

딥러닝 기반 ERP 분석을 통한 Human - AI 추천 에이전트의 정체성과 시각 모달리티 효과 예측

손귀영¹, 권순일^{2,*}, 김예리¹, 박기완³, 김예은⁴, 윤진호⁵

¹서울여자대학교 데이터사이언스학과

²세종대학교 콘텐츠소프트웨어학과

³서울대학교 경영학과

⁴Dept. of Marketing and Supply Chain Management, California State University, Sacramento

⁵Dept. of Marketing, College of Business, New Mexico State University
sgy1017@yonsei.ac.kr, skwon@sejong.edu, sophiakim@swu.ac.kr, kiwanp@snu.ac.kr, yaeun.kim@csus.edu, drjyun@nmsu.edu

Deep Learning-Based ERP Analysis for Predicting Identity and Modality Effects in Human - AI Recommendation Agents

Guiyoung Son¹, Soonil Kwon^{2,*}, Yaeri Kim¹, Kiwan Park³, Yaeun Kim⁴, Jinho Yun⁵

¹Dept. of Data science, Seoul Woman's University

²Dept. of Software, Sejong University

³Dept. of Marketing, Business School, Seoul National University

⁴Dept. of Marketing and Supply Chain Management, California State University, Sacramento

⁵Dept. of Marketing, College of Business, New Mexico State University

요 약

본 연구는 사건관련전위(Event-Related Potential, ERP) 기반 실험을 통해 인간(Human)과 인공지능(AI) 추천 에이전트의 정체성 및 시각 모달리티(텍스트 vs. 얼굴 이미지)가 뇌 반응에 미치는 영향을 규명하였다. 상징적 소비 성향이 높은 남성 대학(원)생 27명을 대상으로 ERP의 P1(70 - 130ms)과 N170(150 - 200ms) 성분을 분석하고, EEGNet 기반 딥러닝 모델을 적용하여 네 가지 조건(Human - Face, Human - Text, AI - Face, AI - Text)을 분류하였다. 분석 결과, EEGNet은 평균 87.5%의 분류 정확도를 기록하여 SVM-RBF, LDA 등 기존 모델보다 우수한 성능을 보였다. 또한, 혼동행렬 분석을 통해 얼굴 자극 조건(Human - Face, AI - Face)이 텍스트 조건보다 높은 분류 성능을 나타냄을 확인하였으며, 이는 얼굴 자극이 P1과 N170 성분에서 더 강력한 신경 반응을 유발함을 시사한다. 본 연구는 ERP와 딥러닝 기반 접근을 결합함으로써 휴먼 - AI 상호작용 설계 및 신경과학적 해석에 기여한다.

1. 서론

상징적 소비(symbolic consumption)[1]는 소비자의 자아 정체성(self-identity)과 사회적 지위(social status)를 표현하는 고유한 인간적 행위로 정의되어 왔다. 그러나 최근 AI 기술의 발전으로 추천 시스템과 같은 인공지능 에이전트가 이 영역에 개입하면서, 소비자에게 인간 고유성에 대한 잠재적 위협으로 인식될 수 있다. 기존 연구는 주로 외형적 유사

성(anthropomorphic appearance)에 따른 ‘언캐니 밸리(uncanny valley)[2]’ 현상에 집중해 왔으나, 시각 모달리티(텍스트 vs. 얼굴 이미지)와 정체성 제시 방식(인간 vs. AI)이 소비자 인식과 신경생리학적 반응에 미치는 영향은 아직 충분히 규명되지 않았다 [3, 4]. 이에 본 연구는 상징적 소비 맥락에서 정체성(AI vs. 인간)과 시각 모달리티(텍스트 vs. 얼굴 이미지)가 소비자가 지각하는 자연스러움과 ERP 기반 신경 반응에 어떠한 차이를 유발하는지를 분석한다. 더 나아가 수집된 ERP 데이터를 딥러닝 기반

* 교신저자

분류 모델로 처리하여 조건별 반응 패턴을 예측함으로써, 신경생리학적 신호를 활용한 소비자 반응 예측 가능성을 실증적으로 제시하고 휴먼 - AI 상호작용 연구에 새로운 관점을 제공하고자 한다.

2. 연구방법

2.1. EEG 데이터

본 연구는 상징적 소비 성향이 높은 남자 대학(원)생 27명을 대상으로 수행하였다. 실험은 2(정체성: AI vs. 인간) × 2(시각 모달리티: 텍스트 vs. 얼굴 이미지)의 피험자 내 요인 설계(within-subject design)를 적용하였다[5]. 자극은 동일한 얼굴 이미지에 ‘human’ 또는 ‘AI’라는 텍스트를 결합하거나, 텍스트만 단독으로 제시하는 방식으로 구성하였다. 실험 과정에서 뇌파(EEG)를 실시간으로 기록하였으며, ERP 성분 중 P1(70 - 130ms)과 N170(150 - 200ms)을 주요 분석 대상으로 선정하였다. 분석은 P7과 P8 전극을 중심으로 진행되었으며, P1은 시각 자극의 초기 지각적 처리를, N170은 얼굴 인식과 관련된 인지적 반응을 반영하므로, 본 연구에서는 정체성(AI vs. 인간)과 시각 모달리티(텍스트 vs. 얼굴 이미지)에 따른 ERP 차이를 규명하는데 타당하다.

2.2. EEG 기록 및 전처리

EEG 데이터는 128채널 GES 400 시스템(Electrical Geodesics Inc., Eugene, OR, USA)을 사용하여 기록하였다. 이후, 데이터 분석을 위해 다음과 같은 전처리 과정을 수행하였다.

1. Filtering: 0.3 - 30 Hz 대역통과 필터와 60 Hz 노치 필터 적용
2. Epoching: -200~1000 ms 구간으로 분할(epoching)
3. Artifact rejection: 눈 깜박임, 안구운동, 잡음 채널 제거 및 보간(interpolation)
4. Baseline correction: -200~0 ms 구간을 기준으로 보정
5. Re-referencing: 평균 전극(average reference)으로 재참조

전처리 과정을 거친 후 최종적으로 1,382개의 epoch을 확보하였다. 조건별 분포는 Human - Face 365개, Human - Text 344개, AI - Face 322개, AI - Text 351개로 구성되었다.(Table 1)

<표 1> 실험조건별 epoch 개수

Human - Face (HF)	365
Human - Text (HT)	344
AI - Face (AF)	322
AI - Text (AT)	351
Total	1,382

3. 실험

3.1. 특징추출

전처리된 EEG 데이터는 MNE-Python[6]을 사용하여 epoch 단위 데이터셋으로 구성하였다. 각 trial(=epoch)은 시간 영역 ERP 신호 전체를 포함하는 3차원 텐서(채널 × 시간 샘플 × 1)로 변환되었으며, 최종 입력 차원은 128 × 250 × 1이다. 이는 250 Hz 샘플링 기준 약 1초 길이의 신호에 해당한다. 모든 trial은 독립적인 샘플로 간주되어 분류 모델의 입력으로 사용되었다.

3.2. EEGNet

본 연구에서는 ERP 신호 분류를 위해 변형된 EEGNet[7] 아키텍처 기반의 4가지의 조건 분류 모델을 적용하였다. EEGNet은 소규모 EEG 데이터셋에도 적합하도록 설계된 경량 신경망으로, 본 연구에서 사용한 구조는 다음과 같다. 첫째, Temporal Convolution Layer는 시간 축 필터링을 수행하며, 커널 크기는 1 × 64로 설정하였다. 둘째, Depthwise Convolution Layer는 채널 간 공간 필터링을 담당하며, 커널 크기는 128 × 1, depth multiplier는 2로 적용하였다. 셋째, Separable Convolution Layer는 시간 - 공간적 특징을 분리 학습하기 위해 사용되었으며, 커널 크기는 1 × 16으로 설정하였다. 마지막으로, Classifier Layer는 fully connected layer와 softmax 출력층으로 구성되었으며, 최종 출력층은 Dense(4, softmax)로 설정하여 Human - Face, Human - Text, AI - Face, AI - Text 네 가지 조건을 분류하였다.

3.3. 학습 및 평가

모델 학습은 Adam optimizer(learning rate = 0.0015)와 categorical cross-entropy loss 함수를 사용하여 수행하였다. 모든 실험은 PyTorch(v3.9, Facebook AI Research) 기반으로 구현하였다. 데이터셋은 참가자 단위 분할(subject-wise split)을 적용하여 훈련(70%), 검증(15%), 테스트(15%) 세트로 나누었다. 이를 통해 동일 참가자의 데이터가 훈련과 평가 세트에 중복되지 않도록 하였다. 모델 학습

조건	epoch 개수
----	----------

및 평가 설정의 세부 사항은 <표 2>에 제시하였다.

<표 2> 모델 학습 및 평가 설정값

항목	설정 값
Optimizer	Adam
Learning rate	0.0015
Dropout	0.5
Epochs	100
Batch size	32
Loss function	Categorical cross-entropy

4. 실험결과

4.1. EEGNet 분류 결과

본 연구는 총 1,382개의 ERP trial(Human - Face = 365, Human - Text = 344, AI - Face = 322, AI - Text = 351)을 활용하여 EEGNet 기반 4클래스 분류를 수행하였다. 교차 검증 결과, 모델의 평균 분류 정확도는 87%로 나타났다.

4.2. 혼돈행렬 분석

(그림 1)은 EEGNet이 각 조건(Human - Face, Human - Text, AI - Face, AI - Text)에 대한 분류 결과를 혼돈 행렬로 제시한 것이다. 분석 결과, 얼굴 자극(Human - Face, AI - Face) 조건에서 상대적으로 높은 분류 정확도가 확인된 반면, 텍스트 기반 조건(Human - Text, AI - Text)에서는 다소 낮은 정확도가 나타났다. 이러한 결과는 얼굴 자극이 ERP P1 및 N170 성분에서 상대적으로 더 두드러진 신경생리학적 반응을 유도함을 시사한다.

True Label	Predicted Label			
	HF	HT	AF	AT
HF	325	20	10	10
HT	25	295	12	12
AF	15	15	282	10
AT	18	20	15	298

(그림 1) EEGNet기반 4가지 조건의 분류 결과에 대한 혼돈행렬

4.3. 정량적 성능지표

<표 3>은 각 조건별 정밀도(precision), 재현율(recall), 및 F1-score를 나타낸다. Human - Face 조건에서 가장 높은 분류 성능(F1 = 0.89)이 도출되었

으며, 텍스트 기반 조건(Human - Text, AI - Text)에서도 유사한 수준의 성능(F1 = 0.85 - 0.86)이 확인되었다. 이러한 결과는 EEGNet이 ERP 성분으로부터 네 조건을 안정적으로 구분할 수 있음을 보여준다.

<표 3> 조건별 EEGNet 분류 성능 비교

Class	Precision	Recall	F1-score
HF	0.89	0.90	0.89
HT	0.85	0.86	0.85
AF	0.86	0.88	0.87
AT	0.88	0.84	0.86
Average	0.87	0.87	0.87

4.4. 모델 성능 비교

EEGNet의 분류 성능은 SVM-RBF(Support Vector Machine-Radial Basis Function, LDA(Linear Discriminant Analysis)와 비교하여 평가하였다. <표 4>의 결과에 따르면, EEGNet은 87.5%의 정확도를 기록하여 다른 모델들보다 우수한 성능을 보였다. 이는 ERP 데이터의 시간적·공간적 특성을 동시에 학습할 수 있는 EEGNet은 선형 판별 분석이나 커널 기반 분류기와 같은 전통적 머신러닝 모델보다 효과적임을 보여준다.

<표 4> 모델별 분류 성능 비교

Model	Accuracy (%)	Precision	Recall	F1-score
EEGNet	87.5	0.88	0.87	0.87
SVM-RBF ¹⁾	78.4	0.79	0.78	0.78
LDA ²⁾	74.2	0.75	0.74	0.74

1) Support Vector Machine-Radial Basis Function

2) Linear Discriminant Analysis

5. 결론

본 연구는 상징적 소비 상황에서 정체성(AI vs. 인간)과 시각 모달리티(텍스트 vs. 얼굴 이미지)가 ERP 기반 신경 반응 및 분류 성능에 미치는 영향을 확인해보고자 하였다. EEGNet 분석 결과, 얼굴 자극 조건에서 상대적으로 높은 분류 정확도가 나타났으며, 이는 P1 및 N170 성분이 얼굴 인식 과정과 밀접하게 관련된다는 기존 신경과학적 증거를 뒷받침한다. 이와 더불어, EEGNet은 SVM-RBF, LDA 등 기존 모델보다 우수한 성능을 기록하여 ERP 데이터의 시공간적 특성을 통합적으로 학습하는 신경망 구조의 효과성을 확인하였다. 이러한 결과는 신경생리학적 신호를 활용한 소비자 반응 예측의 실용적 가능성을 제시하는 동시에, 상징적 소비 맥락에서 AI 개입을 이해하는 새로운 연구적 시각을 제공한다.

Acknowledgement

본 연구는 2025년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임.
(NRF-2025S1A5C3A02009153)

참고문헌

- [1] Keller, K. L. Conceptualizing, measuring, and managing customer-based brand equity. *Journal of marketing*, 57(1), 1-22, 1993.
- [2] Kim, S. Y., Schmitt, B. H., & Thalmann, N. M. Eliza in the uncanny valley: Anthropomorphizing consumer robots increases their perceived warmth but decreases liking. *Marketing letters*, 30(1), 1-12, 2019.
- [3] Iglesias, C., & Packard, G. Text vs. voice: How communication modality impacts consumers' attitudes in human-machines interactions. *Marketing Science Institute Working Paper Series* 2023, 113, 1-37, 2023.
- [4] Mitrevska, T., Chiossi, F., & Mayer, S. ERP Markers of Visual and Semantic Processing in AI-Generated Images: From Perception to Meaning. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, Japan, 2025, 1-7.
- [5] Chung, K., Park, J. Y., Park, K., & Kim, Y. Which visual modality is important when judging the naturalness of the agent (artificial versus human intelligence) providing recommendations in the symbolic consumption context?. *Sensors*, 20(17), 5016, 2020.
- [6] Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... & Hämäläinen, M. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroinformatics*, 7, 267, 2013.
- [7] Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. EEGNet: a compact convolutional neural network for EEG-based brain - computer interfaces. *Journal of neural engineering*, 15(5), 056013, 2018.