

Gradient boosting Classifier와 Feature importance를 이용한 EEG 신호처리에 따른 Hyponogram예측 모델에 관한 연구

류진수, 김규형, 조위덕*

아주대학교, 아주대학교, *아주대학교

s8620494@ajou.ac.kr, kviw0220@gmail.com, *chowd@ajou.ac.kr

A Study on the Hyponogram Prediction Model according to EEG Signal Processing Using Gradient Boosting Classifier and Feature importance

Ryu Jin Su, Kim Kyu Hyeong, Cho We Duke*

Ajou Univ., Ajou, Univ, *Ajou Univ.

요 약

본 논문은 기존의 뇌파, 호흡 수, 심박 수 등의 여러 데이터를 사용하여 수면의 깊이를 예측하는 모델과 다르게 2-channel의 EEG 데이터를 통해 수면의 깊이를 의미하는 Hyponogram을 예측하는 모델에 관한 연구로 뇌파의 주파수 대역에 따라 나눠 총 8개의 feature를 생성하여 머신러닝을 진행하였으며, 여러 방법으로 신호처리를 진행하였다. 신호처리를 진행한 데이터를 Gradient boosting Classifier를 사용하여 생성된 모델의 각 Feature importance를 확인하여 모델별 예측 결과에 대해 비교, 분석을 진행하여 신호처리 방법에 따른 예측률 차이를 확인하였다.

I. 서 론

본 논문에서는 수면 시 2-channel의 EEG 뇌파 데이터[1][2]의 여러 신호 처리 방법을 이용하여, 수면 상태를 나타내는 Hyponogram의 Wake(W), 렘수면 상태(R), 비 렘수면 상태(NR1, NR2, NR3, NR4)를 머신러닝을 이용해 예측 모델을 만드는 연구에 관한 내용이다. 머신러닝 알고리즘은 Gradient Boosting Classifier 알고리즘을 이용하였다. 각 수면 상태별로 신호처리 방법에 따른 예측률 차이를 확인하기 위해 신호처리 방법을 다르게 하여 모델을 생성하고, 예측률에 따른 이유를 feature importance를 비교하여 분석하는 연구를 진행하였다.

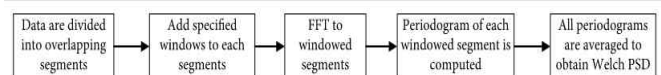
II. 본론

본 논문에서는 뇌파(EEG)의 신호처리를 진행하였다. EEG 신호처리의 sequence는 다음과 같다. EEG signal data를 FFT로 변환한 데이터, welch-PSD으로 변환한 데이터 셋으로 나눈다. welch-PSD[3]으로 변환 시 windows 함수를 hann, hamming, blackman 세 가지 함수를 사용하여 데이터 셋을 나눈다. 각 windows 함수의 수식은 아래와 같다.

$0.5 - 0.5 \cdot \cos\left(\frac{2 \cdot \pi \cdot k}{N-1}\right)$	$0.54 - 0.46 \cdot \cos\left(\frac{2 \cdot \pi \cdot k}{N-1}\right)$	$0.42 - 0.5 \cdot \cos\left(\frac{2 \cdot \pi \cdot k}{N-1}\right) + 0.08 \cdot \cos\left(\frac{4 \cdot \pi \cdot k}{N-1}\right)$
1. Hann Window	2. Hamming Window	3. Blackman Window

Table 1. 사용한 window 함수 수식

EEG signal의 데이터를 python을 이용하여 FFT, Welch PSD 두 가지 방법으로 주파수 영역으로 변환한다. 그 후 EEG의 각 wave 별로 주파수 영역에서 나누어 평균을 구한다. Welch PSD 같은 경우 변환 시 Window 함수를 사용하는데, 이 Window 함수에 따라 그래프의 특성이 변한다. Welch 함수를 구하는 sequence와 알고리즘은 아래와 같다.[4]



- $x_i = x \left[i \times \frac{L}{2} + n \right]$, where $n = 0, \dots, L-1$, $i = 0, 1, 2, \dots, N-1$.
- $N = L + (L - N_D) (N_D - 1)$.
- $W_i = x_i(n) \times w(n)$.
- $A_i(k) = x_i(n) w(n) e^{-j(2\pi/N)nk}$.
- $\phi_i = \frac{1}{L} |A_i(k)|^2$, $U = (1/L) \sum_{n=0}^{L-1} w^2(n)$, $S(k) = \frac{1}{L} \sum_{i=0}^{L-1} \phi_i(k)$.

Figure 1. Welch sequence와 알고리즘

위 알고리즘에 대한 파라미터는 다음과 같다. n은 EEG signal의 개수, 는 EEG signal을 segment로 나눈 것이다. 이 segment의 길이는 L이며 각 segment는 겹치는 부분이 존재한다. 는 겹치는 점의 수는 segment의 수이다.

w[n]은 segment에 대한 window 함수이다. 사용하는 window 함수의 종류에 따라 수식을 변경할 수 있다.

는 각 segment의 Periodogram, U는 window 함수의 평균전력, S(k)는 PSD이다.

위 알고리즘을 따라 welch PSD를 구한 후 EEG의 wave 별 구간으로 나누어야 한다. EEG wave의 구간은 Delta wave : 1~4Hz, Theta wave : 4~7Hz, Alpha wave : 8~12Hz, Beta wave : 12~30Hz이다. welch PSD를 구한 후 각 wave의 구간별로 나누어 평균을 구한 값을 machine Learning의 parameter로 사용하였다. 한 channel당 4개씩 총 8개의 feature를 통해 machine Learning을 진행하였다.

본 연구에서 사용한 machine Learning 알고리즘은 Gradient boosting classifier로 여러 개의 학습기를 연결하여 강한 학습기를 만드는 앙상블 방법인 부스팅 방법 중 하나로 이전 예측기가 만든 잔여 오차에 새로운

예측기를 학습시키는데 Negative gradient(residual)를 최소화시키면서 학습시키는 알고리즘이다.

```


$$F_0(\mathbf{x}) = \arg\min_p \sum_{i=1}^N L(y_i, p)$$

For m = 1 to M do:
    
$$\hat{y}_i = - \left[ \frac{\partial L(y_i, F_{m-1}(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)} \right] F(\mathbf{x}_i) = F_{m-1}(\mathbf{x}_i) + \hat{y}_i, \quad i = 1, N$$

    
$$\mathbf{a}_m = \arg\min_{\mathbf{a}_m} \sum_{i=1}^N L(y_i, F_{m-1}(\mathbf{x}_i) + \hat{y}_i)$$

    
$$\hat{p}_m = \arg\min_p \sum_{i=1}^N L(y_i, F_{m-1}(\mathbf{x}_i) + p h(\mathbf{x}_i; \mathbf{a}_m))$$

    
$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \hat{p}_m h(\mathbf{x}; \mathbf{a}_m)$$

endFor
end Algorithm

```

Figure 2. Gradient boosting classifier algorithm

Hyponogram 상태별 Gradient boosting classifier를 통한 Machine Learning 결과 Test score는 아래와 같다.

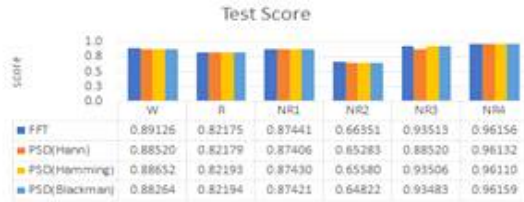


Figure 3. Hyponogram 상태별 Test score

신호처리별로 같은 Hyponogram 일 때의 test score는 크게 차이가 나지 않았으며 Non-rem 수면상태 중 NR2 상태 모델의 test score는 0.65 정도로 다른 상태에 비해 낮은 예측률이 나왔으며 NR4일 때 0.96 정도로 높은 예측률을 보여주었다. 예측률이 낮게 나오는 이유를 확인하기 위해 FFT로 신호처리한 경우 각 Hyponogram 상태별 모델의 Feature importance[6]를 확인하였으며 Feature importance는 아래의 과정을 통해 구해.

- Inputs: fitted predictive model m , tabular dataset (training or validation) D
- Compute the reference score s of the model m on data D (for instance the accuracy for a classifier or the R^2 for a regressor).
- For each feature j (column of D):
 - For each repetition k in $1, \dots, K$:
 - Randomly shuffle column j of dataset D to generate a corrupted version of the data named $\hat{D}_{k,j}$.
 - Compute the score $s_{k,j}$ of model m on corrupted data $\hat{D}_{k,j}$.
 - Compute importance i_j for feature j defined as:

$$i_j = s - \frac{1}{K} \sum_{k=1}^K s_{k,j}$$

Figure 4. Scikit Learn Gradient boosting의 Classifier Feature importance



Figure 5. FFT처리 데이터 model의 Feature importance

상대적으로 높은 예측률이 나타나는 Non-rem 수면 상태와 다르게 NR2의 Feature importance는 FFT - r_delta가 FFT - f_delta에 비해 높게 나타나지만 다른 상태에서는 반대로 FFT - f_delta의 feature importance가 높게 나타난다.

하지만 이것이 model의 score가 낮게 나오는 직접적인 원인이 아닐 수 있기 때문에 Non-rem 수면 상태 중 신호 처리 방법에 따른 모델별 예측 값의 차이가 다소 발생한 NR3의 feature importance를 비교해 본 결과 아래와 같다.

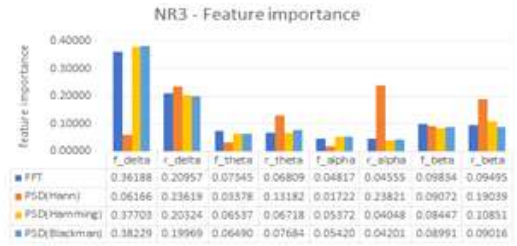


Figure 6. NR3상태 신호처리 방법에 따른 Feature importance

NR3에서 PSD(Hann)의 score가 다른 model 들에 비해 낮게 score가 나왔는데 feature 별 importance를 보면 PSD(Hann)의 f_delta에 대한 importance가 낮고 활동 중에 나타나는 머리 뒤쪽(r) alpha 파와 beta 파의 importance가 높게 설정되어 결과적으로 다른 model 들에 비해 score가 낮게 나타났다. 이를 통해 Non-rem 수면 예측 모델에서 f_delta의 importance가 높아야 예측률이 높게 나올 것이라는 것을 알 수 있다.

III. 결론

본 논문에서는 2-channel의 수면 시 EEG 데이터를 통해 Gradient boosting classifier로 machine learning 학습을 진행하고 신호처리 방법에 따른 score의 차이와 score가 낮은 경우 낮게 나온 원인을 모델별 feature importance를 통해 확인하였다. 대부분의 뇌파 연구 데이터는 많은 channel의 데이터를 사용하기 때문에 일상적으로 사용 가능한 제품에 적용하는 데에는 한계가 있다. 하지만 신호처리를 통해 한 개의 channel의 데이터를 여러 개의 feature로 나누면 적은 수의 channel의 데이터로도 활용이 가능하다는 것을 확인하였다. 또한 신호처리 방법이 머신러닝 결과 차이를 통해 뇌파를 이용한 머신러닝 진행 시 머신러닝 알고리즘 뿐만 아니라 뇌파 신호를 어떻게 처리하는지도 중요한 요소임을 확인할 수 있다.

ACKNOWLEDGMENT

이 논문은 한국연구재단의 이공분야기초연구(기본연구)의 지원을 받아 수행된 연구임(NRF-2019R1F1A1063128).

참고 문헌

- [1] B Kemp, AH Zwinderman, B Tuk, HAC Kamphuisen, JJJ Oberyé. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. IEEE-BME 47(9):1185-1194 (2000).
- [2] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215 - e220.
- [3] Anis Ameera, A. Saidatul, Z Ibrahim. Analysis of EEG Spectrum Bands Using Power Spectral Density for Pleasure and Displeasure State. Anis Ameera et al 2019 IOP Conf. Ser.: Mater. Sci. Eng. 557 012030
- [4] Qi Xiong, Xinman Zhang, Wen-Feng Wang, Yuhong G. A Parallel Algorithm Framework for Feature Extraction of EEG Signals on MPI. Computational and Mathematical Methods in Medicine, vol.2020, 10 pages, 2020.
- [5] Friedman, J. H. (February 1999). "Greedy Function Approximation: A Gradient Boosting Machine".
- [6]

scikit-learn.org,
https://scikit-learn.org/stable/modules/permutation_importance.html