6/17/2016

Today's Goal: Find literature about features for classifying EEG/EMG data.

FASTER: an unsupervised fully automated sleep staging method for mice

Properties of the sleep states:
NREM- Large and slow EEG waves, low EMG amplitude
REM- Lower and faster EEG waves, very low EMG amplitude
Wake- EEG similar to NREM, but high EMG amplitude

For features, some automated systems use specific bands of the EEG power spectrum:
Delta, 0.5-4 Hz
Theta, 6-10 Hz
Sigma, 11-15 Hz

Other features include coefficients from Wavelet analysis, bispectral density, parameters of multichannel autoregressive modeling, and matching pursuit method with slow wave patterns.

Feature dimensionality can be reduced through Principal Component Analysis without loss of critical information.

Unsupervised methods have "hard rules" for classifying. Since the late 1980s, supervised sleep staging programs incorporating sophisticated machine learning algorithms have become more popular, due to their ability to improve the accuracy of staging by including the variances between subject animals.

When taking "epochs" (4-10s time windows), sometimes they have borderline characteristics of 2 sleep states. To solve this, some unsupervised approaches list multiple classifications with the degree of certainty for each class.

Used non-parametric density estimation clustering, which clusters by distribution of data rather than distance between data. This was to keep the data "as objective as possible" and not cluster with a known number of clusters beforehand.

Used a wide range signal bin and used PCA to reduce dimensionality to improve computation time.

Overview of steps in their method:
1. Power spectrum of each epoch is calculated by Fast Fourier Transform.
2. EEG/EMG power spectra are transformed into Principal Components via PCA. Top 4 are used.
3. Components are clustered through non-parametric density estimation clustering method. Estimates probability density of dataset and defines a cluster has a high density area connected with Delaunay triangulation.
4. Clusters are annotated based on avg EEG/EMG powers.

Parameter optimization:
1. Character Extraction- Optimized number of principal components for high performance and low computation time.
2. Clustering- Bandwidth of density estimation kernel (smoothing factor) and number of grids for dataset size.
3. Annotations- Optimize the threshold values determined manually.

They used the entire EEG/EMG power instead of certain bands. This was intended to account for individual variability.

Does not take into account the transitions.

Automatic Sleep Stages Classification Using EEG Entropy Features and Unsupervised Pattern Analysis Techniques

Aims to use an unsupervised approach to classifying. Calculates a bunch of features, uses Q-a algorithm to select relevant ones, and then uses J-means clustering to classify.

The aim of the feature extraction stage was to compare the performance of different entropy estimators when applied to the experimental set, and find out if some metrics are more sensitive to EEG patterns. Below were the features used:

Fractal Dimension-statistically quantifies how well a fractal matches the input data at different scale.

Detrended Fluctuation Analysis- allows the detection of long-range power-law correlations in a time series

Shannon Entropy- Measure of data spread, where p(xi)=xi^2

Approximate Entropy- Regularity of a time series

Sample Entropy- evolution of ApEn devised to solve the bias of ApEn due to counting self-matches.

Multi scale Entropy- Estimator of the complexity of the time series

Q-a took these metrics and found the most important ones.

J means algorithm:

(1) Initialization: A standard k-means clustering is used to set an initial partition of the feature vectors and the centroids. This reduces the temporal cost of the partition calculation.
(2) Search: Given a tolerance threshold (4 standard deviations of the intra-cluster distance), find the unoccupied points (feature vectors that do not belong to any cluster).
(3) Update: Add a new cluster centroid at some unoccupied location and find the index of the best centroid to delete. Update the partition according to the new centroids.
(4) Finalize: If a local minimum is found in the previous iteration, stop. For each resulting cluster, a sleep stage can be assigned as the most frequent class (using a k-Neighbors method), which in clinical practice could be done by a whole cluster manual scoring. Otherwise return to step 2.

Next time: Read Unsupervised Online Classifier in Sleep Scoring for Sleep Deprivation Studies