

사전학습 언어모델을 사용한 다국어 속성기반 감성분석: 호텔 리뷰를 중심으로

장호준¹ · 전다원¹ · 이희상^{2*}

¹성균관대학교 일반대학원 산업공학과 / ²성균관대학교 시스템경영공학과

Multilingual Aspect-based Sentiment Analysis Using Pre-trained Language Models: A Focus on Hotel Reviews

Hojun Jang¹ · Dawon Jeon¹ · Heesang Lee²

¹Department of Industrial Engineering, Sungkyunkwan University

²Department of Systems Management Engineering, Sungkyunkwan University

Consumers tend to rely on user reviews when making purchasing decisions, and reviews play a crucial role in corporate marketing and service development. The latest natural language processing technique, ABSA (Aspect-Based Sentiment Analysis), was employed to understand consumer sentiments in multilingual reviews in detail. The training data consisted of Korean, English, and Japanese reviews to detect model performance by linguistic typology. A multilingual pre-trained language model, XLM-RoBERTa (Cross-Lingual Language Model of Robustly optimized BERT pretraining approach), was used to analyze sentiments by aspects. The performance evaluation of the model showed that the model using all three languages—Korean, English, and Japanese—performed the best regardless of linguistic differences. Based on the proposed model, we analyzed the aspects of overseas hotel reviews on a domestic hotel service platform. Through the results of this study, we proposed directions for companies aiming to leverage multilingual reviews in constructing AI training data and implementing ABSA effectively.

Keywords: ABSA, Pre-trained Language Model, Multilingual BERT, XLM-RoBERTa, Customer Review Analysis

1. 서론

다양한 분야의 온라인 서비스 플랫폼이 등장함에 따라 미디어에서 발생하는 데이터의 양은 기하급수적으로 늘어나고 있다. 많은 소비자는 온라인 서비스 이용 시 기업의 일방적 정보에 의존하지 않고 인터넷 사용 후기를 보고 구입하고 있으며, 한국소비자연맹(CUK, 2022)의 설문조사 결과에 따르면 소비자의 97.2%가 구매 전 이용 후기를 확인하고 있고 82.4%는 이용 후기의 누적 수가 구매 선택에 영향을 미친다고 응답했다. 이처럼 작성된 이용 후기는 소비자의 구매 결정에 큰 영향을 미

치며, 기업과 조직은 소비자 후기 분석을 통한 마케팅 기법 및 서비스 개발을 위해 제품과 서비스에 대한 소비자 의견을 파악하기를 원한다(Liu, 2012). 하지만 대량의 소비자 리뷰 데이터를 사람이 직접 관리하고 분석하는 것은 시간, 비용의 문제가 발생하며, 따라서 이를 효율적으로 분석하기 위한 체계적인 방법이 필요하다. 이러한 맥락에서 텍스트 분석은 대량의 비정형 데이터를 효율적으로 분석하는 방법으로 시장 조사, 마케팅 계획 및 상품에 대한 전략적 의사결정 등에 좋은 방법으로 사용되고 있다(Hassani *et al.*, 2020). 텍스트 분석에는 텍스트 마이닝, 토픽 모델링, 감성 분석, 자연어 분석 등 다양한

본 논문은 24년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원 지원을 받아 수행되었음(RS-2024-00439045).

* 연락저자 : 이희상 교수, 16419 경기도 수원시 장안구 서부로 2066 성균관대학교 시스템경영공학과, Tel : 031-290-7628, E-mail : leehee@skku.edu
2024년 7월 19일 접수; 2024년 10월 7일 게재 확정.

분석 모델이 존재하며, 이 중 감성 분석(Sentiment Analysis)은 텍스트를 분석하여 감성의 극성을 예측하고 가치를 창출하는 방법으로 최근 증가하는 소비자 리뷰 데이터를 분석하고 이해하기 위한 효과적인 수단으로 등장하여 소비재, 서비스, 금융 등 거의 모든 영역으로 확산되고 있다(Liu, 2012), 하지만 소비자 생성 리뷰의 경우 제품의 가격, 배송, 만족도, 서비스 등 여러 가지 속성으로 나타나는 복합적인 감성으로 인해 하나의 리뷰를 하나의 감성으로 판별하는 것은 정확한 분석이 어렵다는 한계가 존재한다(Zhang *et al.*, 2022). 이러한 문제를 해결하기 위해 최근 리뷰 내의 다양한 속성과 그 속성별 감성을 예측할 수 있는 Aspect-Based Sentiment Analysis(이하: ABSA)에 대한 연구들이 보고되고 있다(Zhang *et al.*, 2022). ABSA의 주요 요소는 <Figure 1>와 같이 속성어(Asspect Term), 속성(Asspect Category), 감성어(Opinion Term), 극성(Sentiment Polarity)의 4가지로 구분되며 검출 혹은 예측하려는 항목에 따라 단일 혹은 복합 과업으로 나뉘어진다(Zhang *et al.*, 2022).

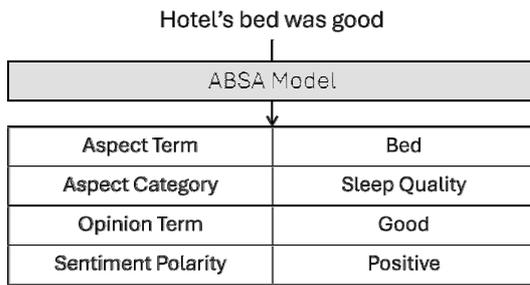


Figure 1. Key Elements of ABSA: Example of Hotel Review

ABSA는 Amazon, Yelp 등과 같은 E-commerce와 웹사이트의 소비자 리뷰에서 활용되기 시작했고, 다양한 응용 분야에서 ABSA 모델 개발을 위한 벤치마크 데이터 세트도 증가하고 있다(Do *et al.*, 2019). 또한 ABSA 모델의 분석 대상이 가지고 있는 각 속성에 대한 감성을 세부적으로 분석한다는 이점이 비즈니스 측면에서 실질적인 가치를 제공하기 때문에 학계뿐만 아니라 산업계의 주목을 받고 있다(Araque *et al.*, 2019; Peng *et al.*, 2018; Park and Shin, 2020). 최근에는 다중 과업 학습(Multi-Task Learning) 및 딥러닝을 활용한 ABSA 연구(Hua *et al.*, 2024)와 사전학습 언어모델(Pre-trained Language Model: PLM)을 활용한 ABSA 성능 향상 연구 등도 활발히 진행되고 있다. 그러나 대부분의 ABSA 연구는 데이터의 풍부함으로 인해 영어 데이터를 사용한 연구가 주로 진행되었으며(Zhang *et al.*, 2021; Kandhro *et al.*, 2024), 데이터가 희소한 아시아권 국가 언어를 기반으로 한 연구는 부족한 실정이다. 또한 여러 언어를 동시에 사용한 다국어 ABSA 연구는 더 지능적이고 우수한 성능의 ABSA 방법론 구축에 있어 여전히 도전적인 문제로 간주되고 있으며(Zhang *et al.*, 2022), 이는 전 세계 인터넷 사용자의 28.6%만이 영어권 국가에 거주하고 있어(Lo *et al.*, 2017), 실제 E-commerce에서 나타나는 다국어 소비자 리뷰 분석을

위해서는 다수의 관련 연구가 진행되어야 할 필요가 있다. 이에 본 연구는 온라인 플랫폼에서 생성된 사용자 리뷰 데이터의 효과적인 분석을 위해 최신 자연어 처리 기법인 사전학습 언어모델을 사용해 비영어권 국가의 언어 데이터를 활용한 다국어 ABSA 모델을 구축하였고, 구축한 모델을 통해 실제 다국어 사용 환경의 호텔에 대한 소비자 리뷰 데이터를 분석하여 본 연구에서 제안하는 모델의 활용 가능성을 제시하였다.

본 논문의 제2장에서는 사전학습 언어모델과 국내외의 ABSA 선행 연구 및 사전학습 언어모델의 다국어 학습 능력과 관련된 선행 연구를 검토하였으며, 제3장에서는 분석에 사용한 데이터를 서술하였고, 제4장에서는 분석 방법론에 대해 언급하였다. 제5장에서는 수행한 다국어 ABSA 실험 결과와 제안 모델을 통한 다국어 리뷰 데이터 분석 내용을 서술하였으며, 마지막으로 제6장에서 연구의 기여도와 한계점, 향후 개선 방향에 관해 서술하였다.

2. 선행연구

2.1 사전학습 언어모델

BERT(Bidirectional Encoder Representations from Transformers)(Devlin *et al.*, 2019)는 Google AI Language에서 공개한 초거대 언어모델(Large Language Model: LLM)로, Transformer(Vaswani *et al.*, 2017)의 Encoder 구조를 기반으로 텍스트의 좌우 문맥을 동시에 고려한 양방향 자연어 학습 능력이 특징이다. BERT의 입력 형식은 단일 문장 또는 문장 쌍을 하나의 입력 형태로 구성할 수 있으며, 모든 문장의 첫 번째 토큰과 문장 쌍의 구분을 위해 각각 특수 분류 토큰인 [CLS], [SEP] 토큰을 사용한다. 각 토큰의 최종 입력 형태는 해당 토큰 임베딩, 세그먼트 임베딩, 위치 임베딩의 합으로 구성된다(Devlin *et al.*, 2019). BERT는 입력 토큰의 일부를 무작위로 마스킹한 후, 마스킹된 토큰을 예측하는 MLM(Masked Language Model), 두 문장이 주어졌을 때, 두 번째 문장이 첫 번째 문장의 다음 문장인지를 예측하는 NSP(Next Sentence Prediction)의 두 가지 학습 과업을 통해 사전학습을 진행하며, 사전학습된 모델을 미세조정(Fine-Tuning) 하는 방식으로 다양한 자연어 처리 과업에서 성능 향상을 이끌어내는 장점을 가지고 있다(Devlin *et al.*, 2019). 이후 등장한 Multilingual BERT는 BERT와 동일한 구조의 사전학습 언어모델로 사전학습 시 영어 단일어 데이터가 아닌 104개 언어로 구성된 다국어 데이터를 사용하여 다양한 언어 간의 전이 학습(Cross-Lingual Transfer Learning)이 가능하게 하였다(Pires *et al.*, 2019). Multilingual BERT는 단일 언어모델의 한계를 극복하고 다양한 언어를 동시에 분석할 수 있는 장점이 있어 다국어 데이터 활용이 필요한 다양한 산업 분야에서 실질적인 가치를 제공할 수 있다고 알려져 있다(Pires *et al.*, 2019).

2.2 다국어 사전학습 언어모델의 언어적 특성에 따른 전이 능력

다국어 사전학습 언어모델의 교차 언어 전이 학습(Cross-Lingual Transfer Learning)에 관한 연구는 다수 수행되었다. Multilingual BERT(Pires *et al.*, 2019)는 언어 간 학습 전이에 관한 성능을 미세 조정과 학습 시 언어가 다른 평가 방법인 Zero-shot Learning 결과를 통해 Multilingual BERT가 언어 간 교차 학습이 가능함을 입증하였다. 그러나 Pires *et al.*(2019)은 SOV(주어, 목적어, 동사) 어순을 가진 언어 집합과 SVO(주어, 동사, 목적어) 어순을 가진 언어 집합 간 학습의 전이 능력이 떨어짐을 밝혀 다국어 모델의 학습에 언어적 성질이 유의미한 영향을 미침을 주장하였다. 이후로 Wu and Dredze(2019), Conneau *et al.*(2020a), Wang *et al.*(2020), Chai *et al.*(2022) 등 여러 연구를 통해 언어적 유사성과 다국어 BERT의 성능 간 관계를 규명하려는 시도가 이어졌으며, Wu and Dredze(2019)는 언어 간 공유되는 Subwords의 수가, Conneau *et al.*(2020a)과 Wang *et al.*(2020) 등은 언어적 동질성이 언어 간 전이에 주요한 영향을 미침을 주장하였다. 특히, Chai *et al.*(2022)은 SOV 어순을 가진 영어 인공 문장을 생성하여 어순과 교차 언어 능력 간 상관관계가 없음을 밝혀냈으며, 다국어 모델의 언어 간 전이 능력은 언어적 특성 중 단어의 조합(Composition)에 가장 큰 영향을 받음을 주장하였다. 그러나 Nikolaev and Padó(2022)가 수행한 종속절 예측 실험 결과에 따르면 SOV 어순을 가진 언어와 SVO 어순을 가진 언어 간 성능 차이는 유의미하게 나타났다. 이처럼 다국어 사전학습 언어모델의 언어 간 전이 능력의 원인은 연구마다 주장이 상이하나, 비슷한 언어적 특성을 가진 언어들 사이의 미세 조정을 통한 전이 능력은 그렇지 않은 경우보다 향상된다는 것이 선행 연구의 공통된 의견이다.

2.3 ABSA

ABSA는 예측하려는 주요 요소(속성어, 속성, 감성어, 극성)에 따라 단일 요소를 예측하는 단일 ABSA와 다수의 요소를 예측하고 이들 간의 관계를 규명하는 복합 ABSA로 분류되어(Zhang *et al.*, 2022), 다양한 도메인에서 다양한 과업의 형태로 다양한 연구가 수행되었다. Essebbbar *et al.*(2021), Pathak *et al.*(2021), Abdelgwad *et al.*(2022) 등은 각각 프랑스어, 힌디어, 아랍어로 구성된 벤치마크 데이터 세트를 사용한 ABSA 연구를 수행하였으며, 특히 Pathak *et al.*(2021)의 연구의 경우 51.22%의 Accuracy를 기록한 모바일 앱 도메인부터 79.77%의 Accuracy를 기록한 영화 도메인까지 4가지 도메인별 성능 평가를 진행해 도메인 간 성능 차이를 확인하였다. 국내에서도 관련 연구가 진행되었는데, Park and Shin(2020)은 SemEval2016(Pontiki *et al.*, 2016) 등의 영어 벤치마크 데이터를 사용하여 개선된 성능의 BERT 구조를 제안하였다. 다국어 사전학습 언어 모델을 사용한 연구로는 한국어, 일본어, 영어 3가지 언어를 사용해 다국어 ABSA 모델을 구축하고 화장품 도메인의 리뷰 데

이터 분석을 진행한 연구가 존재한다(Ji *et al.*, 2023). 그러나 해당 연구는 사전학습 언어모델의 언어적 특성에 따른 언어 간 전이는 고려하지 않았다. ABSA에 관련한 기존 연구를 살펴본 결과 한국어를 포함한 다국어 데이터를 활용한 ABSA 연구의 수는 절대적으로 적어 이와 관련된 체계적인 연구가 필요한 실정이다.

2.4 연구의 차별성

선행연구 검토 결과 언어적 유사성은 다국어 사전학습 언어 모델의 언어 간 전이 능력과 깊은 연관이 있음을 확인하였다. 그러나 검토한 사전학습 언어모델의 언어 간 전이 능력 관련 선행 연구는 전이 능력의 원인을 밝히는 것에 중점을 두어 실험을 진행했기에 미세 조정(Fine-Tuning) 시 주로 단일 언어를 사용하였으며, 이는 서로 다른 특성을 가지는 언어들 동시 사용에 사용하며 학습할 경우 모델 성능 변화에 대한 예측을 제공하기 어렵다는 약점을 가진다. 더욱이 광범위한 도메인과 언어로 수행되고 있는 ABSA 연구에서 여전히 한국어 및 한국어를 포함하는 다국어 데이터를 사용한 연구의 수는 매우 적으며 한국어의 언어적 특성을 고려한 다국어 ABSA 연구가 이루어질 필요가 있다. 따라서 본 연구는 언어적 이질성이 큰 리뷰 데이터를 보유한 경우의 활용 방안에 대한 새로운 ABSA 모델을 제공하고자, 한국어를 포함한 다국어 데이터를 수집하고 언어별 유사도(어순)에 따라 서로 다른 총 7가지의 데이터 세트를 구축하여 언어 간 전이를 고려한 다국어 ABSA 모델을 학습하고 성능 비교를 진행하였다. 이를 통해 영어와 한국어, 일본어 등 서로 언어적 이질성이 큰 리뷰 데이터를 동시에 활용하고자 하는 기업에 AI 학습을 위한 적절한 데이터 세트 구축 방향을 제공하고 이를 활용한 다국어 ABSA 모델과 활용방안을 제안하였다.

3. 데이터

본 연구에서 목표하는 다국어 사전학습 언어모델을 사용한 ABSA 모델을 위해 수집하는 데이터는 두 가지로 구성된다. 첫째로 사전학습 언어모델의 미세 조정을 위한 다국어로 구성된 학습 데이터(Tripadvisor)와 둘째로 미세 조정된 모델을 바탕으로 활용 방안을 제안하기 위한 국내 해외 숙소 플랫폼의 호텔 리뷰 데이터(야놀자)로 구성된다. 모델 미세 조정을 위한 데이터는 언어별로 데이터가 풍부하며, 호텔 속성별로 라벨링이 되어있어야 한다. 종합 여행 서비스 기업인 Tripadvisor(tripadvisor.com)는 <Figure 2>와 같이 세계 각국의 호텔에 대한 소비자 리뷰를 Value(가격), Rooms(객실), Location(장소), Cleanliness(청결도), Service(서비스), Sleep Quality(침대의 퀄리티) 등 6가지 속성으로 세분화해 5점 리커트 척도로 제공 중이기에 본 연구의 학습용 데이터로 선정하였다.

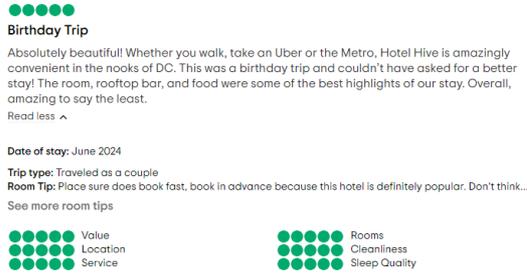


Figure 2. Tripadvisor Review Sample

본 연구에서 AI 모델 학습을 위해 수집하는 언어는 SVO(주어, 동사, 목적어) 어순을 따르는 영어(En)와 SOV(주어, 목적어, 동사)의 어순을 갖는 한국어(Ko)와 일본어(Ja)로 선정했다. 해당 3개 언어는 최종적으로 분석하고자 하는 국내 1위 숙박플랫폼인 야놀자(yanolja.com)의 해외 해외호텔 리뷰 데이터의 대다수를 구성한다. 2023년 코로나-19가 엔데믹으로 전환되면서 국내 해외여행에 대한 수요는 코로나-19 이전인 2019년의 79% 수준까지 회복하였고(KTO, 2024) 이에 대응하고자 국내 주요 숙박 서비스 제공 플랫폼들은 해외 숙소 서비스를 고도화하여 제공하기 시작하였다. 그러나 서비스 기간이 짧아 국내 호텔에 비해 해외호텔 리뷰 수가 빈약하고 속성별 평점을 제공하지 않아 소비자들의 숙소 선택에 충분한 정보가 제공되지 못하고 있다. 따라서 2023년 한 해 동안 약 700만 명의 한국인 해외여행객이 방문한 일본과 75만 명의 여행객이 방문한 대만 그리고 각각 360만, 60만 명이 방문한 베트남, 싱가포르 등 해외여행객이 집중된 동남아(KTO, 2024) 숙소들의 리뷰 분석을 통한 속성별 정보 제공을 목적으로 미세 조정을 위한 학습 데이터 구성 시 한국어, 영어, 일본어 데이터를 Tripadvisor로부터 수집하였다. 2024년 3월 27일부터 2주에 걸쳐 무작위로 호텔을 선정하며 한국어 리뷰 47,134개, 영어 리뷰 22,726개 일본어 리뷰 59,922개를 Tripadvisor로부터 수집하였다. 이 중 6가지 속성 모두 공란인 리뷰 데이터는 소비자의 불성실 응답으로 판단하고 제거하여 최종적으로 한국어 16,303개, 영어 16,180개, 일본어 16,233개로 구성된 학습 데이터 세트를 구축하였다.

구축한 학습 데이터 세트는 다음과 같이 전처리를 수행하였다. 첫째로 사용하려는 언어모델이 받을 수 있는 최대 입력 길이를 넘어서는 데이터 제거 및 모델 학습 시간 단축을 위해 너무 많은 정보를 잃지 않는 선에서 길이가 긴 리뷰를 제거하였다. 토큰나이징 처리 후 평균 리뷰 길이가 가장 긴 일본어 데이터의 75%까지를 포함하는 수준인 길이 155를 기준으로 이보다 긴 리뷰 데이터를 제거했으며 제거 후 학습 데이터의 평균 길이, 리뷰에 명시된 속성의 평균 개수는 <Table 1>과 같다.

Table 1. Train Data Description

	Data	Length mean(std)	Aspect per Review
Ko	14,153	75.25(34.28)	4.51
En	13,434	73.23(28.38)	4.65
Ja	12,269	81.03(34.67)	4.65

다음으로 개수가 적은 레이블의 병합을 통해 수집한 학습 데이터의 불균형을 보완하고 모델 예측 성능 향상을 목적으로 <Figure 3>와 같이 5점 척도와 결측치로 구성된 레이블을 4개의 멀티 레이블로 변환하였다. Tripadvisor에서 제공하는 평점 기준을 근거로 하여 5점과 4점은 ‘긍정(Positive)’으로, 3점은 ‘중립(Neutral)’으로 2점과 1점은 ‘부정(Negative)’으로 변환했으며, 리뷰 속성값이 결측치(nan)라면 ‘해당 없음(None)’으로 레이블을 기입하였다. 즉, 해당 과정을 통해 하나의 리뷰 데이터는 속성에 따라 레이블값이 각기 다른 6개의 데이터로 증폭된다. 마지막으로 증폭된 데이터를 언어별로 7:1:2 비율로 분할 한 후 병합하여 학습, 검증, 평가 데이터 세트를 구축했으며 각 데이터 세트의 언어별 데이터 개수를 <Table 2>로 나타냈다. 구축한 학습, 검증, 평가 데이터는 추후 보조 문장 생성(4.1절)을 통해 최종 데이터 세트로 구축되어 모델에 입력된다.

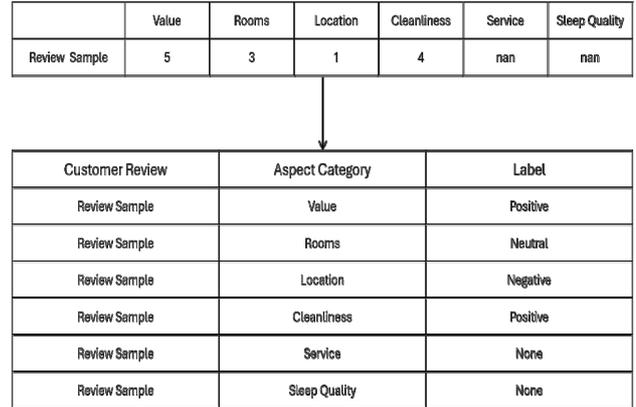


Figure 3. Label Processing Example

Table 2. Train, Validation, Test Set Description

	Korean	English	Japanese	Total
Train Set	59,442	56,422	51,529	167,393
Validation Set	8,407	7,980	7,288	23,675
Test Set	17,069	16,202	14,797	48,068

Tripadvisor에서 수집한 데이터를 바탕으로 미세 조정된 모델을 통해 분석하고자 하는 데이터는 국내 플랫폼인 야놀자에서 제공하는 해외 숙소 리뷰 데이터로 해외여행객이 집중된 대만, 베트남, 싱가포르에서 하나씩, 일본에서 두 개의 호텔을 선별하여 수집하였다. 총 4개국, 5개의 호텔 리뷰 데이터는 모두 2024년 6월 12일에 수집하였으며 레이블 변환 과정을 제외하고 학습 데이터와 동일한 전처리 과정으로 정제하였다. 수집된 리뷰의 언어 감지는 Google Translate API를 활용해 진행하였으며, 정제된 데이터의 기술통계량은 <Table 3>과 같으며 각 호텔 리뷰의 언어 분포는 <Figure 4>와 같다.

Table 3. Target Data Description

	Data Language Info					Length mean(std)
	Total	Ko	En	Ja	Etc	
Japan A Hotel	2,160	1,102	104	949	5	40.57(31.96)
Japan N Hotel	861	613	44	202	2	45.91(36.35)
Taiwan C Hotel	2,340	1,827	362	134	17	42.51(33.67)
Vietnam M Hotel	1,112	938	114	57	3	48.79(36.21)
Singapore B Hotel	1,297	558	681	48	10	36.10(32.34)

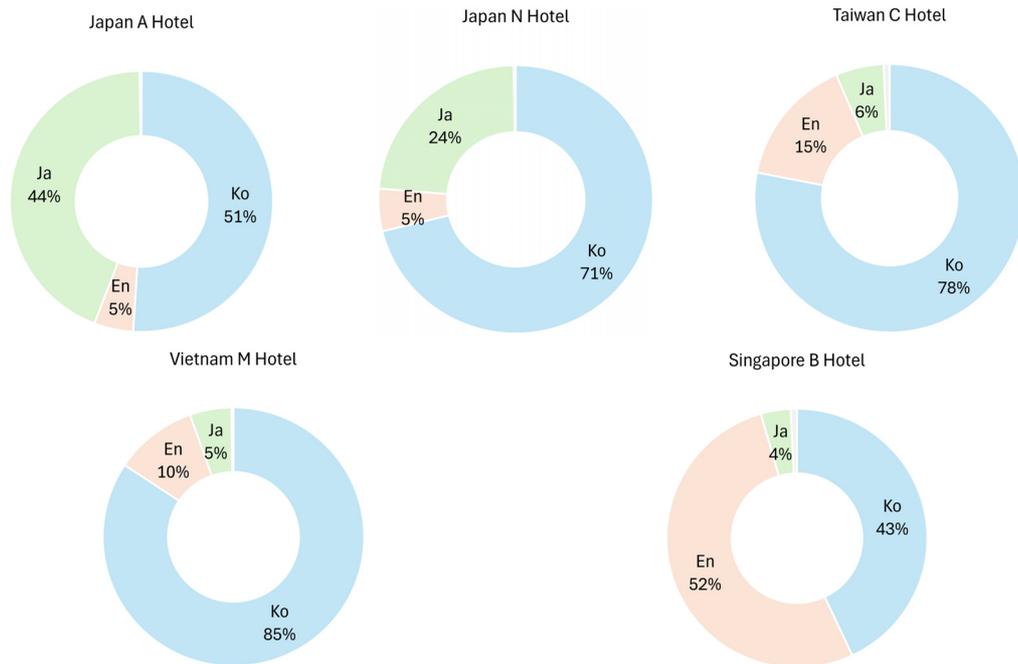


Figure 4. Language Distribution of Target Data

4. 분석 방법론

4.1 사전학습 언어모델을 사용한 ABSA

본 연구에서 수행한 과업은 6개 속성, 4개 레이블을 예측하는 복잡한 과업으로 우수한 예측 성능 달성을 위해 사전학습 언어 모델을 미세 조정(Fine-Tuning)하여 사용하였다. 본 과업에서 사용한 모델은 BERT(Devlin *et al.*, 2019)의 파생 모델인 XLM-RoBERTa(Conneau *et al.*, 2020b)로 동적 마스킹을 적용하여 성능을 향상한 BERT 기반 사전학습 모델인 RoBERTa(Liu *et al.*, 2019)를 100개 국어로 구성된 다국어 데이터로 학습시킨 모델이다. 이는 기존의 우수한 다국어 모델인 Multilingual BERT(Pires *et al.*, 2019)와 비교해 다수의 벤치마크 데이터 세트에서 성능 우위를 보인 모델이다(Conneau *et al.*, 2020b). 따라서 본 과업은 XLM-RoBERTa 모델을 사용하되 <Figure 5>와 같이 기존 국내 선행연구인 Park and Shin(2020)이 제안한 구조인 [CLS] 토큰과 속성에 해당하는 토큰을 동시에 활용하여 분류를 수행하는 모델 최상단 구조를 채택하는 미세 조정을 실시하였

다. 해당 방식은 제안 논문인 Park and Shin(2020) 뿐만 아니라 Ji *et al.*(2023)에서도 [CLS] 토큰만 사용하는 기존 방법론 대비 우수한 성능을 보였다. 수행한 모든 실험의 미세조정 시 옵티마이저(Optimizer)는 AdamW(Loshchilov and Hutter, 2019)를 사용하였고, 학습률(Learning rate)과 배치 사이즈(Batch size)는 각각 $2e-5$, 16으로 설정하였다. 에포크(Epoch)는 10으로 설정하였으며 과적합 방지를 위해 에포크마다 검증 데이터 세트 평가를 진행하여 가장 우수한 결과를 보인 에포크의 모델을 각 실험의 최종 모델로 선택하였다.

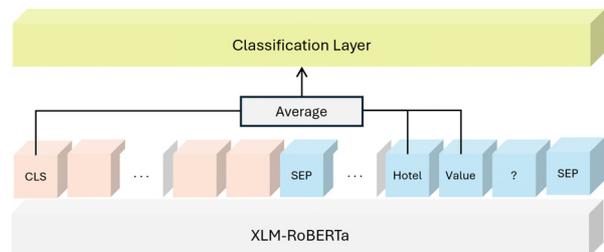


Figure 5. Model Top Layer Architecture

Table 4. Data Samples after Constructed Auxiliary Sentence

Review	Aspect	QA
Aubrey was fantastic...	Cleanliness	What do you think of the Hotel Cleanliness?
Aubrey was fantastic...	Value	What do you think of the Hotel Value?
전반적으로 다 좋았습니다...	Location	호텔 장소에 대한 만족도는 얼마입니까?
전반적으로 다 좋았습니다...	Service	호텔 서비스에 대한 만족도는 얼마입니까?
今回親子2人で1泊利用しました...	Rooms	ホテルの客室に対する満足度はどのくらいですか?
今回親子2人で1泊利用しました...	Sleep Quality	ホテルの寝心地に対する満足度はどのくらいですか?

본 과업은 다양한 ABSA 과업 중 극성에 해당하는 Sentiment Polarity를 예측하는 ASC(Aspect Sentiment Classification) 과업에 속하지만, 예측 레이블에 ‘해당 없음(None)’을 추가하여 속성에 해당하는 Aspect Category도 간접적으로 검출할 수 있어 ACD(Aspect Category Detection) 과업의 성격도 띈다. 본 연구에서 사용한 방식은 Sun *et al.*(2019)가 제안한 방법론으로 동시에 두 문장을 입력받는 BERT-pair 모델을 기반으로 미세조정 시 속성 정보를 포함하는 보조 문장을 생성하여 소비자 리뷰와 함께 입력데이터로 사용하였다. 이를 통해 학습된 분류기는 속성어(Aspect Term)를 통해 속성이 명시적으로 제공되는 ‘명시적 속성’ 뿐만 아니라 속성이 명시적으로 제공되지 않는 ‘암시적 속성’까지도 판별할 수 있어 실제 상품 리뷰 데이터 분석에 유리한 방법이다(Park and Shin, 2019). 생성되는 보조 문장의 형태로는 문장 형태로 질의하는 QA(Question Answering) 형식과 속성 키워드로 제공하는 NLI(Natural Language Inference)방식이 존재하며 정답 레이블의 형태에 따라 ‘Yes’/‘No’의 이진 형태(B)와 ‘긍정(Positive)’/‘부정(Negative)’/‘중립(Neutral)’/‘해당 없음(None)’ 등의 다중 레이블(M) 형태로 나누어진다(Sun *et al.*, 2019). 즉, QA-M, QA-B, NLI-M, NLI-B의 4가지 형태가 존재한다. 본 연구에서는 Ji *et al.*(2023)에서 가장 우수한 성능을 보인 QA-M 방식을 선택하였다. <Table 4>는 QA-M 방식으로 처리한 데이터 샘플을 나타낸다.

4.2 데이터 불균형 처리

본 연구에서 사용한 학습 데이터는 레이블 간 불균형이 심한 데이터로 소수 범주에 대한 분류 성능이 간과되지 않게 하는 것이 중요하다. 따라서 학습 데이터 샘플링을 통해 데이터 자체의 불균형을 보완하고 Precision, Recall, macro F1, Balanced Accuracy 등 불균형 데이터 평가에 적합한 지표들을 Accuracy와 함께 사용하여 모델 성능 평가를 진행하였다. <Table 5>는 4개 레이블의 혼동행렬을 나타내고 있으며 사용한 평가 지표는 식 (1)~(6)과 같이 혼동행렬을 기반으로 산출된다.

Table 5. Confusion Matrix for 4 Labels

	Predicted A	Predicted B	Predicted C	Predicted D
Actual A	P_{AA}	P_{AB}	P_{AC}	P_{AD}
Actual B	P_{BA}	P_{BB}	P_{BC}	P_{BD}
Actual C	P_{CA}	P_{CB}	P_{CC}	P_{CD}
Actual D	P_{DA}	P_{DB}	P_{DC}	P_{DD}

$$Accuracy = \frac{P_{AA} + P_{BB} + P_{CC} + P_{DD}}{\sum_i \sum_j P_{ij}} \quad (1)$$

$$Precision_A = \frac{P_{AA}}{P_{AA} + P_{BA} + P_{CA} + P_{DA}} \quad (2)$$

$$Recall_A = \frac{P_{AA}}{P_{AA} + P_{AB} + P_{AC} + P_{AD}} \quad (3)$$

$$F1_A = \frac{2 \times Precision_A \times Recall_A}{Precision_A + Recall_A} \quad (4)$$

$$macro\ F1 = \frac{F1_A + F1_B + F1_C + F1_D}{4} \quad (5)$$

$$Balanced\ Accuracy = \frac{Recall_A + Recall_B + Recall_C + Recall_D}{4} \quad (6)$$

Table 6. Train Set Number of Label per Language

	Label			
	None	Neutral	Negative	Positive
Ko	14,801	4,600	2,082	37,959
En	12,659	3,189	2,383	38,191
Ja	11,629	5,095	1,237	33,568

Table 7. Validation Set Evaluation Result per Language

	Sampling	Data	Validation		
			Accuracy	macro F1	Balanced accuracy
Ko	1.0	59,442	0.761151	0.596653	0.601800
En	1.0	56,422	0.834085	0.684702	0.685127
Ja	1.0	51,529	0.717206	0.545427	0.543341
Ko	0.5	40,463	0.739860	0.591841	0.611243
En	0.5	37,327	0.816040	0.692617	0.718364
Ja	0.5	34,745	0.685510	0.542615	0.566313
Ko	0.3	32,871	0.712739	0.620314	0.579182
En	0.3	29,688	0.788596	0.673958	0.716447
Ja	0.3	28,031	0.634330	0.525193	0.576141

데이터 샘플링은 언어별로 진행하였다. 학습 데이터 세트의 언어별 레이블 개수는 <Table 6>과 같으며 다수 범주인 긍정 레이블 데이터를 무작위로 추출하는 랜덤 언더샘플링을 사용

하였다. 샘플링 시 원본 데이터(1.0), 0.5배수, 0.3배수 세 가지 경우로 언어별 모델을 구축하여 검증 데이터 평가를 진행하였으며, 비율별로 샘플링한 데이터 개수와 미세 조정된 모델의 평가 결과를 <Table 7>로 나타냈다. 최종적으로 모든 언어에 대해 전체적인 분류 성능과 소수 범주에 대한 분류 성능 사이의 균형을 유지하는 샘플링 비율인 0.5배수를 선택하였다.

5. 실험 결과

5.1 모델 예측 성능

다국어 ABSA 모델 구축을 위해 SOV(주어, 목적어, 동사) 어순을 가진 한국어와 일본어, SVO(주어, 동사, 목적어) 어순을 가져 한국어, 일본어와의 이질성이 큰 영어 데이터의 조합에 따라 총 7가지 학습 데이터 세트를 구성하였다. 즉, 한국어(Ko), 영어(En), 일본어(Ja) 단일어 데이터로 학습한 3종류의 모델과 한국어+일본어(KJ), 한국어+영어(KE), 일본어+영어(JE)와 같이 두 언어씩 학습 데이터로 사용한 3가지 모델, 그리고 모든 언어를 학습 데이터로 사용한 모델(KJE)까지 총 7가지 모델을 학습하였다. 미세 조정된 7가지 모델에 대해 언어별 성능 평가를 진행하여 언어적 동질성과 성능 간 상관관계를 판별하고자 하였다. 해당 실험의 결과는 <Table 8>~<Table 10>까지로 나타냈으며 순서대로 Accuracy, macro F1, Balanced Accuracy를 평가 지표로 사용한 결과이다.

Table 8. Test Set Evaluation Results(Acc)

Fine-Tuning Data	Evaluation(Accuracy)		
	Ko	Ja	En
Ko	0.735954	0.506995	0.577460
Ja	0.587615	0.677908	0.644612
En	0.590896	0.525917	0.816072
Ko+Ja	0.734079	0.686355	0.651031
Ko+En	0.733728	0.499223	0.810455
Ja+En	0.600211	0.687910	0.813356
Ko+Ja+En	0.740817	0.690140	0.818171

Table 9. Test Set Evaluation Results(macro F1)

Fine-Tuning Data	Evaluation(macro F1)		
	Ko	Ja	En
Ko	0.598051	0.379340	0.498883
Ja	0.460560	0.531846	0.489483
En	0.442540	0.376746	0.691198
Ko+Ja	0.604673	0.548506	0.531836
Ko+En	0.600783	0.386699	0.691162
Ja+En	0.470737	0.557255	0.687421
Ko+Ja+En	0.611705	0.558409	0.693874

Table 10. Test Set Evaluation Results(Balanced Accuracy)

Fine-Tuning Data	Evaluation(Balanced Accuracy)		
	Ko	Ja	En
Ko	0.617617	0.405094	0.526468
Ja	0.465966	0.551046	0.496913
En	0.443909	0.364489	0.716965
Ko+Ja	0.630983	0.566728	0.539563
Ko+En	0.615400	0.414770	0.717765
Ja+En	0.477372	0.571658	0.711981
Ko+Ja+En	0.636265	0.574340	0.720549

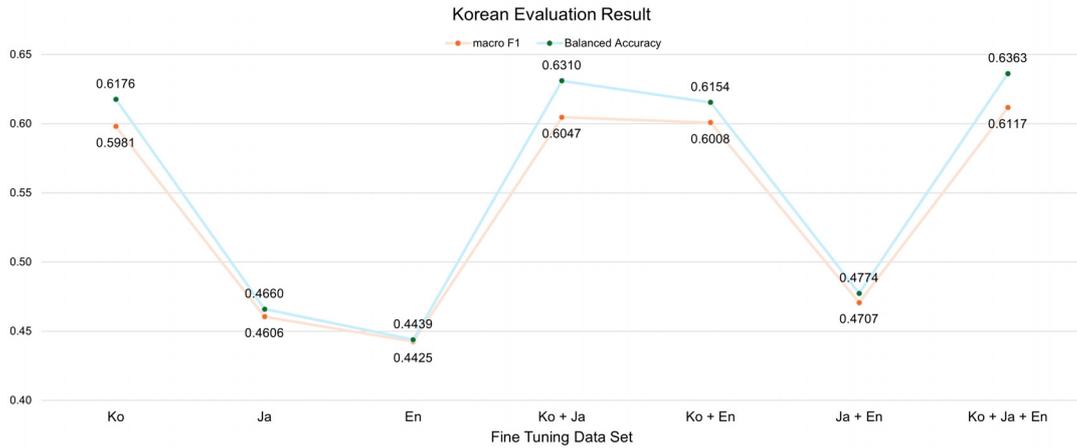
<Table 8> ~ <Table 10>에 따르면 모든 언어, 모든 평가 지표에 대해 3가지 언어를 모두 학습 데이터로 사용하여 미세 조정을 실시한 모델(KJE)이 가장 좋은 성능을 보였으며 해당 모델은 평가 데이터 세트에서 영어, 한국어, 일본어 순으로 좋은 성능을 보였음을 확인할 수 있다. 또한 모든 모델은 한 번도 보지 못한 언어에 대한 평가(Zero-Shot)보다 그렇지 않은 경우가 좋은 성능을 보였다.

<Figure 6>은 불균형 데이터 평가에 적합한 결과인 <Table 9>와 <Table 10>을 언어별로 가시화한 그래프이며 <Figure 6(a)> ~ <Figure 6(c)>를 통해 언어별 성능을 자세히 살펴볼 수 있다. <Figure 6(a)>에 따르면 한국어의 경우 macro F1은 KJE, KJ, KE, Ko 순으로 좋은 결과를, Balanced Accuracy는 KJE, KJ, Ko, KE 순으로 좋은 결과를 나타낸다. <Figure 6(b)>에 따르면 일본어는 두 평가 지표 모두에 대해서 KJE, JE, KJ, Ja의 순서로 우수한 성능을 기록했으며, 영어는 <Figure 6(c)>와 <Table 9>을 통해 macro F1은 KJE, En, KE, JE 순으로, Balanced Accuracy는 KJE, KE, En, JE 순서로 좋은 결과를 나타낼 수 있었다. 해당 결과를 토대로 언어 간 교호작용을 판별하면 한국어는 일본어와 같이 학습한 경우가 영어와 학습한 경우보다 좋은 성능을 나타내 학습 데이터 구성의 언어적 동질성과 모델 성능 간 양의 상관관계를 보임이 확인되었다. 반면에 일본어는 한국어보다 영어와 같이 학습한 경우가 더 좋은 결과를 보여 한국어와 반대의 경우를 보이며, 영어는 두 평가 지표에서 일관되지 못한 결과를 보여주어, 언어적 동질성과 성능 간 일관된 추세를 확인할 수 없었다. 이러한 결과는 <Table 8>을 근거로 해도 마찬가지인데, KJE 모델을 제외하고 단일어 모델과 두 언어를 사용한 모델을 비교할 시 한국어의 경우 일본어를 같이 사용하는 경우가 영어와 같이 사용해 학습한 경우보다 나은 성능을 보여주었으며 영어는 한국어와 일본어를 같이 구성해 학습하는 경우보다 영어만 따로 사용한 편이 더 좋은 성능을 보여주었다. 그러나 일본어는 이질적인 언어 데이터와 같이 구성된 편이 더 좋은 성능을 보여주어 한국어와 영어에서 확인할 수 있는 언어적 특성과 모델 교차 언어 학습 간 결과의 방향성이 서로 반대의 경우를 나타낸다. 무엇보다 세 언어의 평가 결과 모두 모든 지표에서 언어적 특성과 관계

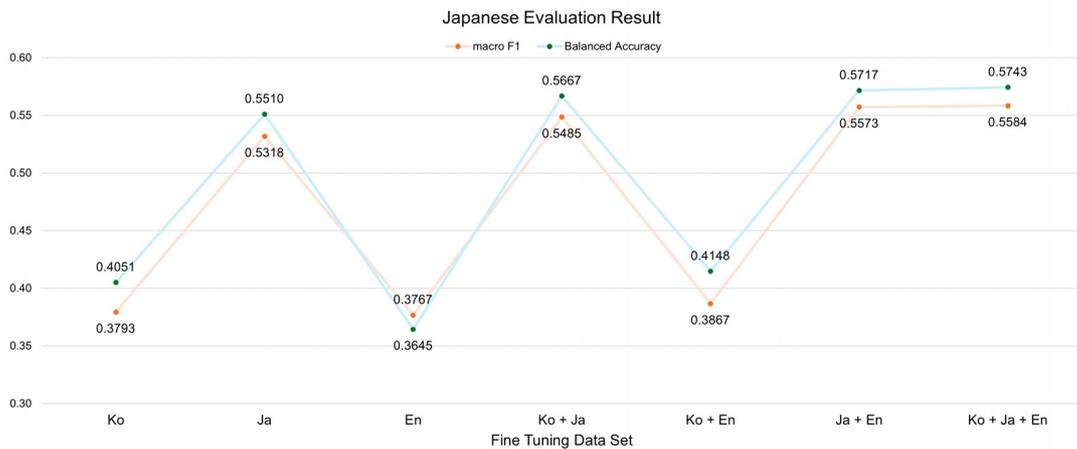
없이 모든 데이터를 사용하여 미세 조정을 진행한 경우가 가장 우수한 성능을 보여주었다. 따라서 본 실험을 통해 다국어 리뷰 데이터 구성 시 언어적 이질성이 존재하더라도 같은 도메인의 데이터라면 함께 사용하여 절대적인 학습 데이터양을

늘리는 편이 더 좋은 모델 분류 성능을 끌어낼 수 있다.

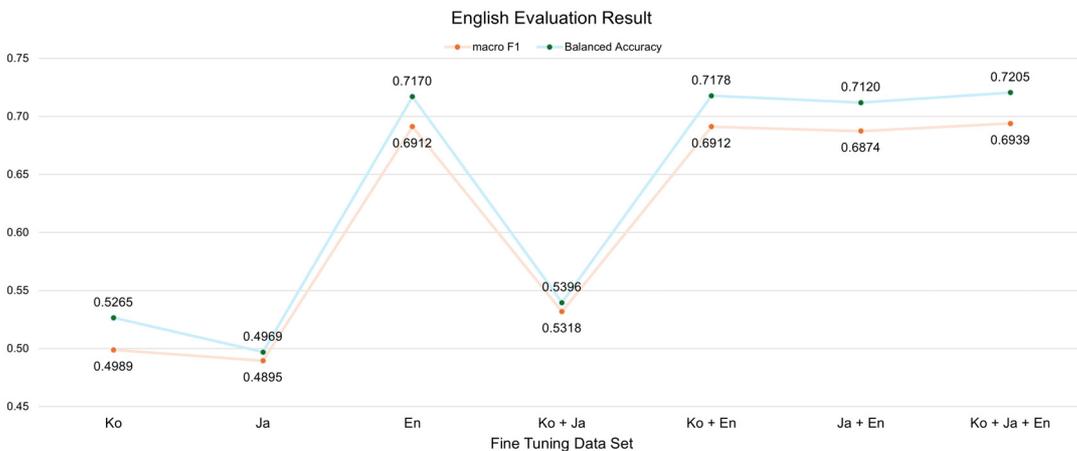
실험 결과를 토대로 본 연구에서 제안한 최종 모델은 모든 언어에 대해 가장 우수한 평가 결과를 보인 세 언어를 모두 사용하여 미세 조정된 모델(KJE)로 해당 모델의 언어별, 속성별



(a) Korean Evaluation Result



(b) Japanese Evaluation Result



(c) English Evaluation Result

Figure 6. Evaluation Result per Different Language Data Set

평가 결과를 <Table 11>로 나타냈다.

<Table 11>에 따르면 제안 모델은 Accuracy 0.7513, macro F1 0.6279, Balanced Accuracy 0.6512의 우수한 성능을 보였으며, 언어별로는 영어, 한국어, 일본어 순으로 좋은 분류 성능을, 속성별로는 Service, Rooms, Cleanliness, Sleep Quality, Location, Value 순서로 우수한 성능을 보였다. 다만 Service 속성의 경우 대부분의 레이블이 ‘긍정(Positive)’으로 답하여(F1 0.9449) 불균형이 심해 성능의 과대평가가 나타났으며 ‘해당 없음(None)’의 경우 모델 미세 조정 시 적은 데이터 개수(한국어 11, 영어 41, 일본어 13)로 인해 모델이 학습하지 못한 모습이 확인되었다. <Figure 7>은 평가 데이터 세트의 정규화된 혼

동행렬로, 이에 따르면 ‘해당 없음(None)’, ‘긍정(Positive)’, ‘부정(Negative)’, ‘중립(Neutral)’이 각각 Recall 88%, 76%, 58%, 39%의 분류 성능을 기록하였다. 미세 조정 시 ‘부정(Negative)’은 ‘중립(Neutral)’의 40% 수준으로 더 적은 데이터를 가지고 있었으나 실제 분류 성능은 20%가량 높게 측정되어 학습한 모델이 극성을 띠지 않는 리뷰 분류에 어려움을 가짐이 확인되었으며, 두 레이블 모두 사전에 불균형 처리를 진행하였음에도 절대적인 데이터의 양이 부족해 상대적인 성능 저하가 나타남이 확인되었다. 그러나 레이블이 풍부한 ‘해당 없음(None)’과 ‘긍정(Positive)’의 분류 성능은 모두 Recall 75% 이상으로 우수한 모습을 보여주었다.

Table 11. Proposed Model Evaluation Result

Total									
Class	Precision			Recall			Support		
None	0.598138			0.881593			11224		
Negative	0.585952			0.580232			1639		
Neutral	0.395137			0.386486			3700		
Positive	0.906936			0.756610			31505		
							48068		
					Accuracy		0.751290		
					macro F1		0.627885		
					Balanced Acc		0.651230		
Class	Korean			Japanese			English		
	Precision	Recall	Support	Precision	Recall	Support	Precision	Recall	Support
None	0.6054	0.9181	4250	0.5103	0.8086	3339	0.6849	0.9059	3635
Negative	0.5471	0.5334	598	0.4882	0.4073	356	0.6554	0.7109	685
Neutral	0.3719	0.3648	1321	0.3904	0.3738	1463	0.4349	0.4377	916
Positive	0.9081	0.7286	10900	0.8733	0.7075	9639	0.9327	0.8275	10966
Accuracy	0.740817			0.690140			0.818171		
macro F1	0.611705			0.558409			0.693874		
Balanced Acc	0.636265			0.574340			0.720549		
Class	Value			Service			Location		
	Precision	Recall	Support	Precision	Recall	Support	Precision	Recall	Support
None	0.5011	0.7798	1644	0.0000	0.0000	19	0.4965	0.7963	1650
Negative	0.6269	0.5734	422	0.6317	0.6925	322	0.3360	0.2818	149
Neutral	0.3884	0.3482	1005	0.4234	0.5139	683	0.3405	0.3275	577
Positive	0.8426	0.7098	4911	0.9557	0.9345	7117	0.8838	0.7384	5773
Accuracy	0.671511			0.887483			0.712725		
macro F1	0.586753			0.517510			0.514217		
Balanced Acc	0.602840			0.535245			0.536059		
Class	Sleep Quality			Rooms			Cleanliness		
	Precision	Recall	Support	Precision	Recall	Support	Precision	Recall	Support
None	0.6388	0.9149	2705	0.6531	0.9304	2546	0.6396	0.9229	2660
Negative	0.4979	0.5401	224	0.6333	0.6263	273	0.6178	0.6104	249
Neutral	0.3816	0.3186	521	0.4220	0.4139	517	0.4081	0.4030	397
Positive	0.8914	0.6722	4482	0.9184	0.6992	4476	0.9211	0.6940	4746
Accuracy	0.728064			0.753200			0.752732		
macro F1	0.596087			0.667516			0.641743		
Balanced Acc	0.611503			0.652338			0.657614		

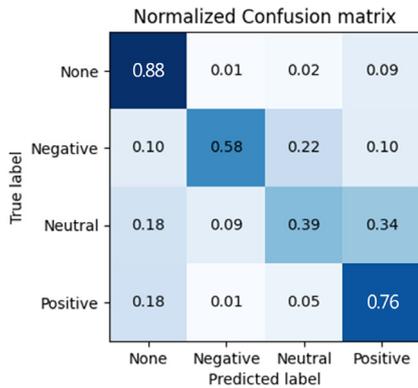


Figure 7. Model Prediction Normalized Confusion Matrix

5.2 분석 결과

최종 제안 모델(KJE)을 통해 야놀자에서 수집한 호텔 리뷰의 속성별 극성 분석을 실시하였다. 수집한 5개의 호텔 중 <Figure 4>를 참고하여 한국어 중심(Vietnam M Hotel), 영어 중심(Singapore B Hotel), 일본어 중심(Japan A Hotel)으로 리뷰가 구성된 3가지 호텔을 선별하여 실시하였으며, 호텔별 리뷰 데이터는 학습 데이터와 동일하게 QA-M 형식으로 보조 문장을 생성하였다. 보조 문장 생성 시 한국어, 일본어는 동일한 언어로, 영어를 포함한 기타 언어 리뷰는 영어로 진행하였다. 모델의 분류 결과는 <Table 12>와 같으며, 속성별로 ‘해당 없음 (None)’을 제외하고 백분율로 환산한 수치를 같이 표기했으며,

Table 12. Target Data(Rating) Prediction(%) Results

Japan A Hotel (4.2/5.0)					
Value		Service		Location	
Positive	1,082(70%)	Positive	1,820(84%)	Positive	1,246(83%)
Neutral	397(26%)	Neutral	274(13%)	Neutral	204(14%)
Negative	61(4%)	Negative	66(3%)	Negative	44(3%)
None	620	None	0	None	666
Rooms		Sleep Quality		Cleanliness	
Positive	1,176(79%)	Positive	1,242(87%)	Positive	1,295(85%)
Neutral	201(13%)	Neutral	138(10%)	Neutral	108(7%)
Negative	117(8%)	Negative	43(3%)	Negative	119(8%)
None	666	None	737	None	638
Vietnam M Hotel (4.8/5.0)					
Value		Service		Location	
Positive	687(91%)	Positive	1,091(98%)	Positive	688(96%)
Neutral	56(7%)	Neutral	15(1%)	Neutral	23(3%)
Negative	15(2%)	Negative	6(1%)	Negative	7(1%)
None	344	None	0	None	394
Rooms		Sleep Quality		Cleanliness	
Positive	483(92%)	Positive	484(94%)	Positive	500(95%)
Neutral	33(6%)	Neutral	23(5%)	Neutral	20(4%)
Negative	9(2%)	Negative	6(1%)	Negative	4(1%)
None	587	None	599	None	588
Singapore B Hotel (4.0/5.0)					
Value		Service		Location	
Positive	660(63%)	Positive	989(76%)	Positive	880(88%)
Neutral	236(22%)	Neutral	202(16%)	Neutral	91(9%)
Negative	156(15%)	Negative	106(8%)	Negative	31(3%)
None	245	None	0	None	295
Rooms		Sleep Quality		Cleanliness	
Positive	573(59%)	Positive	677(71%)	Positive	820(83%)
Neutral	213(22%)	Neutral	164(17%)	Neutral	93(9%)
Negative	187(19%)	Negative	113(12%)	Negative	77(8%)
None	324	None	343	None	307



Figure 8. Proposed Review Service Design

수집일을 기준으로 5점 만점으로 제공 중인 호텔별 평점을 호텔명 옆에 명시하여 속성별 분류 결과와 실제 호텔 평점 간 유사성을 확인할 수 있게 하였다.

<Figure 8>은 분석 대상 중 Japan A Hotel에 대해 본 연구에서 최종 제안하는 속성별 리뷰의 예시이다. 소비자가 추구하는 가치가 다양해짐에 따라 호텔 선택 시에 추구하는 가치의 형태도 다양해지고 있으며(Lee & Chung, 2011), 호텔 속성별 정보는 여행자가 호텔을 평가하고 선택하는 중요한 기준으로 작용한다(Knutson, 1988). 따라서 본 연구에서는 구축한 AI 모델을 통해 속성별 호텔 리뷰를 제안하며, 이는 본 연구에서 구축한 모델이 강점을 가지는 속성의 검출과 긍정 리뷰 판별 능력에 기반하여 속성별 긍정 비율을 제공하는 안으로 설정하였다. 이를 통해 플랫폼에서 제공 중인 평점 비율을 대체하여 서비스 제공 기업 별도의 속성별 질의를 통한 데이터 수집 과정 없이 소비자에게 호텔 선택 시 필요한 고품질 정보 제공이 가능할 것으로 기대된다.

6. 결론

본 연구는 다음과 같은 기여를 하였다. 첫째, 다국어 사전학습 언어모델을 사용한 ABSA 방법론 구축을 목표로 호텔 도메인의 소비자 리뷰 데이터를 Tripadvisor로부터 한국어, 영어, 일본어의 언어별로 수집한 후 언어적 특성에 따른 7가지 데이터 세트 구성과 이를 기반으로 한 7가지의 모델 평가를 통해 언어적 성질에 따른 언어 간 전이 능력 간의 상관관계가 실제 다국어 모델 운용 시 유의미한 영향을 미치지 않는다는 결과를 도출하였다. 이는 기존 선행 연구들(Pires *et al.*, 2019; Wu and Dredze, 2019; Conneau *et al.*, 2020a; Wang *et al.*, 2020; Chai *et al.*, 2022; Nikolaev and Padó, 2022)이 일관적으로 주장하는 언어적 특성과 다국어 사전학습 언어모델의 전이 성능 간 상관관계를 실제 다국어 리뷰 분석에 적용하기 위한 한층 더 고차원의 연구로서 의의가 있다. 둘째, 다국어 모델 운용 시 언어적 특성에 상관없이 동일한 도메인이라면 다국어로 구성된 데이터 세트를 구성하여 학습 데이터의 양을 늘려 하나의 모델로 구축하는 편이 더 좋은 성능을 보임을 확인하였다. 이를 통해 실제 다국어 리뷰를 수집하고 운용하는 기업들의 AI 학습 데이터 관리 및 모델 운용에 실질적인 도움을 제공할 수 있을 것으로 기대된다. 셋째, 본 연구는 구축한 다국어 ABSA 모델을 호텔 도메인의 실제 소비자 리뷰 데이터 분석에 적용하여 학습 데이터 구축부터 모델 활용까지 속성 기반 감성분석의 한 가지 응용 사례로 제안하여 본 연구를 비롯한 다수의 ABSA 관련 연구(Park and Shin, 2020; Essebbat *et al.*, 2021; Pathak *et al.*, 2021;

Abdelgwad *et al.*, 2022)들이 이룬 연구에 그치지 않고 실제 기업 및 조직의 비즈니스에 충분히 활용될 수 있음을 주장하였다.

본 연구는 다음과 같은 한계점을 가진다. 첫째로 미세 조정에서 사용한 데이터의 불균형 문제로 인해 일부 속성의 극성 분류에 어려움이 존재한다. 이는 수집한 데이터 자체의 불균형에서 기인한 것으로 실험 과정에서 데이터 샘플링, 다양한 평가 지표 사용 등을 통해 극복하고자 하였지만 여전히 분류 결과의 편향이 존재하였다. 추후 실험을 통해서 데이터 추가 수집, 샘플링 등의 데이터 측면의 보완과 손실함수 변경, 모델 구조 변경 등의 모델 측면의 보완을 복합적으로 고려하여 개선할 필요가 존재한다. 둘째로 다국어 데이터 구성 시 아시아권 기업에 초점을 두어 한국어와 영어, 일본어의 3가지 데이터를 수집하였지만, 이는 반대로 아시아권을 벗어난 기업에 일반화하기 어렵다는 단점을 가진다. 빅데이터 시대에서 E-commerce의 발달과 상품, 서비스 등에 존재하던 국가 간 장벽이 사라진 현대의 세계화로 인해 글로벌적인 리뷰 수집 및 운용이 중요해진 만큼 더 다양한 언어로 구성된 데이터에 대한 연구를 추가 수행하여 보다 일반화된 모델을 제안할 필요가 있다. 마지막으로 본 연구 결과는 다른 도메인에 적용하기 어렵다는 단점이 존재한다. ABSA는 Pathak *et al.*(2021)의 연구 결과와 같이 도메인에 따라 학습 능력이 현저히 달라지는 특징이 존재하는데, 이에 따라 여러 논문(Zhang *et al.*, 2022; Kandhro *et al.*, 2024)에서 범-언어적(Cross-Lingual)뿐만 아니라 범-도메인(Cross-Domain)적인 ABSA 모델 구축의 필요성을 제시하고 있다. 본 연구도 이와 같은 ABSA의 근본적인 약점을 가지고 있으므로 추후 연구를 통해 다양한 도메인의 데이터를 동시에 활용해 타 도메인에서도 일반화가 가능한 다국어 모델 구축이 요구된다.

참고문헌

- Abdelgwad, M. M., Soliman, T. H. A., and Taloba, A. I. (2022), Arabic Aspect Sentiment Polarity Classification Using Bert, *Journal of Big Data*, **9**, 115
- Araque, O., Zhu, G., and Iglesias, C. A. (2019), A Semantic Similarity-Based Perspective of Affect Lexicons for Sentiment Analysis, *Knowledge-Based Systems*, **165**, 346-359.
- Chai, Y., Liang, Y., and Duan, N. (2022), Cross-lingual Ability of Multilingual Masked Language Models: A Study of Language Structure, *60th Annual Meeting of the Association for Computational Linguistics*, **1**, 4702-4712.
- Conneau, A., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020b), Unsupervised Cross-

- Lingual Representation Learning at Scale, *58th Annual Meeting of the Association for Computational Linguistics*, **1**, 8440-8451.
- Conneau, A., Wu, S., Li, H., Zettlemoyer, L., and Stoyanov, V. (2020a), Emerging Cross-lingual Structure in Pretrained Language Models, *58th Annual Meeting of the Association for Computational Linguistics*, **1**, 6022-6034.
- CUK(Consumers Union of Korea) (2022), Survey of Consumer Online Shopping Reviews and Perceptions.
- Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2019), Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding, *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171-4186.
- Do, H. H., Prasad, P., Maag, A., and Alsadoon, A. (2019), Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review, *Expert Systems With Applications*, **118**, 272-299.
- Essebbar, A., Kane, B., Guinaudeau, O., Chiesa, V., Quénel, I., and Chau, S. (2021), Aspect Based Sentiment Analysis Using French Pre-trained Models, *13th International Conference on Agents and Artificial Intelligence 2021*, **1**, 519-525.
- Hassani, H., Beneki, C., Unger S., Mazinani, M. T., and Yeganegi, M. R. (2020), Text Mining in Big Data Analytics, *Big Data and Cognitive Computing*, **4**(1), 1-34.
- Hua, Y. C., Denny, P., Taskova, K., and Wicker, J. (2023), A Systematic Review of Aspect-Based Sentiment Analysis(ABSA): Domains, Methods, and Trends, arXiv preprint arXiv:2311.10777.
- Ji, S. Y., Lee, S. Y., Choi, D. W., and Kang, K. H. (2023), A Study on The Aspect-Based Sentiment Analysis of Multilingual Customer Reviews, *The Korean Journal of Applied Statistics*, **36**(6), 515-528.
- Kandhro, I. A., All, F., Uddln, M., Kehar, A., and Manlckam, S. (2024), Exploring Aspect-Based Sentiment Analysis: An In-Depth Review of Methods and Prospects for Advancement, *Knowledge and Information Systems*, **66**, 3639-3669.
- Knutson, B. J. (1988), Frequent Travelers: Making Them Happy and Bringing Them Back, *Cornell Hotel and Restaurant Administration Quarterly*, **29**(1), 82-87.
- KTO(Korea Tourism Organization) (2024), Statistics of National International Tourist Arrivals by Top Destinations.
- Lee, C. E. and Chung, K. Y. (2011), Differences of Determinant Attributes of Hotel Selection by Segmentation of Consumption Value: Focused on Foreign Hotel Customer's of First Grade Hotels in Seoul, *Korean Journal of Hospitality and Tourism*, **20**(4), 33-56.
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov V. (2019), Roberta: A Robustly Optimized Bert Pretraining Approach, arXiv preprint arXiv:1907.11692.
- Lo, S. L., Cambria, E., Chiong, R., and Cornforth, D. (2017), Multilingual Sentiment Analysis: From Formal to Informal and Scarce Resource Languages, *Artificial Intelligence Review*, **48**(4), 499-527.
- Loshchilov, I. and Hutter, F. (2019), Decoupled Weight Decay Regularization, *7th International Conference on Learning Representations*.
- Nikolaev, D. and Padó, S. (2022), Word-Order Typology in Multilingual Bert: A Case Study in Subordinate-Clause Detection, *4th Workshop on Computational Typology and Multilingual NLP*, 11-21.
- Park, H. J. and Shin, K. S. (2020), Aspect-Based Sentiment Analysis Using Bert: Developing Aspect Category Sentiment Classification Models, *Journal of Intelligence and Information Systems*, **26**(4), 1-25.
- Pathak, A., Kumar, S., Roy, P. P., and Kim, B. G. (2021), Aspect-Based Sentiment Analysis in Hindi Language by Ensembling Pre-Trained mBert Models, *Electronics 2021*, **10**(21), 2641.
- Peng, H., Ma, Y., Li, Y., and Cambria, E. (2018), Learning Multi-Grained Aspect Target Sequence for Chinese Sentiment Analysis, *Knowledge-Based Systems*, **148**, 167-176.
- Pires, T., Schlinger, E., and Garrette, D. (2019), How Multilingual is Multilingual Bert?, *57th Annual Meeting of the Association for Computational Linguistics*, **1**, 4996-5001.
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., Clercq, O. D., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S. M., and Eryigit, G. (2016), Semeval-2016 task 5: Aspect Based Sentiment Analysis, *10th International Workshop on Semantic Evaluation 2016*, 19-30.
- Sun, C., Huang, L., and Qiu, X. (2019), Utilizing Bert for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence, arXiv preprint arXiv:1903.09588.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkorett, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017), Attention Is All You Need, *31st Conference on Neural Information Processing Systems 2017*, 6000-6010.
- Wang, Z., Mayhew, S., and Roth, D. (2020), Cross-lingual Ability of Multilingual Bert: An Empirical Study, *International Conference on Learning Representations 2020*.
- Wu, S. and Dredze, M. (2019), Beto, Bentz, Becas: The Surprising Cross-Lingual Effectiveness of Bert, *9th International Joint Conference on Natural Language Processing*, 833-844.
- Zhang, W., He, R., Peng, H., Bing, L., and Lam, W. (2021), Cross-Lingual Aspect-Based Sentiment Analysis with Aspect Term Code-Switching, *Conference of the Empirical Methods in Natural Language Processing 2021*, 9220-9230.
- Zhang, W., Li, X., Deng, Y., Bing, L., and Lam, W. (2022), A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges, *IEEE Transactions on Knowledge and Data Engineering*, **35**(11), 11019-11038.

저자소개

장호준: 성균관대학교 시스템경영공학과에서 학사학위를 취득하고, 현재 산업공학과 석사과정 재학 중이다. 관심분야는 데이터 분석이다.

전다원: 경기대학교에서 2022년 응용통계학과, 산업경영공학과 학사학위를 취득하고, 현재 성균관대학교 산업공학과 석사과정 재학 중이다. 관심분야는 데이터 분석, 비즈니스 애널리틱스다.

이희상: 서울대학교 산업공학과에서 학사학위와 석사학위를 취득하고 Georgia Tech에서 Industrial & Systems Engineering 박사학위를 취득하였다. KT 선임연구원, 한국외국어대학교 조교수/부교수를 역임하고 2004년부터 성균관대학교 시스템경영공학과에서 교수로 재직 중이다. 연구분야는 경영과학, 비즈니스 애널리틱스, 기술경영이다.