fMRI Analysis with Sparse Weisfeiler-Lehman Graph Statistics

Katerina Gkirtzou, Jean Honorio, Dimitris Samaras, Rita Goldstein, Matthew Blaschko

fMRI Analysis with Sparse Weisfeiler-Lehman Graph Statistics

Katerina Gkirtzou[1,2], Jean Honorio[3], Dimitris Samaras[4], Rita Goldstein[5], and Matthew B. Blaschko[1,2] [⋆]

[1] Center for Visual Computing, École Centrale Paris, France
[2] Équipe Galen, INRIA Saclay, Île-de-France, France
[3] CSAIL, Massachusetts Institute of Technology, Cambridge, USA
[4] Computer Science Department, Stony Brook University, NY, USA
[5] Icahn School of Medicine, Mount Sinai, USA

Abstract. fMRI analysis has most often been approached with linear methods. However, this disregards information encoded in the relationships between voxels. We propose to exploit the inherent spatial structure of the brain to improve the prediction performance of fMRI analysis. We do so in an exploratory fashion by representing the fMRI data by graphs. We use the Weisfeiler-Lehman algorithm to efficiently compute subtree features of the graphs. These features encode non-linear interactions between voxels, which contain additional discriminative information that cannot be captured by a linear classifier. In order to make use of the efficiency of the Weisfeiler-Lehman algorithm, we introduce a novel pyramid quantization strategy to approximate continuously labeled graphs with a sequence of discretely labeled graphs. To control the capacity of the resulting prediction function, we utilize the elastic net sparsity regularizer. We validate our method on a cocaine addiction dataset showing a significant improvement over elastic net and kernel ridge regression baselines and a reduction in classification error of over 14%. Source code is also available at `https://gitorious.org/wlpyramid`.

Keywords: fMRI, graph kernels, Weisfeiler-Lehman, sparse regression, cocaine addiction

1 Introduction

Functional magnetic resonance imaging (fMRI) is a wide-spread, non-invasive modality used for studying brain activity, under the objectives of localizing brain regions participating in a specific task, determining connectivity networks or making predictions about disease states. Numerous approaches have been proposed for analyzing fMRI data. Most of them select features either by a predefined set of regions of interest (ROIs) using prior knowledge [1, 2], statistical methods [3, 4] or thresholding [5]. Recently, there has been an exploration of

---

sparsity regularization methods in fMRI analysis [6, 7]. fMRI analysis is particularly suited to sparsity regularization due to the intrinsic high dimensional nature of fMRI data, the small size of datasets and the lack of requirement for predefined ROIs. Although the aforementioned methods perform well in analyzing fMRI data, they tend to ignore the 3D structure of the brain. On the other hand, graph-theoretic methods can model structural information, and have also been used in fMRI analysis. The most common use is modeling the network of brain connectivity [8, 9]. Graph kernel methods have also been used in fMRI connectivity graphs for brain decoding [10].

The majority of previous approaches treat prediction in fMRI as a linear combination of functions over individual voxels. Such approaches do not capture potentially complex interactions between voxels. We enrich our capacity to model such dependencies by representing fMRI recordings as graphs, and we design an algorithm that learns from the interconnections between voxels. Through efficient graph algorithms, we are able to learn in a fully exploratory fashion without restricting our prediction, e.g., to a pre-defined region of interest or a connected component. We control the complexity of our prediction while modeling non-linear interactions between voxels by adding a sparsity regularizer using the elastic net [11]. We represent fMRI data as graphs over voxels, and compare the resulting graphs with a novel method that combines elements of the Weisfieler-Lehman graph kernel [12] and the pyramid match kernel [13]. This method achieves the computational advantages of efficient graph kernels while extending the representation to continuous node labels, which represent a contrast map at a given voxel. We validate the approach on a clinically important cocaine addiction dataset [14], achieving a reduction in classification error from an elastic net baseline of over 14% with statistical significance.

2 Methods

Graphs are rich representations of networks of data, and are consequently a promising representation for neural populations. A typical approach for exploring graph representation is to consider a measure of similarity between graphs as a kernel [15, 12]. The most popular strategy to define similarity over graph representations is to aggregate statistics of subgraphs and compare them. Here we agregate statistics of subtree patterns from a graph representation constructed from fMRI data. We then use these statistics with a sparsity regularizer to make predictions from graphs representing brain contrast maps. In order to make use of a graph representation for learning from fMRI data, several design choices must be made: (i) the learning algorithm, (ii) the graph construction, (iii) the graph statistics employed as a feature representation, and (iv) the node labeling. We address each of these choices in the following sections.

2.1 Sparsity Regularization

In the sequel, we will assume that we have access to a training set of data $\{(x_i, y_i)\}_{1 \leq i \leq n}$ of size n, where $x \in \mathcal{X}$ is a fMRI recording and $y \in \mathcal{Y} \equiv \{-1, +1\}$

Algorithm 1 The statistical learning pipeline for fMRI analysis with sparse subgraph statistics.

Require: Training set $\mathcal{D} = \{(x_i, y_i), i = 1, \ldots, n\}$.

    Compute $\hat{\beta}_{\mathrm{lin}}$ from the objective in Equation (1) with $\phi_{\mathrm{lin}}(x) \equiv \mathrm{vec}\, x$.

    Construct k-nearest neighbor graphs for all training samples from the voxels associated with non-zero $\hat{\beta}_{\mathrm{lin}}$.

    for each level in the quantization pyramid **do**

        Label the nodes of all graphs according to the quantization of the voxel value.

        Compute the Weisfeiler-Lehman statistics for the given quantization level over all graphs and aggregate them into the feature vector $\phi_{\mathrm{graph}}(x)$.

    end for

    Compute $\hat{\beta}_{\mathrm{graph}}$ from the objective in Equation (1) with $\phi_{\mathrm{graph}}(x)$.

---

is a label to be predicted (cf. Section 2.5). As our statistical estimator, we have made use of the elastic net [11]. The elastic net combines ℓ_1 with ℓ_2 regularization in order to appropriately trade off sparsity with a low variance estimator in the case of correlated signals. Formally, if $\phi(x) \in \mathbb{R}^d$ is a feature vector computed from x, the elastic net computes

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^d} \lambda_2 \|\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \frac{1}{n} \sum_{i=1}^{n} \left(\langle \beta, \phi(x_i) \rangle - y_i \right)^2 , \qquad (1)$$

where λ_1, $\lambda_2 \geq 0$ are scalar regularization parameters. This method is particularly appropriate in fMRI where nearby voxels are likely to be correlated, and regions responsible for a given function or behavior distributed across multiple voxels. Furthermore, it is typical that the majority of voxels in the brain are not discriminative of a specific output, y. We make use of the elastic net twice in our learning pipeline (Algorithm 1). In the first instance, we use the elastic net on the raw voxel values to determine a subset of voxels on which we build a graph representation, specifically those with non-zero $\hat{\beta}_{\mathrm{lin}}$. Our model selection step has typically chosen approximately 10^3 voxels for this stage. We subsequently compute subgraph statistics over this graph to generate a feature vector, $\phi_{\mathrm{graph}}(x)$. Finally, we use the elastic net on these subgraph statistics in order to determine our final prediction function, with a model selection step to determine appropriate values for λ_1 and λ_2.

2.2 Graph Construction

To construct the graph representation, we have made use of k-nearest neighbor graphs on the voxels that were selected by an initial training of the elastic net. We symmetrize the k-nn relationship by considering the edges to indicate an undirected graph structure. While other models of connectivity are of interest [16, 17], we have found that the use of k-nearest neighbors to determine the graph topology yields good performance in general. Furthermore, the subtree statistics considered in the next section implicitly account for longer distance connections for sufficiently deep subtree patterns. We set $k = 5$ in all experiments.
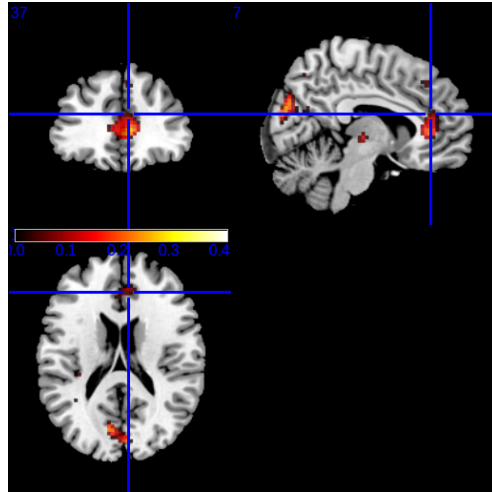
2.3 The Weisfeiler-Lehman Algorithm

The subgraph statistics considered here are subtree patterns from the Weisfeiler-Lehman algorithm [18], which has recently been employed in the construction of a graph kernel [12, Algorithm 2]. These statistics are linear-time in the number of edges in the graph, and make use of an efficient hashing scheme in order to only enumerate relevant (non-zero) dimensions of an exponentially sized feature space. In addition to these computational benefits, Weisfeiler-Lehman graph kernels have been shown to perform comparably to or better than a number of more computationally complex kernels [12]. Although Weisfeiler-Lehman statistics over subtree patterns have most recently been used in the construction of a kernel function after which ℓ_2 regularized learning is applied [12], we retain the vectors of subtree pattern counts and apply a sparsity regularizer in order to control model complexity in a high dimensional representation (Section 2.1).

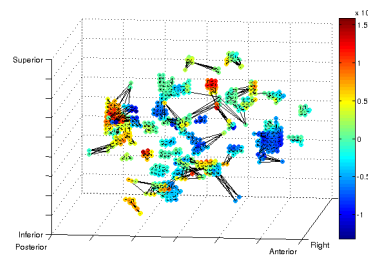### 2.4 Pyramid Quantized Weisfeiler-Lehman for Continuous Labels

The Weisfeiler-Lehman algorithm is efficient precisely because it makes use of a discrete labeling over nodes, which enables an efficient hashing scheme in order to scale linearly in the number of edges. This presents a problem when extending this method to continuous vector labeled graph vertices. Our vertices (fMRI voxels) are naturally labeled by their (continuous) contrast values. To retain the computational efficiencies of the graph representation, and to extend the method to continuous node labels, we consider a pyramid quantization strategy to determine a logarithmic number of discrete labelings for which we run the Weisfeiler-Lehman algorithm. This approximates a graph representation with continuous labels as a sequence of graphs with increasingly granular discrete labels. To construct our pyramid quantization, we recursively partition the space as proposed by [13]. In this manner, we have a series of nested vector quantizations of increasing granularity. We believe our method is the first to propose a pyramid quantization scheme for the efficient application of graph kernels to continuous (multi-dimensional) node labels. We expect that this will additionally be of high interest to the machine learning community, so code is available at `https://gitorious.org/wlpyramid`.
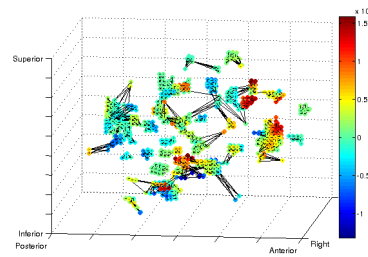
2.5 Cocaine Addiction Dataset

Our dataset [14] contains an approximately equal number of cocaine addicted individuals and control subjects performing a neuropsychological experiment of block design known as drug Stroop. Each subject performs six sessions with two varying conditions, a monetary reward (50¢, 25¢ and 0¢) and the cue shown (drug-related or a neutral word). We focus here on the 50¢ condition based on previous analysis of the data [5]. For the subjects that complied to motion $< 2\text{mm}$ translation, $< 2°$ rotation and at least 50% performance of the subject in an unrelated task [14], contrast maps were computed using the SPM package (`http://www.fil.ion.ucl.ac.uk/spm/`), which were used to determine the
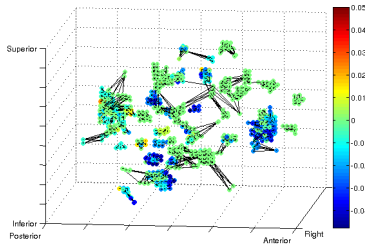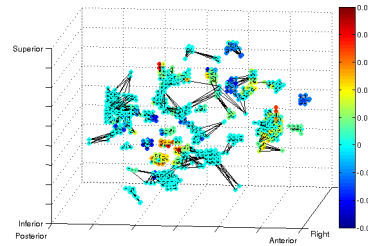
(a) Regions selected by Elastic Net.

(b) Elastic Net - Control

(c) Elastic Net - Cocaine

(d) Weisfeiler-Lehman - Control

(e) Weisfeiler-Lehman - Cocaine

Fig. 1: 1(a): A visualization of the areas of the brain selected by the elastic net. The selected regions correspond to areas previously implicated as being related to addiction [14]. 1(b)-1(c): A visualization of the function learned by Elastic net for control and cocaine subjects. 1(d)-1(e): A visualization of the function learned by the proposed system applied to control and cocaine addicted subjects.

node labels of the graphs. Finally, the task of interest is the classification of a subject as cocaine addicted or control.

3 Results

We evaluate our proposed pyramid quantized Weisfeiler-Lehman method for continuous node labels using the cocaine addiction dataset. We perform a random

Table 1: Mean accuracy over the hold-out data of 50 trials of the pyramid quantized Weisfeiler-Lehman algorithm for four different subtree pattern depths, $h \in \{0, 1, 2, 3\}$. Maximum performance is achieved with subtree patterns up to depth two.

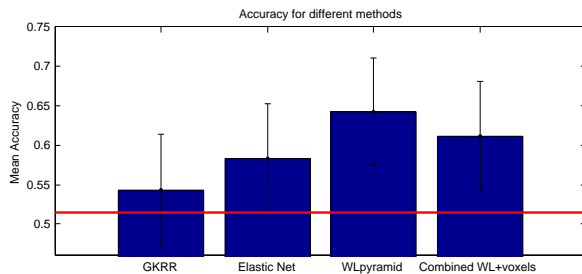| Pyramid Quantized Weifeiler-Lehman				
h	0	1	2	3
Accuracy	54.00%	57.14%	**64.28%**	63.42%

Fig. 2: Mean accuracy and standard error on the cocaine addiction dataset. The compared methods are (left to right) Gaussian kernel ridge regression (GKRR), the elastic net on raw voxels, pyramid quantized Weisfeiler-Lehman (WLpyramid), and the elastic net with a concatenation of the raw voxels and the pyramid Weisfeiler-Lehman features (Combined EN+WL). The horizontal red line indicates chance performance. The pyramid quantized Weisfeiler-Lehman features perform better than GKRR and the elastic net on raw voxels with statistical signficance.

splitting scheme with 50 trials to estimate the classification performance. In each trial, a random selection of 80% of the data are used for training, while the remaining 20% are used to estimate the performance. In Table 1 we show the performance of the pyramid quantized Weisfeiler-Lehman model for four different depths of subtree patterns (cf. Section 2.3). Our approach achieves a mean accuracy of 64.28% for subtree patterns up to depth two. We also compare our proposed technique with three other methods on the same dataset: (i) Gaussian kernel ridge regression, (ii) the elastic net with raw voxels as features, and (iii) elastic net with raw voxels and pyramid Weisfeiler-Lehman features concatenated in a joint feature vector. In Fig. 2 we show the mean accuracy of the final system and the standard error. Pyramid quantized Weisfeiler-Lehman outperforms the rest of the methods. With a Wilcoxon signed rank test between elastic net with raw voxels and pyramid quantized Weisfeiler-Lehman we determine that our proposed method is statistically significantly better ($p = 0.02$). Additionally, a reduction of over 14% in classification error is recorded between elastic net on the raw voxels and our method.

4　Discussion

Fig. 1 shows the areas selected by the elastic net and visualizations of the learned functions. One of the main areas is the rostral anterior cingulate cortex (rostral ACC), which has been previously shown to deactivate in cocaine users in comparison to controls [14]. Moreover, it is an area whose activity is normalized by oral methylphenidate [19], which blocks the dopamine transporters, similar to cocaine, increasing extracellular dopamine–an escalation that has correlated with lower task-related impulsivity (errors of commission). Furthermore, this region has shown reduction in drug cue reactivity as a response to pharmacotherapeutic interventions in cigarette smokers [20, 21] and might be a marker of treatment response in other psychopathology (e.g. depression).

Although our method works in an implicitly high dimensional space, we empirically observe that Elastic Net regularization controls the complexity at each stage of the pipeline. The first learning step selects approximately 1100 voxels. Using the Weifeiler-Lehman algorithm, we generate a feature vector of length 6×10^5, but with a sparsity of $\sim 2\%$. The second application of Elastic Net selects only $\sim 2K$ dimensions. In each step, the method retains complexity much lower than a "simple" linear function over tens of thousands of voxels as has been proposed in previous works.

Several broad observations are apparent from our quantitative results. From Table 1, we note that subtree patterns up to depth two seem to perform best, and that deeper subtree patterns begin to reduce average performance. This indicates that the big-$\mathcal{O}$ complexity of the graph representation is only slightly higher than using a simple linear function. The proposed method performs significantly better than the ridge regression and elastic net baselines (Table 1 and Figure 2). In our final experiment of combining the raw voxel values with the subtree pattern features, we found that performance decreased slightly from that of only considering subtree pattern features.

In this work, we have presented a fully automated, statistically sound method for classification of brain states with graph representations. It combines the Weisfeiler-Lehman graph kernel [12] with the pyramid match kernel [13], in order to learn from graphs with continuous labels through a pyramid quantization scheme. The method was evaluated on a real world dataset and outperformed other machine learning techniques with statistical significance, including kernel ridge regression and the elastic net. This validates the primary hypothesis of this work: that the interconnections between voxels can contain additional information about brain structure that is not apparent in a linear function on the raw voxel values.

References

1. Demirci, O., Clark, V., Calhoun, V.: A projection pursuit algorithm to classify individuals using fMRI data: Application to schizophrenia. Neuroimage **39** (2008)
2. Wang, X., Hutchinson, R., Mitchell, T.M.: Training fMRI classifiers to discriminate cognitive states across multiple subjects. In: NIPS. (2003)

3. 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** (2004) 145–175

4. Tahmasebi, A.M., Artiges, E., Banaschewski, T., Barker, G.J., Bruehl, R., Bchel, C., Conrod, P.J., Flor, H., Garavan, H., Gallinat, J., Heinz, A., Ittermann, B., Loth, E., Mareckova, K., Martinot, J.L., Poline, J.B., Rietschel, M., Smolka, M.N., et al.: Creating probabilistic maps of the face network in the adolescent brain: A multicentre functional mri study. Human Brain Mapping **33** (2012) 938–957

5. Honorio, J., Tomasi, D., Goldstein, R., Leung, H., Samaras, D.: Can a single brain region predict a disorder? IEEE Transactions on Medical Imaging (2012)

6. Carroll, M., Cecchi, G., Rish, I., Garg, R., Rao, A.: Prediction and interpretation of distributed neural activity with sparse models. NeuroImage **44** (2009) 112–122

7. Gkirtzou, K., Honorio, J., Samaras, D., Goldstein, R., Blaschko, M.B.: fMRI analysis of cocaine addiction using k-support sparsity. In: ISBI. (2013)

8. Venkataraman, A., Kubicki, M., Golland, P.: From brain connectivity models to identifying foci of a neurological disorder. In: MICCAI. (2012) 715–722

9. Liu, Y., Liang, M., Zhou, Y., He, Y., Hao, Y., Song, M., Yu, C., Liu, H., Liu, Z., Jiang, T.: Disrupted small-world networks in schizophrenia. Brain **131** (2008)

10. Mokhtari, F., Hossein-Zadeh, G.A.: Decoding brain states using backward edge elimination and graph kernels in fMRI connectivity networks. Journal of Neuroscience Methods **212** (2013) 259–268

11. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B **67** (2005) 301–320

12. Shervashidze, N., Schweitzer, P., van Leeuwen, E.J., Mehlhorn, K., Borgwardt, K.M.: Weisfeiler-lehman graph kernels. JMLR **12** (2011) 2539–2561

13. Grauman, K., Darrell, T.: The pyramid match kernel: Efficient learning with sets of features. J. Mach. Learn. Res. **8** (2007) 725–760

14. Goldstein, R., Alia-Klein, N., Tomasi, D., Carrillo, J., Maloney, T., Woicik, P., Wang, R., Telang, F., Volkow, N.: Anterior cingulate cortex hypoactivations to an emotionally salient task in cocaine addiction. PNAS **106** (2009) 9453

15. Vishwanathan, S.V.N., Schraudolph, N.N., Kondor, R.I., Borgwardt, K.M.: Graph kernels. JMLR **11** (2010) 1201–1242

16. Sporns, O.: Networks of the Brain. MIT Press (2010)

17. Wee, C.Y., Yap, P.T., Li, W., Denny, K., Browndyke, J.N., Potter, G.G., Welsh-Bohmer, K.A., Wang, L., Shen, D.: Enriched white matter connectivity networks for accurate identification of MCI patients. NeuroImage **54** (2011) 1812 – 1822

18. Weisfeiler, B., Lehman, A.A.: A reduction of a graph to a canonical form and an algebra arising during this reduction. Nauchno-Tech. Informatsia, Ser. 2 **9** (1968)

19. Goldstein, R.Z., Woicik, P.A., Maloney, T., Tomasi, D., Alia-Klein, N., Shan, J., Honorio, J., Samaras, D., Wang, R., Telang, F., Wang, G.J., Volkow, N.D.: Oral methylphenidate normalizes cingulate activity in cocaine addiction during a salient cognitive task. PNAS **107** (2010) 16667–72

20. Culbertson, C., Bramen, J., Cohen, M., London, E.D., Olmstead, R.E., Gan, J.J., Costello, M.R., Shulenberger, S., Mandelkern, M.A., Brody, A.L.: Effect of bupropion treatment on brain activation induced by cigarette-related cues in smokers. Archives of General Psychiatry **68** (2011) 505–515

21. Franklin, T.R., Wang, Z., Li, Y., Suh, J.J., Goldman, M., Lohoff, F.W., Cruz, J., Hazan, R., Jens, W., Detre, J.A., Berrettini, W., O'Brien, C.P., Childress, A.R.: Dopamine transporter genotype modulation of neural responses to smoking cues: confirmation in a new cohort. Addiction Biology **16** (2011) 308–322