Sparse Multimodal Manifold-Regularized Transfer Learning for MCI Conversion Prediction

Bo Cheng[1,2], Daoqiang Zhang[1,*], Biao Jie[1], and Dinggang Shen[2,*]

[1] Dept. of Computer Science and Engineering,
Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[2] Dept. of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599
dqzhang@nuaa.edu.cn, dgshen@med.unc.edu

Abstract. Effective prediction of conversion of mild cognitive impairment (MCI) to Alzheimer's disease (AD) is important for early diagnosis of AD, as well as for evaluating AD risk pre-symptomatically. Different from most traditional methods for MCI conversion prediction, in this paper, we propose a novel sparse multimodal manifold-regularized transfer learning classification (SM^2TLC) method, which can simultaneously use other related classification tasks (e.g., AD vs. normal controls (NC) classification) and also the unlabeled data for improving the MCI conversion prediction. Our proposed method includes two key components: (1) a criterion based on the maximum mean discrepancy (MMD) for eliminating the negative effect related to the distribution differences between the auxiliary (i.e., AD/NC) and the target (i.e., MCI converters/MCI non-converters) domains, and (2) a sparse semi-supervised manifold-regularized least squares classification method for utilization of unlabeled data. Experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) database show that the proposed method can significantly improve the classification performance between MCI converters and MCI non-converters, compared with the state-of-the-art methods.

1 Introduction

Alzheimer's disease (AD) is the most common type of dementia, which can be characterized by the progressive impairment of neurons and their connections, leading to the loss of cognitive function and the ultimate death. Mild cognitive impairment (MCI) is a prodromal stage of AD, with high likelihood of conversion to AD. Thus, effective prediction of MCI-to-AD conversion is of great significance for early diagnosis of AD and also for evaluating AD risk pre-symptomatically. So far, more and more machine learning methods have been developed for addressing the MCI conversion prediction, i.e., classification between MCI converters (MCI-C) and MCI non-converters (MCI-NC) based on the baseline imaging data [1-6].

One challenge in MCI conversion prediction is the small number of MCI (including both MCI-C and MCI-NC) subjects available for training, while the dimensionality of data is often very high. This makes it very difficult to train a robust and accurate classifier. To address this issue, many advanced machine learning methods have been

* Corresponding authors.

G. Wu et al. (Eds.): MLMI 2013, LNCS 8184, pp. 251–259, 2013.

proposed for MCI conversion prediction. For example, Zhang *et al.* used the multi-task learning for joint regression and classification from multimodal data, including magnetic resonance imaging (MRI), fluorodeoxyglucose positron emission tomography (FDG-PET) and cerebrospinal fluid (CSF), achieving an accuracy of 73.9% on classifying between 43 MCI-C and 48 MCI-NC subjects [2]. Cho *et al.* adopted a manifold harmonic transform method using the cortical thickness data, achieving a sensitivity of 63% and a specificity of 76% on 72 MCI-C and 131 MCI-NC subjects [1]. Duchesne *et al.* used the morphological factor method with the MRI data, achieving an accuracy of 72.3% on 20 MCI-C and 29 MCI-NC subjects [5].

On the other hand, some recent studies based on semi-supervised learning (SSL) methods have shown that the task of classifying between MCI-C and MCI-NC is related to the task of classifying between AD and normal controls (NC) [7]. In machine learning community, a new learning methodology called *transfer learning* has been developed to deal with the problems involving cross-domain learning. Unlike SSL, transfer learning does not assume the auxiliary data (i.e., unlabeled data in SSL) have the same distribution as the target data (i.e., labeled data in SSL). Recently, this transfer learning technique has been introduced into medical imaging area. For example, a domain transfer support vector machine (DTSVM) has been used for classification between MCI-C and MCI-NC subjects, showing greatly improved classification performance for MCI conversion prediction with help of AD and NC subjects used as the auxiliary data [8].

In this paper, we propose a novel sparse multimodal manifold-regularized transfer learning classification (SM^2TLC) framework for MCI conversion prediction. Specifically, we employ a criterion based on maximum mean discrepancy (MMD) for eliminating the negative effect due to the distribution differences between the auxiliary (AD/NC) and the target (MCI-C/MCI-NC) domains. Then, we combine with the sparse semi-supervised manifold-regularized least squares method for utilization of unlabeled subjects. Our proposed method can be applied to not only the single-modality data, but also the multimodal data. We validate our method on both single-modality and multimodal data, including MRI, FDG-PET and CSF, from the ADNI database.

2 Method

In this section, we introduce our proposed sparse multimodal manifold transfer learning classification method (SM^2TLC) for classifying MCI-C from MCI-NC. Specifically, in Section 2.1, we will first introduce a criterion based on the maximum mean discrepancy (MMD) for reducing the mismatch of distributions between the auxiliary (i.e., AD/NC) and the target (i.e., MCI-C/MCI-NC) domains. Then, we will derive the objective function and its corresponding optimization algorithm for the proposed SM^2TLC method in Sections 2.2 and 2.3, respectively.

2.1 Maximum Mean Discrepancy (MMD) Criterion

Assume that we have N_A samples with class labels in the auxiliary domain, denoted as $\boldsymbol{A} = \{\mathbf{x}_i^A, \mathrm{y}_i^A\}_{i=1}^{N_A}$, where $\mathbf{x}_i^A \in \boldsymbol{R}^d$ is the i-th sample and $\mathrm{y}_i^A \in \{+1, -1\}$ is its

corresponding class label (i.e., with the AD labeled as 1 and the NC labeled as -1). Also, assume that we have N_T^L *labeled target* samples with class labels, denoted as $\boldsymbol{T}^L = \{\mathbf{x}_i^L, \mathrm{y}_i^L\}_{i=1}^{N_T^L}$, where $\mathbf{x}_i^L \in \boldsymbol{R}^d$ is the i-th sample and $\mathrm{y}_i^L \in \{+1, -1\}$ is the corresponding class label (i.e., with the MCI-C labeled as 1 and the MCI-NC labeled as -1). Similarly, we have N_T^U *unlabeled target* samples, denote as $\boldsymbol{T}^U = \{\mathbf{x}_i^U\}_{i=1}^{N_T^U}$, where $\mathbf{x}_i^U \in \boldsymbol{R}^d$. We use $N_T = N_T^L + N_T^U$ to represent the total number of *target* samples $\boldsymbol{T} = \{\boldsymbol{T}^L \cup \boldsymbol{T}^U\}$. Because of the distribution differences between the auxiliary and the target domains, direct training with samples from the auxiliary domain may degrade the classification performance in another target domain. Therefore, we cannot directly add auxiliary data to target domain for training. To address this issue, we employ a criterion based on the maximum mean discrepancy (MMD) [9, 16]:

$$dist^2(\boldsymbol{A}, \boldsymbol{T}) = tr(\mathbf{KS}) \tag{1}$$

where

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}^{A,A} & \mathbf{K}^{A,T} \\ \mathbf{K}^{T,A} & \mathbf{K}^{T,T} \end{bmatrix} \tag{2}$$

and $\mathbf{S} = \mathbf{s}\mathbf{s}^T$, $\mathbf{s} = [\frac{1}{N_A}, \ldots, \frac{1}{N_A}, \frac{-1}{N_T}, \ldots, \frac{-1}{N_T}]^T$, and $tr(.)$ is the matrix trace [16]. Here, $\mathbf{K}$ denotes a compound cross-domain kernel matrix over both $\boldsymbol{A}$ and $\boldsymbol{T}$, with $\mathbf{K}^{A,A} = \left[\boldsymbol{k}\left(\mathbf{x}_i^A, \mathbf{x}_j^A\right)\right] \in \boldsymbol{R}^{N_A \times N_A}$, $\mathbf{K}^{T,T} = \left[\boldsymbol{k}\left(\mathbf{x}_i^T, \mathbf{x}_j^T\right)\right] \in \boldsymbol{R}^{N_T \times N_T}$, $\mathbf{K}^{A,T} = [\boldsymbol{k}\left(\mathbf{x}_i^A, \mathbf{x}_j^T\right)] \in \boldsymbol{R}^{N_A \times N_T}$, and $\mathbf{K}^{T,A} = [\boldsymbol{k}\left(\mathbf{x}_i^T, \mathbf{x}_j^A\right)] \in \boldsymbol{R}^{N_T \times N_A}$.

2.2 Objective Function of SM^{2}TLC

To explore related auxiliary domains data and the manifold structure of both labeled and unlabeled data from the target domain, we employ the semi-supervised manifold-regularized least squares method [15] and combine the criterion of MMD (i.e., Eq. 1) to design a sparse multimodal manifold-regularized transfer learning classification method (denoted as SM^{2}TLC), which can simultaneously use the multimodal biomarkers for learning a sparse multimodal weight matrix $\mathbf{W}$. To this end, SM^{2}TLC solves the following optimization problem with L_1/L_2-norm regularization:

$$\begin{aligned} \min_{\mathbf{W}} \frac{1}{M} \sum_{m=1}^{M} (d_m \cdot tr(\mathbf{K}_m \mathbf{S}) + (\mathbf{Y} - \boldsymbol{J}\mathbf{K}_m \mathbf{W}_m)^T (\mathbf{Y} - \boldsymbol{J}\mathbf{K}_m \mathbf{W}_m) \\ + \gamma \mathbf{W}_m^T \mathbf{K}_m^T \boldsymbol{\Lambda}_m \mathbf{K}_m \mathbf{W}_m) + \mu \|\mathbf{W}\|_{2,1} \end{aligned} \tag{3}$$

where $\mathbf{Y} = [\mathrm{y}_1^A, \ldots, \mathrm{y}_{N_A}^A, \mathrm{y}_1^L, \ldots, \mathrm{y}_{N_T^L}^L, 0_1, \ldots, 0_{N_T^U}]^T$, $\mathbf{K}_m$ is the compound cross-domain kernel matrix defined on the m-th modality according to Eq. 2 and d_m is a nonnegative weight parameter with $\sum_{m=1}^{M} d_m = 1$. And $\boldsymbol{J}=diag(1, \ldots, 1, 0, \ldots, 0)$ is diagonal matrix with the first $N_A + N_T^L$ diagonal entries as 1 and the rest as 0, and $\gamma, \mu > 0$ are the two regularization parameters. Specifically, there are three modalities (e.g., $M = 3$) used in this paper, including MRI, PET and CSF. In Eq. 3, $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, .., \mathbf{w}_M] \in \boldsymbol{R}^{N \times M}$, where $N = N_A + N_T$, is the weight matrix whose i-th

row $\mathbf{w}^i$ is the vector of coefficients associated with the i-th training sample across different modalities. It's worth noting that a 'group sparsity' regularization term is used for joint selection of samples across different modalities based on the L_1/L_2-norm, i.e., $\|\mathbf{W}\|_{2,1} = \sum_{i=1}^{N} \left\|\mathbf{w}^i\right\|_2$. Also, $\mathbf{\Lambda}_m$ is the compound cross-domain Laplacian matrix on the m-th modality, which is defined as:

$$\mathbf{\Lambda}_m = \begin{bmatrix} \mathbf{\Lambda}_m^A & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_m^T \end{bmatrix} \tag{4}$$

where $\mathbf{\Lambda}_m^A = \boldsymbol{D}_m^A - \boldsymbol{C}_m^A$ and $\mathbf{\Lambda}_m^T = \boldsymbol{D}_m^T - \boldsymbol{C}_m^T$ are the Laplacian matrices [15] over the auxiliary domain and the target domain, respectively. Here, $\boldsymbol{C}_m^A = [c_{ij}^A] \in \boldsymbol{R}^{N_A \times N_A}$ and $\boldsymbol{C}_m^T = [c_{ij}^T] \in \boldsymbol{R}^{N_T \times N_T}$ are the similarity matrices that respectively define the similarity of subjects in the auxiliary domain and the target domain, while $\boldsymbol{D}_m^A = [d_{ii}^A] \in \boldsymbol{R}^{N_A \times N_A}$ and $\boldsymbol{D}_m^T = [d_{ii}^T] \in \boldsymbol{R}^{N_T \times N_T}$ are the diagonal matrices with $d_{ii}^A = \sum_j c_{ij}^A$ and $d_{ii}^T = \sum_j c_{ij}^T$. By minimizing Eq. 3, many rows of $\mathbf{W}$ will be zeros, thus we can obtain a sparse solution for $\mathbf{W}$. Then, the decision function $f^*(\mathbf{x})$ for the predicted label in the target domain can be obtained as follows:

$$f^*(\mathbf{x}) = sign\left(\sum_{m=1}^{M} d_m \mathbf{K}_m^* \mathbf{w}_m^*\right) \tag{5}$$

where we assume that $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$ is a testing sample's multimodal data in the target domain, $\mathbf{K}_m^* = [\boldsymbol{k}(\mathbf{x}_m, \mathbf{x}_m^i)]_{i=1}^N \in \boldsymbol{R}^{1 \times N}$ is the testing sample's kernel vector on the m-th modality (between the testing sample $\mathbf{x}_m$ and the i-th training samples $\mathbf{x}_m^i$ in the cross-domain on the m-th modality), and $\mathbf{W}^* = [\mathbf{w}_1^*, \mathbf{w}_2^*, \ldots, \mathbf{w}_M^*]$ is the optimal solution obtained from Eq. 3. Thus, we can get the predicted label $f^*(\mathbf{x})$ for the testing sample $\mathbf{x}$ via the Eq. 5 to compute.

2.3 Optimization Algorithm for SM^{2}TLC

To optimize the problem in Eq. 3, the accelerated gradient descent (AGD) method is employed to solve the optimization problem with L_1/L_2-norm regularization [10]. According to the AGD algorithm, the objective function (denoted as $F(\mathbf{W})$) of Eq. 3 can be separated into a smooth part:

$$G(\mathbf{W}) = \frac{1}{M}\sum_{m=1}^{M}(d_m \cdot tr(\mathbf{K}_m\mathbf{S}) + (\mathbf{Y} - \boldsymbol{J}\mathbf{K}_m\mathbf{w}_{\boldsymbol{m}})^T(\mathbf{Y} - \boldsymbol{J}\mathbf{K}_m\mathbf{w}_m) + \gamma\mathbf{w}_m^T\mathbf{K}_m^T\mathbf{\Lambda}_m\mathbf{K}_m\mathbf{w}_m) \tag{6}$$

and a non-smooth part:

$$H(\mathbf{W}) = \mu\|\mathbf{W}\|_{2,1} \tag{7}$$

So, $F(\mathbf{W}) = G(\mathbf{W}) + H(\mathbf{W})$. Then, we define the generalized gradient update step to solve the Eq. 3 as follows:

$$Q_h(\mathbf{W}, \mathbf{W}_t) = G(\mathbf{W}_t) + \langle \mathbf{W} - \mathbf{W}_t, \nabla G(\mathbf{W}_t) \rangle + \frac{h}{2} \|\mathbf{W} - \mathbf{W}_t\|_F^2 + H(\mathbf{W})$$
$$q_h(\mathbf{W}_t) = \arg min_{\mathbf{W}} Q_h(\mathbf{W}, \mathbf{W}_t) \tag{8}$$

Where $\nabla G(\mathbf{W}_t)$ denotes the gradient of $G(\mathbf{W})$ at point $\mathbf{W}_t$ for the t-th iteration, h is the step size, $\|\cdot\|_F$ denotes the Frobenius norm, and $\langle \mathbf{W} - \mathbf{W}_t, \nabla G(\mathbf{W}_t) \rangle = tr((\mathbf{W} - \mathbf{W}_t)^T \nabla G(\mathbf{W}_t))$ is the matrix inner product. According to [10], the generalized gradient update step of Eq. 8 can be decomposed into N separate subproblems with the gradient mapping update approach. In summary, the AGD algorithm to solve the objective function of Eq. 3 is presented in Algorithm 1.

Algorithm 1. AGD algorithm for SM^{2}TLC

Initialization: $h_0 > 0, \eta > 1, \mathbf{W}_0 \in \boldsymbol{R}^{N \times M}, \bar{\mathbf{W}}_0 = \mathbf{W}_0, h{=}h_0$ and $\alpha_0 = 1$.
for $t = 0,1,2,\ldots$ until convergence of $\mathbf{W}_t$ do:
Set $h = h_t$
While $F\left(q_h(\bar{\mathbf{W}}_t)\right) > Q_h(q_h(\bar{\mathbf{W}}_t), \bar{\mathbf{W}}_t)$, $h = \eta h$
Set $h_{t+1} = h$ and compute
$\mathbf{W}_{t+1} = \arg min_{\mathbf{W}} Q_{h_{t+1}}(\mathbf{W}, \bar{\mathbf{W}}_t)$, $\alpha_{t+1} = \frac{2}{t+3}$, $\boldsymbol{\beta}_{t+1} = \mathbf{W}_{t+1} - \mathbf{W}_t$ and
$\bar{\mathbf{W}}_{t+1} = \mathbf{W}_{t+1} + \frac{1-\alpha_t}{\alpha_t} \alpha_{t+1} \boldsymbol{\beta}_{t+1}$
end for

3 Experiments

In this section, we evaluate the effectiveness of our proposed SM^{2}TLC method on multimodal data, including MRI, PET and CSF, from the Alzheimer's disease Neuroimaging Initiative (ADNI) database.

3.1 Experimental Settings

In our experiments, the baseline ADNI subjects with all corresponding MRI, PET, and CSF data are included, which leads to a total of 202 subjects (including 51 AD patients, 99 MCI patients, and 52 normal controls (NC)). For the 99 MCI patients, it includes 43 MCI converters and 56 MCI non-converters. Also, for each of the three modalities, we include another set of data from 153 randomly selected subjects as unlabeled data. We use 51 AD and 52 NC subjects as auxiliary domain, and 99 MCI subjects as target domain.

The same image pre-processing as used in [11] is adopted here. First, we do AC-PC (anterior commissure-posterior commissure) correction on all images using MIPAV software, and then resample the images to $256 \times 256 \times 256$. Automated skull-stripping is performed [12], followed by manual review to ensure clean skull and dura removal. Then the cerebellum is removed based on registration, in which we use a manually labeled cerebellum as template. After intensity inhomogeneity correction (which was done for three times by the N3 algorithm), we use FAST in

FSL to segment the human brain into three different tissues: grey matter (GM), white matter (WM) and Cerebrospinal fluid (CSF). We use HAMMER [13] to do the registration. After registration, we get the subject-labeled image based on the Jacob template, which was manually labeled into 93 ROIs. For each of the 93 ROI regions in the labeled image of one subject, we compute the GM/WM/CSF tissue volumes in this ROI region by combining the segmentation result of this subject. For each subject, we first align the PET image to its respective T1 MR image by affine registration. Then, we get the average intensity of every ROI in the PET image as feature.

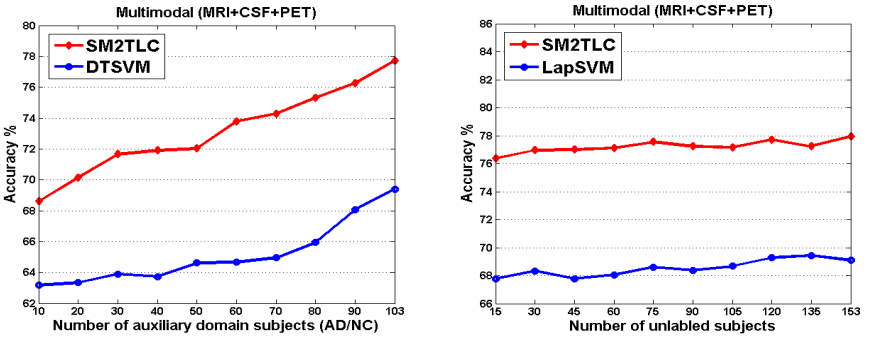
To evaluate the performance of different classification methods, we use a 10-fold cross-validation strategy to compute the classification AUC (area under the ROC curve), accuracy, sensitivity, and specificity. In particular, in the target domain, 99 labeled MCI subjects are equally partitioned into 10 subsets, and then one subset is successively selected as the testing samples and all remaining subsets are used for training classifiers. This process is repeated 10 times. The SVM classifier is implemented using the LIBSVM toolbox [14], with a linear kernel and a default value for the parameter C (i.e., $C = 1$). For comparison, LapSVM is also adopted in this paper. LapSVM is a typical semi-supervised learning method based on manifold hypothesis [15]. For LapSVM settings, we use linear kernel, and the Laplacian matrix $\mathbf{\Lambda}$ with N_T nodes are connected using k (i.e., k=5) nearest neighbors, and their edge weights are calculated using the *Euclidean* distance among samples. For regularization parameters γ and β, $\gamma, \beta \in \{0.001, 0.01, 0.03, 0.06, 0.09, 0.1, 0.2, 0.4, 0.6, 0.8\}$, they are learned based on the training samples by cross-validation, respectively. In addition, our previous transfer learning based method (i.e., DTSVM) [8] is also included here for comparison. We use both single-modality and multimodal data to evaluate our method. For combining multimodal data in both standard SVM and LapSVM methods and also for computing d_m, we specifically use a multi-kernel combination technique [11], with the weights learned from the training samples through a grid search, using the range from 0 to 1 at a step size of 0.1. Also, for features of each modality, the same feature normalization scheme as used in [11] is adopted here.

3.2 Results

We compare our SM^{2}TLC method with DTSVM [8], LapSVM, and standard SVM (SVM) for both single-modality and multimodal cases. Table 1 shows their classification performance measures on different modalities. Note that Table 1 shows the averaged results of 10 independent experiments. As we can see from Table 1, SM^{2}TLC consistently achieves better results than DTSVM, LapSVM and SVM methods on each performance measure, which validates the efficacy of our SM^{2}TLC method on using AD and NC subjects as auxiliary domains for helping SSL classification. Specifically, for multimodal case, SM^{2}TLC can achieve a classification accuracy of 77.8%, which is significantly better than DTSVM, LapSVM and SVM

Table 1. Comparison of performance measures of SM^{2}TLC, DTSVM, LapSVM, and SVM for MCI-C/MCI-NC classification using different modalities. (ACC= Accuracy, SEN=Sensitivity, SPE= Specificity).

Modality	Methods	ACC %	SEN %	SPE %	AUC
MRI+CSF+PET	SM^{2}TLC	**77.8**	**83.9**	69.8	**0.814**
	DTSVM	69.4	64.3	**73.5**	0.736
	LapSVM	69.1	74.3	62.1	0.751
	SVM	63.8	58.8	67.7	0.683
MRI	SM^{2}TLC	**72.1**	**75.1**	**68.2**	**0.768**
	DTSVM	63.3	59.8	66.0	0.700
	LapSVM	65.9	69.6	61.0	0.686
	SVM	53.9	47.6	57.7	0.554
CSF	SM^{2}TLC	**66.7**	**74.6**	60.5	0.668
	DTSVM	66.2	60.3	**70.8**	**0.701**
	LapSVM	62.1	66.2	56.8	0.660
	SVM	60.8	55.2	65.0	0.647
PET	SM^{2}TLC	**68.1**	**71.5**	63.7	**0.734**
	DTSVM	67.0	59.6	**72.7**	0.732
	LapSVM	61.6	65.7	56.1	0.661
	SVM	58.0	52.1	62.5	0.612



(a) (b)

Fig. 1. Comparison of classification accuracy (a) between SM^{2}TLC and DTSVM with respect to the use of different number of subjects in the auxiliary domain, and (b) between SM^{2}TLC and LapSVM with respect to the use of different number of unlabeled subjects

which achieve only 69.4%, 69.1% and 63.8%, respectively. Finally, in Fig. 1, we compare SM^{2}TLC with DTSVM [8] for classification accuracy, with respect to the use of different number of subjects in the auxiliary domain, and compare SM^{2}TLC with LapSVM for classification accuracy with respect to the use of different number of unlabeled subjects, respectively. As we can see from Fig. 1, in most cases, the performance of SM^{2}TLC is significantly better than DTSVM as the number of subjects

in the auxiliary domain increases. On the other hand, the performance of SM^{2}TLC is significantly better than LapSVM as the number of unlabeled subjects increases. These results further validate the efficacy of our proposed SM^{2}TLC method.

4 Conclusion

This paper addresses the problem of exploiting the use of auxiliary domain data (i.e., AD/NC) and unlabeled subjects for helping classifying MCI converters (MCI-C) from MCI non-converters (MCI-NC). By integrating the criterion based on the maximum mean discrepancy (MMD) and the sparse semi-supervised manifold-regularized least squares classification, we developed a sparse multimodal manifold-regularized transfer learning classification method, namely SM^{2}TLC, for MCI conversion prediction. Our method does not require the auxiliary domain data and the target domain data to be from the same distribution. Experimental results on the ADNI database validate the efficacy of our proposed method.

Acknowledgements. This work was supported in part by NIH grants EB006733, EB008374, EB009634, and AG041721, SRFDP grant (No. 20123218110009), NUAAFRF grant (No. NE2013105), JiangsuSF for Distinguished Young Scholar, CQESTP grant (Nos. KJ121111, KJ131108), and UNSFA grant (No. KJ2013Z095).

References

1. Cho, Y., Seong, J.K., Jeong, Y., Shin, S.Y.: Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. NeuroImage 59, 2217–2230 (2012)
2. Zhang, D., Shen, D., Alzheimer's Disease Neuroimaging Initiative: Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage 59, 895-907 (2012)
3. Lehmann, M., Koedam, E.L., Barnes, J., Bartlett, J.W., Barkhof, F., Wattjes, M.P., Schott, J.M., Scheltens, P., Fox, N.C.: Visual ratings of atrophy in MCI: prediction of conversion and relationship with CSF biomarkers. Neurobiology of Aging 34, 73–82 (2012)
4. Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q.: Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiology of Aging 32, 2322.e19–2322.e27 (2011)
5. Duchesne, S., Mouiha, A.: Morphological factor estimation via high-dimensional reduction: prediction of MCI conversion to probable AD. International Journal of Alzheimer's Disease 2011, 914085 (2011)
6. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O.: Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. NeuroImage 56, 766–781 (2011)
7. Filipovych, R., Davatzikos, C.: Semi-supervised pattern classification of medical images: Application to mild cognitive impairment (MCI). NeuroImage 55(3), 1109–1119 (2011)

8. Cheng, B., Zhang, D., Shen, D.: Domain transfer learning for MCI conversion prediction. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 82–90. Springer, Heidelberg (2012)
9. Borgwardt, K.M., Gretton, A., Rasch, M.J., Kriegel, H.P., Scholkopf, B., Smola, A.J.: Integrating structured biological data by kernel maximum mean discrepancy. Bioinformatics 22, e49–e57 (2006)
10. Chen, X., Pan, W., Kwok, J.T., Carbonell, J.G.: Accelerated gradient method for multi-task sparse learning problem. In: Proceedings Ninth IEEE International Conference on Data Mining, ICDM 2009, pp. 746–751 (2009)
11. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage 55(3), 856–867 (2011)
12. Wang, Y., Nie, J., Yap, P.-T., Shi, F., Guo, L., Shen, D.: Robust Deformable-Surface-Based Skull-Stripping for Large-Scale Studies. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 635–642. Springer, Heidelberg (2011)
13. Shen, D., Davatzikos, C.: HAMMER: Hierarchical attribute matching mechanism for elastic registration. IEEE Transactions on Medical Imaging 21, 1421–1439 (2002)
14. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001)
15. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. Journal of Machine Learning Research 7, 2399–2434 (2006)
16. Duan, L., Xu, D., Tsang, I., Xu, D.: Domain Transfer Multiple Kernel Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(3), 465–479 (2012)