What makes a good multivariate model for fMRI-based decoding?

Bertrand Thirion,
INRIA Saclay-Île-de-France, Parietal team
http://parietal.saclay.inria.fr
bertrand.thirion@inria.fr
MVPA is an inference procedure

• Said to be more powerful than standard brain mapping experiments
  What does this mean?

• Generalization across protocols (transfer learning)
• Individual prediction, diagnosis problem
Outline

• Sample size issues in MVPA
• How to interpret pattern maps in MVPA?
• MVPA and functional specificity
Sample size & multivariate analysis

Classification
- SVC Ensemble Classifiers
- KNeighbors Classifier
- SGD Classifier
- Naive Bayes
- Text Data
- Linear SVC
- kernel approximation
- NOT WORKING

Clustering
- Spectral Clustering GMM
- KMeans
- MiniBatch KMeans
- MeanShift VBGMM
- NOT WORKING

Regression
- SGD Regressor
- ElasticNet Lasso
- SVR(kernel='rbf') Ensemble
- RidgeRegression SVR (kernel='linear')
- NOT WORKING

Dimensionality Reduction
- Randomized PCA
- Isomap Spectral Embedding
- LLE
- NOT WORKING

Predicting a category
- >50 samples
- <100K samples
- few features should be important

Predicting a quantity
- <100K samples
- NOT WORKING

Just looking
- <10K samples
- NOT WORKING

Predicting a structure
- tough luck
- NOT WORKING

Get more data
- >50 samples
- NOT WORKING

June 2017
Decoding cognitive information – Bertrand Thirion
Multivariate analysis
Learning curve: how prediction improves with n

- Predict the age of a subject given gray matter density maps (OASIS dataset, n=403)
The weight map depends on the batch of subject considered (bootstrap):
One question, different datasets, different answers

Variability actually worse than for univariate analysis!
Weight maps for age prediction / OASIS

The weight map depends on the batch of subject considered (bootstrap):
One question, different dataset, different answers
Weight maps for age prediction / OASIS

The weight map depends on the batch of subject considered (bootstrap):
One question, different dataset, different answers

Summarized into a z image:
(effect size) / (effect std)
Weight maps for age prediction / OASIS

(z = 5, n = 10, 20, 50)

(z = 0, n = 100, 200, 300)

(z = -5, n = 10, 20, 50)
Outline

- Sample size issues in MVPA
  - Some remarks on cross-validation
- How to interpret pattern maps in MVPA?
- MVPA and functional specificity
Statistical mapping vs decoding?

- “the increased sensitivity of multivariate analyses” [Haynes neuron 2015]

- Decoding patterns should not be interpreted as activation patterns [Haufe et al., img 2013]

- Decoding rejects only a global null hypothesis

- Localization with decoding is an ill-posed problem

- Decoding maps represent conditional evidence ≠ SPMs represent marginal evidence
The multivariate miracle

• Individual voxels corrupted by a noise source → weakly significant

• Their difference is strongly task related: MVPA is very sensitive

[Haufe et al. nimg 2013, Haynes neuron 2015]

→ The configuration where the noise has more correlation than the signal is unlikely in fMRI!

→ Global noise reduced with e.g. Compcorr [Behzadi et al. Nimg 2007]
The multivariate failure

- Consider the decoder as an estimator: good at finding brain regions?
- [Haufe et al. 2013]: no, due to the unmodeled noise covariance
- Problem: noise covariance not invertible / manageable in fMRI
The multivariate failure

- Consider the decoder as an estimator: good at finding brain regions?
- [Haufe et al. 2013]: no, due to the unmodeled noise covariance
- Problem: noise covariance not invertible / manageable in fMRI
Decoders behave however much differently whenever the signal is smoother than the noise.
• Spatial regularization + sparsity recover the pattern.
(Non-)identifiability of the model?

- When solving the inverse problem, you don't recover the true pattern but an approximation.
- True model: $y = Xw_0 + e$
- Estimated model: $\hat{w} = \arg\min_w \ell(w) + J(w)$
- $\hat{w}$ cannot be equal to $w_0$, as $X$ is non-invertible.
(Non-)identifiability of the model ?

- Can $\hat{w}$ have at least the correct support ?
  - $= \text{Non-zero voxel set}$

- No: the encoding model violates the conditions for accurate reconstruction with sparse model [Varoquaux et al. 2012]

- Better support recovery by introducing relevant priors on the decoder [Varoquaux et al. 2012]
  - Smoothness
  - Small variations
Why bother with full-brain decoding?

- Focus on pre-defined ROIs?
  - How do you define the ROI (position, size, boundaries)? Consistently across subjects?
- Consider all possible ROIs: searchlight
  - Computational cost
  - Correction for multiple comparisons
  - Miss long-range interactions
- Any other idea?
One question, one dataset, many answers

- What regions are involved in face perception?
  - [Haxby et al. Science 2001], subject 2
- Univariate models
  - 'face-rest' contrast
  - 'face-others' contrast:
    7 * face – (place + … + cat)
  - 'face > others' conjunction:
    (face – place) \( \land \ldots \land \)(face-cat)

\( p < .001 \) uncorrected
One question, one dataset, many answers

- MVPA approach
  - SVM classifier, Leave-2-session out cross-validation (12 sessions)
  - Discriminate face vs rest
  - Discriminate face vs others
One question, one dataset, many answers

- Can one conclude that the discriminating pattern is \textit{distributed}?
- No
Sparsity and interpretability

- 'face' versus 'rest' classifier: the pattern is ugly

- Why not thresholding it, like an SPM (z>3), and showing:
Sparsity and interpretability

- No!

is the pattern of another classifier, that may work or not

(here, it does)

Then, how do I do to get a map with blobs?

- Prior **feature selection** by univariate screening (fast and accurate)
  - decide how many features in advance → nested cross-validation
- **sparse** classifier (spacenet)
Sparsity and interpretability

Prior **feature selection**

**sparse** classifier

face vs others (svm, thresholded)

face vs others (univariate + svm)

face vs others, spacenet_tv-l1
A good model on a budget

Subsampling

Fitting with each hyperparameter

Select the best model per CV fold

Averaging (final model)

[Hoyos Idrobo et al. PRNI 2015 NeuroImage in Press]
Good model with a small budget

- Inter-subject settings: clustering step

Accurate and fast clustering-based data compression

[Thirion et al. 2015, Hoyos-Idrobo et al. IEEE PAMI in Press]
# A good model on a budget

<table>
<thead>
<tr>
<th>Model</th>
<th>Relative prediction score</th>
<th>Relative weight stability</th>
<th>Relative computation time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-10% -5% 0% 5% 10% 15%</td>
<td>-0.4 -0.2 0.0 0.2</td>
<td>0.4 $\frac{1}{16}$ x</td>
</tr>
<tr>
<td>Graph-net</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TV-$\ell_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-enet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-$\ell_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-$\ell_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-$\ell_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM-$\ell_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FReM + clustering</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Hoyos Idrobo et al. PRNI 2015 NeuroImage in Press]
A good model on a budget

(Haxby: objects / scrambled)

Classifiers

- Graph-net
- TV-\(\ell_1\)
- Log-enet
- SVM-\(\ell_2\)
- SVM-\(\ell_1\)
- FReM: SVM-\(\ell_2\)
- FReM: SVM-\(\ell_1\)
- FReM: SVM-\(\ell_2\) + clustering
- FReM: SVM-\(\ell_1\) + clustering

[Hoyos Idrobo et al. PRNI 2015 NeuroImage in Press]
A good model on a budget

State of the art solution: nice but costly

State of the art solution: not very stable, but cheap

[Hoyos Idrobo et al. PRNI 2015 NeuroImage in Press]
Outline

- Sample size issues in MVPA
  - Some remarks on cross-validation
- How to interpret pattern maps in MVPA?
- MVPA and functional specificity
Functional specificity and MVPA

A vs B

A vs B, C, D, E, F
A note on linear decodability

• Can we conclude from successful decoding w. linear classifier that brain activity encodes stimulus information linearly?
  
• No

• Counter-example: position ((x, y) or (r, θ) coordinates) of an object in the visual field
  
  • not encoded linearly
  
  • can be decoded linearly
A note on linear decodability

Visual field

neural response (population receptive fields)

Visual field

neural response (population receptive fields)
A note on linear decodability

Decoding = linear summation

[Thirion et al. Neuroimage 2006]
Conclusion

- MVPA relies on complex estimators
  - Results are unstable under small pertubations
  - Use large test sets (no LOO)
- Pattern maps are highly sensitive to model assumptions and structured noise
  - A thresholded map is another model
  - State precisely what comparison is performed
  - Use relevant priors
- Need comprehensive models
  - Oool data across datasets: OpenfMRI, NeuroVault
The power of scikit learn for MVPA

- Machine learning for neuroimaging [http://nilearn.github.io]
- Scikit-learn-like API
- BSD, Python, OSS
  - Classification of neuroimaging data (decoding)
  - Functional connectivity analysis
Acknowledgements

Parietal
V. Michel
G. Varoquaux
A. Gramfort
F. Pedregosa,
Andés H. idrobo
V. Fritsch,
Y. Schwartz,
E. Dohmatob,
M. Rahim,
M. Eickenberg,
L. Estève,
O. Grisel,
A. Abadie

Other partners and collaborators
(thanks for the data!)
S. Dehaene
R. Poldrack,
C. F. Gorgolewski
K. Jimura,
J. Haxby
Thank you for your attention

http://parietal.saclay.inria.fr