

Integration of Sparse Bayesian Learning and Random Subspace for fMRI Multivariate Pattern Analysis

Shulin Yan¹, Xian Yang¹, Chao Wu¹ and Yike Guo^{1,2}

Abstract—Multivariate Pattern Analysis (MVPA) is frequently used to decode cognitive states from brain activities in fMRI study. Due to the discrepancy between sample and feature size, MVPA methods are suffered from the overfitting problem. This paper addresses this issue by introducing sparse modelling along with its advanced decoding method, Compressive Sensing (CS). As brain voxels have highly correlated in spatial domain, the prerequisite of CS methods are not well satisfied. We therefore propose a novel MVPA method to integrate linear Sparse Bayesian Learning (i.e. Bayesian Compressive Sensing) with random subspace method. Benefiting from the random subspace method, spatial correlation and feature-to-sample ratio are largely reduced. The experimental results from a real fMRI dataset demonstrate that our method has distinct prediction power compared to three other popular MVPA methods, and the detected relevant voxels are located in informative brain areas.

I. INTRODUCTION

Functional MRI (fMRI) is a neuroimaging technique to investigate the relationship between brain regions with specific cognitive functions by measuring changes in brain blood flow signal (BOLD). Conventional fMRI analyses focus on investigating the interpretation of neural activities with univariate analysis, such as General Linear Model (GLM) [1]. The univariate analysis methods work on isolated voxels and they determine active brain regions with the most statistically significant voxels in response to a cognitive task.

In contrast to univariate analyses, Multivariate Pattern Analysis (MVPA) of fMRI attempts to informatively decode patterns of brain activities [2]. By measuring multi-voxels simultaneously, MVPA is more sensitive and informative to the brain activities and robust to noises. Recent advanced MVPA methods [3], [4], [5] employ sparse modelling and have shown distinct advantages: 1) they can directly work on the whole brain without introducing feature

selection methods beforehand, 2) they only use a small subset of input voxels for prediction so that overfitting problem can be eliminated, 3) when linear sparse model is used, neural activities can be interpreted by studying the selected voxels. On the other hand, random subspace method has also been proven to alleviate the overfitting problem [6].

In this paper, we propose a robust sparse modelling method for fMRI analysis; it is implemented by incorporating linear Sparse Bayesian Learning (SBL) with the random subspace method. We show that with the introduction of random subspace method, the performance of linear SBL will be improved based on the following facts: 1. the correlation among features in a subspace can be tremendously reduced by random selection so that better conditioned matrix can be produced; 2. discrepancy between sample size and feature size in a subspace is highly reduced. As a consequence, linear SBL can provide robust predictors for subspaces, and a final strong predictor can be constructed via an aggregating process. Moreover, benefiting from the linear sparse model, prediction maps are provided so that interpretation of neural activities can be investigated.

II. LINEAR SPARSE MODELLING

The results of fMRI analysis show that active brain regions responding to a cognitive task are just a small part of the entire investigated area (e.g. whole brain). Motivated by this observation, the decoding problem can be formalized by linear sparse modelling. This gives:

$$y = Xw, \quad (1)$$

where $X \in R^{n \times p}$ composes of n observations and p features corresponding to a response vector $y \in R^{n \times 1}$, and the non-zero elements of $w \in R^{p \times 1}$ indicate their corresponding voxels are relevant to the cognitive state y . The total number of relevant voxels, k , should be far less than the total number of voxels, that is $k \ll p$. As in fMRI study, $n \ll p$, estimation of w is ill posed.

Compressive Sensing (CS) technique [7] offers an opportunity for solving this problem. However, the

¹Data Science Institute, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom {shu.yan09, xian.yang08, chao.wu, y.guo} at imperial.ac.uk

²Corresponding Author

spatial correlation of voxels leads to poor conditioned design matrix. When several columns of the design matrix X are highly correlated, the estimation of their corresponding non-zero elements is difficult that only part of them may be detected [5]. To address this problem, the elastic net method is proposed to provide a relaxation on poor conditioned matrix where the correlated voxels are highly grouped [4]. Another alternative approach, randomized ward lasso [5], is developed to reduce the spatial correlation by grouping correlated voxels together and using the average value of each cluster to construct design matrix. This method is focused on the relevant voxel recovery rather than the prediction power of the linear model. Moreover, the above methods involve penalty parameters that must be fixed beforehand using a hold-out method such as cross-validation.

III. METHODS

In this work, we integrate Sparse Bayesian Learning (SBL) [8] with the random subspace method to construct a robust prediction model composed of multiple SBL predictors. SBL is adopted to build the linear sparse model. The SBL algorithm has been demonstrated that it can empirically provide a useful sparse solution even when the design matrix is in poor condition and no penalty parameters need to be defined via cross-validation.

Random subspace (RS) method aims to generate multiple predictors, from which a strong aggregated predictor can be produced. It starts with generating L subspaces, each of which contains M voxels randomly sampled (with replacement) from the whole input space. Then, a sparse linear model is constructed in each subspace. The relevant voxels in subspaces can be more easily detected by SBL. That is because, in each subspace, both the relevant-to-sample ratio and the spatial correlation are reduced due to the decreased size of relevant voxels and random sampling process respectively. However, even when the relevant features can be correctly detected, the majority of predictors are weak as each of them only involves partial information. Therefore, an aggregating process is employed to produce a strong predictor involving all detected relevant features. Here, we use a common aggregating method, majority voting, which counts the largest number of predicted results that agree with each other. The final result benefits from the random sampling process that it can highly reduce the probability of involving biased voters.

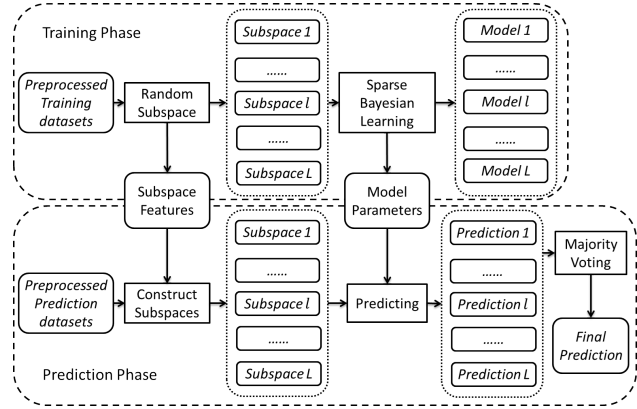


Fig. 1: The framework of RS-SBL in training and prediction phases.

Figure 1 shows the framework of our method, random subspace SBL with linear model (RS-SBL), by illustrating how our method works in both training and prediction phases. In the training process, the RS-SBL method generates a set of prediction models along with their selected features from a training dataset. In the prediction process, the prediction models learned from training process are applied to a prediction dataset, where the final prediction result is obtained by majority voting. Note that both training and prediction datasets are constructed with preprocessed whole brain fMRI images.

IV. RESULTS

The fMRI data used in our experiments were provided by Human Connectome Project (HCP) [9] (see reference for more details). In our work, we adopted our algorithm to analyse the relationship processing task that was conducted to study the active brain regions for processing internally and externally generated information. Under the relationship processing experiment, we focus on three classification tasks within individual subjects to check whether the subject: 1. did relation processing or stayed relaxed, 2. did match processing or stayed relaxed, 3. did relation processing or match processing. We performed the experiments on the preprocessed whole brain fMRI data of a group of 10 unrelated subjects selected from 77 candidates. For each subject, we performed training process on one session and prediction on the other. Each session consisted of 232 samples and each of the three response vectors (relation, match, and rest) had approximately one third of the samples. As a result, there were about 155 training samples for each classification task. As

the total number of voxels ($\approx 228,000$) was far more than the number of samples, training classification models was a underdetermined problem.

In these experiments, we focus on constructing binary classifiers for cognitive task using the classification algorithm of the SBL, SBC [8]. We compared the performance of our proposed algorithm, RS-SBC, to three other MVPA methods:

- *Sparse Bayesian Classification with linear model (SBC)*. This is the fundamental classification algorithm of our method, where the design matrix is constructed by the features in voxel space other than any projected spaces (e.g. projected using linear kernel function).
- *Support Vector Machine with linear kernel (SVM-Lin)*. SVM has been the most popular classifier for fMRI data analysis, and the most useful version is the one with linear kernel [6]. The C parameter of linear kernel is optimized by 10-fold cross-validation in our experiment.
- *Sparse Bayesian Classification with linear kernel (SBC-Lin)*. SBC has been demonstrated having similar (or even better) classification performance to SVM on some applications (e.g. hyperspectral image classification [10]). However, no comparison has been made on fMRI analysis. We constructed this competitor using SBC with linear kernel to make it comparable to the above SVM classification method.

All the competitors and our proposed algorithm were applied on whole brain fMRI images rather than a subset of voxels selected with a feature selection algorithm. The classification performance was evaluated by classification accuracy. We set ($L = 100, M = 10\%p$) for both Relation vs. Rest and Match vs. Rest and ($L = 30, M = 33\%p$) for Relation vs. Match. Due to the randomness of sampling, the results were averaged over 50 experiments.

A. Classification Accuracy

Figure 2 shows the average classification accuracies across 10 subjects using four different classifiers. Each of the three classification tasks contained two experiments: using session 1 to predict session 2 and vice versa. The results demonstrate that our proposed method has the highest classification accuracy, while SVM has the worst performance as it cannot handle overfitting problem with huge number of voxels.

SBC-Lin performs better than SBC in the first two classification tasks but worse in the last one. For the first two tasks, although the number of relevant voxels is small compared to the total number of input

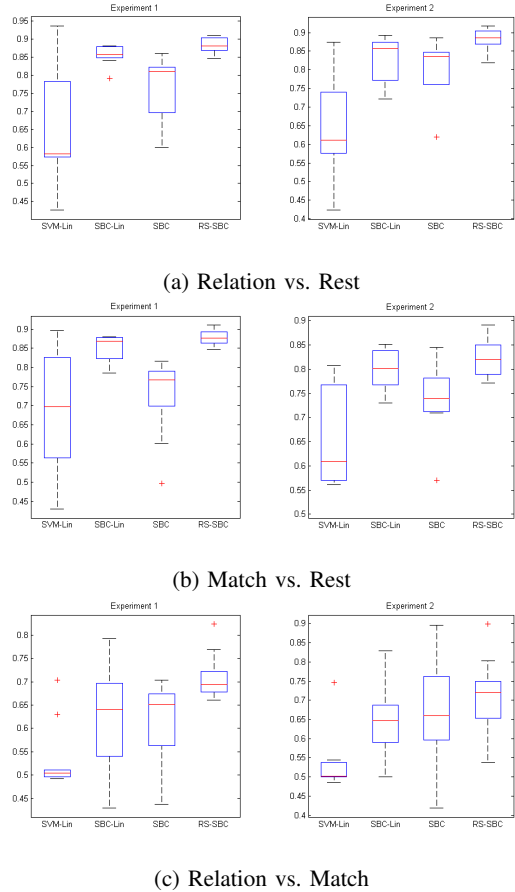


Fig. 2: Classification accuracies across 10 subjects with four different classifiers: SVM-Lin, SBC-Lin, SBC and RS-SBC.

voxels, compressive sensing theory cannot detect all relevant ones with such small sample size. The number of relevant voxels found by SBC is only several dozen which was extremely sparse compared to the significant voxels found by GLM (around 1000), leading to inaccurate estimation. RS-SBC increases estimation accuracy by reducing the number of input voxels in each subspace. As the difference between the relation and match tasks only lies in the internal information generation region, the important voxels is much sparser in the last task. Samples are relatively sufficient to find the most important voxels using SBC. Therefore SBC outweigh the performance of SBC-Lin.

Compared to SBC, RS-SBC can further increase classification accuracy by detecting more relevant voxels. On average, the classification accuracy is improved 15% over the SBC method, and 3% better than SBC-Lin method. The number of relevant voxels determined by RS-SBC is larger than SBC but much

smaller than SBC-Lin. Approximately 1500 and 300 voxels are found for the first two and last tasks. Along with the highest average accuracy, our method can construct the most stable classifier with the smallest variance across all the cases.

B. Prediction Map

We illustrate some examples of 'individual' prediction maps obtained by our method and active maps generated by GLM analysis. The GLM analysis, implemented by SPM first-level analysis [11], returns the active maps containing voxels with $FWE \leq 0.05$ (empirical value). Our method produces prediction maps by selecting voxels that are supposed to be relevant on average across 50 repeated experiments. The maps of both methods are created by registering relevant voxels to the subject's T1w structural image. Figure 3 shows an example of the prediction maps of

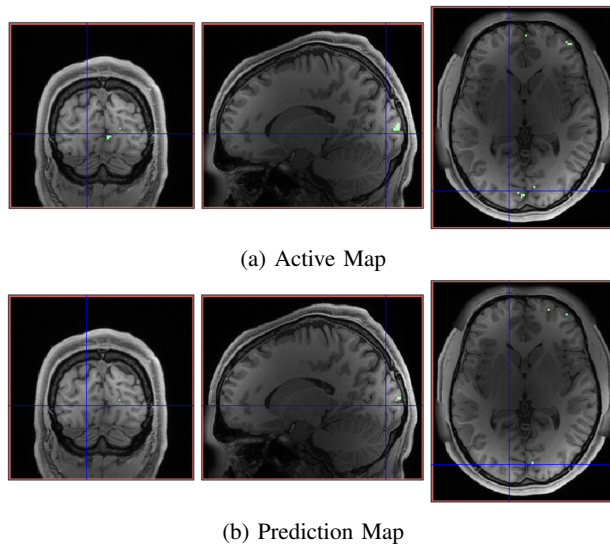


Fig. 3: An example of 'individual' prediction map vs. active map for Relation vs. Match task.

Relation vs. Match. We can see that relevant voxels detected by different methods are not identical, but similar brain regions are clearly outlined. The located brain regions with the relevant voxels are consistent with the findings of the work in [12] that the task activates both Rostrolateral Prefrontal Cortex (RLPFC) and primary visual cortex.

V. CONCLUSIONS

This paper proposes the robust sparse modelling method (RS-SBL) to generate brain behaviour predictors and prediction maps from whole brain fMRI images. Our work is the first attempt to integrate

random subspace method with SBL. We have shown that the performance of our method outweighs three other methods using the dataset provided by HCP. Among those methods, SVM-Lin and SBC-Lin involve feature space transformation so that all voxels are used for classification. Hence, relevant voxels are difficult to be selected and the prediction models tend to overfit the training data. SBC, on the other hand, directly detect relevant voxels but the resulted prediction map is extremely sparse. Our method, benefiting from the implementation of random subspace, is able to provide meaningful prediction maps with the strongest prediction power.

ACKNOWLEDGMENT

This research is partially supported by the Innovative R&D Team Support Program of Guangdong Province (NO. 201001D0104726115), China

REFERENCES

- [1] Friston, K.; Jezzard, P. and Turner, R., Analysis of functional MRI time-series Human brain mapping, Wiley Online Library, 1994, 1, 153-171
- [2] Haxby, J. V.; Gobbini, M. I.; Furey, M. L.; and others, Distributed and overlapping representations of faces and objects in ventral temporal cortex Science, American Association for the Advancement of Science, 2001, 293, 2425-2430
- [3] Yamashita, O.; Sato, M.-a.; Yoshioka, T.; Tong, F. and Kamitani, Y., Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns NeuroImage, Elsevier, 2008, 42, 1414-1429
- [4] Carroll, M. K.; Cecchi, G. A.; Rish, I.; Garg, R. and Rao, A. R., Prediction and interpretation of distributed neural activity with sparse models NeuroImage, Elsevier, 2009, 44, 112-122
- [5] Varoquaux, G.; Gramfort, A. and Thirion, B., Small-sample brain mapping: sparse recovery on spatially correlated designs with randomization and clustering arXiv preprint arXiv:1206.6447, 2012
- [6] Kuncheva, L. I.; Rodriguez, J. J.; Plumpton, C. O.; Linden, D. E. and Johnston, S. J., Random subspace ensembles for fMRI classification Medical Imaging, IEEE Transactions on, IEEE, 2010, 29, 531-542
- [7] Donoho, D. L., Compressed sensing Information Theory, IEEE Transactions on, IEEE, 2006, 52, 1289-1306
- [8] Tipping, M. E., Sparse Bayesian learning and the relevance vector machine The journal of machine learning research, JMLR. org, 2001, 1, 211-244
- [9] Barch, D. M.; Burgess, G. C.; Harms, M. P.; and others, Function in the human connectome: Task-fMRI and individual differences in behavior NeuroImage, Elsevier, 2013, 80, 169-189
- [10] Demir, B. and Erturk, S., Hyperspectral image classification using relevance vector machines Geoscience and Remote Sensing Letters, IEEE, IEEE, 2007, 4, 586-590
- [11] Ashburner, J.; Barnes, G.; Chen, C.; and others, SPM8 manual Functional Imaging Laboratory, Institute of Neurology, Citeseer, 2008
- [12] Christoff, K.; Ream, J. M.; Geddes, L. P. T. and Gabrieli, J. D. E., Evaluating self-generated information: anterior prefrontal contributions to human cognition Behavioral Neuroscience, PscARTICLES, 2003, 117(6), 1161-1168