Lexical Access During Sentence Comprehension: (Re)consideration of Context Effects." *Journal of Verbal Learning and Verbal Behavior*. 645–659. [https://doi.org/10.1016/S0022-5371\(79\)90355-4](https://doi.org/10.1016/S0022-5371(79)90355-4).
- Schrimpf, Martin, et al (2021). "The Neural Architecture of Language: Integrative Modeling Converges on Predictive Processing." *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.2105646118>.
- Steenbergh, Kristine (2011). "Emotion, Performance and Gender in Shakespeare's Hamlet." *Sexed Sentiments: Interdisciplinary Perspectives on Gender and Emotion*, edited by Ruberg, Willemijn; Steenbergh, Kristine. Rodopi. 93–116.
- Taboada, Maite, et al (2011). "Lexicon-Based Methods for Sentiment Analysis." *Computational Linguistics*. 267–307.
- Tarwani, Kanchan M.; Edem, Swathi (2017). "Survey on Recurrent Neural Network in Natural Language Processing." *International Journal of Engineering Trends and Technology*. 301 –304.
- Toneva, Mariya; Wehbe, Leila. "Interpreting and Improving Natural-Language Processing (in Machines) with Natural Language-Processing (in the Brain)." *Advances in Neural Information Processing Systems*. 14954–14964.
- Wang, Shaonan, et al (2023). "Language Cognition and Language Computation—Human and Machine Language Understanding." *arXiv preprint arXiv:2301.04788*. 1–31.
- Wilcox, Ethan, et al (2018). "What Do RNN Language Models Learn about Filler–Gap Dependencies?" *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. 211–221. <https://aclanthology.org/W18-5423>.