

사전 학습 언어 모델을 활용한 감정 말뭉치 구축 연구

장연지⁰, 비립¹, 강예지¹, 강혜린¹, 박서윤¹, 김한샘[†]

국립국어원, 연세대학교 언어정보연구원^{1†}

yeonji3547@korea.kr, {feili0820, yjkang5009, hyerink, seoyoon.park, khss}@yonsei.ac.kr

A Study on the Construction of an Emotion Corpus Using a Pre-trained Language Model

Yeonji Jang⁰, Fei Li¹, Yejee Kang¹, Hyerin Kang¹, Seoyoon Park¹, Hansaem Kim[†]

National Institute of Korean Language

Institute of Language and Information Studies, Yonsei University^{1†}

요약

감정 분석은 텍스트에 표현된 인간의 감정을 인식하여 다양한 감정 유형으로 분류하는 것이다. 섬세한 인간의 감정을 보다 정확히 분류하기 위해서는 감정 유형의 분류가 무엇보다 중요하다. 본 연구에서는 사전 학습 언어 모델을 활용하여 우리말샘의 감정 어휘와 용례를 바탕으로 기쁨, 슬픔, 공포, 분노, 혐오, 놀람, 흥미, 지루함, 통증의 감정 유형으로 분류된 감정 말뭉치를 구축하였다. 감정 말뭉치를 구축한 후 성능 평가를 위해 대표적인 트랜스포머 기반 사전 학습 모델 중 RoBERTa, MultiDistilBert, MultiBert, KcBert, KcELECTRA, KoELECTRA를 활용하여 보다 넓은 범위에서 객관적으로 모델 간의 성능을 평가하고 각 감정 유형별 정확도를 바탕으로 감정 유형의 특성을 알아보았다. 그 결과 각 모델의 학습 구조가 다중 분류 말뭉치에 어떤 영향을 주는지 구체적으로 파악할 수 있었으며, ELECTRA가 상대적으로 우수한 성능을 보여주고 있음을 확인하였다. 또한 감정 유형별 성능을 비교를 통해 다양한 감정 유형 중 기쁨, 슬픔, 공포에 대한 성능이 우수하다는 것을 알 수 있었다.

주제어: 말뭉치, 감정 분석, 감정 유형, 언어 모델

1. 서론

자연어처리에서 감정 분석은 복잡하고 섬세한 인간의 감정을 기쁨, 슬픔, 분노, 혐오 등의 여러 유형으로 분류하여 분석하는 것이다. 인간의 감정은 매우 주관적이다. 같은 상황에서도 사람마다 느끼는 감정의 종류와 정도가 다르고 감정을 표현하는 방식도 다르다. 뿐만 아니라 감정은 목소리 톤이나 표정으로도 함께 표현되기 때문에 텍스트만으로 특정 감정을 포착하기는 더욱 쉽지 않다. 어려운 난도에도 불구하고 그동안 감정 분석 연구는 SNS나 블로그 등의 데이터를 대상으로 활발하게 이루어져 왔으며, 감정 분류 체계에 대한 연구도 다양하게 이루어져 왔다.

난도 높은 감정을 보다 정확히 분석하기 위해서는 감정을 몇 가지의 유형으로 나눌 것인지를 결정하는 것도 중요한 요소가 된다. 많은 연구에서 대표적으로 사용되는 감정 유형에는 [1]에서 제시하는 6가지 감정과 [2]에서 제시하는 8가지 감정이 있는데, 두 연구에서 제시한 감정 유형 중 하나를 택해 감정 유형을 그대로 따르거나 몇 가지 감정 유형을 더 추가하여 분석하기도 한다. 그러나 사람이 느끼는 감정은 특정 몇 가지로 분류할 수 없고, 선택한 감정 유형에 포함되지 않는 것처럼 느껴지

거나 여러 유형에 속하는 경우도 있다. 따라서 어떤 감정 유형으로 분석할 것인지 감정의 범주를 정한다고 하더라도 문장에서 드러나는 감정을 명확히 포착하기는 쉽지 않다. 주관적인 성격이 강한 감정 분석의 기계 학습을 위한 데이터셋의 구축은 기계 학습의 발달에 꼭 필요한 부분이다.

대부분의 감정 분석, 오피니언 마이닝, 감정 분석에서 사용되는 데이터는 SNS의 게시글이나 댓글 등 특정 도메인에 대한 개인의 의견이 드러난 경우가 많다. 해당 데이터들은 어떤 대상에 대한 좋고 싫음 또는 특정 감정을 포착하기에 적합한 데이터라고 볼 수 있지만, 어떤 상품이나 영화에 대한 표현과 그 표현에서 드러나는 감정 유형이 인간이 느끼는 모든 감정을 다 아우른다고 보기는 어렵다. 도메인의 영향을 받지 않고 인간의 감정을 폭넓게 인식하기 위해서는 풍부한 감정 어휘가 포함된 감정 어휘 사전과 감정 말뭉치의 구축이 필요하고, 한국어 감정 분석에 적절한 감정 유형에 대한 분석이 반드시 필요하다. 이에 본 연구에서는 네이버 영화 리뷰 감정 말뭉치로 사전 학습된 모델을 기반으로 감정 말뭉치를 구축하였다. 또한, 감정 말뭉치가 트랜스포머 기반의 모델 중 어떤 모델에서 가장 높은 성능을 보이는지 확인하고, 모델에 학습된 데이터의 성격이나 학습 방법에 따라 그

성능이 어떻게 달라지는지를 확인하였다. 또한, 감정 유형의 정확도를 통해 트랜스포머 기반의 모델들이 어떤 감정 유형을 잘 포착하고, 어떤 감정 유형을 잘 포착하지 못하는지 살펴보았다.

2. 관련 연구

2.1. 언어 자원

감정 분석을 위해 국내외로 다양한 연구들이 진행되었는데 주로 감정 분석을 위한 감정 어휘 사전을 구축하거나, SNS 데이터나 일상 대화, 댓글이나 대댓글을 바탕으로 구축한 감정 말뭉치에 관한 연구들이 있다. 감정 어휘 사전은 긍정/부정의 극성값을 가진 데이터[3],[4],[5]부터 여러 감정 유형으로 분류된 데이터[6],[7],[8],[9]까지 다양하게 연구되었다.

먼저 [6]에서는 감정 데이터셋을 ‘행복’, ‘슬픔’, ‘분노’, ‘중립’으로 나누고 각 문장에서 드러나는 감정에 대한 정도를 0-1 사이의 값으로 표현하여 BERT 기반 감정 사전을 구축하였고, [7]에서는 감정 동사에 대해 통사/의미적 특성을 바탕으로 나누었다. 이외에도 심리학적 모델을 적용하여 다중 감정 결정 모델을 제안하는 연구도 있었다.[8] 심리학적 모델을 기반으로 한 기본 감정에 HCI에서 빈번하게 사용되는 감정을 더하여 9가지 감정 유형을 제안하고, 총 428개의 감정 어휘 사전을 구축한 연구도 있었는데[9], 감정 어휘 사전의 감정은 ‘기쁨’, ‘슬픔’, ‘공포’, ‘분노’, ‘혐오’, ‘놀람’, ‘흥미’, ‘지루함’, ‘통증’, ‘중성’, ‘기타’ 범주로 구성되어 있다.

감정 어휘에 대한 연구 뿐만 아니라 최근에는 감정 말뭉치를 구축하려는 시도도 있다. 한국지능정보사회진흥원에서 구축한 감성 대화 말뭉치는 자연어 이해 모델인 ALBERT의 모델링을 위해 구축된 AI기반 챗봇용 감성대화 말뭉치이다. 말뭉치는 60가지 감정이 포함된 대화 텍스트 17만 문장으로 구성되어 있다. 감정은 6개의 기본 감정(분노, 슬픔, 불안, 상처, 당황, 기쁨)을 기준으로 각각 9개의 세부 감정을 포함하여 총 60가지의 감정이 반영되어 있으며, 총 60가지의 감정 상태에 대한 개인의 페르소나를 생성하여 말뭉치를 구축하였다. 국가에서 구축하는 경우 이외에 개인적으로 감정 말뭉치를 구축하는 연구도 있다. [10]에서는 NSMC(Naver Sentiment Movie Corpus)¹의 댓글 원문에 [8]에서 공개한 감정 어휘 사전을 부착하여 총 77,375건의 네이버 영화 리뷰 감정 말뭉치를 구축하였다. 네이버 영화 리뷰 감정 말뭉치는 [8]에서 공개한 감정 어휘 사전의 감정 유형에서 ‘중립’과 ‘기타’를 제외한 9가지 감정 유형을 사용하였는데, 감정 분석 결과 ‘기쁨(54%)’, ‘분노(14.6%)’, ‘슬픔(10%)’ 등 감정 유형별 데이터의 분포가 큰 차이가 있었다. 지금까지 소개된 감정 말뭉치와 같이 하나의 문장을

하나의 감정으로 분석하는 경우 뿐만 아니라, 하나의 문장을 두 가지 이상의 감정으로 분석하는 연구도 진행되고 있다. [11]은 세분화된 감정 분류를 위한 데이터셋으로 인기 사이트 Raddit의 데이터를 27가지의 감정 유형으로 분류한 58,000건의 감정 데이터이다. GoEmotion의 특징은 하나의 문장에 두 가지 이상의 감정이 드러나는 경우에도 그 감정을 모두 분석하는 Multi-label 감정 말뭉치라는 것이다.

본 연구에서는 [10]과 같이 특정 도메인의 영향을 받지 않고 인간이 느끼는 감정을 포괄적으로 아우를 수 있는 감정 어휘 사전과 감정 말뭉치 구축의 필요성을 인지하고 우리말샘의 감정 어휘를 대상으로 감정 말뭉치를 구축하였다.

2.2. 언어 모델

사전 학습 언어 모델을 활용한 연구가 많아지면서 다양한 언어 모델이 개발되고 있다. 자연어처리 연구의 활발함과 함께 여러 키워드가 등장하고 있지만 트랜스포머를 기반으로 한 사전 학습 기법들이 높은 성능을 보이면서 다양한 방식으로 응용된 모델들이 계속해서 나오고 있다. 트랜스포머 기반의 언어 모델은 크게 인코더를 사전 학습한 모델과 디코더를 사전 학습한 모델로 나누어 볼 수 있는데, 인코더를 사전 학습한 모델에는 BERT, RoBERTa, ELECTRA 등이 있다.

먼저 Google에서 발표한 BERT는 앞뒤로 등장하는 단어를 바탕으로 빈칸을 표시하여 예측하는 방식으로 학습을 진행한다. BERT는 위키피디아와 BooksCorpus를 사전 학습시킨 언어 모델로 문맥을 반영한 임베딩을 사용한다. 양방향 학습을 가능하게 한 BERT의 학습 방식은 이전에 먼저 개발된 GPT보다 좋은 성능을 보였고, 이후 BERT를 개량한 모델들이 발표되었다, BERT가 GPT보다 좋은 성능을 보이지만 영어 데이터를 사전 학습한 모델이기 때문에 한국어에 대해서는 정확도가 떨어진다. Bert는 base와 large 모델을 공개하였는데, BERT 모델의 크기는 1억 1,100만개이며, large 모델의 크기는 3억 4,000만개이다. [12] BERT 중에서 대표적으로 Google의 multilingual Bert와 SKT의 KoBERT²가 문어를 학습한 언어 모델이라고 할 수 있다. Google의 multilingual Bert는 한국어를 포함해 100개의 언어로 이루어진 언어로 작성된 위키피디아 문서를 학습한 모델이며, KoBERT는 SKTBrain에서 BERT에 한국어 데이터를 추가로 학습시켜 한국어에도 잘 활용할 수 있도록 KoBERT를 개발하였다. KoBERT는 위키피디아나 뉴스 등에서 수집한 대규모 말뭉치를 학습하였으나 비교적 정제된 데이터를 학습하여 문어에 특화되어 있다. 이에 댓글이나 일상 대화 등 비정형 데이터에도 좋은 성능을 낼 수 있도록 15GB의 한국어 뉴스 댓글을 추가로 사전 학습하여 비정형 데이터에도 적합한 결과를 보이는 KcBERT가 개발되었다. KcBERT는 2019년 1월부터 2020년 6

¹ <https://github.com/e9t/nsmc>

² <https://github.com/SKTBrain/KoBERT>

월까지의 뉴스 댓글 약 1억 1천만 건을 수집하고 전처리 작업을 거쳐 Huggingface의 Bert WordPiece Tokenizer를 학습한 모델이다.[13] BERT의 후속으로 등장한 RoBERTa는 BERT를 개선하기 위해 모델 설계에 변화를 주고, 새로운 데이터셋을 사용하여 데이터셋과 학습량의 중요성을 보여준 모델이다.[14] 다음으로 ELECTRA는 기존의 학습 단계 손실을 개선하기 위해 mask된 단어를 다른 단어로 치환하고, 각 단어의 치환 여부를 분류하는 모델이다. ELECTRA는 모든 단어에서 손실값이 계산되어 기존의 학습 방법보다 높은 성능을 보이는데, 모델이 작을수록 성능 차이를 크게 보인다. ELECTRA는 small과 base, large 모델을 공개하였고, 영어 텍스트를 학습하였다. [15]

ELECTRA의 한국어 버전인 KoELECTRA³는 14GB/34GB의 한국어 텍스트로 학습한 모델로 v3는 약 20GB의 모두의 말뭉치를 추가 학습하였다. ELECTRA는 정확도와 함께 학습의 효율성도 주목을 받는 모델인데, Replaced Token Detection(RTD)이라는 새로운 사전 학습 방법을 제안하며 기존의 GPT나 BERT보다 월등한 성능을 보여주었다.

KoELECTRA⁴의 경우 한국어 위키 백과, 뉴스 기사 등 정제된 데이터를 기반으로 학습한 한국어 트랜스포머 기반의 모델과는 달리 구어체 특징에 적용하기 위해 네이버 뉴스에서 댓글과 대댓글을 수집하여 학습한 Pretrained ELECTRA 모델이다. 이름에서도 알 수 있듯 한국어 댓글을 학습한 KoELECTRA는 잘 정제되어 있지 않고 신조어와 오타자가 많은 데이터셋을 학습하는데 적합하다.

[15]에서는 트랜스포머 기반의 다양한 인코더 사전 학습 언어 모델 중 대표적인 모델로 BERT, XLNet, RoBERTa, ALBERT, ELECTRA를 소개하고 있으며, [16]에서는 KoBERT, HanBERT, KoELECTRA, KcBERT, KcELECTRA, DistilKoBERT를 소개하고 있다. 본 연구에서는 선행 연구에서 대표적으로 소개된 트랜스포머 기반의 언어 모델 중 RoBERTa, MultiDistilBert, MultiBert, KcBert, KcELECTRA, KoELECTRA를 중심으로 구축한 말뭉치의 성능을 평가하고, 각각의 모델의 특성을 비교 분석하였다.

3. 감정 말뭉치 구축 절차

3.1. 데이터 선정

본 연구에서는 국립국어원의 개방형 국어사전 우리말샘의 용례를 활용하여 감정 말뭉치를 구축하였다. 먼저 감정 어휘를 선정하기 위해 우리말샘의 정의문에 ‘마음’, ‘느낌’, ‘감정’ 등의 어휘가 포함된 감정 어휘 중 동사와 형용사, 감탄사 8,965개의 감정 어휘 중 용례가 있는 4,139개의 어휘를 분석 대상으로 삼았다. 감정 어휘 중 용례가 있는 어휘를 분석 대상으로 삼은 이유는 본 연구에서 구축한 감정 말뭉치가 감정 어휘의 용례로 구성되어 있기 때문이다. 특정 감정 어휘에 대한 정의문

은 ‘어떠한 마음’, ‘어떠한 느낌’, ‘어떠한 감정’에 대한 설명이므로 정의문 자체에서는 감정을 추측하기 어렵다. 이와 달리 용례의 경우에는 어휘가 어떠한 맥락에서 사용되는지 실제 예시를 보여주므로 감정이 보다 잘 드러난다. 따라서 감정 어휘가 사용된 예시를 대상으로 감정 말뭉치를 구축하였다. 이때 하나의 감정 어휘에 하나 이상의 용례가 있는 경우가 있어 전체 감정 말뭉치 구축에 사용된 문장은 총 20,634개이다.

3.2. 사전 학습 모델

본 연구에서 구축한 감정 말뭉치는 KcBert에 여러 차례 학습시켜 구축한 감정 말뭉치이다. 뉴스 댓글 데이터로 학습을 진행한 KcBert는 2019년부터 2020년 6월까지의 기사 댓글과 대댓글을 수집하여 모은 약 15GB의 댓글 데이터셋을 학습한 모델이다. KcBert는 뉴스 댓글 데이터로 학습시켰기 때문에 NSMC와 같은 댓글 데이터에서 높은 성능을 보인다. [10]에서는 감정 어휘 사전을 기반으로 네이버 영화평 말뭉치에 9가지의 감정 유형을 부착하여 감정 말뭉치를 구축하고, 구축한 감정 말뭉치를 KcBert에 학습시켜 성능을 평가하였다.

본 연구에서는 [10]에서 구축한 네이버 영화 리뷰 감정 말뭉치를 재학습시킨 pre-train 모델에 우리말샘의 용례를 넣어 감정 유형을 부착하였다.

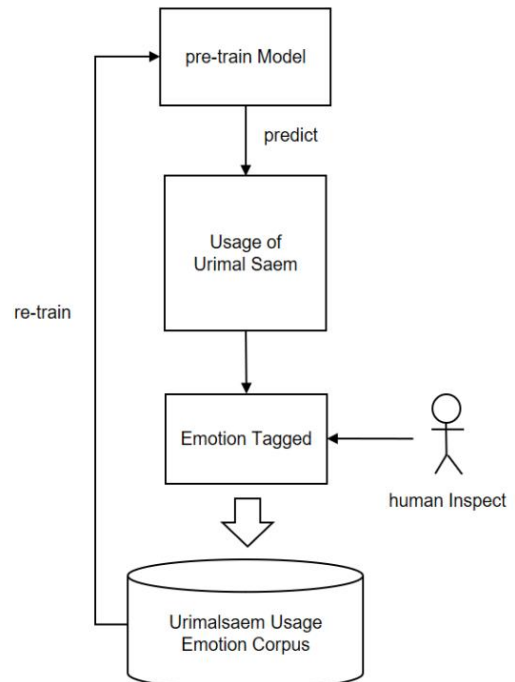


그림 1. 우리말샘 감정 말뭉치 구축 과정

기존 학습 모델에는 ‘기쁨’, ‘슬픔’, ‘공포’,

³ <https://github.com/monologg/KoELECTRA>

⁴ <https://github.com/Beomi/KcELECTRA>

‘분노’, ‘혐오’, ‘놀람’, ‘흥미’, ‘지루함’, ‘통증’의 9가지의 감정 유형이 학습되어 있으므로 우리말샘 용례도 9가지 감정 유형이 부착되어 나오게 된다. 따라서 학습 모델에 의해 예측된 감정 유형에 대해서 사람이 개입하여 해당 감정 유형이 잘 부착되었는지 판단하는 작업이 필요하다. 이에 본 연구에서는 사전 학습된 모델이 9가지 감정으로 분류한 우리말샘 데이터의 감정 유형을 일부 수정하였고, 수정한 우리말샘 감정 말뭉치를 다시 사전 학습된 모델에 넣어 감정 유형을 태깅하였다. 우리말샘 감정 말뭉치 구축 과정을 정리하면 그림 1과 같다.

본 연구에서는 우리말샘 감정 어휘를 기반으로 선정된 감정 어휘의 용례를 대상으로 우리말샘 감정 말뭉치를 구축하였다. 학습된 모델을 통해 9가지 유형의 감정을 부착하고, 일부 잘못된 결과는 사람의 검수를 통해 감정 유형을 바로잡는 절차를 거쳤다.

우리말샘 감정 말뭉치는 [8]에서 공개한 감정 어휘 목록 428개를 활용하여 구축된 네이버 영화 리뷰 말뭉치를 추가로 학습한 KcBert를 기반으로 감정 유형을 분류한 감정 말뭉치이다. 따라서 우리말샘 감정 말뭉치의 감정 유형도 네이버 영화 리뷰 감정 말뭉치와 마찬가지로 9가지(기쁨, 슬픔, 공포, 분노, 혐오, 놀람, 흥미, 지루함, 통증)로 분류된다.

우리말샘 감정 말뭉치의 감정 유형을 살펴보면 기쁨 16.9%, 분노 13.45%, 놀람 13.24%, 흥미 11.46%, 공포 11.24%, 통증 10.97%, 혐오 8.82%, 슬픔 8.17%, 지루함 5.75% 순으로 나타나고 있다.

감정 말뭉치에 포함된 감정 유형의 분포는 아래와 같다.

표 1. 감정 말뭉치의 감정 분포

감정 유형	데이터 수 (건)
기쁨(0)	3,488
슬픔(1)	1,685
공포(2)	2,319
분노(3)	2,776
혐오(4)	1,819
놀람(5)	2,732
흥미(6)	2,365
지루함(7)	1,186
통증(8)	2,264
전체	20,634

4. 실험 및 결과

우리말샘 감정 어휘를 기반으로 구축한 감정 말뭉치는 네이버 영화 리뷰 감정 말뭉치를 학습한 KcBert로 감정을 태깅한 감정 분석 말뭉치이다.

본 연구에서는 여러 유형으로 태깅된 우리말샘 감정 분석 말뭉치를 다양한 언어 모델에 학습시켜 멀티 레이블 데이터의 학습 성능을 평가해 보았다. 구축한 말뭉치는 총 20,634건이며 훈련 데이터는 16,504건, 테스트 데이터는 4,130건이다.

4.1. 모델 학습

사전 학습된 모델로 구축된 우리말샘 감정 말뭉치를 트랜스포머 기반의 모델 중 RoBERTa, MultiDistilBert, MultiBert, KcBert, KcELECTRA, KoELECTRA를 사용하여 성능을 평가하였다. 사전 학습된 모델은 HuggingFace의 Tokenizers 라이브러리를 통해 Bert WordPiece Tokenizer를 학습시킨 모델에 [10]에서 구축한 네이버 영화 리뷰 감정 말뭉치를 추가로 학습시킨 것으로 학습된 감정 말뭉치의 9가지 유형에 따라 출력 라벨 수가 9개로 설정되어 있다.

본 연구에서는 사전 학습된 모델을 기반으로 우리말샘 감정 말뭉치를 구축하고 트랜스포머 기반의 언어 모델인 RoBERTa, MultiDistilBert, MultiBert, KcBert, KcELECTRA, KoELECTRA를 중심으로 성능을 평가하였다. 이들은 트랜스포머 기반의 언어 모델 중 널리 쓰이고 있는 대표적인 모델로, 모델들 간의 성능을 비교해야 보다 넓은 범위에서 모델 성능을 정확성, 효율성, 범용성 등의 차원에서 객관적으로 평가할 수 있기 때문이다.

성능 평가에 사용한 모델의 파라미터는 다음과 같다.

표 2. 모델별 하이퍼파라미터

모델	하이퍼파라미터 (epoch, learning rate, batch size)
KoELECTRA	15, 2e-5, 128
KcELECTRA	20, 5e-5, 128
KcBert	10, 5e-5, 128
MultiBert	10, 2e-5, 64
MultiDistilBert	10, 1e-4, 128
RoBERTa	30, 5e-5, 128

4.2. 학습 결과

표 3은 본 연구에서 구축한 우리말샘 감정 말뭉치를 학습시켜 성능을 평가한 트랜스포머 기반 언어 모델들의 F1 Score이다.

표 3. 모델별 F1 점수

	Macro	Weighted
KoELECTRA	0.7857	0.793
KcELECTRA	0.6751	0.687
KcBert	0.6139	0.6244
MultiBert	0.5471	0.5627

MultiDistilBert	0.5264	0.5358
RoBERTa	0.4161	0.4265

표 3의 F1 점수를 통해 가장 먼저 확인할 수 있는 것은 우리말샘 감정 말뭉치를 학습한 모델 중 가장 높은 성능을 보이는 모델은 KoELECTRA라는 것이다. ELECTRA의 학습 방법이나, ELECTRA에 사전 학습된 데이터의 특성을 고려했을 때 ELECTRA가 가장 좋은 성능을 보일 것이라는 기대는 있었으나 KcELECTRA와 KoELECTRA의 성능이 다소 차이를 보였다. 이는 우리말샘 감정 말뭉치가 구어가 아닌 문어로 구성되어 있어 구어나 신조어, 댓글 등으로 학습된 KcELECTRA보다 위키백과나 뉴스와 같이 정제된 데이터를 학습한 KoELECTRA에서 좋은 성능을 보인 것을 추측해 볼 수 있다. 두 모델이 이러한 성능 차이를 보이고는 있으나, BERT를 계량한 다른 모델들보다 성능이 우수함을 알 수 있다.

두 번째로 KcBert와 MultiBert의 성능을 비교해 보면 KcBert의 성능이 더 우수함을 알 수 있다. 두 모델은 Bert에 학습시킨 데이터와 학습 방법에 차이가 있다. MultiBert(Multilingual Bert)는 여러 언어들을 같이 학습한 모델로 100개의 언어를 지원하지만, 모델 내에 포함된 어휘의 수가 많지 않아 좋은 성능을 기대하기엔 무리가 있다. 반면 KcBert는 한국어 댓글을 학습한 모델이므로 다국어 지원 모델보다 성능이 좋게 나오는 것은 당연한 현상이다. 다만 KcBert 역시 KcELECTRA와 마찬가지로 댓글과 같은 구어를 학습한 모델이기 때문에 본 연구에서 구축하여 학습시킨 문어 말뭉치에는 일반적인 경우보다 성능이 떨어져 보이는 경향이 있다.

세 번째, KcBert로 [10]에서 구축한 네이버 영화 리뷰 말뭉치와 본 연구에서 구축한 우리말샘 감정 말뭉치를 학습시켰을 때의 성능을 비교해 볼 수 있다. 두 말뭉치의 차이는 구어와 문어라는 데에서 차이가 있고, 포함하고 있는 어휘의 난도에도 차이가 있다. 네이버 영화 리뷰 말뭉치의 경우 구어 작성된 영화 리뷰라는 특성상 사용된 어휘의 난도가 높지 않다는 특징이 있다. 반면, 우리말샘 감정 말뭉치는 사전에 나타나는 동사, 형용사, 감탄사 중 ‘~ 느낌’, ‘~ 마음’ 등으로 풀이되는 감정 어휘와 감정 어휘가 드러난 용례를 바탕으로 구성되어 있어 일반적으로 사용되는 문장보다 더 정제되어 있다. 따라서 같은 Bert 모델로 학습했을 때, 특히 네이버 댓글로 사전 학습된 KcBert의 경우에는 더욱 네이버 영화 리뷰 감정 말뭉치의 성능이 우수하게 나타난다.

4.2. 감정 유형별 성능 평가

트랜스포머 기반의 모델들이 감정 말뭉치 내에 분류된 각각의 감정을 얼마나 정확히 포착하고 있는지 확인해 보았다. 그림 2는 [10]에서 구축한 네이버 영화 리뷰 감정 말뭉치를 학습한 KcBert의 성능과 본 연구에서 구축한 우리말샘 감정 말뭉치를 학습한 트랜스포머 기반의 모델의 성능을 감정 유형별로 살펴본 그래프이다. 점선으로 표현된 부분이 [10]의 KcBERT 성능 평가 결과인데, 본 연구에서 구축한 우리말샘 감정 말뭉치의 감정 유형

에서 보이는 성능과 조금 다른 결과를 보여주고 있다.

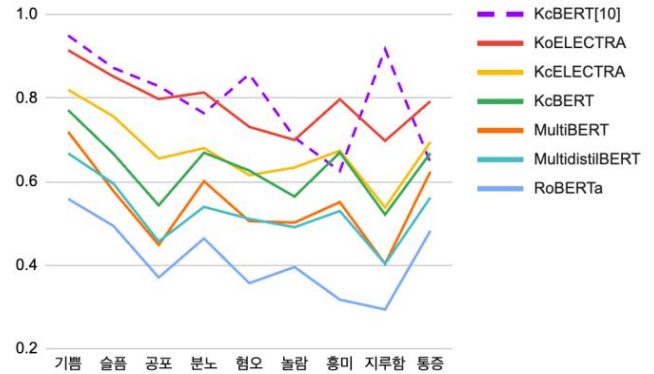


그림 2. 감정 유형별 정확도

우선 전체적으로 살펴보면 공통적으로 드러나는 것은 ‘기쁨’에 대한 정확도가 높은 반면 ‘기쁨’에 비해 상대적으로 ‘슬픔’에서 ‘공포’로 갈수록 정확도가 낮아지는 것이다. 즉, ‘기쁨’과 ‘슬픔’, ‘공포’ 유형에 대한 감정 분석은 말뭉치의 특성이 어떠한 학습 모델들의 평가가 잘 나오고 있다고 볼 수 있다.

이와 달리 ‘분노’나 ‘혐오’, ‘지루함’은 데이터의 특성에 따라 정확도가 보이는 경향이 크다. 이는 모델이 ‘기쁨’이나 ‘슬픔’, ‘공포’에 대한 감정은 상대적으로 명확하게 포착하는 반면, ‘분노’, ‘혐오’, ‘지루함’ 등의 감정에 대해서는 명확하게 포착하지 못한다는 것으로 볼 수 있다. 일반적으로 ‘기쁨’이나 ‘슬픔’은 기본적으로 쉽게 판단할 수 있고, 감정 표현도 상대적으로 명확한 편에 속하는 반면, ‘분노’나 ‘혐오’, ‘지루함’의 경우 각 감정 유형의 경계가 생각보다 명확하지 않아 특정 감정으로의 분류가 어려운 경우가 있다.

표 4. 감정 유형 분류 예시

유형	우리말샘 용례
분노	그는 그를 따돌리는 친구들에게 성미나서 집에 가 버렸다.
	불통불통하게 대하는 점원의 태도가 기분을 상하게 했다.
혐오	자칭 도사라고 하는 그의 모습은 기괴스럽고 주술적이다.
	예기, 이 배은망덕한 놈.
지루함	하루 종일 좁은 방에만 있으려니 답답하다.
	몇 시간을 쪼그리고 장롱 안에 있으려니 나중에는 답답해서 미칠 것만 같았다.

표 4는 ‘분노’, ‘혐오’, ‘지루함’으로 분석된 실제 용례를 제시한 것이다. 먼저 ‘분노’로 분석된 ‘그는 그를 따돌리는 친구들에게 성미나서 집에 가 버렸다’를 보면 명확히 분노의 감정이 느껴지지만, ‘불통불통하게 대하는 점원의 태도가 기분을 상하게 했다’의 경우

‘기분이 상했다’는 표현이 분노의 감정인지 슬픔의 감정인지 그 정도를 알기 모호하다. 다음으로 ‘혐오’로 분석된 ‘자칭 도사라고 하는 그의 모습은 기괴스럽고 주술적이다.’는 ‘기괴스럽다’라는 어휘가 포착되어 ‘혐오’로 분류되었는데, ‘예기, 이 배은망덕한 놈.’의 경우에는 혐오의 감정보다는 분노의 감정에 가깝다고 판단할 가능성도 있어 어느 감정이 더 명확한지 판단하기 어렵다. ‘지루함’의 경우 ‘하루 종일 좁은 방에만 있으려니 갑갑하다’에서는 지루한 감정이 느껴지지만, ‘몇 시간을 쪼그리고 장롱 안에 있으려니 나중에는 갑갑해서 미칠 것만 같았다’는 지루함보다는 짜증, 분노의 감정이 더 많이 느껴진다. 이렇게 특정 감정이라고 판단하기 어려운 데이터의 분류는 감정 분석에 있어서 감정 유형 설정이 얼마나 중요한지를 다시 한 번 생각하게 한다. 같은 감정으로 분석되었다고 하더라도 포착되는 감정의 정도가 다르거나, 하나 이상의 감정이 포착되어 하나의 감정으로 분류하기 어려운 경우, 또는 감정 유형들 간의 경계에 있어 특정 감정으로 판단이 모호한 경우를 고려하여 어떻게 감정을 분석할 것인지 결정하는 것이 감정 분석에 중요한 요소가 될 것이다.

5. 결론

본 연구에서는 인간이 느낄 수 있는 다양하고 섬세한 감정을 포괄적으로 아우를 수 있는 감정 말뭉치를 구축하기 위해 우리말샘의 감정 어휘와 용례를 기반으로 감정 말뭉치를 구축하였다. 말뭉치 구축은 사전 학습된 모델을 기반으로 이루어졌는데, 사전 학습된 모델은 9가지 감정 유형(기쁨, 슬픔, 공포, 분노, 혐오, 놀람, 흥미, 지루함, 통증)으로 기구축된 네이버 영화 리뷰 감정 말뭉치가 학습된 KcBert이며, 사전 학습된 모델을 기반으로 감정 태깅을 한 후 일부 내용 검수를 통해 재학습시켜 말뭉치 구축의 정확도를 높이고자 하였다. 감정 말뭉치를 구축한 뒤 한국어 언어 모델로 성능을 평가하였는데, 성능 평가를 위해 트랜스포머 기반의 언어 모델 중 대표적인 RoBERTa, MultiDistilBert, MultiBert, KcBert, KcELECTRA, KoELECTRA를 활용하였다. 해당 모델을 선정하는 이유는 대표적인 트랜스포머 기반의 언어 모델 간의 성능 분석을 통해 보다 넓은 범위에서 객관적으로 모델 성능을 평가하고, 각 감정 유형별 정확도를 바탕으로 도메인별 감정 유형의 특성을 알아보기 위함이다. 성능 평가 결과 BERT나 BERT를 개량한 모델보다 ELECTRA의 성능이 가장 우수함을 알 수 있었고, 다국어로 학습된 모델보다 한국어로 학습된 모델에서 성능이 우수함을 알 수 있었다. 또한 구어를 학습한 모델은 구어에 강하고, 문어를 학습한 모델은 문어에 강하다는 것을 통해 사전 학습 데이터의 유형의 중요성을 다시 한 번 확인할 수 있었다.

이외에도 감정 유형별 모델의 성능을 통해 한국어 언어 모델이 어떤 감정에 더 강하고, 어떤 감정에 취약한지 확인할 수 있었는데, 구어와 문어를 아울러 기쁨, 슬픔, 공포에 대한 성능이 우수했으며, 분노, 혐오, 지루함에 대한 분류가 쉽지 않다는 것을 알 수 있었다.

인간의 다양하고 섬세한 감정은 인간의 판단으로도 쉽

지 않은 일이고 어떠한 감정으로 느껴지는지에 대한 문제에 정답이 없는 경우도 있다. 때문에 감정을 몇 가지 유형으로 분류할 수 있는가에 대한 문제는 앞으로도 연구되어야 할 것이다. 뿐만 아니라 하나의 문장에 하나 이상의 감정이 들어 있는 경우도 있고, 하나의 문장에 하나의 감정이 느껴지더라도 그 감정의 정도가 다른 경우도 많다. 본 연구에서는 하나의 문장에 하나의 감정 유형을 부착하여 감정 말뭉치를 구축하였으나, 단순히 몇 가지의 감정 유형으로 분류하는 것에서 그치지 않고 각 감정 유형에서 나타나는 감정의 정도를 분석하는 것이 한국어 감정 분석 언어 자원 구축의 향후 과제이다.

참고문헌

- [1] 황유선, “페이스북 일상담화의 감정 탐색.”, 한국콘텐츠학회논문지, 16.2, 2016.
- [2] Ekman, P., “Universals and cultural differences in facial expressions of emotion”, Proceedings of the 1971 Nebraska Symposium on Motivation, 207-283, 1971.
- [3] Stefano Baccianella, Andrea Esuli, Fabrizio Sebastiani, “SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining”
- [4] 박상민, 나철원, 최민성, 이다희, 은병원, “Bi-LSTM 기반의 한국어 감성사전 구축 방안.”, 지능정보연구, vol 24, No 4, pp. 219-240, 2018.
- [5] 신동혁, 조동희, 남지순, “한국어 감성 사전 DecoSelex 구축을 위한 영어 SentiWordNet 활용 및 보완 논의”, 한국사전학, No.28, pp.75-111. 2016.
- [6] 이가은, “Bert 기반 한국어 감정 사전을 이용한 감정 예측기 개발”, 석사, 서강대학교 정보통신대학원, 2020.
- [7] 김은영, “국어 감정동사 연구”, 박사, 전남대학교 대학원, 2004
- [8] 김해준, 도준호, 선주오, 정서희, 이현아, “감정 분석에서의 심리 모델 적용 비교 연구”, 제32회 한글 및 한국어 정보처리 학술대회 논문집, pp.450-452. 2020.
- [9] 손선주, 박미숙, 박지은, 손진훈, “한국어 감정표현 단어의 추출과 범주화”, 감성과학, Vol. 15, No. 1, pp. 105-120, 2012.
- [10] 장연지, 최지선, 김한샘, “감정 어휘 사전을 활용한 KcBert 기반 영화 리뷰 말뭉치 감정 분석”, 정보과학회논문지 49, no.8 (2022) : 608-616
- [11] Demszky, Dorottya and Movshovitz-Attias, Dana and Ko, Jeongwoo and Cowen, Alan and Nemade, Gaurav and Ravi, Sujith, “GoEmotions: A Dataset of Fine-Grained Emotions”
- [12] Devlin, Jacob and Chang, Ming-Wei and Lee, Kenton and Toutanova, Kristina, “BERT: Pre-training of Deep Bidirectional Transformers for Language

- Understanding” , <https://arxiv.org/abs/1810.04805>
- [13] 이준범, “KcBert: 한국어 댓글로 학습한 Bert” , 제 32회 한글 및 한국어 정보처리 학술대회 논문집, pp.437-440. 2020.
- [14] Y.Liu,M.Ott,N.Goyal,J.Du,M.Joshi,D.Chen,O.Levy, M. Lewis, L. Zettlemoyer and V. Stoyanov, “RoBERTa: A Robustly Optimized BERT Pretraining Approach” , arXiv:1907.11692, 2019.
- [15] Kevin Clark, Minh-Thang Luong, Quoc V. Le, Christopher D. Manning, “ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators” , In International Conference on Learning Representations, 2020.
- [16] 임준호, “딥러닝 사전학습 언어모델 연구 동향” , 정보과학회지,39(4),9-14. 2021.
- [17] Yang, Kichang, “Transformer-based Korean Pretrained Language Models: A Survey on Three Years of Progress” , arXiv:2112.03014, 2021.