NeuroImage xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Full Length Articles

Representational similarity encoding for fMRI: Pattern-based synthesis to predict brain activity using stimulus-model-similarities

Andrew James Anderson *, Benjamin D. Zinszer, Rajeev D.S. Raizada

Brain and Cognitive Sciences, University of Rochester, NY 14627, USA

A R T I C L E I N F O

Article history:
Received 13 August 2015
Accepted 19 December 2015
Available online xxxx

Keywords:
Representational Similarity Analysis
Encoding
Decoding
Semantic model
Semantic memory
fMRI

A B S T R A C T

Patterns of neural activity are systematically elicited as the brain experiences categorical stimuli and a major challenge is to understand what these patterns represent. Two influential approaches, hitherto treated as separate analyses, have targeted this problem by using model-representations of stimuli to interpret the corresponding neural activity patterns. Stimulus-model-based-encoding synthesizes neural activity patterns by first training weights to map between stimulus-model features and voxels. This allows novel model-stimuli to be mapped into voxel space, and hence the strength of the model to be assessed by comparing predicted against observed neural activity. Representational Similarity Analysis (RSA) assesses models by testing how well the grand structure of pattern-similarities measured between all pairs of model-stimuli aligns with the same structure computed from neural activity patterns. RSA does not require model fitting, but also does not allow synthesis of neural activity patterns, thereby limiting its applicability. We introduce a new approach, representational similarity-encoding, that builds on the strengths of RSA and robustly enables stimulus-model-based neural encoding without model fitting. The approach therefore sidesteps problems associated with overfitting that notoriously confront any approach requiring parameter estimation (and is consequently low cost computationally), and importantly enables encoding analyses to be incorporated within the wider Representational Similarity Analysis framework. We illustrate this new approach by using it to synthesize and decode fMRI patterns representing the meanings of words, and discuss its potential biological relevance to encoding in semantic memory. Our new similarity-based encoding approach unites the two previously disparate methods of encoding models and RSA, capturing the strengths of both, and enabling similarity-based synthesis of predicted fMRI patterns.

Introduction

The brain represents different categories as spatially distributed and overlapping activity patterns, and a major challenge is to crack this representational code (Haxby et al., 2001; Haxby et al., 2014). Neural activity can be elicited by presenting participants with various stimuli (e.g. words, images, sounds) and recorded by neuroimaging techniques such as functional Magnetic Resonance Imaging (fMRI). Two approaches targeting the problem of explaining the resultant neural codes are stimulus-model-based-encoding and Representational Similarity Analysis (RSA). Stimulus-model-based-encoding forms models of stimuli as vectors of feature-weights. For pictorial stimuli, model-features may correspond to visual filters (e.g. Kay et al., 2008; Naselaris et al., 2009), for words, features may be the association of the word with senses used to experience the word's referent (e.g. Mitchell et al., 2008; Fernandino et al., 2015; Anderson et al., submitted for publication). Synthesized neural activity patterns corresponding to new model-stimuli are predicted by a mapping from model-features to voxels trained by fitting weights to features with supervised learning. In contrast, RSA assesses models by comparing the grand structure of similarities between all pairs of stimulus-model feature-vectors and neural activity patterns, and does not require model fitting but cannot synthesize predicted voxel-space activation patterns.

We present a new approach, similarity-encoding, that bridges between stimulus-model-based-encoding and RSA. The new method is illustrated in Fig. 1. This approach achieves similar accuracy in synthesizing predicted neural activity patterns to standard regression-based strategies, but without model fitting. Hence unlike standard regression we observe that similarity-encoding robustly manages situations where there are many more stimulus-model dimensions than stimuli. We also show how this new approach enables stimulus-model-based-decoding of novel fMRI data to be entirely abstracted to representational-similarity space (Fig. 2). Thus, like regression there is generalization from trained to untrained stimuli. However, the generalization here stems from exploiting the structure of similarity-space.

Encoding and decoding (discussed in detail in the context of fMRI by Naselaris et al., 2011) are of broad relevance to assess the value of models/and or neural data to making practical decisions, e.g., clinically

* Corresponding author.
E-mail address: andrewanderson@bcs.rochester.edu (A.J. Anderson).

http://dx.doi.org/10.1016/j.neuroimage.2015.12.035
1053-8119/© 2015 Published by Elsevier Inc.

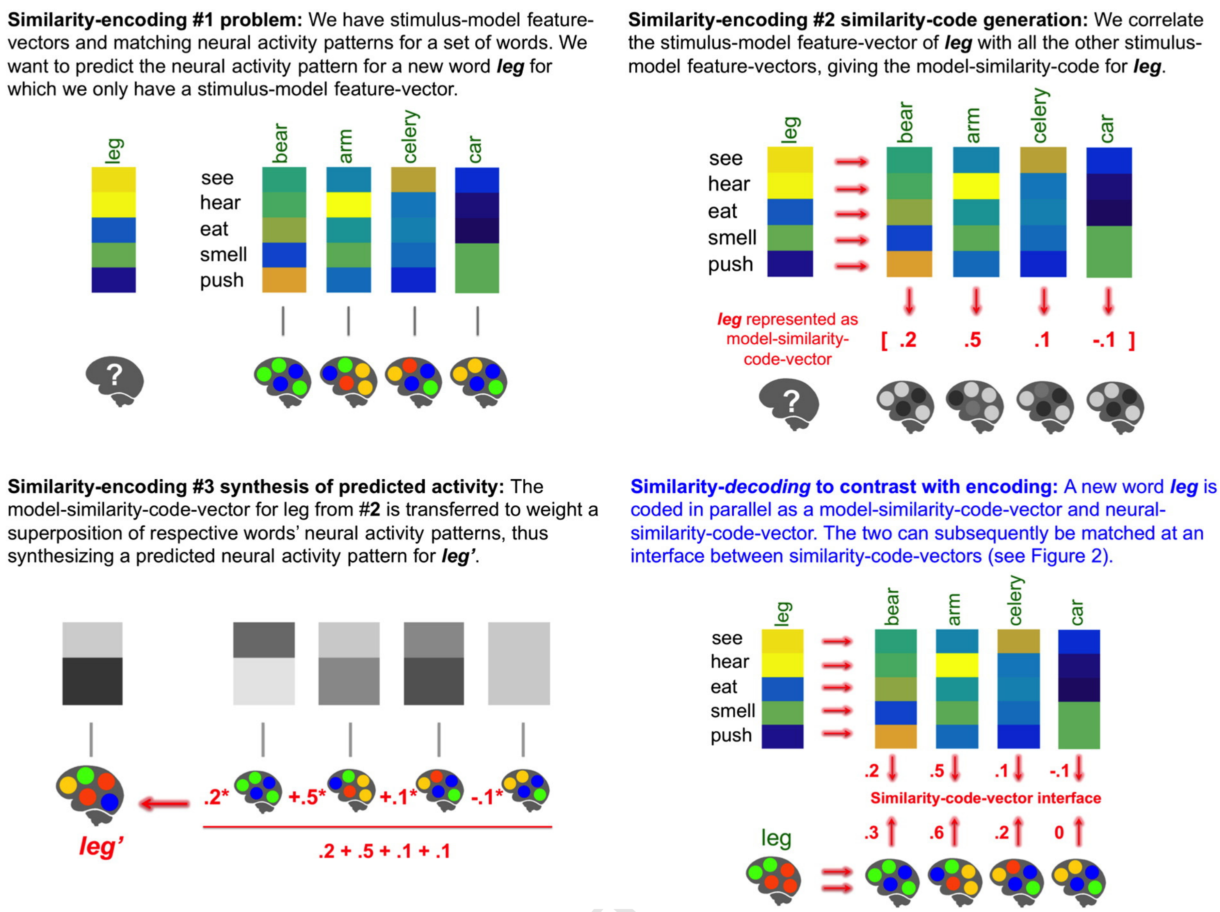

Fig. 1. The three stages of similarity-based neural-activity-pattern encoding. Separate to this the fourth panel illustrates similarity-based-decoding for contrast with encoding in the other three panels (see Fig. 2 for further details of the new similarity-based decoding algorithm).

in distinguishing healthy and unhealthy samples (e.g., Just et al., 2014; Matthews et al., 2006), in brain–computer-interfaces and neuroprosthetics (e.g. Sulzer et al., 2013; deCharms, 2008), or from an ecological perspective to estimate whether measured neural activity patterns could actually be the grounds of decision making within an individual. As such whilst RSA and neural encoding and decoding have tended to be treated as separate analyses with different properties and benefits (e.g. Haxby et al., 2014), the extension introduced here provides a means for all types of analyses to be easily undertaken within the same similarity based framework. Where previous analyses have decoded neural activity patterns using representational-similarity methods (e.g. Raizada and Connolly, 2012, Nili et al., 2014; Anderson et al., 2015; Zinszer et al., 2015), none have considered encoding (synthesis of predicted neural activity patterns from stimulus-models).

Methodologically, the new similarity-encoding strategy is a natural development to the Representational Similarity Analysis (RSA) Q3 framework (Kriegeskorte et al., 2008a,b; Kriegeskorte and Kievit, 2013; Nili et al., 2014), building on theories that visual-object categories are partially represented in terms of similarities in the brain (Edelman, 1998, Edelman et al., 1998) and (as we will return to in the Discussion) follows a computational architecture reminiscent of distributed associative memory neural networks (e.g. Willshaw et al., 1969). RSA takes a matching set of stimulus-feature-vectors and neural activity patterns and measures the degree of association between the stimulus models and neural modalities by (1) inter-correlating all pairs of stimulus-feature-vectors to produce a square model-correlation matrix; (2) likewise inter-correlating all pairs of neural activity patterns to produce an equivalent square neural-correlation matrix. (3) Quantifying the association between the model-correlation matrix and the neural-correlation matrix by extracting the lower below diagonal triangle (or upper) of unique pairwise comparisons from each matrix, vectorizing both to produce similarity-structure-vectors, and correlating model and neural-similarity-structure-vectors to quantify the association. By vectorizing the similarity-structure, conventional RSA treats an entire data set holistically. This strategy has proved extremely successful e.g. in interpreting pictorially induced representations in the brain, as in Kriegeskorte et al. (2008a,b) and Connolly et al. (2012), and demonstrating that the semantic structure embedded within neural activity patterns associated with comprehending concrete nouns matches sets of semantic models of those nouns (e.g. Q4 Bruffaerts et al., 2013; Carlson et al., 2014; Anderson et al., 2013, 2015). However this holistic comparison does not allow synthesis of predicted voxel-space activation patterns, and it is here that our approach introduces new capabilities.

As opposed to manipulating the representational similarity-structure holistically, we use inter-correlations between stimulus-model feature-vectors as a secondary code to represent stimuli. Therefore under our approach a stimulus is modeled with two codes, the first is the standard stimulus-model feature-vector, the second – the similarity-code – is a vector of correlations with other stimulus-model feature-vectors. The similarity-code is an independent representation that defines the similarity between one stimulus and other stimuli and adheres to theories that consider similarities to underpin object categories in the brain (Edelman, 1998; Edelman et al., 1998).

Encoding – the synthesis of a predicted neural activity pattern – is achieved by: taking a new stimulus-model feature-vector for which we would like to predict the associated neural activity; generating a new similarity-code for that stimulus-model feature-vector;

Please cite this article as: Anderson, A.J., et al., Representational similarity encoding for fMRI: Pattern-based synthesis to predict brain activity using stimulus-model-similarities, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.12.035