

LSTM과 Bi-LSTM을 이용한 개인 성향분석

Personality Analysis using LSTM and Bi-LSTM

Ryu Ha Yeong, Moon Seong Uk,
Cho Young Wan⁺, Kim Jae Hyun⁺

Department of Computer Engineering Seokyeong University
+Corresponding Author (E-mail: ywcho@skuniv.ac.kr, statsr@skuniv.ac.kr)



요 약

- 본 논문은 딥러닝 기법 **LSTM**과 **Bi-LSTM**을 이용해 **MBTI의 E와 I 성격 유형**을 다양한 구조로 학습하고, 그에 대한 **성능 평가 및 성격 유형을 분류하는 모델**을 제안한다.

연구 배경

- 심리분석 분야에서는 자연어 처리를 기반으로 한 인공지능 기법이 새롭게 주목받고 있다. 과거의 인공지능은 긍정 및 부정 분류와 같은 한정된 문제만 해결 가능했지만, 기술이 발전함에 따라 고도화된 언어모델을 통해 사람의 심리와 성향까지 탐색할 수 있게 되었다. 최근, 사용자의 선호 경향을 기반으로 16가지 성격으로 분류하는 **성격 유형 분석 도구인 MBTI**가 등장하여 인간의 심리 이해에 큰 지표를 제시하고 있다.
- 전통적인 MBTI 검사 방식**은 사용자가 객관식 질문에 응답하도록 설계되어 있어 **인간의 복잡한 심리를 묘사하기 어렵지만**, 사용자가 주관식으로 자신의 생각과 감정을 표현할 수 있다면 **MBTI의 성격 유형을 더욱 정밀하게 파악할 수 있을 것이다**.
- 이에 MBTI 네 가지성격 유형 중 '**EI preference (Extraversion or Introversion, 이하 유형 E/유형 I)**'을 대상으로, 자연어 처리의 기본적인 모델인 **LSTM** 및 **Bi-LSTM**을 이용해 학습을 진행한다. 학습시에는 사전 학습된 임베딩 벡터를 사용한 것과 이를 사용하지 않은 경우와의 차이를 비교하고, 학습 데이터 선택 범위를 적용하여 그에 따른 성능 변화를 분석하였다.

실험 방법

- 데이터 수집에 앞서, 16 Personalities와 같은 대중적인 검사에서 사용되는 E와 I유형 질문들을 수집하고 분석하여 총 **20개의 질문**을 새롭게 제작하였다.
- 위의 질문에 대한 **기초 응답**은 각 질문에 대해 대략 **30개씩** 작성했다. 이후 **GPT-4**를 이용해 E와 I 유형의 특징이 분명하게 드러나는 약 **70개의 모범답안**을 생성하였다. **앞서 제작한 응답을 바탕으로 KoGPT-2 모델을 기초 응답과 모범답안을 결합한 데이터 셋으로 fine-tuning한 후**, 응답 데이터의 다양성을 확장하여 최종적으로 총 **1,375개의 질문-응답 데이터 셋**을 만들었다.
- 데이터 과적합을 방지하기 위해서, 수집한 데이터 셋 중 **1,000개를 Data Augmentation** 과정을 거쳐 각 응답 데이터를 **3개의 문장으로 확장하였다**. 이후 중복을 제거하여 최종적으로 총 **2,965개의 데이터를 학습에 사용**하였다. 나머지 **375개**는 Validation set으로 할당하였고, 모델의 Test 성능은 사람들이 자유롭게 작성한 **38개**의 응답을 수집하여 평가하였다.
- 위의 데이터를 바탕으로 LSTM과 Bi-LSTM 중 MBTI 성향 분석에 더 효과적인 딥러닝 모델을 제안하는 것을 목표로 다양한 접근법을 시도하였다.
 - 첫 번째로, 임베딩 레이어의 선택에서 **기본 임베딩과 사전학습된 FastText 임베딩 중 어느 것이 더 효과적인지를 평가**하였다. FastText 임베딩은 3만 개의 한국어 단어로 사전에 학습된 벡터(**FastText-KO**)를 활용한다.
 - 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 추출한 extracted word와 추출하지 않은 전체 문장 데이터인 whole word**에 따른 성능 비교를 진행했다.
- ✓ 본 연구에서는 위의 두 가지 기준을 바탕으로 **LSTM**과 **Bi-LSTM**을 사용하여 총 **5가지 조합의 모델**을 실험하였다.

실험 결과

- 모델의 정확성 측면에서 제시된 모델들은 Train 및 Validation 과정에서 **90% 이상의 정확성**을 보였다. E와 I 유형을 구분하는 학습 데이터 셋은 주로 유사한 문장구조를 가지고, MBTI의 유형이 '외향'과 '내향'이라는 명확한 특징을 가지고 있어 높은 성능을 보였다.

| Classification model | Use of FastText | Training Data Scope | Model name | Train | Validation | Test |
|----------------------|-----------------|---------------------|---------------|---------------|------------|---------------|
| LSTM | X | extracted word | Simple-LSTM | 95.10% | 93.60% | 92.11% |
| LSTM | O | extracted word | Fa-LSTM | 97.50% | 93.30% | 86.84% |
| Bi-LSTM | X | whole word | Simple-BLSTM | 98.00% | 94.40% | 89.29% |
| Bi-LSTM | O | extracted word | Fa-extr-BLSTM | 95.50% | 93.60% | 94.74% |
| Bi-LSTM | O | whole word | Fa-whl-BLSTM | 98.00% | 95.30% | 94.74% |

< 분류 모델별 학습 방식에 따른 성능 비교 >

- LSTM 기반의 모델 중** Simple-LSTM은 제시된 7개의 예시 중 4개를 올바르게 예측했고, Fa-LSTM은 5개를 정확하게 예측하였다. 이를 보아 사전 학습된 임베딩 벡터를 사용한 **Fa-LSTM이 Simple-LSTM에 비해 E/I 성격 유형을 잘 판단한다는 것을 알 수 있다**.
- Bi-LSTM 기반의 세 모델** 모두 대부분의 case를 정확하게 예측하였다. 하지만 문장이 정제되지 않아 비교적 예측이 어려운 **case 4**에 대해서는 Simple-BLSTM이 2개 중 1개만 정답을 맞췄고, **사전 학습된 임베딩 벡터를 이용한 Fa-extr-BLSTM와 Fa-whl-BLSTM은 모두 정확하게 예측했다**. **case 2의 중립단어**는 예측값이 50%에 가까울수록 잘 예측하는 것인데, extracted word에 대해서만 학습한 Fa-extr-BLSTM 모델에 비해 **전체 단어에 의해 학습한 Simple-BLSTM과 Fa-whl-BLSTM이 50%에 가까운 예측 성능을 보이는 것을 확인할 수 있다**.
- 결과적으로 **GPT를 활용한 양질의 데이터 생성**과 학습 시 **사전에 대량으로 학습된 임베딩 벡터를 가중치로 사용하는 것**이 성능 향상에 유리한 점, **Bi-LSTM**이 LSTM보다 예측 능력이 향상된 점, 그리고 **Bi-LSTM은 whole word로 학습하는 것이 효과적**이라는 점을 시사할 수 있다.

| case | | | model | | | | | | | | | |
|----------|--|---|---------------|-----------|---------|-----------|---------------|-----------|---------------|-----------|---------------|-----------|
| | | | Simple-LSTM | | Fa-LSTM | | Simple-BLSTM | | Fa-extr-BLSTM | | Fa-whl-BLSTM | |
| | | | Pred | Pred MBTI | Pred | Pred MBTI | Pred | Pred MBTI | Pred | Pred MBTI | Pred | Pred MBTI |
| case 1.1 | 혼자 | I | 80.47% | I | 95.67% | I | 73.15% | I | 91.37% | I | 80.59% | I |
| case 1.2 | 친구 | E | 78.10% | E | 97.18% | E | 93.99% | E | 99.33% | E | 90.54% | E |
| case 2 | 사람 (N) | | 52.85% | I | 94.65% | E | 70.57% | E | 97.31% | E | 61.94% | E |
| case 3.1 | 친한 친구와만 노는 편이다. | I | 61.89% | E | 97.66% | I | 72.75% | I | 66.29% | I | 58.92% | I |
| case 3.2 | 많은 친구와만 노는 편이다. | E | 83.60% | I | 74.04% | E | 99.87% | E | 98.79% | E | 98.56% | E |
| case 4.1 | 주목받은 후 축시는 살짝 부끄럽지만, 나를 주목받는 것을 좋아합니다. | E | 97.79% | I | 75.82% | I | 93.76% | I | 70.19% | E | 91.92% | E |
| case 4.2 | 예기치 못한 상황이면 살짝 당황스러움.. | I | 84.09% | I | 99.85% | I | 99.87% | I | 98.83% | I | 68.57% | I |

< 모델별 성격 유형 예측 성능 분석 >

결 론

- 제안된 5개의 모델 중, FastText를 활용해 whole word에 대해서 Bi-LSTM으로 학습한 Fa-whl-BLSTM 모델이 특히 좋은 예측 성능을 보였다.
- 하지만 LSTM 기반의 모델은 **학습 데이터에 없는 새로운 단어인 OOV(Out Of Vocabulary)에 대한 예측 능력이 제한적이다**. 사용자가 학습되지 않은 단어를 포함하여 응답을 작성한다면 해당 단어들은 모두 OOV로 처리되어 예측 성능에 부정적인 영향을 줄 수 있지만, 더 발전된 모델인 Transformer 아키텍처와 단어를 더 작은 단위로 나누어서 학습하는 Subword Tokenization 기법으로 한계를 해결하고자 한다.