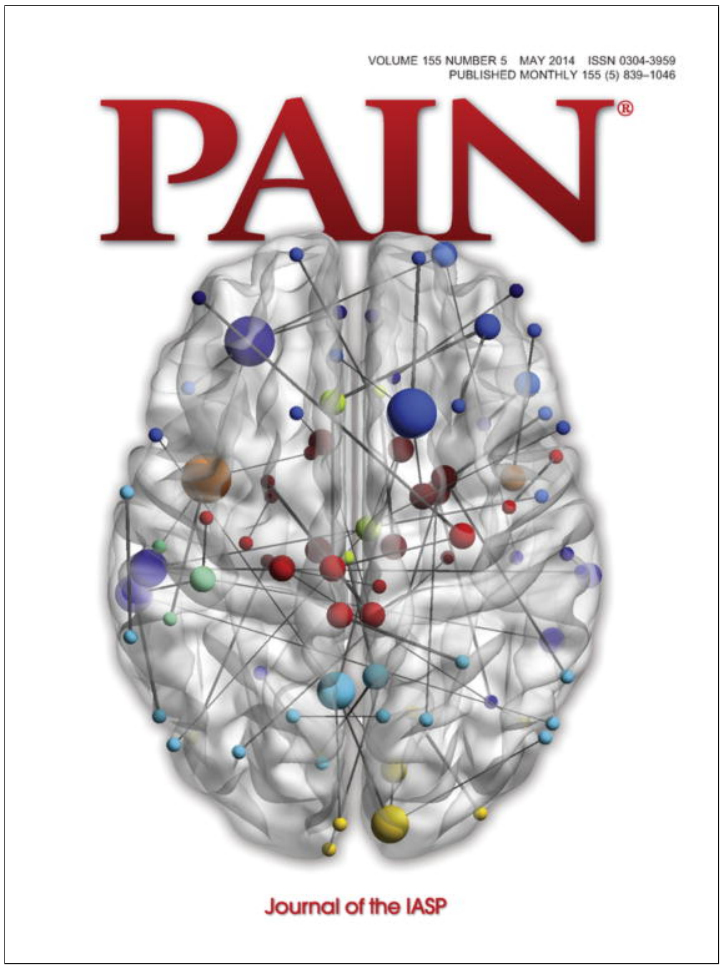

PAIN® 155 (2014) 864–867

PAIN®

www.elsevier.com/locate/pain

Topical review

Decoding the matrix: Benefits and limitations of applying machine learning algorithms to pain neuroimaging

Maria Joao Rosa [a,b,c,*], Ben Seymour [b,d]

[a] Centre for Computational Statistics and Machine Learning, Computer Science Department, University College London, London, UK
[b] Center for Information and Neural Networks, National Institute of Information and Communications Technology, Osaka, Japan
[c] Department of Neuroimaging, Centre for Neuroimaging Sciences, Institute of Psychiatry, King's College London, London, UK
[d] Computational and Biological Learning Lab, Department of Engineering, Cambridge University, Cambridge, UK

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

1. Introduction

A unique and puzzling observation about pain is its distributed processing amongst multiple cortical and subcortical brain areas, informally called the pain matrix. This is challenging and frustrating—challenging because it puts the onus on neuroscientists to determine how each area and its interactions contribute to pain processing, but frustrating given that no individual area can act as an objective biomarker for subjective pain. Recently sophisticated new analytical methods that can decode complex patterns in data offer potentially more promising ways to interrogate pain related brain activity. Here we consider whether and how these methods might contribute to basic and clinical pain neuroscience, and whether their limitations outweigh their promise.

2. Principles of decoding

In principle, any physiological measure that correlates with pain, such as heart rate, can be used for prediction (decoding). The problem is that most measures lack sensitivity and specificity, such that we might need to pool the predictive capabilities of multiple data. Neuroimaging techniques, such as functional magnetic resonance imaging (fMRI) and magneto-/electroencephalography (M/EEG) allow noninvasive recording of brain signals from a large number of spatially distributed sensors (eg, voxels/electrodes) and time points. Traditionally, these data have been analysed using simple linear regression models, treating the signals from each sensor independently [8]. Accordingly, the approach has been to predict signal changes, averaged over trials/subjects, at the level of individual sensors from a set of known variables (eg, experimental conditions). Such analyses are called univariate: the analysis of one sensor does not affect the analysis of any other (Fig. 1A). Although univariate analyses have proven powerful to test which brain regions respond to different aspects of pain processing, they do not capture the complex spatiotemporal dynamic between regions that characterises brain function.

By contrast, multivariate statistical analyses do not treat each sensor independently and are in principle more suitable than univariate methods for assessing information encoded in spatial and temporal dependencies among regions. For this reason, the use of multivariate techniques for neuroimaging, in particular multivariate pattern analyses (MVPA), also known as brain decoding [13,25], has been growing over recent years [19]. Brain decoding is usually based on machine learning (ML) techniques [3]. ML is a branch of artificial intelligence concerned with learning from data and focussed on making predictions. ML has an extremely broad application domain, from computer vision to stock market analysis. In neuroimaging, ML tries to identify spatial and/or temporal patterns in the data from multiple sensors that best discriminate between 2 or more conditions (classification) or predict continuous variables (regression) on unobserved data. Brain decoding therefore seems particularly well suited for clinical applications, having the potential to identify diagnostic biomarkers for different medical conditions [14] and for the development of brain–machine interfaces [24].

3. Decoding methodology

ML-based decoding analysis of brain imaging data entails a series of steps that include feature extraction and selection, training and testing, and pattern interpretation (Fig. 1B).

Feature extraction consists of selecting the data (features) that will be used to learn the ML model. In neuroimaging, it is common to treat the sensors' data as features and use whole-brain signals as input to the model [21]. An alternative is to first run a sensorwise

* Corresponding author at: Department of Neuroimaging, Centre for Neuroimaging Sciences, Institute of Psychiatry, King's College London, London, UK. Tel.: +44 (0) 20 3228 3066.
E-mail address: maria.rosa@kcl.ac.uk (M.J. Rosa).

http://dx.doi.org/10.1016/j.pain.2014.02.013

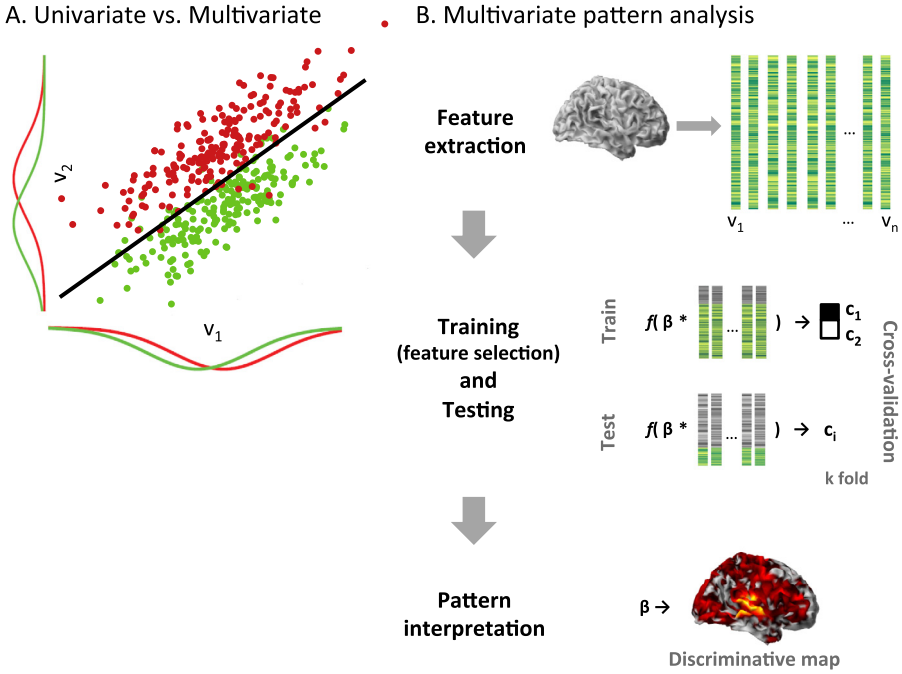

Fig. 1. (A) Toy example of the response of 2 sensors (v_1 and v_2) for 2 different conditions (green and red). If each sensor is analysed independently (univariate analysis), it is not possible in this case to reliably distinguish between the 2 conditions. Distributions of the responses of each sensor for the 2 conditions (green and red curves) overlap significantly (particularly for v_1). However, using a linear classifier (multivariate pattern analysis or brain decoding), it is possible to estimate a linear function (black line) that can reliably discriminate between the 2 conditions using data from the 2 sensors simultaneously. (B) Multivariate pattern analysis (brain decoding) steps. These steps include feature extraction and possibly selection (extracting and selecting a subset of brain signals from sensors v_1 to v_n) as well as training and testing a machine learning (ML) model using cross-validation (CV). It is important to note here that when there is no a priori reason for choosing a particular subset of features, feature selection should be performed using training data only (ie, within CV) to avoid double dipping (ie, the use of both training and testing data to select relevant features), which can strongly bias estimates of the generalization error [16]; for a comprehensive discussion on circularity in feature selection, see [16]. The k-fold CV breaks the data into k blocks of observations. The model is then trained using $k-1$ blocks and tested on the remaining block. This procedure is repeated k times. Common choices for k are $k = 5$ and 10. When k equals the total number of samples, it is called leave-one-out CV (LOOCV). LOOCV is approximately unbiased for the true (expected) prediction error because the data are trained on all observations minus 1. However, because the training sets are almost identical, the predictions are highly correlated, and therefore the LOOCV prediction error estimate tends to have higher variance than does the error estimate from k-fold CV. CV can also be used for choosing the value of tuning parameters from the ML algorithm or feature selection approach, as long as we hold out a validation set (independent from the training and testing data used for CV) to assess the model generalization error. However, in situations where there are not enough data, model selection and parameter tuning can be done in an inner CV performed on every training set of the main CV (known as nested CV). This approach can, however, be computationally expensive. We illustrate here a linear classifier (classification model) where the predictions of 2 classes, c_1 (eg, chronic pain patients) and c_2 (eg, healthy control), are given by a linear combination of the input features, ie, $f(\beta \times \text{features}) \rightarrow c_i$, where β are the model parameters, and grey regions indicate the blocks of data not being used in this cross-validation fold. The final step refers to interpreting the predictive patterns by displaying them, for example, in the form of discriminative maps. Note that the model predictions are based on all features.

univariate analysis and then use the estimated regression coefficients as features [9]. More recently, connectivity-based features, such as correlation between time series of different regions, have also been used [2,1].

Feature selection is an optional step where nonmeaningful/noisy features can be precluded from the analysis. Feature selection can rely on dimensionality reduction methods [22] or on prior knowledge (eg, regions of interest) [12]. It can also rely on measures such as the predictive accuracy/error of the model [10] or be part of the model itself [30].

After extracting features, the next step is to train and test the model. By training, we mean learning the parameters of the model using data. By testing, we mean estimating the model's extrasample/prediction error, ie, the average generalization error when the algorithm is applied to an independent test sample from the same distribution of the training data (eg, a different sample of subjects or experimental conditions). A common approach to do this is to use cross-validation (Fig. 1B). Cross-validation uses part of the available data to train the model and part to test it. It is important to emphasize that if the test set data are not independent from the training data, the cross-validation error does not reflect the true generalization error of the model. The choice of algorithm depends on the problem, and different properties—such as its predictive accuracy, interpretability, and reproducibility—may be more or less important depending on the application domain [20].

Finally, in neuroscience, it is important to assess the role that each feature played in the predictions (pattern interpretation). For linear algorithms, it is easy to represent the parameters of the model in the original data space and to visualize the importance of each feature (magnitude of its parameter) in what is called a discriminative map (Fig. 1B).

4. Decoding pain

The first study to use brain decoding in the context of pain [18] demonstrated the feasibility of using ML on whole-brain fMRI data from healthy individuals to predict self-reported thermal pain. Using similar approaches to decode different aspects of pain perception, Brodersen et al. [4] demonstrated that the joint activation of a set of brain regions, including the pain matrix, provided more accurate trial-by-trial predictions of thermal pain perception than any single region. The spatial patterns of activity in insular and cingulate cortex have been found to highly overlap for pain felt on one's hand and observed on another person's hand [7]. Cecchi et al. [6] illustrated the advantage of combining ML and psychophysics to predict the temporal evolution of thermal pain perception from within-subject fMRI data. Despite these successes, the translation of brain decoding into real-world clinical applications depends on its ability to make between-subject predictions. One of the first studies to address this issue [5] showed that ML

algorithms trained on whole-brain patterns of activity from a group of subjects could discriminate painful from nonpainful thermal stimulation on an independent set of individuals. In addition, Prato et al. [26] proposed a regression algorithm based on fMRI for making between-subject predictions of pain intensity following an injection of ascorbic acid. Schulz et al. [31] showed that ML models trained on EEG-based time-frequency patterns from a group of subjects can predict the sensitivity of a new individual to cutaneous pain. Perhaps the most comprehensive demonstration of an objective neural code of subjective phasic pain was from Wager et al. [35]. The authors first identified patterns of brain activity that could distinguish between heat-induced painful from nonpainful stimulation in healthy subjects. In 3 independent groups of subjects, these patterns were then used to successfully discriminate painful heat from nonpainful warmth and physical from social pain, and measure the effect of analgesics. More recently, Liang et al. [17] showed how pain can be distinguished from other sensory stimuli which also involve activation in areas of the classical pain matrix. Finally, Ung et al. [33] demonstrated that decoding can go beyond predicting acute pain in healthy subjects, using MRI-based measures of grey-matter volume to successfully discriminate patients with chronic low back pain from controls.

5. Neurobiological considerations

Pain neurobiology is generally concerned with the question of pain encoding (ie, how the nervous system gives rise to the percept of pain); despite the impressive results of decoding experiments to date, how much they inform our understanding of pain is less clear [23]. One of the problems is knowing what is being decoded. Depending on the model and data, ML models can be powerful and robust. When used as black boxes, they will maximize the predictive accuracy given any property of the data, whether it is biologically meaningful and/or interpretable (eg, condition-related neuronal activity) or not (eg, artefacts). This issue is particularly relevant for neuroscience applications, where it is important to understand which brain regions/features contain information about the experimental condition. In fact, although it is possible to create sensorwise maps from multivariate model parameters (Fig. 1B), spatial inferences on these maps are not straightforward. Contrary to univariate models, multivariate maps do not naturally provide a null hypothesis (and corresponding statistical test) associated with each sensor/feature, and therefore information mapping usually relies on computationally expensive techniques, such as permutation tests [21,27] and local sensorwise multivariate (searchlight) approaches [15]. The parameters are also dependent on the characteristics of the algorithm (eg, regularization) and data (eg, signal-to-noise ratio and brain physiology), which further hinders the interpretation of multivariate maps [11]. Even though the latter issues also affect the univariate approach to some extent, the success of decoding methods for neuroscience needs further investigation and possibly confirmation from lesion/invasive data.

6. Clinical applications

Clinical applications are often argued to have the greatest need for MVPA-based decoding (Fig. 2). In a sense, the issue about what is being decoded is less problematic for diagnostic purposes, as the problem is defined in terms of clinical outcomes, not biological mechanisms. However, there are many other issues looming: to be useful, an MVPA-based biomarker will need to differentiate chronic pain from comorbidities (eg, depression) and previous treatments (eg, opiates), and predict disease severity as well as spontaneous and treatment-induced remission. Ideally, it would also distinguish different aetiologies of chronic pain and predict future outcomes [32,1]. Methodologically, the greatest problem may be operational: to be clinically useful, an MRI-based biomarker would need to be independent of geographic site, scanner type, and operator use. So whereas the existing decoding results are highly encouraging, the challenges ahead for a clinically usable biomarker are enormous.

A more realistic application might be in clinical trials, in which many of the above factors can be controlled. Indeed, MVPA could in principle be applied to both human and animal models of chronic pain used for drug development. As well as providing new and robust biologically based outcome measures, it might even be possible to address other issues, such as differentiating placebo from true therapeutic effects.

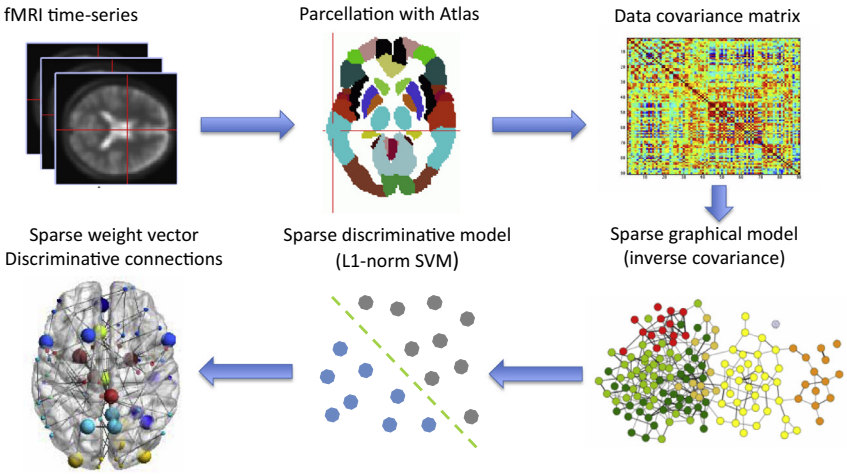

Fig. 2. Using multivariate pattern analyses for diagnostic classification. In principle, imaging-based diagnostic classification can be based either on studying brain responses to phasic pain stimuli or on baseline activity. This example is based on using baseline brain connectivity to classify patients and control subjects; it is very similar to the method we have previously used to successfully classify patients with major depressive disorder [28]. Subjects are scanned with functional magnetic resonance imaging to produce standard voxel-based time series of activity. The brain is then parcellated into regions of interest (ROIs) using a standardized brain atlas. The average activity in each ROI is covaried with every other to generate a covariance matrix of pairwise connectivities. To reduce the high dimensionality of the feature space, a sparse graphical model (gLASSO) can be used to estimate sparse pairwise partial correlations (inverse covariances), which become the features used in decoding. The sensitivity maps in this case correspond to the brain networks that best discriminate between patients and healthy subjects.

7. Neuroengineering applications

There are broader potential uses of decoding—for example, in the development of brain–machine interfaces for pain. However, these neuroengineering applications need to solve the difficult problem of real-time decoding [34], in which spontaneous (not event locked) changes in pain are decoded. If this can be achieved, one could, for example, develop communication systems for those unable to reliably report pain, such as patients with impaired consciousness and infants. Decoding methods could also be useful for closed-loop neuromodulation, whereby stimulation (eg, thalamic stimulation) is directly tied to decoded subjective pain. This is useful for 2 reasons: first, it optimizes the timing and level of stimulation (avoiding under- or overstimulation). Second, with more complex neuromodulatory methods, which have multiple stimulation parameters (eg, current frequency/amplitude), it allows search for optimal parameters to be computerized.

Clearly the logistical limitations of current neuroimaging methods (ie, fMRI and MEG) constrain their use to proof-of-principle therapeutic neuroengineering applications. At present, there are no successful decoding examples using practically implementable systems (eg, wireless EEG/functional near-infrared spectroscopy [fNIRS]), and it is possible that the best signals will be either too deep for surface-based recording or will not require multivariate decoding at all, as is also the case for closed-loop deep brain stimulation for Parkinson's disease [29].

8. Conclusion

The appeal of MVPA-based decoding is that it intuitively captures the multivariate nature of how the brain processes information—something that has particular resonance with current concepts of pain processing. Not only does this offer a formalization of the concept of a decodable brainwise pain matrix, but it also has allowed identification of the most accurate biological signature of subjective pain to date. This leads to justifiable optimism about clinical diagnostic/prognostic applications. However, the problem of what is being decoded plagues their interpretability from a neurobiological perspective and arguably leaves us little closer to understanding how chronic pain is processed in the brain. With this in mind, it is likely that data-driven multivariate approaches will gradually give way entirely to hypothesis-driven multivariate approaches that mechanistically describe network-level pain processing.

Conflict of interest statement

The authors report no conflict of interest.

Acknowledgements

We thank Lauren Atlas and Shinji Nishimoto for their helpful comments. Maria Joao Rosa is funded by a Japanese Society for the Promotion of Science (JSPS) fellowship. Ben Seymour is funded by the National Institute of Communications Technology (NICT) Japan and the Wellcome Trust (UK).

References

[1] Baliki MN, Bogdan P, Torbey S, Herrmann KM, Huang L, Schnitzer TJ, Fields HL, Apkarian AV. Corticostriatal functional connectivity predicts transition to chronic back pain. Nat Neurosci 2012;15:1117–9.
[2] Billinger M, Brunner C, Müller-Putz GR. Single-trial connectivity estimation for classification of motor imagery data. J Neural Eng 2013;10:046006.
[3] Bishop CM, Nasrabadi NM. Pattern recognition and machine learning. New York: Springer; 2006.
[4] Brodersen KH, Wiech K, Lomakina EI, Lin CS, Buhmann JM, Bingel U, Ploner M, Stephan KE, Tracey I. Decoding the perception of pain from fMRI using multivariate pattern analysis. Neuroimage 2012;63:1162–70.
[5] Brown JE, Chatterjee N, Younger J, Mackey S. Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation. PloS One 2011;6:e24124.
[6] Cecchi GA, Huang L, Hashmi JA, Baliki M, Centeno MV, Rish I, Apkarian AV. Predictive dynamics of human pain perception. PLoS Comput Biol 2012;8:e1002719.
[7] Corradi-Dell'Acqua C, Hofstetter C, Vuilleumier P. Felt and seen pain evoke the same local patterns of cortical activity in insular and cingulate cortex. J Neurosci 2011;31:17996–8006.
[8] Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Map 1994;2:189–210.
[9] Fu CH, Mourao-Miranda J, Costafreda SG, Khanna A, Marquand AF, Williams SC, Brammer MJ. Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. Biol Psychiatry 2008;63:656–62.
[10] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3:1157–82.
[11] Haufe S, Meinecke F, Görgen K, Dähne S, Haynes JD, Blankertz B, Bießmann F. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 2014;87:96–110.
[12] Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 2001;293:2425–30.
[13] Haynes JD, Rees G. Decoding mental states from brain activity in humans. Nat Rev Neurosci 2006;7:523–34.
[14] Klöppel S, Abdulkadir A, Jack Jr CR, Koutsouleris N, Mourão-Miranda J, Vemuri P. Diagnostic neuroimaging across diseases. Neuroimage 2012;61:457–63.
[15] Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. Proc Natl Acad Sci U S A 2006;103:3863–8.
[16] Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 2009;12:535–40.
[17] Liang M, Mouraux A, Hu L, Iannetti GD. Primary sensory cortices contain distinguishable spatial patterns of activity for each sense. Nat Commun 2013;4:1979.
[18] Marquand A, Howard M, Brammer M, Chu C, Coen S, Mourão-Miranda J. Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. Neuroimage 2010;49:2178–89.
[19] McIntosh AR, Mivsic B. Multivariate statistical analyses for neuroimaging data. Annu Rev Psychol 2013;64:499–525.
[20] Misaki M, Kim Y, Bandettini PA, Kriegeskorte N. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. Neuroimage 2010;53:103–18.
[21] Mourão-Miranda J, Bokde AL, Born C, Hampel H, Stetter M. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. Neuroimage 2005;28:980–95.
[22] Mourao-Miranda J, Reynaud E, McGlone F, Calvert G, Brammer M. The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data. Neuroimage 2006;33:1055–65.
[23] Naselaris T, Kay KN, Nishimoto S, Gallant JL. Encoding and decoding in fMRI. Neuroimage 2011;56:400–10.
[24] Nicolas-Alonso LF, Gomez-Gil J. Brain computer interfaces: a review. Sensors 2012;12:1211–79.
[25] Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 2006;10:424–30.
[26] Prato M, Favilla S, Zanni L, Porro CA, Baraldi P. A regularization algorithm for decoding perceptual temporal profiles from fMRI data. Neuroimage 2011;56:258–67.
[27] Rasmussen PM, Hansen LK, Madsen KH, Churchill NW, Strother SC. Model sparsity and brain pattern interpretation of classification models in neuroimaging. Pattern Recognit 2012;45:2085–100.
[28] Rosa MJ, Portugal L, Shawe-Taylor J, Mourao-Miranda J. Sparse network-based models for patient classification using fMRI. In: Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on, IEEE 2013. p. 66–9.
[29] Rosin B, Slovik M, Mitelman R, Rivlin-Etzion M, Haber SN, Israel Z, Vaadia E, Bergman H. Closed-loop deep brain stimulation is superior in ameliorating parkinsonism. Neuron 2011;72:370–84.
[30] Ryali S, Supekar K, Abrams DA, Menon V. Sparse logistic regression for whole-brain classification of fMRI data. Neuroimage 2010;51:752–64.
[31] Schulz E, Zherdin A, Tiemann L, Plant C, Ploner M. Decoding an individual's sensitivity to pain from the multivariate analysis of EEG data. Cereb Cortex 2012;22:1118–23.
[32] Tracey I. Can neuroimaging studies identify pain endophenotypes in humans? Nat Rev Neurol 2011;7:173–81.
[33] Ung H, Brown JE, Johnson KA, Younger J, Hush J, Mackey S. Multivariate classification of structural MRI data detects chronic low back pain. Cereb Cortex 2014;24:1037–44.
[34] Van Gerven MA, Kok P, de Lange FP, Heskes T. Dynamic decoding of ongoing perception. Neuroimage 2011;57:950–7.
[35] Wager TD, Atlas LY, Lindquist MA, Roy M, Woo CW, Kross E. An fMRI-based neurologic signature of physical pain. N Engl J Med 2013;368:1388–97.