NeuroImage 84 (2014) 466–475

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification

Feng Liu [a,b], Chong-Yaw Wee [b], Huafu Chen [a], Dinggang Shen [b,c,*]

[a] Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China
[b] Image Display, Enhancement, and Analysis (IDEA) Laboratory, Biomedical Research Imaging Center (BRIC) and Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[c] Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Accepted 8 September 2013
Available online 14 September 2013

Keywords:
Alzheimer's Disease
Mild cognitive impairment
Multi-kernel support vector machine
Multi-task feature selection
Inter-modality relationship

ABSTRACT

Previous studies have demonstrated that the use of integrated information from multi-modalities could significantly improve diagnosis of Alzheimer's Disease (AD). However, feature selection, which is one of the most important steps in classification, is typically performed separately for each modality, which ignores the potentially strong inter-modality relationship within each subject. Recent emergence of multi-task learning approach makes the joint feature selection from different modalities possible. However, joint feature selection may unfortunately overlook different *yet* complementary information conveyed by different modalities. We propose a novel multi-task feature selection method to preserve the complementary inter-modality information. Specifically, we treat feature selection from each modality as a separate task and further impose a constraint for preserving the inter-modality relationship, besides separately enforcing the sparseness of the selected features from each modality. After feature selection, a multi-kernel support vector machine (SVM) is further used to integrate the selected features from each modality for classification. Our method is evaluated using the baseline PET and MRI images of subjects obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Our method achieves a good performance, with an accuracy of 94.37% and an area under the ROC curve (AUC) of 0.9724 for AD identification, and also an accuracy of 78.80% and an AUC of 0.8284 for mild cognitive impairment (MCI) identification. Moreover, the proposed method achieves an accuracy of 67.83% and an AUC of 0.6957 for separating between MCI converters and MCI non-converters (to AD). These performances demonstrate the superiority of the proposed method over the state-of-the-art classification methods.

Introduction

Alzheimer's Disease (AD) is a genetically complex and irreversible neurodegenerative disorder which is clinically characterized by progressive dementia and neuropsychiatric symptoms (Blennow et al., 2006). The number of subjects with AD has been predicted to quadruple by 2050 (Brookmeyer et al., 2007). Although numerous efforts have been made in the past decades to develop new treatment strategies, there is no effective treatment or effective diagnostic instrument until now. This causes substantial financial burden to the society, as well as psychological and emotional burden to patients and their families. Mild cognitive impairment (MCI), an intermediate stage between normal cognition and dementia, has a high risk of progressing to AD (Petersen et al., 1999). While the annual incidence rate of healthy subjects to develop AD is 1% to 2% (Bischkopf et al., 2002), the conversion rate from MCI to AD is reported to be approximately 10% to 15% per year (Grundman et al., 2004). Thus, it is of great interest to identify MCI and also predict its risk of progressing to AD.

Accumulating evidence demonstrates that individuals with AD have both functional and structural changes in the brain, such as loss of gray matter volume (Karas et al., 2003) and metabolic abnormalities (Matsuda, 2001). However, these findings are mainly obtained based on group-level statistical comparison, and thus are of limited value for individual-based disease diagnosis. To overcome this limitation, pattern classification methods have been used in recent years, and have shown great potential in neuroimaging studies (Fan et al., 2007; Wee et al., in press). Unlike group-based comparison approaches, pattern classification methods are able to detect the fine-grained spatial discriminative patterns, which are critical for individual-based disease diagnosis. Moreover, some studies have shown that the combination of complementary information from different imaging modalities can improve the accuracy in diagnosis of AD and MCI. For example, Z. Dai et al. (2012) used both structural Magnetic Resonance Imaging (MRI) and

* Corresponding author at: Image Display, Enhancement, and Analysis (IDEA) Laboratory, Biomedical Research Imaging Center (BRIC) and Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
E-mail address: dgshen@med.unc.edu (D. Shen).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.09.015

resting-state functional MRI to classify 16 AD patients from 22 healthy subjects, and achieved a classification accuracy of 89.47%, which is an increase of 2.63% from the single-modality based method. Wee et al. (2012) used both Diffusion Tensor Imaging (DTI) and resting-state functional MRI to identify 10 individuals with MCI from 17 matched Normal Controls (NC), and obtained a very promising classification accuracy of 96.3%, which is an increase of 7.4% from the single-modality based method. Although high classification accuracies were achieved, the small sample and large feature size problem in these studies may still lead to data overfitting. Since the original feature space of neuroimaging data is relatively high compared to the sample size, feature selection is one of the most important steps in neuroimaging classification. However, in the literature, feature selection in multimodal classification studies is often performed separately for each imaging modality without considering the potentially strong relationship among different modalities, thus possibly affecting the final classification results. Hence, it is reasonable to consider preserving the inter-modality relationship during the feature selection for final improvement of classification.

Recently, multi-task learning approach has attracted the increasing attention in machine learning, computer vision, and artificial intelligence (Evgeniou and Pontil, 2007; Zhou et al., 2011). The main goal of this approach is to capture the intrinsic relationship among different tasks with the assumption that these tasks are related to each other. Learning multiple related tasks simultaneously has been shown to often perform better than learning each task separately (Evgeniou and Pontil, 2007). Specially, recent emergence of multi-task learning method enables joint feature selection via group sparsity (i.e. using $L_{2,1}$ norm) (Liu et al., 2009). However, this constraint may be too strong, since it forces common features to be selected for different tasks, without considering that different tasks may need different features.

In this paper, a novel multi-task learning based feature selection method is proposed to better preserve the complementary information conveyed by different modalities. More specifically, selection of features from different modalities is treated as different tasks. To better capture the complementary information among different modalities, it is also important to preserve the relationship between the feature vectors derived from different modalities, especially after their projection onto the lower dimensional feature space. To this end, a new constraint is imposed to preserve the inter-modality relationship, besides enforcing the sparseness of the selected features from each modality as popularly used in the literature. A multi-kernel support vector machine (SVM) is finally used to combine the selected features from each modality for predicting the classification labels.

The remainder of this paper is organized as follows. Materials and methods section furnishes information on the image dataset, preprocessing pipeline, and details of the proposed framework. Experimental results of the proposed method and some state-of-the-art methods on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset are summarized in the Experiment and results section. The findings of proposed framework are extensively discussed in the Discussion section, which is followed by the Conclusion section.

Materials and methods

Subjects

The data were taken from the ADNI dataset (www.adni.loni.ucla.edu/ADNI). 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 nonprofit organizations, as a $60 million, 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 Dr. Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators 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 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years.

All the patients met the following inclusion criteria: 1) diagnosis of AD was made if the subject had a Mini-Mental State Examination (MMSE) score of 20–26, a Clinical Dementia Rating (CDR) of 0.5 or 1.0, and meets the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria for probable AD. 2) Individuals were categorized as amnestic MCI if they had an MMSE score of 24–30, a CDR of 0.5, a memory complaint, objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, absence of significant levels of impairment in other cognitive domains, while essentially preserved activities of daily living, and an absence of dementia. All NC individuals met the following criteria: an MMSE score of 24–30, a CDR of 0, nondepressed, non-MCI, and nondemented. The research protocol was approved by each local institutional review board, and written informed consent was obtained from each subject at the time of enrollment for imaging and genetic sample collection.

Two hundreds and two subjects from ADNI dataset: 51 patients with AD, 99 patients with MCI (43 MCI converters who had converted to AD within 18 months and 56 MCI non-converters who had not converted to AD within 18 months), and 52 NC are analyzed in this study. Table 1 presents a summary of the demographic characteristics of the used subjects.

Data acquisition and preprocessing

All structural MRI scans used were acquired using 1.5 T scanners. MRI acquisitions were performed according to the ADNI acquisition protocol (Jack et al., 2008). For the image preprocessing, we first performed Anterior Commissure–Posterior Commissure (AC–PC) correction on all images, and then used N3 algorithm (Sled et al., 1998) to correct intensity inhomogeneity. Skull-stripping (Wang et al., 2011) was then performed, followed by the registration-based cerebellum removal. We used FAST in FSL (Zhang et al., 2001) to segment brain into three different tissue types: gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). We registered the template into subject specific space using HAMMER (Shen and Davatzikos, 2002) to preserve the absolute image volume of each subjects. Subsequently, we parcellated the brain in 93 regions-of-interest, based on the Jacob template (Kabani, 1998). GM volume of each ROI was extracted as the feature for the MRI modality. PET images were acquired 30–60 min post-injection, averaged, spatially aligned, interpolated to a standard voxel size, intensity normalized, and smoothed to a common resolution of 8-mm full width at half maximum. For each subject, we aligned the preprocessed PET image to its respective MRI image using affine registration. Then we calculated the average intensity of each regions-of-

Table 1
Characteristics of the subjects used in this study.

Characteristics	AD ($n = 51$)	MCI ($n = 99$)	NC ($n = 52$)
Gender (M/F)	33/18	67/32	34/18
Age (mean ± SD)	75.2 ± 7.4	75.3 ± 7.0	75.3 ± 5.2
Education (mean ± SD)	14.7 ± 3.6	15.9 ± 2.9	15.8 ± 3.2
MMSE (mean ± SD)	23.8 ± 2.0	27.1 ± 1.7	29.0 ± 1.2