

Sentiment Dictionary Construction Based on Reason-Sentiment Pattern Using Korean Syntax Analysis

Woo Hyun Kim* · Heejung Lee**†

*Department of Industrial Data Engineering, Hanyang University

**School of Interdisciplinary Industrial Studies, Hanyang University

한국어 구문분석을 활용한 이유-감성 패턴 기반의 감성사전 구축

김우현* · 이희정**†

*한양대학교 산업데이터엔지니어링학과

**한양대학교 산업융합학부

Sentiment analysis is a method used to comprehend feelings, opinions, and attitudes in text, and it is essential for evaluating consumer feedback and social media posts. However, creating sentiment dictionaries, which are necessary for this analysis, is complex and time-consuming because people express their emotions differently depending on the context and domain. In this study, we propose a new method for simplifying this procedure. We utilize syntax analysis of the Korean language to identify and extract sentiment words based on the Reason-Sentiment Pattern, which distinguishes between words expressing feelings and words explaining why those feelings are expressed, making it applicable in various contexts and domains. We also define sentiment words as those with clear polarity, even when used independently and exclude words whose polarity varies with context and domain. This approach enables the extraction of explicit sentiment expressions, enhancing the accuracy of sentiment analysis at the attribute level. Our methodology, validated using Korean cosmetics review datasets from Korean online shopping malls, demonstrates how a sentiment dictionary focused solely on clear polarity words can provide valuable insights for product planners. Understanding the polarity and reasons behind specific attributes enables improvement of product weaknesses and emphasis on strengths. This approach not only reduces dependency on extensive sentiment dictionaries but also offers high accuracy and applicability across various domains.

Keywords : Aspect-Level Sentiment Analysis, Sentiment Dictionary, Opining Mining

1. 서론

텍스트 마이닝은 디지털화된 텍스트 데이터에서 유용한 정보나 지식을 추출하는 과정이다[42]. 텍스트 마이닝의 중요한 연구 분야 중 하나인 감성분석(sentiment analy-

sis)은 온라인 채널의 상품 리뷰 등에서 사용자의 경험을 나타내는 의견이나 감성표현을 추출하는 것을 의미한다[26]. 감성분석은 문서 수준, 문장 수준, 속성 수준으로 분류될 수 있으며, 각각은 문서, 문장, 특정 속성에 대한 감정 극성을 분석한다[1, 5 7, 11]. 특히 속성 수준의 감성분석은 온라인 채널 상품 리뷰에서 속성에 따른 감성 표현을 추출하는 연구들이 활발한데[29, 47], 이 분야의 연구는 감성사전에 크게 의존하기 때문에 감성사전의 수준이 분석 결과에 큰 영향을 미친다[41]. 감성 사전을 구축하는 것은 많은

Received 1 December 2023; Finally Revised 14 December 2023;
Accepted 15 December 2023

† Corresponding Author : stdream@hanyang.ac.kr

시간과 비용이 들며, 각 도메인별로 다른 감성 표현을 사용하기 때문에 새로운 감성 사전을 매번 구축해야 한다. 따라서, 기존의 감성사전을 재활용하거나, 도메인별로 수정하여 구축하는 방식이 주로 사용되었다. 하지만 이는 도메인이 다를 경우 분석 성능이 저하되는 한계가 있고, 도메인별 감성 사전을 수정하는 작업도 역시 적지 않은 시간과 비용이 소요된다.

속성 수준 감성 분석에서 속성과 관련된 감성 표현을 추출하는 것은 매우 중요하지만, 현재 국내에서 구축된 한국어 감성 사전은 도메인에 따라 감성 표현이 달라질 수 있어 완벽하지 않다. 이를 해결하기 위해 본 논문에서는 단독으로 사용될 때도 극성을 파악할 수 있는 단어를 감성 단어로 정의하고, 도메인과 사용자 특성에 따라 극성이 달라지는 단어는 감성 단어에서 제외하는 방법을 제안한다. 예를 들어, “좋다”는 단독으로 사용해도 긍정적인 극성이 명확하지만, “작다”는 상황에 따라 긍정적일 수도 부정적일 수도 있어 명확한 극성을 가진 감성 단어로 분류하기 어렵다. 이처럼 감성 표현이 내재되어 있거나, 감성 단어를 보조하는 단어들은 ‘이유(reason)’로 분류하고, 명확한 극성을 가진 단어만을 감성 단어(sentiment)로 분류한다.

명확한 극성을 가진 감성 단어로만 구성된 감성 사전을 구축하면, 속성 수준의 감성분석 결과는 상품 기획자에게 유용한 정보를 제공할 수 있다. 속성이 갖는 극성과 그 이유를 파악하면, 상품을 기획할 때 단점이 되는 속성은 개선하고, 장점이 되는 속성을 강조할 수 있기 때문이다. 따라서 본 논문에서는 온라인 쇼핑몰 리뷰에서 단어 사이의 관계를 바탕으로 감성 단어의 출현 패턴을 찾고, 이를 통해 명확한 극성을 가진 감성 단어를 추출하여 정확한 감성 분석을 할 수 있는 감성 사전을 구축하는 방법론을 제안한다. 제안된 방법론은 감성 사전의 의존도를 줄이면서도 높은 정확도의 감성 분석을 가능하게 하며, 다양한 도메인에서 적용할 수 있는 범용성을 갖추고 있다. 또한, 본 제안된 방법론으로 구축된 사전은 순수 감성 단어에 초점을 맞춘 독립적인 감성 사전으로, 동일 도메인 내 재사용 및 업데이트가 용이하다. 대부분의 기존 감성 사전은 합성어와 파생어 등이 섞여 있어 방대한 양을 이루며, 새로운 도메인에 즉시 적용하기가 어렵다. 이로 인해 감성 단어를 새로 수집하고 재구축한데 많은 비용과 노력이 든다. 그러나 본 논문에서 제안하는 방법론은 각 도메인의 특성을 반영해 어근화된 감성 단어를 추출함으로써, 동일 도메인 내에서 사용되는 다양한 합성어와 파생어를 모두 추출할 필요가 없고, 오직 극성이 명확한 단어만을 추출하기 때문에 업데이트와 재구축이 간편하다는 장점이 있다. 이러한 접근 방식은 기업에서 상품 리뷰, 소셜 미디어, 기타 온라인 텍스트 자료로부터 비자의 요구와 기대를 더 잘 이해하고, 그

에 따라 상품을 개선하거나 마케팅 전략을 조정하는 데 큰 도움이 될 것이다.

2. 이론적 배경 및 관련 연구

2.1 감성사전 구축

감성 사전을 구축하기 위한 전체 절차는 “데이터 수집, 데이터 전처리, 데이터 가공, 사전구축, 데이터 후처리” 순으로 이루어진다. 그 중 사전구축에 대한 연구는 수작업 기반, 통계 기반, 기계 학습 기반 및 하이브리드의 네 가지 접근 방법이 있다.

수작업 기반 접근 방식은 주로 전처리와 가공을 거쳐 얻어진 단어 후보를 직접 수정 및 제거하는 방법론이다. Kwon et al.[21]과 Song et al.[39]은 각각 온라인 쇼핑몰 리뷰와 영화 리뷰 데이터를 품사 태깅하고, 특정 품사를 갖는 단어만 추출하여 수작업으로 감성 사전을 구축하는 연구를 하였다. Shin[37]과 Steinberger et al.[41]은 기존에 존재하던 감성 사전 및 범용 사전[4, 15]을 바탕으로 고빈도 단어를 추출하거나 번역 작업을 거쳐 수작업으로 감성 사전을 구축하는 연구를 진행하였다. Ahn and Kim[2]은 국립국어원 사전을 대상으로 전문가들의 투표를 통해 감성 단어를 추출하고, 해당 단어를 바탕으로 감성 사전 구축 연구를 진행했다. 이러한 연구들은 다른 방법론에 비해 정확도는 높지만, 연구자들의 일관성과 명확한 기준이 요구된다는 점과 상당한 돈과 시간이 소요된다는 특징이 있다. 통계 기반 접근 방식은 기존에 구축된 감성 사전의 감성 단어들의 빈도수를 바탕으로 단어를 추출하는 방법론이다. Kim et al.[18]은 연세대학교에서 독자적으로 구축한 사전인 오픈한글을 바탕으로 등장 빈도수가 높은 단어를 추출하여 감성 사전 구축 연구를 진행했다. 통계 기반 접근법은 보다 적은 시간과 비용이 들지만, 기존에 잘 구축된 감성 사전이 존재해야 한다는 특징이 있다. 기계 학습 기반 접근 방식은 주로 품사태깅과 임베딩을 통해 전처리된 데이터를 다양한 방법론을 통해 전파하거나 분류 및 군집 작업을 거쳐 감성 사전을 구축하는 방법론이다. Tsai et al.[44]은 기존 사전인 ANEW SenticNet[8, 9, 36]을 시드단어로 사용하여 ConceptNet[12, 25]의 노드에 극성을 부여하고 랜덤 워크 기법을 통하여 극성을 전파하는 방법론을 통해 감성 사전 구축 연구를 진행했다. Lee[22]는 Plutchik[35]의 선행 연구로부터 얻어진 시드 단어를 한국어로 번역 및 재해석하여 새로운 시드 단어를 구축하였고, SNS 해시태그를 수집하여 TF-IDF(Term Frequency - Inverse Document Frequency)로 가중치를 추출 및 NMF(Nonnegative Matrix Factorization) 알고리즘을 통해 차원 축소 기법을 사용하여 감성 사전을 구축했다. Heo

and On[13]과 Kim et al.[19]은 영화 리뷰 데이터를 바탕으로 단어와 구문을 분류하여 추출하고 특히, 단어는 명사, 형용사, 동사만을 추출하였다. Heo and On[13]은 word2vec을 바탕으로 추출한 단어를 LPA(label propagation algorithm)를 통하여 극성을 라벨링하였고, Kim et al.[19]은 SVM(support vector machine)을 바탕으로 분류한 단어를 LPA를 통하여 극성을 라벨링 하였다. Alshari et al.[3], Park et al.[34]은 각각 영화 리뷰 데이터와 표준국어대사전을 입력으로 품사 태깅하여 감성 사전을 구축하였다. Alshari et al.[3]은 기존 감성사전인 SentiWordNet을 학습한 Senti2Vec을 통하여 영화 리뷰 데이터 내 감성 단어를 추출하고, SVM을 통하여 극성을 분류하여 감성 사전을 구축하는 연구를 진행하였다. Park et al.[34]은 표준국어대사전에서 수집된 모든 단어를 대상으로 품사 태깅하여 형용사, 부사, 동사, 명사의 뜻풀이를 추출하였다. 그 후, 양방향 LSTM(bi-long term short term)을 사용하여 뜻풀이의 극성을 판단하여 극성이 존재하는 뜻풀이에서 수작업으로 감성 단어를 추출하여 감성 사전을 구축하였다. Ahmed et al.[3]과 Li et al.[23]은 각각 감성 분석 데이터셋(SemEval ABSA dataset)과 기존 감성 사전[6, 16, 40, 45]을 임베딩 하여 감성 사전을 구축했다. Ahmed et al.[3]은 레스토랑, 랩탑, 핸드폰, 카메라의 리뷰 데이터를 모아 놓은 감성 분석 데이터셋을 K-clustering 기법을 통해 군집하여 감성 단어를 추출하고, 해당 감성 단어로 감성 사전을 구축했다. Li et al.[23]은 기존 감성 사전과 금융도메인 말뭉치(corpus)의 중복 어휘를 추출하여 감성 단어 후보 집합을 만든 후, word2vec 기법을 통해 금융 감성 시드 단어와 유사도를 계산하여 임계점을 넘는 감성 단어 후보만을 따로 추출한다. 그 후 추출된 후보들을 DNN(deep neural net)과 양방향 LSTM을 통해 극성을 분류하여 감성 사전을 구축했다. Cho et al.[10]은 여행과 관련된 SNS 말뭉치와 리뷰를 품사 태깅 하여 형용사, 부사, 동사, 명사, 이모티콘을 추출하고 임베딩 하였다. 임베딩된 특정 품사들을 SVM을 통하여 감성 단어로 분류하고 전문가들의 집단 지성에 의해 투표에 통과한 단어들로 감성 사전을 구축했다. 기계 학습 기반 연구들은 대부분 높은 정확도를 보여주지만, 시드 단어 혹은 기존 감성 사전이 필요하다. 또한, 높은 접근성을 보이기에 쉽게 구현하기 어렵다는 특징이 있다. 마지막으로 하이브리드 접근 방식은 전처리로 번역과 품사 태깅한 후 기계학습과 수작업을 앙상블(ensemble)하여 감성 사전을 구축하는 방법론이다. Sivasankar et al.[38]은 아마존 리뷰를 타밀(Tamil)어로 번역한 후 품사 태깅을 하여 형용사, 부사, 동사, 명사에 해당하는 단어들을 따로 추출한다. 추출된 단어들은 기존에 존재하는 시드 단어와 PMI(pointwise mutual information)를 통해 높은 값을 가진 단어만을 초기 감성 사전으로 구축한다. 구축된 초기 감성 사전 내 단어는 번역된 아마존 리뷰 내 등장 빈도로 랭크 점수를 도출하여 임계점을

넘는 단어만을 최종 감성 사전으로 구축하는 연구를 진행하였다.

지금까지 수작업 기반, 통계 기반, 기계 학습 기반, 하이브리드 접근 방식을 활용한 다양한 방법론을 통하여 감성 사전 구축 연구를 살펴보았다. 본 연구에서는 구문 분석과 룰베이스(rule based) 방법론을 통하여 감성 사전을 구축한다. 구문 분석은 단어와 단어 사이의 관계를 분석하는 의존 구문 분석을 사용하였고, 룰베이스 방법론은 오피니언 스키마 기반의 패턴 매칭(pattern matching)을 활용하였다. 이를 통해, 누구나 쉽게 구축할 수 있는 높은 접근성을 지니고, 도메인 별 특징에 따라 서로 다른 패턴을 적용하기에 객관적이고 준수한 정확도를 보이는 장점이 있다. 하지만, 감성 단어의 정의가 다른 방법론 및 사전과 다르기에 비교가 불가능하다는 한계가 존재한다. 다음 2.2장과 2.3장에서 의존 구문 분석과 오피니언 스키마에 대해 간략하게 소개한다.

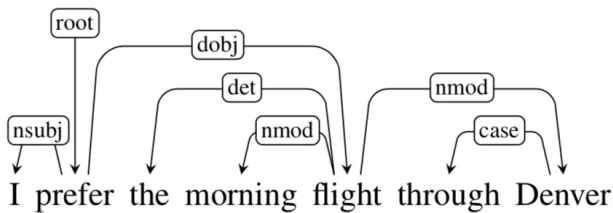
2.2 의존 구문 분석

비정형 텍스트로 이루어진 자연어에서 필요한 정보를 추출하기 위해선 자연어의 형태를 파악하고, 구성 요소 간 관계를 파악해야 한다. 형태를 파악하는 것은 문장을 어절로 분할하고, 각 어절의 품사를 태깅하는 것을 의미한다. 어절에 형태소를 태깅함으로써 문장의 형태를 파악하고 문장을 구성하는 요소들을 알 수 있다. 그 후 구성 요소 간의 관계 파악은 분할된 어절 간의 의존 관계를 나타내고 태깅하는 것을 의미한다. 각 어절 간의 의존 관계를 통해 문장 내 중요한 어절을 판단하고 추출할 수 있기 때문이다. 이와 같은 작업을 하는 것을 의존 구문 분석이라 한다.

의존 구문 분석은 유럽의 전통적인 문법학자들, 특히 슬라브 영역에서 사용되는 구문적 표현의 형태로 발전했다[28]. 이후, Tesnière[43]의 연구로 현대적 이론의 의존 구문 분석이라는 명칭이 등장하게 되었다[30]. 이 당시 의존 구문 분석은 문장의 구조가 종속성이라고 불리는 이진 비대칭 관계로 연결된 어휘 요소로 구성된다고 가정하였다. 현대적 의존 구문 분석은 계층적 구조를 최소화한 문법이며, 단어 간의 관계가 이진법의 문법적 관계를 가지고, 단어 간 관계가 직접적으로 이어진다는 특징이 있다. 또한 형태소가 많고 어순이 자유로운 언어에 적합하고, 단어 간의 관계를 하나의 선으로 표시함으로써 필요한 정보만 보여준다[17].

<Figure 1>은 문장을 의존 구문 분석하여 나온 의존 구문 트리이다. 의존 구문 트리는 꼭지점과 꼭지점을 연결하는 호로 이루어져 있다. 이 때, 꼭지점은 각 어절을 의미하고, 호는 어절 간 관계를 나타내는 선이다. 호에 표기된 화살표는

의존 관계의 방향을 나타내며, 화살표가 들어가는 꼭지점이 의존소(dependent)가 되고 화살표가 나오는 꼭지점이 헤드(head)가 된다. 즉, 의존 구문 분석의 특징인 이진법적 특징을 갖는다는 것이다. 이때의 태깅된 값은 두 꼭지점 간의 의존 관계를 나타낸다. 예를 들어, ‘I’(dependent)와 ‘prefer’(head)는 nsubj(주어)의 이진법적 관계를 갖는다. 의존 구문 트리는 다음의 조건을 반드시 만족해야 한다. 각 꼭지점(어절)간 최소한 한 개의 선(호)은 연결되어야 하며, root를 부여받은 꼭지점이 존재해야 한다. 그리고 평면적이어야 하며, 선(호)이 교차되지 않아야 한다[17]. 여기서 root란, 그 무엇의 의존소도 되지 않는 지배소를 일컫는다. 위와 같은 조건에 따라 root를 부여받은 꼭지점에는 호가 들어오지 못하며, root를 제외한 나머지 꼭지점에서는 단 하나의 호만 나갈 수 있고, root부터 각각의 꼭지점에는 고유한 경로가 존재한다는 특징이 존재한다.



<Figure 1> Example of Dependency Parsing

한국어는 어순이 자유롭고, 체언과 용언에 붙는 접사와 어미에 따라 단어의 형태가 변하는 교착어의 특성에 위와 같은 의존 구문 분석을 하기 적합하다.

본 논문은 한국어 의존 구문 분석을 활용하여 연구를 진행하였으며, 한국어 의존 구문 분석 값은 영어와 다르기에 아래 <Table 1>의 기준을 따른다. 즉, 한국어는 크게 주성분, 부속성분, 독립성분의 세 가지로 분류할 수 있다. 주성분에는 주어, 서술어, 목적어, 보어가 속하고, 부속성분에는 관형어와 부사어가 속한다. 마지막으로 독립성분에는 감탄, 부름, 응답 등을 나타내는 독립어가 포함된다. 위 <Table 1>은 이러한 문장 성분을 관계에 따라 태깅할 수 있도록 한국정보통신기술협회에서 정의한 한국 의존 관계 태그 세트이다[24]. 구문 태그는 대상 어절이 체언인지, 용언인지, 관형사인지, 부사인지 등을 구분하고, 기능 태그 세트는 지배소와 가지는 관계가 주격인지, 목적격인지, 관형격인지 등을 구분한다. 각 어절은 의존 관계를 맺으면서 관계 태그를 반드시 부여받게 되는데, 이 때 부여받는 관계 태그는 두 가지 종류가 존재한다. 첫 번째는 구문 태그가 단독으로 부여되는 경우이고, 두 번째는 구문 태그와 기능 태그가 결합하여 부여되는 경우이다.

<Table 1> Korean Dependency Tag

Syntax Tag	Description	Function Tag	Description
NP	체언	SBJ	주어
VP	용언	OBJ	목적어
AP	부사구	MOD	관형어
VNP	긍정 지시사구	AJT	부사어
DP	관형사구	CMP	보어
IP	감탄사구	CNJ	접속어
X	의사구		
L	부호(왼쪽)		
R	부호(오른쪽)		

2.3 오피니언 스키마

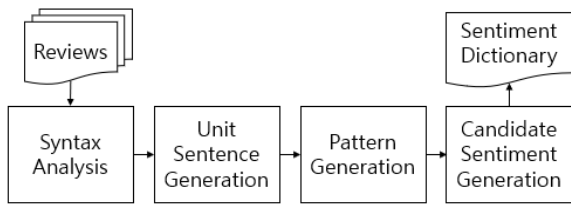
오피니언이란 개인의 평가 또는 감정을 포함하는 의견을 포함하는 광범위한 개념이다. 이 안에서의 감정은 의견이 암시하는 근본적인 긍정 혹은 부정만을 의미한다[26]. Liu[26]는 문장을 (g, s, h, t)와 같은 네 가지의 요소로 정의하였다. g(sentiment target)는 감정의 대상을 나타내며, s(sentiment)는 타겟에 대한 오피니언의 감정, h(holder)는 오피니언의 발화자, t(time)는 오피니언의 발화 시점을 의미한다. 예를 들어, 철수라는 사람이 2023년 1월 1일에 ‘전통시장은 가격이 싸서 좋아요.’라는 리뷰를 작성했다면, g는 ‘가격’처럼 감정의 대상을 나타내며, s는 ‘좋음’과 같이 타겟의 감정을 나타낸다. h는 작성자인 철수가 되며, t는 작성 시간인 2023년 1월 1일이 될 것이다. 이와 같이 리뷰(문장)는 네 가지의 요소로 정의할 수 있다. 하지만 위 예시를 보면, 사실 ‘가격’은 ‘전통시장’이라는 객체에 대한 대상을 나타낸다. 이처럼 g(sentiment target)는 객체와 속성이라는 두 가지 측면을 포함하고 있다. Liu[27]는 두 측면을 다시 분리하여 문장의 구조를 (e, a, s, h, t)로 정의하였다. e(entity)는 감정의 대상 객체이며, a(aspect)는 객체의 속성을 나타낸다. 위의 예시를 다시 보면, e는 감정의 대상 객체가 되는 ‘전통시장’이 될 것이며, a는 객체의 속성인 ‘가격’이 된다. 이 단계부터 우리는 속성 수준 감정 분석이라 부를 수 있고, h와 t는 선택적으로 사용할 수 있다. 위 예시를 더 살펴보면, ‘가격’이라는 속성(aspect)에 대한 감정(sentiment) ‘좋아요’, 그리고 그 때의 이유(reason)는 ‘싸서’임을 알 수 있다. 이와 같이 속성에 대한 감정의 이유를 나타내는 것이 r(reason)이다. 이러한 r을 통해 기업이나 상품 기획자는 감정에 대한 피드백이 가능하다. 그렇기에 r이 없는 문장보다 r이 존재하는 문장이 훨씬 많은 정보를 담고 있다고 할 수 있다. 따라서, 최종 오피니언 스키마는 (e, a, r, s, h, t)로 정의할 수 있다.

한편 본 연구에서는 기존 Liu[27]를 확대하여 e(entity)

와 a(aspect)를 하나의 A(aspect)로 병합하고, h(holder)와 t(time)는 제외하고, O(others)가 추가한 오피니언 스키마(A, R, S, O)를 활용하고자 한다. 이는 리뷰(문장)의 구조를 파악하는 데에 h와 t는 필수가 아니므로 삭제하고 a와 e는 하나의 A로 병합하여 간결성을 높였다. 그리고 O(others)를 통해 A, R, S 중 어디에도 속하지 않는 문장 요소들을 포함시켰다. 위 네 가지 스키마의 장점은 속성 수준 감성 분석에 필수적 요소인 속성과 감정 그리고 그 때의 이유에 대해 정확히 파악할 수 있고, 비교적 사용이 적은 나머지 정보는 따로 모아두었기 때문에 중요한 정보를 손쉽게 찾아볼 수 있다는 점이다.

3. 감성사전 구축 방법론

본 장에서는 제2장에서 다루었던 이론적 배경을 바탕으로 의존 구문 분석과 오피니언 스키마를 통해 감성 사전 구축에 대한 방법론을 제안한다.



<Figure 2> Sentiment Dictionary Construction Procedure

<Figure 2>는 제안 방법론의 전체 과정을 보여준다. 먼저 온라인 쇼핑몰 리뷰를 구문 분석하여 의존 구문 정보를 얻게 된다. 의존 구문 정보만으로는 의미 있는 결과를 도출하기 어렵기 때문에, 본 논문에서는 의존 구문 정보를 통해 단어 간의 의존 관계를 고려한 체인 형태의 리스트인 단위 문장을 생성하는 과정을 추가로 진행하였다. 이렇게 생성된 단위 문장을 바탕으로 패턴 분석을 거쳐 감성 단어 후보를 추출하고 후처리 작업을 거쳐 감성 사전을 구축한다. 특히, 기존의 감성 사전과 다르게 감성 표현과 그 이유를 분류하여 오직 감성 표현만을 감성 단어로 갖는 감성 사전을 구축한다.

3.1 구문분석

구문 분석에서는 온라인 쇼핑몰 리뷰를 문장 단위로 입력으로 받아서 의존 구문 분석을 진행하게 된다. 예를 들어 ‘성분이 순하고 피부가 촉촉해져서 사용하기 좋아요.’라는 문장이 있다고 하자. 이 문장은 6개의 어절로 구성되며, <Table 2>와 같은 결과를 얻을 수 있다.

<Table 2> Results of Dependency Parsing

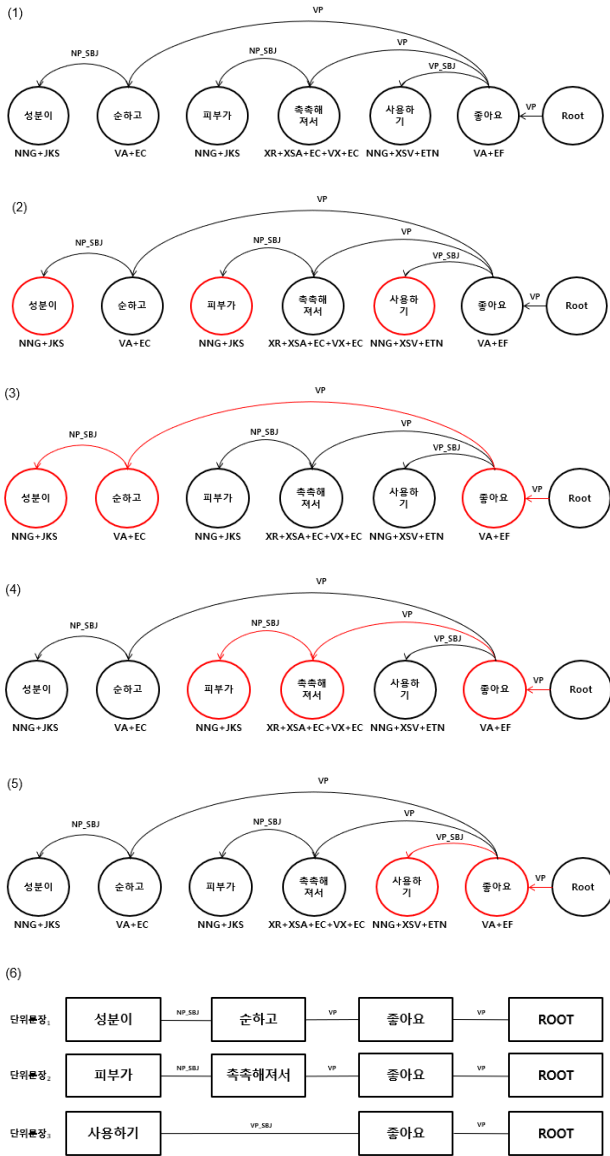
Token ID	Head	Token	POS Tag	DP Tag
1	2	성분이	NNG+JKS	NP_SBJ
2	6	순하고	VA+EC	VP
3	4	피부가	NNG+JKS	NP_SBJ
4	6	촉촉해져서	XR+XSA+EC+VX+EC	VP
5	6	사용하기	NNG+XSV+ETN	VP_SBJ
6	0	좋아요	VA+EF	VP

‘Token ID’는 각 단어에 할당된 일련번호로, 각 토큰의 개수만큼 1부터 n까지 부여된다. ‘Head’는 문장 내 토큰의 지배소에 해당하는 ‘Token ID’를 나타내며, 0은 어떠한 토큰의 지배도 받지 않는 root를 의미한다. ‘Token’은 문장을 어절로 분할한 단위이며, ‘POS Tag’와 ‘DP Tag’는 현재 어절의 품사태깅 결과와 지배소와 의존 관계 태깅 결과를 나타낸다. 예를 들어 Token ID’ 2의 토큰은 ‘순하고’이고, 이때의 품사 태깅 결과는 ‘VA+EC’임을 알 수 있다. 또한 ‘Token ID’ 2의 지배소는 6번 토큰 ‘좋아요’이며, 의존 관계 태깅은 ‘VP’이다. 다음으로 ‘Token ID’ 1,3,5를 보면 ‘Head’에 해당 ID 번호가 등장하지 않았는데, 이 3개의 토큰을 순수 의존소라고 한다. 즉, 모든 토큰은 의존소와 지배소의 역할을 하며 반드시 다른 토큰과 이항 관계를 갖게 되는데, 이들 중 오직 의존소의 역할만 하는 토큰이 순수 의존소이다. 반대로 오직 지배소의 역할만 하는 토큰이 root가 된다[30].

3.2 단위 문장 생성

단위 문장(unit sentence)이란, 문장 내 토큰들 간의 관계를 연쇄적으로 연결한 모양이며, 의존 관계가 존재하는 토큰들의 리스트라 정의하겠다. 즉, 하나의 문장에서 순수 의존소를 시작으로 root로 끝나는 의존 관계들의 사슬이다. 예를 들어 앞서 제시된 예시 ‘성분이 순하고 피부가 촉촉해져서 사용하기 좋아요.’를 사용하여 단위 문장을 생성하는 과정은 <Figure 3>과 같다. (1)은 리뷰 문장이 의존 구문 분석을 거쳐 만들어지는 의존 구문 트리이다. (2)는 (1)에서 순수 의존소를 찾는 과정으로, 일련의 과정을 거쳐 3개의 순수 의존소를 찾은 것을 보여준다. (3), (4), (5)는 각각의 순수 의존소로부터 root(head_n=0)까지 슬라이스(slice)하는 과정으로, 순수 의존소의 토큰 노드부터 head 노드로 이어지며 노드간의 의존 관계는 edge에 태깅되어 있다. 최종적으로 (6)에서 3개의 단위 문장이 생성되는 결과 예시를 볼 수 있다. 단위 문장의 시작 노드는 항상 순수 의존소에 해당하므로, 단위 문장은 순수 의존소의 개수만큼 생성되게 된다. 즉 <Figure 3>의 해당 리뷰문장의

순수의존소는 3개이므로, 3개의 단위 문장을 생성한다.



<Figure 3> Example of Unit Sentence Generation

3.3 의존 관계 패턴 분석

앞에서 생성한 단위 문장에서 감성 단어를 추출하고, 감성 사전을 구축하기 위해서는 패턴 분석이 필요하다. 본 논문에서의 패턴은 두 토큰 사이의 관계에 해당하는 단위 의존 관계 Tag를 통해 이루어지며, 정의는 다음과 같다.

(정의1. 패턴) 단위 문장이 주어질 때, 단위 문장 내에서 연속으로 등장하는 두 개의 의존관계의 연결을 의미한다. (의존관계 k , 의존관계 $k+1$). 단, $k=1,2,\dots, n$.

온라인 쇼핑몰 리뷰를 감성 분석하여 패턴화를 진행한 선행 연구에 따르면, 감성 단어가 등장하는 패턴에서의 감성 단어는 단위 문장의 가장 뒤에 위치할 확률이 높았다 [46]. 또한, 온라인 쇼핑몰 리뷰의 특성 상 감성 단어는 주로 객체나 감성 단어의 이유에 해당하는 토큰과 의존 관계를 맺고 있고, 전처리를 거쳐 생성된 단위 문장의 특성 상 최소 두 개 이상의 의존 관계가 생성되기 때문에 root를 토큰으로 포함한 세 개의 토큰을 활용한 패턴을 정의하였다. 한편 본 논문의 패턴에 활용될 단위 의존 관계 Tag는 단위 문장에 등장하는 전체 단위 의존 관계 Tag 중 빈도수가 높은 상위 10개의 의존 관계 Tag와 빈도수가 적은 나머지를 하나로 묶은 Others로 통합하여, 총 11개로 구성하여 패턴의 가능한 조합 수는 $11 \times 11 = 121$ 개가 된다.

단위문장으로부터 패턴의 조합을 발견하기 위해서는 다음과 같은 절차를 따른다. 먼저 하나의 단위 문장 내 두 토큰의 의존관계를 의미하는 edge 들을 추출하여 모두 나열한 후, 두개의 연속된 edge 들의 의존관계를 나열하면 패턴이 된다. 예를 들어 예를 들어, <Figure 3> (2)에서 단위 문장 ‘성분이-순하고-좋아요.’의 edge 들의 의존관계를 나열하면 ‘NP_SBJ, VP, VP’가 되며, 추출된 순서대로 ‘NP_SBJ, VP’, ‘VP, VP’의 패턴을 생성하면 된다.

3.4 감성 단어 후보 도출 및 감성 사전 구축

감성 단어가 존재하는 온라인 쇼핑몰 리뷰를 구문 분석하여 문장 내 토큰들의 관계를 얻었고, 단위 문장 생성을 통해 해당 토큰들과 그 의존 관계를 나타내었다. 다음으로 단위 의존 관계 Tag를 통해 의존 관계 조합을 생성하였고, 해당 의존 관계 조합에서 상위 빈도 조합을 추출하여 본 연구의 패턴으로 정의했다. 이번 절에서는 위와 같은 방법을 통해 얻어진 패턴을 활용하여 감성 단어 후보를 도출하고 감성 사전을 구축한다.

감성 사전을 구축하기 위해 두 가지 과정이 필요하다. 본 논문에서는 이를 단어 쌍(word-pair) 추출과 후처리 작업으로 구분하였다. 먼저 단어 쌍 추출은 위 과정을 통해 도출된 패턴에 해당하는 토큰을 추출하는 것을 말한다. 단어 쌍추출을 통해 얻어진 각각의 토큰들은 상위 의존 관계 조합 패턴에 해당하는 단어들로, 감성 단어가 아닐 확률이 존재한다. 본 논문에서는 추출된 토큰들의 의존 관계 Tag를 바탕으로 감성 단어 후보를 분류한다. 이모티콘을 제외한 감성 단어는 ‘기대하다’, ‘좋다’, ‘최고다’와 같이 각각 ‘동사’, ‘형용사’, ‘명사+이다’의 문법적 형태를 갖는 경우가 대부분이며, 의존 관계 Tag는 오직 용언(동사, 형용사, 보조용언)에 해당하는 ‘VP’와 긍정 지정사구(명사+이다)에 해당하는 ‘VNP’ 두 개만 존재한다. 따라서 본 논문은 감성 단어 후보로 도출할 때 의존 관계 Tag가 ‘VP’와

‘VNP’를 갖는 패턴의 토큰을 사용한다.

도출된 감성 단어 후보는 교차 검증과 후처리 작업을 거쳐 감성 사전으로 구축된다. 감성 사전 구축은 아래와 같이 두 가지 과정을 따른다. 먼저 상위 의존 관계 조합을 통해 생성된 패턴 내에 ‘VP’ 혹은 ‘VNP’가 존재한다면 해당 패턴에서 추출된 단어 쌍을 교차 검증하여 중복 제거 작업을 거친 후, 남은 의존 관계 Tag가 ‘VP’ 혹은 ‘VNP’에 해당하는 토큰들을 어근화하고 정렬과 일부 레이블링 작업을 진행하여 감성 사전으로 구축한다. 만약 ‘VP’ 혹은 ‘VNP’가 연속 등장하였을 경우, 해당 패턴을 완전단위 문장 패턴으로 분류하여 처리한다.

(정의2. 완전 단위 문장) 단위 문장 내에 속성(A: aspect), 이유(R: Reason), 감성(R: Sentiment)의 세 가지 토큰들을 모두 포함할 때 해당 단위 문장은 완전 단위 문장이다.

예를 들어 ‘성분이 순해서 좋아요.’라는 단위 문장은 ‘성분’이라는 속성과, ‘좋아요’라는 감성표현과, ‘순해서’라는 감성표현의 이유를 모두 포함하고 있는 완전 단위 문장이다. 한편, 속성은 명사, 대명사, 수사 등 체언의 문법적 형태를 가지므로 의존 관계 태그는 체언을 뜻하는 ‘NP’가 되고, 이유와 감성의 표현은 용언 혹은 궁정 지정사구를 뜻하는 ‘VP’와 ‘NP’가 태깅 된다. 이에 본 연구에서는 다음과 같이 완전패턴을 정의하고, 감성단어의 후보를 찾도록 한다.

(정의3. 완전 패턴) 완전 단위 문장이 주어질 때, 완전 단위 문장 내에서 연속으로 등장하는 두 개의 의존관계가 (VP or VNP, VP or VNP)되는 패턴이다.

본 연구에서는 이유와 감성 표현을 분류하여 감성 표현만 감성 단어로 정의하고 감성 사전을 구축하기 위해서 완전 패턴을 정의하여 감성 표현을 추출하고자 한다. 이때, 감성 표현은 단독으로 등장하여도 극성을 알 수 있다고 정의하였는데, 예를 들어 ‘예쁘다’와 ‘귀엽다’는 예뻐서 좋을지 귀여워서 나쁠지가 사용자에게 따라 달라지기 때문에 감성 표현이 될 수 없고 감성 표현에 대한 이유에 해당한다. 따라서 완전 단위 문장과 완전 패턴을 발견할 때, 감성 단어의 후보를 완전 패턴에서 후자에 해당하는 토큰이 된다.

(정의4. 감성단어) 완전 패턴 (VP or VNP, VP or VNP)이 주어질 때, 두 번째 등장한 의존관계 Tag인 ‘VP’ 혹은 ‘VNP’를 갖는 토큰이다.

따라서, 리뷰를 구성하는 모든 문장은 의존 구문 분석을 통해 단위 문장을 생성할 수 있고, 위 절차에 따라 온라인 쇼핑몰 리뷰로부터 감성 사전을 구축할 수 있다.

4. 사례 연구

본 연구에서 사용된 의존 구문 분석은 Mecab-ko 와 KLUE-DP를 결합한 한국어 의존 구문 분석기이다. Mecab[20]은 일본어나 중국어 등 띄어쓰기가 없는 언어에서의 단어 범위를 찾기 위해 CRFs(Conditional Random Fields)을 바탕으로 만들어진 토큰화 모델이며, 이를 한국어에 적용한 것이 Mecab-ko이다[32]. KLUE[33]는 국내 10개 기관에서 연구를 진행한 한국어 자연어 평가 데이터셋으로, 주제 분류/문장 유사도 비교/형태소 및 의존 구문 분석 등 여덟 가지 tasks로 이루어져 있으며, KLUE-DP는 문장에 나타나는 단어들의 지배소와 의존소의 관계를 예측하고 관계 정보를 찾아 분류하는 모델이다

사례연구를 위해 네이버 온라인 쇼핑몰 사이트에서 화장품 스킨토너 제품 대상 2018년 11월11일부터 2022년 3월 31일까지의 리뷰데이터 30,000개를 크롤링한 결과를 사용하였다. 30,000개의 리뷰문장을 입력으로 의존구문을 분석하여 총 52,955개의 단위문장이 생성되었으며, 단위 문장 내 등장한 의존관계 태그는 총 240,291개이고, 각 태그가 등장한 횟수와 비율은 <Table 3>과 같다. 전체 의존관계 태그 49개 중 총 28개가 등장하였고, 상위 10위까지의 누적비율은 약 97%이다. 이를 통해 패턴 조합생성에 사용될 11개의 태그를 도출하였다: VP, AP, NP_SBJ, NP_AJT, NP, VNP, VP_MOD, VP_AJT, VP_SBJ, Others.

전체 121개의 패턴조합 중 상위 세 개의 패턴은 각각 (VP, VP) 패턴이 15,423회(15.4%), (AP, VP) 패턴이 11,190회(11.2%), (NP_SBJ, VP) 패턴이 6,921회(6.9%) 등장하였으며, 해당되는 단어쌍의 예시는 <Table 4>와 같다. 한편, 위 세 가지 패턴 중 (VP, VP) 패턴만 완전패턴에 해당하므로, 패턴 중 두 번째 VP에 해당하는 토큰을 추출하여 화장품 분야의 감성사전을 <Table 5>와 같이 구축하였다.

본 논문의 제안방법론은 감성 표현의 이유를 감성 단어로 정의하지 않고 따로 분리하였기 사례분석을 통해 도출한 감성단어가 적절한지 여부를 평가하기는 쉽지 않다. 따라서 본 연구에서는 화장품 도메인의 리뷰로부터 미리 완전단위문장 1,000개를 구축하고, 이를 통해 17개의 감성단어로 이루어진 기준 감성사전을 <Table 6>과 같이 도출하고 평가를 실시하였다. 평가결과 F1-Score는 0.59이며, Recall과 Precision은 각각 0.71, 0.50이다. F1-Score는 0.59로 그리 높지 않은 수치를 보여준다. 이는 기준 감성사전을 만들 때 사용한 완전단위문장의 샘플 크기가 30,000개 대비 너무 적어서 Precision이 낮아진 이유이며, Recall 또한 ‘ㅎㅎ’이나 ‘ㅋㅋ’와 같은 단어가 감성단어 뒤에 추가되어, 정작 감성단어가 되었어야 할 VP가 감성단어 대상에서 제외된 결과에 대부분 기인한다. 추후 ‘ㅎㅎ’, ‘ㅋㅋ’와 같은 온라인 상품 리뷰에 자주 등장하는 부사어를 전처리

하여 연구를 진행한다면 훨씬 높은 Recall을 얻을 수 있을 것으로 예상된다.

<Table 3> Frequency of Tags in 52,955 Unit Sentences

dp_tag	frequency	%	dp_tag	frequency	%
VP	101,105	42.08	NP_CNJ	690	0.29
AP	34,822	14.49	VP_CMP	614	0.26
NP_SBJ	22,639	9.42	VNP_MOD	438	0.18
NP_AJT	20,422	8.50	VNP_CMP	191	0.08
NP	13,712	5.71	IP	161	0.07
VNP	11,929	4.96	X_AJT	48	0.02
VP_MOD	11,738	4.88	AP_MOD	33	0.01
NP_OBJ	10,382	4.32	X	30	0.01
VP_AJT	3,094	1.29	VNP_OBJ	30	0.01
VP_SBJ	2,820	1.17	VP_CNJ	28	0.01
dp	2,358	0.98	VNP_SBJ	26	0.01
VP_OBJ	1,137	0.47	VNP_AJT	16	0.01
NP_CMP	1,082	0.45	X_SBJ	16	0.01
NP_MOD	722	0.30	X_OBJ	8	0.00

<Table 4> Word-pair Extraction Results by Pattern

Pattern	Word-Pair
VP-VP	순하고-좋아요, 족족해서-좋아요 많아서-좋아요, 써서-만족합니다 자극없고-괜찮아요
AP-VP	확실히-맡아졌어요, 잘-샀네요 더-써봐야, 늘-구비하고, 너무-족족해요
NB_SUBJ-VP	피부도-족족해지고, 용량이-많아서 토너-가볍고, 토너-순하고, 토너-족족해서

<Table 5> Sentiment Dictionary (Predicted)

Sentiment words	Polarity	Sentiment words	Polarity
좋다	긍정	부담되다	부정
ㅎㅎ	중립	아쉽다	부정
만족하다	긍정	놀라다	중립
괜찮다	긍정	걱정되다	부정
모르다	부정	굿이다	긍정
감사하다	긍정	고민하다	부정
ㅍㅍ	중립	후회하다	부정
믿다	긍정	안심하다	긍정
기대하다	긍정	답답하다	부정
추천하다	긍정	고맙다	긍정
나쁘다	부정	지겹다	부정
무난하다	긍정	번잡하다	긍정

<Table 6> Sentiment Dictionary (Actual)

Sentiment Words	Polarity	Sentiment Words	Polarity
좋다	긍정	아쉽다	부정
만족하다	긍정	궁금하다	중립
감사하다	긍정	놀라다	중립
괜찮다	긍정	나쁘다	부정
추천하다	긍정	강추이다	긍정
믿다	긍정	사랑하다	긍정
무난하다	중립	충분하다	긍정
기대하다	긍정	짱이다	긍정
굿이다	긍정		

5. 결론

온라인 채널 상품 리뷰에서 속성에 따른 감성 표현을 추출하는 연구들이 활발한 가운데, 감성 사전을 구축하는 것은 많은 시간과 비용이 들며, 각 도메인별로 다른 감성 표현을 사용하기 때문에 새로운 감성 사전을 매번 구축해야 한다. 또한, 영어권 국가는 SenticNet, WordNet-Affect, VADER 등 어느 도메인에도 사용될 수 있는 감성 사전이 존재한다. 이는 국가적 차원에서 많은 지원이 있기에 가능한 일인데, 현재 국내에서는 이제 시작되는 단계이기에 아직 미약한 현실이다. 따라서, 한글이 가진 언어적 특징을 반영할 수 있는 구문 분석을 통하여 단어와 단어 사이의 관계를 파악하고, 그 관계의 패턴을 발견하여 감성 단어를 추출하는 방법이 필요하다. 본 논문에서는 속성 수준 감성 분석에 필요한 감성 단어와 그 이유를 분리하여 감성 사전을 구축하는 새로운 방법론을 제안하였다. 즉, 의존 구문 분석을 활용하여 문장 내 감성 단어와 그 이유 사이의 관계를 파악하고, 감성 단어가 등장하는 패턴을 추출하여 감성 사전을 구성할 수 있다. 사례 연구를 통해 적은 노력으로 의미있는 감성사전 구축의 가능성을 확인하였으며, 향후 감성 단어의 형태소 분석, 복잡한 패턴 정의, 전처리 개선 등을 통해 감성 분석의 정확도를 높이는 연구가 필요하다

Acknowledgement

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2023S1A5A2A03083440).

References

[1] Aggarwal, C.C. and Aggarwal, C.C., *Machine learning*

- for text: *An introduction*, Springer International Publishing, 2018.
- [2] Ahn, J. and Kim, H., Building a Korean Sentiment Lexicon Using Collective Intelligence, *Journal of Intelligence and Information Systems*, 2015, Vol. 21, No. 2, pp. 49-67.
- [3] Ahmed, M., Chen, Q., and Li, Z., Constructing domain-dependent sentiment dictionary for sentiment analysis, *Neural Computing and Applications*, 2020, Vol. 32, pp. 14719-14732.
- [3] Alshari, E.M., Azman, A., Doraisamy, S., Mustapha, N., and Alkeshr, M., Effective method for sentiment lexical dictionary enrichment based on Word2Vec for sentiment analysis, In *fourth international conference on information retrieval and knowledge management*, IEEE, 2018.
- [4] Baccianella, S., Esuli, A., and Sebastiani, F., Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, In *Lrec*, 2010.
- [5] Behdenna, S., Barigou, F., and Belalem, G., Sentiment analysis at document level, In *Smart Trends in Information Technology and Computer Communications: First International Conference*, 2016.
- [6] Bian, S., Jia, D., Li, F., and Yan, Z., A new Chinese financial sentiment dictionary for textual analysis in accounting and finance, 2021, Available at SSRN 3446388.
- [7] Birjali, M., Kasri, M., and Beni-Hssane, A., A comprehensive survey on sentiment analysis: Approaches, challenges and trends, *Knowledge-Based Systems*, 2021, Vol. 226, p. 107134.
- [8] Cambria, E., Speer, R., Havasi, C., and Hussain, A., Senticnet: A publicly available semantic resource for opinion mining, In *2010 AAAI fall symposium series*, 2010.
- [9] Cambria, E., Havasi, C., and Hussain, A., Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis, 2012, In *Twenty-Fifth international FLAIRS conference*.
- [10] Cho, S.H. and Kang, H.B., Text sentiment classification for SNS-based marketing using domain sentiment dictionary, In *IEEE International Conference on Consumer Electronics*, 2012.
- [11] Do, H.H., Prasad, P.W., Maag, A., and Alsadoon, A., Deep learning for aspect-based sentiment analysis: a comparative review, *Expert Systems with Applications*, 2019, Vol. 118, pp. 272-299.
- [12] Havasi, C., Speer, R., and Alonso, J., ConceptNet: A lexical resource for common sense knowledge, *Recent advances in natural language processing V: selected papers from RANLP*, 2007, Vol. 309, p. 269.
- [13] Heo, C. and Ohn, S., A Novel Method for Constructing Sentiment Dictionaries using Word2vec and Label Propagation, *Journal of Korean Institute of next Generation Computing*, 2017, Vol. 13, No. 2, pp. 93-101.
- [14] Hong, Y., Lee, J., and Lee, G., A Korean Syntactic Analyzer based on the Dependency Grammar, In *KIISE Conference*, 1993, pp. 781-784.
- [15] Hutto, C. and Gilbert, E., Vader: A parsimonious rule-based model for sentiment analysis of social media text, In *Proceedings of the International AAAI Conference On Web and Social Media*, 2014.
- [16] Jiaheng, H., Yonghua, C., and Chengyao, W., Constructing sentiment dictionary with deep learning: Case study of financial data, *Data Analysis and Knowledge Discovery*, 2018, Vol. 2, No. 10, pp. 95-102.
- [17] Jurafsky, D. and James H.M., *Speech and language processing*, Pearson Education India, 2019.
- [18] Kim, J., Remarks on "Sentence", *Poetics and Linguistics*, 2003, Vol. 6, pp. 65-113.
- [19] Kim, J., Oh, Y., and Chae, S., Construction of a Domain-Specific Sentiment Dictionary Using Graph-based Semi-supervised Learning Method, *Science of Emotion and Sensibility*, 2015, Vol. 18, No. 1, pp. 97-104.
- [20] Kudo, T., Yamamoto, K., and Matsumoto, Y., Applying conditional random fields to Japanese morphological analysis, In *Proceedings of the conference on empirical methods in natural language processing*, 2004.
- [21] Kwon, O., Kim, J., Cho, H., Hong, K., Han, J., Jung J., Kim, Y., and Choi S., KHU-SentiwordNet: Developing A Korean SentiwordNet Combining Empty Morpheme, In *Korea Society of IT Service Conference*, 2019, pp. 194-197.
- [22] Lee, J., Research on Designing Korean Emotional Dictionary using Intelligent Natural Language Crawling System in SNS, *The Journal of Information Systems*, 2020, Vol. 29, No. 3, pp. 237-251.
- [23] Li, S., Shi, W., Wang, J., and Zhou, H., A deep learning-based approach to constructing a domain sentiment lexicon: A case study in financial distress prediction, *Information Processing and Management*, 2021, Vol. 58, No. 5, 102673.
- [24] Lim, J., Bae, Y., Kim, H., Kim, Y., and Lee, K., Korean

- Dependency Guidelines for Dependency Parsing and Exo-Brain Language Analysis Corpus, *In Hangul and Korean Information Processing Conference*, 2015.
- [25] Liu, H., and Singh, P., ConceptNet—a practical commonsense reasoning tool-kit, *BT Technology Journal*, 2004, Vol. 22, No. 4, pp. 211-226.
- [26] Liu, B., *Sentiment analysis and subjectivity*, Handbook of natural language processing, 2010.
- [27] Liu, B., *Sentiment analysis and opinion mining*, Springer Nature, 2022.
- [28] Mel'cuk, I. A., *Dependency syntax: theory and practice*, SUNY press, 1988.
- [29] Nazir, A., Rao, Y., Wu, L., and Sun, L., Issues and challenges of aspect-based sentiment analysis: A comprehensive survey, *IEEE Transactions on Affective Computing*, 2020, Vol. 13, No. 2, pp. 845-863.
- [30] Nivre, J. Dependency grammar and dependency parsing, *MSI report*, 2005, Vol. 5133, No. 1959, pp. 1-32.
- [31] Manning, C.D. and Zeman, D., Universal dependencies v1: A multilingual treebank collection, *In Proceedings of the Tenth International Conference on Language Resources and Evaluation*, 2016.
- [32] Park, K., Lee, J., Jang, S., and Jung, D., An empirical study of tokenization strategies for various Korean NLP tasks, arXiv preprint arXiv:2010.02534, 2020.
- [33] Park, S., Moon, J., Kim, S., Cho, W. I., Han, J., Park, J., ... and Cho, K., Klue: Korean language understanding evaluation, arXiv preprint arXiv:2105.09680, 2021.
- [34] Park, S., Na, C., Choi, M., Lee, D., and On, B., Korean Sentiment Lexicon: Bi-LSTM-based Method for Building a Korean Sentiment Lexicon, *Journal of Intelligence and Information Systems*, 2018, Vol. 24, No. 4, pp. 219-240.
- [35] Plutchik, R., *A general psychoevolutionary theory of emotion*, In Theories of emotion, Academic press, 1980.
- [36] Poria, S., Gelbukh, A., Hussain, A., Howard, N., Das, D., and Bandyopadhyay, S., Enhanced SenticNet with affective labels for concept-based opinion mining, *IEEE Intelligent Systems*, 2013, Vol. 28, No. 2, pp. 31-38.
- [37] Shin D., Cho, D., and Nam, J., Building the Korean Sentiment Lexicon DecoSelex for Sentiment Analysis, *Journal of Korealex*, 2016, Vol. 28, pp. 75-111.
- [38] Sivasankar, E., Krishnakumari, K., and Balasubramanian, P., An enhanced sentiment dictionary for domain adaptation with multi-domain dataset in Tamil language (ESD-DA), *Soft Computing*, 2021, Vol. 25, pp. 3697-3711.
- [39] Song, J. and Lee, S., Automatic Construction of Positive/Negative Feature-Predicate Dictionary for Polarity Classification of Product Reviews, *Journal of Computing Science and Engineering*, 2011, Vol. 38, No. 3, pp. 157-168.
- [40] Song, K., Feng, S., Gao, W., Wang, D., Chen, L., and Zhang, C., Build emotion lexicon from microblogs by combining effects of seed words and emoticons in a heterogeneous graph, *In Proceedings of the 26th ACM Conference on Hypertext and Social Media*, 2015.
- [41] Steinberger, J. Ebrahim, M., Ehrmann, M., Hurriyetoglu, A., Kabadjov, M., Lenkova, P., Steinberger, R., Tanev, H., Vázquez, S., Zavarella, V., Creating sentiment dictionaries via triangulation, *Decision Support Systems*, 2012, Vol. 53, No. 4, pp. 689-694.
- [42] Tan, A.H., Text mining: The state of the art and the challenges, *In Proceedings of the Pakdd Workshop on Knowledge Discovery from Advanced Databases*, 1999.
- [43] Tesnière, L., *Elements de syntaxe structural*, Editions Klincksieck, 1959.
- [44] Tsai, A.C.R., Wu, C.E., Tsai, R.T.H., and Hsu, J.Y.J., Building a concept-level sentiment dictionary based on commonsense knowledge, *IEEE Intelligent Systems*, 2013, Vol. 28, No. 2, pp. 22-30.
- [45] Yang, L., Li, Y., Wang, J., and Sherratt, R.S., Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning, *IEEE access*, 2020, Vol. 8, pp. 23522-23530.
- [46] Yang, Y., Aspect-Level Opinion Pattern Mining Using Dependency Relations in Product Reviews, [Master's thesis], Hanyang University 2023.
- [47] Yue, L., Chen, W., Li, X., Zuo, W., and Yin, M., A survey of sentiment analysis in social media, *Knowledge and Information Systems*, 2019, Vol. 60, pp. 617-663.

ORCID

Woo Hyun Kim | <http://orcid.org/0009-0002-1593-8061>

Heejung Lee | <http://orcid.org/0000-0001-7548-9291>