Rolling window clustering analysis for Electroencephalograms of Athletes

Abstract

1. Introduction

2.Method

2.1 Method\_overview

In this paper, the proposed method to cluster channels comprises of three stages. In the first stage, rolling window FFT(Fast Fourier Transform) spectral analysis has been used to extract important frequency features from EEG signal, and after this stage, those features were then combined for each subject. In the second stage, auto-Kmeans clustering has been used to cluster channels on each individual subject using the features produced by first stage, and a [cluster ring] was produced from the auto kmeans. In the third stage, Jaccard Distance measure has been used to determine the distance between two channels cross the whole subject set, and 4 clusters were determined. We have explained the subsections of proposed method in following sections.

?? Where do I put different time-interval settings?

We have applied our method on different time-interval settings, and have obtained the same results for adjusted threshold on auto-kmeans.

?? Where do I put different wavelengths settings?

We have applied our method to pruned data on different wavelengths, namely delta (1–3.5 Hz), theta (4–7.5 Hz), alpha (8–12 Hz),beta(13-32 Hz), gamma(32-45 Hz).

[A flow chart should be added here explain the above process]

2.1.1 Rolling window fft spectral analysis

As a feature extraction process, we have used rolling window fft analysis method to transform EEG signals from time domain to frequency domain.

[Transform equation here]

[Explain transform equation here]

[Show the before-transform window raw data here, with correct labeling]

[Show the after-transform window fft here, with correct labeling]

2.1.2 auto-Kmeans clustering on each athelet

Clustering algorithms are used widely to collect similar or dissimilar data. The most used clustering algorithms are k-means clustering, and we built on that to automate the k selection process.

[kmeans algorithm here in pseudo code]

The challenge here is to determine when we should stop the kmeans. We applied ?? to automatically determine the stopping of our kmeans iterations.

[Stopping increase k distance code here] (sscluster)

pF = (distortion/(clusterNumber-1))/((ssBetween)/(n-clusterNumber)) --- simplify this thing

[Why we chose it] : We have compared this threshold function with pure distortion, standardized distortion, BIC?, AIC? And we found this to be the best as shown on scree plot.

[Select some scree plot to show]

A [cluster ring] is obtained for each subject, and we group them into a dictionary.

2.1.3 Cross\_athelet distance definition for all channels

For the obtained [cluster ring], namely, grouped channel dictionary, we have defined distance between pairs of channels by jaccard distance.

[Equation here for cross\_athelet\_jaccard\_distance]

Basically, the time two channels are together / two channels present.

Software??

Analysis

Conclusion

Future direction

References