Journal of Neuroscience Methods xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Computational Neuroscience

Visualization and unsupervised predictive clustering of high-dimensional multimodal neuroimaging data

Benson Mwangi [a,*], Jair C. Soares [a], Khader M. Hasan [b]

[a] UT Center of Excellence on Mood Disorders, Department of Psychiatry and Behavioral Sciences, UT Houston Medical School, Houston, TX, USA
[b] The University of Texas Health Science Center at Houston, Department of Diagnostic & Interventional Imaging, Houston, TX, USA

HIGHLIGHTS

- A multi-modal neuroimaging approach is used to investigate the utility of *t*-distributed stochastic neighbour embedding (*t*-SNE) in identifying 'unseen' population patterns.
- Method able to detect gender related brain differences in a healthy population.
- Ability to detect relevant patterns improves with additional imaging modalities.
- An *unsupervised* predictive classifier developed from detected clusters able to predict *individual* subject's gender with high accuracy.

ARTICLE INFO

Article history:
Received 27 May 2014
Received in revised form 30 July 2014
Accepted 1 August 2014
Available online xxx

Keywords:
Multimodal neuroimaging
Unsupervised machine learning
t-Distributed stochastic neighbour embedding (*t*-SNE)
Big data
Research domain criteria

ABSTRACT

Background: Neuroimaging machine learning studies have largely utilized *supervised* algorithms – meaning they require both neuroimaging scan data and corresponding target variables (e.g. healthy vs. diseased) to be successfully 'trained' for a prediction task. Noticeably, this approach may not be optimal or possible when the global structure of the data is not well known and the researcher does not have an a priori model to fit the data.
New method: We set out to investigate the utility of an *unsupervised* machine learning technique; *t*-distributed stochastic neighbour embedding (*t*-SNE) in identifying 'unseen' sample population patterns that may exist in high-dimensional neuroimaging data. Multimodal neuroimaging scans from 92 healthy subjects were pre-processed using atlas-based methods, integrated and input into the *t*-SNE algorithm. Patterns and clusters discovered by the algorithm were visualized using a *2D* scatter plot and further analyzed using the *K*-means clustering algorithm.
Comparison with existing methods: *t*-SNE was evaluated against classical principal component analysis.
Conclusion: Remarkably, based on unlabelled multimodal scan data, *t*-SNE separated study subjects into two very distinct clusters which corresponded to subjects' gender labels (cluster silhouette index value = 0.79). The resulting clusters were used to develop an *unsupervised* minimum distance clustering model which identified 93.5% of subjects' gender. Notably, from a neuropsychiatric perspective this method may allow discovery of data-driven disease phenotypes or sub-types of treatment responders.

1. Introduction

Machine learning techniques have recently shown promise in decoding *individual* subjects' brain state using high-dimensional neuroimaging scan data. Notable applications include; diagnostic or categorical predictions (e.g. healthy vs. diseased) (Doehrmann et al., 2013; Ecker et al., 2010a,b; Johnston et al., 2014; Mwangi et al., 2012, 2014; Zeng et al., 2012), and predictions of continuous demographic, cognitive and clinical variables (e.g. illness severity scores or chronological age) (Dosenbach et al., 2010; Feis et al., 2013; Marquand et al., 2010; Mwangi et al., 2012, 2013a; Wang et al., 2010). In addition, applications utilizing these techniques to fuse multi-scale biological measurements (e.g. genetics and imaging) to make *individualized* brain state predictions have also been reported (Brodersen et al., 2014; Kohannim et al., 2012; Mwangi et al., 2013b; Wang et al., 2012; Ziegler et al., 2012).

* Corresponding author at: Department of Psychiatry and Behavioral Sciences, The University of Texas Health Science Center, 1941 East Road, Houston, TX 77054, USA. Tel.: +1 7134862507; fax: +1 7134862507.
E-mail addresses: benson.irungu@uth.tmc.edu, benson.mwangi@gmail.com (B. Mwangi).

http://dx.doi.org/10.1016/j.jneumeth.2014.08.001
0165-0270/

However, whilst these brain decoding studies have significantly advanced our understanding of neuropsychiatric disorders and even set pace for potential diagnostic and prognostic applications, this has not been without practical limitations. First, a majority of these studies have noticeably been '*supervised*' – meaning target labels (e.g. healthy vs. disease) are used in 'training' the classifier. Second, *supervised* techniques are not capable of elucidating natural groupings of data points in a high-dimensional space without making a priori hypotheses (Jain, 2010; Mwangi et al., 2013b). Third, *supervised* techniques are largely confirmatory – meaning that the researcher wants to confirm the validity of a hypothesis or set of assumptions (Jain, 2010). For example, a popular assumption in neuroimaging predictive classification studies (e.g. health vs. disease) is that brain differences (e.g. structure or function) between 'healthy' and 'diseased' groups may allow accurate predictions at an *individual* subject level.

In view of the above, although *supervised* techniques have made significant contributions in elucidating the pathophysiology of neuropsychiatric disorders (Johnston et al., 2012; Linden, 2012; Mwangi et al., 2012); this has not been without criticism. For example, Savitz and Colleagues (2013) recently raised doubts on the popular assertion that *supervised* machine learning applications in neuroimaging may lead to objective biomarkers of psychiatric disorders given that these algorithms are often 'trained' using target labels (e.g. patient vs. control) largely derived from subjective clinical diagnoses (e.g. diagnostic and statistical manual – DSM-IV or DSM 5). Notable exemptions include those utilizing post-mortem ground-truth labels to train and test a machine learning model (Klöppel et al., 2008).

Conspicuously, *unsupervised* machine learning techniques are practical alternatives which may allow discovery of biologically relevant groupings or clusters in the data without requiring *user defined* target labels. These clusters may correspond to 'new' data-driven phenotypic subtypes or demographic groupings and there is active research work in this area. For example Aljabar et al. (2011) utilized an unsupervised manifold learning approach to characterize neonatal brain development. Most recently, Brodersen and Colleagues (2014) reported an unsupervised machine-learning proof-of-concept study examining the feasibility of defining subgroups in psychiatric spectrum disorders using generative embedding techniques. Fair and Colleagues (2012) utilized a graph theory approach to identify unique data-driven neuropsychological subgroups in children with attention deficit hyperactivity disorder (ADHD).

In this report we studied the utility of an *unsupervised* machine learning technique *t*-distributed stochastic neighbour embedding (*t*-SNE) (Van der Maaten and Hinton, 2008) in analyzing high-dimensional multimodal neuroimaging scan data. The main objectives of this study were; first to investigate *t*-SNE's ability to reduce the dimensionality of multimodal neuroimaging scan data into a visually plausible 2D space. Second, to investigate *t*-SNE's ability to reveal natural groupings or clusters in the study sample whilst concurrently capturing the local structure of the neuroimaging scan data (e.g. subjects' anatomical differences). Notably, *t*-SNE is an '*unsupervised*' data-driven technique which translates a high-dimensional dataset into a pair-wise similarity matrix whilst simultaneously capturing both local and global structure of the data (e.g. population clusters) (Van der Maaten and Hinton, 2008).

Specific benefits of using *t*-SNE over traditional dimensionality reduction or manifold learning techniques such as principal component analysis (PCA) (Jolliffe, 1972), multidimensional scaling (MDS) (Kruskal, 1964) and local linear embedding (LLE) (Roweis and Saul, 2000) should be noted. First, PCA and MDS are linear transformations of the high-dimensional data and may not always be able to capture non-linear relationships in a high-dimensional dataset (Amir et al., 2013; Bishop, 1995; Ji, 2013; Mwangi et al., 2013b; Van der Maaten and Hinton, 2008). Second, a majority of dimensionality reduction techniques are not capable of retaining both *local* and *global* structures of the data simultaneously during the dimensionality reduction process (Van der Maaten and Hinton, 2008). Most notably, *t*-SNE recently showed a superior performance in visualizing subjects' neuroimaging scan data from a multi-centre study as compared to MDS (Ridgway and Ashburner, 2012; Ridgway et al., 2012). However, for reviews and subsequent empirical evaluations of these dimensionality reduction techniques, the reader is pointed elsewhere (Lee and Verleysen, 2007; Mwangi et al., 2013b; Van der Maaten and Hinton, 2008).

In this study, high-dimensional multimodal neuroimaging scan data (T_1-weighted, diffusion tensor imaging, T_2-relaxation time and proton density) from 92 healthy subjects were pre-processed, fused and input into the *t*-SNE algorithm. The algorithm returned a new set of variables for each subject in a 'reduced' 2D space. The *K*-means clustering algorithm and a quantitative cluster evaluation metric (silhouette width index value) were used to evaluate population clusters present in the new 2D space. A major objective here was to elucidate whether the technique is able to discover biologically relevant clusters from neuroimaging scan data. Markedly, a major motivation for utilizing multimodal scan data was to establish whether neuroimaging data from multiple modalities would lead to improved detection of relevant sub-groups or clusters.

2. Materials and methods

This study was approved by the University of Texas Health Science Center at Houston local institute review board (IRB) and was compliant with the Health Insurance Portability Accountability Act (HIPAA) guidelines. A total of 92 (44 Males, 48 Females) healthy subjects with age ranging from 18.7 to 61.8 years; mean/SD = 37.09/10.55 and identified as neurologically normal before scanning were included in this study. There were no significant age differences between genders.

2.1. MRI acquisition protocol

Subjects were scanned using a 3.0 T Philips Intera scanner. T_1-weighted scans were acquired using a 3D-spoiled gradient-echo with a field-of-view = $240 \times 240\,mm^2$, isotropic voxel size = 0.94 mm, and 2-dimensional dual spin echo images ($T_{E1}/T_{E2}/T_R$ = 10/90/5000 ms), and in the axial plane with 3 mm slice thickness. Diffusion-weighted data were acquired using a single-shot spin echo diffusion sensitized echo-planar imaging (EPI) sequence with a balanced *icosa21* encoding scheme as demonstrated elsewhere (Hasan et al., 2007; Walimuni et al., 2011). The diffusion-weighted sequence had a sensitization of $b = 1000\,s\,mm^{-2}$, repetition and echo times of T_R = 6.1 s and T_E = 84 ms respectively with a slice thickness of 3 mm and 44 axial slices covering the whole brain. Similarly, the sequence had a square field-of-view = $240 \times 240\,mm^2$, an image matrix of 256×256 and 8 non-diffusion weighted images.

2.2. Image pre-processing and feature extraction

The image pre-processing pipeline followed in this study is shown in Fig. 1 and also detailed elsewhere (Walimuni et al., 2011; Walimuni and Hasan, 2011).

T_1-weighted scans were automatically segmented into both cortical and subcortical anatomical regions using the Freesurfer software library Version 5.0 (http://surfer.nmr.mgh.harvard.edu/) (Fischl, 2012) by following an atlas of 169 anatomical regions also available with Freesurfer (Desikan et al., 2006). These anatomical regions included superior temporal gyrus, middle temporal gyrus,

Please cite this article in press as: Mwangi B, et al. Visualization and unsupervised predictive clustering of high-dimensional multimodal neuroimaging data. J Neurosci Methods (2014), http://dx.doi.org/10.1016/j.jneumeth.2014.08.001