안녕하세요 Feature analysis using interpretable artificial intelligence of EEG-based emotion classification에 대해 소개하겠습니다.

프로젝트의 개요는 다음과 같습니다. EEG가 무엇인지 간단하게 이해한 후, SVM, Random Forest, Extreme Gradient Boosting을 EEG 데이터 맞게 설계한 뒤 여러 지표를 종합적으로 평가하여 적절한 모델을 찾습니다.  
그 후 선정된 모델의 ICE, LIME을 통한 중요한 feature를 분석합니다.

EEG에 대한 간단한 설명입니다. EEG는 두피에 부착된 전극을 사용하여 뇌의 전기적 활동을 기록하고 측정하는 것입니다. 즉 뇌파입니다.  
EEG는 다양한 감정 상태와 관련된 뇌 활동 패턴을 모니터링하여 감정 표현의 신경 상관 관계를 연구하는 데 도움이 될 수 있습니다.  
이러한 EEG의 특성에 따라서 모델을 설계 및 분석하겠습니다.

데이터에 대한 설명은 다음과 같습니다. 라벨은 positive, neutral, negative로 구성되어 있습니다.  
왼쪽 그림이 뇌파의 상태를 바꾸기 위한 영상의 종류입니다. 오른쪽 그림은 뇌파를 수집하는 위치를 표현한 것입니다.

다음으로는 Method입니다. 단순한 뇌파는 노이즈와 각종 정보가 섞여 있어 판단하기 어렵기에 프리 프로세싱이 필요합니다. 이러한 프리 프로세싱 과정을 통해 여러 feature를 만들어냅니다.

사용한 모델중 첫 번째인 SVM입니다. SVM은 이상값에 강력하기에 이상값이나 노이즈가 많은 EEG데이터 분류에 적합하다고 생각했습니다. 또한 여러 클래스를 분류해야 하기에 적절하다고 생각했습니다. 반환된 확률이 잘 보정되지 않는 LinearSVC 문제를 해결하기 위해 CalibratedClassifierCV를 사용하여 보정했습니다. 모델의 과대 과소 적합을 막기 위한 적절한 parameter는 다음과 같습니다. 위에서부터 반복횟수, 정규화 파라미터, l1규제입니다.

다음 모델은 랜덤포레스트입니다. 랜덤포레스트는 EEG 데이터에서 다양한 특징 간의 복잡한 상호 작용을 고려하여 분류 모델을 구축할 수 있습니다. 또한 feature importance를 추정하여 EEG를 어떻게 처리했을 때 효과적인지 볼 수 있습니다. 과적합을 방지하기에 적합하여 노이즈가 많은 EEG 데이터에 적합하다고 생각했습니다. 모델의 과대 과소 적합을 막기 위한 적절한 parameter는 다음과 같습니다. 위에서부터 의사 결정 트리의 최대 깊이, 하나의 리프 노드에 필요한 최소 샘플 수 입니다.

다음 모델은 XGBoost입니다. XGBoost는 오류를 줄이면서 약한 모델을 순차적으로 훈련하여 예측 성능을 향상시킵니다. 따라서 EEG 데이터의 다양한 특징을 활용하여 보다 정확한 분류를 할 수 있습니다. 또한 이상값에 덜 민감하기에 EEG에 이상값이 있는 경우에도 강력한 분류를 수행할 수 있습니다. feature importance를 추정할 수 있기에 EEG에서 가장 유용한 기능을 식별하고 모델을 해석하는 데 도움이 될 수 있습니다. 위에서부터 개별 트리 최대 깊이, 학습률, 앙상블에 사용되는 트리 수, 각 트리별 학습할 데이터의 비율, L1 정규화 계수, L2 정규화 계수입니다.

해석 방법에 대해서 설명하겠습니다.  
먼저 PDP, ICE입니다. 앙상블 모델은 개별 나무 간의 복잡한 상호 작용과 특성의 영향을 이해하기 어렵게 만들 수 있습니다. PDP 및 ICE는 모델의 예측에 대한 기능의 영향을 시각화하고 설명하는 데 유용합니다. 기능이 변경됨에 따라 모델의 예측을 그래프로 표시함으로써 각 기능이 예측에 미치는 영향을 이해할 수 있습니다.

다음은 LIME입니다. 개별 데이터 포인트에 대한 모델의 예측을 해석할 수 있는 설명 모델을 생성합니다.  
EEG는 개별 관찰이며 각 관찰에는 고유한 특성이 있습니다. LIME은 개별 EEG 포인트의 예측을 설명하는 데 사용할 수 있습니다. LIME은 모델이 특정 데이터 포인트를 분류하는 방법을 이해하는 데 사용할 수 있습니다.

모델에 대한 분석 결과는 다음과 같이 나왔습니다. SVM입니다.

Random Forest와 XGBoost입니다.

SVM의 학습 곡선은 다음과 같습니다. 교차 검증과 트레인 학습의 일치율이 높고 정확도가 높은것을 확인할 수 있습니다.

Random Forest의 feature importance와 학습 곡선은 다음과 같습니다. 일치율은 상대적으로 더 낮지만 정확도는 높은 것을 확인할 수 있습니다.

XGBoost의 feature importance와 학습 곡선은 다음과 같습니다. 일치율이 낮으며 과소적합이 살짝 발생했습니다. 하지만 정확도는 높은 것을 확인할 수 있습니다.

SVM의 ICE입니다. feature들의 값이 증가할 때 Negative, 감소할 때, Neutral과 Positive인 것을 확인할 수 있고 Positive는 Neutral에 비해 판단이 어려운 것을 확인 가능합니다.

SVM의 LIME입니다. LIME은 로컬에 대해서 진행하였기에 부정확할 수 있음을 밝힙니다. Negative와 Neutral이 mean\_0\_b에 의해서 구분이 가능하다고 보여집니다. Positive또한 mean\_0\_b의 영향을 받습니다. (mean\_0\_b는 b번째 실험자의 0번째 실험 평균값입니다.)

Random Forest의 ICE입니다. feature들의 값이 증가하면 Negative일 확률이 줄어들고, Neutral일 확률이 증가, Positive일 확률이 증가합니다. 여전히 Positive는 Neutral에 비해 판단이 어려운 것을 확인 가능합니다. 하지만 feature의 값이 증가함에 따라서 Positive일 확률이 증가한다는 것은 증가하는 feature의 개수가 더 많기에 확인 가능합니다.

Random Forest의 LIME입니다. 뚜렷하게 관여하는 feature는 없고 여러 feature들이 동시에 판단을 돕습니다. 여기서 특이하게 Positive는 SVM과 동일한 Feature가 가장 크게 관여한다는 것을 알 수 있습니다.

XGBoost의 ICE입니다. feature들의 값이 증가하면 Negative일 확률이 줄어들고, Neutral일 확률이 증가, Positive일 확률이 감소합니다.

XGBoost의 LIME입니다. 마찬가지로 가장 중요하게 봐야 할 것이 mean\_0\_b입니다. SVM에서 중요한 feature가 여전히 관여하는 것을 볼 수 있습니다. Negative를 보시면 mean\_2\_b, mean\_0\_a등 EEG의 평균이라는 것이 감정에 크게 관여한다는 것을 볼 수 있습니다.

결론입니다. ICE를 통해 분석했을 때, SVM보다 두 앙상블 모델의 정확도가 높은 이유가 feature의 경향성이 뚜렷하기 때문임을 확인하였습니다. Negative와 Neutral은 구분이 뚜렷하지만, Positive의 구분이 상대적으로 힘든 것을 확인하였습니다. 평균이라는 값이 감정 분류에 큰 영향을 미친다는 것을 확인하였습니다. 평균을 취하게 되면 Moving Average Filter처럼 작용하게 되어 노이즈가 어느정도 제거되기에 큰 영향이 있는 것으로 보입니다. 추후 Feature들을 PDP를 통해 하나하나 분석해보며 정확히 어떠한 Feature가 감정 표현에 큰 영향을 주는지를 알아낼 수 있을 것입니다.

Hello, I will introduce feature analysis using interpretable artificial intelligence of EEG-based emotion classification.

An outline of the project is as follows: After a simple understanding of what EEG is, SVM, Random Forest, and Extreme Gradient Boosting are designed to fit the EEG data, and various indicators are comprehensively evaluated to find an appropriate model.

Then, important features of the selected model are analyzed through ICE and LIME.

A brief explanation of the EEG. EEG is the recording and measurement of electrical activity in the brain using electrodes placed on the scalp. That is, brain waves.

EEG can help study neural correlates of emotional expression by monitoring patterns of brain activity associated with different emotional states.

A model will be designed and analyzed according to the characteristics of these EEGs.

A description of the data follows. Labels are composed of positive, neutral, and negative.

The picture on the left is the type of image used to change the state of brain waves. The picture on the right shows the location where EEG is collected.

Next is Method. Pre-processing is required because simple brain waves are mixed with noise and various information, making it difficult to judge. Through this pre-processing process, several features are created.

The first of the models used is SVM. Since SVM is robust to outliers, we thought it was suitable for classifying EEG data with a lot of outliers or noise. We also thought it was appropriate as we need to classify several classes. I calibrated it using CalibratedClassifierCV to solve the problem of LinearSVC where the returned probabilities were not well calibrated. The appropriate parameters to prevent overfitting of the model are as follows. From the top, iteration count, regularization parameter, l1 regulation.

The next model is a random forest. Random Forest can build a classification model by considering complex interactions between various features in EEG data. You can also estimate feature importance to see how effective EEG processing is. I thought it was suitable for noisy EEG data as it was suitable to prevent overfitting. The appropriate parameters to prevent overfitting of the model are as follows. From top to bottom, the maximum depth of the decision tree, the minimum number of samples required for one leaf node.

The next model is XGBoost. XGBoost improves predictive performance by sequentially training a weak model while reducing errors. Therefore, it is possible to perform more accurate classification by utilizing various characteristics of EEG data. It is also less sensitive to outliers, allowing robust classification even when there are outliers in the EEG. Being able to estimate feature importance can help identify the most useful features in the EEG and interpret the model. From top to bottom, these are the individual tree maximum depth, the learning rate, the number of trees used in the ensemble, the fraction of data to be trained for each tree, the L1 regularization factor, and the L2 regularization factor.

I will explain how to interpret it.

First, PDP and ICE. Ensemble models can make it difficult to understand the complex interactions between individual trees and the influence of traits. PDP and ICE are useful for visualizing and explaining the effect of a feature on a model's predictions. By graphing the model's predictions as features change, you can understand how each feature affects the predictions.

LIME is next. Create explanatory models that can interpret the model's predictions for individual data points.

EEG is an individual observation and each observation has its own characteristics. LIME can be used to describe predictions of individual EEG points. LIME can be used to understand how a model classifies a particular data point.

The results of the analysis of the model were as follows. It is SVM.

Random Forest and XGBoost.

The learning curve for SVM is: You can see that cross-validation and train learning have a high concordance rate and high accuracy.

The feature importance and learning curve of a random forest are as follows: It can be seen that the match rate is relatively low, but the accuracy is high.

XGBoost's feature importance and learning curve are as follows: The agreement rate is low and there is a slight underfitting. However, it can be confirmed that the accuracy is high.

This is the ICE of SVM. It can be confirmed that when the value of the features increases, it is Negative, and when it decreases, it is Neutral and Positive, and it is difficult to judge Positive compared to Neutral.

LIME of SVMs. Please note that LIME may be inaccurate because it was performed locally. Negative and Neutral are shown to be distinguishable by mean\_0\_b. Positive is also affected by mean\_0\_b. (mean\_0\_b is the average value of the 0th experiment of the bth experimenter.)

This is ICE from Random Forest. As the value of a feature increases, the probability of being negative decreases, the probability of being neutral increases, and the probability of being positive increases. It is still possible to confirm that positive is more difficult to judge than neutral. However, it can be confirmed that the probability of being positive increases as the value of the feature increases because the number of increasing features increases.

This is LIME from Random Forest. There are no distinctly involved features, and several features assist in the decision at the same time. In particular, it can be seen that positives are most significantly involved in the same features as SVMs.

This is XGBoost's ICE. As the value of a feature increases, the probability of being negative decreases, the probability of being neutral increases, and the probability of being positive decreases.

This is the LIME of XGBoost. Likewise, the most important thing to look at is mean\_0\_b. In SVM, we can see that important features are still involved. If you look at Negative, you can see that the average of EEG, such as mean\_2\_b and mean\_0\_a, is greatly involved in emotion.

Conclusion. When analyzed through ICE, it was confirmed that the reason for the higher accuracy of the two ensemble models than SVM was that the feature tendencies were clear. The distinction between Negative and Neutral is clear, but it was confirmed that it is relatively difficult to distinguish between Positive. It was confirmed that the value of mean has a great influence on emotion classification. If you take the average, it acts like a moving average filter, so it seems to have a big impact on removing noise to some extent. In the future, we will be able to analyze the features one by one through the PDP and find out which feature has a great influence on the expression of emotion.