DTU

Methods, Physiology, Cognition # 179

Artificial Neural Network Model for fMRI timeseries and a Framework for Comparison of Convolution Models

Finn Årup Nielsen, Cyril Goutte and Lars Kai Hansen

Department of Mathematical Modelling, Technical University of Denmark

International Neuroimaging Consortium

INC

Abstract

- fMRI-response represented by a nonlinear filter implemented by an artificial neural network. Previous work [8, 11, 7].
- Comparison between convolution models (Poisson, FIR) done within the generalization framework.

Convolution models

Our approach models the fMRI-response as a nonlinear time-invariant (stationary) system. Output of the model with the paradigm (activation) as $\mathbf{x}_{u,n}$ and the model parameters as $\boldsymbol{\theta}_u$:

$$y_{u,n} = f(\mathbf{x}_{u,n}; \boldsymbol{\theta}_u)$$

We assume a Gaussian noise model between the model output and the fMRI-signal:

$$p(t|\mathbf{x};\boldsymbol{\theta}) = \mathcal{N}(y, \sigma^2)$$

Our *cost function* is the negative log-likelihood:

$$-\log p(t|\mathbf{x};\boldsymbol{\theta})$$

n Sample (or scan) index

u Voxel index

$\mathbf{x}$ Input, the activation (paradigm)

y Predicted output

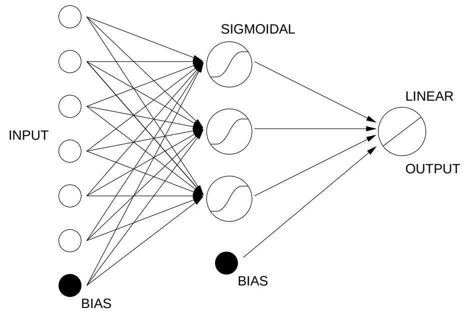
t Target, the measured fMRI time series

Nonlinear models

		Parameters	
		linear	nonlinear
Input	linear	FIR	Gamma, Poisson
	nonlinear	Polynomial	Neural net

Models can be nonlinear w.r.t the cost function in two senses: The parameters in the model can be nonlinear, e.g., the Poisson parameter in the Poisson model [2] is nonlinear, or the input can be nonlinear, e.g., in a polynomial model.

Neural network



Two-layer feed-forward neural network:

$$\mathbf{y} = [\tanh([\mathbf{X}, \mathbf{1}]\mathbf{V}), \mathbf{1}]\,\mathbf{w} \qquad (1)$$

Parameters of the network $\boldsymbol{\theta}$ are $\mathbf{V}$ and $\mathbf{w}$.

Comparison

In [3] the significance of the nonlinearity is assessed by evaluating the fit when the linear part of the model is treated as a confound.

Here we stress instead the use of an independent *test set*. When the fit is done on a separate training set the estimate of the performance in the test set provides an unbiased estimate of the *generalization*.

The *generalization* can be derived from a probabilistic framework by the following definition of the generalization:

$$\Gamma = -\int \log p(t|\mathbf{x};\boldsymbol{\theta})\; p(\mathbf{x},t)\; d\mathbf{x}\; dt$$

Γ can be estimated from a test set:

$$\hat{\Gamma} = -\frac{1}{N}\sum \log p(t|\mathbf{x};\boldsymbol{\theta})$$

For a Gaussian noise model this is proportional to the simple squared error ($\mathcal{T}$ = test set):

$$\hat{\Gamma} \propto \sum_{n\in\mathcal{T}} (y_n - t_n)^2$$

Although there is no bias on the generalization estimate there is variance, due to the finite size of the test set.

This framework is well-known in pattern-recognition [1, 5].

Learning curve

Within the generalization framework we can also evaluate the performance of the model as the number of scans used to fit the model is varied.

The decrease in the generalization error as the number of scans is increased indicates whether it would be advisable to obtain new scans to improve the fit or if the number of scans aquired so far is adequate.

In this case we would usually be interested in the *average* performance over *all possible* training set, i.e. experiments, of a certain size. This adds an extra source of variance to the generalization estimate.

Artificial neural network techniques

We employ standard neural network techniques, see e.g. [9]:

- Preprocessing of input: Singular value decomposition. (PCR)
- Nonlinear optimization: gradient descent, Levenberg-Marquardt, ...
- Architecture optimization: Regularization by *weight decay* (ridge regression) and pruning by *Optimal Brain Damage*.
- Generalization assessment for architecture. Control the bias-variance trade-off between a simple and a complex model, — between underfitting and overfitting: Empirical on a validation set or analytically, e.g., Akaike's FPE.

Analysis

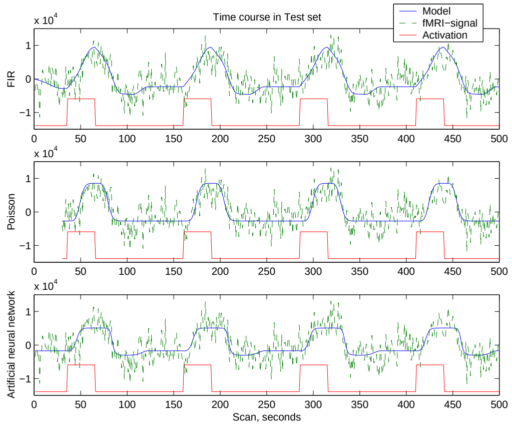
We analyzed two different fMRI data sets, with the neural network, a FIR model [10] and a Poisson convolution model [2].
We used singular value decomposition as preprocessing for the neural network and the FIR model. This gives the model the ability to have a large lag while keeping the number of parameters down.
All the models were run with the `lyngby` Matlab toolbox [6].
The hyperparameters in the FIR model (lag, ridge parameter) and the neural network (pruned architecture) were choosen with a single validation set.

Data

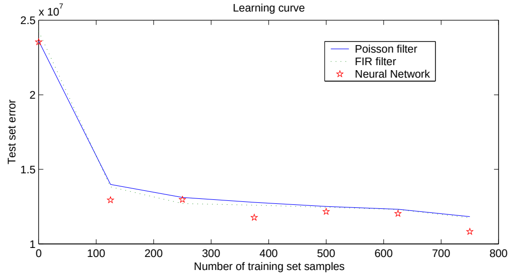
Data : MGH (Robert Savoy)
Paradigm : Multiple runs of 72 images. 24 baseline, 24 activation, 24 baseline. Left-handed finger-to-thumb opposition.
Frequency : 0.4 images / second.
Location : Axial slice through primary motor cortex and SMA.
Voxel size : 3.1 x 3.1 x 8 mm.
Preprocessing : Masked, motion alignment, centering.

Data : Hvidovre Hospital (Egill Rostrup)
Stimulus : 10 runs of 150 images (60-30-60). Visual stimulus 8Hz full-field checkerboard.
Modality : 1.5T EPI (T2* weighted)
Frequency : 3 images / second
Location : Para-axial orientation parallel to calcarine sulcus.
Voxel size : 1.8 x 1.8 x 8 mm
Preprocessing : First 25 scans from each run removed (contained a transient), centering

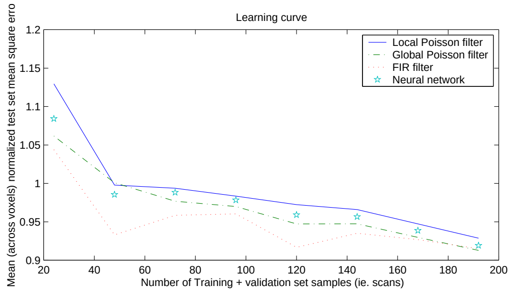
Results



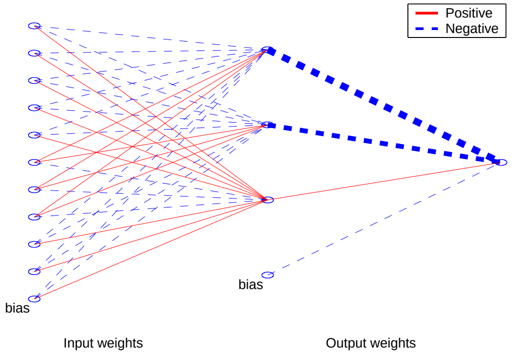
Time course for a single voxel in the visual cortex. (Hvidovre data)



Learning curve for an analysis with the Hvidovre data.



Learning curve for an analysis with the MGH data.



Example of an artificial neural network obtained for the MGH-data.

Discussion

- More flexible kernels enable better modeling of the fMRI-signal, e.g., they are able to model an asymmetry between the onset of the activation and the transition back to base line. This is in contrast to the linear models.
- Artificial neural network parameters can be difficult to interpret and postprocess, in e.g. feature clustering [4].
- The neural network optimization is slow. For a whole volume it is not feasible to optimize one neural network for each voxel. However, it is possible to train a single neural network for the whole volume or a region of interest and use a separate amplitude parameter for each voxel.
- We would expect that the neural network would perform significantly better than linear models with short stimuli: [13] used stimulations with 1–8 seconds duration to assess the nonlinearity in the visual cortex and found strong nonlinear effects for durations less than 4 seconds.
- Future work will include developing tests for statistical significance of the difference in model performance [12].

References

[1] Bishop, C. M., Neural Networks for Pattern Recognition, Clarendon Press Oxford, 1995.

[2] Friston, K. J., et al., Human Brain Mapping, 1994, 1:153–171.

[3] Friston, K. J., et al., Magnetic Resonance in Medicine, 1998, 39:41–52.

[4] Goutte, C., et al., NeuroImage, vol. 7(part 2), 1998 S610.

[5] Hansen, L. K., et al., NeuroImage 1999, 9(5):534–544.

[6] Hansen, L. K., et al., 5th Intl. Conf. Func. Map. Hum. Brain, 1999 S241.

[7] Knowles, A. J., et al., MRM 6th Annual Meeting, 1998 1477.

[8] Lange, N., et al., NeuroImage, vol. 3(part 2), 1996 S75.

[9] Mørch, N., et al., Information Processing in Medical Imaging, 1997 259–270.

[10] Nielsen, F. Å., et al., NeuroImage, vol. 5(part 2), 1997 S473.

[11] Pedersen, M. W., Optimization of Recurrent Neural Networks for Time Series Modeling, Ph.d. thesis, Tech. Univ. Denmark, 1997.

[12] Rasmussen, C. E., et al., The DELVE Manual, University fo Toronto, first edn. 1996.

[13] Vazquez, A. L., Noll, D. C., NeuroImage 1998, 7(2):108–118.

Acknowledgement

This project has been funded by The Danish Research Councils and the Human Brain Project P20 MH57180 “Spatial and Temporal Patterns in Functional Neuroimaging”. The data were acquired by Egill Rostrup at Hvidovre Hospital and Robert Savoy at MGH.