Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2011 March 30; : 1117–1120. doi:10.1109/ISBI.2011.5872597.

BRAIN PATTERN ANALYSIS OF CORTICAL VALUED DISTRIBUTIONS

Shantanu H. Joshi, Ian Bowman, Arthur W. Toga, and John D. Van Horn

Laboratory of Neuro Imaging, Department of Neurology, University of California, Los Angeles, 90095, USA

Abstract

We introduce a new representation of cortical regions via distribution functions of their features. The distribution functions are estimated non-parametrically from the data and are observed to be non Gaussian. Cortical pattern matching is enabled by using the information-based Jensen-Shannon divergence as a measure between features. Our approach explicitly avoids pairwise registrations between brains, but instead focuses on modeling and discriminating between the cortical structural patterns. We demonstrate our approach on 120 subject brains from an Alzheimer's dataset, and present applications to clustering, classification, and dimension reduction.

Index Terms

cortical distributions; Jensen-Shannon divergence; clustering; dimension reduction

1. INTRODUCTION AND BACKGROUND

Brain mapping studies are concerned with localization and evaluation of structural changes caused by various developmental or even disease processes. Cortical morphometry is a branch of brain mapping that studies changes in morphology on the periphery of the cerebrum. There are several approaches for achieving this goal. Most of these approaches [1, 2] start with a geometrical representation for the cortical surface, and establish a dense correspondence of homologous regions and points across a group of subjects. This correspondence is further used to drive the construction of a population-based atlas that maps all the subjects in a common reference frame. Accurate registration methods allow the reconciliation of various brain measures and enable consistent inter-subject comparisons both within and across different populations.

In this paper we follow a different approach. Instead of focusing on local geometric changes, and spatial registration, we attempt to represent and analyze global structural patterns on the cortex via continuous density functions that capture the variation of different cortical valued features. Our goal here is not to achieve a precise localized map of brain changes, but rather discover a set of pertinent global features that attempt to model the overall structural patterns. Lately, there is an emerging interest in understanding the underlying low dimensional structure at a population level [3, 4, 5]. This idea is also exploited in Dalal et al. [6], who propose a shape-based approach for cortical similarity. We are interested in extracting functions of cortical features that serve as compact representations of the underlying brain populations. In this paper, we focus on the cortex, and its regional segmentations in particular, and represent different parcellations using distributions of their features. These distributions at a cortical level, implicitly capture the global location of the feature, and at a regional level capture the underlying trend of the feature valued data.

Additionally we use the Jensen-Shannon (JS) divergence [7] for discriminating between the feature functions. There are several advantages to the proposed approach. i) The distribution functions are invariant to brain scale, translation, and orientation, ii) they are easily and efficiently computable, and iii) finally this approach does not require pairwise registrations, and hence can be effectively applied to sizable populations. The rest of the paper is organized as follows. We introduce the feature function representation in Sec. 2.1, and present the cortical matching approach in Sec. 2.2. Section 3 presents applications of this method to clustering, classification, and dimension reduction on a population from the ADNI dataset [8], followed by discussion and conclusion.

2. MATERIALS AND METHODS

2.1. Representation of Cortical Features as Distributions

We represent cortical regions by distributions of their continuous feature values. These features are typically derived from projecting biological measures such as volume, or thickness on the cortex, or calculated using intrinsic properties of the cortical surface, and can be as diverse as shape index, curvedness, Gaussian or the mean curvature respectively. One can either define a single distribution for the entire cortex, or represent individual regions of the cortex by multiple distributions. In the first case the neighborhood information of features is merged into an all-inclusive distribution, whereas in the second case, one takes into account the regional locality by separating the feature contributions due to each area. We will follow the latter approach and use the cytoarchitectonic organization of the cortex to obtain its parcellation into different areas. The main advantage of this scheme, is that in addition to the underlying structural composition, it also reveals important aspects of functional organization of the brain. We now establish relevant mathematical notation. For a total of N regions, let each parcellation be denoted by $\mathfrak{x}_i: u \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3, i = 1, \dots, N$. Furthermore, we represent the a given cortical feature, by a mapping $f_{\mathfrak{x}_i}: u \rightarrow \mathbb{R}$. For example, if $f_{\mathfrak{x}_i}$ represents the projected cortical thickness for region i , it assumes all permissible values roughly from 0.5 to 4.5 mm. Next, we calculate the relative frequencies of the observed feature and reduce the two dimensional variation of the cortical attribute to a univariate quantity. From this frequency distribution, We then estimate a probability density function from this frequency distribution in a non-parametric manner using kernel density estimation. There are numerous approaches for estimating non-parametric densities, and the kernel bandwidth plays a major role in either over-smoothing or under-smoothing the resultant function. Recently Botev et al. [9] have proposed an adaptive approach for kernel density estimation. They solve a stochastic differential equation that models a linear diffusion process. Unlike a data-driven bandwidth selection method, they use a novel plug-in selection method that demonstrates a better performance compared to the existing methods. We use the method proposed in [9] for constructing a non-parametric density function modeling the variability in the cortical features. Thus for a single individual cortex, the set $\{p_i: [a, b] \rightarrow \mathbb{R} \mid i = 1, \dots, N\}$ represents the univariate cortical patterns that represent the feature distribution. Here the interval $[a, b]$ is the range of allowable values associated with that attribute. Figure 1 shows a schematic of the proposed distributions. It shows different regions of the cortical surfaces, and histogram of cortical thickness associated with each parcellation.

2.2. Distances between Distributions using Jensen-Shannon Divergence

In order to discriminate between cortical feature patterns, the choice of the distance between distributions becomes important. Under a simplistic approach, one can treat the distribution functions as elements of a Hilbert space, and compute Euclidean inner products to get a distance between features. On the other hand, one can use an entropy based criteria to exploit the information present in the feature distributions to perform matching. Following

this approach, we will use the Jensen-Shannon divergence [7] as a dissimilarity measure between feature functions. The JS divergence is a symmetrized version of the Kullback-Leibler divergence, and is defined as

$$J(p_k, p_m) = KL(p_k, \overline{p_{km}}) + KL(p_m, \overline{p_{km}}) \quad (1)$$

Here the quantity KL is the Kullback-Leibler divergence, and is given by

$$KL(p_k, p_m) = \int_{-\infty}^{\infty} p_k(x) \log \frac{p_k(x)}{p_m(x)} dx \quad (2)$$

The term $\overline{p_{km}}$ is simply the average of p_k , and p_m . The Jensen-Shannon divergence can be interpreted as the difference between the entropy of the average of the two distributions, and the mean of the entropies of the two distributions. Other properties of the JS divergence are $J(p_i, p_j) \geq 0$, $J(p_i, p_j) = J(p_j, p_i)$, and $J(p_i, p_j) = 0 \Leftrightarrow p_i = p_j$. The JS divergence however is not a true metric since it does not satisfy triangle inequality.

Having established a distance between feature functions, we now define a distance between two cortical patterns as follows. Given two pairs of feature functions, $\{p_k^i\}$, $\{p_k^j\}$, $k = 1, \dots, N$, the distance between patterns i , and j is given by,

$$D(i, j) = \sum_{k=1}^N w_k J(p_k^i, p_k^j) \quad (3)$$

Here w_k are the weights assigned to each divergence, and can be either optimized or adjusted according to prior information. In this paper, we assume all weights are equal to 1.

3. RESULTS AND APPLICATIONS

In this section, we present several results and applications of the above pattern representation approach. First, we describe the data set and various operations performed on the dataset below.

3.1. Data

Our data comprised of a sample of $N = 117$ subjects from the ADNI dataset [8]. Specifically they consisted of 39 Alzheimer's disease (AD) patients, 38 mild cognitively impaired (MCI) subjects, and 40 normal control (NC) subjects (mean age 76 years $\pm$ 7.7 years). All images in this example were scanned using a 3T MR scanning platform using an MPRAGE data acquisition sequence. All the MRI images were processed by skull removal, segmentation, cortical extraction, and topology correction using Freesurfer [1]. Additionally, the cortex was parcellated into 34 regions using the same tool. Several volumetric features such as gray matter thickness, volume, and surface area can be computed on the processed data. Gray matter thickness is computed as the mean distance from the gray matter/cerebrospinal fluid interface to the gray/white matter surface, and vice versa. Gray matter volume is defined as the product of the thickness and the area of the surface layer midway between the gray/CSF and white/gray matter boundaries. Lastly, a mean area measure can be calculated at each point on the cortical surface by averaging the areas of triangles including that point. In this

paper, we present results on the cortical thickness feature, since we are analyzing Alzheimer's subjects.

In the next subsections, we outline two applications that make use of the above dataset.

3.2. Clustering and Classification

In order to test the discriminative properties of our approach, we selected cortical thickness as a feature, and constructed 34 distribution functions for each of the subjects. As an example, Figure 2 shows the thickness distribution functions for a single subject. We then computed pairwise Jensen-Shannon divergences between all 117 subjects using Eq. 3, and using average linkage performed hierarchical agglomerative clustering. Figure 3 shows a dendrogram of the resulting cluster configuration. It is observed that only 20% of Alzheimer's subjects were grouped together with the normal controls, whereas nearly 8% of the normal controls were grouped together with Alzheimer's patients. This result is obtained without performing pair-wise registrations and deformation morphometry, and in a completely unsupervised manner. It is further observed that the mild cognitively impaired subjects were almost equally scattered throughout both groups. This was expected, since we are simply performing a group-wise analysis without any longitudinal component. Additionally, we also compared the clustering solution by utilizing Euclidean distances between distributions as opposed to the JS divergences (Figure 4). This time there was a reduced discriminative effect for the normal controls; nearly 40% normal controls grouped with the Alzheimer's subjects, and only 15% AD subjects were grouped with the normal controls. This is largely due to the choice of the distance between distributions.

3.3. Dimension Reduction and Visualization

From an informatics perspective [3], it is of great interest to perform statistical analysis, as well as inference on large samples of data. With reduced scan times, and the cost of storage, the size of neuro imaging datasets are increasing rapidly. The advantages of dimension reduction of such datasets are twofold. Firstly, they can help discover the manifold structure of such data. Combined with meta-data and other nonimaging information, this can be helpful in swift classification and recognition of important parameters for disease. Secondly, the dimension reduction can also help in visualization of large archives for data mining or analytics purposes. For this purpose we use the Isomap algorithm [10] that achieves non-linear dimension reduction by performing classical MDS on an adjacency matrix which approximates the geodesic distance between data points. Figure 5 shows the scatter plot of the 3D eigen-projection of the pairwise dissimilarity matrix. This projection can be rendered to display this reduced-dimensional manifold representation of the whole data and used as the basis for user interaction.

4. DISCUSSION AND CONCLUSION

We presented a pattern analysis approach for cortical valued features. Our approach avoided the need for pairwise registration for pattern matching. From the clustering results, it was observed that the Jensen-Shannon divergence exhibited better discriminative properties than the Euclidean distances. Thus the choice of the metric plays an important role in matching distribution functions. Since our data consisted of Alzheimer's subjects, we selected cortical thickness as a feature for our analysis. It should also be noted that the age variation of the subjects was in the same range, and the method was able to separate the natural atrophy of the normal control subjects, as well as the abnormal atrophy owing to the disease. We plan to explore several choices of metrics between distributions, as well as features for different types of pathology, stages of brain development, as well as demographics such as age, gender and genetics.

Acknowledgments

This work is supported in part by NIH grants RC1MH088194, U54 RR021813, P41 RR013642.

References

1. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. segmentation and surface reconstruction. *NeuroImage*. Feb; 1999 9(2):179–94. [PubMed: 9931268]
2. Thompson P, Toga AW. A surface-based technique for warping 3-dimensional images of the brain. *IEEE Tran Med Imaging*. Jan; 1996 15(4):402–417.
3. Joshi SH, Van Horn JD, Toga AW. Interactive exploration of neuroanatomical meta-spaces. *Frontiers in Neuroinformatics*. 2009; 3(38):1–10. [PubMed: 19198661]
4. Gerber S, Tasdizen T, Fletcher PT, Joshi S, Whitaker R. Alzheimers Disease Neuroimaging Initiative (ADNI). Manifold modeling for brain population analysis. *Medical Image Analysis*. Oct; 2010 14(5):643–53. [PubMed: 20579930]
5. Verma R, Khurd P, Davatzikos C. On analyzing diffusion tensor images by identifying manifold structure using isomaps. *IEEE Tran Med Imaging*. Jun; 2007 26(6):772–8.
6. Dalal P, Shi F, Shen D, Wang S. Multiple cortical surface correspondence using pairwise shape similarity. *Med Image Comput Comput Assist Interv*. Jan; 2010 13(Pt 1):349–56. [PubMed: 20879250]
7. Lin J. Divergence measures based on the shannon entropy. *IEEE Tran Information Theory*. 1991; 37(1):145–151.
8. Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, et al. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*. 2008; 27(4):685–691. [PubMed: 18302232]
9. Botev ZI, Grotowski JF, Kroese DP. Kernel density estimation via diffusion. *Ann Stat*. Jan; 2010 38(5):2916–2957.
10. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science*. Dec; 2000 290(5500):2319–23. [PubMed: 11125149]

Fig. 1. Distribution of features on a parcellated cortex. The distribution of the cortical measure (thickness in this case) is shown by histograms. (Best viewed in color).

Fig. 2.
Distribution functions of cortical thickness for 34 regions.

Fig. 3.
Dendrogram of the pairwise Jensen-Shannon divergence between thickness patterns for subjects of the ADNI dataset.

Fig. 4. Dendrogram of the pairwise Euclidean distances between thickness patterns of subjects of the ADNI dataset.

Scatter plot of MDS applied to pairwise dissimilarity matrix

Fig. 5. Scatter plot of multidimensional scaling applied to the pairwise dissimilarity matrix. (Best viewed in color).