Generative embedding and model-based classification

Kay H. Brodersen\textsuperscript{1,2}

\textsuperscript{1} Department of Computer Science, ETH Zurich, Switzerland
\textsuperscript{2} Department of Economics, University of Zurich, Switzerland
Classification approaches by data representation

Model-based classification
How do patterns of hidden quantities (e.g., connectivity among brain regions) differ between groups?

Structure-based classification
Which anatomical structures allow us to separate patients and healthy controls?

Activation-based classification
Which functional differences allow us to separate groups?
Colleagues & collaborators

Thomas Schofield
University College London

Justin R Chumbley
University of Zurich

Cheng Soon Ong
ETH Zurich

Jean Daunizeau
University of Zurich · University College London

Kate Lomakina
University of Zurich · ETH Zurich

Joachim M Buhmann
ETH Zurich

Alexander Leff
University College London

Klaas Enno Stephan
University of Zurich · University College London

Christoph Mathys
University of Zurich · ETH Zurich
Model-based classification through generative embedding

- **step 1 — model inversion**: measurements from an individual subject
- **subject-specific inverted generative model**: A \(\rightarrow\) B, A \(\rightarrow\) C, B \(\rightarrow\) B, B \(\rightarrow\) C
- **subject representation in the generative score space**

- **step 2 — kernel construction**: A \(\rightarrow\) B, A \(\rightarrow\) C, B \(\rightarrow\) B, B \(\rightarrow\) C

- **step 3 — classification**: jointly discriminative connection strengths
  - separating hyperplane to discriminate between groups

Choosing a generative model: DCM for fMRI

haemodynamic forward model
\[ x = g(z, \theta_h) \]

neural state equation
\[ \dot{z} = (A + \sum u_j B^{(j)})z + Cu \]

Stephan & Friston (2007) Handbook of Brain Connectivity
Example: diagnosing stroke patients

To illustrate our approach, we aimed to distinguish between stroke patients and healthy controls, based on non-lesioned regions involved in speech processing.
Example: diagnosing stroke patients

$y = -26 \text{ mm}$

anatomical regions of interest
Example: diagnosing stroke patients
Univariate analysis: parameter densities
Multivariate analysis: connectional fingerprints

patients
controls
Full Bayesian approach to performance evaluation

Brodersen, Chumbley, Mathys, Daunizeau, Ong, Buhmann & Stephan (in preparation)
Classification performance

Activation-based analyses
- a anatomical feature selection
- c mass-univariate contrast feature selection
- s locally univariate searchlight feature selection
- p PCA-based dimensionality reduction

Correlation-based analyses
- m correlations of regional means
- e correlations of regional eigenvariates
- z Fisher-transformed eigenvariate correlations

Model-based analyses
- o gen.embed., original full model
- f gen.embed., less plausible feedforward model
- l gen.embed., left hemisphere only
- r gen.embed., right hemisphere only
Biologically less plausible models perform poorly
The generative projection

Voxel-based activity space

Model-based parameter space
Discriminative features in model space

![Diagram showing connectivity between different brain regions such as PT, HG, and MGB with stimulus input indicated.]
Discriminative features in model space

- **MGB**
- **PT**
- **HG (A1)**

- **L**
- **R**

- **stimulus input**

- Highly discriminative
- Somewhat discriminative
- Not discriminative
Model-based inference on individual pathophysiology

1. Model of neuronal (patho)physiology

2. Application to brain activity data from individual patients

3. Diagnostic classification
   - Type 1
   - Type 2
   - Type 3

Spectrum disease

Generative embedding
Summary: generative embedding for fMRI

1. **Strong classification performance.** Generative embedding exploits the rich discriminative information encoded in ‘hidden’ quantities, such as coupling parameters.

2. **Creation of an interpretable feature space.** High-dimensional fMRI data are replaced by low-dimensional subject-specific fingerprints with biologically interpretable axes.

3. **Future applications.** Generative embedding could help dissect spectrum disorders into physiologically defined subgroups (*in preparation*).