Design of MVPA Studies

Martin Hebart
Department of Systems Neuroscience Hamburg
Goals of this Presentation

MVPA Design
- What are requirements for the experimental design of MVPA studies?

Minimize Noise, Maximize Signal
- How can we maximize the information we extract by our experimental design?

Confounds in MVPA studies
- What are possible confounds that we have to consider?
- How can we avoid and eliminate confounds? (→ Kai Görgen’s talk)
Important Notes

• How you later analyze your data can have important consequences for your design

• This presentation: strong focus on fMRI classification within subject (e.g. different conditions)

• **But**: Principles apply also to between subject classification or other modalities (e.g. EEG)
MVPA DESIGN: GENERAL
Design choices for Multivariate Decoding

Most design choices identical to univariate GLM

- duration of experiment (longer = better)
- scanner settings (TR, TE, flip angle, descending acquisition, …)
- high spatial resolution: unclear if specific benefit for multivariate

What you need to consider

- classification requires independence (or balanced dependence) of training and test data
- often data dependencies exist that invalidate this assumption
- there may be hidden confounds
- design may need special treatment whether it is a block design, slow event-related design, or fast event-related design
Within-subject Designs

What data enters the classifier?

- single-image data
- parameter estimate (beta)

When possible: parameter estimate preferred

Question: trial-wise or run-wise beta estimates?
Within-subject: Trial-wise or Run-wise Betas?
## Within-subject: Trial-wise or Run-wise Decoding?

<table>
<thead>
<tr>
<th></th>
<th>Trial-wise</th>
<th>Run-wise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training data</td>
<td>many samples</td>
<td>few samples</td>
</tr>
<tr>
<td>Data quality</td>
<td>noisy</td>
<td>stable</td>
</tr>
<tr>
<td><strong>Consequence</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (Ku et al., 2008)</td>
<td>lower, but stable</td>
<td>higher, but variable</td>
</tr>
<tr>
<td>Statistical Power</td>
<td>slightly lower</td>
<td>slightly higher</td>
</tr>
<tr>
<td>(Allefeld &amp; Haynes, 2014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>slow</td>
<td>fast</td>
</tr>
<tr>
<td>Slight imbalance</td>
<td>requires correction</td>
<td>no problem</td>
</tr>
<tr>
<td>Strong imbalance</td>
<td>requires correction</td>
<td>confound!</td>
</tr>
<tr>
<td>Statistical estimation</td>
<td>more samples</td>
<td>maybe too few samples</td>
</tr>
</tbody>
</table>

Leave-one-run out cross-validation

Typical Analysis: Leave-one-run-out cross-validation

Reason: Non-independence within run can bias results

Mumford et al (2014) – Neuroimage
How many runs for leave-one-run out?

<table>
<thead>
<tr>
<th></th>
<th>Many short runs</th>
<th>Few long runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data variability</td>
<td>more variable</td>
<td>more stable</td>
</tr>
<tr>
<td>Amount of training data</td>
<td>more training data</td>
<td>less training data</td>
</tr>
</tbody>
</table>

- Usually 4 to 12 runs
- Number often determined by condition balancing within run

*Coutanche & Thompson-Schill (2012) – Neuroimage*
Interim Summary

• Most design choices are identical between univariate GLM and multivariate classification
• Design choice may depend on trial-wise or run-wise decoding
• Non-independence within run suggests preference for leave-one-run out classification
• Ideal number of runs unclear and probably depends on task
HOW TO MINIMIZE NOISE AND MAXIMIZE SIGNAL
Minimize Noise

Example: Head motion vs. classification accuracy


Test pairs across these nine participants. Accuracy across participants was strongly correlated \( r = -0.66 \) with estimated head motion (i.e., the less the participant’s head motion, the greater the prediction accuracy), suggesting that the variation in accu-

Wutte et al (2011) – Front Psych

Alink et al (2013) – Front Psych

head motion - decoding accuracy correlation

rho = -0.71

p = 0.0015
Minimize Noise

- Decoding usually done on unsmoothed data
- Reason: Part of information may lie in fine-scale patterns
- Effects of noise larger on unsmoothed data
- Most important sources of noise that affect fine-scale patterns:
  - head motion
  - physiological noise
  - scanner-related artifacts

Example: Effect of displacement on sampling of orientation columns

→ motion correction only interpolates

Minimize Noise (Example: Head motion)

Solution 1: Avoid head motion

- Pad the head and instruct the subject well
- Use a bite-bar
- Use prospective motion correction (Weiskopf et al)
- Create 3D-printed version of skull (Gallant et al)

Todd et al (2015) - Neuroimage
Minimize Noise (Example: Head motion)

Solution 2: Remove head motion after acquisition

- Motion scrubbing (e.g. removing images with spikes > 0.2 mm)
- Add head motion and physiological parameters to GLM
- Use other methods for removing artifacts, e.g. MELODIC

Minimize Noise (Example: Head motion)

Example application of MELODIC to classification

choice (original)  choice (denoised)

$p < 0.001$ (uncorrected)  $p < 0.001$ (cluster corrected)
Minimize Noise (Example: Head motion)

Example application to dissimilarity matrix

region: 251 voxel sphere in pre-SMA in one subject
dissimilarity measure: Euclidean distance (scaled between 0 and 10)
trial-by-trial dissimilarity (trials spaced regularly in steps of 2.5s)
Maximize Signal: Design Efficiency

Design Efficiency

Goal: Maximize the pattern distinctness, i.e. increase between-class scatter or decrease within-class scatter
→ Improving design efficiency of GLM for univariate contrast will maximize the distinctness before data acquisition
Maximize Signal: Design Efficiency

Design Efficiency

\[ e(C, Z) = \text{trace}(C' \text{inv}(Z'Z) C)^{-1} \]

where \( Z \) is the pre-whitened and high-pass filtered design matrix \( X \)

What does this mean?

- We maximize the **unique** variance of the regressor of interest
- If we are interested in a contrast (e.g. \( C = [1 \ -1] \)), we maximize the variability of the *contrast* of the regressors

Example: How well can we classify betas based on such regressors?

\[ r = -0.9 \]

- each regressor alone: terrible
- classification: excellent

http://imaging.mrc-cbu.cam.ac.uk/imaging/DesignEfficiency
Maximize Signal: Design Efficiency

How do we optimize 2-class classification?

1. Create balanced randomized trial order (and possibly ISI), where
   \[ p(A|A) = p(A|B) = p(B|A) = p(B|B) \]
2. Create design and apply pre-whitening
3. Calculate efficiency using contrast \([1 -1]\)
4. Repeat and pick design with largest efficiency

Additional advice:
- For single-trial event-related analysis and short TR (< 2.5s), time-locking onsets to TR can reduce between trial variability
- Need to consider equal individual estimability, i.e. \(e([1 0],X) = e([0 1],X)\)
THE PERVERSIVENESS OF CONFOUNDS IN MVPA STUDIES
Confounds

Two classes of confounds:
1. Confounds in the experimental design
2. Confounds in the results

→ Best practice is to avoid confounds before they happen
→ We can avoid confounds in experimental design
→ There are some confounds we cannot avoid, but can only control
Confounds in the Experimental Design

Typically disregarded issues can become a confound

Example: Classification of top-down visual attention

Confound: visual cue
Confounds in the Experimental Design

Typically disregarded issues can become a confound

Example: Classification of choice?

Confound: motor response
Confounds in the Experimental Design

Solution 1: Exclude brain regions that confound responds to
  • For motor confound: Exclude motor-related brain regions
  • For visual confound: Exclude visually responsive regions

But: Maybe unexpected brain region responds to confound?
And: Sometimes approach not possible

not recommended
Confounds in the Experimental Design

Solution 2: Separate confound in time or jitter
- For motor confound: Wait 20s with motor response
- For visual confound: Show cue jittered and model separately

But: Pattern autocorrelation can last very long
And: Jitter only reduces confound, never eliminates it!

not recommended
Confounds in the Experimental Design

Solution 3: Balance confound between runs
  • For motor confound: Switch response after each run
  • For visual confound: Invert cue direction

But: The brain may respond more to the confound in some runs or better to one version of the confound
And: Potential bias, task-switching costs

definitely not recommended
Confounds in the Experimental Design

Solution 4: Cross-classification

- For motor confound: Switch response modality, e.g. after each run
- Train classifier on data with one confound, test on data with other
- If above chance, then classifier generalizes across confound
- For visual confound: Use different cue

But: Less data available for training, i.e. possibly reduced sensitivity
And: Possible task-switching costs

Recommended if no better possibility
Confounds in the Experimental Design

Solution 5: Cue Trick
- For motor confound: Use response-mapping rule
- Controls for confound by balancing
- For visual confound: Not always possible

But: Requires training of subject

recommended when possible

Hebart et al (2014) – Cerebral Cortex
Summary

• Reducing noise in the acquisition can have a stronger benefit for multivariate analyses than univariate analyses
• Improving the design efficiency can improve the pattern distinctness
• Confounds are difficult to deal with
• It is not easy to follow the own intuition to avoid confounds
• We provide some recipes for avoiding or eliminating the influence of confounds
• Don’t worry: Slight confounds not always a problem
Maximize Signal

Software for Design Efficiency

- Doug Greve: OptSeq
  http://surfer.nmr.mgh.harvard.edu/optseq/

- Wager & Nichols: Genetic algorithm
  https://github.com/canlab/CanlabCore

- No optimization algorithm, but easy and more flexible method to set up design matrix:
  http://martin-hebart.de/webpages/code/matlab.html

  Suggestion: Brute-force repetition works well in general