

Functional MRI Analysis with Sparse Models

Irina Rish

IBM T.J. Watson Research Center
Computational Biology Center - Neuroscience
1101 Kitchawan Rd. Yorktown Heights, NY 10598
`rish@us.ibm.com`

Abstract. Sparse models embed variable selection into model learning (e.g., by using l_1 -norm regularizer). In small-sample high-dimensional problems, this leads to improved generalization accuracy combined with interpretability, which is important in scientific applications such as biology. In this paper, we summarize our recent work on sparse models, including both sparse regression and sparse Gaussian Markov Random Fields (GMRF), in neuroimaging applications, such as functional MRI data analysis, where the central objective is to gain a better insight into brain functioning, besides just learning predictive models of mental states from imaging data.

Keywords: neuroimaging, fMRI, l_1 -norm regularization, Lasso, Elastic Net, sparse Gaussian Markov Random Fields (GMRF).

1 Introduction

Predicting person's mental state based on his or her brain imaging data, such as functional MRI (fMRI), is an exciting and rapidly growing research area on the intersection of neuroscience and machine learning. A mental state can be cognitive, such as viewing a picture or reading a sentence [8], emotional, such as feeling happy, anxious, or annoyed while playing a virtual-reality videogame [1], reflect person's perception of pain [11,12,3], or person's mental disorder, such as schizophrenia [2,10], drug addiction [6], and so on.

In fMRI, an MR scanner non-invasively records a subject's blood-oxygenation-level dependent (BOLD) signal, known to be related to neural activity, as a subject performs a task of interest or is exposed to a particular stimulus. Such scans produce a sequence of 3D images, where each image typically has on the order of 10,000-100,000 subvolumes, or *voxels*, and the sequence typically contains a few hundreds of time points, or TRs (time repetitions). Thus, each voxel is associated with a time series representing the average BOLD signal within that subvolume (i.e., voxel activity) at each TR; a task or a stimulus is associated with the corresponding time series over the same set of TRs.

2 Sparse Regression

Our work is motivated by the traditional fMRI goal of discovering task-relevant brain areas. However, we wish to avoid limitations of traditional mass-univariate

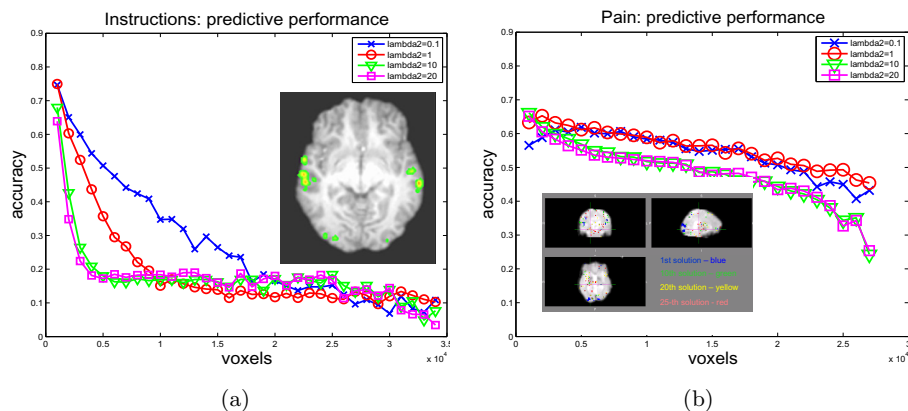


Fig. 1. Predictive accuracy of the Elastic Net for the task of predicting (a) “Instructions” task in PBAIC and (b) thermal pain perception. The insets visualize the sparse solutions found by the Elastic Net.

approaches such as GLM [4] that essentially performs *filter-based* variable selection based on individual voxel correlations with the task, and thus can miss important multivariate interactions, as noted by [5] and others. Thus, we focus instead on **sparse** multivariate models capable of identifying a relatively small subset of variables (voxels) that (*jointly*) predict the task well. In [1], we were among the first ones to apply sparse methods to fMRI, presenting our analysis of the PBAIC 2007 competition data [9] that we obtained using the Elastic Net (EN) approach [15]. EN improves upon the basic LASSO [14] by using a convex combination of l_1 - and l_2 -norm regularizers instead of just l_1 . The effect of such combined penalty is that, on top of sparsity (voxel selection), a *grouping effect* is encouraged, i.e. joint inclusion (or exclusion) of groups of highly correlated variables (such as spatial clusters of voxels). The grouping property is particularly important from the interpretability perspective, since we hope to discover relevant brain areas rather than their single-voxel representatives sufficient for accurate prediction, as the basic LASSO does. We investigate the effects of both l_1 and l_2 regularization parameters on the predictive accuracy and stability, measured here as a support overlap between the regression coefficients learned for different runs of the experiment. We conclude that, (a) EN can be highly predictive about various mental states, achieving 0.7–0.9 correlation between the true and predicted response variables (see Figure 1a), and (b) even among equally predictive models, increasing the l_2 -regularization weight can help to improve the model stability.

Furthermore, our subsequent work presented in [11], demonstrates that the Elastic Net can be highly predictive about such subjective and seemingly hard-to-quantify experience as pain, achieving up to 0.75–0.8 correlation between

the predicted and actual pain perception ratings, drastically outperforming unregularized linear regression, and identifying novel areas undiscovered by GLM¹.

However, given a brain map of task-relevant voxels, does this imply that the rest of the brain voxels is irrelevant? Not necessarily, since multiple sparse predictive solutions are possible in presence of highly-correlated predictors. In [12], we explore the space of sparse solutions, by first finding the best EN solution with 1000 voxels, removing those voxels from the set of predictors, and repeating the procedure until there are no more voxels left. Interestingly, for multiple tasks we considered, including pain perception and others, *no clear separation between relevant and irrelevant areas was observed*, as shown in Figure 1b, suggesting highly non-localized, “holographic” distribution of task information in the brain. The only task which demonstrated fast (exponential) performance degradation, and clear separation of relevant vs irrelevant areas, was a relatively simple auditory task from PBAIC (Figure 1a)². Note that standard GLM method does not reveal such phenomenon, since, as shown in [12], individual voxel-task correlations always seem to decay exponentially, and for many reasonably predictive (but not best) sparse solutions, their voxel would not even pass 0.1 correlation threshold. Thus, multivariate sparse regression is a much better tool than GLM for exploring actual distribution of task-relevant information in the brain.

3 Sparse Gaussian Markov Random Fields (GMRFs)

Though task-relevant brain areas are still the most common type of patterns considered in fMRI analysis, they have obvious limitations, since the brain is an interconnected, dynamical system, whose behavior may be better captured by modeling interactions across different area. For example, in our recent study of schizophrenia [2,10], task-based voxel activations are dramatically outperformed by network-based features, extracted from the voxel-level correlation matrices (“functional networks”), and yielding from 86% to 93% classification accuracy on schizophrenic vs. control subjects. Furthermore, we investigate structural differences of sparse Gaussian Markov Random Fields, or GMRFs, constructed from fMRI data via l_1 -regularized maximum likelihood (inverse covariance estimation). Used as probabilistic classifiers, GMRFs often outperform state-of-art classifiers such as SVM (e.g., see Figure 2a from [2]). In [13], we proposed a simple, easily parallelizable greedy algorithm SINCO, for Sparse INverse CO-variance estimation.

Next, we developed a *variable-selection structure learning* approach for GMRFs in [7]. A combination of (ℓ_1, ℓ_p) group-Lasso penalty with the l_1 -penalty selects the most-important variables/nodes, besides simply sparsifying the set of

¹ The predictive accuracy can be further improved by combining EN for predicting the intensity of the painful stimulus (e.g., the temperature) from fMRI data with a novel analytic, differential-equation model that links temperature and pain perception [3].

² A possible hypothesis is that, while “simple” tasks are localized, more complex tasks/experiences (such as pain) tend to involve much more distributed brain areas (most of the brain, potentially).

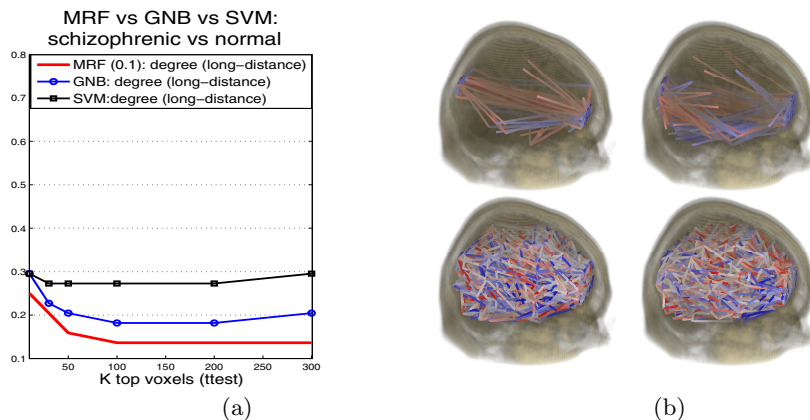


Fig. 2. (a) Gaussian MRF classifier predicts schizophrenia with 86% accuracy using just 100 top-ranked (most-discriminative) features, such as voxel degrees in a functional network. (b) Structures learnt for cocaine addicted (left) and control subjects (right), for sparse Markov net learning method with variable-selection via $\ell_{1,2}$ method (top) and without variable-selection, i.e., standard graphical lasso approach (bottom). Positive interactions are in blue, negative – in red. Structure density on top is 0.0016, while on the bottom it is 0.023 (number of edges in a complete graph is ≈ 378000).

edges. Our main advantage is a better interpretability of the resulting networks due to elimination of noisy variables (see below), combined with improvements in model likelihood and more accurate recovery of ground-truth structure. From an algorithmic point of view, we show that a block coordinate descent method generates a sequence of positive definite solutions. Thus, we reduce the original problem into a sequence of strictly convex (ℓ_1, ℓ_p) regularized quadratic minimization subproblems for $p \in \{2, \infty\}$. Our algorithm is well founded since the optimal solution of the maximization problem is unique and bounded. Figure 2b shows the network structures learnt for cocaine addicted vs healthy control subjects, comparing the two methods. The disconnected variables are not shown. The variable-selection sparse Markov network approach yields much fewer connected variables but a higher log-likelihood than graphical lasso, as reported in [7], which suggests that the discarded edges from the disconnected nodes are not important for accurate modeling of this dataset. Moreover, removal of a large number of nuisance variables (voxels) results into a much more interpretable model, clearly demonstrating brain areas involved in structural model differences that discriminate cocaine addicts from healthy control subjects. Cocaine addicts show increased interactions between the visual cortex (back of the brain, on the left here) and the prefrontal cortex (front of the brain image, on the right), while at the same time decreased density of interactions between the visual cortex with other brain areas. Given that the trigger for reward in this experiments was a visual stimulus, and that the prefrontal cortex is involved in higher-order cognitive functions such as decision making and reward processing,

the alteration in this pathway in the addict group is highly significant from a neuroscientific perspective.

References

1. Carroll, M.K., Cecchi, G.A., Rish, I., Garg, R., Rao, A.R.: Prediction and interpretation of distributed neural activity with sparse models. *Neuroimage* 44(1), 112–122 (2009)
2. Cecchi, G., Rish, I., Thyreau, B., Thirion, B., Plaze, M., Paillere-Martinot, M.L., Martelli, C., Martinot, J.L., Poline, J.B.: Discriminative network models of schizophrenia. In: *Proc. of NIPS 2009* (2009)
3. Cecchi, G.A., Huang, L., Hashmi, J.A., Baliki, M., Centeno, M.V., Rish, I., Apkarian, A.V.: Predictive dynamics of human pain perception. *PLoS Computational Biology* 8(10) (2012)
4. Friston, K.J., et al.: Statistical parametric maps in functional imaging - a general linear approach. *Human Brain Mapping* 2, 189–210 (1995)
5. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P.: Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science* 293(5539), 2425–2430 (2001)
6. Honorio, J., Ortiz, L., Samaras, D., Paragios, N., Goldstein, R.: Sparse and locally constant gaussian graphical models. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C.K.I., Culotta, A. (eds.) *Advances in Neural Information Processing Systems* 22, pp. 745–753 (2009)
7. Honorio, J., Samaras, D., Rish, I., Cecchi, G.A.: Variable selection for gaussian graphical models. In: *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2012* (2012)
8. Mitchell, T.M., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M., Newman, S.: Learning to decode cognitive states from brain images. *Machine Learning* 57, 145–175 (2004)
9. Pittsburgh EBC Group. PBAIC Homepage (2007), <http://www.ebc.pitt.edu/2007/competition.html>
10. Rish, I., Cecchi, G., Thyreau, B., Thirion, B., Plaze, M., Paillere-Martinot, M.L., Martelli, C., Martinot, J.L., Poline, J.B.: Schizophrenia as a network disease: Disruption of emergent brain function in patients with auditory hallucinations. *PLoS ONE* 8(1) (2013)
11. Rish, I., Cecchi, G.A., Baliki, M.N., Apkarian, A.V.: Sparse regression models of pain perception. In: *Proc. of Brain Informatics (BI 2010)* (August 2010)
12. Rish, I., Cecchi, G.A., Heuton, K., Baliki, M.N., Apkarian, A.V.: Sparse regression analysis of task-relevant information distribution in the brain. In: *Proc. of SPIE Medical Imaging 2012* (February 2012)
13. Scheinberg, K., Rish, I.: Learning sparse gaussian markov networks using a greedy coordinate ascent approach. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) *ECML PKDD 2010, Part III. LNCS*, vol. 6323, pp. 196–212. Springer, Heidelberg (2010)
14. Tibshirani, R.: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58(1), 267–288 (1996)
15. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B* 67(2), 301–320 (2005)