Decoding Cognitive Processes from Functional MRI

Oluwasanmi Koyejo[1] and Russell A. Poldrack[2]

[1] Imaging Research Center, University of Texas at Austin
sanmi.k@utexas.edu
[2] Depts. of Psychology and Neuroscience, University of Texas at Austin
poldrack@utexas.edu

Abstract. The goal of cognitive neuroscience is to understand the the brain processes that underlie cognitive function. These brain processes are studied by examining neural responses to experimental tasks and stimuli. While most experiments are designed to isolate a single cognitive process, the resulting brain images often encode multiple processes simultaneously. Thus standard classification methods are inappropriate for decoding cognitive processes. We propose a multilabel classification approach for decoding, and present empirical evidence that multilabel classification can accurately predict the set of cognitive processes associated with an experimental contrast image.

1 Introduction

An important hypothesis in modern cognitive neuroscience is that brain function is decomposable into a set of elementary cognitive processes - representing the basis set of brain functions recruited for cognitive tasks [13]. For example, recognizing a face may require the cognitive process of vision, working memory and retrieval, while the music comprehension may require, in addition to the shared cognitive processes of working memory and retrieval, additional cognitive processes of rhythm and intonation. Cognitive neuroscientists and other researchers measure these processes in the laboratory setting by developing experiments that allow (e.g., via cognitive subtraction) the isolation of a specific cognitive process from other recruited processes. Unfortunately, despite careful selection of the stimuli and control tasks, the measured brain function often captures multiple cognitive processes simultaneously [8].

Functional magnetic resonance imaging (fMRI) has enabled the non-invasive measurement of brain function in response to experimental stimuli at fine spatial scales. From initial studies that used classifiers to discriminate between different classes of visual objects [4] to more recent studies showing large scale classification across experiments [11], decoding from brain images has become an important research tool [7]. Decoding performance can be used to test hypothesis about the cognitive content of the brain images. Further, the classifier parameters can be used to localize predictive voxels [3], or select regions of interest for additional processing. In addition to the general scientific utility of decoding, the specific application to cognitive processes may help address additional scientific questions, such as which cognitive processes outlined in the literature represent true differences in brain function, and which merely reflect theoretical distinctions [10]. Despite these potential insights, direct decoding of cognitive processes from brain function has not been attempted before.

Table 1. Cognitive Process Labels Sorted by Prevalence in Data.

Code	Process Label
A	Vision
B	Action Execution
C	Decision Making
D	Orthography
E	Shape Vision
F	Audition
G	Phonology
H	Conflict
I	Semantics
J	Reinforcement Learning
K	Working Memory
L	Feedback
M	Response Inhibition
N	Reward
O	Stimulus-driven Attention
P	Speech
Q	Emotion Regulation
R	Mentalizing
S	Punishment
T	Error Processing
U	Memory Encoding
V	Spatial Attention

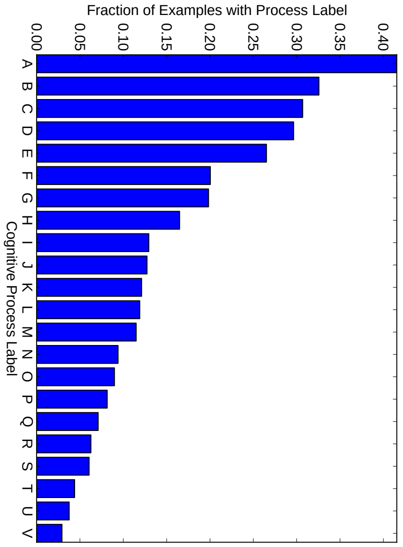


Fig. 1. Fraction of Data with each Cognitive Process Label.

We study the decoding of cognitive processes from brain function measured via functional magnetic resonance imaging (fMRI) contrasts using a multilabel classification approach. Multilabel classifiers are designed to solve classification problems where each example may be associated with multiple processes, and are popular in several domains such as image processing and text processing [14]. We focus on the subclass of multilabel classification methods known as label decomposition methods [14], where the multilabel classification problem is decomposed into multiple binary classification problems. Our work is enabled by the recent availability of a large public fMRI database (OpenFMRI[3]) [9] and a large cognitive ontology labeled by domain experts (Cognitive Atlas[4]) [12]. Our results provide empirical evidence that the set of cognitive processes associated with an experimental contrast can be accurately decoded.

Notation: We denote vectors by bold-face lower case letters $\mathbf{x}$ and matrices by bold-face capital letters $\mathbf{X}$. The set of real valued D dimensional vectors are denoted by $\mathbb{R}^D$. Label sets are denoted by script capital letters $\mathcal{S}$ with cardinality $|\mathcal{S}|$.

2 Methods

Let $\mathbf{x}_n \in \mathbb{R}^D$ denote the n^{th} brain volume with voxels collected into a real valued D dimensional vector. The total number of brain volumes is represented by N. Each brain volume is associated with a set of process labels $\mathcal{S}_n = \{s_1, \dots s_K\}$ chosen from the full set of possible process labels $\mathcal{L} = \bigcup_{n=1,\dots,N} \mathcal{S}_n$ with $|\mathcal{L}| = L$. Multilabel classification

[3] openfmri.org

[4] www.cognitiveatlas.org

involves estimating a predictive mapping $\mathbf{f} : \mathbf{x}_n \mapsto \mathcal{S}_n$. There are several approaches in the literature for multilabel classification including label decomposition, label ranking, and label projection methods [14]. We focus on label decomposition methods due to their simplicity, scalability and ease of interpretation. Label decomposition methods separate the multilabel classification task into a set of binary classification tasks. A popular approach in this family is the *One-Vs-All* decomposition, where the multilabel classification is decomposed into binary classification tasks. Each binary classification model is trained to predict the presence or absence of each each label independently.

We experimented with the multilabel decomposition approach using the following base classifiers: (i) l_2 regularized support vector machine (**SVM**) [2] , (ii) l_2 regularized logistic regression (**Logistic**) [1] , and l_2 regularized squared loss classifier (**Ridge**) [1]. Each sub-classifier was implemented using a linear model of the form $f_l(\mathbf{x}_n) = \mathbf{w}_l^\top \mathbf{x}_n$ where $\mathbf{w}_l \in \mathbb{R}^D \; \forall \, l = 1 \ldots L$ is a real valued weight vector. In addition, we experimented with a baseline multilabel classifier (**Popularity**) designed to approximate the dataset label statistics. To this end, the set of predicted process labels were determined based on prevalence in the training set. Specifically, the indicator denoting the presence of each label was drawn independently from a Bernoulli distribution with probability given by the fraction of examples in the training data containing that label.

3 Empirical Results

We compiled brain image data from the publicly available openfMRI database [9]. OpenfMRI contains pre-extracted z-statistic contrasts for each subject computed using a generalized linear model. This data extraction was implemented using the FMRIB Software Library (FSL). Combining the whole brain data with the standard brain mask resulted in $D = 174,264$ extracted voxels. We extracted $N = 479$ contrast images associated with 26 contrasts in the database. Further details on data preprocessing may be found in [9]. In addition to the brain volumes, we extracted a list of cognitive process labels associated with each experimental contrast. The list was curated starting from processes in the Cognitive Atlas [12] and refined by domain experts. The final set of $L = 22$ cognitive process labels are provided in Table 1. It is clear from Fig. 1 that some process labels are significantly more prevalent in the data than other process labels. For example vision is more than 20 times more prevalent than spatial attention. The data samples included an average of 3.5 process labels per example with a maximum of 9 process labels per example and a minimum of 1 process label per example.

We evaluated the models using (label) *Accuracy*, *Precision*, *Recall*, *Hamming loss* and *F1Score*, metrics commonly applied for evaluating multilabel classification [14]. Let $\mathcal{S}_n$ represent the true process labels and $\mathcal{Z}_n$ represent the predicted process labels associated with the n^{th} example. The metrics are computed as:

$$\text{Precision} = \frac{1}{N}\sum_{n=1}^{N}\frac{|\mathcal{S}_n \cap \mathcal{Z}_n|}{|\mathcal{Z}_n|}, \text{Recall} = \frac{1}{N}\sum_{n=1}^{N}\frac{|\mathcal{S}_n \cap \mathcal{Z}_n|}{|\mathcal{S}_n|}, \text{Accuracy} = \frac{1}{N}\sum_{n=1}^{N}\frac{|\mathcal{S}_n \cap \mathcal{Z}_n|}{|\mathcal{S}_n \cup \mathcal{Z}_n|},$$

$$\text{Hamming Loss} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{L}|\mathcal{S}_n \ominus \mathcal{Z}_n|, \quad \text{F1Score} = \frac{1}{N}\sum_{n=1}^{N}\frac{2.0 * \text{Precision}_n \times \text{Recall}_n}{\text{Precision}_n + \text{Recall}_n},$$

Table 2. Mean (var) of Aggregated Performance Metrics. *- represents models where all metrics are statistically significant $(p < 10^{-3})$ wrt. the permutation based null distribution for the model.

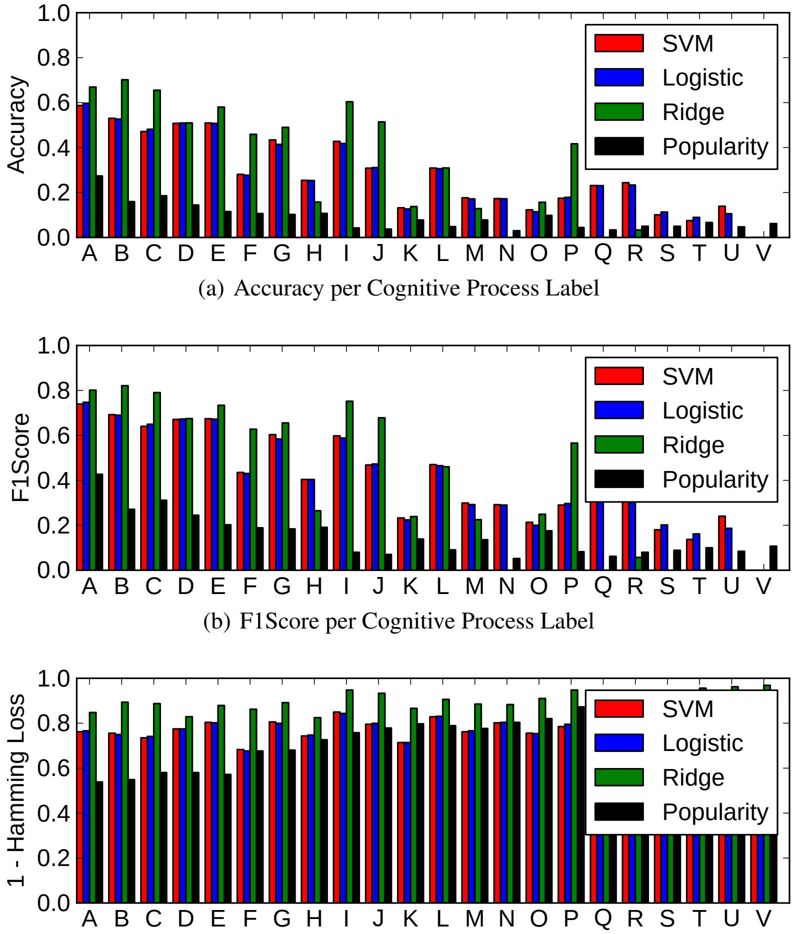
	Accuracy	Precision	Recall	F1Score	1 - Hamming Loss
SVM*	0.43 (0.03)	**0.53 (0.03)**	0.68 (0.03)	0.51 (0.03)	0.79 (0.01)
Logistic*	**0.44 (0.03)**	**0.53 (0.02)**	0.68 (0.03)	**0.52 (0.03)**	0.79 (0.01)
Ridge*	0.34 (0.02)	0.47 (0.02)	0.37 (0.02)	0.39 (0.02)	**0.91 (0.00)**
Popularity	0.12 (0.01)	0.21 (0.02)	0.18 (0.03)	0.18 (0.02)	0.76 (0.01)

where $\mathcal{A} \ominus \mathcal{B}$ represents the symmetric difference of set $\mathcal{A}$ and $\mathcal{B}$. Label *Accuracy* measures the average fraction of process labels that are predicted correctly with respect to the cardinality of the union of true and predicted process labels. *Precision* measures the fraction of predicted process labels that are relevant, and *Recall* measures the fraction of relevant process labels that are predicted. The *F1Score* combines *Precision* and *Recall* into a single score. Higher scores indicate superior performance for *Accuracy*, *Precision*, *Recall* and *F1Score*, and the best possible score is 1. The *Hamming Loss* directly penalizes both false positives and false negatives equally. Lower cores indicate superior performance for *Hamming Loss*, and the best possible score is 0. To simplify comparison with other scores, we present results as 1 - *Hamming Loss*. Further details on the metrics are available in [14].

All models were trained using 5-fold double loop cross validation. The inner loop was used for parameter selection, and the outer loop was used to estimate the generalization performance. The l_2 regularization parameter for all models was selected from the set $\{10^2, 10^{1.5}, 10^1, \ldots, 10^{-2.5}, 10^{-3}\}$. We used the *Hamming Loss* metric for parameter selection. We evaluated the use of the other metrics for parameter selection and found the results to be qualitatively equivalent. In addition to performance comparisons, we were interested in evaluating the statistical significance of the results. Hence, we computed an empirical null distribution by randomly permuting the process labels 1000 times and retraining the model. Note that the empirical null distribution was estimated separately for each trained model, so the presented statistical significance is model dependent. We computed statistical significance using a threshold of $p = 10^{-3}$, suggesting high confidence in rejecting the hypothesis that the performance scores were the result of chance.

We found that the performance of **SVM** and **Logistic** were almost identical in aggregate (Table 2). **Ridge** was comparable to **SVM** and **Logistic** in terms of *Precision*, but performed worse in terms of *Accuracy* and *Recall*. On the other hand, **Ridge** significantly outperformed all other models in terms of *Hamming Loss*. To investigate these observations further, we computed per-label performance metrics as shown in Fig. 2. As expected, the overall trend of most of the metrics was correlated with the label imbalance i.e. more common process labels were easier to predict. Our results show that **Ridge** was the most accurate model for prevalent process labels. However, **Ridge** was not accurate for rare process labels. Surprisingly, some cognitive process labels such as *Speech* were well predicted by **Ridge** despite their rarity.

To investigate any systematic bias in classifier mistakes, we computed the classifier confusion matrices as shown in Fig. 3 (only confusion matrices for **Ridge** and **Logistic**



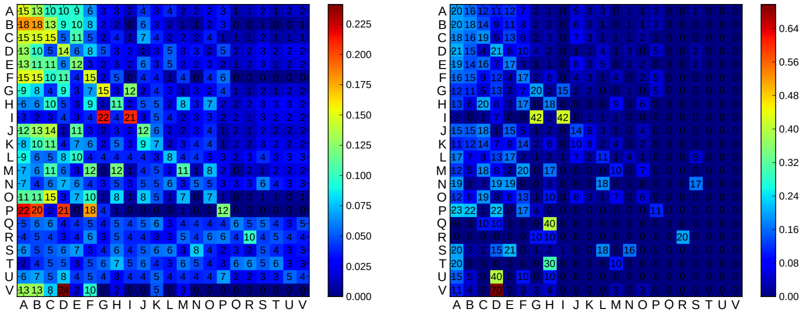
(a) Accuracy per Cognitive Process Label

(b) F1Score per Cognitive Process Label

(c) 1 - Hamming Loss per Cognitive Process Label

Fig. 2. Model Performance per Cognitive Process Label. For metrics other than *Hamming Loss*, label prevalence is highly correlated with classification performance. **Ridge** was especially accurate for the most prevalent process labels, but was relatively less accurate for rare process labels. The cognitive process labels are coded as capital letters $A, \ldots, V$ (see Table 1). Figure is best viewed in color.

are shown due to limited space). Each row represents the average fraction of examples where the cognitive process label associated with the row was predicted as the cognitive process label associated with the column. Across process labels, it was clear that labeling mistakes were systematically in the direction of more prevalent process labels i.e. the confusion matrices are brighter towards the left side. The cooler color in **Ridge** was mostly due to the high proportion of mistakes made for *Spatial Attention* - the rarest process label. Examining the right side of the confusion matrices, we note that **Logis-**



(a) Logistic (b) Ridge

Fig. 3. Avg. of Normalized Confusion Matrices for **Logistic** and **Ridge**. The true process labels are along the row, and the predicted process labels are along the columns. Each row represents the average fraction of examples where the cognitive process label associated with the row was predicted as the cognitive process label associated with the column. The matrix entries are scaled $\times 10^2$ to improve readability. Cognitive process labels are coded as capital letters $A, \ldots, V$ (see Table 1). Figure is best viewed in color.

tic sometimes classified prevalent process label examples as rare process labels, while **Ridge** rarely made such mistakes at the expense of low accuracy for rare process labels. Recall that the cost of ignoring rare process labels is relatively low for the *Hamming Loss* as compared to the other losses. This explains the relatively high performance of **Ridge** for the *Hamming Loss*. The empirical results suggest that a multi-classifier approach combining the advantages of the different classifiers may be effective. For example, **Ridge** could be used for predicting the most prevalent process labels, and combined with **Logistic** for predicting rare cognitive process labels.

4 Conclusion

The decoding of cognitive processes is an important first step towards evaluating and verifying the latent processes the brain employs to complete various tasks. We have provided experimental evidence that cognitive processes can be accurately decoded from brain function using a multilabel classification approach. We also studied some of the trade-offs that arise due to the imbalance of the process labels. We intend to continue further verification of the decoding performance by evaluating various multilabel classification methods in the literature [14]. This will also aid in understanding the trade-offs between different methods in the specific application to neuroimaging data. In addition, we plan to incorporate structured regularizers such as the total variation regularization [5], or Bayesian models for structured sparsity [6] that may help to localize the sources of classification performance, improving the interpretability of the results.

Bibliography

[1] Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006)

[2] Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov. 2, 121–167 (1998)

[3] De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., Formisano, E.: Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. NeuroImage 43, 44–58 (2008)

[4] 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)

[5] Michel, V., Gramfort, A., Varoquaux, G., Eger, E., Thirion, B.: Total variation regularization for fmri-based prediction of behavior. Medical Imaging, IEEE Transactions on 30, 1328–1340 (2011)

[6] Park, M., Koyejo, O., Ghosh, J., Poldrack, R.R., Pillow, J.W.: Bayesian structure learning for functional neuroimaging. In: International Conference on Artificial Intelligence and Statistics (AISTATS) (2013)

[7] Pereira, F., Mitchell, T., Botvinick, M.: Machine learning classifiers and fMRI: A tutorial overview. NeuroImage 45, S199–S209 (2009)

[8] Poldrack, R.A.: Subtraction and beyond: The logic of experimental designs for neuroimaging. In: Hanson, S.J., Bunzl, M. (eds.) Foundational Issues in Human Brain Mapping, pp. 147–160. MIT Press, Cambridge, MA (2010)

[9] Poldrack, R.A., Barch, D.M., Mitchell, J.P., Wager, T.D., Wagner, A.D., Devlin, J.T., Cumba, C., Koyejo, O., Milham, M.P.: Towards open sharing of task-based fMRI data: The OpenfMRI project. Frontiers in Neuroinformatics (2013)

[10] Poldrack, R.A.: Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. Neuron 72(5), 692–697 (2011)

[11] Poldrack, R.A., Halchenko, Y.O., Hanson, S.J.: Decoding the large-scale structure of brain function by classifying mental states across individuals. Psychological Science 20, 1364–1372 (2009)

[12] Poldrack, R.A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., Parker, D.S., Sabb, F.W., Bilder, R.M.: The cognitive atlas: Towards a knowledge foundation for cognitive neuroscience. Frontiers in Neuroinformatics 5 (2011)

[13] Posner, M.I., Petersen, S.E., Fox, P.T., Raichle, M.E.: Localization of cognitive operations in the human brain. Science 240(4859), 1627–31 (Jun 1988)

[14] Zhang, M.L., Zhou, Z.H.: A review on multi-label learning algorithms. IEEE Transactions on Knowledge and Data Engineering 99(PrePrints), 1 (2013)