Manifold-based classification incorporating subject metadata

Robin Wolz[1]
http://www.doc.ic.ac.uk/~rw1008/
Paul Aljabar[1]
http://www.doc.ic.ac.uk/~pa100/
Joseph V. Hajnal[2]
jo.hajnal@imperial.ac.uk
Jyrki Lötjönen[3]
Jyrki.Lotjonen@vtt.fi
Daniel Rueckert[1]
http://www.doc.ic.ac.uk/~dr/

[1] Biomedical Image Analysis
Imperial College London
London, UK

[2] MRC Clinical Sciences Centre
Imperial College London
London, UK

[3] Signal and Image Processing
VTT Technical Research Center Finland
Tampere, Finland

Abstract

Recent work suggests that the space of brain magnetic resonance (MR) images can be described by a nonlinear and low-dimensional manifold. In the context of classifying Alzheimer's disease (AD) patients from healthy controls, we propose a method to incorporate subject meta-information into the manifold learning step. Information such as gender, age or genotype is often available in clinical studies and can inform the classification of a given query subject. In the proposed method, such information, whether discrete or continuous, can be used as an additional input to manifold learning and to enrich a distance measure derived from pairwise image similarities. We use the ApoE genotype, the CSF-concentration of Aβ_{42} and hippocampal volume as meta-information to achieve significantly improved classification results.

1 Introduction

Many of the well-established biomarkers for dementia from MR images are based on morphometric measures, such as the volume or shape of brain structures [3, 7]. More recently, models based on machine learning techniques have been proposed which seek discriminating features over the whole brain or within a defined region of interest [4].

In recent work, a nonlinear manifold representation for serial MR data [6] of subjects undergoing normal aging and neurodegeneration was proposed which was then used together with training data to define a classifier in the low-dimensional space. To incorporate additional information into the manifold learning process, and to seek potentially more reliable biomarkers, we propose a strategy to enrich the pairwise image similarities used to learn the manifold representation with metadata about a subject's state.

We evaluate the proposed method on brain MR images from healthy controls, subjects with mild cognitive impairment (MCI) and AD taken from the ADNI study[1]. We use the

[1] www.loni.ucla.edu/ADNI

420 subjects for which a measurement of the CSF-concentration of the Aβ_{42} protein and the subject's ApoE genotype are currently available and we also test the power of automatically derived hippocampal volumes as meta-information.

2 Method

2.1 Classification using manifold learning

Given a set of images $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\} \in \mathbb{R}^n$ with each image $\mathbf{x}_i$ defined as a vector of intensities, the goal is to derive biomarkers which discriminate between clinically relevant subject groups. Assuming $\mathbf{x}_1, ..., \mathbf{x}_N$ lie on or near an l-dimensional manifold $\mathscr{M}$ embedded in $\mathbb{R}^n$, we learn a low dimensional representation $\mathbf{Y} = \{\mathbf{y}_1, ..., \mathbf{y}_N\}$ with $\mathbf{y}_i \in \mathbb{R}^l$ of the input images in $\mathscr{M}$.

Laplacian eigenmaps (LE) [1] may be used to achieve a nonlinear dimensionality reduction $f : \mathbf{X} \to \mathbf{Y}, \mathbf{y}_i = f(\mathbf{x}_i)$. An undirected weighted graph $G =< V, E >$ with N nodes V representing the images and edges E connecting neighboring nodes is defined on $\mathbf{X}$. The weights of E are defined based on a weight matrix $\mathbf{W}$ of pairwise image similarities w_{ij}. Based on w_{ij}, and with certain constraints on $\mathbf{y}_i$, the Laplacian eigenmaps may be viewed as minimizing the following objective function

$$\sum_{ij} (\mathbf{y}_i - \mathbf{y}_j)^2 w_{ij}. \tag{1}$$

This problem can be formulated as solving the generalized eigenvector problem $\mathbf{L}v = \lambda \mathbf{D}v$, where the graph Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is defined from the weight matrix $\mathbf{W}$ and the diagonal degree matrix $D_{ii} = \sum_j w_{ij}$ [1].

As previously proposed in [6], a classifier can be defined in the low-dimensional space by using training data to identify a separating hyperplane between two subject groups.

2.2 Manifold learning incorporating non-imaging information

In [2], the graph G is extended by two nodes, each representing one of two classes available for training data. Connecting each training subject with its respective class node imposes the distribution of the training data on the manifold structure.

Metadata available in a clinical setting can be defined by discrete categories (two or more), or by a continuous variable. To deal with such data, we propose an extension to the method described in [2], dealing with more than two classes and a fuzzy-class membership to represent continuous metadata. In the discrete and the continuous case, additional nodes are introduced in graph G, representing discrete groups or sub-intervals of a continuous interval. In the discrete setting, equally weighted edges connect each subject only to the node representing its group. In the continuous setting, each subject is connected to all additional nodes with the weights defined according to the distances of the subject's metadata to the centers of the defined intervals. The two weighting schemes are illustrated in Figure 1.

Extending graph G by $\hat{N}$ nodes $\hat{V}$, each associated with a possible state z of a metadata variable Z, gives the objective function

$$\sum_{ij} (\mathbf{y}_i - \mathbf{y}_j)^2 w_{ij} + \gamma \sum_{ik} (\mathbf{y}_i - \hat{\mathbf{y}}_k)^2 c_{ik} \tag{2}$$

which defines the constrained low-dimensional embedding space

$$\hat{\mathbf{Y}} = \{\hat{\mathbf{y}}_1, ..., \hat{\mathbf{y}}_{\hat{N}}, \mathbf{y}_1, ..., \mathbf{y}_N\} \quad , \quad \hat{\mathbf{y}}_k, \mathbf{y}_i \in \mathbb{R}^l. \tag{3}$$

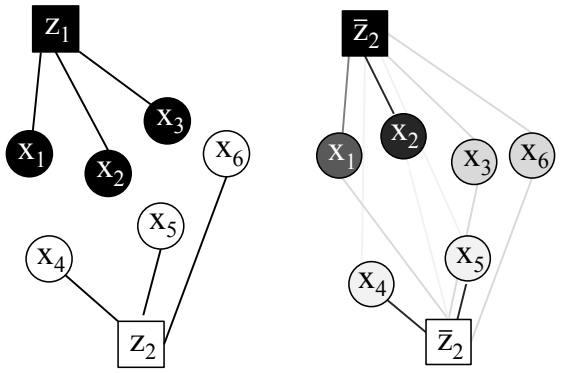

Figure 1: Weights defined between image nodes $\mathbf{x}_i$ and additional nodes representing metadata Z. In the discrete setting (left), equally weighted edges are defined according to group membership (black or white). In the continuous setting, weights to both additonal nodes are defined according to the continuous metadata (grey-value).

The additional nodes $\hat{V}$, with embedding coordinates $\hat{\mathbf{y}}_k$, represent the different groups defined by variable Z in the low-dimensional space. Minimizing the distance of every subject to the additional nodes according to the defined weighting, incorporates the information from the metadata into the learned manifold. The weights c_{ik} affect the proximity of subject i to the k^{th} group and its centroid $\hat{\mathbf{y}}_k$. The c_{ik} can be binary in a discrete setting with $\hat{N}$ possible states and variables $Z = \{z_1, ..., z_{\hat{N}}\}$. To represent continuous metadata, $\hat{N}$ variables $\bar{z}_1, ..., \bar{z}_{\hat{N}}$ are defined, representing discretized sub-intervals of the continuous interval $a < z < b$ in which the variable Z is defined. The weights c_{ik} are then defined according to the probability of a given subject of belonging to each discretized group. A high weight of the parameter γ clusters the subjects mainly according to the metadata, whereas $\gamma = 0$ results in the standard embedding with Laplacian eigenmaps which uses only pairwise image similarities. With the $N \times \hat{N}$ matrix $\mathbf{C}$ defining the weights between subject i and the additional nodes, an extended weight matrix

$$\mathbf{W}' = \begin{pmatrix} \mathbf{I} & \frac{\gamma}{2}\mathbf{C}^T \\ \frac{\gamma}{2}\mathbf{C} & \mathbf{W} \end{pmatrix} \tag{4}$$

is derived, where $\mathbf{I}$ is the $\hat{N} \times \hat{N}$ identity matrix. As above, solving the generalized eigenvector problem associated with the extended weight matrix gives the embedding coordinates which optimize the objective function in Equation 2.

3 Data and Results

3.1 Subjects

Images used to evaluate the proposed method were obtained from the ADNI database [5]. In the ADNI study, brain MR images were acquired at regular intervals after an initial baseline scan from approximately 200 cognitively normal older subjects (CN), 400 subjects with mild cognitive impairment (MCI), and 200 subjects with early AD.

ADNI provides the ApoE genotype (determined by the ApoE alleles carried) for all subjects. Humans carry two out of three possible ApoE alleles ($\varepsilon 2$, $\varepsilon 3$, $\varepsilon 4$). Carriers of $\varepsilon 4$ have been shown to have a higher risk of developing AD, while $\varepsilon 2$ carriers have a lower risk [5]. In addition an Aβ_{42} protein analysis of cerebrospinal fluid (CSF) is available for a subset of ADNI subjects. A decrease in the concentration of this protein has been shown to be associated with a development of AD [5].

In this work, we used the 1.5T T1-weighted baseline images of the 420 subjects for which

a CSF analysis was available. Out of 201 MCI subjects, 89 were progressive, i.e. were diagnosed as converting to AD as of October 2010. We therefore independently analyzed stable (S-MCI) and progressive (P-MCI) subjects. Table 1 presents an overview of the subjects studied and their metadata as well as their scores on the Mini Mental State Examination (MMSE) used for clinical diagnosis.

Table 1: Number (female) of study subjects. Carriers of the ApoE $\varepsilon 2/\varepsilon 4$ alleles are shown. The remaining subjects only carry the $\varepsilon 3$ allele. Average Aβ_{42} concentration, MMSE score and the derived biomarker hippocampal volume are shown. There is no significant difference in age between the different groups with an average age of 74.95±7.03 years.

	N (F)	$\varepsilon 2/\varepsilon 4$	Aβ_{42} (pg/ml)	MMSE	Hippo. vol.
CN	116 (56)	16/28	202.3 ± 57.5	29.12 ± 1.02	4.53 ± 0.55
S-MCI	112 (36)	9/49	178.9 ± 61.6	27.16 ± 1.75	4.26 ± 0.59
P-MCI	89 (33)	1/52	146.3 ± 46.3	26.64 ± 1.80	3.93 ± 0.65
AD	103 (43)	4/63	147.5 ± 45.8	23.55 ± 1.87	3.92 ± 0.73

3.2 Experiments

All images were aligned to a template and pairwise similarities w_{ij} were estimated as cross correlation over a region of interest around the hippocampus [7]. The proposed method was then used to incorporate ApoE genotype and Aβ_{42} concentration into the manifold learning process. In a third experiment, we used automatically determined hippocampus volumes [7] as a derived biomarker to enrich the manifold learning process (right column of table 1).

For each of the experiments, $\hat{N} = 3$ additional nodes with embedding coordinates $\hat{\mathbf{y}}_k$, $k = 1,2,3$ are used in Equation 2. With ApoE, these nodes are trivially associated with the three genotypes. c_{ik} is set to one if subject i has a genotype associated with node k, otherwise it is set to zero. For the two continuous variables, Aβ_{42} concentration and hippocampus volume, a continuous weighting c is defined. In both cases, the defined variable interval Z is subdivided into three sub-intervals Z_1, Z_2, Z_3 defined by the limits of the variable data and its 33^{rd} and 67^{th} percentiles. Weights $c_{ik}, k = 1,2,3$ are inversely proportional to the distances of z_i to the three mean values $\bar{z}_1, \bar{z}_2, \bar{z}_3$ of Z_1, Z_2, Z_3 and normalized to sum to one.

The number of embedding coordinates was optimized on the 418 ADNI baseline images not used in the evaluation. When varying $l \in [1,...,50]$ with standard LE, stable results were achieved for $l \in [6,15]$. Performance was accordingly evaluated for these 10 dimensions in all experiments and averaged results are reported. The weighting factor γ defining the influence of metadata was set for all experiments to $\gamma = 0.125$. This choice is based on the classification accuracy on the 418 training images with hippocampal volume as metadata. Manifold coordinates were corrected for subject age using linear regression before performing a leave-25%-out cross-validation on the test data (420 images). Average classification rates after 1,000 runs are determined for every dimension $l \in [6,15]$ and then averaged.

3.3 Results

Table 2 presents the correct classification rates for four different experiments which are summarized as follows: (A) Classic LE using pairwise cross correlation (CC), (B) Extended LE using CC and ApoE genotype, (C) Extended LE using CC and Aβ_{42}, (D) Extended LE using CC and hippocampal volume.

For each experiment, the multiple runs provides a distribution estimate for the corresponding classification rate which were used to carry out unpaired t-tests between the results of method A and each of methods B-D to test the significance of any improvements.

Table 2: Classification rates (%) with LE (A) and extensions incorporating different types of meta-information (B-D). **Bold** indiciates significant difference from method A ($p<0.001$).

	A	B	C	D
AD vs CN	84	84	**86**	**87**
S-MCI vs P-MCI	62	**63**	**65**	**65**
P-MCI vs CN	80	**81**	**83**	**83**

4 Discussion and conclusion

We have proposed a method to include clinical meta-information associated with subjects into a manifold learning framework. Our results show that incorporating such information can help to achieve improved classification rates when using the low-dimensional embedding coordinates. We have validated the proposed method on a large and diverse clinical dataset (ADNI) using ApoE genotype, the concentration of $A\beta_{42}$ and hippocampal volume as metadata. The effectiveness of the proposed approach is evidenced by a significant improvement in classification rates compared to standard manifold learning. Our results are in the range of what was reported on ADNI in a recent comparison of different high-dimensional classification approaches and methods based on the volumes of brain structures [3].

References

[1] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[2] J.A. Costa and A.O. Hero III. Classification constrained dimensionality reduction. *ICASSP '05*, pages 1077 – 1080, 2005.

[3] R. Cuingnet, E. Gerardin, J. Tessieras, and et al. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage*, In Press:–, 2010. ISSN 1053-8119.

[4] Y. Fan, N. Batmanghelich, C. M. Clark et al., and et al. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict cognitive decline. *NeuroImage*, 39(4):1731 – 1743, 2008. ISSN 1053-8119.

[5] S. G. Mueller, M. W. Weiner, L. J. Thal, and et al. The Alzheimer's Disease Neuroimaging Initiative. *Neuroimaging Clinics of North America*, 15(4):869 – 877, 2005.

[6] R. Wolz, P. Aljabar, J. Hajnal, and D. Rueckert. Manifold learning for biomarker discovery in MR imaging. volume 6357 of *LNCS*, pages 116–123. Springer Berlin / Heidelberg, 2010.

[7] R. Wolz, P. Aljabar, J. V. Hajnal, and et al. LEAP: Learning embeddings for atlas propagation. *NeuroImage*, 49(2):1316 – 1325, 2010.