A Markov Random Field Approach to Neural Encoding and Decoding

Marcel A.J. van Gerven[1,2], Eric Maris[1], and Tom Heskes[1,2]

[1] Donders Institute for Brain, Cognition and Behaviour
[2] Institute for Computing and Information Sciences
Radboud University Nijmegen, Nijmegen, The Netherlands
m.vangerven@donders.ru.nl, e.maris@donders.ru.nl, t.heskes@science.ru.nl

Abstract. We introduce a new approach to neural encoding and decoding which makes use of sparse regression and Markov random fields. We show that interesting response functions were estimated from neuroimaging data acquired while a subject was watching checkerboard patterns and geometrical figures. Furthermore, we demonstrate that reconstructions of the original stimuli can be generated by loopy belief propagation in a Markov random field.

Keywords: Encoding, decoding, sparse regression, Markov random field.

1 Introduction

Neural encoding and decoding are two topics which are of key importance in cognitive neuroscience. Neural encoding refers to the representation of certain stimulus features by particular neuronal populations as reflected in measured brain activity. Conversely, neural decoding refers to the prediction of such stimulus features from measured brain activity. Encoding is a classical topic in (cognitive) neuroscience and can be tackled by reverse correlation methods [1]. Decoding has gained much recent popularity with the adoption of multivariate analysis methods by the cognitive neuroscience community [2]. While the first decoding studies have focused mainly on the prediction of discrete states such as stimulus orientation [3] or object category [2], more recent decoding studies have focused on the prediction of more complex stimulus properties, culminating in the reconstruction of the contents of visual scenes [4–6].

In this paper, we tackle the encoding problem using elastic net linear regression and show how the resulting parameter estimates can be used to solve the decoding problem. This is achieved by incorporating the estimated regression coefficients within a pairwise Markov random field (MRF) and by using loopy belief propagation [7] to approximate its MAP solution. We apply our methods to a neuroimaging dataset that has previously been used in [5] where subjects have been shown checkerboard patterns and simple geometrical figures. We analyzed the encoding models that have been learned by the employed sparse regression method and find interesting response functions. Furthermore, it is shown that a suitable stimulus prior leads to better stimulus reconstructions.

T. Honkela et al. (Eds.): ICANN 2011, Part II, LNCS 6792, pp. 1–8, 2011.

2 Methods

Let (s, r) denote a stimulus-response pair, say, an image $s = (s_1, \ldots, s_I)^\top$, characterized by pixel values s_i, and its associated measured response vector $r = (r_1, \ldots, r_K)$. The stimulus can be either discrete or continuous and the response is typically continuous, e.g., the BOLD response in multiple voxels.

2.1 Encoding

In an encoding analysis, both stimulus and response are observed and we are only interested in estimating the parameters $\hat{\theta}$ of the encoding distribution $p(r|s)$ given i.i.d. data $D = \{(s^n, r^n)\}_{n=1}^N$. This can be realized by taking the MAP estimate

$$\hat{\theta} = \arg\min_\theta \left\{ -\log p(\theta) - \sum_n \log p(r^n|s^n, \theta) \right\} . \quad (1)$$

We assume that the individual responses r_k are conditionally independent and given by a linear function of s with additive Gaussian noise, such that $r_k = \alpha_k + \beta_k^\top s + \epsilon_k$ where α_k is an offset term, β_k is a vector of regression coefficients for response k and ϵ_k is a zero mean Gaussian random variable with variance σ_k^2. Thus, we have $p(r|s, \theta) = \mathcal{N}\left(r; \alpha + B^\top s, \Sigma\right)$ with $\theta = (\alpha, B, \Sigma)$ where $\alpha \equiv (\alpha_1, \ldots, \alpha_K)^\top$, $B \equiv (\beta_1, \ldots, \beta_K)$ and $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_K^2)$. Also assuming that $p(\theta)$ factorizes accordingly, we obtain a set of minimization problems of the form

$$\hat{\theta}_k = \arg\min_{\alpha_k, \beta_k, \sigma_k^2} \left\{ -\log p(\alpha_k, \beta_k, \sigma_k^2) - \sum_n \log \mathcal{N}\left(r_k^n; \alpha_k + \beta_k^\top s^n, \sigma_k^2\right) \right\} . \quad (2)$$

We use the prior to express indifference about σ_k^2 as well as our preference for sparse regression vectors β_k containing just a small number of non-zero elements. This can be realized by choosing $-\log p(\alpha_k, \beta_k, \sigma_k^2) \propto R_{\lambda,\tau}(\beta_k)$ where

$$R_{\lambda,\tau}(\beta_k) = \lambda \sum_{k=1}^K \{(1-\tau)\frac{1}{2}||\beta_k||_2^2 + \tau||\beta_k||_1\} \quad (3)$$

is the elastic net regularizer [8]. The parameter λ determines the amount of regularization and τ determines the balance between L_1 and L_2 regularization. Let $S_k(a, b) = \sum_n \left(r_k^n - a - b^\top s_k^n\right)^2$ denote the sum of squares error function. Minimization of Eq. (2) with respect to (α_k, β_k) then boils down to computing $(\hat{\alpha}_k, \hat{\beta}_k) = \arg\min_{\alpha_k, \beta_k} \frac{1}{2} S_k(\alpha_k, \beta_k) + R_{\lambda,\tau}(\beta_k)$ which can be achieved using an efficient coordinate gradient descent algorithm [9]. Minimization of Eq. (2) with respect to the variance yields $\hat{\sigma}_k^2 = \frac{1}{N} S_k(\hat{\alpha}_k, \hat{\beta}_k)$.

We use the elastic net algorithm to test how the responses are predicted by a small number of stimulus features. In our problem setting this amounts to probing how the responses r_k are encoded by which pixels via sparse vectors β_k. We refer to β_k as the response function of r_k. Encoding performance is quantified in terms of the coefficient of determination $R_k^2 = 1 - S_k(\hat{\alpha}_k, \hat{\beta}_k) / \sum_n \left(r_k^n - \bar{r}_k\right)^2$ where $\bar{r}_k$ is the mean of the observed response data. We use R^2 to denote the average over all responses.

2.2 Decoding

Once the parameters (α, B, Σ) are estimated they may be used for decoding. That is, they are used to approximate the most likely configuration

$$s^* = \arg\max_s \{p(r|s)p(s)\} \tag{4}$$

given an observed response r. In the following, we show how s^* can be approximated by means of inference in a MRF. We start by assuming that the prior can be specified in terms of a MRF

$$p(s) = \frac{1}{Z}\prod_i \phi_i(s_i)\prod_{i\sim j}\phi_{i,j}(s_i, s_j) \tag{5}$$

where $\phi(s_i)$ and $\phi(s_i, s_j)$ are unary and pairwise potential functions, Z is the partition function and $i \sim j$ denotes (unordered) pairs (i, j) that are neighbors in some undirected graph G. The resulting model, together with the encoding of the responses r, is shown in Fig. 1.A.

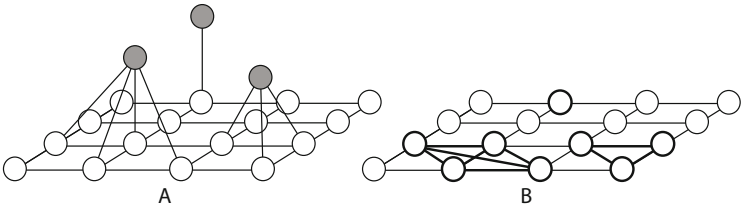

Fig. 1. Panel A: The correlation between stimulus elements s_i, represented by white discs, is given by a MRF. Measured responses r_k, represented by grey discs, are each assumed to be determined by a few stimulus elements. Panel B: Estimated regression coefficients can be absorbed into the MRF yielding a new, typically more dense, MRF. Fat lines indicate which node and edge potentials have changed due to this procedure.

Since $p(r|s) \propto \exp\left(-\frac{1}{2}r^\top\Sigma^{-1}r + r^\top\Sigma^{-1}\mu - \frac{1}{2}\mu^\top\Sigma^{-1}\mu\right)$, given a fixed response r, we can drop terms not depending on s and obtain

$$\begin{aligned} p(r \mid s) &\propto \exp\left(r^\top\Sigma^{-1}(\alpha + B^\top s) - \frac{1}{2}(\alpha + B^\top s)^\top\Sigma^{-1}(\alpha + B^\top s)\right) \\ &\propto \exp\left(r^\top\Sigma^{-1}B^\top s - \alpha^\top\Sigma^{-1}B^\top s - \frac{1}{2}s^\top B\Sigma^{-1}B^\top s\right) \\ &= \exp\left(c^\top s + s^\top U s\right) \end{aligned} \tag{6}$$

with $c^\top \equiv (r - \alpha)^\top\Sigma^{-1}B^\top$ and $U = -\frac{1}{2}B\Sigma^{-1}B^\top$. Equation (6) may also be written as a pairwise MRF $p(r|s) = \frac{1}{Z}\prod_i \psi_i(s_i)\prod_{i\sim j}\psi_{i,j}(s_i, s_j)$ where $\psi_i(s_i) = \exp(s_i\sum_k \frac{\beta_{ki}}{\sigma_k^2}(r_k - \alpha_k - \frac{1}{2}\beta_{ki}))$ encodes local information about the probability

of observing pixel i in state s_i and $\psi_{i,j}(s_i, s_j) = \exp(-s_i s_j \sum_k \frac{\beta_{ki}}{\sigma_k^2}\beta_{kj})$ couples pixels s_i and s_j whenever there exists a response r_k that is encoded (in part) by s_i and s_j (see Fig. 1.B). Since the prior is expressed in the same form, the decoding problem can be solved by approximating the mode of a pairwise Markov random field $p(s|r) = \frac{1}{Z}\prod_i(\phi_i(s_i)\psi_i(s_i))\prod_{i\sim j}(\phi_{i,j}(s_i, s_j)\psi_{i,j}(s_i, s_j))$.

In order to decode the most likely stimulus given a response, we require the maximizer s^* of $p(s|r)$. In case of discrete stimuli, as used in this paper, we have no analytical expression for the mode and we need to rely on approximate inference methods in order to estimate it. We use loopy belief propagation as implemented in the UGM toolbox[1] to compute approximate marginals $q(s_i)$. These marginals are used to approximate the MAP estimate as $s^* \approx \arg\max_s \prod_i q(s_i)$. In order to quantify decoding performance, we use the Manhattan distance between the real stimulus s and its reconstruction s^*, separately averaged over N_0 inactive pixels and N_1 active pixels in order to penalize reconstruction errors of figure or background equally: $M(s, s^*) = \frac{1}{2}\sum_{j=0}^{1}\frac{1}{N_j}\sum_{i:\ s_i=j}|s_i - s_i^*|$.

3 Experiments

3.1 Stimuli, Acquisition and Data Analysis

We are interested in the encoding and decoding of neural responses to simple visual stimuli. We make use of neuroimaging data for one subject which has been made available by the authors[2] and was used in [5]. Neuroimaging data was acquired while a subject was viewing checkerboard patterns or simple geometrical figures on a 10×10 grid. For the checkerboard patterns, 440 trials were presented which each lasted six seconds. For the geometrical figures, 120 trials were presented which each lasted twelve seconds. Geometrical figures consisted of six repetitions of twenty different patterns. Functional images were acquired with a 3T MRI scanner using an interleaved T2*-weighted gradient-echo echo-planar imaging scan which covered the entire occipital lobe (TR, 2000 ms; TE, 30 ms; flip angle, 80°; FOV, 192 × 192 mm; voxel size, 3 ×3 × 3 mm; slice gap, 0 mm; number of slices, 30). Additionally, functional localizer scans were used to delineate the borders between visual cortical areas. Functional images were slice-timing corrected, motion corrected, coregistered with a high-resolution structural scan and reinterpolated to $3 \times 3 \times 3$ mm using SPM2 software. Data was linearly detrended, shifted by 3 volumes to take the HRF lag into account and standardized such that the BOLD response in each voxel had zero mean and unit variance. Finally, data was averaged over three consecutive volumes for the checkerboard patterns and six consecutive volumes for the geometrical figures.

3.2 Encoding Analysis

In the encoding analysis we were interested in examining how pixels within the 10×10 grid encode the BOLD response of individual voxels in visual areas

[1] http://www.cs.ubc.ca/∼schmidtm/Software/UGM
[2] http://www.cns.atr.jp/∼yoichi_m

V1, V2 and V3. To this end, we made use of the elastic net algorithm with $\tau = 0.99$ while varying λ. Encoding performance was computed by training on randomly selected trials (75% of the checkerboard pattern data) and testing on the remaining trials. The value of λ was selected by computing performance as a function of λ using an inner cross-validation and taking the λ with maximal performance to produce the results on the test data. The same procedure was followed using the geometric figure data. This allowed us to compare encoding performance and response functions between the two datasets.

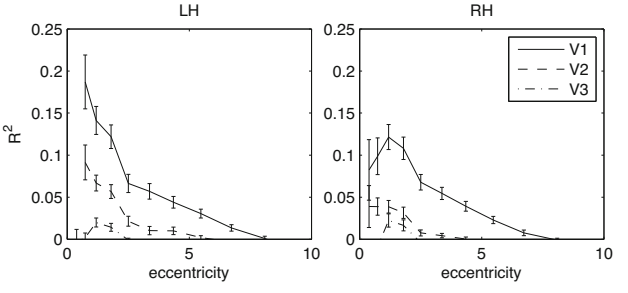

Fig. 2. Average R^2 as a function of visual field eccentricity for voxels in left (LH) and right (RH) hemispheres. Errorbars denote standard error of the mean.

Figure 2 shows average encoding performance of voxels in left and right hemisphere visual cortex while subjects were looking at checkerboard patterns. Voxels are ordered according to their visual field eccentricity as determined by the functional localizer. Maximal encoding performance ranged up to $R^2 = 0.75$, $R^2 = 0.53$ and $R^2 = 0.40$ for voxels that are exclusively assigned to area V1, V2 or V3, respectively. Encoding results degraded when moving to higher visual areas or to areas that code for pixels in the periphery. Spearman rank correlations between the R^2 values for responses to checkerboard patterns and geometric figures were $r = 0.35$, $r = 0.17$ and $r = 0.06$ for areas V1, V2 and V3, respectively.

Figure 3 depicts some of the properties of the estimated response functions. Figure 3.A shows the distribution of the number of pixels used in the response functions of voxels in primary visual areas when using either the checkerboard or the geometric data. Most voxels use no pixels whatsoever, indicating that their BOLD response to visual input is negligible or could not be properly detected. Most of the voxels which do show a response use just one pixel in the encoding. Some voxels, however, are dependent on a large number of pixels. Interestingly, the graphs for the geometric data have heavier tails than those for the checkerboard data. Hence, on average, more pixels are used when voxels are trained on structured images. We examined what the response functions look like for voxels that show large R^2 values for the checkerboard patterns and small R^2 values for the geometric figures (Fig. 3.B) and vice versa (Fig. 3.C). Figure 3.B shows that more diffuse encoding models are found using the geometric figure data for voxels that perform well on checkerboard pattern data. If we examine those voxels in Fig. 3.C which show good performance on the geometric figure

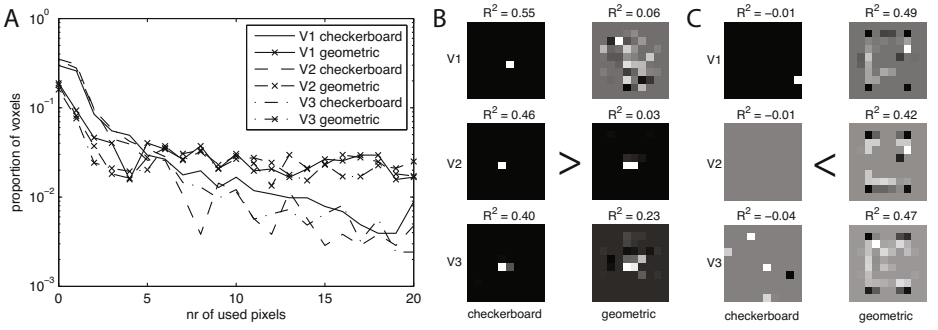

Fig. 3. Panel A: Distribution of the number of used pixels to encode the response for checkerboard and geometric data. Panel B: Examples of response functions $\hat{\beta}_k$ for voxels that show good encoding performance for the checkerboard patterns and bad performance for the geometric figures. Panel C: Examples of response functions $\hat{\beta}_k$ for voxels that show good encoding performance for the geometric figures and bad performance for the checkerboard patterns.

data and not on the checkerboard pattern data, we find quite striking response functions, resembling the presented geometric figures.

3.3 Decoding Analysis

In the decoding analysis, we were interested in reconstructing the stimuli which were most likely to have caused the observed responses by computing the most likely state of a pairwise MRF. We used the checkerboard pattern data to estimate the encoding distributions (training data) and to compute the residual variances (test data). Subsequently, we applied the constructed MRF to decode the geometric figure data. The rationale here is that we wanted to examine whether it is possible to build a generic decoder that is trained on random data. The decoding was achieved using voxels in areas V1, V2 and V3 which were sorted according to their R^2 values in decreasing order. Figure 4. A shows how

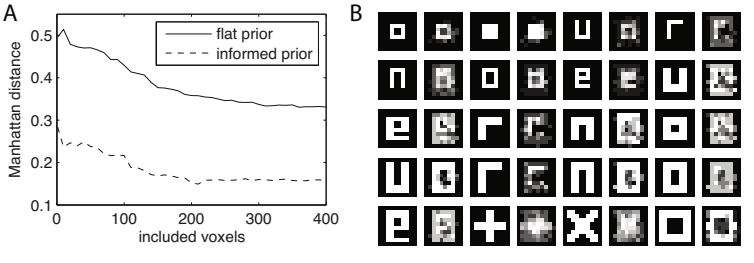

Fig. 4. Panel A shows the decrease in Manhattan distance between a stimulus and its reconstruction averaged over all trials as a function of the number of included voxels. Panel B shows the average reconstruction for each of the twenty structured patterns.

the distance between the stimuli and their reconstructions decreases as a function of the number of included voxels using a flat prior. A minimal average error of 0.33 was obtained by including the responses of 410 voxels.

In order to examine the effect of the prior, we have also constructed an informed prior. Specifically, we created a fully connected pairwise MRF and used the geometric figure data to estimate the MRF parameters. Figure 4.A shows that the structured prior gives much better decoding performance with a minimal error of 0.15 using 210 voxel responses. Figure 4.B depicts the average reconstructions obtained under this regime.

4 Conclusions

In this paper, we have shown that the elastic net model is capable of generating good predictions of the BOLD response as induced by the presentation of checkerboard patterns and geometric figures. The best predictions for checkerboard patterns were obtained by using single pixels in the central visual field. For the geometric figures a subset of the responsive voxels have complex response functions. By absorbing the estimated regression coefficients in a MRF over the stimulus pixels we were able to decode the presented stimuli to some degree.

The encoding results show that the response functions of certain voxels are quite complex (Fig. 3.C). Note however that the geometrical figures on which the encoding models were trained consisted of twenty figures which were repeated six times each. Furthermore, these patterns show strong correlations between pixels. This may lead one to believe that the response functions are just an artifact of these correlations and the voxels are in reality just sensitive to a small number of pixels. Note, however, that if this were the case, then (a) these voxels should also show good performance on the checkerboard patterns and (b) the sparseness constraint of the employed elastic net model would favor response functions which rely on a small number of voxels, even in the presence of strong correlations. In conclusion, we are led to believe that some voxels are truly responsive to complex inputs, although additional analyses are warranted.

Obtained reconstructions show that our MRF approach to decoding is feasible and the inclusion of an informed prior is shown to lead to better decoding performance as compared to the use of a flat prior. Still, single trial decoding results remain quite noisy, which can be due to various reasons. One important reason is that the encoding results actually show that the responses of voxels to checkerboard patterns and geometric figures are only weakly correlated. Hence, the encoding models learned from checkerboard pattern data will not generalize perfectly to the geometric figure data. Another observation is that the decoding results for the informed prior, shown in Fig. 4.B, are biased towards the reconstruction of 'O' shapes. This is due to the fact that many of the geometric figures share features with these shapes. This is taken into account by the prior and will bias the reconstructions towards such shapes. Decoding is also influenced by the employed approximate inference method. We used loopy belief propagation and estimated the MAP solution as a max product over the marginals. More sophisticated inference methods may further improve decoding performance.

Various extensions of the framework introduced in this paper are possible. One such extension is to explicitly incorporate the hemodynamic response within the model instead of collapsing over a number of successive measured volumes. Encoding could be improved by including latent variables that explicitly represent the complex features to which particular voxels are sensitive (cf. [10]). Finally, in this paper, we used Bayes' rule to decode using a (Markov random field) model that has been optimized for encoding. An alternative could be to train or at least fine-tune the model's parameters using discriminative training, specifically geared towards improving the decoding performance. We expect this to lead to better reconstructions, possibly at the expense of encoding.

Acknowledgements. The authors gratefully acknowledge the support of the Netherlands Organization for Scientific Research NWO (Vici grant 639.023.604) and the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science.

References

1. Ringach, D., Shapley, R.: Reverse correlation in neurophysiology. Cogn. Sci. 28, 147–166 (2004)
2. Haxby, J., Gobbini, M., Furey, M., Ishai, A., Schouten, J., Pietrini, P.: Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430 (2001)
3. Kamitani, Y., Tong, F.: Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685 (2005)
4. Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J., Lebihan, D., Dehaene, S.: Inverse retinotopy: inferring the visual content of images from brain activation patterns. NeuroImage 33, 1104–1116 (2006)
5. Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M., Morito, Y., Tanabe, H.C., Sadato, N., Kamitani, Y.: Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron. 60, 915–929 (2008)
6. Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M., Gallant, J.L.: Bayesian reconstruction of natural images from human brain activity. Neuron. 63, 902–915 (2009)
7. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Francisco (1988)
8. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. Roy. Stat. Soc. Series B 67, 301–320 (2005)
9. Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1–22 (2010)
10. van Gerven, M.A.J., de Lange, F.P., Heskes, T.: Neural decoding with hierarchical generative models. Neural Comput. 22, 1–16 (2010)