EEG expert features for GP classification [May 17]

This text continues the results of May 9th. The discussed plan was:

1. Check if the onset is stated properly
2. Make the data zero-mean and Energy-normalized
3. Engineer the interpretable features, according to the SchRes paper
4. Run GP and compare various kernels with all features (many channels)
5. Select the features according to the ERP appearance time; first goes first
6. Select the features with PCA, and reduce dimensionality with the jackknife

1. Check if the onset is stated properly

Ok

2. List of used scalers

MinMaxScaler, MaxAbsScaler, StandardScaler, RobustScaler, Normalizer, QuantileTransformer, PowerTransformer.

3. Engineer the interpretable features

See Section List of the expert-engineered features and Appendix below.

4. List of Kernels for Gaussian Process Classification

Basic kernels: RBF, Maternel, Rational Quadratic, Dot Product, Exponential Sine Quadratic.
Composite kernels: White, Ridge, Periodic, Irregular, Noisy, Rumble.

5. Select the features according to the appearance time

See Section List. It is aligned by the starting time.

6. Principal Component Analysis

The optimal number of components is 4, according to the CV accuracy.

Preliminary results for the expert features from Table 1.

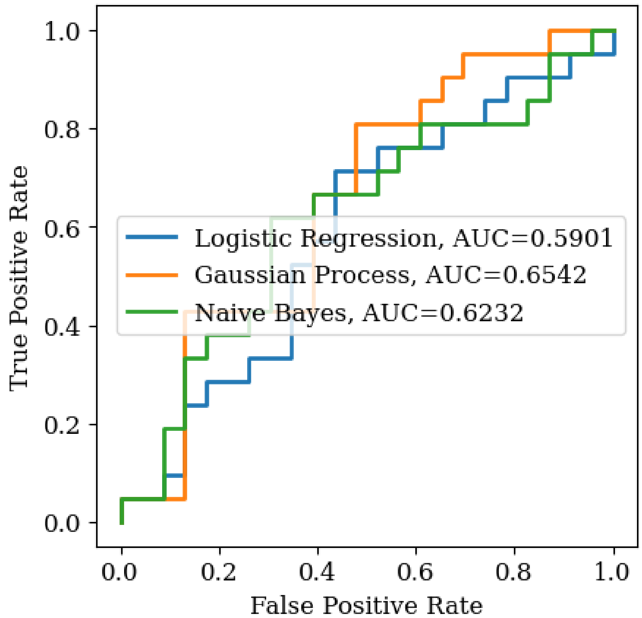

Figure 1. The AUC curve for expert-engineered features is from Table 1. The task is recognition; classes are Old versus New words, Incorrect responses dropped

Warning, as earlier. Until we estimate the sufficient sample size, the accuracy and AUC are unstable and meaningless: overtraining on the test set.

List of the expert-engineered features

Table 1. Features are the average time segment collected from the expert note and the SchRes paper. For the sources, see Appendix and the cap map.

Electrode	Peak	After onset	End	Comment
PO9	P1	110	160	SchRes paper starts, recognition
PO10	P1	110	160	recognition
PO7	P2	200	410	recognition
PO8	P2	200	410	recognition
PO7	P3	410	650	recognition
PO8	P3	410	650	recognition
AF7	LNP	500	800	encoding, item-specific
AF7	LNP	400	750	recognition, associative

Other sources: Notes and Wikipedia

Electrode	Peak	After onset	End	Comment
O1	N1	150	200	Notes start
C2	N400	250	500	or
C2	N400	300	350	
P3	LPC	400	500	or
P3	LPC	400	800	
P2	LPC	500	700	right
P4	LPC	500	700	
P8	LPC	500	700	
Fz	LPC	450	700	
F7	LPC	450	700	
F8	LPC	450	700	
any	N1	80	120	wikipedia starts
any	N170	130	200	faces, familiar objects or words
any	N1 visual	150	200	
any	N2	200	350	
any	N400	250	500	including visual and auditory
any	P200	150	275	
any	P300	300	400	
any	P3a	250	280	
any	LPC	400	500	explicit recognition memory
any	P600	500	800	

Code shared for review

The talk on May 10th suggested sharing the reproducible code. This code uses proprietary data. The code itself and, later, some adequate demo open data are https://github.com/vadim-vic/EEG-ERP-precog.

Appendix

List of the engineered features, sources

Collected from the manuscript note
O1, N1: 150 to 200 ms, the first negative peak, take average in this window
C2, N400: 250-500 ms (300-350 ms), take average in this window
LPC (late positive component), P3: beginning 400-500 ms and lasting for a few hundred milliseconds, average between 460-800 ms
P2, P4, P8, Fz, F7, F8 left-right, 500-200 ms
Last LPC component, F8 right, F7 left
Collected from SchRez https://doi.org/10.1016/j.schres.2023.02.019
See also https://doi.org/10.1093/arclin/acx082
Encoding
P1 at PO9/PO10 (110-160ms),
P2 at PO7/PO8 (240-320ms),
and the late negative potential (LP) at AF7 (500-800ms).
For recognition, we quantified P1 at PO9/PO10 (110-160ms),
P2 at PO7/PO8 (200-410ms),
P3 at PO7/PO8 (410-650ms), and the
LNP at AF7 from 500-800 ms during item-specific recognition and 400-750 ms during associative recognition.

Scott R Sponheim: A late negative potential (LNP) over left frontal brain regions during recognition was larger for relationally encoded objects than new and item-specific encoded objects in HCs. This pattern was absent for SZ and SZr. Smaller P2 and LNP components were associated with greater self-reported cognitive-perceptual abnormalities. Early posterior brain responses likely relevant to perceptual functions supporting memory formation were diminished in schizophrenia. Late frontal electrophysiological responses associated with relational aspects of memory appear diminished in SZ and SZr, potentially reflecting the influence of genetic liability for schizophrenia on the brain.

An early posterior component (P2) during encoding predicted later recognition and was diminished in SZ. A late negative potential (LNP) over left frontal brain regions during recognition was larger for relationally encoded objects than new and item-specific encoded objects in HCs. This pattern was absent for SZ and SZr. Smaller P2 and LNP components were associated with greater self-reported cognitive-perceptual abnormalities.

During both retrieval tasks, after each response of “old” or “new,” there was a 1000ms pause

Miscellaneous remarks

Collect all ERP-related features and statistics
Peaks before 200 are too early (160 ms is too fast)
Weight the electrodes with the biggest noise
Emotions in the front temple, frontal might be more informative
Peaks moving a little in time
Compare two classes' latent spaces of minimum dimensionality and make classifications on them in the style of semi-supervised learning
Hypothesis: each patient might have a different informative set of electrodes
Do not make filtering when averaging
The promised models: Linear dynamics SSM and Gaussian process from GPPython
AW: *…(and eventually Ali and Vadim), the correct thing to be decoding is on the REcognition block of the VerbMem dataset. For now, take only the trials where the participant answered correctly. The binary classification I want to get out is "old previously seen word" vs. "new not ever seen word."*

Code sources

1. Pipelining: chaining a PCA and logistic regression [**Grid search PCA+Logistic**]
2. AUC plot for several models
3. Scalers in pipelines [Usage before or after CV is discussable for EEG]
4. Comparing randomized search and grid search for hyperparameter estimation
5. Examples using sklearn.model_selection.RandomizedSearchCV
6. Tuning the hyper-parameters of an estimator
7. Pipelines and composite estimators [**Example of PCA dim reduction + logistic**]
8. Sample pipeline for text feature extraction and evaluation
9. Displaying Pipelines
10. Selecting dimensionality reduction with Pipeline and GridSearchCV
11. Concatenating multiple feature extraction methods [useful too]
12. List of sklearn examples, and Machine learning examples

Feature extraction notes

Feature selection principles

Select electrodes
Select features
Select time segments
Collinearity analysis

Two variants of data:

Collected from the raw data [name the module]
Collected from the prepared data [point to the data structure]
For both datasets, we have 1) all electrodes to select and 2) a small non-reducible set assigned by the experts.

The data does not keep records with no user responses since, in this case, nobody guarantees the user's attention.
Cut to 800 ms
Hyperparameters
Along the time of one electrode
Mean, variance, histogram, histogram of the difference between smooth and error
Average in bins, see for example Shrez
Distance between electrodes time convoluted
Distance between time decimated
Over a group of electrodes
Between groups of electrodes
Movement of the activation zone over time and space

Links to GP code and references

Gaussian Process Regression With Python
Youtube GP explanation
Jwangjie Gaussian-Processes-Regression-Tutorial
Sklearn Gaussian Processes regression: basic
GP sklearn kernels

Feature extraction

1. Fast Fourier Transform (FFT), Wavelet Transform(WT), Time-Frequency Distribution (TFD)
2. SoftwareX, eeglib: A Python module for EEG feature extraction {List of composite features} FFT, Band Power, Synchronization Likelihood, Petrosian and Higuchi Fractal Dimensions, Hjorth Parameters, Detrended Fluctuation Analysis, Sample Entropy, Lempel-Ziv Complexity, Cross Correlation Coefficient
3. Data Science for Psychology and Neuroscience in Python, Course {Very introduction level}
4. MNE-Python data loading and preprocessing {no feature engineering discovered}
5. Linear classifier on sensor data with plot patterns and filters {**Check it in depth!!!**}
6. Machine learning examples {**Check!!**}
7. Analyzing continuous features with binning and regression in sensor space {**Check it in depth!!!**}
8. https://mne.tools/stable/auto_tutorials/evoked/30_eeg_erp.html
9. Averaging across channels with regions of interest
10. eeg signal classification python code {Nice search }
11. **multitaper** is a spectral density estimation
12. Global field power (GFP) Global field power [1][2][3] is, generally speaking, a measure of agreement of the signals picked up by all sensors across the entire scalp: if

Feature engineering list Adelph, see github

Amplitude
Mean
Median
Variance
Standard deviation
Root mean square (RMS)
Skewness
Kurtosis
Zero-crossing rate
Peak-to-peak amplitude
Mean absolute deviation (MAD)
Signal energy
Autocorrelation
Hjorth parameters (activity, mobility, and complexity)
Interquartile range
Crest factor
Shape factor
Impulse factor
Frequency Domain:
Amplitude
Mean
Median
Variance
Standard deviation
Root mean square (RMS)
Skewness
Kurtosis
Zero-crossing rate
Peak-to-peak amplitude
Mean absolute deviation (MAD)
Signal energy
Autocorrelation
Hjorth parameters (activity, mobility, and complexity)
Interquartile range
Crest factor
Shape factor
Impulse factor
Time-Frequency:
Short-time Fourier transform (STFT)
Continuous wavelet transform (CWT)
Discrete wavelet transform (DWT)
Wavelet packet transform (WPT)
Wavelet coherence
Wavelet phase coherence
Wavelet entropy
Wavelet energy
Scalogram
Time-frequency distributions (e.g., Wigner-Ville distribution, Choi-Williams distribution)
Hilbert-Huang transform (HHT)
Empirical mode decomposition (EMD)
Stockwell transform (S-transform)
Spectrogram
Time-frequency reassignment (TFR)
Synchrosqueezed wavelet transform (SWT)
Complex Morlet wavelet transform
Other features:
The directed transfer function (DTF)
Partial directed coherence (PDC)
Granger causality
Multivariate autoregressive modeling (MVAR)
Symbolic dynamic analysis (SDA)
Sample entropy
Fuzzy entropy
Permutation entropy
Lyapunov exponent
Detrended fluctuation analysis (DFA)
Recurrence quantification analysis (RQA)
Approximate entropy (ApEn)
Lempel-Ziv complexity (LZC)
Independent component analysis (ICA)
Principal component analysis (PCA)
Canonical correlation analysis (CCA)
Joint time-frequency entropy
Mutual information
Phase-amplitude coupling (PAC)
Cross-frequency coupling (CFC)