Abstract

The brain works in a distributed and dynamic manner. Information is represented and processed dynamically in multiple interconnected neural populations in the brain. One of the key aspects of cognitive and computational neuroscience is to understand how different information is represented in the brain. With the development of neural recording techniques, activity from larger and denser neural populations can be recorded simultaneously in vivo. This leads to new challenges for the statistical analysis of such large scale dataset. A specific question that we address in this project is that how to model and decode the neural activity associated with specific categorical visual input. In this project, we begin with a heuristic template projection method that implicitly utilizes the transient temporal profile and the spatial distributed property of the signal. We then try to approach and outperform this heuristic method with more general data-driven graphical models that account for the intrinsic spatial and temporal structure in the data. We show that the proposed graphical models can be used to effectively model and inference spatiotemporal neural data.

1. Introduction

One of the central questions in cognitive neuroscience is to understand how information is represented in the brain through different neural activities. The brain is a distributed and highly interconnected system, and information is usually represented and processed through the recurrent interactions between different neural populations across different scales. The structures and the properties of these networks of neural populations are closely related to sensory, motor, and cognitive functioning. Therefore identifying the interacting neural populations and understanding their patterns of activity associated with specific information processing are the main aim of many neuroscience studies.

The marvelous advances in data acquisition techniques, such as optical imaging (Ahrens et al., 2013), multielectrode array (Kipke et al., 2008), intracranial electroencephalography (iEEG) (Niedermeyer & da Silva, 2005), magnetoencephalography (MEG) (Hämäläinen et al., 1993), functional magnetic resonance imaging (fMRI) (Huettel et al., 2004) etc., have made large datasets of human brain activity with broad spatial coverage and fine temporal resolution available during resting state and during the execution of various cognitive functions. On one hand, the higher dimensionality of the data make it possible to go beyond univariate statistics and to ask questions from a population level (Abbott & Dayan, 1999; Averbeck et al., 2006). However, it also introduces new challenges for modeling and inference of the data. First, how do we model the data such that the both the spatial and the temporal properties of the signal are well described? Second, how do we make inference on the interactions between populations? In other words, how can we use the data to answer the ‘what’, ‘where’, and ‘when’ question (Ungerleider & Haxby, 1994), such as, “what information is being processed?” “where and when does this processing happen?” Fundamentally, we want to give a statistical model of such spatiotemporal data that accounts for both the spatial and the temporal characteristics of the recording, and use the model to predict cognitive states given observations of the neural activity. Furthermore, we would like to make statistical inference about the neural mechanism that allow for such prediction.

Specifically, our problem can be formulated as follows: given observations \( x = \{x_1, ..., x_T\} \), and states \( y = \{y_1, ..., y_T\} \), where \( x_i \in \mathbb{R}^p \) is the observation vector from \( p \) electrodes/sensors, and \( y_i \in \{C_1, C_2, ..., C_m\} \) is a discrete number identifying the corresponding stimulus condition/cognitive state at the moment, we would like to learn a model \( \theta = \arg \max_y \mathcal{L}(x, y; \theta) \), so that with new observations \( x' \), we can predict what is the corresponding state sequence, i.e. \( y' = \arg \max_y P(y|y'; \theta) \). The important feature in a desired model is that it has temporal structure in both \( x \) and \( y \), in addition, \( x \) also has
smooth priors that taken into account the spatial structures of the sensors/electrodes. Moreover, each unique type of state/stimulus would trigger a response with specific spatiotemporal profile that is only within a certain spatial and temporal range. For example, a stimulus image of a human face would trigger a strong response in fusiform face area that would last several hundred milliseconds (Ghuman et al., 2014). Therefore, there are two important aspects of the desired model. First, the model should be able to detect such cognitive events from the resting state when different stimuli are presented to the subject. Second, the model should be able to decode the contents of the stimuli that trigger the events.

In this project, we try to approach such a model from both generative and discriminative perspectives. We will try to evolve from a heuristic approach proposed by a recent study (Miller et al., 2016), and design graphical models based on generalized linear model (GLM), Gaussian process (GP), and conditional random field (CRF) that account for the spatiotemporal properties of the signal.

2. Related works
2.1. Generative models

With the rich spatial dimensions in the data, a natural way of modeling the data is using latent model. Specifically, we can assume that there is an ongoing hidden process that is driving the observed variables, which represent the changing states of the cognitive process. For cognitive processes with distinct stages and relatively longer time scale (this is often true for fMRI data), it can be modeled as a hidden Markov model (HMM). For example, HMM and HMM-like models are used to model the cognitive process of math problem solving using fMRI data (Anderson & Fincham, 2014). In addition to discrete sequences, continuous state space models are also used to model the neural signal with finer time scale, especially in eletrophysiology data from sensorimotor cortex. This can be as simple as a spatial latent model with no enforced temporal structure, such as principle component analysis or factor analysis (Nicolelis et al., 1995; Churchland et al., 2012), a temporal smoothing model, such as Kalman filter (Sadtler et al., 2014), or a more complex model that takes into account both dimensionality reduction and the continuity in temporal sequence (Yu et al., 2009). However, these state space models mainly serve as a dimensionality reduction approach, which would give a continuous state variable rather than a discrete prediction of the state condition as HMM-like methods do. In addition, most of these models do not fully exploit the spatial structure of the signal. This is partly because most of the models are applied to multielectrode array data, in which only the neural activity from a small local population is recorded and the geometry property of the spatial structure is not crucial for the model. Incorporating the geometric structure is more common when dealing with data on a full brain scale, such as fMRI, or scalp EEG, where topographical structures imposed are imposed in the latent space models. Specifically, radial basis functions can be used to represent latent variables and to include extra constraints on the latent variables based on the geometric structure of the brain (Gershman et al., 2011; 2014; Manning et al., 2017). However, the choice of prior distributions plays an important role in such latent state space models. It usually requires some heuristics to pick a proper prior distribution, which is often ad-hoc and limited to specific application cases.

2.2. Discriminative models

In addition to modeling the data itself, the supervised learning problem is also of great importance. This is also known as the ‘decoding’ problem: given the observed neural activities, can we predict the corresponding cognitive states or input visual stimulus? Instead of building a generative model as we mentioned in the previous section, decoding can be done by using discriminant model, which directly estimates the posterior \( P(y|x) \). Decoding can be done using nonparametric kernel density methods, or simply sliding window in time domain. In Ghuman et al. (2014), sliding time windows is applied to univariate time sequence of neural activity to decode the content of the stimuli. In Miller et al. (2016), template projection features are used, and a generative model (linear discriminant analysis) is used to decode the neural activity. These decoding methods can also be viewed as variants of (naive) Bayes model, with a temporal structure imposed by sliding time window. However, these studies only consider the temporal structure of the signal, without taking the information about spatial structure into their predictions. In fact, all the state space models have the inherent mechanism to serve as ‘decoder’, in which the hidden states are associated with the stimuli or cognitive states. As discussed above, for HMM-like methods, the inference problem is directly addressed by evaluating the posterior likelihood. For continuous state space models, an extra step of inference should be imposed in order to give a discrete state prediction.

2.3. Inference

There are two aspects of the inference problems that we are interested in. The first aspect is to make inference of some hidden states based on the observed variables. This part is overlapping with the previous part in many cases. For example, when we have discrete hidden variables representing the cognitive states or stimulus category, the inference for such hidden variables is fundamentally a supervised classification problem as described in the previous section.
Another aspect of the inference problem tries to probe the scientific interpretation of the data and asks which feature or subset of features actually enable the model to do predictions? This is related to graph structure estimation problem. The lasso-based approaches for recovering the graph structure of a static/dynamic Bayesian network have been proposed in several related works, such as Meinshausen & Bühlmann (2006); Friedman et al. (2008); Kolar et al. (2010). More importantly, we are interested in understanding which part of the brain is contributing to encoding the discriminative information about the stimuli and how the information is encoded.

3. Methods

3.1. Problem formulation

We first formalize the problem by describing the corresponding probabilistic model. Let’s consider the following problem:

Assume that we observe the recorded neural activity as a sequence of random vectors, and the cognitive states of the brain is also represented as a sequence of scalars. The observations are noted as \( x = \{ x_1, ..., x_T \} \) and \( y = \{ y_1, ..., y_T \} \), where \( x_i \in \mathbb{R}^p \) is the observation vector from \( p \) electrode contacts, and \( y_i \in \{ C_1, C_2, ..., C_m \} \) is a discrete number identifying the corresponding stimulus condition/cognitive state at the moment, including a baseline state with respect to the inter-stimulus intervals. And we also know the onset of the stimuli are marked at time stamps \( \{ t_1, ..., t_n \} \). We would like to learn a model \( \hat{\theta} = \arg \max L(x, y; \theta) \), so that with new observations \( x' \), we can identify the onsets of the stimuli from the baseline as \( \{ t_1', ..., t_m' \} \), and predict what is the corresponding state sequence, i.e. \( y' = \arg \max_y P(y|x'; \theta) \).

The graphical model for the problem is a Markov random field (MRF), as shown in Figure 1(a). Each observation \( x^{(i)} \) is a clique and there is horizontal links between the states. The cognitive states is dependent on the entire sequence of \( x \). This full model is in general hard to make inference. Therefore, we rely on specific simplifications to make the model tractable. In the following sections, we propose different levels of simplifications that approach the problem from different perspectives.

3.2. Template feature matching: the heuristic approach

This is a heuristic approach towards a discriminative model (Miller et al., 2016). According to the prevalent model of sensory information processing in the brain, visual stimuli can be represented through both the spatial and the temporal pattern of neural activity in the relevant brain regions (Kandel et al., 2000). The heuristic approach first estimates the temporal templates of evoked responses associated with each condition by taking the mean over a square window \( [t_i-l_b, t_i+l_e] \) around stimulus onset time \( \{ t_i \} \). Therefore the template for category \( j \) is

\[
M_j = \frac{1}{n_j} \sum_{y_i = C_j} [x_{t_i-l_b}, ..., x_{t_i+l_e}]^T
\]

where \( n_j \) is the total number of training samples from category \( j \). Then the distance to the template is used as the new feature

\[
z_{i,j} = D([x_{t_i-l_b}, ..., x_{t_i+l_e}]^T, M_j)
\]

For each set of \( \{ (y_i, z_{i,j}) \} | y_i = C_j \), we can learn conditional likelihood model \( P(z|y = C_j) \). Then for the testing phase, \( y' = \arg \max_y P(y|x'; \hat{\theta}) \).

It can be seen that this is close to a naive mean field approximation for a linear chain graph, where it ignores all the links between consecutive \( y_i \)’s and turns \( \{ x_j \} \) into \( \{ z_j \} \), where each \( z_j \) corresponds to an approximation of a clique of \( x_j \)’s around the corresponding \( y_j \). Therefore the model
becomes simply \( Z \leftarrow Y \), as shown in Figure 1. And after estimating the posterior probability \( P(y_i|z) \), the corresponding condition can be determined by setting a threshold on the posterior and find the condition that passes the threshold.

### 3.3. Generalized linear model + Gaussian process: a two-step approach

The heuristic approach actually works reasonably well in practice (Miller et al., 2016). However, it apparently oversimplifies the original model. Here we try to reverse those simplifications and solve the model in a more principled manner. We first try to break it down to a two-step approach. In the first step, a generalized linear model is used to detect binary events in \( y \). In the second step, we model the signal around the event onset time as a Gaussian process, and train a classifier to decode the associated conditions.

One critical simplification in the heuristic approach is that we only consider the conditional mean \( M \) as the template and convert \( x \) to \( z \) with the correlation distance to the mean. To lift this assumption, in the first step, we directly model \( P(y_i|x) \) with a discriminative model, where \( y_i \in \{0, 1\} \) is a binary variable indicating whether there is a stimulus being presented. As an intermediate approach, a generalized linear model can be used here \( P(y_i|x) = \log(\text{vec}(\tilde{x}_i)^T \beta) \), where \( \tilde{x}_i = [x_{i-l_0}, \ldots, x_{i-l_1}] \in \mathbb{R}^{p \times (l_0+1)} \) is the time-windowed observation with respect to \( y_i \). The corresponding graphical model is shown in Figure 1(c). To enforce temporal and spatial structure, a group lasso penalty is applied on \( \beta \), where the time points for each dimension in \( x \) is grouped together, and therefore we get sparsity in space. This can be written as

\[
\min_{\beta \in \mathbb{R}^{p+1}} -\ell(y, \tilde{x}, \beta) + \lambda \sum_{j=1}^{J} ||\beta_{(j)}||_2 \tag{1}
\]

where \( \beta_{(j)} \) is the group of weights corresponding to the \( j \)-th dimension (row) in \( \tilde{x}_i \), and \( \ell \) is the log likelihood. Here we use a log-linear logistic model as the discriminant model, therefore

\[
\ell(\beta) = \sum_{i=1}^{T} y_i(\text{vec}(\tilde{x}_i)^T \beta) - \sum_{i=1}^{T} \log(1 + \exp(\text{vec}(\tilde{x}_i)^T \beta)) \tag{2}
\]

After detection of the events in \( y \), denoted as \( \{v_i\}_{i=1}^m, v_i \in \{C_1, \ldots, C_m\} \), and the corresponding time stamp series, denoted as \( \{t_{v_i}\} \). We can then identify the non-overlapping time intervals in \( x \) that are around each event (Figure 1(d)), denoted as \( \{w_i\}_{i=1}^m, w_i = [x_{t_{v_i}-l_0}, \ldots, x_{t_{v_i}+l_1}] = [w_{1i}, \ldots, w_{li}] \in \mathbb{R}^{p \times (l_0+1)} \). At this point, the problem becomes a classification problem for short clips of sequences with a fixed length. One straightforward way for classification is to train another GLM as in (1).

On the other hand, since \( w_i \) has smooth temporal structure in each dimension, we can use Gaussian process (GP) model, and apply Gaussian process classification to solve the decoding problem (Rasmussen & Williams, 2006). Specifically, this is a multivariate Gaussian process, with each \( w_i \) being a sample from the \( p \)-dimensional vector-valued function \( f \). And we have GP prior \( f \sim \mathcal{GP}(0, K(w, w')) \), where \( K(w, w') = \text{diag}\{K^{(1)}(w, w'), \ldots, K^{(p)}(w, w')\} \) is a block-diagonal that each block corresponding to the GP kernel on each observation dimension. In other words, the prior GP enforces smoothness along time and independence across space. With the GP prior, we can get posterior \( P(f|w, y, w_s) \) for any test inquiry \( w_s \). The inference can be written as

\[
P(f|w, y, w_s) = \int p(f|w, y, w_s) p(f|w) df \tag{3}
\]

where \( p(f|w, y) = p(y|f)p(f|w)/p(y|w) \) is the posterior over the latent variables. And the prediction given test inquiry \( w_s \) is the posterior mean over the link function

\[
p(y_s|w, y, w_s) = \mathbb{E}[P(f|w, y, w_s)](\sigma(f_s)) \tag{4}
\]

Because \( y \) is discrete, the exact inference is intractable. We use approximate inference such as Laplace approximation and expect propagation (Rasmussen & Williams, 2006) to get the posterior inference.

### 3.4. Conditional random field: exploiting the temporal structure

Another important simplification in the previous models is that we break down the linear chain structure in \( \{y_i\} \). To fully account for the temporal structure, the full model becomes a conditional random field (CRF). Given the fact that each stimulus would only trigger a transient response that can be considered lasting for only a few hundred milliseconds, the model can be simplified as Figure 2(a).

The CRF likelihood can be written as

\[
P(y|x) = \frac{1}{Z(x, \lambda, \mu)} \exp \left( \sum_{i=1}^{n} \left( \lambda^T f(y_i, y_{i-1}, x) + \mu^T g(y_i, x) \right) \right) \tag{5}
\]

where \( f \) is the feature function, which is a generalized indicator function of the state transitions, and \( g \) is the emission function. Specifically, we parameterized the CRF as a log-linear model. The model can be learned through maximum likelihood estimation, and the inference can be done using forward-backward algorithm.
In order to capture non-linear input-output dependencies, we can further extend the CRF model (Peng et al., 2009; Baltrušaitis et al., 2014). One way to incorporate non-linear dependencies is to add a hidden layer into the graphical model of CRF. In the model shown in Figure 2(b), a layer of sigmoid units $h$ are introduced to the model, as we define

$$h(\theta, x) = \frac{1}{1 + \exp(-\theta^T x)}$$  

This transformation effectively maps $x$ to $h$ and therefore turns the CRF between $x$ and $y$ into a CRF between $h(x)$ and $y$. As a result, we parameterize the graphical model as

$$f_k(y_i, x, \theta) = -(y_i - h(\theta_k, x_i))^2$$  

$$g_k(y_i, y_j) = -\frac{1}{2} S_{i,j}^{(g_k)}(y_i - y_j)^2$$  

Plug (7) and (8) into (5), then the potential function is quadratic in $y$. Therefore, the likelihood function is transformed into a multivariate Gaussian form.

$$P(y|x) = \frac{1}{(2\pi)^{T/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (y - m)^T \Sigma^{-1} (y - m)\right)$$  

the mean can be represented as

$$m = 2\Sigma \lambda^T (1 + \exp(-\Theta^T X))$$  

and the covariance can be written as

$$\Sigma^{-1} = 2(A + B)$$

$$A_{i,j} = \left\{ \begin{array}{ll} \sum_k \lambda_k & \text{if } i = j \\ 0 & \text{if } i \neq j \end{array} \right.$$  

$$B_{i,j} = \left\{ \begin{array}{ll} (\sum_k \mu_k \sum_r S_{i,r}^{(g_k)}) - (\sum_k \mu_k S_{i,j}^{(g_k)}) & \text{if } i = j \\ -\sum_k \mu_k S_{i,j} & \text{if } i \neq j \end{array} \right.$$  

The diagonal matrix $A$ represents the contribution of the $\lambda$ terms (vertex features) of the potential function to the covariance matrix, and matrix $B$ represent the contribution of the $\mu$ terms of the potential function (edge features) to the covariance matrix.

The model parameters can be learned through maximum likelihood estimation using gradient descent or second order method, such as L-BFGS (Baltrušaitis et al., 2014). The formulation in Gaussian form makes the inference straightforward. With observed samples $x'$, the conditions $y'$ can be estimated as the posterior mean in (10). Specifically, the gradient can be derived as

$$\frac{\partial \log(P(y|x))}{\partial \lambda_k} = -y^T y + 2y^T H_{k,*} - 2H_{*,k} m + m^T m + \text{tr}(\Sigma)$$  

$$\frac{\partial \log(P(y|x))}{\partial \mu_k} = -y^T \frac{\partial B}{\partial \mu_k} y + m^T \frac{\partial B}{\partial \mu_k} m + \text{tr}(\Sigma \frac{\partial B}{\partial \mu_k})$$  

$$\frac{\partial \log(P(y|x))}{\partial \theta_{i,j}} = (y - m)^T \frac{\partial b}{\partial \theta_{i,j}}$$

where $H = 1/(1 + \exp(-\Theta^T X))$ is a matrix representing the hidden layer.

4. Experiments

In this section, we evaluate and compare the proposed models using real electrophysiology dataset collected from human participant. The experimental protocols were approved by the Institutional Review Board of the University of Pittsburgh and Carnegie Mellon University. Written informed consent was obtained from the participant.

4.1. Data

intracranial EEG: The data was collected from electrodes implanted in human patients when they were receiving medical treatment for epilepsy. During the experiment, pictures of different categories were shown to the patients, and the corresponding brain activity was simultaneously recorded from 96 distributed electrodes in temporal lobe, as shown in Figure 3. A total of 480 pictures from 6 different categories, including human bodies, faces, words, tools, houses, and scrambled images, were presented to each subject in a random order (see Figure 3(b) for example stimuli). Each picture was presented for 900ms with 900ms (plus a random jittering in [0,400ms]) of inter-stimulus interval between two consecutive pictures. The signal recorded by each electrode is the local field potential (LFP) of the neurons near the electrode contact with
a sampling rate of 1000 Hz. Stimuli from different visual categories would evoke neural activity with different temporal profiles in different brain regions in temporal lobe. As a result, we recorded the spatiotemporal signal with tens of spatial dimensions and 1044.61 sec long (1044610 time points).

Figure 3. The iEEG experiment: (a) illustration of the locations of electrodes implanted in the brain; (b) example images from the 6 categories used in the experiment

Preprocessing: The data was filtered to remove line noise and DC component. Wavelet transformation was applied to estimate the power spectrum density (PSD) at each time point with temporal resolution of 100Hz. The mean normalized PSD for 40-100 Hz broadband activity is then used for the analysis (Miller et al., 2016). Therefore, we have the data matrix $X$ with 96 rows (96 electrode channels) and 104461 columns (1044.61 sec in time). And 480 events corresponding to the 480 stimuli were also recorded as $y$.

4.2. Results

We implemented and applied different methods proposed in the previous section on the dataset. As mentioned before, we specifically interested in two questions:

(1) detect the onset of the events;
(2) decode the contents of the events.

For (1), the metrics that we use to evaluate the performance are the precision and recall rate for the event detection. For (2), currently we are interested in a binary case, in which we want to separate human faces versus everything else. Therefore, we would like to see the precision and recall for face category.

Since each session of the experiment is broken down into 3 blocks, we use a three-fold cross-validation approach to evaluate the performance of the methods. Each time two-thirds of the data are used as training set, and the rest one-third are used as testing set.

For each event, a 200 ms time window around the exact onset time is used as tolerant window, any detection within the [-100 ms, 100 ms] interval around the onset time is treated as true positive.

The results are summarized in Table 1.

Table 1. Performance summary of the heurisitic baseline and the proposed methods

<table>
<thead>
<tr>
<th>metric</th>
<th>heuristic</th>
<th>two-step</th>
<th>CRF</th>
<th>CRF-extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>event detection precision</td>
<td>0.4823</td>
<td>0.8518</td>
<td>0.9501</td>
<td>0.9836</td>
</tr>
<tr>
<td>event detection recall</td>
<td>0.8500</td>
<td>0.9146</td>
<td>0.9975</td>
<td>0.9979</td>
</tr>
<tr>
<td>face classify precision</td>
<td>0.3200</td>
<td>0.8315</td>
<td>0.7024</td>
<td>0.8732</td>
</tr>
<tr>
<td>face classify recall</td>
<td>0.9523</td>
<td>0.9867</td>
<td>0.7024</td>
<td>0.7381</td>
</tr>
</tbody>
</table>

5. Discussion

5.1. Comparison of the different methods

From Table 1, the two-step approach outperforms the heuristic approach proposed by Miller et al. (2016) in both event detection and categorical classification (faces vs. rest). Since the two-step approach is fundamentally a formalized version of the heuristic approach with probabilistic models that account for the variance in a principled manner, this improvement in the performance should be well-expected.

On the other hand, the CRF and CRF-extension models dominate the heuristic and the two-step approaches in the event detection tasks. The fundamental difference between CRF-based approaches and the two-step or heuristic approaches is that the CRF models are designed to capture temporal dependencies in the sequence through horizontal links in $y$. From the results, we can see that the CRF models are efficient in modeling the temporal dependencies, therefore succeed in detecting events from the baseline of inter-stimulus intervals.

The additional neural layer helps to further improve the performance of CRF models in both event detection and face category classification. The introduction of the hidden layer adds nonlinearity into the log-linear model and enriches the representation structure of the emission part of the potential function. As a result, the extended model dominates the original CRF model.

Although CRF-based models have the advantage of exploiting the correlations between the different labels, and, as a result, significantly improve the accuracy over approaches that classify instances independently when detecting the events. However, CRF-based models do not optimize for the discrimination between different conditions. As a result, in the face classification task, which aims at
separating different types of evoked responses from each other, the CRF-based approaches do not outperform the two-step approach, which has additional supervised training on the discriminant task. One way to make up for this shortcoming is to adopt max-margin models, such as the max-margin Markov networks (Taskar et al., 2004).

Another possible drawback of the extended CRF model is that it is not convex with the hidden neural layer. However, this may not be a fatal issue since the local minimum obtained by gradient-based methods performs relatively well in the listed metrics, compared to the global optimum of the other models.

5.2. Inference and neuroscientific insights

As we mentioned, there are two aspects of inference that we are interested in. The CRF-based approaches perform well in inferring the hidden states. However, because of the complexity of the model, it is not straightforward to make inference on the exact structure of the spatio-temporal signal that gives rise to the discriminant information.

On the other hand, from the GLM in the two-step approach, we can actually make inference on the contribution of different electrodes to the event detection task. In other words, the groups of variables with non-zero weights selected by the group lasso are representing the electrodes that actually contribute to the encoding of the visual stimuli. Formally speaking, this requires post-selection inference techniques (Lee et al., 2016; Taylor & Tibshirani, 2017), which is out of the scope of this project. However, by looking at the non-zeros in the group lasso, we can still shed light upon the underlying neural process that gives rise to the discriminant information.

On the other hand, from the GLM in the two-step approach, we can actually make inference on the contribution of different electrodes to the event detection task. In other words, the groups of variables with non-zero weights selected by the group lasso are representing the electrodes that actually contribute to the encoding of the visual stimuli. Formally speaking, this requires post-selection inference techniques (Lee et al., 2016; Taylor & Tibshirani, 2017), which is out of the scope of this project. However, by looking at the non-zeros in the group lasso, we can still shed light upon the underlying neural process that gives rise to the discriminant information.

5.2. Inference and neuroscientific insights

As we mentioned, there are two aspects of inference that we are interested in. The CRF-based approaches perform well in inferring the hidden states. However, because of the complexity of the model, it is not straightforward to make inference on the exact structure of the spatio-temporal signal that gives rise to the discriminant information.

On the other hand, from the GLM in the two-step approach, we can actually make inference on the contribution of different electrodes to the event detection task. In other words, the groups of variables with non-zero weights selected by the group lasso are representing the electrodes that actually contribute to the encoding of the visual stimuli. Formally speaking, this requires post-selection inference techniques (Lee et al., 2016; Taylor & Tibshirani, 2017), which is out of the scope of this project. However, by looking at the non-zeros in the group lasso, we can still shed light upon the underlying neural process that gives rise to the discriminant information.

On the other hand, from the GLM in the two-step approach, we can actually make inference on the contribution of different electrodes to the event detection task. In other words, the groups of variables with non-zero weights selected by the group lasso are representing the electrodes that actually contribute to the encoding of the visual stimuli. Formally speaking, this requires post-selection inference techniques (Lee et al., 2016; Taylor & Tibshirani, 2017), which is out of the scope of this project. However, by looking at the non-zeros in the group lasso, we can still shed light upon the underlying neural process that gives rise to the discriminant information.

6. Conclusion

In this project, we investigate the modeling and analysis of spatiotemporal neural signal. We build graphical models that capture the structure and distribution of the time-evolving dynamics in the signal. We compare our proposed methods to the heuristic method in the literature and show that our methods outperform the heuristic method in event detection and classification.

Based on the heuristic approach, we propose a two-step model based on GLM/GP. The supervised model captures the spatiotemporal structure of the data by group lasso penalty and Gaussian process priors. To further encorporate temporal dependencies in the signal, we propose two CRF-based models: the vanilla CRF and the conditional neural field model. In general, the CRF-based models better capture the temporal dynamics in the stimulus-evoked response, and therefore have better performance in the event detection task. The two-step approach have a specific emphasis on the supervised training of discriminant information between categories and therefore performs better in the category classification task. In addition, GLM with group lasso penalty gives insights on how the representation structure of the spatiotemporal signal.

References


Ahrens, Misha B, Orger, Michael B, Robson, Drew N, Li, Jennifer M, and Keller, Philipp J. Whole-brain functional imaging at cellular resolution using light-sheet mi-
Spatiotemporal Dynamics of Neural Signals: Modeling and Inference


Rasmussen, Carl Edward and Williams, Christopher KI. Gaussian processes for machine learning. the MIT Press, 2006.


