LSTM과 Bi-LSTM을 이용한 개인 성향분석

유하영, 문성욱, 김재현*, 조영완*
서경대학교
bluebarry37@naver.com, uky6202@naver.com,
statsr@skuniv.ac.kr*, ywcho@skuniv.ac.kr*

Personality Analysis using LSTM and Bi-LSTM

Ryu Ha Yeong, Moon Seong Uk, Kim Jae Hyun*, Cho Young Wan*
Seokyeong Univ

요 약

본 논문은 딥러닝 기법(LSTM과 Bi-LSTM)을 이용해 MBTI의 E와 I 성격 유형을 다양한 구조로 학습하고, 그에 대한 성능 평가 및 성격 유형을 분류하는 모델을 제안한다. 학습 데이터의 사용 범위와 사전학습된 임베딩 벡터의 사용에 따른 영향, 두 가지 측면에서 5가지의 학습 환경으로 모델을 훈련하였다. 실험 결과 모든 모델이 약 90% 이상의 학습 정확도를 보였으며, 실제 성격 예측 단계에서는 Bi-LSTM이 LSTM에 비해 높은 예측률을 보였다. Bi-LSTM 모델 중에서도 사전학습된 임베딩 벡터와 전체 단어학습을 결합한 Bi-LSTM 모델이 가장 많은 단어를 예측하였다.

I. 서 론

사람의 성향이나 심리를 이해하기 위해서는 의학적, 언어적 분석 등이 존재한다. 그 중, 언어적 심리분석 분야에서는 자연어 처리를 기반으로 한 인공지능 기법이 새롭게 주목받고 있다. 과거에는 인공지능의 문장 분석 능력이 긍정 및 부정 분류와 같은 한정된 문제만 해결 가능했다. 하지만 인공지능이 발전함에 따라 고도화된 언어모델을 통해 사람의 심리와 성향까지 탐색할 수 있게 되었다.

최근, 새로운 성격 유형 분석 도구인 MBTI가 등장하여 인간의 심리 이해에 큰 지표를 제시하고 있다. MBTI(Myers-Briggs Type Indicator)는 사용자의 4가지의 선호 경향을 기반으로 16개의 성격 유형으로 분류하는 도구이다. 전통적인 MBTI 검사 방식은 사용자가 객관식 질문에 응답하도록 설계되어 있다. 이러한 방식의 검사는 사용자의 응답 범위를 제한하며, 인간의 복잡한 심리를 묘사하기 어렵다는 한계점이 존재한다. 다시 말해, 사용자는 '그렇다'와 '그렇지 않다'라는 단순한 선택만 할 수 있어, 검사 결과의 정확성에 의문을 제기할 수 있다. 만약 사용자가 주관식으로 자신의 생각과 감정을 표현할 수 있다면, MBTI의 성격 유형을 더욱 정밀하게 파악할 수 있을 것이다. 이러한 문제점을 극복하기 위하여, 텍스트 데이터를 활용하여 MBTI 성격 유형을 분석하고 예측하는 모델을 제안한다.

본 논문은 MBTI 성격 유형 중 'EI preference (Extraversion or Introversion, 이하 유형 E/유형 I)'을 중심으로 분석을 수행한다.[1] E 유형과 I 유형을 구분하는 방법으로는 binary-class를 적용하였다. 실험에 사용할 데이터는 GPT-4와 GPT-2를 혼용해 생성하고, 자연어 처리에서의 데이터 증강 기법을 이용해 데이터의 수를 늘렸다.

제작된 데이터 셋을 기반으로 각 유형의 특징을 분석하기 위해 자연어처리의 기본적인 모델인 LSTM 및 Bi-LSTM을 활용했다.

모델 학습에는 사전학습된 임베딩 벡터를 사용한 것과 이를 사용하지 않은 경우와의 차이를 비교하고, 학습 데이터 선택 범위를 적용하여 그에 따른 성능 변화를 분석하였다. 마지막으로 개발된 모델을 바탕으로 실제 텍스트 데이터에 대한 예측을 진행해보고, 이를 바탕으로 적합한 성격 유형을 판단하는 모델을 제시한다.

II. LSTM을 이용한 MBTI 유형 분석

2.1 선행 연구

국내 인터넷 커뮤니티 사이트의 게시글을 활용한 선행 연구에서는 LSTM 신경망으로 약 33,000개의 데이터를 학습하였고, 그 결과 예측 정확도는 20.29%로 나타났다.[2] 선행 연구는 크롤링을 사용하여 데이터를 수집하였지만, 본 논문에서는 GPT를 활용해 새로운 학습 데이터를 생성한다.

타 연구에서는 MBTI 성격 유형 분류를 위한 방식을 제안한다.[3] 16개의 multi-class 성격 유형을 개별적으로 분류하는 16-class classifier이고, 두 번째는 4가지의 선호 경향을 binary-class로 분류한 뒤 조합하는 4 binary classifier이다. 전자의 방식은 23%의 Test Accuracy를 보였고, 후자의 방식에서는 38%로 전자의 방식보다 높은 정확도를 보였다. 이러한 결과는 4 binary classifier 방식이 MBTI의 특징을 더 정확하게 반영하고 있음을 알 수 있으며, 본 논문에서도 이와 같은 방식을 활용한다.

2.2 LSTM

LSTM(Long Short-Term Memory)은 시퀀스 데이터에서 장기 의존성 문제를 해결하기 위해 고안된 RNN의 한 형태이다.[4] LSTM은 과거의 정보를 통해 현재의 컨텍스트를 이해하는데 유용하며, 이 특징은 Vanilla RNN에 비해 텍스트와 같은 시퀀스 데이터 처리에 효율적이다.

그러나 입력 시퀀스가 길어질수록 초기 정보를 잃어버리는 장기 의존성 문제를 해결하기 위해 역방향으로도 정보를 처리하는

Bi-LSTM(Bidirectional LSTM)을 도입하였다.[5] 이 구조는 과거와 미래의 정보를 동시에 활용하여 정확도를 향상시킨다.

본 논문은 LSTM과 Bi-LSTM 두 모델에 대한 실험을 진행하고 성능을 비교한다.

Ⅲ. 분류 모델 비교 실험

3.1 데이터 수집

본 연구는 16개의 성격 유형 중 E 유형과 I 유형을 중점적으로 분류하는 것을 주목표로 설정하였다. 분류 방식은 이전 연구와 마찬가지로 binary-class를 사용하였다. 성격 유형 E에 해당하면 1, 유형 I면 0으로 labeling 하였다.

성격 유형의 분류를 위해 각 유형 사용자의 특성을 정확하게 파악해야 한다. 이에 두 유형에 대한 응답 데이터를 학습용으로 구성하였다. 학습 데이터 생성에 앞서, 주어진 질문을 바탕으로 응답 데이터를 구성할 것이기 때문에 질문 데이터 셋을 먼저 생성했다. 이를 위해 16 Personalities와 같은 대중적인 검사에서 사용되는 질문들을 수집하고 분석한 뒤, 총 20개의 질문을 새롭게 제작하였다.[6]

기초 응답은 각 질문에 대해 대략 30개씩 작성했다. 이후 GPT-4를 이용해 E와 I 유형의 특징이 분명하게 드러나는 약 70개의 모범답안을 생성하였다. 앞서 제작한 응답을 바탕으로 KoGPT-2 모델에 기초 응답과 모범답안을 결합한 데이터 셋으로 fine-tuning한 후, 응답 데이터의 다양성을 확장하여 최종적으로 총 1,375개의 질문-응답 데이터 셋을 만들었다. KoGPT-2는 GPT-4만을 이용하여 데이터를 생성한 방법보다 비용적 면에서 합리적이고 실험목적에 맞게 수정 가능하다는 장점이 있기에, 이와 같은 방법을 적용하였다. 데이터 셋 중 1,000개는 Training set에 이용하였고, 나머지 375개는 Validation set으로 사용하였다.

과적합을 방지하기 위하여 1,000개의 Training set에 대하여 Data Augmentation을 진행하였다.[7] 본 논문은 한국어에 적합하다고 판단되는 RS(Random Swap)와 RD(Random Deletion)를 사용했다. 이를 통해 한 응답 데이터당 3개의 문장으로 확장하여 총 3,000개의 문장을 생성하고 중복된 문장을 제거하여 2,965개의 증강된 데이터를 학습에 사용하였다. 성능 테스트를 위해서 사람들이 자유롭게 작성한 38개의 응답을 수집하여 활용하였다.

구축된 데이터는 모두 한국어 정보처리 패키지 Konlpy의 Mecab을 이용하여 형태소 분석을 진행하였다.[8]

3.2 모델 제안

본 연구에서는 더 적합한 딥러닝 모델을 제안하는 것을 목표로 다양한 접근법을 시도하였다.

첫 번째로, 임베딩 레이어의 선택에서 기본 임베딩과 사전학습된 FastText 임베딩 중 어느 것이 더 효과적인지를 평가하였다. FastText 임베딩은 3만 개의 한국어 단어로 사전에 학습된 벡터(FastText-KO)를 활용한다.[9]

두 번째로, 학습 데이터 선택 범위에 따른 성능 변화를 관찰하였다. 문법적 역할을 하는 조사나 어미와 같은 형태소는 제거하고, 주요 정보를 담고 있는 명사, 동사, 형용사, 그리고 어간을 추출하여 사용한 extracted word와 추출하지 않은 전체 문장 데이터인 whole word에 따른 성능 비교를 진행했다.

따라서, 이 연구에서는 위의 두 가지 기준을 바탕으로 LSTM과 Bi-LSTM을 사용하여 총 5가지 조합의 모델을 실험하였다. (표 1)에는 각 모델의 조합과 그에 따른 특성을 요약하여 제시하였다. 이후 본 논문에서는 5가지 조합 모델을 (표 1)의 model name에 정의되어있는 이름으로 명칭 한다.

(표 1) 분류 모델별 학습 방식 비교

Classification model	Use of FastText	Training Data Scope	Model name
LSTM	X	extracted word	Simple-LSTM
LSTM	O	extracted word	Fa-LSTM
Bi-LSTM	X	whole word	Simple-BLSTM
Bi-LSTM	O	extracted word	Fa-extr-BLSTM
Bi-LSTM	O	whole word	Fa-whl-BLSTM

3.3 성격 유형 정의

제안한 5개 모델에 대해 학습을 진행한 후, 최종적으로 실제 입력한 응답을 바탕으로 사용자의 MBTI를 잘 예측하는지 실험하였다. 사용자가 주어진 질문에 대답하면, Konlpy의 mecab을 이용해 응답을 형태소 단위로 분해하며 전처리 과정을 거친다. 전처리 된 데이터는 모델에 입력해 예측값을 출력한다. 예측값이 0.5 이상이면 사용자의 성격 유형을 E 유형으로 판단하고, 0.5 미만이면 I 유형으로 판단한다. 이와 함께 어느 정도로 해당 성격 유형에 가까운지 출력한다. 만일 예측값이 0.1564라면, 이는 사용자의 MBTI E/I 유형이 84.36% 정도로 I 유형임을 의미한다.

Ⅳ. 실험결과

본 연구의 목표는 MBTI의 E와 I 유형을 분류하는 것이다. 사전학습된 임베딩 벡터의 사용 여부를 기준으로 FastText 임베딩 벡터와 결합한 Fa-LSTM, Fa-BLSTM 모델을 사용하였다. 추가로 Bi-LSTM에서는 학습 데이터에서 extracted word을 사용해 학습한 Fa-extr-BLSTM과 whole word을 사용한 Fa-whl-BLSTM에 따른 성능 비교를 진행했다.

그 외에 과적합을 방지하기 위한 전략을 적용했다. 초기 10 epoch 동안에는 Adam 최적화 함수의 학습률을 0.0001로 낮추어 안정적인 학습이 이루어질 수 있도록 하였다. 이후의 epoch에서는 기본 학습률로 학습을 진행하였다.

또한, 모델에는 드롭아웃, 배치 정규화, 그리고 Leaky ReLU 활성화 함수를 사용하였다. 학습 과정 중에 Validation 데이터의 loss 값이 증가하는 경우, Early Stopping을 활용하여 학습을 중단해주었다. 이러한 접근법들을 통해 모델의 성능을 최적화해 나갔다. (표 2)에서는 5개의 각 모델에 대한 정확도를 나타낸다.

(표 2) 분류 모델별 성능 비교

Model name	Train	Validation	Test
Simple-LSTM	95.10%	93.60%	92.11%
Fa-LSTM	97.50%	93.30%	86.84%
Simple-BLSTM	**98.00%**	94.40%	89.29%
Fa-extr-BLSTM	95.50%	93.60%	**94.74%**
Fa-whl-BLSTM	**98.00%**	95.30%	**94.74%**

제시된 모델들은 Train 및 Validation 과정에서 90% 이상의 정확성을 보였다. 이러한 높은 성능의 원인은 두 가지로 볼 수 있다. 첫째로, E와 I 유형을 구분하는 학습 데이터 셋은 주로 유사한 문장구조를 가지기 때문이다. 대부분의 질문이 성격 유형을 구분하는 특징적인 질문들이기 때문에, 답변도 비슷한 패턴을 보이는 경향이 있다. 둘째로, MBTI의 각 유형은 서로가 상반되는 유형이다.

'외향'과 '내향'이라는 명확한 특징은 모델 학습에서 강력한 분류 기준을 제공한다. 따라서 이러한 특징을 기반으로 모델은 각 유형을 효과적으로 구분할 수 있다.

그러나 단순히 정확성만으로 모델의 성능을 판단하는 것은 한계가 있다. 이에 모델들에 대한 세부적인 능력을 검증하기 위해 단어와 문장에 대한 성격 유형 예측을 추가로 시행하였다.

(표 3) 성격 유형 분류를 위한 대표적인 단어와 문장 예시

Classification Criteria	Word/Sentence	Label	case Identifier
1. Clear Distinction Word	혼자	I	**case 1.1**
	친구	E	**case 1.2**
2. Neutral Word	사람	Neutral	**case 2**
3. Classification Indicator Sentence	친한 친구와만 노는 편이다.	I	**case 3.1**
	많은 친구와만 노는 편이다.	E	**case 3.2**
4. User-defined Sentence	주목받은 후 즉시는 살짝 부끄럽지만, 나름 주목받는 것을 좋아합니다.	E	**case 4.1**
	예기치 못한 상황이면 살짝 당황스러움..	I	**case 4.2**

(표 3)에서는 모델의 성능 평가를 위한 다양한 텍스트 데이터를 제시한다. case 1은 E 유형과 I 유형을 명확히 구분하는 단어를 기반으로 한다. case 2는 E와 I 유형 사이에 있는 중립적인 단어를 선정하였다. 예를 들어 '사람'이라는 단어는 문맥에 따라 다르게 해석될 수 있다. '사람들을 만나는 것을 즐긴다.'라는 문장은 E 유형을 나타내지만, '사람들을 만나는 것을 꺼린다.'는 I 유형을 더 잘 반영하기 때문이다. case 3은 case 1에 따른 단어를 문장으로 확장하여 제시한다. case 4에서는 실제 사용자가 작성한 원본 문장을 사용했다. 여기에는 '당황스러움..'과 같은 정제되지 않은 문장구조도 포함된다. (표 3)에서 제시된 단어와 문장을 통해 각 모델의 MBTI E/I 유형 예측 능력을 평가하였다.

(표 4) 모델별 성격 유형 예측 성능 분석 - LSTM

case	y	model: Simple-LSTM Pred	Simple-LSTM $\hat{y}$	Fa-LSTM Pred	Fa-LSTM $\hat{y}$
case 1.1	I	80.47%	**I**	95.67%	**I**
case 1.2	E	78.10%	**E**	97.18%	**E**
case 2	(N)	**52.85%**	I	94.65%	E
case 3.1	I	61.89%	E	97.66%	**I**
case 3.2	E	83.60%	I	74.04%	**E**
case 4.1	E	97.79%	I	75.82%	I
case 4.2	I	84.09%	**I**	99.85%	**I**

(표 4)에서 1행의 Pred는 case에 대한 모델의 예측값을 의미하고, y는 각 Case의 실제 MBTI이며 $\hat{y}$는 해당 Case를 학습된 모델에 적용했을 때의 예측 MBTI 결과를 표시한다. 2열의 (N)은 Neutral word를 의미한다. LSTM 기반의 모델 중 Simple-LSTM은 제시된 7개의 예시 중 4개를 올바르게 예측했고, Fa-LSTM은 5개를 정확하게 예측하였다. 이를 보아 사전학습된 임베딩 벡터를 사용한 Fa-LSTM이 Simple-LSTM에 비해 E/I 성격 유형을 잘 판단한다는 것을 알 수 있다.

(표 5) 모델별 성격 유형 예측 성능 분석 - Bi-LSTM

case	y	model: Simple-BLSTM Pred	Simple-BLSTM $\hat{y}$	Fa-extr-BLSTM Pred	Fa-extr-BLSTM $\hat{y}$	Fa-whl-BLSTM Pred	Fa-whl-BLSTM $\hat{y}$
case 1.1	I	73.15%	**I**	91.37%	**I**	80.59%	**I**
case 1.2	E	93.99%	**E**	99.33%	**E**	90.54%	**E**
case 2	(N)	**70.57%**	E	97.31%	E	**61.94%**	E
case 3.1	I	72.75%	**I**	66.29%	**I**	58.92%	**I**
case 3.2	E	99.87%	**E**	98.79%	**E**	98.56%	**E**
case 4.1	E	93.76%	I	70.19%	**E**	91.92%	**E**
case 4.2	I	99.87%	I	98.83%	**I**	68.57%	**I**

(표 5)에 따르면 Bi-LSTM 기반 모델들은 LSTM보다 성능이 우수한 것을 확인할 수 있다. 세 모델 모두 대부분의 case를 정확하게 예측하였다. 하지만 비교적 예측이 어려운 case 4에 대해서는 Simple-BLSTM이 2개 중 1개만 정답을 맞혔고, 사전 학습된 임베딩 벡터를 이용한 Fa-ext-BLSTM와 Fa-whl-BLSTM은 모두 정확하게 예측했다.

case 2의 중립단어를 예측할 때는 예측값이 50%에 가까울수록 잘 예측하는 것이다. 위에서 정의한 '사람'이란 용어는 E, I 유형 모두 사용할 수 있는 단어이기 때문이다. case 2에 대해서 Simple-BLSTM과 Fa-whl-BLSTM 모델은 중립단어에 대해 거의 50%에 가까운 예측 성능을 보였으나, Fa-ext-BLSTM은 97.31% 정도로 E 유형에 가까운 예측을 하였다. 이는 extracted word에 대해서만 학습한 Fa-ext-BLSTM 모델은 중립단어나 모호한 단어와 같이 문맥에 의해 좌우되는 단어 잘 반영하지 못한다는 것을 알 수 있다.

종합적으로 Fa-whl-BLSTM 모델은 모든 단어에 대해 정답을 맞히며 좋은 예측 성능을 보여주었다. 이를 통해 네 가지 주요한 결론을 도출하였다. 첫째, 이전 연구 [2]는 크롤링을 통한 데이터 수집 방법을 사용했으나, 본 연구에서는 GPT를 활용해 양질의 데이터를 생성해 학습하였다. 데이터가 양이 적었음에도 유의미한 결과가 나타난 것으로 보아, 데이터의 질이 양보다 중요하다는 것을 시사할 수 있다. 둘째, 학습 시 사전에 대량으로 학습된 임베딩 벡터를 가중치로 사용하는 것이 유리하다. 셋째, 기존의 연구대로 LSTM보다는 Bi-LSTM을 사용했을 때 예측 능력이 향상된다.[10] 마지막으로 Bi-LSTM은 whole word로 학습하는 것이 extracted word로 학습하는 것보다 문맥 정보를 더욱 잘 파악하는 데 있어 효과적이었다.

Ⅴ. 결론

본 논문은 MBTI의 E/I preference를 대상으로 E와 I 유형 분류

에 LSTM과 Bi-LSTM 기법을 적용하여 분석하였다. 각 유형의 데이터는 GPT를 이용해 고품질의 데이터를 생성하였고, 자연어 처리에서의 증강 기법을 이용해 데이터의 양을 늘려주었다. 학습 과정에서는 내부 파라미터의 조정을 통해 성능을 조절하였고, 한국어로 사전학습된 임베딩 벡터를 사용해 수집한 데이터가 한국어의 다양한 의미적·구문적 정보를 학습할 수 있도록 하였다. 그 뒤 기존 모델과 개선한 모델까지 총 5가지 모델을 제안하여 다양한 환경에서 학습을 시도하였다. 그 결과, 약 1,000개의 데이터만을 수집했음에도 불구하고 대부분의 모델이 90%이상의 정확성을 보였으며, (표 3)의 실제 성격 유형도 대부분의 case에 대하여 잘 예측했다. 결과적으로 제안된 5개의 모델 중, FastText를 활용해 whole word에 대해서 학습한 Fa-whl-BLSTM 모델이 특히 좋은 예측 성능을 보였다.

하지만 해당 실험에서도 일부 한계가 존재한다. LSTM 기반의 모델은 학습 데이터에 없는 새로운 단어인 OOV(Out Of Vocabulary)에 대해서는 예측하지 못한다. 만일, 사용자가 학습되지 않은 단어를 포함하여 응답을 작성한다면, 해당 단어들은 모두 OOV 처리가 되어 예측 성능에 영향을 줄 수 있다. 이를 극복하기 위해 더 발전된 모델인 Transformer와 단어를 더 작은 단위로 나누어서 학습하는 Subword Tokenization 기법을 이용하여 위의 문제를 해결하고, 복잡한 단어 간의 관계를 파악할 수 있다. 후속 연구 주제로 앞서 언급한 문제점을 보완하고, 더욱 정교한 MBTI 유형 예측 모델을 개발하고자 한다.

ACKNOWLEDGMENT

*교신저자, 서경대학교 컴퓨터공학과,

E-mail : statsr@skuniv.ac.kr, ywcho@skuniv.ac.kr

이 성과는 정부(과학기술정보통신부, 교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2019M3E7A1113102).

참 고 문 헌

[1] Isabel Briggs Myers, Peter B. Myers, 『Gifts Differing: Understanding Personality Type』, CPP, 2010.

[2] 김정민, 박지민, 이로운, 조서원, 심재형, “딥러닝 기반의 MBTI 성격유형 분류 연구”, 한국통신학회 하계종합학술대회 논문집, p. 1,740-1,741, 2022.

[3] Cui, Brandon and Calvin Qi., “Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction.”, 2017.

[4] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, p. 1735-1780, 1997.

[5] Graves, Alex, and Jürgen Schmidhuber., “Framewise phoneme classification with bidirectional LSTM and other neural network architectures.”, Neural networks 18, p. 602-610, 2005.

[6] NERIS Analytics Limited, “16-personalities”, https://www.16personalities.com/ko, 2023.

[7] Wei, Jason and Kai Zou., “EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks.”, Conference on Empirical Methods in Natural Language Processing, 2019.

[8] Park, E. L., Cho, S, “KoNLPy: Korean natural language processing in Python.”, In: Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology. Vol. 6, p. 133-136, 2014.

[9] Bojanowski, Piotr, et al., "Enriching word vectors with subword information.", Transactions of the association for computational linguistics 5, 135-146, 2017.

[10] Shewalkar, Apeksha, Nyavanandi, Deepika and Ludwig, Simone A.. "Performance Evaluation of Deep Neural Networks Applied to Speech Recognition: RNN, LSTM and GRU", Journal of Artificial Intelligence and Soft Computing Research, vol.9, no.4, 2019, p. 235-245.