DOI 10.3233/BME-141116
IOS Press

Sparse models for visual image reconstruction from fMRI activity

Linyuan Wang, Li Tong, Bin Yan[*], Yu Lei, Lijun Wang, Ying Zeng and Guoen Hu
National Digital Switching System Engineering & Technological R & D Center, Zhengzhou 450002, China

Abstract. Statistical model is essential for constraint-free visual image reconstruction, as it may overfit training data and have poor generalization. In this study, we investigate the sparsity of the distributed patterns of visual representation and introduce a suitable sparse model for the visual image reconstruction experiment. We use elastic net regularization to model the sparsity of the distributed patterns for local decoder training. We also investigate the relationship between the sparsity of the visual representation and sparse models with different parameters. Our experimental results demonstrate that the sparsity needed by visual reconstruction models differs from the sparsest one, and the l2-norm regularization introduced in the EN model improves not only the robustness of the model but also the generalization performance of the learning results. We therefore conclude that the sparse learning model for visual image reconstruction should reflect the spasity of visual perceptual experience, and have a solution with high but not the highest sparsity, and some robustness as well.

Keywords: sparse learning model, visual image reconstruction, sparsity, elastic net

1. Introduction

Decoding perceptual experience from brain activity is a major challenge in neuroscience, as it is impractical to specify brain activity for all of the possible states of visual perception. There are many early fMRI studies, which have demonstrated that visual features, such as visual object categories [1,2], and orientation and motion direction [3,4] can be decoded from fMRI activity patterns by a statistical learning classifier. Two recent studies presented that a natural image can be identified among a large number of candidate images using the visual encoding model [5,6]. In these fMRI studies, statistical "decoder" is trained to learn the mapping between brain activity pattern and a stimulus, and is used to predict fMRI activity for visual images. Even more challenging study is visual image reconstruction from fMRI activity patterns [7]. To capture the complexity of visual perceptual, it decodes visual perception by utilizing retinotopy into an image that is free from the constraints of categories. A presented visual image can be inferred given the brain activity consistent with retinotopy in the early visual cortex by inverting the receptive-field model [8].

A new approach for visual image reconstruction was recently proposed using multivoxel patterns of fMRI signals [9]. The stimulus state at each local element was predicted by a decoder using

[*]Corresponding author: Bin Yan, National Digital Switching System Engineering & Technological R & D Center, Zhengzhou 450002, China. Tel.: +08681630583; Fax: +08681630531; E-mail: tom.yan@gmail.com.

multivoxel patterns, after which all the outputs of each local decoder were combined to reconstruct the presented image. The training of each local decoder only needs a small number of samples, as the possible states of each local element are fewer than those of the entire image. Another study used a learning algorithm, "sparse logistic regression", for training a local decoder [10]. The estimation of sparse parameter avoids overfitting to noisy training data and thereby help on achieving a high generalization (test) performance [10].

Proper variable selection is widely known to be as critical for prediction as it is for interpretation in statistical analysis [11]. Sparse modeling, in which resulting models only use information from a relatively small subset of predictive variables, has a strong prediction performance. Learning statistical models from fMRI data is particularly difficult, as the training data may easily be overfitted and resulting model may generalize poorly. Many predictor selection techniques [12] use a straightforward filtering stage, but sparse modeling based methods can combine the selection and modeling states into one process, which is usually called as embedded selection. Sparse modeling based methods, such as least absolute shrinkage and selection operator (LASSO), compare favorably with non-sparse methods on prediction performance because they incorporate multivariate information into the selection process.

Recently, powerful methods that simultaneously select discriminant voxels and estimate their weights for classification have been proposed [10,13,14]. These methods incorporate sparse regularization into traditional classifiers, and avoid overfitting by assigning zero weights to be assigned to irrelevant voxels. Carroll et al. [14] recommended elastic net (EN) regularization [15] for fMRI modeling.

2. EN model

In this section, we will formally describe the EN method. Since both the predicted mental states and fMRI data are viewed as real-valued time series, the prediction task is formulatecd as a regression problem, where individual TRs are seen as independent and identically distributed (i. i. d.) samples, the predictive variables (predictors) are the levels of voxel activity , and the predicted variable is the mental state .

EN regularization [15] was first designed to build models that have both sparsity and grouping effect using a weighted combination of l_1- and l_2-norm penalties on a least-squares problem. It can be expressed as minimizing the following problem:

$$L_{\lambda_1,\lambda_2}(\beta) = \|y - X\beta\|_2^2 + \alpha\left(\lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2\right) \quad (1)$$

where X denotes an $M \times N$ data matrix, which consists of predictors. Each predictor consists of the values of the individual voxel activity levels. Vector y is the corresponding values for the mental state.

This equation clearly shows that EN becomes equivalent to LASSO [11, 16] when λ_2=0 and λ_1>0, and is equivalent to Ridge regression when λ_2>0 and λ_1=0. When λ_1=λ_2=0, the EN problem is simply reduced to OLS regression. Carroll et al. [14] underscored that λ_1 controls the sparsity of representation in the EN model, whereas λ_2 controls the robustness of the model. They reported robustness as the percent of the total number of unique voxels included in either model that appeared in both models. The EN model can coordinate λ_1 and λ_2 by creating the entire EN solution path. The algorithms for the EN model perform variable selection efficiently through computing EN regularization paths.

In this work, we study the EN model with different values of λ_1 and λ_2 for the visual image reconstruction experiments of Miyawaki [9] to analyze the sparsity of sparse modeling and visual representation. Friedman et al. [17] proposed fast algorithms for estimating generalized linear models with convex penalties. We used these algorithms and a publicly available R package glmnet (http://www-stat.stanford.edu/~tibs/glmnet-matlab/) to solve the EN model.

3. Experimental procedure

In preliminary scans, four subjects were screened for head motion, and two of them (male adults with normal visual acuity) whose head motions are the least, underwent the full experimental procedure. Informed consents were written by the two subjects, and the study was approved by the local Ethics Committee of the Chinese Academy of Sciences.

3.1. Visual stimulus and experimental design

Visual stimulus was projected onto a mirror placed in the scanner bore using a LCD projector. Three types of experimental sessions were conducted to measure the fMRI activities of the visual cortex similar as [9]: (1) a random image session, (2) a figure image session, and (3) a conventional retinotopy mapping session.

In the random image session, each run consisted of 30 stimulus blocks. Each stimulus block lasted 6s, followed by an intervening rest period lasted 6s. At the beginning (30 s) of each run there were extra rest periods. Fifty runs were repeated, and 1500 random patterns were presented to each subject. Random images were generated by a linear feedback shift register sequence.

In the figure image session, each run consisted of 15 stimulus blocks. Each stimulus block lasted 12s, followed by an intervening rest period lasted 12s. Extra rest periods were added as the same as in the random image session. Stimulus images consisted of flickering checkerboard patches as in the random image session as well, but they formed alphabet letters (“C”, “Y”, “V”, “N”, “I”, “L”, “O”, “A”, “r”, “D”, “E”, “J”, “F”, “T”, and “H”). In each run, five letters were presented and each image was repeated thrice. Each subject underwent 12 figure runs.

In these two sessions, subjects need to view the stimulus sequence while keeping fixation. The retinotopy mapping session, which is used to delineate the borders of visual cortical areas, followed the two sessions using a rotating wedge.

3.2. MRI acquisition and data preprocessing

MRI data for the experiment were all obtained using a 3.0-Tesla Siemens MAGNETOM Trio Tim scanner at the Henan Province Imaging Center. An interleaved T2*-weighted gradient-echo echo-planar imaging (EPI) scan was performed to acquire functional images to cover the entire occipital lobe (TR, 2000 ms; TE, 30 ms; flip angle, 80°; FOV, 220×220 mm; voxel size, 3.4×3.4×3 mm; slice gap, 0 mm; number of slices, 30).

The acquired fMRI data underwent head motion correction and spatial smoothing (FWHM=8 mm) by SPM8 (http://www.fil.ion.ucl.ac.uk/spm). Data from the retinotopy session were used to make a visual cortical mask. After removing voxels of extremely low signal amplitudes, 2534 and 3058 voxels were chosen in the visual cortex for subject 1 (S1) and subject 2 (S2), respectively.

The fMRI data were then subjected to linear trend removal within each run. Response amplitudes of individual voxels were normalized relative to the mean amplitude of the first 20s rest period to minimize the baseline difference across runs. To compensate for hemodynamic delays, the fMRI data for each voxel were shifting by 4s and then averaged within each stimulus block.

4. Results

We formulate the reconstruction task as a regression problem, and then apply EN model to the data. The matrix X in Eq. (1) indicates for each voxel, the average fMRI signal intensity over each stimulus block, while the vector y is set to 1 and -1 for appearance and non-appearance of the visual stimulus, respectively.

The training data sets consisted of 1500 random patterns for both S1 and S2, whereas the testing data sets consisted of 180 alphabet letters for both subjects. The EN model is used for reconstructing the letters. In the training phase, we firstly fix the value of the parameter $\lambda_1=0.05i$ ($i=0,1,2,\ldots,20$) for each local decoder, and use twofold cross-validation to get an optimal value of the parameter α, and then get an optimal value of the parameter λ_1 using the twofold cross-validation again. This training method probably results in some overestimation of predictive performance. Without loss of generality, we set $\lambda_1+\lambda_2=1$. The dimensionality of the matrices X is 750×2534 and 750×3058 for S1 and S2, respectively.

4.1. Sparsity and accuracy rate of EN model

We define the sparsity of the weight vector β by $S=\frac{\text{the number of zeros in } \beta}{\text{the dimension of } \beta}\times 100\%$. To study the relationship between the sparsity S and the parameter λ_1, we fix a same value of λ_1 for each local decoder herein for convenience, and use twofold cross-validation to get an optimal value of the parameter α, in the training phase. Sparsity S increases as λ_1 increases for both subjects.

Table 1

Reconstruction results for three models for S1

models	OLS	LASSO	EN
sparsity	40.84%	89.19%	85.39%
average accuracy rate (test data)	53.80%	69.33%	71.08%
sparsest decoder number	0	100	7

Table 2

Reconstruction results for three models for S2

models	OLS	LASSO	EN
sparsity	50.98%	91.86%	88.69%
average accuracy rate (test data)	52.97%	70.6%	72.08%
sparsest decoder number	0	100	5

Table 3

Average accuracy rates of reconstructed alphabet letters for S1

alphabets	C	Y	V	N	I
average accuracy rate	63.67%	73.42%	69.50%	66.25%	78.33%
alphabets	L	O	A	r	D
average accuracy rate	70.58%	63.50%	68.75%	66.08%	67.67%
alphabets	E	J	F	T	H
average accuracy rate	75.08%	78.08%	72.75%	79.17%	73.5%

Table 4

Average accuracy rates of reconstructed alphabet letters for S2

alphabets	C	Y	V	N	I
average accuracy rate	67.75%	76.08%	74.17%	69.92%	72.75%
alphabets	L	O	A	r	D
average accuracy rate	74.25%	70.17%	67.75%	68.44%	66.42%
alphabets	E	J	F	T	H
average accuracy rate	69.58%	75.33%	69.67%	80.08%	79.92%

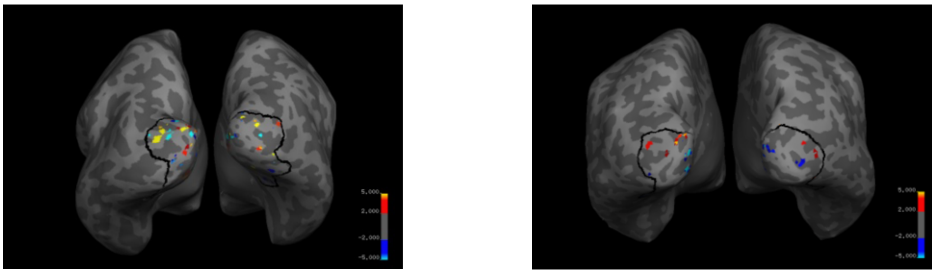
For comparison, the OLS model was explored, and the LASSO model was also used for local decoder training by setting λ_1=1 and λ_2=0 in the EN model. The results for three models are shown in Table 1 for S1 and Table 2 for S2. Accuracy rate is defined as follows: $accuracy\ rate = \left(1 - \frac{the\ number\ of error\ local\ decoders}{the\ total\ number\ of\ all\ local\ decoders}\right) \times 100\%$. Sparsest decoder number denotes the number of local decoder that was trained under the parameter λ_1=1. As is clearly evident in Table 1 and Table 2, the sparsity and the reconstruction average accuracy rates of LASSO and EN models with different values of λ_1 are much higher than that of the OLS model. It is clear that the sparsity and l_1-norm regularization is paramount for improving model performance. Furthermore in the EN model, the optimal value of λ_1 for almost local decoder is not 1. The reconstruction average accuracy rate of EN model is higher than LASSO model, while the sparsity of EN model is lower than LASSO. These imply visual reconstruction model needs sparsity that differs from the sparsest one, and l_2-norm regularization helps model performance.

4.2. Reconstruction results with EN model

This section provides details of the reconstruction results with EN models, which are showed in Table 1 and Table 2 for two subjects. For S1, the average accuracy rate of the 180-alphabet stimulus is 71.08%. The average accuracy rates of all reconstructed alphabet letters are listed in Table 3. For S2, the average accuracy rate of the 180-alphabet stimulus is 72.08%. The average accuracy rates of all reconstructed alphabet letters are listed in Table 4.

4.3. Sparsity of distributed patterns of visual reconstruction

The distributed patterns of the weight vector β of one particular local decoder in the visual areas (outlined by solid black lines) are shown in Figure 1 for S1 and S2. For S1, the sparsity of β is 85.44%, and the reconstruction accuracy rate of the local decoder is 69.44%. The corresponding values for S2 are 87.97% and 71.11%. The sparsity and the reconstruction accuracy rate for both decoders are close to the average level of an individual subject. As noted above, the locations of the selected voxels are



(a) Subject 1

(b) Subject 2

Fig. 1. Distributed pattern of the weight in the visual cortex for S1 and S2.

consistent with the conventional retinotopy mapping. Figure 1 also illustrates that important voxels are widely distributed in the visual cortex. This property of the distributed patterns implies that multi-voxel pattern analysis with sparse learning model is more suitable. Furthermore, true neural response is marked by distributed patterns of localized clusters of activity in the Figures, not the sparsest case, which has only one predictor in each cluster.

5. Discussion

The reconstruction results reveal that sparse learning models are suitable for the machine learning problem in visual reconstruction. It can be figured out that the l_1-norm regularization can be interpreted as a sparsity that enforcing prior over the model parameters that shrinks some parameters to zero and leading to a sparse model structure. Moreover, it is helpful to explore the sparsity of the contribution of each voxel to every classifier, thereby ensuring that the learning results have higher classification accuracy. Furthermore, the experimental results show that different responses arise from different subjects, different alphabet stimulus, or different time. Learning models should tolerate these tiny differences of responses for the same class of data. The l_2-norm regularization tends to leave strongly correlated features either in or out of the model, whereas a model with only l_1-norm regularization tends to select only one such correlated feature [15]. Compared with the LASSO model, the reconstruction results of the EN model exhibited a higher accuracy rate. These results are due to the l_2-norm regularization introduced in the EN model, which improves not only the robustness of the model but also the generalization performance of the learning results.

The results demonstrate that compare for OLS model the sparsity and the reconstruction average accuracy rates of models with l_1-norm regularization are much higher, which means that the sparsity of the model plays a key role in its performance. Furthermore, these results highlight the sparsity of visual representation. The LASSO model has the sparsest solution but does not exhibit the best performance for visual reconstruction, indicating that sparsity is not the only objective of visual reconstruction models. So models need sparsity that differs from the sparsest one for visual reconstruction process. The combination of l_1 and l_2 can balance sparsity and energy.

We find that training time will decrease as λ_1 increases. This result demonstrates that higher sparsity could expedite the training. However, several researchers have observed that a larger λ_1 tends to cause slower convergence when λ_1 is large enough that the solution of the EN model is also the solution of the LASSO model. Thus, we should choose an appropriate λ_1 to ensure that the solution of the EN model differs from that of the LASSO model.

6. Conclusion

Our results highlight the sparsity of visual representation. The EN model exhibits better performance for visual image reconstruction. As researches have suggested that the entries of the solution for the EN model corresponding to the strongly correlated columns of X in Eq. (1) tend to be close and that this may affect the sparsity of the solution.

Acknowledgement

This work was supported by the National High Technology Research and Development Program of China (863 Subject No. 2012AA011603), and the National Natural Science Foundation of China (No. 61372172).

References

[1] J.V. Haxby, M.I. Gobbini, M.L. Furey et al., Distributed and overlapping representations of faces and objects in ventral temporal cortex, Science **293** (2001), 2425–2430.

[2] D.D. Cox and R.L. Savoy, Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex, Neuroimage **19** (2003), 261–270.

[3] Y. Kamitani and F. Tong, Decoding the visual and subjective contents of the human brain, Nat. Neurosci. **8** (2005), 679–685.

[4] Y. Kamitani and F. Tong, Decoding seen and attended motion directions from activity in the human visual cortex, Curr. Biol. **16** (2006), 1096–1102.

[5] K.N. Kay, T. Naselaris, R.J. Prenger et al., Identifying natural images from human brain activity, Nature **452** (2008), 352–355.

[6] T. Naselaris, R.J. Prenger, K.N. Kay et al., Bayesian reconstruction of natural images from human brain activity, Neuron **63** (2009), 902–915.

[7] G.B. Stanley, F. Li and Y. Dan, Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus, J. Neurosci. **19** (1999), 8036–8042.

[8] B. Thirion, E. Duchesnay, E. Hubbard et al., Inverse retinotopy: inferring the visual content of images from brain activation patterns, Neuroimage **33** (2006), 1104–1116.

[9] Y. Miyawaki, H. Uchida, O. Yamashita et al., Visual image reconstruction from human brain activity using a combination of multiscale local image decoders, Neuron **60** (2008), 915–929.

[10] O. Yamashita, M.A. Sato, T. Yoshioka et al., Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns, Neuroimage **42** (2008), 1414–1429.

[11] R. Tibshirani, Regression shrinkage and selection via the Lasso, J. R. Stat. Soc. Ser. B **58** (1996), 267–288.

[12] T. Mitchell, R. Hutchinson, R. Niculescu et al., Learning to decode cognitive states from brain images, Mach. Learn. **57** (2004), 145–175.

[13] S. Ryali, K. Supekar, D.A. Abrams et al., Sparse logistic regression for whole-brain classification of fMRI data, Neuroimage **51** (2010), 752–764.

[14] M.K. Carroll, G.A. Cecchi, I. Rish et al., Prediction and interpretation of distributed neural activity with sparse models, Neuroimage **44** (2009), 112–122.

[15] H. Zou and T. Hastie, Regularization and variable selection via the Elastic Net, J. R. Stat. Soc. Ser. B **67** (2005), 301–320.

[16] D. Gosh and A.M. Chinnaiyan, Classification and selection of biomarkers in genomic data using LASSO, Journal of Biomedicine and Biotechnology **2005** (2005), 147–154.

[17] J. Friedman, T. Hastie and R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, Journal of Statistical Software **33** (2010), 1–22.