

생체신호 기반의 T-SNE 를 활용한 대화 내 감정 인식

임수빈¹, 이병천², 문지훈^{1,2}
¹순천향대학교 의료과학과
²순천향대학교 AI·빅데이터학과

{qlsl0519, qudcjs0208, jmoon22}@sch.ac.kr,

Physiological Signal-Based Emotion Recognition in Conversations Using T-SNE

Subeen Leem¹, Byeongcheon Lee², and Jihoon Moon^{1,2}
¹Department of Medical Science, Soonchunhyang University
²Department of AI and Big Data, Soonchunhyang University

요 약

본 연구는 대화 중 생체신호 데이터를 활용하여 감정 인식 분야에서 더욱 정확하고 범용성이 높은 인식 기술을 제안한다. 이를 위해, 먼저 대화별 길이에 따른 측정값의 개수를 동일하게 조정하고 효과적인 생체신호 데이터의 조합을 비교 및 분석하기 위해 차원 축소 기법인 T-SNE (T-distributed Stochastic Neighbor Embedding)을 활용하여 감정 라벨의 분포를 확인한다. 또한, AutoML (Automated Machine Learning)을 이용하여 축소된 데이터로 감정을 분류 및 각성도와 긍정도를 예측하여 감정을 가장 잘 인식하는 생체신호 데이터의 조합을 발견한다.

1. 서 론

대화 내 감정은 언어적 표현뿐만 아니라 비언어적 요소인 생체신호들로 표현할 수 있다. 이는 상대방의 감정을 인식하고 적절한 대처나 반응을 통해 대화의 효율성을 높이는 데 활용될 수 있다. 그리하여 최근 생체신호를 이용하여 감정을 인식하는 기술은 심리학, 뇌 과학, 인공지능 등 다양한 분야에서 활발히 연구되고 있으며, 국내 연구에서 인간의 감정을 인식하기 위해 뇌파[1, 2], 맥파[3] 및 멀티모달[4, 5]을 활용한 연구가 활발히 진행되고 있다.

생체신호 데이터는 차원의 수가 매우 많으므로, 특징 추출 과정을 통해 데이터를 차원 축소하고 불필요한 정보를 제거하여 분석을 용이하게 한다. 이를 위해 PCA (Principal Component Analysis) [6], LDA (Linear Discriminant Analysis) [7], T-SNE (T-distributed Stochastic Neighbor Embedding) [8] 등 다양한 기법이 특징 추출 모델로 주로 활용되고 있다. 비록 T-SNE 을 활용하여 감정 라벨 시각화를 수행한 연구[9]도 있으나, 이는 음성 데이터에 관한 실험이므로, 생체신호 데이터를 대상으로 해당 기법의 효과를 검증할 필요가 있다.

본 논문에서는 감정 인식 분야에서 더욱 강건하고 범용성 높은 인식 기술을 개발하기 위해 세 종류의 생체신호 데이터를 개별 또는 조합하여 비교 및 분석

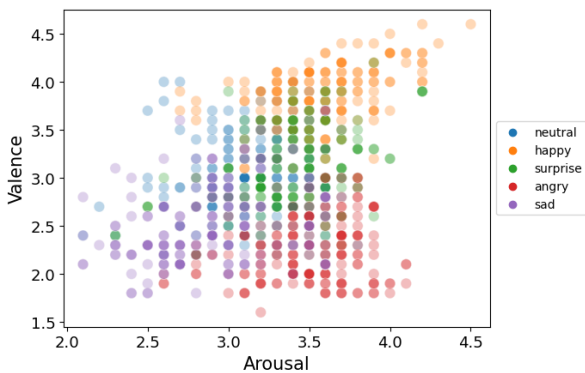
한다. 먼저 고차원의 데이터를 2 차원으로 축소하고 시각화하여 감정 라벨이 잘 분류되었는지 판단한다. 다음으로 축소된 데이터로 감정을 분류하고, 각성도/긍정도를 예측한다. 여기서 AutoML (Automated Machine Learning)을 통해 시간과 비용을 절감하면서 높은 정확도를 도출하는 모델을 구성한다. 본 연구는 감정을 가장 잘 인식하는 생체신호 데이터의 조합을 발견하여 효과적인 감정 인식에 기여할 수 있다.

본 논문의 나머지 구성은 다음과 같다. 2 장에서는 데이터 소개 및 전처리 과정을 소개한다. 3 장에서는 데이터 축소 및 시각화를 다루며, 4 장은 감정 인식 모델을 구성하고 성능을 비교하는 과정에 대해 기술한다. 마지막으로 5 장은 향후 연구와 함께 본 논문의 결론을 맺는다.

2. 데이터 소개 및 전처리

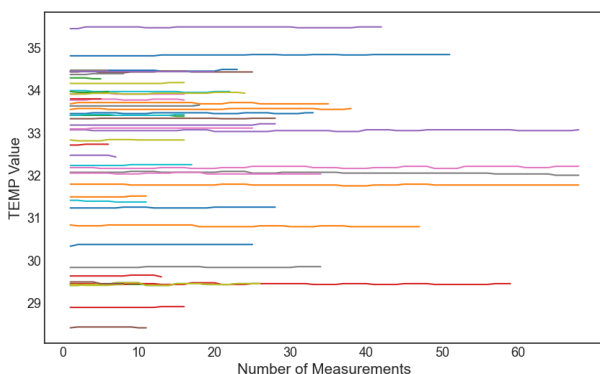
본 연구에서 사용한 데이터는 ‘한국어 멀티모달 감정 데이터셋 2020 (KEMDy20) [10]’으로 80 명의 성우들의 자유로운 대화 과정에서 수집된 맥박(IBI), 피부전도도(EDA), 손목 피부 온도(TEMP)에 대한 생체신호이다. 또한, 대화별로 10 명의 외부 감정 관찰자가 판단한 감정의 최빈값, 각성도(Arousal)와 긍정도(Valence)의 평균값에 대한 정보도 포함되어 있다.

실험에서는 데이터가 풍부한 5 가지 감정 라벨로 중립, 행복, 놀람, 분노, 슬픔에 관련된 데이터를 대상으로 라벨별로 균형(Balance)을 맞추고자 오버샘플링(Oversampling) 및 언더샘플링(Undersampling)하여 200 개씩 데이터를 추출하였다. 또한, 각성도와 긍정도를 동시에 파악하는 것은 복잡한 감정들을 효과적으로 파악하는 데 매우 유용하므로[11], 그림 1 과 같이 두 지표에 따른 감정 분포를 나타내었다. 각 점은 alpha 파라미터를 지정하여 점의 투명도를 제시하였으므로 진한 영역은 해당 부분에 데이터가 주로 분포한다는 것을 의미한다. 두 지표의 중간 범위에는 중립(Neutral) 라벨이 고르게 분포한다. 또한, 가장 높은 범위에는 행복(Happy) 라벨이 분포한다. 이전에 언급한 두 지점값이 약간 낮은 범위에는 놀람(Surprise) 라벨이 있고, 이보다 각성도가 높고 긍정도가 낮은 범위에는 분노(Angry) 라벨이 있다. 두 지점값이 제일 낮은 부분에는 슬픔(Sad) 라벨이 분포한다.



<그림 1> 각성도와 긍정도에 따른 감정 분포

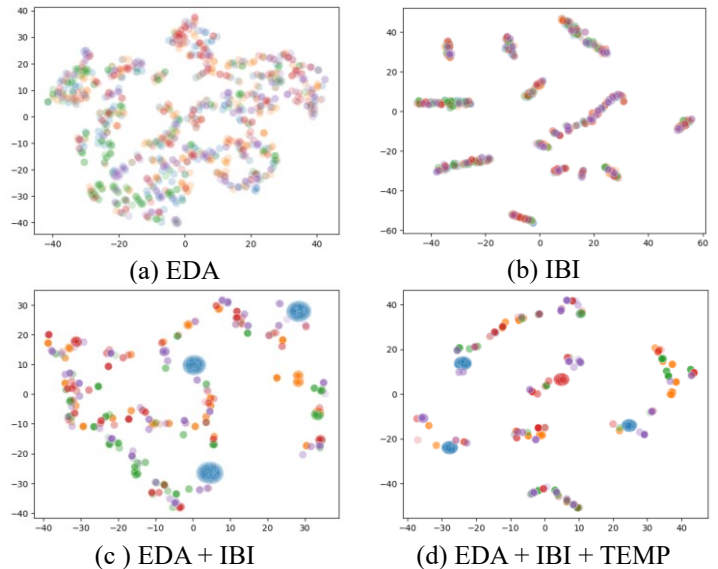
신호 데이터의 특징 추출을 위해서는 모든 대화가 동일한 길이를 가져야 한다. 그러나 수집된 대화의 길이가 일정하지 않으므로, 측정된 신호 데이터 또한 길이가 다양하다. 그림 2 는 TEMP 를 측정할 일부 대화에서 측정 횟수의 다양성을 보여준다. 따라서, 길이를 동일하게 하기 위해, 가장 긴 측정값의 길이에 맞추어 짧은 측정값을 0 으로 채워 결측치를 보완한다.



<그림 2> 대화별 길이에 따른 측정값

3. 축소된 데이터 시각화

T-SNE 는 데이터 포인트 간의 유사성을 유지하면서 차원을 축소하는 기술로, 이를 통해 잘 구분된 클러스터를 형성한다[12]. 그림 3 은 T-SNE 를 통해 2 차원으로 축소된 데이터를 감정 라벨에 따라 그림 1 과 동일한 색으로 구분하여 나타낸 것이다. 그림 3(a), (b) 에서 피부전도도(EDA)와 맥박(IBI)을 각각 사용하여 분석한 결과, 감정에 대한 클러스터를 찾을 수 없었으나 EDA 와 IBI 를 함께 사용한 경우, 중립 감정에 대한 3 개의 클러스터를 확인할 수 있었으며, TEMP 데이터를 함께 사용한 경우, 분노 감정에 대한 클러스터도 발견할 수 있었다. 이를 통해 각 데이터 유형을 복합적으로 사용함으로써 더욱 세부적인 감정 분류가 가능하다는 것을 확인할 수 있다. 이뿐만 아니라 클러스터 간 간격이 좁아져서 각각의 감정에 대한 분류가 더욱 세밀하게 이루어지게 되었음을 확인하였다.



<그림 3> 개별 또는 조합 데이터 별 감정 분포

4. 감정 인식 모델

본 논문에서는 AutoML 기술을 이용하여 기계학습 파이프라인을 자동화하고, 시간과 비용을 절감하면서 다양한 모델들의 분류 성능을 비교하였다. 이를 위해, PyCaret 라이브러리[13]를 활용하여 각 데이터의 감정 분류와 각성도 예측과 긍정도 예측 모델을 학습하고 평가하였다. 또한, F1-score, CCC-A 및 CCC-V 와 같은 성능 지표를 사용하여 각각의 학습 목표에 따른 모델 성능을 평가하며, 이를 표 1 에 나타내었다.

<표 1> 데이터별 분류 성능 평가

| Data | F1-score (%) | CCC-A (%) | CCC-V (%) |
|--------------|--------------------------|--------------------|--------------------|
| EDA | 76.81 | 94.26 | 93.97 ^c |
| IBI | 75.15 | 89.92 ^c | 93.97 |
| TEMP | 73.18 | 91.44 ^r | 94.62 ^r |
| EDA + IBI | 98.15 ^c | 99.38 | 99.69 |
| EDA + TEMP | 86.24 | 97.88 | 98.16 ^x |
| IBI + TEMP | 95.29 ^l | 99.00 | 99.33 |
| EDA+IBI+TEMP | 98.44^g | 99.40 | 99.53 ^x |

c: Catboost, l: LightGBM, g: GBM, r: Random Forest, x: XGBoost

AutoML 의 편리성으로, 입력 데이터별, 학습 목표 등에 따라 최종 모델을 다르게 선택할 수 있다. 대부분의 경우, Extra Tree 모델이 가장 높은 성능을 도출하였다. 이 모델은 트리 기반 모델에 추가적인 랜덤성을 더하여 과적합을 방지하고 성능을 향상시키는 알고리즘으로, 3 개의 신호 데이터를 모두 사용할 때, 3 개의 지표에 대해 높은 성능을 보였다. 또한, CCC-V 지표에서 TEMP 를 사용하지 않았을 때가 가장 높았으나, 이는 0.16% 차이에 불과하다는 것을 확인할 수 있었다

5. 결 론

본 논문은 대화 중 생체신호 데이터를 활용한 강건하고 범용성이 높은 감정 인식 모델을 제안하였다. 이를 위해 길이가 다른 생체신호 데이터를 동일하게 조정하여 실험하였으며, T-SNE 기법을 이용하여 생체신호 데이터를 시각화하였다. 또한, 다양한 생체신호 데이터를 조합하여 감정 인식을 분석하면 단독으로 분석했을 때보다 더욱 정확한 성능을 도출하였으며, AutoML 기술을 통해 최적의 기계학습 모델을 구성할 수 있었다. 결론적으로, 본 연구는 생체신호 데이터를 활용하여 감정 인식 분야에서의 연구 발전을 이끌어 나가는 기초가 될 수 있다.

향후 연구는 본 연구 외 다양한 생체신호 데이터를 활용하여 감정 인식 분석에 새로운 지표를 도입하고, AutoML 기술을 더욱 효율적으로 활용하여 정확도를 높인 모델을 개발하고자 한다. 이뿐만 아니라 감정 인식을 실제 환경에서 적용할 때, 데이터의 신뢰성과 개인정보 보호 문제에 관한 여러 측면들을 고려하여 연구를 발전시켜 나가하고자 한다.

사 사 문 구

본 연구는 한국연구재단 4 단계 두뇌한국 21 사업 (4 단계 BK21 사업)의 지원을 받아 작성되었음(과제 번호: 5199990514663). 또한, 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구 결과로 수행되었음(2021-0-01399).

참 고 문 헌

- [1] D.-Y. Lee and Y.-S. Choi, "Emotion Recognition based on Multivariate Multiscale Fuzzy Entropy Analysis of EEG Recordings," *Journal of KIISE*, vol. 47, no. 3, pp. 227–234, 2020.
- [2] G. Son, W. Lee, J. Lee, and M. Ki, "Machine Learning based Human emotion state classification using EEG signals," *The Journal of KINGComputing*, vol. 18, no. 1, pp. 37–46, 2022.
- [3] W.-D. Chang, "Frequency Analysis for Recognition of Emotional States Using Photoplethysmograms," *Journal of Next-generation Convergence Technology Association*, vol. 6, no. 1, pp. 26–31, 2022.
- [4] J. E. Lee and S. K. Yoo, "Deep Learning based Emotion Classification using Multi Modal Bio-signals," *Journal of Korea Multimedia Society*, vol. 23, no. 2, pp. 146–154, 2020.
- [5] S. Lee and W. Kim, "An emotion classification model based on decision fusion using multi-modal data," in *Proc. of the KCC 2022*, 2022, pp. 2285–2287.
- [6] A. Choi and W. Woo, "Feature extraction for emotion analysis based on physiological signal," in *Proc. of 2005 2005 Conference on the HCI Society of Korea*, 2005, pp. 624–629.
- [7] W.-C. Lim and K.-C. Kwak, "A Multilinear LDA Method of Tensor Representation for ECG Signal Based Individual Identification," *Smart Media Journal*, vol. 7, no. 4, pp. 90–98, 2018.
- [8] Y. Ye, G. Xiong, Z. Wan, T. Pan, and Z. Huang, "PPG-based biometric identification: Discovering and identifying a new user," in *Proc. of 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021, pp. 1145–1148.
- [9] Y. Ahn, S. Han, S. J. Lee, and J. W. Shin, "Speech emotion recognition based on wav2vec feature embedding," in *Proc. of the KICS Conference*, 2021, pp. 11–12.
- [10] K. J. Noh and H. Jeong, "KEMDy20," https://nanum.etri.re.kr/share/kjnoh/KEMDy20/update?lang=ko_KR.
- [11] P. Kuppens, F. Tuerlinckx, J. A. Russell, and L. F. Barrett, "The relation between valence and arousal in subjective experience." *Psychological Bulletin*, vol. 139, no. 4, p. 917, 2013.
- [12] G. C. Linderman and S. Steinerberger, "Clustering with t-SNE, provably," *SIAM Journal on Mathematics of Data Science*, vol. 1, no. 2, pp. 313–332, 2019.
- [13] X. He et al., "AutoML: A survey of the state-of-the-art," *Knowledge-Based Systems*, Vol. 212, p. 106622, 2021.