HHS Public Access

Author manuscript

Trends Cogn Sci. Author manuscript; available in PMC 2016 October 01.

Published in final edited form as:

Trends Cogn Sci. 2015 October ; 19(10): 551–554. doi:10.1016/j.tics.2015.07.005.

Resolving ambiguities of MVPA using explicit models of representation

Thomas Naselaris[1] and **Kendrick N. Kay**[2]

[1]Medical University of South Carolina, Charleston, SC

[2]Washington University in St. Louis, St. Louis, MO

Abstract

We advocate a shift in emphasis within cognitive neuroscience from multivariate pattern analysis (MVPA) to the design and testing of explicit models of neural representation. With such models it becomes possible to identify the specific representations encoded in patterns of brain activity and to map them across the brain.

Keywords

representation; encoding model; multivariate pattern analysis; computational modeling; fMRI

Multivariate pattern analysis (MVPA) is a powerful analysis tool that is replacing activation (or subtraction-based) analysis as the go-to method for interpreting fMRI data [1]. MVPA refers to classification of patterns of brain activity into discrete experimental conditions (e.g. different stimuli, tasks, or cognitive states). The major appeal of MVPA is its sensitivity: it can identify populations of voxels that encode information about experimental conditions, even when the average amplitude of activity in the population does not vary across conditions.

Despite its appeal, MVPA has critical limitations as a tool for identifying the representations that are encoded in patterns of brain activity. There are three distinct kinds of ambiguity inherent to MVPA. The most benign kind is *geometric ambiguity*. This refers to the fact that activity patterns, interpreted as multivariate vectors, can be discriminated by MVPA on the basis of either their length (overall activation) or orientation. Although the pooling of length and orientation provides statistical sensitivity, these distinct features of the activity pattern cannot be disentangled when using MVPA alone. For example, the overall activity in a region may simply be larger in one condition compared to another, a fact that is missed in MVPA. Geometric ambiguity can be resolved by performing additional analyses (such as activation analysis).

Corresponding author: Naselaris, T., tnaselar@musc.edu.

More problematic is *spatial ambiguity*. MVPA provides little information about how representations are organized across the cortical surface (e.g. the retinotopic organization of visual cortex [2]). This ambiguity arises from the fact that different cortical organizations can give rise to identical classification performance [3]. Techniques for resolving spatial ambiguity in MVPA, such as the use of searchlights or examination of classifier weights, can be misleading [4–6]. For example, significant nonzero classifier weights can be obtained for voxels whose average response is the same across the experimental conditions of interest [5].

The most serious limitation of MVPA is *representational ambiguity*. Even moderately complex stimuli or task paradigms contain many distinct sources of variation. Each source corresponds to different stimulus features or cognitive states that might be encoded in brain activity. MVPA does not provide a framework for testing and distinguishing between different sources of variation [3].

Here is an illustration of how representational ambiguity can arise. Suppose we hypothesize that a brain region of interest (ROI) specializes in representing the genre of movie one is watching (Figure 1). To test the hypothesis, we conduct an experiment in which subjects are scanned while viewing segments of movies of two different genres—say action movies and romantic comedies. If it turns out that a classifier is able to accurately discriminate the movie segments of each genre on the basis of measured brain activity, we will have established that *something* about the movie segments is indeed encoded in the activity patterns of our ROI. However, we will not have determined what that *something* actually is. A variety of alternative features correlated with movie genre might be encoded in the activity patterns, including visual and auditory energy (e.g. action movies contain more energy than romantic comedies), amount of social interaction, humor, amount of spoken language, etc.

In some cases it may be possible to discriminate features by using an experimental design that varies one and only one feature at a time. Although careful experimental design will always play an important role in studying brain representations, in the case of MVPA studies, experimental design can be surprisingly difficult. Consider for example a highly controlled experiment in which sinusoidal gratings of different orientations are presented: successful orientation decoding may not necessarily derive from an encoding of the orientation of the stimulus, but may instead derive from an encoding of edge artifacts in the stimulus [3].

One might attempt to use MVPA to compare the movie-genre hypothesis against alternative hypotheses by comparing classification performance obtained for different features. For example, we might divide movie segments into low/high stimulus energy, low/high social interaction, low/high humor content, etc., and then train a separate classifier to discriminate each of these. According to this logic, if movie genre is discriminated with higher performance than other features, then the movie-genre hypothesis is affirmed. However, this approach is problematic because decoding performance does not directly indicate the amount of variance in activity that is attributable to a given feature. A feature may be perfectly decoded from population activity even though it is responsible for very little

variance in activity. For example, a purely visual representation might support highly accurate decoding of genre, even though genre *per se* explains little variance in the brain responses. Comparing classification performance across different kinds of features is therefore an "apples-to-oranges" comparison that is likely to mislead.

Many researchers have adopted an alternative approach for identifying representations encoded in brain activity [2,3,7–11]. We refer to this approach as voxelwise-modeling (VM). The hallmark of VM is an explicit model of representation, known as an *encoding model*. Formally, an encoding model proposes a set of sensory or cognitive *features* and specifies how these features are transformed into a prediction of brain activity for the experiment under consideration. A given set of features represents an explicit hypothesis about the representation encoded in the brain. This hypothesis is tested by evaluating how much variance in measured activity the encoding model explains (Box 1). Competing hypotheses can be adjudicated by comparing the amount of variance explained by different encoding models. Alternatively, hypotheses can be assessed by comparing how well a representational similarity matrix (e.g. a matrix with correlations between pairs of experimental conditions) constructed from a set of features matches the representational similarity matrix constructed from the measured activity. This approach, called *representational similarity analysis*, imposes fewer constraints on the mapping between features and brain activity [12]. In both cases, hypotheses are tested by evaluating explicit models of representation.

Box 1

Steps in building encoding models

(1) *Design the experiment.* Typically, a large number of conditions are used in order to sample a variety of features, postponing commitment to the specific features that may be relevant to a given brain area. (2) *Collect the data.* Physiological responses are measured using multiple repetitions of each condition so that response variability (i.e. noise level) can be quantified. (3) *Select a model.* The features hypothesized to be encoded in a given brain area are formally specified. (4) *Fit the model.* Free parameters of the model (e.g. weights in a linear model) are adjusted to best fit the data. This can entail ordinary least-squares estimation or regularized estimation procedures such as ridge regression or the lasso. (5) *Summarize model parameters.* Parameters are summarized and compared across brain areas using simple metrics (e.g. mean, median) or more sophisticated methods (e.g. principal components analysis, model-based decoding). Reliability of parameter estimates is also assessed (e.g. by bootstrapping trials or subjects). (6) *Quantify model accuracy.* To control for overfitting, model accuracy is assessed by cross-validating on new data (e.g. new trials, experimental conditions, or subjects). Accuracy is quantified as percent variance explained. (7) *Consider alternative models.* The modeling procedure is repeated to determine whether the data might be better explained by a simpler or completely different model.

VM offers important advantages over MVPA. There is no notion of geometric or spatial ambiguity. Analyses are performed on individual voxels, so the length and orientation of an

activity pattern are naturally separated and individual parameters can be mapped to the cortical surface at the native resolution of the data.

Importantly, VM provides the means to resolve representational ambiguity. Encoding models predict activity based on explicitly defined representations. This makes it possible to enumerate different potential sources of variation, test the explanatory power of each source of variation, and identify specific data points that are well or poorly predicted by a given model [7].

Finally, VM provides a quantitative benchmark of our understanding of neural representation. In an experiment where responses are measured to a wide range of stimulus or task conditions, a model that perfectly explains the observed variance in voxel activity in an ROI (or, alternatively, a similarity matrix constructed from the observed activity patterns [12]) could be offered as a complete theory of the ROI.

In conclusion, by improving detection sensitivity MVPA is a powerful tool that has served the fMRI community well. In situations where prediction of stimulus or task states is of primary importance, MVPA will continue to play a useful role. However, MVPA provides fundamentally ambiguous results regarding the nature of brain representations. As research in cognitive neuroscience moves forward, we suggest that MVPA should be replaced by explicit models of representation.

Acknowledgments

This work was supported by the McDonnell Center for Systems Neuroscience and Arts & Sciences at Washington University (K.N.K.) and by grant NEI R01 EY023384 (T.N.).

References

1. Haxby JV, et al. Decoding neural representational spaces using multivariate pattern analysis. Annual review of neuroscience. 2014; 37:435–456.
2. Dumoulin SO, Wandell B. Population receptive field estimates in human visual cortex. NeuroImage. 2008; 39:647–660. [PubMed: 17977024]
3. Carlson TA, Wardle SG. Sensible decoding. NeuroImage. 2015; 110:217–218. [PubMed: 25680521]
4. Etzel JA, et al. Searchlight analysis: promise, pitfalls, and potential. NeuroImage. 2013; 78:261–269. [PubMed: 23558106]
5. Haufe S, et al. On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage. 2014; 87:96–110. [PubMed: 24239590]
6. Davis T, et al. What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. NeuroImage. 2014; 97:271–283. [PubMed: 24768930]
7. Kay KN, et al. A two-stage cascade model of BOLD responses in human visual cortex. PLoS computational biology. 2013; 9:e1003079. [PubMed: 23737741]
8. Naselaris T, et al. A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. NeuroImage. 2015; 105:215–228. [PubMed: 25451480]
9. Mitchell TM, et al. Predicting human brain activity associated with the meanings of nouns. Science. 2008; 320:1191–1195. [PubMed: 18511683]
10. Santoro R, et al. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. PLoS computational biology. 2014; 10:e1003412. [PubMed: 24391486]

11. Brouwer GJ, Heeger DJ. Decoding and reconstructing color from responses in human visual cortex. J Neurosci. 2009; 29:13992–14003. [PubMed: 19890009]

12. Kriegeskorte N, Kievit RA. Representational geometry: integrating cognition, computation, and the brain. Trends in cognitive sciences. 2013; 17:401–412. [PubMed: 23876494]

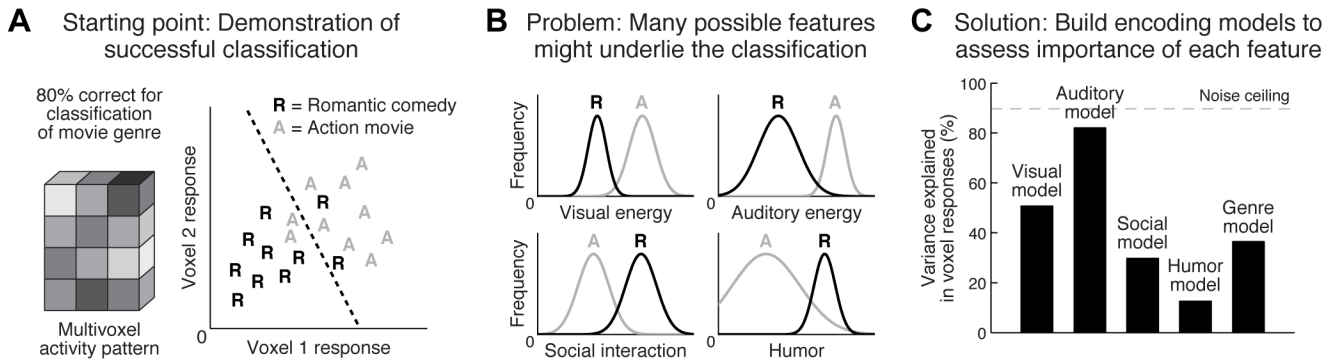

Figure 1. Resolving the representational ambiguity of MVPA

(A) Example MVPA experiment. Responses to several movie clips are measured. It is demonstrated that a linear classifier can predict the genre of a movie clip based on the multivoxel activity pattern elicited by that movie clip. (B) Representational ambiguity. Movie clips from different genres may differ with respect to one or more features. For example, clips from action movies (gray) may have larger amounts of visual energy than clips from romantic comedies (black). Thus, the voxels under consideration might represent a feature that is correlated with, but distinct from, movie genre. (C) Building encoding models. To adjudicate between competing hypotheses about the features that are represented, each feature of interest is used to build an encoding model and the various models are fit to the data (see Box 1). By directly comparing the accuracy (variance explained) of different models, it is possible to determine the features encoded in the population activity.