트랜스포머 기반 MBTI 성격 유형 분류 연구 : 소셜 네트워크 서비스 데이터를 중심으로

정재준O, 임희석*
O고려대학교 컴퓨터정보통신대학원 인공지능융합학과,
*고려대학교 컴퓨터정보통신대학원
e-mail: thezombies@korea.ac.krO, limhseok@korea.ac.kr*

Research on Transformer-Based Approaches for MBTI Classification Using Social Network Service Data

Jae-Joon JungO, Heui-Seok Lim*
ODept. of Applied Artificial Intelligence, Graduate School of Computer & Information Technology, Korea University,
*Graduate School of Computer & Information Technology, Korea University

● 요 약 ●

본 논문은 소셜 네트워크 이용자의 텍스트 데이터를 대상으로, 트랜스포머 계열의 언어모델을 전이학습해 이용자의 MBTI 성격 유형을 분류한 국내 첫 연구이다. Kaggle MBTI Dataset을 대상으로 RoBERTa Distill, DeBERTa-V3 등의 사전 학습모델로 전이학습을 해, MBTI E/I, N/S, T/F, J/P 네 유형에 대한 분류의 평균 정확도는 87.9181, 평균 F-1 Score는 87.58를 도출했다. 해외 연구의 State-of-the-art보다 네 유형에 대한 F1-Score 표준편차를 50.1% 낮춰, 유형별 더 고른 분류 성과를 보였다. 또, Twitter, Reddit과 같은 글로벌 소셜 네트워크 서비스의 텍스트 데이터를 추가로 분류, 트랜스포머 기반의 MBTI 분류 방법론을 확장했다.

키워드: 트랜스포머(Transformer), 성격 탐지(Personality Detection), MBTI (Myers-Briggs Type Indicaton)

I. Introduction

자연어 처리 기반의 자동 성격 탐지 (Automated NLP-Based Personality Detection)는 자연어 처리 분야의 발전과 함께 성장한 분야다. 초기 나이브 베이즈부터 최근의 PLM (Pretrained Language Model)까지, 텍스트로 성격을 분류하고자 하는 연구는 발전해 기업 채용에도 활발히 응용되고 있다[1].

COVID-19 이후, 사람들과의 상호작용을 통해 찾던 정체성을 MBTI 유형 검사로 대신하는 것이 한국 사회의 밈이 되었기에[2], 텍스트 기반 MBTI 성격 분류 연구가 국내에 필요한 적시가 될 수 있을 것이다.

II. Preliminaries

1. Related works

1.1 국내 동향

한국 사회에서의 MBTI 유행은, 최근 자연어 처리 (이하 NLP)와 딥러닝을 통한 성격 감지에 대한 연구에도 영향을 주었다. 그러나 최근 다양한 자연어처리 분야에 성공적으로 활용되는 트랜스포머 계열의 모델이나, 전이학습 방법론이 적용된 연구는 부재했다.

'22년 6월, 한글 텍스트에 딥러닝을 적용해, MBTI 성격 유형을 분류한 국내 최초의 연구가 있었다. 논문 [3]은 MBTI 심리 온라인 커뮤니티의 글을 라벨링해 3.3만 개의 데이터셋를 정리했다. 이를 LSTM을 활용, MBTI를 구성하는 4개의 카테고리를 이진 분류해, E/I 유형 65.33%, N/S 유형 66.88%, T/F 유형 66.07%, J/P 유형 70.32%이라는 유형별 정확도를 도출했다. 또, 논문 [4]는 직무를

추천하는 알고리즘 에, 구직자의 MBTI 성격 유형을 분류하는 자연어 처리 모델을 제안했다. SDG, NB, KNN 등의 알고리즘을 활용했고, 이는 각각 67%, 21%, 28%, 69%의 정확도를 보였다.

1.2 국제 동향

국내와는 달리, '21년 12월, [5]는 트랜스포머 계열 모델 RoBERTa Distil 등을 성격 감지 분야에 성공적으로 적용했다. 이는 Kaggle에 공개된 성격 유형 분류 데이터 (이하 Kaggle MBTI)를 대상으로 88.63%의 정확도 및 88.97%의 F1-Score를 기록, LSTM 기반[6]의 최고기록을 경신했다. 1) MBTI 유형별 정확도 및 F1-Score의 일부 높은 편차와 2) Kaggle MBTI 데이터 외에는 적용하지 않았다는 한계는 있었다.

이후, 논문 [7]은 트랜스포머 계열의 언어모델 전이학습에만 의존하지 않고, 언어심리학을 기반의 특성 선택 (Feature Selection) 방법론을 제안했다. Improved Distiributed Gray Wolf Optimization (IDGWOFS)는 타 최적화 방법론에 비해, Kaggle MBTI에 대한 평균 F1-Score 5.8%라는 유의미한 상회를 보였다. 그러나, 평균 F1-Score는 78.44%로 [5]의 기록을 경신하지는 못했다.

III. The Proposed Scheme

1. Methodology

NLP에서의 성격 감지 분야 중, 국내 연구에 부재했던, 트랜스포머 계열 모델의 활용을 논문 [5]를 바탕으로 실험했다. [5]의 state-of-the-art (이하 SOTA)를 경신하기 위해, SST-2 [8] 감성 분류 문제에 RoBERTa 를 상회한 DeBERTa 등, 최근의 모델을 추가 선정했다. 또, [5]의 두 한계를 해결하기 위해, 1) MBTI의 E/I, N/S, T/F, J/P 유형별 모델 학습과 2) Twitter, Reddit 등으로부터 수집된 MBTI 분류 데이터셋에[9,10]도 적용했다. 토크나이즈부터 모델 평가까지 HuggingFace [11]의 오픈소스 라이브러리인 transformers, tokenizer, trainer를 활용했다. 실험의 상세는 다음과 같다.

2. Experiments

2.1 Dataset

Table 1. MBTI Datasets

기준 \ 명칭	Kaggle MBTI	Twitter MBTI	Reddit MBTI
기존 연구 적용	Yes	No	No
데이터 크기 (행)	8,675	7,811	10,607
텍스트 평균 길이	1,270	1,305	500
MBTI 표준편차	576.70	358.22	8162.51
영어 비중	99.98%	95.59%	99.89%
언어 갯수	3	13	9

본 연구에 사용한 데이터셋은 3개다. [5, 6]에서 실험한 Kaggle MBTI 외, 소셜 네트워크인 Twitter와 Reddit에서 수집된 데이터셋을 추가했다. (이하 트위터 MBTI 및 레딧 MBTI) Kaggle MBTI는 작성자가 정의한 MBTI 유형으로, 트위터 MBTI는 작성자가 글에 명시한 MBTI 유형으로 라벨링 했다. 레딧 MBTI는 두 방법을 함께 활용했다.

각 데이터셋의 크기는 7.8천 행 ~ 10.6천 행으로, 유사한 실험 조건을 위해 레딧 MBTI는 원본의 10%만을 무작위 추출해 데이터셋으로 활용했다. 영어 비중과 언어 갯수는 fasttext[12]를 통해 도출했다.

2.2 Model

연구 [5]에서 가장 성과가 좋았던 RoBERTa Distil [13], DistilBERT [14], XLNet [15]과 더불어, 본 연구에는 DeBERTa-V3 [16]와 Twitter-RoBERTa-base for Sentiment Analysis - UPDATED (2022) (이하 Twitter-RoBERTa) [17]를 추가로 실험했다.

DeBERTa-V3는 SST-2 감성 분류 문제에서 RoBERTa Large 보다 더 높은 정확도를 기록한 V2 [17]의 다음 모델이다. Twitter-RoBERTa는 124백만 개의 트윗을 학습, 감성 분류를 위해 파인튜닝한 모델이다.

2.3 Hyperparameters and Setting

Table 2. Hyperparameters

Hyperparameter	Value
Sequence Max size	512
Number of Epochs	5
Batch Size	8
Learning Rate	2E-05
Initial Seed	7

주요 파라미터는 Table 2.와 같다. 학습과 평가는 Google Colab Pro+에서 진행, RAM은 최대 49.09 GB 사용 가능했으며, GPU는 Tesla T4를 활용했다.

3. Results

Table 2. Results : Accuracy & F1 Score on Kaggle MBTI

Models	Accuracy %		F1 Score %	
	SOTA	This Paper	SOTA	This Paper
RoBERTa Distil	**88.6286**	87.8893	**88.9716**	87.58
DistilBERT - Base	88.6000	86.1304	88.9697	85.6948
XLNet - Base	87.8572	86.9089	87.9523	86.5211
DeBERTa-V3-Base	N/A	87.0242	N/A	86.6712
Twitter-Roberta	N/A	87.9181	N/A	87.5692

Table 3. Results : Std. of F1 Score on Kaggle MBTI

Models	Std. of F1 Score	
	SOTA	This Paper
RoBERTa Distil	6.6575	3.3213
DistilBERT - Base	**6.4263**	3.9278
XLNet - Base	7.6959	4.0842
DeBERTa-V3-Base	N/A	5.1694
Twitter-Roberta	N/A	3.2259

Kaggle MBTI 데이터셋에 대한 실험 결과, [5]의 SOTA를 상회하는 정확도 및 F1-Score를 기록하지는 못했다. 정확도의 경우, 본 연구에는 Twitter-RoBERTa가 87.9181로 가장 높았고, [5]의 SOTA와 0.7105%p 차이였다. F-1 Score의 경우, 본 연구도 RoBERTa Distil이 87.9181로 가장 높았고, [5]의 SOTA와 1.3916%p 차이였다.

MBTI의 E/I, N/S, T/F, J/P에 대한 평균 정확도 및 F1-Score는 연구 [5] 보다 낮았으나, 네 유형에 대한 F1-Score의 표준편차는 본 연구가 3.2259로 연구 [5]의 6.4263보다 50.1% 낮았다. 즉, 본 연구의 방법론 - E/I, N/S, T/F, J/P 유형을 예측하는 각 모델을 학습 - 이 기존 연구 [5]의 한계인, 유형 마다 상대적으로 분류 성과의 편차가 있다는 점을 완화하는데 유의미하다고 볼 수 있다.

Table 4. Results : Accuracy & F1 Score on the others

Models	MBTI Twitter		MBTI Reddit	
	Accuracy	F1 Score	Accuracy	F1 Score
RoBERTa Distil	73.5916	**71.3497**	91.7531	91.7533
DistilBERT - Base	72.6953	70.6053	91.5881	91.5123
XLNet - Base	72.1511	69.1574	91.1169	90.5681
DeBERTa-V3-Base	**73.7196**	69.7605	90.9991	90.8809
Twitter-Roberta	73.4635	71.2889	**92.2244**	**92.1609**

Kaggle MBTI 외 트위터 MBTI와 레딧 MBTI에도 같은 방법론으로 적용하였다. 그 결과, 트위터 MBTI에는 DeBERTa-V3-Base가 가장 높은 정확도를 보였고 (73.7196%), F1-Score는 RoBERTa Distil이 가장 높았다. (71.3497%) 레딧 MBTI는 정확도와 F1 Score 모두 Twitter-Roberta가 가장 높았다. (92.2244%, 92.1609%)

세 데이터셋 중, 정확도 및 F1-Score가 평균적으로 높은 것은 레딧 MBTI, Kaggle MBTI, 트위터 MBTI 순이었다. 이는 Table 1에서 보듯, 레딧 MBTI가 16개 MBTI 유형에 대한 표준편차가 가장 높아 분류에 유리했고, 데이터셋이 가장 많고 평균 길이가 가장 짧아 학습에도 유리했으리라 보여진다.

Ⅳ. Conclusions

본 연구는 소셜 네트워크 이용자의 텍스트를 대상으로 트랜스포머 계열의 언어모델을 전이학습해 MBTI 성격 유형을 분류한 국내 첫 연구다. Kaggle MBTI를 대상으로 평균 정확도는 87.9181, 평균 F-1 Score는 87.58로 해외 연구 SOTA보다 낮았으나 E/I, N/S, T/F, J/P 네 유형별 정확도와 F-1의 표준편차는 SOTA보다 50.1% 낮아 수적으로 불균형한 유형을 분류하는데에 더 유리함을 보였다. 또, 트랜스포머 계열의 MBTI 유형 분류에 대한 기존 연구들에서 다루지 않았던 두 모델 DeBERTa-V3와 Twitter-RoBERTa를 트위터 그리고 레딧의 두 데이터셋에도 확장해 SOTA를 기록한 RoBERTa 모델과 근사한 결과를 보였다는 의의도 있다.

SOTA를 상회하기 위해, MBTI 유형간의 불균형 데이터에 대한 증강을 통한 Over-sampling이나 상대적으로 비중이 높은 유형을 무작위로 제외하는 Under-sampling의 데이터 정제 방법론을 적용하는 것도 후속 연구에 다룰 수 있을 것이다. 또, 영어 중심의 텍스트 데이터셋 외, 한글 텍스트 데이터셋에서의 적용도 추후 이뤄진다면, 한국 사회에서의 MBTI 성격 유형 분류의 긍정적인 효과 - 자신과 타인을 이해하는 도구 - 를 배가하는데에 도움이 될 것이다.

REFERENCES

[1] Perera,H. Costa,L. “Personality Classification of text through Machine learning and Deep learning A Review”, 2023.

[2] Youngju, L. “Korean society's perception of MBTI using big data“, A Review 2023 Journal of Learner-Centered Curriculum and Instruction Vol. 22, No. 17, pp. 797-809, 2022.

[3] J. Kim, J. Park, R. Lee, S. Cho, J. Sim, ”Deep Learning Based MBTI Personality Type Classification Study”, The Journal of Korean Institute of Communications and Information Sciences., pp.1,740-1,741, 2022.

[4] J. Kim, Y. Cho, “Design of MBTI Job Recommendation Algorithm Based on Deep Learning.”, Proceedings of the Korean Society of Computer Information Conference, Vol. 31, No. 1, pp.13-15, 2023.

[5] Vasquez, R. L., Ochoa-Luna, J., “Transformer-based Approaches for Personality Detection using the MBTI Model”, 2021 XLVII Latin American Computing Conference (CLEI), pp. 1-7, 2021.

[6] Yash Mehta et al., “Recent trends in deep learning based personality detection”, Artificial Intelligence Review, pp. 1-27, 2019.

[7] Hao, L. Chundong, W. Qingbo. H., “A novel personality detection method based on high-dimensional psycholinguistic features and improved distributed Gray Wolf Optimizer for feature selection”, Information Processing & Management, Vol. 60, No. 2, 2023.

[8] Wang et al., “GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding”, In Proceedings of the 2018 EMNLP Workshop

BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP. Association for Computational Linguistics, Brussels, Belgium, pp. 353-355, 2018.

[9] MBTI Personality Type Twitter Dataset, https://www.kaggle.com/datasets/mazlumi/mbti-personality-type-twitter-dataset

[10] MBTI Personality Types 500 Dataset, https://www.kaggle.com/datasets/zeyadkhalid/mbti-personality-types-500-dataset

[11] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., et al., “Huggingface’s transformers: State-of-the-art natural language processing”, arXiv preprint arXiv:1910.03771, 2019.

[12] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov, “Bag of Tricks for Efficient Text Classification”, A. Joulin, E. Grave, P. Bojanowski, T. Mikolov pp. 427-431, 2017.

[13] Yinhan Liu et al., “Roberta: A robustly optimized bert pretraining approach”, arXiv preprint arXiv:1907.11692, 2019.

[14] Victor S., Lysandre D., Julien C., Thomas W., “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter”, arXiv preprint arXiv:1910.01108, 2020.

[15] Zhilin Y. et al., “Xlnet: Gen- eralized autoregressive pretraining for language understanding.”, In Advances in neural information processing systems, pp. 5753-5763, 2019.

[16] Pengcheng He, Jianfeng Gao, Weizhu Chen, “” DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing“, arXiv preprint arXiv:2111.09543 , 2021.

[17] Francesco Barbieri, Luis Espinosa Anke, and Jose Camacho-Collados., “Xlm-t: Multilingual language models in twitter for sentiment analysis and beyond. ”, 2022 Language Resources and Evaluation Conference, LREC 2022, pp. 258–266, 2021.