MRI based dementia classification using semi-supervised learning and domain adaptation

Elaheh Moradi[1], Christian Gaser[2], Heikki Huttunen[1], Jussi Tohka[1] and Alzheimer's Disease Neuroimaging Initiative[**]

1. Department of Signal Processing , Tampere University of Technology, P.O. Box 553 FIN-33101 Tampere, Finland
jussi.tohka@tut.fi
2. Departments of Neurology and Psychiatry, Jena University Hospital Jena, Germany

Abstract. We propose a method for dementia classification based brain magnetic resonance images (MRIs). The method learns to recognize patients with Alzheimer's disease or Mild Cognitive Impairment from healthy controls. The features used are extracted with sparse logistic regression from a large pool of voxel-wise gray matter densities computed based on MRIs registered to stereotactic space. The classifier uses a Low Density Separation algorithm, which can take advantage of both labeled and unlabeled samples. The differences between the training and test sets are compensated based on an algorithm for unsupervised domain adaptation. The method is fully automatic. The proposed method participated in the 2014 CADDementia competition, with an estimated accuracy of 0.767 for the public test data. The training data was extracted from ADNI database.

Key words: Semi-supervised learning, Alzheimer's disease, mild cognitive impairment, domain adaptation, low density separation

1 Introduction

Alzheimers disease (AD) is the most common form of dementia. More than 30 million people worldwide suffer from AD and, due to the increasing life expectancy, this number is expected to triple by 2050 [1]. Therefore, it is extremely important to identify subjects in a risk of getting the disease.

In this paper, we propose an approach to automatically categorize subjects into three classes: Subjects with Alzheimer's disease (AD), subjects with mild

[**] Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

cognitive impairment (MCI) and cognitively normal subjects (NC). Our method relies on voxel-based morphometry (VBM) style preprocessing [2]. The MRI features used for the classification are selected from a larger pool of features using sparse logistic regression [3, 4]. Then, relying on these features, we construct a hierarchical classification framework utilizing binary classifiers as component classifiers. Each binary classifier is trained in a semi-supervised manner, meaning that they can utilize unlabeled data in addition to labeled data. The labeled data is obtained from the ADNI database and CADDementia data is used as unlabeled data. The semi-supervised learning is performed by the Low Density Separation (LDS) algorithm [5], and we have demonstrated its efficiency for classifying subjects with stable and progressive MCI [6, 7]. Here, we extend the methods [4, 7] to the three-category classification problem (AD vs. MCI vs. NC). Moreover, because our training and test sets have different characteristics we utilize an unsupervised domain adaptation method to normalize the samples [8]. The method is fully automatic.

The rest of this paper is organized as follows. Section 2 describes the image data used for the training and validating the classifier. Also the image preprocessing and the used features are described in Section 2. Section 3 introduces the classification scheme. Section 4 presents experimental validation results with CADDementia data and Section 5 concludes the paper.

2 Material

2.1 Training data: ADNI

Data used in this work to train the classifier is obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database `http://adni.loni.usc.edu/`. The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a 60 million US dollar, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in

the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

Data used include baseline MRI of 835 subjects (T1-weighted MP-RAGE sequence at 1.5 Tesla, typically 256 x 256 x 170 voxels with the voxel size of approximately 1 mm x 1 mm x 1.2 mm). 835 subjects are grouped as

1. *AD (Alzheimers disease)*, if diagnosis was Alzheimers disease at baseline (n = 200);
2. *NC (Normal Cognitive)*, if diagnosis was normal at baseline (n = 231);
3. *sMCI (stable MCI)* if diagnosis was MCI at all available time points, but at least for 36 months (n = 100);
4. *pMCI (progressive MCI)*, if diagnosis was MCI at baseline but conversion to AD was reported after baseline within 1, 2 or 3 years, and without reversion to MCI or NC at any available follow-up (n = 164);
5. *uMCI (unknown MCI)*, if diagnosis was MCI at baseline but they are not diagnosed at the end of the project. These subjects' data were not used for the classifier training.

We used different training sets based on this labeling to build the component classifiers of our hierarchical scheme. We explain the details in Section 3.4.

2.2 Test data:CADDementia

The method was validated with 30 labeled images from CADDementia data described in the challenge homepage [1]. These images were not used for training nor parameter tuning. The essential difference between this data and the training data is that the CADDementia images were acquired with 3 Tesla scanners.

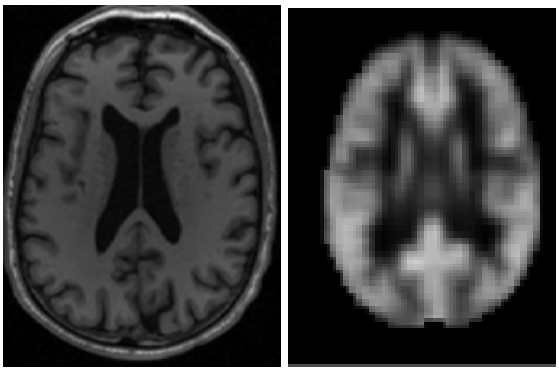
2.3 Image preprocessing and features

All the images (train and test) were preprocessed in a fully automatic manner by a pipeline similar to that described in [2]. Preprocessing of the T1-weighted images was performed using the SPM8 package[2] and the VBM8 toolbox[3], both running under MATLAB. All T1-weighted images were corrected for bias-field inhomogeneities, then spatially normalized and segmented into grey matter (GM), white matter, and cerebrospinal fluid (CSF) within the same generative model [9]. The segmentation procedure was further extended by accounting for partial volume effects [10], by applying adaptive maximum a posteriori estimations [11], and by using an hidden Markov random field model [12] as described previously [13]. Only the GM images were used. Note that these images represent GM tissue fractions in each voxel. Following the pipeline proposed by [14], the GM images

[1] `http://caddementia.grand-challenge.org/home/`
[2] `http://www.fil.ion.ucl.ac.uk/spm/`
[3] `http://dbm.neuro.uni-jena.de/`

were processed with affine registration and smoothed with 8-mm full-width-at-half-maximum smoothing kernels. After smoothing, images were resampled to 4 mm spatial resolution. See Figure 1 for an example slice of the original image and preprocessed image.

Masking of the GM tissue fraction images results in aligned GM tissue fractions from 29852 voxels. As discussed in [4, 7], the number of voxels significantly exceeds the number of the available training data. Although our classifier is relatively tolerant to high-dimensional data, it is still unable to process this high number of features. Therefore, we initially reduce the number of features by using a regularized logistic regression classifier, that has an inherent feature selection property, on data from AD and NC classes (ADNI) [4, 7, 3]. The classifier produces a set of good candidate subsets with different cardinalities, and the most appropriate subset is selected by cross-validation. As a result, 309 voxels were selected. Finally, age related effects were removed from the data by using linear regression [15, 6].



(a) (b)

Fig. 1. An example slice of the original image (a) and preprocessed image (b)

3 Classifier

3.1 Overview

A simplified flowchart of the training of the classifier is shown in Figure 2. Details of this procedure are explained in Section 3.4. Important algorithmic components, low density separation classifier and domain adaptation are briefly described in Sections 3.2. and 3.3, respectively.

Our classification scheme is hierarchical. In the first step, we aim to separate AD and NC subjects to different classes without caring to which class the MCI subjects are classified. We term these disease classes (AD + MCI and NC +

MCI) as the first level classes. Before the classification, we perform the domain adaptation by using only AD and NC subjects in both datasets. In second step, the first level classes are further divided to AD and MCI classes (AD + MCI) and MCI and NC classes (NC + MCI) giving us the desired three class classification. Note that the label of the subjects classified as MCI is independent of whether the first level class was AD + MCI or NC + MCI. The rationale of this hierarchical scheme is that the MCI is a transitional stage between the AD and normal aging, however, not all the MCI subjects convert to AD and this makes the MCI class very heterogeneous.

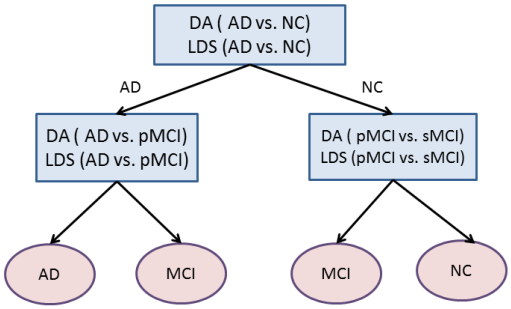


Fig. 2. Classifier structure. DA stands for domain adaptation. LDS stands for low density separation.

3.2 Low density separation

We apply a semi-supervised learning (SSL) technique called Low Density Separation (LDS) to train the three required component classifies in our hierarchical scheme. We next explain the main ideas of the algorithm briefly; see [5] for further details. Semi-supervised learning (SSL) approaches are able to use unlabeled data in conjunction with labeled data in a learning procedure for improving the classification performance. LDS is a semi-supervised learning algorithm which relies on the assumption that there is low density region with little (if any) data, which is where the decision boundary should lie.

The algorithm consists of two stages. First, it constructs a graph distance derived kernel with the aim of increasing class separability and on the other hand to increase the clustering within the classes. Heuristically, the distance between the two nearest neighbors in feature space is incremented if they are far from each other and decremented if they are close to each other. The definition of the distance depends on the parameter ρ which we tune by cross-validation (see [5] for details).

The second step consists of training a transductive support vector machine with the graph-distance derived kernel to obtain the parameters for the dis-

criminant function $y = sign(\mathbf{w}^T\mathbf{x} + b)$, where $\mathbf{w}$ is the weight vector (in the transformed space), $\mathbf{x}$ is the image to be classified (in the transformed space), $y = \{-1, 1\}$ is the label of $\mathbf{x}$ and b is the bias of the classifier. Vector $\mathbf{w}$ and bias b are found by minimizing

$$\frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{n=1}^{N} L(y_n(\mathbf{w}^T\mathbf{x}_n - b)) + C^*\sum_{n=N+1}^{N+M} L(|\mathbf{w}^T\mathbf{x}_n - b|),$$

where $(\mathbf{x}_n, y_n), n = 1, \ldots, N$ are the labeled data, $\mathbf{x}_n, n = N+1, \ldots, N+M$ are the unlabeled data, $L(\cdot)$ is the hinge loss function, and C and C^* are scalar parameters. The value of C is selected by cross-validation within the training data, and C^* is set as in [5].

3.3 Domain adaptation

It is unlikely that the data from ADNI and CADDementia would follow the same distributions conditioned on the disease labels. For example, the ADNI data we use has been acquired with 1.5 Tesla scanners while the CADDementia data has been acquired with 3 Tesla scanners. Techniques for addressing learning problems with mismatched distributions are often referred as *domain adaptation* or *transfer learning*. The idea of these algorithms is try to improve the similarity of the data from source (ADNI in our case) and target domains (CADDementia in our case). When there is no labeled data from the target domain to help learning classifiers, the problem setting is termed *unsupervised domain adaptation*. Here, we utilize an information theoretic approach for the unsupervised domain adaptation [8]. We did not compensate for possible differences between the different acquisition sites in the CADDementia data.

3.4 Detailed procedure

In the first level, the ADNI (AD and NC) subjects are used as source data and all CADDementia data are used as target data for domain adaptation. After domain adaptation, the AD and NC subjects from ADNI are used as training data for AD + MCI and NC + MCI classes, respectively. In order to design the first level classifier and the most important LDS parameters (C and ρ) are tuned using 10-fold cross-validation inside the training (ADNI) data. All CADDementia data are used as unlabeled data for this classifier and eventually divided into two groups. This is repeated 101 times to obtain the best possible parameter values and the final class of the subjects is decided based on the majority vote. Based on this procedure all the CADD data are divided into two groups, i.e., AD + MCI group and NC + MCI group.

In the second level, to design AD vs. MCI classifier, the AD and pMCI subjects of ADNI are used as the source data and the CADDementia subjects classified to the AD + MCI class during the first level are used as target data for the domain adaptation. After the domain adaptation the AD and pMCI

subjects of ADNI are used as labeled data for training in order to design the AD vs MCI classifier and the LDS parameters (C and ρ) are tuned using 10-fold cross-validation inside training data. The CADDementia subjects classified to the AD + MCI class during the first level are used as unlabeled data for the LDS classifier and subsequently classified to AD and MCI groups. Again this is repeated 101 times and the final label of the subject is decided based on the majority vote.

In the second level, to design NC vs. MCI classifier, the sMCI and pMCI subjects of ADNI are used as the source data and the CADDementia's NC + MCI class from the first level are used as the target data for the domain adaptation. After the domain adaptation the sMCI and pMCI subjects of ADNI are used for training in order to design the NC vs. MCI classifier and the LDS parameters (C and ρ) are tuned using cross-validation inside training data, again repeating the procedure for 101 times. The classifier divides CADDementia's NC + MCI class into two subclasses, i.e. NC and MCI subclasses. The decision to use sMCI subjects' data as the training data for NC class was made based on experimental grounds. In this phase, we balanced the numbers of pMCI and sMCI subjects in the ADNI data by resampling in both domain adaptation and classification phases because the number of sMCI subjects (100) is smaller than the number of pMCI (164) subjects in the ADNI data.

3.5 Computation time

The total running time was approximately 9 minutes per image with a Matlab based implementation. The computationally most heavy part was the image pre-processing implemented in VBM8 that required approximately 8 minutes per image. The domain adaptation required, on average, 29.73 seconds per 100 images (domain adaptation cannot be performed for a single image). The LDS classification and age removal required under one second per image. The domain adaptation and classification were performed twice for each image (in the first and second level).

4 Results

The classification accuracy on 30 labeled examples in the CADDementia dataset was 0.767. Since we did not use the label information on the CADDementia data, we consider this to be unbiased estimate of the classification accuracy. The confusion matrix with this data is shown in Table 1. The confusion matrix shows that there were no NC subjects mislabeled as AD subjects or vice versa. However, there were mislabelings between MCI and AD and NC and MCI. This is consistent with the MCI being a transitional stage between normal aging and AD.

True class	Predicted class		
	NC	MCI	AD
NC	9	3	0
MCI	2	7	0
AD	0	2	7

Table 1. Confusion matrix

5 Discussion

We have proposed a method for dementia classification based brain MRIs. The fully automated method recognized patients with AD or MCI from healthy controls. The features for the classification were extracted from voxel-wise gray matter densities computed based on aligned MRIs [2]. The classification method was hierarchical, utilizing the fact that the MCI is a translational stage between the AD and normal aging. The main novelties of the classifier were the utilization of unlabeled data in additional to labeled data in order to improve the classification [5] and the use of unsupervised domain adaptation to try to compensate between the differences of the training and test data [8]. We have previously demonstrated the utility of unlabeled data for predicting MCI-to-AD conversion [4]. The method trained with ADNI data achieved an accuracy of 0.767 with the public part of the CADDementia test data.

Acknowledgments. Financially supported by Academy of Finland grants 130275 and 263785. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimers Association; Alzheimers Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; And Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

1. Barnes, D.E., Yaffe, K.: The projected effect of risk factor reduction on Alzheimer's disease prevalence. The Lancet Neurology **10**(9) (2011) 819–828
2. Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., Sauer, H., et al.: BrainAGE in mild cognitive impaired patients: predicting the conversion to alzheimers disease. PloS ONE **8**(6) (2013) e67346
3. Friedman, J.H., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. J. Stat. Software **33**(1) (2010) 1–22

4. Moradi, E., Gaser, C., Tohka, J.: Semi-supervised learning in MCI-to-AD conversion prediction - when is unlabeled data useful? In: IEEE Pattern Recognition in Neuro Imaging, IEEE (2014) 121 – 124
5. Chapelle, O., Zien, A.: Semi-supervised classification by low density separation. In: AISTATS. (2005) 57–64
6. Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J.: Semi-supervised learning for early MRI-based MCI-to-AD conversion prediction. In: Organization for human brain mapping annual meeting. (2014)
7. Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J.: Semi-supervised learning and random forests for MRI-based prediction of AD conversion in MCI subjects. Submitted (2014)
8. Shi, Y., Sha, F.: Information-theoretical learning of discriminative clusters for unsupervised domain adaptation. ICML 2012 (2012)
9. Ashburner, J., Friston, K.J.: Unified segmentation. Neuroimage **26**(3) (2005) 839–851
10. Tohka, J., Zijdenbos, A., Evans, A.: Fast and robust parameter estimation for statistical partial volume models in brain MRI. Neuroimage **23**(1) (Sep 2004) 84–97
11. Rajapakse, J.C., Giedd, J.N., Rapoport, J.L.: Statistical approach to segmentation of single-channel cerebral mr images. Medical Imaging, IEEE Transactions on **16**(2) (1997) 176–186
12. Cuadra, M.B., Cammoun, L., Butz, T., Cuisenaire, O., Thiran, J.P.: Comparison and validation of tissue modelization and statistical classification methods in t1-weighted MR brain images. IEEE Trans Med Imaging **24**(12) (Dec 2005) 1548–1565
13. Gaser, C.: Partial volume segmentation with adaptive maximum a posteriori (MAP) approach. NeuroImage **47** (2009) S121
14. Franke, K., Ziegler, G., Klöppel, S., Gaser, C.: Estimating the age of healthy subjects from t1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. Neuroimage **50**(3) (2010) 883–892
15. Dukart, J., Schroeter, M.L., Mueller, K., et al.: Age correction in dementia–matching to a healthy brain. PloS one **6**(7) (2011) e22193