

CNN을 활용한 MMTIC 성격분석 예측

김경열¹, 양영보², 김미라³, 박지수⁴, 김지희⁵

¹동국대학교 인공지능학과, 동국대학교 AI융합학부,

²국민대학교 데이터 사이언스학과

³호남대학교 교양융합학과

⁴전주대학교 컴퓨터공학과

⁵동국대학교 AI융합학부

2021121170@dgu.ac.kr, diddudgh79@kookmin.ac.kr, kim1004w@hanmail.net,

jisupark@jj.ac.kr, jihie.kim@dgu.edu

Prediction of MMTIC Personality Analysis using CNN

Kyungy-Yeul Kim¹, Cheol-Soo Kim², Mi-ra Kim³, Ji Su Park⁴, Jihie Kim¹

¹Dept. of Artificial Intelligence, Dongguk University

²Dept. of Data Science, Kookmin University

³Dept. of Liberal Arts Convergence, Honam University

⁴Dept. of Computer Science, Jonju University

요 약

청소년의 성격유형을 분석할 때 소셜미디어 데이터를 활용하여 텍스트 처리로 분석하는 연구는 많이 알려져 있다. 그러나 이미지를 사용하여 성격유형을 분석한 연구는 미비하다. 본 연구는 청소년의 발태그 그림검사로 표현된 이미지를 데이터로 사용하고, CNN을 활용하여 MMTIC의 16가지 청소년의 성격유형을 예측한다. 연구 대상은 중학교 재학생을 대상으로 한다. MMTIC에서 U-band를 제외한 340명의 학생으로 2012년 4월부터 2013년 3월까지 조사하였다. 연구 결과 CNN을 사용하였을 때 21.6% 예측율을 보였으며, CNN Ensemble을 적용하였을 때 23.1%로 2.5%가 증가한다.

1. 서 론

청소년의 성장 과정은 생물학적, 신체적, 인지적, 심리적, 정신적 변화가 일어나는 중요한 시기이다 [1]. 청소년은 현실에 대한 적응의 어려움으로 반항하고, 스스로 고민하며, 수업에 제대로 집중하지 못하여, 이로 인해 건강관리에도 문제가 생길 수 있다 [2]. 성격유형을 분석하면 청소년들은 자신의 특성과 성격을 알게 되고, 비슷한 고민을 하는 다른 사람을 이해할 수 있는데 도움을 준다. 또한 교수자들은 학생의 행동에 대한 특성을 이해하며 장점과 단점을 지도할 수 있기에 교수자와 학생과의 관계에 도움이 된다[3].

현재 청소년 성격 진단 도구에 사용되는 검사는 대부분 자기 보고형 질문지인 설문지로 되어 있다. 자기 보고형 설문지의 단점은 불성실한 응답과 자기 방어 등으로 인하여 정확도가 떨어지므로, 유용한 정보를 얻는 데 한계가 따를 수밖에 없다[4]. 이러한

문제점을 보완하기 위하여 투사검사를 시행한다. 투사는 피험자가 무의식 속에 있는 직접적 경험을 그림으로 표현하는 것이다. 투사검사는 정신과 영역에서 이상 성격을 파악하는 데 사용하고 다양하게 발전한다. 청소년의 그림에서는 청소년의 성향에 따라 특정한 패턴이 존재하며 그림의 특정 부분에서는 보편적이고 유사한 패턴이 나타난다[5]. 투사검사를 위하여 발태그 그림검사(Wartegg-Zeichentest)를 선정한다. 발태그 그림검사는 독일 심리학자인 에리히 발태그(Ehrig Wartegg)에 의해 창안되었다. 발태그 그림검사의 대상은 학령 전 유아에서부터 성인에 이르기까지 광범위하게 적용되고 있다.

발태그 그림 검사는 8가지 그림이 제시되는데 놀이의 요소가 내포되어 검사를 의식하지 않는 장점이 있다. 또한 검사의 결과는 청소년 성격유형 평가에 도움을 줄 수 있으며, 성격유형 예측은 경영에 실제 적용되는 심리학에서 잘 알려진 사용법이다.

청소년의 성격유형을 알아보기 위하여 청소년 성격유형을 측정하는데 잘 알려진 MMTIC (Murphy-Meisgeir Type Indicator for Children)을 사용하였으며, 성격유형을 예측하는 것이 건강관리에 유용하다는 것을 보여주었다[6]. 그러나 성격유형 평가는 설문조사와 투사법으로 평가할 수 있는 자격을 갖춘 전문가를 포함하기 때문에 결과가 나오기까지 시간이 걸리며, 비용도 발생한다[7].

본 연구는 청소년의 발테그 그림검사로 나타난 그림 이미지를 CNN을 활용하여 MMTIC의 16가지 성격유형을 예측한다. 이것은 딥러닝 기술을 이용하여 전문가를 포함하는 시간과 비용의 프로세스를 개선하였다. 특히 CNN (Convolutional Neural Network)은 다양한 이미지 데이터 분석과 성격을 감지하는 작업에서 우수한 성과를 나타낸 방법이다[8].

2. 관련연구

성격유형 예측은 경영에서 실제 적용되는 잘 알려진 사용법이다. 그리고 자동 성격 예측은 자연어 처리 및 사회 과학 커뮤니티에서 많은 관심을 받고 있다[9]. 또한 성격분석은 매년 거의 2백만 명이 경영에서 사업 목적으로 사용한다[10]. 심리학자인 카를 융(Carl Jung)은 심리 유형론에서 똑같은 지는 않더라도 분류할 수 있을 만큼의 유사성과 동질적인 특징으로 개인의 태도와 정신기능을 분류할 수 있다고 언급하였다. 이것은 청소년의 그림 중 특정 부분에서 형태적으로 유사한 패턴이 나타나는데 개개인의 독특한 유형들 가운데는 어느 정도 보편성이 내재되어 있기 때문이다[11].

MMTIC 모델에는 4개의 차원 또는 이분법이 있으며 각각은 상호 배타적인 두 범주로 구성된다. 예를 들어 맨 처음 나타나는 외향적(Extraversion) 대 내향적(Introversion)은 몇 가지 공통점을 찾아 집단적인 유형으로 분류할 수 있다. 일반적으로 성격유형은 MBTI(Myers-Briggs Type Indicator)로 나타낸다, 그러나 청소년들은 U-band (Undetermined band, 미결정 영역)가 존재하는데, 심리적 선호가 아직 발달하고 있어서 중간치(cut-point)에 가까운 점수를 분명한 유형으로 할당하여 범할 오류를 방지할 수 있는 MMTIC를 사용한다.

2.1 MMTIC의 속성분류

MMTIC은 다음과 같이 4가지 속성으로 분류할 수 있다.

첫째는 자신의 태도에 관한 방법의 속성분류를 나타낼 때 **E(Extraversion, 외향형)** -**U-I(Introversion, 내향형)**로 나타낸다.

둘째는 정보를 처리하는 방법을 속성분류를 나타낼 때 **S(Sensing, 감각)** -**U- N(INTuition, 직관)**으로 나타낸다.

셋째는 결정을 내리는 방법을 속성분류를 나타낼 때 **T(Thinking, 사고)** -**U- F(Feeling, 감정)**으로 나타낸다.

넷째는 자신을 외부에 제시하는 방법으로 분류를 나타낼 때 **J(Judging, 판단)** -**U- P(Perceiving, 인식)**로 나타낸다.

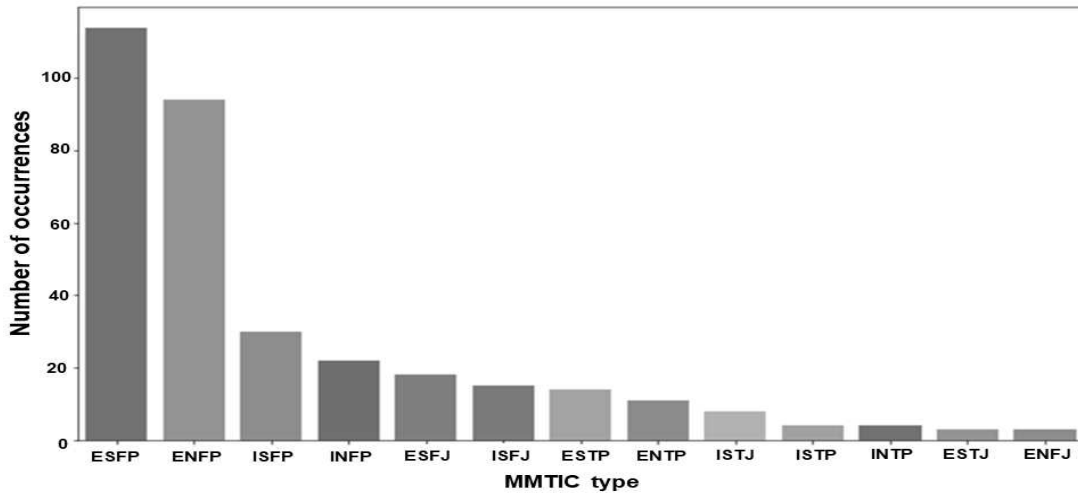
2.2 MMTIC 성취유형

MMTIC의 성격유형 16가지의 형태를 그림으로 나타내면 다음과 같다.



(그림 1) MMTIC 16 Personality Type Characteristics

본 연구는 청소년을 대상으로 발테그 그림 검사에 나타난 이미지를 활용하여 청소년의 성격유형을 예측하기 위한 목적이 있다. CNN은 이미지를 분석하고 성격을 감지하는 작업에 잘 알려져 있으며 CNN Ensemble을 사용하여 성능을 향상할 수 있다. 브랜든과 켈빈은 소셜미디어의 text를 기반으로 MMTIC 성격유형을 분류한 결과 Deep Learning의 결과 38%로 가장 높았다. 전체 정확도는 50% 미만으로 높지 않았지만 앞으로 더 풍부한 데이터와 기능을 통합하여 성능을 개선할 수 있을 것이다[12]. 다양한 데이터를 통한 성격특성 예측은 감성 컴퓨팅 분야에서 삶의 선택, 웰빙, 건강에 영향을 미치기 때문이다. 다중모드 딥러닝 기술은 신뢰할 수 있는 성격 예측을 하기 시작했지만, 성격 탐지를 위해



(그림 2) MMTC 13 Personality type Dataset

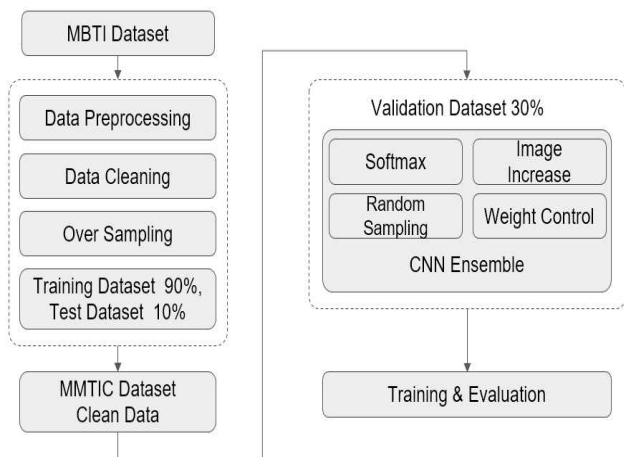
더 크고 정확하며, 다양한 데이터 세트가 필요하다 [13].

3. 연구 방법

연구의 대상은 중학교 813명을 대상으로 2012년 4월부터 2013년 3월까지 조사하였다. 16가지 성격유형을 나타내기 위하여 U-band를 제외한 340명의 학생의 13가지 성격유형을 나타낸다. 그림 2는 340명 학생의 13가지 성격유형의 분포이다.

데이터 세트는 상당히 편향되어 있으며 16가지 성격유형에 균일하게 분포되어 있지 않다. 예를 들어 ESFP는 114개(33.5%)로 가장 많았지만, ENFJ는 3개(0.8%)로 가장 적었으며, INFJ, INTJ, ENTJ의 성격유형은 나타나지 않았다.

연구 방법은 다음과 같은 흐름으로 한다.



(그림 3) Overview of the pipeline of the method

1) Data Preprocessing은 몇몇 발테그 그림 검사에서 나타난 이미지는 형태가 기존유형을 따르지 않

아 제거하였고, U-band를 제외한다.

2) Data Cleaning은 학번과 그림을 비워 놓은 데이터는 제외한다.

3) Over Sampling은 편향된 데이터 세트에서 상대적으로 적은 개수에 데이터를 증폭한다.

4) Softmax는 원본 데이터 세트를 CNN으로 다중 분류한다.

5) Image 증식은 원본 데이터 세트와 이미지 증식을 결합하여 CNN으로 다중 분류한다.

6) 랜덤 샘플링은 원본 데이터 세트와 랜덤 샘플링 데이터를 결합하여 CNN으로 다중 분류한다.

7) 가중치 조절은 원본 데이터 세트와 랜덤 샘플링 데이터를 결합 후 가중치 조절을 적용하여 CNN으로 다중 분류한다.

8) CNN Ensemble은 4),5),6),7)에 결괏값을 합산하여 Ensemble Voting을 활용하고 다중 분류한다.

4. 결과

본 논문은 청소년의 발테그 그림 검사에서 이미지를 활용하여 청소년의 성격유형을 예측한다. 이미지를 CNN으로 분석하였을 때 예측율은 21.6%였으며, ENFP, ESFP 위주로 예측하는 편향된 분류를 나타내었다. 그러나 이미지를 CNN Ensemble로 분석하였을 때 예측율은 23.1%로 높아졌을 뿐만 아니라, ESFP, ENFP를 포함한 ESTP, ESFJ, ISFP, ISTJ, ESTP를 예측하여 상대적으로 적은 데이터셋도 분류할 수 있는 개선되었다. 340명 학생들은 16개의 속성이 나타나지 않고 13의 성격유형만 나타난다. 이는 풍부하지 않은 데이터의 문제이며, 향후 더 풍부한 데이터를 확보하여 청소년이 그린 이

미지를 자동으로 분석하여 성격유형을 예측하는데 다양한 방법이 논의되어야 한다.

<표 2> Comparison of different methods for MMTIC classification

Approach	Classifier	The best result	
		Accuracy	True MMTIC type
Multiclass	Softmax	21.6%	ESFP, ENFP
	Image Increase	22.7%	ESFP, ENFP
	Softmax Random Sampling	20.5%	ESFP, ENFP, ESTP, ISFP
	Softmax Random Sampling Class Weight	18.7%	ESFP, ENFP, ISFP, ISTJ, ESTP
	Softmax		
	Ensemble	23.1%	ESFP, ENFP, ESTP, ESFJ, ISFP

5. 결론

청소년의 발테그 그림검사로 표현된 이미지를 CNN Ensemble을 활용하면 MMTIC의 16가지 성격유형 예측을 향상시킬 수 있으며, 적은 수의 성격유형도 예측할 수 있다. 본 연구는 데이터 세트 수가 소량으로 MMTIC을 MBTI 로 변형하여 더 많은 데이터를 확보하고, CNN 예측시스템을 개선하기 위하여 data augmentation 과 효과적인 이미지 생성 연구가 필요하다.

Acknowledgments:

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음" (IITP-2022-2020-0-01789)

참고문헌

[1] 김미라, & 김정필, "청소년의 발테그 그림검사 (WZT) 와 MMTIC 성격유형간의 관계", 미술치료연구, 21, 409-429, 2014.
 [2] 임지향, 안윤지, & 신지현, "몰입을 활용한 집단 미술치료가 고등학생의 학습동기와 학업적 자기효능감에 미치는 효과", 특수교육재활과학연구, 54(3), 427-448, 2015.
 [3] 이정희, "청소년의 성격특성과 '빛속의 사람' 그림 반응특성의 관계", 석사학위논문, 영남 대학교 환경보건대학원, 2010.

[4] 손현숙, "부모 및 또래에 대한 아동의 애착안정성과 교우관계문제", 석사학위논문, 진주교육대학교, 2008.
 [5] 노덕영, "아동의 그림표현과 행동특성에 관한 유형화 연구 -6, 7세 아동을 대상으로", 박사 학위논문, 대구대학교 디자인대학원, 2012.
 [6] M. Mitchell, K. Hollingshead, and G. Coppersmith, "Quantifying the language of schizophrenia in social media," in Proc. 2nd Workshop Comput. Linguistics Clin. Psychol., From Linguistic Signal Clin. Reality, pp. 11 - 20, 2015.
 [7] Cerkez, N., Vrdoljak, B., & Skansi, S., "A Method for MBTI Classification Based on Impact of Class Components", IEEE Access, 9, 146550-146567, 2021.
 [8] Kim, M. K., "Epileptic Seizure Detection Using CNN Ensemble Models Based on Overlapping Segments of EEG Signals", KIPS Transactions on Software and Data Engineering, 10(12), 587-594, 2021.
 [9] Gjurkovic, M., & Snajder, J., "Reddit: A Gold Mine for Personality Prediction", In PEOPLES@NAACL-HTL pp. 87-97, 2018.
 [10] S. E. S. Duane and P. Schultz "Theories of Personality" 11th ed. Boston, MA, USA: Cengage Learning, 2016.
 [11] C.G. Jung "Psychological Types The Collected Works of C. G. Jung" Vol. 6, London: Routledge & Kegan Paul, 1971.
 [12] Cui, B., & Qi, C., "Survey analysis of machine learning methods for natural language processing for MBTI personality type prediction", 2017.
 [13] Mehta, Y., Majumder, N., Gelbukh, A., & Cambria, E., "Recent trends in deep learning based personality detection", Artificial Intelligence Review, 53(4), 2313-2339, 2020.