NeuroImage 62 (2012) 852–855

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Review

Multivariate pattern analysis of fMRI: The early beginnings

James V. Haxby *

Center for Cognitive Neuroscience, Dartmouth College, Hanover, NH, USA
Center for Mind/Brain Sciences, University of Trento, Italy

ARTICLE INFO

Article history:
Accepted 5 March 2012
Available online 9 March 2012

Keywords:
fMRI
Multivariate pattern analysis (MVPA)
Vision
Decoding
Machine learning
Pattern classification

ABSTRACT

In 2001, we published a paper on the representation of faces and objects in ventral temporal cortex that introduced a new method for fMRI analysis, which subsequently came to be called multivariate pattern analysis (MVPA). MVPA now refers to a diverse set of methods that analyze neural responses as patterns of activity that reflect the varying brain states that a cortical field or system can produce. This paper recounts the circumstances and events that led to the original study and later developments and innovations that have greatly expanded this approach to fMRI data analysis, leading to its widespread application.

Contents

Multivariate pattern analysis (MVPA) of fMRI data has proven to be more sensitive and more informative about the functional organization of cortex than is univariate analysis with the general linear model (GLM). MVPA refers to a set of methods that analyze neural responses as patterns of activity, thus affording investigation of the varying brain states that a cortical field or system can produce, thus increasing the amount of information that can be decoded from brain activity, in contrast to simpler univariate measures that indicate the extent to which a cortical field or system is globally engaged. We first devised a prototype MVPA method in the course of investigation of the functional architecture for face and object recognition in ventral temporal cortex (Haxby et al., 2001). The results of this study demonstrated the basic concept and power of MVPA. Initially, however, the paper attracted attention primarily for its neuroscientific content, namely as a strong argument for distributed representation of high-level visual percepts that stood in contrast to modular accounts (Kanwisher et al., 1997). Recognition of the significance of the methodological innovation was slower. The initial disinclination of others to use MVPA in fMRI research was due, in part, to unfamiliarity and the perceived complexity of these methods. MVPA does not provide simple answers to the kinds of questions people were asking — Where is the motion area? Where is the face (or place or body parts) area? Where is the numbers area? and so forth; and partly because they addressed questions that people hadn't thought of investigating — What are the varying brain states in an area and how do they encode different types of information? These methods were not simply another method to answer the same questions but, rather, challenged cognitive neuroscientists to consider a different model of cortical organization.

In 1991 and 1994, working with Cheryl Grady, Barry Horwitz, Leslie Ungerleider, Mort Mishkin, and Stanley Rapoport, we published two positron emission tomography (PET) studies on the division of visual processing in the human brain into the ventral object vision pathway and the dorsal spatial vision pathway (Haxby et al., 1991, 1994). We had chosen a face-matching task as our proxy for object vision in the ventral pathway. Without fail, wherever we presented these data, we were asked why we had made this decision, given the neuropsychological, developmental, and behavioral evidence that face processing has a special status that is distinct from processing of other objects.

In the meantime, I had moved from the National Institute on Aging to the National Institute of Mental Health (down two floors) where I developed an independent cognitive neuroscience research program working with Leslie Ungerleider and Alex Martin. With our entry into fMRI research, one of my first goals was to address whether

* Center for Cognitive Neuroscience, Dartmouth College, Hanover, NH, USA.
E-mail address: james.v.haxby@dartmouth.edu.

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.03.016

face and object perception activate the same or different cortical regions. As we were still puzzling over our initial results, others published reports on a specialized area in the fusiform gyrus that responded more during face perception tasks than during perception of other objects (Kanwisher et al., 1997; McCarthy et al., 1997; Puce et al., 2006). We decided to conduct two further experiments — one looking at the effect of face inversion (Haxby et al., 1999) and the other looking at whether two different categories of objects evoked equivalent patterns of response in the non-face-selective cortex (Ishai et al., 1999). The results of both studies confirmed the existence of a patch of fusiform cortex that responded more to faces than to other stimuli but led us to doubt the interpretation that this region is specialized for face processing and nothing else. As I analyzed the results of the face inversion paper, I also devised a new method of correlating patterns of response to different conditions, which became the basis for our paper in 2001 that introduced MVPA (Haxby et al., 2001).

The principal finding from these two studies that made us doubt that the face area was specialized for face perception and nothing else, and that face processing was restricted to the face area, came from our analysis of responses to faces, houses, and chairs, relative to a no stimulus baseline, in ventral temporal cortex (Haxby et al., 1999, Fig. 2; Ishai et al., 1999, Fig. 3). The voxels in the face area clearly showed significant responses to the non-preferred categories. Similarly, chair-selective voxels and house-selective voxels also showed positive, and usually significant, responses to non-preferred categories (with the exception of a negligible response to faces in medial fusiform house-selective voxels). The strong modularity hypothesis, and the logic of univariate analysis of contrasts in neuroimaging, implied that these weaker, albeit significant, brain responses played no role in representation. In essence, they are discarded because they indicate that the stimulus is a suboptimal fit for the function in those cortical fields. This seemed improbable to us on two counts. First, discarding information in submaximal responses seemed suboptimal and conflicted with methods for analyzing population response representations. Second, the proposal that every possible face, animal, and object category has a specialized region or set of neurons dedicated to its representation didn't seem possible. There are too many ways that faces and objects can be categorized. Moreover, category-dedicated systems do not capture the relationships, or similarities, among categories. Instead, it seemed more probable that the weak responses play an important role in representation, implying that face and object categories are encoded as patterns of neural activity, rather than as peak responses in category-specific modules, and that these patterns involve neural populations that play a role in the representation of multiple categories.

We set out to test this hypothesis and began collecting data in 1998 on neural responses to eight categories of objects, animals, and faces — human faces, cats, chairs, shoes, scissors, bottles, houses, and phase-scrambled images. The broad outlines of the project were clear from the beginning but working out the details of the analysis took a lot of time. We hypothesized that each category would evoke a distinct pattern of response in ventral temporal cortex. We hypothesized further that these distinctive patterns would not be restricted to category-selective regions, such as the FFA and PPA, leading to the prediction that distinctive patterns could be detected if such regions were excluded from the analysis. We also hypothesized that neural activity patterns within category-selective regions would carry information that discriminates between categories other than those regions' preferred stimuli. The analytic approach would be based on the pattern correlation method that I had devised for our paper on face inversion (Haxby et al., 1999). In hindsight, the project was straightforward from the beginning and was completed mostly as planned. At the time, I felt as if I were groping around in the dark working out how to actually execute these analyses. My next-door neighbor at NIH, Alex Martin, thought that I would never emerge from these labors.

The idea was straightforward and based on a concept from conventional statistics, namely split-sample cross-validation. If a given stimulus category evoked a distinct pattern of activity, then independent observations of the response to that category should be more similar to each other than to responses to different categories. Correlation of patterns was the chosen measure of similarity, and I made independent observations by dividing the data for each subject into two halves — even-numbered and odd-numbered runs. Thus, I predicted that within-category correlations would be higher than between-category correlations. The subsidiary hypotheses were to be tested in two further analyses. In the first, I identified the voxels that responded maximally to each category and eliminated those voxels for each possible pair of categories when comparing within- and between-category correlations for that pair. In the second, I identified the FFA and PPA and tested for category-specific patterns for all categories within those areas.

In retrospect, it is hard to remember why devising and executing the analysis took so long. Working with fMRI analysis systems that were designed for univariate analyses, primarily AFNI (Cox) and, to a lesser extent, our own home-grown FIDAP (Functional Imaging Data Analysis Program, written by José Maisog in my group, and decommissioned years ago) and UNIX programming, each step was constructed using tools not designed for this application. For example, the correlations were all calculated using a little-known function in AFNI, 3ddot, that calculated one correlation at a time.

At the time, I was unaware of other pattern classification methods from machine learning. The relevance and utility of these methods were demonstrated through collaborations with colleagues who reanalyzed my data using neural net classifiers (Hanson et al., 2004) and linear discriminant analysis (LDA) (O'Toole et al., 2005) and through subsequent papers from groups that analyzed similar datasets collected independently, most prominently Cox and Savoy's (2003) application of support vector machines (SVM). Tom Mitchell later dignified the split-sample, correlation-based method, which I had devised based on conventional statistics, by giving it a respectable machine learning pattern classifier name, calling it a 'one-nearest neighbor' classifier using a correlation-based distance measure (correlation-based 1NN). Curiously, this method is still used in many reports (e.g. Peelen et al., 2010), presumably because of its conceptual simplicity and straightforward interpretation. It also proved to be surprisingly sensitive, although we have found that other classifiers, in particular SVM (e.g. Connolly et al., 2012; Haxby et al., 2011), consistently outperform correlation-based 1NN. Later, my colleague at Princeton, Ken Norman, took to calling this approach 'multivoxel pattern analysis' (MVPA), which we subsequently changed to 'multi*variate* pattern analysis', to acknowledge, with no need for a new acronym, its application to feature sets other than voxels.

The general adoption of MVPA gained momentum very slowly. I believe that three factors underlay this inertia. First and foremost, it was unclear what the results of an MVPA classification analysis meant in terms of neural representation. Second, it seemed complicated and software for implementation was not available. Third, MVPA analyses typically were done separately for each individual because the pattern structure that carries subtle distinctions appeared to be based on fine-grained topographies that did not align well across brains based on anatomy. Consequently, the cortical topographic features that carry these subtle distinctions were unspecified and somewhat mysterious.

Enthusiasm for MVPA ticked up significantly with two reports that appeared back-to-back in *Nature Neuroscience* in 2005 (Haynes and Rees, 2005; Kamitani and Tong, 2005), both of which showed that MVPA could be applied to a visual feature with a well-understood neural basis, namely edge orientation. Kamitani and Tong (2005) and later papers from Haynes' group showed further that MVPA

could decode cognitive states, such as the target of selective attention (Kamitani and Tong, 2005), and the intention to perform one task rather than another (Haynes et al., 2007). Shortly thereafter, a series of review papers began to appear that presented the principles of MVPA in a clear and accessible manner, further demystifying the basis of this new analytic approach (Haynes and Rees, 2006; Mur et al., 2009; Norman et al., 2006; O'Toole et al., 2007; Pereira et al., 2009; Tong and Pratte, 2012).

A parallel development was the realization that multivariate response patterns also could be analyzed in terms of strength of similarities among response patterns, rather than simply as binary distinctions, affording analysis of the structure of representational spaces. This work actually began with an underappreciated paper by Edelman et al. (1998). Reanalyses of our data by Steve Hanson et al. (2004) and Alice O'Toole et al. (2005) revealed similarity structures that conformed to intuitions about semantic relationships — namely, a major distinction between animate and inanimate objects and a secondary distinction between houses and small inanimate objects. This approach was amplified greatly in a landmark study by Kiani et al. (2007) of similarity structure in the population responses of large numbers of single cells in monkey IT cortex, followed by a second landmark study (Kriegeskorte et al., 2008a), which showed that the similarity structure in monkey IT cortex (from Kiani et al.'s single unit data) and human VT cortex (using fMRI) is remarkably similar, characterized by major distinctions between animate and inanimate stimuli and, within the animate domain, between faces and bodies. In a later study, Connolly et al. (2012) found even more structure in the representation of animate entities in human VT cortex, with a similarity structure that embodies semantic relationships among classes of animal species, similar to what Kiani et al. (2007) had found in monkey IT population responses. Kriegeskorte et al. (2008b) has proposed that analysis of similarity structure could provide a common basis for comparing representational spaces across multiple neuroscience approaches, such as fMRI, single unit physiology, computational models, and stimulus models, and gave this approach a new name, 'representational similarity analysis' or RSA, which has stuck.

Meanwhile, another new approach showed that multivoxel responses to new stimuli could be predicted based on high-dimensional feature models of stimuli (Kay et al., 2008; Mitchell et al., 2008; Naselaris et al., 2009; Nishimoto et al., 2011). These methods make the relationships between stimulus attributes and response patterns explicit, bringing us closer to understanding how patterns of activity in fMRI data and stimulus representation are related.

MVPA is more complicated than conventional univariate analysis, and this complexity has been a major barrier for many. Good software is now readily available, however, that makes MVPA more accessible to investigators. When I was at Princeton, a graduate student in Ken Norman's laboratory, Greg Detre, took on the project of developing an MVPA toolbox using Matlab (the Princeton MVPA toolbox) (http://code.google.com/p/princeton-mvpa-toolbox/). Around the same time, Michael Hanke, who at the time was a graduate student studying with Stefan Pollman at the University of Magdeburg, came to Princeton to learn MVPA in my laboratory. Michael Hanke turned out to be a brilliant software developer and a fervent advocate of free and open-source software (FOSS). He decided to develop PyMVPA, a Python-based MVPA toolbox, in collaboration with Yaroslav Halchenko (www.pymvpa.org) (Hanke et al., 2009), who was working with Steve Hanson at Rutgers University. By using Python, this toolbox incorporates software from a vast library of machine learning algorithms.

Another source of wariness about MVPA came from its reliance on fine-scale topographic features to detect subtle distinctions between patterns. These fine-scale features cannot be aligned well across brains based on anatomy. Consequently, MVPA generally builds classifier models for an individual subject based on that subjects' own data. As a result, the population response topographies that carry fine distinctions are difficult to characterize and appear idiosyncratic. Ten years after our 2001 paper, we published a paper that, we believe, solves this problem (Haxby et al., 2011). In this report, we presented a new method, hyperalignment, that enabled us to derive common models of cortical representational spaces that capture all of the information for MVP classification in a set of a few dozen basis functions. These basis functions have stimulus tuning functions that are common across brains and model individual voxel tuning functions as weighted sums. These basis functions also are associated with individual-specific pattern basis functions that model individual cortical response patterns as weighted sums. These models thus account for the fine-scale topographies that underlie the sensitivity of MVPA in a straightforward linear model and show that these topographies have a basis that is common across individuals.

Ten years after our paper that introduced MVPA in 2001, the methods have become vastly more sophisticated, software has been developed that make implementation with standardized algorithms widely accessible, and the conceptual foundations have been clarified and amplified. The methods are more complicated than those of conventional univariate analysis, requiring more sophistication and computational knowledge on the part of its users. MVPA, however, allows one to investigate questions about the how information is encoded in patterns of neural activity, rather than simpler questions about where functions simply are performed. Many investigators are now finding that the scientific payoff justifies the effort necessary to use these methods effectively. Now that this approach has engaged a large community of sophisticated investigators, the rate of development and innovation is, if anything, accelerating.

References

Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.-C., Abdi, H., Haxby, J.V., 2012. The representation of biological classes in the human brain. J. Neurosci. 32, 2608–2618.

Cox, D.D., Savoy, R.L., 2003. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19, 261–270.

Edelman, S., Grill-Spector, K., Kushnir, T., Malach, R., 1998. Toward direct visualization of the internal shape space by fMRI. Psychobiol. 26, 309–321.

Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J.V., Pollman, S., 2009. PyMVPA: a Python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics 7, 37–53.

Hanson, S.J., Toshihiko, M., Haxby, J.V., 2004. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area. Neuroimage 23, 156–167.

Haxby, J.V., Grady, C.L., Horwitz, B., Ungerleider, L.G., Mishkin, M., Carson, R.E., Herscovitch, P., Schapiro, M.B., Rapoport, S.I., 1991. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. Proc. Natl. Acad. Sci. U. S. A. 88, 1621–1625.

Haxby, J.V., Horwitz, B., Ungerleider, L.G., Maisog, J.M., Pietrini, P., Grady, C.L., 1994. The functional organization of human extrastriate cortex: A PET-rCBF study of selective attention to faces and locations. J. Neurosci. 14, 6336–6353.

Haxby, J.V., Ungerleider, L.G., Clark, V.P., Schouten, J.L., Hoffman, E.A., Martin, A., 1999. The effect of face inversion on activity in human neural systems for face and object perception. Neuron 22, 189–199.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430.

Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., Ramadge, P.J., 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72, 404–416.

Haynes, J.D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in primary visual cortex. Nat. Neurosci. 8, 686–691.

Haynes, J.D., Rees, G., 2006. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. 7, 523–534.

Haynes, J.D., Sakai, K., Rees, G., Gilbert, S., Frith, C., Passingham, R.E., 2007. Reading hidden intentions in the human brain. Curr. Biol. 17, 323–328.

Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V., 1999. Distributed representation of objects in the human ventral visual pathway. Proc. Natl. Acad. Sci. U. S. A. 96, 9379–9384.

Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685.

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.

Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L., 2008. Identifying natural images from human brain activity. Nature 452, 352–355.
Kiani, R., Esteky, H., Mirpour, K., Tanaka, K., 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J. Neurophysiol. 97, 4296–4309.
Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Bandettini, P.A., 2008a. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60, 1126–1141.
Kriegeskorte, N., Mur, M., Bandettini, P., 2008b. Representational similarity analysis — connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4.
McCarthy, G., Puce, A., Gore, J.C., Allison, T., 1997. Face-specific processing in the human fusiform gyrus. J. Cogn. Neurosci. 9, 605–610.
Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.-M., Malave, V.L., Mason, R.A., Just, M.A., 2008. Predicting human brain activity associated with the meanings of nouns. Science 320, 1191–1195.
Mur, M., Bandettini, P.A., Kriegeskorte, N., 2009. Revealing representational content with pattern-information fMR: an introductory guide. Soc. Cogn. Affect. Neurosci. 4, 101–109.
Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M., Gallant, J.L., 2009. Bayesian reconstruction of natural images from human brain activity. Neuron 63, 902–915.
Nishimoto, S., Vu, A.T., Naselaris, T., Bejamini, Y., Yu, B., Gallant, J.L., 2011. Reconstructing visual experience from brain activity evoked by natural movies. Curr. Biol. 21, 1–6.
Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10, 424–430.
O'Toole, A.J., Jiang, F., Abdi, H., Haxby, J.V., 2005. Partially distributed representations of objects and faces in ventral temporal cortex. J. Cogn. Neurosci. 17, 580–590.
O'Toole, A.J., Jiang, F., Abdi, H., Pénard, N., Dunlop, J.P., Parent, M.A., 2007. Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. J. Cogn. Neurosci. 19, 1735–1752.
Peelen, M.V., Atkinson, A.P., Vuilleumier, P., 2010. Supramodal representations of perceived emotions in the human brain. J. Neurosci. 30, 10127–10134.
Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45 (Suppl. 1), S199–S209.
Puce, A., Allison, T., Asgari, M., Gore, J.C., McCarthy, G., 2006. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. J. Neurosci. 16, 5205–5215.
Tong, F., Pratte, M.S., 2012. Decoding patterns of human brain activity. Annu. Rev. Psychol. 63, 483–509.