SEMI-SUPERVISED MULTIMODAL CLASSIFICATION OF ALZHEIMER'S DISEASE

Daoqiang Zhang, Dinggang Shen

Dept. of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599

ABSTRACT

One challenge in identification of Alzheimer's disease (AD) is that the number of AD patients and healthy controls (HCs) is generally very small, thus difficult to train a powerful AD classifier. On the other hand, besides AD and HC subjects, we often have MR brain images available from other related subjects such as those with mild cognitive impairment (MCI), a prodromal stage of AD, or possibly the unrelated subjects whose cognitive statuses may be not known. These images may be helpful for building a powerful AD classifier, although their cognitive status may not belong to AD or HC. Accordingly, in this paper, we investigate the potential of using MCI subjects to aid classification of AD from HC subjects via multimodal imaging data and CSF biomarkers. In particular, a multimodal Laplacian Regularized Least Squares (mLapRLS) method, based on semi-supervised learning, is proposed for achieving this purpose. In the objective function of mLapRLS, there are two terms: a term involving only AD and HC subjects for supervised learning, and another term involving all AD, HC, and MCI subjects for unsupervised estimation of intrinsic geometric structure of the data. Experimental results show that our proposed method can significantly improve AD classification, with aid from MCI subjects.

Index Terms— Semi-supervised, Alzheimer's disease, MCI, disease classification, multimodal imaging

1. INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia in elderly people worldwide. Early diagnosis of AD is very important for possible delay of the disease. At present, several biomarkers have been proved to be sensitive to AD, including brain atrophy measured by magnetic resonance imaging (MRI), hypometabolism measured by functional imaging (e.g., positron emission tomography (PET)), and quantification of specific proteins measured through cerebrospinal fluid (CSF) [1-2]. Many AD classification methods have been developed based on one or two imaging modalities [3-7]. Recently, some methods have also been proposed for combining multimodal biomarkers to improve classification performance, which can generally do better than using only a single type of biomarkers [6-7].

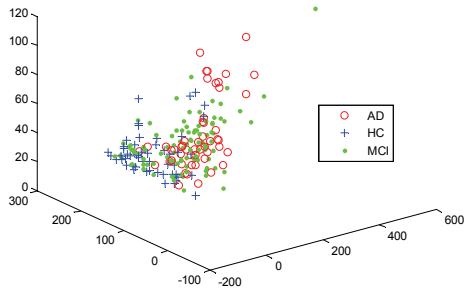

Fig. 1. Distributions of AD, HC and MCI subjects with CSF features.

One challenge in AD classification is that the number of AD patients and healthy controls (HCs) available for training is generally very small, while the dimensionality of data is often very high. For instance, structural MRI or functional PET images may contain hundreds of thousands of voxels in each image. To reduce the dimensionality of imaging data, the dimensionality reduction techniques are always used, which can be categorized into (1) feature extraction and (2) feature selection methods. The former is used to extract some features from original imaging data. For example, we can extract various features from structural MRI image, i.e., local structural volumes (from the region of interests (ROIs)), cortical thickness, and hippocampal volumes [3, 5]. The latter is used to select the most discriminative features from those extracted features, according to some criteria. In the literature, many feature extraction and selection methods have been used for AD classification [3, 5-6].

Besides AD and HC subjects, in many cases we often have MR brain images available from other related subjects such as those with mild cognitive impairment (MCI), a prodromal stage of AD, or possibly the unrelated subjects whose cognitive statuses may be not known. For MCI subjects, although they are currently not belonging to the classes of AD and HC subjects, they may eventually convert to AD within some years, or never convert to AD. Figure 1 plots the distributions of AD, HC, and MCI subjects with CSF features. As we can see from this figure, there exists a large overlap between MCI subjects and AD/HC subjects. It implies that MCI subjects are similar to some extent as AD or HC subjects, and may contain helpful information for guiding the classification between AD and HC. To the best

of our knowledge, this type of study was not done previously.

In this paper, we exploit the potential of using MCI subjects to aid the classification of AD from HC subjects with multimodal imaging and CSF biomarkers. To this end, semi-supervised learning techniques that have been successfully applied to protein classification, aggressive prostate cancer identification, and skull stripping [8-10] will be employed. Specifically, we propose a multimodal Laplacian Regularized Least Squares (mLapRLS) algorithm, based on the recent LapRLS method [11], for better integration of knowledge from different types of training samples. In the objective function of mLapRLS, only AD and HC subjects will be used for supervised learning, while all AD, HC, and MCI subjects are used for unsupervised estimation on intrinsic geometric structure of the whole data.

The rest of this paper is organized as follows. Section 2 briefly reviews LapRLS and then introduces our proposed mLapRLS method. Section 3 presents experimental results on comparison of different classification methods using the Alzheimer's Disease Neuroimaging Initiative (ADNI) data. This paper is concluded in Section 4.

2. METHODS

To exploit the potential of using MCI subjects to aid classification between AD and HC subjects, we treat MCI subjects as *unlabeled* data (i.e., having no class labels either as AD or HC), and then use a semi-supervised learning technique to solve the classification problem. In the following, we will first introduce the Laplacian Regularized Least Squares (LapRLS) method [11], and then derive our multimodal extension (mLapRLS).

2.1. Laplacian Regularized Least Squares (LapRLS)

Assume we have l labeled data (from AD and HC samples), (x_i, y_i), i=1,…,l, and u unlabeled data (from MCI samples), (x_j, y_j), j=l+1,…,l+u. Suppose $k(.,.)$ is a Mercer kernel function, and let H be the associated Reproducing Kernel Hilbert Space (RKHS) and $\|.\|$ be the corresponding norm. The LapRLS algorithm solves the following least-squared loss function [11]:

$$\min_{f\in H} \frac{1}{l}\sum_{i=1}^{l}\left(y_i - f(x_i)\right)^2 + \gamma_A \|f\|^2 + \frac{\gamma_B}{(u+l)^2}\mathbf{f}^T L\mathbf{f} \qquad (1)$$

Where $\mathbf{f} = [f(x_1), \ldots, f(x_{l+u})]^T$. L is the graph Laplacian given as L=D-W, where W_{ij}s are the edge weights in the adjacency graph defined on both labeled and unlabeled data and the diagonal matrix D is given by D_{ii} =$\sum_j W_{ij}$. Symbols γ_A and γ_B are the two regularization parameters. Intuitively, the first two terms in Eq. 1 are for the supervised learning on only the labeled data (AD and HC samples), while the last term in Eq. 1 involves both labeled and unlabeled data (AD, HC and MCI samples) for unsupervised estimation on intrinsic geometric structure of the whole data.

According to the Representer Theorem [11], the solution of Eq. 1 is an expansion of kernel functions over both labeled and unlabeled data:

$$f(x) = \sum_{i=1}^{l+u} \alpha_i k(x, x_i) \qquad (2)$$

Substituting Eq. 2 into Eq. 1, we arrive at the dual form of Eq. 1 with respect to the (l+u)-dimensional variable vector $\alpha = [\alpha_1, \ldots, \alpha_{l+u}]^T$:

$$\min_{\alpha\in R^{l+u}} \frac{1}{l}(Y - JK\alpha)^T(Y - JK\alpha) + \gamma_A \alpha^T K\alpha + \frac{\gamma_B}{(u+l)^2}\alpha^T KLK\alpha \qquad (3)$$

Where K=$\{k(x_i, x_j)\}$ is an (l+u)×(l+u) kernel matrix over all labeled and unlabeled data; Y=[y_1, …, y_l, 0, … 0] is an (l+u)-dimensional label vector, and J=*diag* (1, …, 1, 0, …, 0) is an (l+u)×(l+u) diagonal matrix with the first l diagonal entries as 1 and the rest as 0.

By computing the derivative of Eq. 3 with respect to α, and let it be zero, we obtain the following solution:

$$\alpha = \left(JK + \gamma_A lI + \frac{\gamma_B l}{(u+l)^2} LK\right)^{-1} Y \qquad (4)$$

Where I is the identity matrix.

It is worth noting that, when γ_B=0, Eq. 4 gives zero coefficients over the unlabeled data, and the coefficients over the labeled data are exactly those in the standard Regularized Least Squares (RLS) method. In that case, LapRLS degenerates to RLS.

2.2. Multimodal LapRLS (mLapRLS)

Now, we extend LapRLS to the multimodal case (mLapRLS) for multimodal classification between AD and HC. Given l labeled data (from AD and HC samples), (x_i, y_i), i=1,…,l, and u unlabeled data (from MCI samples), (x_j, y_j), j=l+1,…,l+u, we assume each data x_i is now composed of M modalities of data, i.e., x_i=$\{x_i^{(1)}, \ldots, x_i^{(M)}\}$, i=1,…,l+u.

Define the distance function between two multimodal data x_i and x_j as

$$d(x_i, x_j) = \sum_{m=1}^{M} \beta_m d^{(m)}(x_i^{(m)}, x_j^{(m)}) \qquad (5)$$

Where $d^{(m)}(.,.)$ denotes the distance function on the m-th modality, and β_ms are the nonnegative weighting parameters used to balance the contributions of different modalities. All β_ms are constrained by $\sum_m \beta_m$=1.

According to Eq. 5, we can compute the adjacency graph for the multimodal data, and then obtain the corresponding edge weights matrix W and graph Laplacian L on the multimodal data.

Next, we can define the kernel function on two multimodal data x and x_i as

$$k(x, x_i) = \sum_{m=1}^{M} \beta_m k^{(m)}(x^{(m)}, x_i^{(m)}) \qquad (6)$$

Where $k^{(m)}$ denotes the kernel matrix over the m-th modality, similar to the definition given above for the single modality case. With the definition of $k(x, x_i)$, the $(l+u)\times(l+u)$ kernel matrix K on the multimodal data can be straightforwardly obtained as $K=\{k(x_i, x_j)\}$.

Once we have gotten the graph Laplacian L, the definition of the kernel function $k(x, x_i)$ on the multimodal data, and the kernel matrix K, the mLapRLS solution to the multimodal data can be obtained exactly the same as LapRLS in Eq. 4.

Similar to LapRLS, mLapRLS will degenerate to the corresponding multimodal RLS (mRLS) when $\gamma_B=0$. In this case, mRLS uses only AD and HC samples for training its model on the multimodal data.

3. RESULTS

To evaluate the effectiveness of our proposed mLapRLS method, we perform various experiments on the multimodal data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). We focus on multimodal classification in this paper, because it usually achieves better results than single-modal classification.

3.1. Subjects

The ADNI database contains approximately 200 cognitively normal elderly subjects to be followed for 3 years, 400 subjects with MCI to be followed for 3 years, and 200 subjects with early AD to be followed for 2 years. In this paper, only ADNI subjects with all corresponding MRI, PET, and CSF data at baseline are included. This yields a total of 202 subjects, including 51 AD patients, 99 MCI patients, and 52 healthy controls (HCs). Table 1 lists the subject characteristics, where MMSE and CDR are acronyms of Mini-Mental State Examination and Clinical Dementia Rating, respectively.

Image pre-processing is performed for all MRI and PET images. Specifically, we do anterior commissure (AC) - posterior commissure (PC) correction, skull-stripping, removal of cerebellum, and segmentation of structural MR images into three different tissues: grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). With atlas warping, we can partition each subject image into 93 ROIs. For each of the 93 ROIs, we compute the GM tissue volume from the subject's MRI image. For PET image, we first rigidly align it with its respective MRI image of the same subject, and then compute the average value of PET signals in each ROI. Therefore, for each subject, we can totally obtain 93 features from MRI image, other 93 features from PET image, and 3 features ($A\beta_{42}$, t-tau, and p-tau) from CSF biomarkers. Paired t-test is applied on both MRI and PET data for feature selection according to the p-values ($p<0.002$).

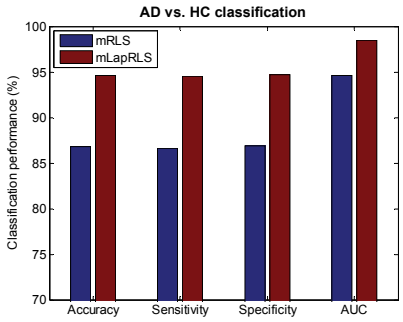

Fig. 2. Classification results on multimodal data.

Table 1. Subject characteristics (mean ± standard deviation)

	AD (n=51)	MCI (n=99)	HC (n=52)
Age	75.2 ± 7.4	75.3 ± 7.0	75.3 ± 5.2
Education	14.7 ± 3.6	15.9 ± 2.9	15.8 ± 3.2
MMSE	23.8 ± 2.0	27.1 ± 1.7	29 ± 1.2
CDR	0.7 ± 0.3	0.5 ± 0.0	0 ± 0.0

3.2. Experimental Setup

To evaluate the performances of different classification methods, we use 10-fold cross-validation strategy to compute the classification accuracy (for measuring the proportion of subjects correctly classified among the whole population), as well as the sensitivity (i.e., the proportion of AD patients correctly classified) and the specificity (i.e., the proportion of HC subjects correctly classified).

We compare mLapRLS with mRLS on the multimodal (MRI, PET, and CSF) data. Specifically, 10-fold cross-validation is performed on 51 AD patients and 52 HC subjects to get the labeled training data and testing data. Unlabeled data are obtained from those 99 MCI subjects. A linear kernel is used for both algorithms. Following [11], for mRLS, we set the parameters as $\gamma_A=0.05/l$ and $\gamma_B=0$; for mLapRLS, we set $\gamma_A=0.05/l$ and $\gamma_B=0.05(u+l)^2/l$. Here, l denotes the number of AD and HC subjects, and u is the number of MCI subjects. The Euclidean distance is used for each modality in Eq. 5. For both algorithms, the values of the weighting parameters β_ms are gotten through cross-validation using grid search.

3.3. Experimental Results

Figure 2 shows the classification results of both algorithms on the multimodal data, which include classification accuracy, sensitivity, specificity, and Area Under the ROC Curve (AUC). The results in Figure 2 indicate that, by using the MCI subjects as additional unlabeled data, mLapRLS significantly improves the performances of classifying AD from HC subjects, compared to those by mRLS that uses only AD and HC subjects as samples for training classifier.

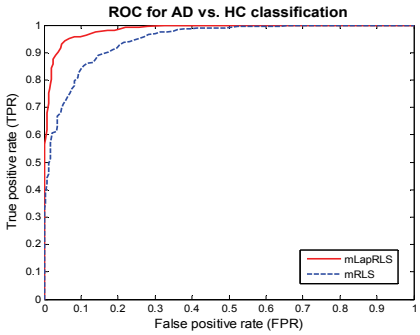

Fig. 3. ROC curves of different methods for AD vs. HC classification.

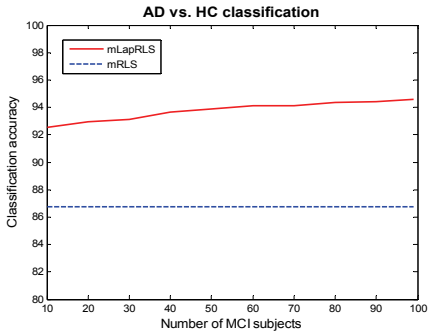

Fig. 4. Classification accuracy with respect to the different number of MCI subjects used to help train the multimodal classifier.

Specifically, the AUC measures of mLapRLS and mRLS are 98.5% and 94.6%, respectively. These results validate the effectiveness of mLapRLS in using additional data (i.e., MCI subjects) to enhance the AD classification.

Figure 3 plots the ROC curves of mLapRLS and mRLS for AD vs HC classification. As we can see from this figure, mLapRLS consistently outperforms mRLS. Figure 3 also shows that mLapRLS has a very steep curve, with the corresponding AUC measure being close to 1.0. Therefore, the mLapRLS algorithm achieves a very good performance on multimodal classification.

Finally, in Fig. 4, we show the classification accuracy of the mLapRLS algorithm with respect to different number of MCI subjects used for helping training. As we can see from Fig. 4, as the number of included MCI subjects increases, the classification accuracy of mLapRLS also steadily increases, which again validates the usefulness of using MCI subjects for helping classification between AD and HC.

4. CONCLUSION

This paper proposes using MCI subjects as additional data to enhance the classification between AD and HC. Specifically, a multimodal Laplacian Regularized Least Squares (mLapRLS) method has been proposed for semi-supervised multimodal classification, by including MCI subjects to aid the unsupervised estimation of intrinsic geometric structure of the data. The experimental results on ADNI data show that our proposed method can help significantly improve classification performance between AD and HC.

In the future, we will use other semi-supervised learning methods such as semi-supervised support vector machine (SVM) [10], and investigate selecting MCI subjects for further improvement of classification performance [12]. Moreover, we will apply our semi-supervised multimodal classification method for helping prediction of MCI conversion, by separating between MCI converters (who will convert into AD within some years) and MCI non-converters (who will not convert).

ACKNOWLEDGEMENT

This work was supported in part by NIH grants EB006733, EB008374, EB009634 and MH088520.

REFERENCES

[1] K.B. Walhovd, A.M. Fjell, et al., "Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease," *AJNR Am J Neuroradiol*, vol. 31, pp. 347-354, 2010.

[2] P. Vemuri, H.J. Wiste, "MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change," *Neurology*, vol. 73, pp. 294-301, 2009.

[3] R. Cuingnet, E. Gerardin, et al., "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *Neuroimage*, 2010, doi:10.1016/j.neuroimage.2010.06.013.

[4] S. Kloppel, C.M. Stonnington, et al., "Automatic classification of MR scans in Alzheimer's disease," *Brain*, vol. 131, pp. 681-689, 2008.

[5] Y. Fan, D. Shen, R.C. Gur, R.E. Gur, and C. Davatzikos, "COMPARE: Classification Of Morphological Patterns using Adaptive Regional Elements," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 93-105, 2007.

[6] J. Ye, K. Chen, et al., "Heterogeneous data fusion for Alzheimer's disease study," in *KDD*, pp. 1025-1033, 2008.

[7] C. Hinrichs, V. Singh, G. Xu, and S. Johnson, "MKL for robust multi-modality AD classification," in *MICCAI*, pp. 786-794, 2009.

[8] P. Tiwari, J. Kurhanewicz, M. Rosen, and A. Madabhushi, "Semi Supervised Multi Kernel (SeSMiK) Graph Embedding: Identifying Aggressive Prostate Cancer via Magnetic Resonance Imaging and Spectroscopy," in *MICCAI*, pp.666-673, 2010.

[9] J.E. Iglesias, C.Y. Liu, P. Thompson, and Z. Tu, "Agreement-Based Semi-supervised Learning for Skull Stripping," in *MICCAI*, pp. 147-154, 2010.

[10] O. Chapelle, B. Scholkopf, and A. Zien, *Semi-Supervised Learning*, MIT Press, Cambridge, MA, 2006.

[11] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples," *Journal of Machine Learning Research*, vol. 7, pp. 399- 2434, 2006.

[12] O. Querbes, F. Aubry, et al., "Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve," *Brain*, vol. 132, pp. 2036-2047, 2009.