

Nonlinear Metric Learning for Alzheimer’s Disease Diagnosis with Integration of Longitudinal Neuroimaging Features

Bibo Shi¹
bs354409@ohio.edu
Yani Chen¹
yc147311@ohio.edu
Kevin Hobbs²
hobbsk@ohio.edu
Charles D. Smith³
csmith@mri.uky.edu
Jundong Liu¹
liu@cs.ohio.edu

¹ School of Electrical Engineering
and Computer Science
Ohio University
Athens OH, USA
² Department of Biological Sciences
Ohio University
Athens OH, USA
³ Department of Neurology
University of Kentucky
Lexington KY, USA

Identifying neuroimaging biomarkers of Alzheimer’s disease (AD) is of great importance for diagnosis and prognosis of the disease. In this study, we develop a novel nonlinear metric learning method (**ML-TPS**) to improve biomarker identification for Alzheimer’s disease (AD) and its early stage Mild Cognitive Impairment (MCI). Formulated under a constrained optimization framework, the proposed method learns a smooth nonlinear feature space transformation that pulls the samples of the same class closer to each other while pushing different classes further away. The thin-plate spline (TPS) is chosen as the geometric model due to its remarkable versatility and representation power in accounting for sophisticated deformations. In addition, a multi-resolution patch-based feature selection strategy is proposed to extract both cross-sectional and longitudinal features from MR brain images.

Nonlinear ML-TPS Learning a metric from the training input is equivalent to learn a feature transformation [1]. Depending on the feature space transformation to be sought, metric learning can be divided into linear and nonlinear groups. The proposed ML-TPS model is a direct generalization of linear ML through the application of a deformable geometric model – the thin-plate spline (TPS) – to transform the feature space. TPS is chosen due to its remarkable versatility and representation power in accounting for high-order deformations.

With the side information embedded in the class-equivalent constraints $\mathcal{P} = \{(\mathbf{x}_i, \mathbf{x}_j) | \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ belong to the same class}\}$ and class-nonequivalent constraints $\mathcal{N} = \{(\mathbf{x}_i, \mathbf{x}_j) | \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ belong to different classes}\}$, the goal of our ML solution is pulling the similar subjects closer while pushing dissimilar subjects apart, directly through a TPS nonlinear transformation f as described in Eqn. (2). This can be achieved through the following constrained optimization:

$$\begin{aligned} \min_{B, W} \quad & J = \sum_{\mathbf{x}_i, \mathbf{x}_j \in \mathcal{P}} \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|^2 + \lambda \|W\|_F^2 \\ \text{s.t.} \quad & \sum_{\mathbf{x}_i, \mathbf{x}_j \in \mathcal{N}} \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|^2 \geq 1; \\ & \sum_{i=1}^p W_i^k = 0, \quad \sum_{i=1}^p W_i^k \mathbf{x}_i^k = 0, \quad \forall k = 1 \dots m. \end{aligned} \quad (1)$$

where, f is the TPS transformation, as described in matrix format:

$$f(\mathbf{x}) = \mathbf{x} \cdot B + \begin{pmatrix} G(\mathbf{x}, \mathbf{x}_1) \\ \dots \\ G(\mathbf{x}, \mathbf{x}_p) \end{pmatrix} \cdot W = \mathbf{x} \cdot B + \vec{G} \cdot W. \quad (2)$$

B is the linear transformation matrix, and W is the weight matrix for the nonlinear parts. \mathbf{x}_p are the anchor points used to compute the TPS kernel. Compared with a classic linear metric learning model MMC [4], another component $\|W\|_F^2$, the squared Frobenius norm of W , is added to the objective function as a regularizer to prevent overfitting. λ is the weighting factor to control the importance of two components. Similar as in MMC, the nonequivalent constraint $\sum_{\mathbf{x}_i, \mathbf{x}_j \in \mathcal{N}} \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|^2 \geq 1$ is to impose a scaling control to avoid trivial solutions. The other two equivalent constraints with respect to W is to ensure that the elastic part of the transformation is zero at infinity. W^k is the k th column of W , and \mathbf{x}^k is the k th component of \mathbf{x} .

Feature extraction Feature extraction and selection from the ADNI database [3] is also in great need of further exploration. Most of the exist-

ing works are either cross-sectional features obtained at one point in time, or “static” longitudinal volumetric information acquired at two or multiple time points but only through structural segmentation. In part due to the unavailability of deformation data under ADNI, “dynamic” longitudinal information such as the atrophy over time at various gray matter (GM) areas, which is a major hallmark in the progression of AD, has not been fully utilized in the literature. In this paper, we propose a multi-resolution patch selection strategy, with both cross-sectional (baseline) and longitudinal atrophy features extracted from MR brain images.

The neuroimage data used in this work were obtained from the ADNI database. We consider only the subjects for whom the baseline (M0) visits and 12-month follow-up (M12) T1-weighted MRIs, together with their *MIDAS Whole Brain Masks*, are all available. Fig. 1 illustrates an overview of the schematic diagram of our proposed framework. Two sets of class-discriminative patches are first extracted at three different levels of resolutions: one set is based on the gray matter (GM) densities of the baseline MRIs, and the other is selected through the longitudinal deformation magnitudes (DM) estimated between baselines (M0-MRIs) and follow-ups (M12-MRIs). Both selection procedures are conducted through statistical significance tests between groups. With p -value set to 0.05, this step returns more than 1000 patches in each set. To capture the most discriminative patches, a wrapper feature selection [2] is carried out with a greedy forward searching within the pools of GM patches, DM patches, and their combination.

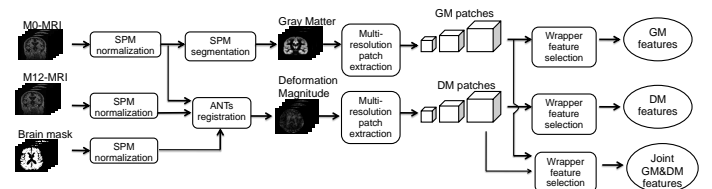


Figure 1: Flowchart of the proposed two-step patch based feature extraction and selection strategy.

Conclusion Implementations of the proposed methods, as well as the detailed experimental evaluations, are described in the paper. The two novel components in our solution – nonlinear metric learning through TPS and integration of longitudinal atrophy information – have both proven to be highly effective in improving the classification of AD/MCI versus NC.

[1] A. Bellet, A. Habrard, and M. Sebban. A survey on metric learning for feature vectors and structured data. *arXiv preprint arXiv:1306.6709*, 2013.
[2] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3: 1157–1182, 2003.
[3] Clifford R Jack, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L Whitwell, Chadwick Ward, et al. The alzheimer’s disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008.
[4] Eric P Xing, Andrew Y Ng, Michael I Jordan, and Stuart Russell. Distance metric learning with application to clustering with side-information. *NIPS*, pages 521–528, 2003.