EEG exhaustive model comparison [Apr 26]

This text continues the results of Apr 19th. The conclusions were:

1. We want to model the situation where class 1 is "old word" (seen before), class 2 is "new word" (not seen before), and we want to decode what is true.
2. Focus on the accuracy level over the whole block vs. the trial – the trial, except as noted above; I am not so interested in the accuracy.

The computational experiment and classification accuracy

In the classification of Task 3 Recognition, class 1 is "old word" (seen before) stimulus, and class 2 is the "new word" stimulus. We classify stimulus (class 1 and 2) independent of the user response (correct, incorrect) and reaction time.

Dataset: Verbal Memory

Objective: to assess classification accuracy per cross-validated sense per user

Model and features: logistic regression with features, generated by the extended covariance matrix of the electrodes' time series. See Section Model and feature generation.

- The classification object is an Event, regardless of any user reaction or lack thereof (see the table titles).
- The input data: the EPR time segment starts in the stimulus onset and lasts 1200 ms. The data include EEG recorded from 128 channels BioSemi headcap at 256 Hz sampling rate.
- The quality criterion: average accuracy on the 10 folded cross-validated data

Table 1. Classification accuracy per user (Recognition, the cases with incorrect answers dropped out). These are non-verified preliminary results, which could change drastically after verification and error analysis.

User	Logistic Regression	Nearest Neighbors	Linear SVM	RBF SVM	Gaussian Process	Decision Tree	Random Forest	Neural Net	AdaBoost	Naive Bayes	Quadratic DA
1037	**0.9**	0.9	0.92	0.5	0.9	0.82	0.82	0.9	0.82	0.92	0.59
1045	**0.99**	0.99	0.99	0.46	0.99	0.97	0.9	0.99	0.93	0.99	0.75
1158	**0.97**	0.97	0.97	0.54	0.97	0.96	0.93	0.97	0.94	0.97	0.64
1363	**0.77**	0.75	0.8	0.47	0.78	0.6	0.7	0.79	0.67	0.81	0.46
1368	**0.5**	0.55	0.41	0.38	0.52	0.5	0.45	0.52	0.56	0.49	0.48
1385	**0.7**	0.74	0.59	0.59	0.69	0.55	0.45	0.69	0.6	0.67	0.59
1034	**0.51**	0.5	0.56	0.6	0.5	0.47	0.54	0.51	0.56	0.53	0.56
2038	**0.89**	0.87	0.9	0.35	0.89	0.72	0.87	0.89	0.78	0.9	0.62
6639	**0.9**	0.78	0.92	0.56	0.84	0.88	0.82	0.88	0.82	0.92	0.74
7974	**0.74**	0.72	0.78	0.52	0.74	0.64	0.64	0.74	0.68	0.78	0.54

User	Logistic Regression	Nearest Neighbors	Linear SVM	RBF SVM	Gaussian Process	Decision Tree	Random Forest	Neural Net	AdaBoost	Naive Bayes	Quadratic DA
7977	**0.84**	0.79	0.85	0.59	0.83	0.76	0.85	0.84	0.76	0.85	0.58
7980	**0.98**	0.98	0.98	0.54	0.98	0.97	0.97	0.98	0.98	0.98	0.78
1327	**0.59**	0.57	0.61	0.61	0.57	0.54	0.47	0.54	0.52	0.57	0.47
mean	0.79	0.77	0.79	0.51	0.78	0.72	0.72	0.78	0.74	0.79	0.6
min	0.5	0.49	0.4	0.34	0.49	0.46	0.44	0.51	0.51	0.49	0.46
max	0.99	0.99	0.99	0.61	0.99	0.96	0.96	0.99	0.98	0.99	0.78

Table 1. (continues for models). Classification accuracy per user (Recognition, the cases with incorrect answers dropped out). These are non-verified preliminary results, which could change drastically after verification and error analysis.

User	GP rbf	GP matern	GP quad	GP linear	SVM	
1037	0.92	0.92	0.92	0.9	0.9	
1045	0.92	0.92	0.92	0.9	0.9	
1158	0.99	0.99	0.99	0.99	0.99	
1363	0.97	0.97	0.97	0.97	0.97	
1368	0.8	0.8	0.8	0.78	0.78	
1385	0.45	0.44	0.45	0.49	0.46	
1034	0.64	0.54	0.64	0.65	0.67	
2038	0.59	0.59	0.59	0.5	0.53	
6639	0.9	0.9	0.9	0.89	0.83	
7974	0.92	0.92	0.92	0.9	0.82	
7977	0.78	0.78	0.78	0.74	0.74	
7980	0.85	0.85	0.85	0.84	0.83	
1327	0.98	0.98	0.98	0.98	0.98	
mean	0.8	0.79	0.8	0.78	0.77	
min	0.44	0.43	0.44	0.49	0.46	
max	0.99	0.99	0.99	0.99	0.99	

Table 2. Classification accuracy per user (Recognition, correct, incorrect, and the other responses are included). These are non-verified preliminary results, which could change drastically after verification and error analysis.

	Logistic Regression	Nearest Neighbors	Linear SVM	RBF SVM	Gaussian Process	Decision Tree	Random Forest	Neural Net	AdaBoost	Naive Bayes	Quadratic DA
1037	0.5	0.43	0.37	0.31	0.46	0.5	0.46	0.49	0.49	0.52	0.43
1045	0.74	0.76	0.75	0.4	0.73	0.52	0.52	0.72	0.64	0.68	0.58
1158	0.51	0.46	0.56	0.49	0.51	0.52	0.58	0.51	0.47	0.61	0.49
1363	0.55	0.57	0.57	0.52	0.57	0.57	0.53	0.58	0.53	0.56	0.44
1368	0.69	0.66	0.61	0.41	0.66	0.52	0.47	0.64	0.71	0.64	0.49
1385	0.5	0.54	0.54	0.47	0.52	0.47	0.52	0.55	0.52	0.55	0.47
2038	0.57	0.63	0.47	0.41	0.56	0.5	0.45	0.51	0.45	0.42	0.47
6639	0.47	0.47	0.43	0.38	0.49	0.38	0.49	0.49	0.42	0.49	0.52
7974	0.7	0.66	0.75	0.4	0.6	0.56	0.59	0.72	0.62	0.75	0.55
7977	0.6	0.56	0.53	0.53	0.61	0.44	0.54	0.61	0.57	0.56	0.44
7980	0.67	0.6	0.64	0.54	0.66	0.59	0.59	0.65	0.58	0.66	0.45
1327	0.73	0.61	0.59	0.5	0.71	0.42	0.53	0.67	0.5	0.61	0.59
1034	0.56	0.48	0.55	0.43	0.54	0.55	0.55	0.58	0.5	0.47	0.46
mean	0.6	0.57	0.57	0.45	0.59	0.5	mean	0.59	0.54	0.58	0.49
min	0.47	0.43	0.37	0.31	0.46	0.38	min	0.49	0.42	0.42	0.43
max	0.74	0.76	0.75	0.54	0.73	0.59	max	0.72	0.71	0.75	0.59

Table 3. The accuracy of classification for Old versus Non word (all responses included, event codes 150 170). Draft.

User	Logistic Regression	Nearest Neighbors	Linear SVM	RBF SVM	Gaussian Process	Decision Tree	Random Forest	Neural Net	AdaBoost	Naive Bayes	Quadratic DA
1034	0.62	0.68	0.66	0.56	0.6	0.42	0.62	0.59	0.59	0.66	0.54
1037	0.56	0.5	0.55	0.53	0.57	0.52	0.5	0.61	0.62	0.58	0.48
1045	0.69	0.67	0.71	0.53	0.64	0.55	0.5	0.67	0.64	0.67	0.47
1158	0.6	0.6	0.54	0.42	0.6	0.46	0.52	0.6	0.49	0.53	0.57
1363	0.53	0.57	0.54	0.47	0.52	0.5	0.49	0.52	0.57	0.49	0.53
1368	0.36	0.46	0.38	0.4	0.37	0.48	0.48	0.38	0.46	0.39	0.49
1385	0.63	0.61	0.61	0.5	0.56	0.54	0.51	0.63	0.65	0.63	0.53
2038	0.46	0.47	0.49	0.56	0.46	0.46	0.49	0.49	0.41	0.38	0.47
6639	0.56	0.54	0.52	0.41	0.54	0.44	0.49	0.53	0.55	0.6	0.46
7974	0.55	0.5	0.56	0.4	0.49	0.55	0.55	0.5	0.57	0.57	0.57
7977	**0.72**	**0.63**	**0.71**	**0.53**	**0.71**	**0.56**	**0.54**	**0.7**	**0.64**	**0.68**	**0.49**
7980	0.57	0.58	0.57	0.51	0.55	0.54	0.5	0.57	0.55	0.52	0.5
1327	0.6	0.61	0.64	0.57	0.6	0.6	0.63	0.63	0.64	0.59	0.5

Table 4. The accuracy of classification for New versus Non word (all responses included, event codes 150 170). Draft.

User	Logistic Regression	Nearest Neighbors	Linear SVM	RBF SVM	Gaussian Process	Decision Tree	Random Forest	Neural Net	AdaBoost	Naive Bayes	Quadratic DA
1034	**0.53**	0.53	0.54	0.45	**0.48**	0.39	0.49	0.53	0.5	0.45	0.48
1037	**0.56**	0.49	0.51	0.46	**0.53**	0.58	0.43	0.54	0.55	0.56	0.46
1045	**0.6**	0.58	0.6	0.58	**0.6**	0.5	0.66	0.58	0.59	0.54	0.45
1158	**0.45**	0.48	0.43	0.57	**0.46**	0.54	0.55	0.47	0.49	0.51	0.48
1363	**0.5**	0.45	0.49	0.41	**0.52**	0.53	0.48	0.5	0.54	0.56	0.64
1368	**0.61**	0.6	0.61	0.55	**0.61**	0.62	0.47	0.6	0.56	0.54	0.49
1385	**0.64**	0.54	0.64	0.54	**0.64**	0.58	0.49	0.62	0.6	0.69	0.52
2038	**0.53**	0.44	0.45	0.35	**0.5**	0.41	0.48	0.52	0.47	0.36	0.52
6639	**0.5**	0.55	0.49	0.46	**0.49**	0.46	0.51	0.49	0.48	0.49	0.45
7974	**0.54**	0.56	0.52	0.41	**0.52**	0.53	0.54	0.55	0.57	0.54	0.53
7977	**0.66**	**0.62**	**0.6**	**0.53**	**0.66**	**0.63**	**0.5**	**0.65**	**0.63**	**0.63**	**0.52**
7980	**0.55**	0.6	0.55	0.47	**0.56**	0.42	0.55	0.56	0.59	0.6	0.52
1327	**0.5**	0.54	0.49	0.48	**0.52**	0.49	0.41	0.53	0.48	0.44	0.54

Table 5. The average accuracy for **all non-aligned users in one set** on all tasks (all responses included), the model is logistic regression

Task	Classes	Class 1 sample size	Class 2 sample size	Accuracy
Encoding	Larger Smaller	657	593	0.53
Lexical	Old New	660	589	0.54
Recognition	Old New	617	598	0.54
Lexical	Old Non	660	645	0.60
Lexical	New Non	589	645	0.56

Table 6. The class balance, each user per task

	Task	Classes	1034	1037	1045	1158	1363	1368	1385	2038	6639	7974	7977	7980	1327
Class 1	Encoding	Large Small	41	50	60	62	43	45	59	38	50	48	61	44	56
	Lexical	Old New	48	51	55	49	57	46	51	48	51	46	56	52	50
	Recognition	Old New	55	50	45	53	54	41	37	44	52	45	55	57	29
	Recognition	Old Non	48	51	55	49	57	46	51	48	51	46	56	52	50
	Recognition	New Non	40	49	50	49	46	42	49	38	48	42	49	53	34
Class 2	Encoding	Large Small	39	47	45	47	45	38	51	38	47	49	55	40	52
	Lexical	Old New	40	49	50	49	46	42	49	38	48	42	49	53	34
	Recognition	Old New	53	46	44	51	55	41	37	46	48	45	54	49	29
	Recognition	Old Non	50	55	56	49	49	43	53	48	50	43	55	49	45
	Recognition	New Non	50	55	56	49	49	43	53	48	50	43	55	49	45

Organizational questions

1. To compare the classification results **can we use** not raw data but **prepared** data in the format of a **3-way matrix** X (in Events x Electrodes x Time) and a class ground truth vector Y (in Events)? The supplementary info is the sample rate, tmin, tmax, and electrode placement, and the other session features. A unified format like FIF is appreciated.
   This helps us avoid misunderstanding what classes we should classify.
2. Is the Verbal Memory data **our final dataset?** Do we expect augmentation of these data or chance the dataset (that could case the nature of data and conserquenty, type of models).

Conclusion and plans

1. This very same dataset shows a significant difference in classification quality, depending on the problem statement (**0.94 to 0.50** in accuracy). It means that
    a. The way of collecting data is much more important than the type of model
    b. The model must be as simple as possible with good generalization ability.
2. Due to the dataset's insufficient size, dozens of samples per user, and hundreds per cohort, the model must be well-interpreted. Deep learning models will cause overtraining and a lack of generalization ability due to the backpropagation of the class error. The only complex part could be feature generation and transformation without the classification part.
3. Our next move is to make the error analysis and exhaustive search for the model's hyperparameters.
4. Despite modeling differences between users, to augment the size of the sample set, we will include an alignment module to make users more similar in the common feature space. We examine time alignment to align the user reaction and self-modeling to align the signal amplitude.
5. The next move is to run models SSM, DGD, and continuous-time and space-time models. We include graph-convolution models to refer to the brain functional groups.
6. If the size of the current dataset remains the same, we shall use the open-source collections of similar datasets.

The model and feature generation (draft)

The goal is to classify an ERP in an EEG time segment. There given a set of M EEG measurements $\mathbf{X} \in \mathbb{R}^{N \times T}$, where N is the number of electrodes and T is the length of the time segment. The target variable $y \in \{0, 1\}$ indicates one of two classes of EPR[1]. The goal is to construct and optimize the classification model $f(\mathbf{w}, \mathbf{x})$ parameters. The model approximates the target variable y according to the ERP description $\mathbf{x}$ obtained by transforming the measurements $\mathbf{X}$. The quality criterion of the model is the binary cross-entropy

$$\mathcal{L}(\mathbf{w}) = -\frac{1}{M} \sum_{m=1}^{M} \Big(y_m \log f(\mathbf{w}, \mathbf{x}_m) + (1 - y_m) \log\big(1 - f(\mathbf{w}, \mathbf{x}_m)\big) \Big) - c\|\mathbf{w}\|_2,$$

where c is some fixed regularization coefficient. The optimization problem is

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w}} \mathcal{L}(\mathbf{w}), \text{ given the data set } \{\mathbf{x}_m = \boldsymbol{\varphi}(\mathbf{X}_m), y_m\} \text{ and the model } f. \tag{1}$$

To solve this problem we use the logistic regression model $f(\mathbf{w}, \mathbf{x}) = \big(1 + \exp(-\mathbf{w}^{\mathsf{T}}\mathbf{x})\big)^{-1}$.

The transformation $\varphi(\mathbf{X})$ produces a vector representation $\mathbf{x}$ by one of two methods. According to the first method, each channel's matrix rows, the time series, are recorded sequentially

$$\mathbf{X}_i = \{\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_N\}, \text{ where } \mathbf{x}_j \in \mathbb{R}^T \text{ for electrodes } j \in \{1, \dots .N\}.$$

The second method uses an extended covariance matrix of the electrodes' time series. It includes the event-related potential covariance matrix and the class average matrices.For one EEG measurement $\mathbf{X}_m$, this covariance matrix

$$\mathbf{C}_m = \frac{1}{T-1} \mathbf{P}_m \mathbf{P}_m^{\mathsf{T}}, \qquad \mathbf{P}_m = \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ \mathbf{X}_m \end{bmatrix}, \qquad \mathbf{C}_m \in \mathbb{R}^{3N \times 3N},$$

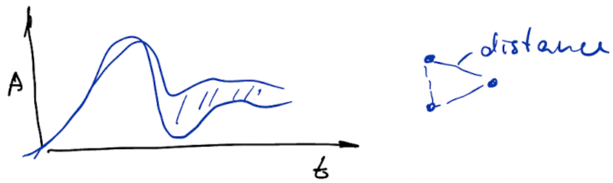
where $\mathbf{P}_0$ and $\mathbf{P}_1$ are average values for classes $\{0, 1\} \ni c$,

$$\mathbf{P}_c = \frac{\sum_{m=1}^{M} [y_m = c]\, \mathbf{X}_m}{\sum_{m=1}^{M} [y_m = c]}.$$

[1] New versus old word stimulus in the recognition task

Feature selection with genetic algorithms (plan)

Select the most informative group of electrodes. Reveal and compare these electrodes to functional groups. Make the importance analysis. Select the electrodes: discrete, the events: discrete, and the time segment weighted. Filter in time and space on the electrodes graph neighborhood. Include physics restrictions in the error function.



Convolution on space or functional groups
Convolution and tensor decomposition
Probabilistic diffusion flow over time

Table 7. Change in user's behavior by the number of incorrect and correct responses

User	**140**	1	2	11	12	**145**	1	2	11	12	**150**	1	2	11	12	**160**	1	2	11	12	**170**	1	2	11	12	**180**	1	2	11	12	**200**	1	2	11	12
1034	**36**	1	0	0	35	**26**	0	3	23	0	**42**	1	0	0	41	**37**	5	0	0	32	**46**	0	7	39	0	**49**	0	10	39	0	**49**	23	0	0	26
1037	**31**	2	0	0	29	**37**	0	0	37	0	**48**	1	0	0	47	**42**	1	0	0	41	**51**	0	3	48	0	**33**	1	4	28	0	**33**	1	0	0	32
1045	**58**	3	0	0	55	**45**	0	7	38	0	**55**	0	0	0	55	**50**	3	0	0	47	**54**	0	5	49	0	**43**	1	2	40	0	**44**	5	0	0	39
1158	**61**	7	0	0	54	**45**	0	6	39	0	**49**	1	0	0	48	**49**	6	0	0	43	**48**	0	2	46	0	**53**	1	8	44	0	**50**	11	0	0	39
1363	**39**	4	0	0	35	**45**	0	3	42	0	**50**	0	0	0	50	**39**	2	0	0	37	**49**	0	2	47	0	**53**	1	4	48	0	**49**	5	0	0	44
1368	**45**	2	0	0	43	**38**	0	2	36	0	**46**	0	0	0	46	**42**	1	0	0	41	**43**	0	2	41	0	**41**	1	3	37	0	**41**	6	0	0	35
1385	**53**	1	0	0	52	**41**	0	0	41	0	**51**	0	0	0	51	**49**	0	0	0	49	**51**	0	0	51	0	**35**	0	0	35	0	**31**	6	0	0	25
2038	**33**	1	0	0	32	**31**	0	2	29	0	**47**	2	0	0	45	**38**	3	0	0	35	**48**	0	1	47	0	**41**	1	6	34	0	**40**	7	0	0	33
6639	**39**	2	0	0	37	**42**	0	2	40	0	**47**	0	0	0	47	**47**	0	0	0	47	**49**	0	4	45	0	**27**	1	4	22	0	**29**	2	0	0	27
7974	**36**	3	0	0	33	**38**	0	3	35	0	**46**	0	0	0	46	**42**	0	0	0	42	**42**	0	3	39	0	**42**	1	1	40	0	**43**	3	0	0	40
7977	**61**	0	0	0	61	**54**	0	6	48	0	**56**	0	0	0	56	**49**	1	0	0	48	**54**	0	6	48	0	**54**	1	7	46	0	**48**	16	0	0	32
7980	**41**	3	0	0	38	**39**	0	3	36	0	**51**	1	0	0	50	**52**	7	0	0	45	**48**	0	1	47	0	**53**	0	10	43	0	**46**	8	0	0	38
1327	**51**	0	0	0	51	**43**	0	0	43	0	**46**	1	0	0	45	**32**	0	0	0	32	**42**	0	1	41	0	**26**	1	0	25	0	**22**	6	0	0	16