Pattern Recognition 63 (2017) 593–600

Contents lists available at ScienceDirect

Pattern Recognition

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

BundleMAP: Anatomically localized classification, regression, and hypothesis testing in diffusion MRI

Mohammad Khatami [a], Tobias Schmidt-Wilcke [b], Pia C. Sundgren [c,d], Amin Abbasloo [a], Bernhard Schölkopf [e], Thomas Schultz [a,*]

[a] Department of Computer Science, University of Bonn, Friedrich-Ebert-Allee 144, 53113 Bonn, Germany
[b] Department of Neurology, University Hospital Bergmannsheil, Buerkle-de-la-Camp-Platz 1, 44789 Bochum, Germany
[c] Institute for Clinical Sciences/Diagnostic Radiology, Lund University, SE.221 85 Lund, Sweden
[d] Department of Radiology, University of Michigan Health Systems, 1500 E Medical Center Court, Ann Arbor MI 48109, USA
[e] Empirical Inference Department, Max Planck Institute for Intelligent Systems, Spemannstr. 38, 72076 Tübingen, Germany

ARTICLE INFO

Article history:
Received 31 January 2016
Received in revised form
23 August 2016
Accepted 21 September 2016
Available online 21 September 2016

Keywords:
Disease detection
Manifold learning
Support vector machines
Classification
Regression
Fiber tracking
Diffusion MRI

ABSTRACT

Diffusion MRI (dMRI) provides rich information on the white matter of the human brain, enabling insight into neurological disease, normal aging, and neuroplasticity. We present BundleMAP, an approach to extracting features from dMRI data that can be used for supervised classification, regression, and hypothesis testing. Our features are based on aggregating measurements along nerve fiber bundles, enabling visualization and anatomical interpretation. The main idea behind BundleMAP is to use the ISOMAP manifold learning technique to jointly parametrize nerve fiber bundles. We combine this idea with mechanisms for outlier removal and feature selection to obtain a practical machine learning pipeline. We demonstrate that it increases accuracy of disease detection and estimation of disease activity, and that it improves the power of statistical tests.

1. Introduction

Even though nerve fibers are much too small to be observed with Magnetic Resonance Imaging (MRI) directly, their coherent organization hinders and restricts the natural Brownian heat motion of water in a characteristic manner. This can be probed at a macroscopic level by diffusion MRI (dMRI), making it a unique mechanism for studying white matter organization of the human brain in vivo [1]. However, dMRI sequences involve a large number of whole-brain images, taken with different measurement parameters, and biologically relevant quantities can only be derived by analyzing the relationship between those images.

This makes it challenging to leverage the rich information that diffusion MRI provides on white matter disease for the detection of disease or assessment of disease severity using supervised machine learning [2–5]. In particular, it is desirable to derive features that are not only effective, but that should also be interpretable, indicating which anatomical structures are particularly relevant, and how they might be affected by a given disease.

We present a practical system that provides mechanisms for supervised classification, regression, and hypothesis testing of dMRI data based on features that we derive from diffusion parameters and anatomical structures whose interpretation is familiar to neurologists. The name of our method, BundleMAP, reflects the fact that it combines manifold learning using the ISOMAP method [6] with registration and clustering to achieve a joint parametrization of the fiber bundles in a group of subjects. Significantly extending a previous conference paper [7], we combine this idea with methods for outlier removal and feature selection, and demonstrate that this allows us to detect disease, predict disease activity, and visualize diffusion parameters along major nerve fiber bundles, highlighting specific segments on fiber bundles that differ most between the populations.

The structure of our paper is as follows: After reviewing related work in Section 2, we present the individual steps of our BundleMAP approach in Section 3. In particular, Section 3.3 describes the core idea of using manifold learning for joint parametrization, and Section 3.5 describes a stable method to decide on the number of sections per bundle, and a method for imputing missing features.

* Corresponding author.
E-mail addresses: khatami@cs.uni-bonn.de (M. Khatami), tobias.schmidt-wilcke@bergmannsheil.de (T. Schmidt-Wilcke), pia.sundgren@med.lu.se (P.C. Sundgren), abbasloo@cs.uni-bonn.de (A. Abbasloo), bs@tuebingen.mpg.de (B. Schölkopf), schultz@cs.uni-bonn.de (T. Schultz).

http://dx.doi.org/10.1016/j.patcog.2016.09.020

Section 4 contains results and evaluations. In Section 4.2, we demonstrate that, compared to an earlier work [8], BundleMAP shows superior accuracy in detecting Systemic Lupus Erythematosus, a neuroinflammatory disease that can affect cerebral white matter. In Sections 4.3 and 4.4, we report results of using our system for regression and localized hypothesis testing. In Section 4.6, we find that the fiber bundles that our method highlights as important for disease detection are in excellent agreement with previous findings from the literature [8,9]. In Section 4.7, we discuss advantages of using manifold learning for bundle parametrization over a previously proposed alternative, and perform a direct comparison [10]. Finally, Section 5 concludes the paper.

2. Related work

Localized comparison of white matter structures requires their joint parametrization, i.e., anatomical correspondences between fiber bundles of different subjects. Previous methods for this require manual specification of start and end points [11,12] or manual alignment of a cutting plane [10], whereas BundleMAP works fully automatically. Some more automated methods fit deformable models [13] or match fibers to a tractography atlas [14] which, unlike our tool, makes prior assumptions on the bundle shape.

Among the existing alternatives, tract-based morphometry (TBM) [15] is the approach most similar to ours. However, it has not been used for supervised machine learning. Moreover, we take a novel perspective on the problem, leading to computational techniques that are completely different from those used in TBM. The exact relationship between BundleMAP and TBM is discussed in more detail in Section 3.3.

3. Method: our proposed bundlemap pipeline

The overall pipeline of our BundleMAP method consists of four steps, which are illustrated in Fig. 1. First, the major nerve fiber bundles are automatically extracted. Since this is not fully reliable, a combination of outlier removal techniques is applied at the second stage. In the third and most important step, manifold learning is used to establish a joint parametrization of the remaining fibers by mapping them to the latent fiber bundle core. Finally, this parametrization is used to map some common diffusion parameters as a function of position along the bundles. Adaptive binning, in which feature selection determines the most suitable number of bins, creates our final feature vectors.

3.1. Tracking the fiber bundles

Different diffusion models are available to reconstruct nerve fiber bundles from dMRI data. Among them, Diffusion Tensor Imaging is the most widely used, and is employed in our study [16]. More elaborate methods, such as constrained spherical deconvolution [17] or multi-tensor models [18], significantly increase accuracy especially in fiber crossings, and could easily replace Diffusion Tensor Imaging in our pipeline. Unfortunately, they cannot be used with our current data due to its limited angular resolution.

Even though rules exist on how to place seed points for tracking the major fiber bundles [19], we want to avoid having to follow them manually for each individual subject. Therefore, we define seed regions in a template that represents the average of a large number of healthy normal brains. Since this template is aligned with a widely used brain atlas created at the Montreal Neurological Institute, its coordinates define the so-called MNI space.

For each subject, tractography is performed in its individual coordinate system, which allows us to sidestep the difficult problem of correctly adjusting local fiber directions while spatially transforming dMRI data [20]. Seed points are automatically transferred from the template using a nonlinear transformation obtained from an established algorithm for volumetric registration of Fractional Anisotropy [21]. The resulting fibers are warped back into MNI space using the inverse of that transformation.

3.2. Eliminating erroneous fibers

There are two main sources of error that can lead to the inclusion of erroneous fibers during tracking. The first is automatic placement of seed points using image registration, which is known to suffer from inaccuracies [22]. The second are imperfections in the tractography itself, which are known to occur where tracts run closely together or where, at the imaging resolution, two or more fiber bundles cross [23].

We follow two strategies to remove erroneous fibers. First, anatomical knowledge imposes natural constraints on many bundles, and we filter out fibers that violate them. For example, bundles that are known to connect ipsilateral regions should not cross the mid-sagittal plane. Second, in the set of remaining fibers from all subjects, it is often quite obvious which of them are erroneous, since they follow trajectories that differ substantially from the majority of all reconstructed fibers. We use a one-class support vector machine (SVM) with a radial basis function kernel [24] to separate out those atypical fibers.

A one-class SVM treats its input data as samples from a probability distribution, and estimates the support of that distribution. In other words, it identifies a region in the input space that should

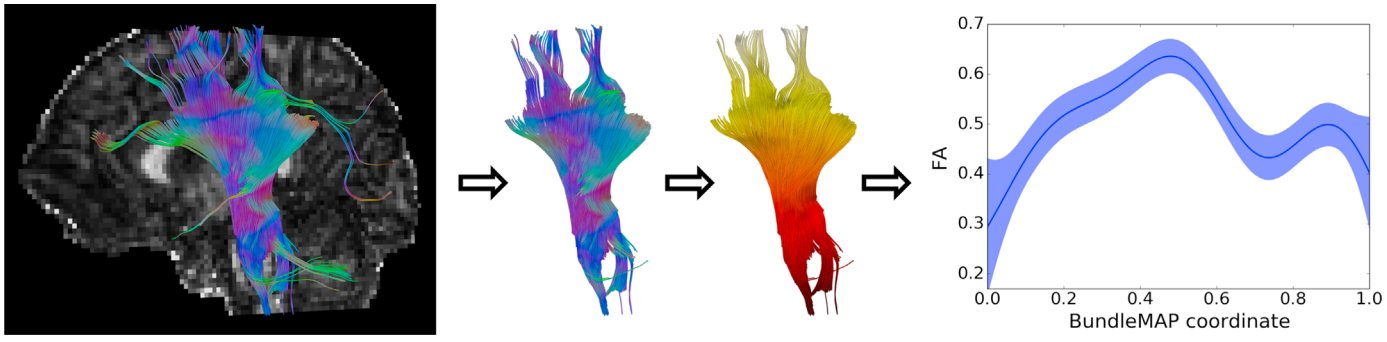

Fig. 1. The four main steps of the BundleMAP pipeline are fiber tractography, outlier removal, joint parametrization of a group of subjects, and derivation of spatially localized features that can be used for supervised learning, hypothesis testing, and visualization.