Class-Aware Hidden Markov Model for simultaneous functional connectivity estimation and classification

Chendi Han[1], Pavithran Pattiam Giriprakash[1], Rajesh Nandy[2], and Dietmar Cordes[1]
[1]Cleveland Clinic, Las Vegas, NV, United States, [2]Department of Biostatistics and Epidemiology, School of Public Health, University of North Texas Health Science Center, Fort Worth, TX, United States

Synopsis

Motivation: Hidden Markov Models (HMM) are widely used for modeling resting-state fMRI data. However, current classification approaches usually involve additional parameter extraction steps, which make the model complicated and reduce sensitivity.

Goal(s): Our goal is to develop a one-step classification method based on HMMs that is more efficient, accurate, and maintains interpretability.

Approach: We utilize Hidden Conditional Random Fields (HCRF), combining HMM and discriminative learning into a unified model that eliminates the need for separate parameter extraction.

Results: In both simulated and real data, the one-step model outperforms the traditional two-step model. Supervised learning could further improve classification accuracy.

Impact: For fMRI classification problems, the one-step Class-Aware HMM is simpler and more accurate compared to two-step classification while maintaining model interpretability. This could help in understanding brain connectivity and disease diagnosis.

Introduction

Several approaches have been adopted to characterize the temporal dynamics of resting-state fMRI connectivity. Compared to sliding window methods[1-2], Hidden Markov Models (HMM) are suitable for rapid state transitions, with the covariance matrix directly measuring brain connectivity[3-4]. In this paper, we propose a framework to classify fMRI data based on HMM. We compare the two-step approach, feature extraction and subsequent classification based on selected features[5-8], and the one-step approach, using the joint probability to compute the posterior probability[9-10]. We evaluate the classification accuracy using both simulated and real fMRI datasets.

Methods

Consider fMRI data labeled as $X \in \mathbb{R}^{N\times p\times T}$, indicating N subjects with length T and p components. Each subject has a ground truth class label among C classes. We assume fMRI data in each class follow a time-independent Hidden Markov Model (HMM). Specifically, the model has $K^{(c)}$ hidden states, with transition probability matrix $A^{(c)} \in \mathbb{R}^{K^{(c)}\times K^{(c)}}$, and the emission probability follows a Gaussian distribution with mean $\mu^{(c)} \in \mathbb{R}^{K^{(c)}\times p}$ and covariance $\Sigma^{(c)} \in \mathbb{R}^{K^{(c)}\times p\times p}$. The covariance matrix is regarded as the dynamic functional connectivity[3-4]. During parameter estimation, we assume that $K^{(c)} = 3$ for all classes, while other parameters $\theta = \{A, \mu, \Sigma\}$ are treated as optimization parameters. For each class, the Expectation-Maximization (EM) algorithm can be used to maximize the joint probability Ψ, as shown in Figure 1(a). One potential problem is that performing the EM algorithm separately for each class may not generate good class separation. We implement additional gradient descent to minimize the Cross Entropy Loss (CEL) between the posterior probability and the ground truth. This method is also called Hidden Conditional Random Field (HCRF)[9-10], as shown in Figure 1(b). Suppose we have a new time series X_{test} with unknown class labeling. Figure 1(c) shows a two-step approach for the classification problem. For each subject in the testing set, we evaluate the model parameters θ_{test} first and then compute the posterior probability based on the evaluated parameters. We consider an unsupervised classifier, assigning class index based on the nearest neighborhood (Method 1), and a supervised Multi-Layer Perceptron (MLP) classifier (Method 2). Figure 1(d) shows a one-step approach: an HMM-based generative model (Method 3) and an HCRF-based discriminative model (Method 4).

Results

We begin with simulated data to test performance with $C = 3, p = 30$, and $T = 100$. All three classes share the same means and covariances but have different transition probabilities. An example is shown in Figures 2. Based on the ground truth, observations X are generated. The number of time series for each class in the training, evaluation, and testing datasets are 100, 100, and 1,000, respectively. The class-level EM algorithm evaluates θ_{EM} from random starting points. Two-step models estimate subject-level parameters starting from θ_{EM}, then perform either unsupervised or supervised learning to evaluate posterior probabilities. In contrast, one-step models evaluate the posterior probability directly using θ_{EM} or perform additional gradient descent. Figure 2(d) shows the average classification accuracy over 100 realizations. We find that across different noise levels, the one-step model outperforms two-step methods. At higher noise levels, additional supervised learning yields better performance compared to the unsupervised method. Next, we apply the same method to real fMRI data containing 292 subjects with 74 Cognitively Normal (CN), 119 with Mild Cognitive Impairment (MCI), and 99 with Alzheimer's Disease (AD). Resting-state fMRI data were acquired with 3T, TR=3000 ms, TE=30 ms, flip angle=90°, FoV=220 mm, slice thickness=3.4 mm, EPI factor=64, echo spacing=0.72 ms. After group independent component analysis, the time series per subject has size $T = 135$ and $p = 54$. The data is split into training, evaluation, and testing sets in a 7/2/1 ratio. Figure 3 shows the transition probabilities, means, and covariance matrices from class-level EM results, and Figure 4 shows the state occupational probability and posterior probability for one selected subject versus time. We find that data augmentation in the time domain and Lasso regularization could help to reduce CEL in the validation dataset[4]. The whole process is repeated 100 times. Figure 5 shows the performance of different methods. Our proposed HCRF shows the highest performance, with classification accuracy close to 50%.

Discussion

The key findings of this study are that one-step posterior probability can be used for disease classification, and that supervised HCRF shows better performance compared with unsupervised HMM.

Acknowledgements

This study was funded by NIH-R01AG071566 and NIH-P20GM109025.

References

[1] Allen, E., et al., Tracking whole-brain connectivity dynamics in the resting state. Cerebral cortex, 2014. 24(3), pp.663-676.

[2] Cai, B., et al., Estimation of dynamic sparse connectivity patterns from resting state fMRI, 2017. IEEE transactions on medical imaging, 37(5), p.1224-1234.

[3] Vidaurre, D., et al., Brain network dynamics are hierarchically organized in time. Proceedings of the National Academy of Sciences, 2017. vol. 114, no. 48: p.12827-12832.

[4] Zhang, G., et al., Estimating dynamic functional brain connectivity with a sparse hidden Markov model. IEEE transactions on medical imaging, 2019. 39(2): p. 488-498.

[5] Ji, J., et al., Convolutional neural network with sparse strategies to classify dynamic functional connectivity. IEEE Journal of Biomedical and Health Informatics, 2021. 26(3), p.1219-1228.

[6] Zhao, C., et al., Abnormal characterization of dynamic functional connectivity in Alzheimer's disease. Neural regeneration research, 2022. 17(9), pp.2014-2021.

[7] Jin, H., et al., Dynamic functional connectivity MEG features of Alzheimer’s disease. NeuroImage, 2023. 281, p.120358.

[8] Canal-Garcia, A., et al., Dynamic multilayer functional connectivity detects preclinical and clinical Alzheimer’s disease. Cerebral Cortex, 2024. 34(2), p.542.

[9] Quattoni, A., et al., Hidden-state conditional random fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007. 29(10), p.1848-1852.

[10] Sutton, C. and McCallum, A., An introduction to conditional random fields. Foundations and Trends® in Machine Learning, 2012 4(4), p.267-373.

Figures

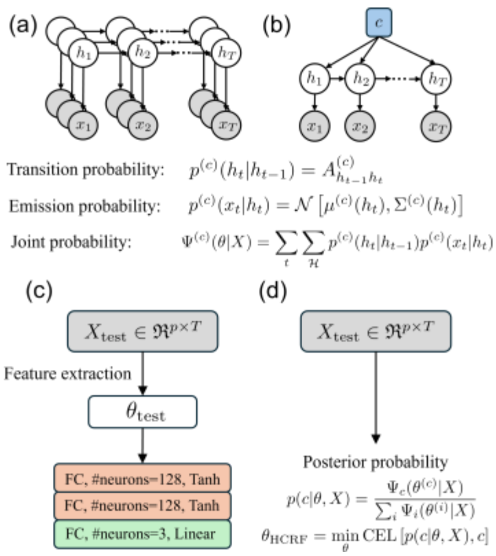

Figure 1: (a) Illustration of Hidden Markov Models (HMMs) corresponding to different classes. Each HMM is performed separately, given an unsupervised approach. (b) Illustration of a Hidden Conditional Random Field (HCRF), a supervised approach where the class information is integrated into the loss function. (c) Illustration of the two-step classification with a multi-layer perceptron (MLP) as the additional classifier. (d) Illustration of the one-step model and the posterior probabilities.

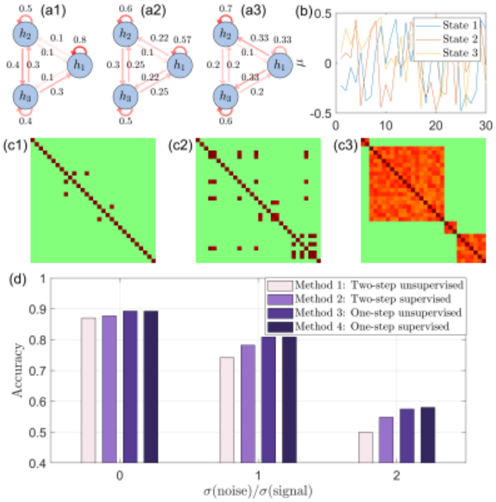

Figure 2: For simulated dataset. The different classes exhibit distinct transition probabilities while maintaining the same means and variances. (a1)-(a3) The transition probabilities for each class. (b) Mean values for each hidden state. (c1)-(c3) The covariance matrices corresponding to each hidden state. (d) Classification accuracy versus noise level for simulated data. Noise level is defined as the ratio of noise variance to signal variance.

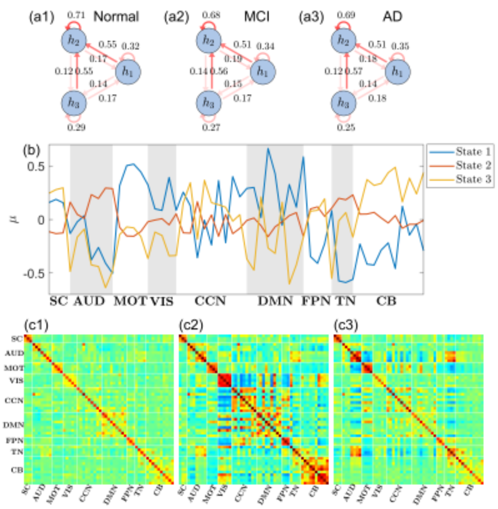

Figure 3: Dynamic functional connectivity analysis for resting-state fMRI data using class-level EM. Different classes share the same mean and covariance. (a1)-(a3) The transition probabilities for CN/MCI/AD. (b) Mean values for each hidden state. (c1)-(c3) Covariance matrices for each hidden state, organized into nine functional groups: subcortical (SC), auditory (AUD), motor (MOT), visual (VIS), cognitive control (CCN), default mode (DMN), frontoparietal (FPN), temporal (TN), and cerebellar (CB).

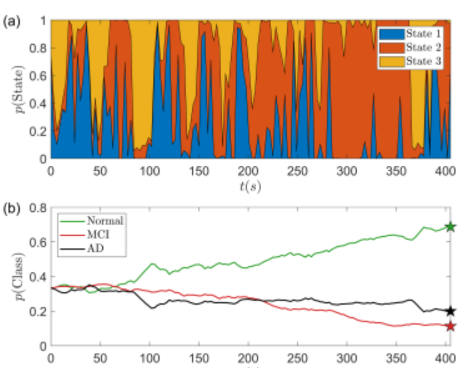

Figure 4: Time-related properties for a normal subject in the testing set; all results are evaluated using class-level EM, the unsupervised one-step method. (a) State occupational probability versus time. (b) Posterior probability versus time. The decision is made at $t = T$, indicated by star samples at the right.

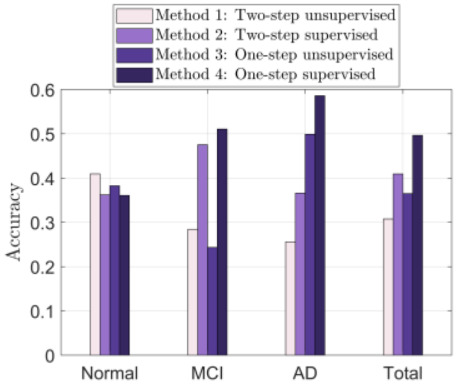

Figure 5: For real resting fMRI data, classification accuracy comparison across different methods and classes. On average, supervised training improves classification accuracy. The mean accuracies for the four methods are 30.77%, 40.93%, 36.50%, and 49.67%, respectively.