NeuroImage 78 (2013) 233–248

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Modeling disease progression via multi-task learning

Jiayu Zhou [a,b], Jun Liu [a,b], Vaibhav A. Narayan [c], Jieping Ye [a,b,*], for the Alzheimer's Disease Neuroimaging Initiative [1]

[a] Center for Evolutionary Medicine and Informatics, The Biodesign Institute, ASU, Tempe, AZ, USA
[b] Department of Computer Science and Engineering, ASU, Tempe, AZ, USA
[c] Johnson & Johnson Pharmaceutical Research & Development, LLC, Titusville, NJ, USA

ARTICLE INFO

Article history:
Accepted 28 March 2013
Available online 12 April 2013

Keywords:
Alzheimer's disease
Disease progression
Multi-task learning
Fused Lasso
MMSE
ADAS-Cog

ABSTRACT

Alzheimer's disease (AD), the most common type of dementia, is a severe neurodegenerative disorder. Identifying biomarkers that can track the progress of the disease has recently received increasing attentions in AD research. An accurate prediction of disease progression would facilitate optimal decision-making for clinicians and patients. A definitive diagnosis of AD requires autopsy confirmation, thus many clinical/cognitive measures including Mini Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale cognitive subscale (ADAS-Cog) have been designed to evaluate the cognitive status of the patients and used as important criteria for clinical diagnosis of probable AD. In this paper, we consider the problem of predicting disease progression measured by the cognitive scores and selecting biomarkers predictive of the progression. Specifically, we formulate the prediction problem as a multi-task regression problem by considering the prediction at each time point as a task and propose two novel multi-task learning formulations. We have performed extensive experiments using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Specifically, we use the baseline MRI features to predict MMSE/ADAS-Cog scores in the next 4 years. Results demonstrate the effectiveness of the proposed multi-task learning formulations for disease progression in comparison with single-task learning algorithms including ridge regression and Lasso. We also perform longitudinal stability selection to identify and analyze the temporal patterns of biomarkers in disease progression. We observe that cortical thickness average of left middle temporal, cortical thickness average of left and right Entorhinal, and white matter volume of left Hippocampus play significant roles in predicting ADAS-Cog at all time points. We also observe that several MRI biomarkers provide significant information for predicting MMSE scores for the first 2 years, however very few are shown to be significant in predicting MMSE score at later stages. The lack of predictable MRI biomarkers in later stages may contribute to the lower prediction performance of MMSE than that of ADAS-Cog in our study and other related studies.

Introduction

Alzheimer's disease (AD), a severe neurodegenerative disorder, is characterized by loss of memory and reduction of cognitive function due to progressive impairment of neurons and their connections, leading directly to death (Khachaturian, 1985). AD accounts for 60–70% of age-related dementia; it currently affects about 5.3 million individuals in United States and more than 30 million worldwide and the number is projected to be over 114 million by 2050 (A. Association, 2010; Wimo et al., 2003). Alzheimer's disease has been not only the substantial financial burden to the health care system but also the psychological and emotional burdens to patients and their families. Currently there is no cure for Alzheimer's and efforts are underway to develop sensitive and consistent biomarkers for AD. In order to better understand the disease, an important area that has recently received increasing attention is to understand how the disease progresses and identify related pathological biomarkers for the progression. Realizing its importance, NIH in 2003 funded the Alzheimer's Disease Neuroimaging Initiative (ADNI). The initiative is facilitating the scientific evaluation of neuroimaging data including magnetic resonance imaging (MRI), positron emission tomography (PET), other biomarkers, and clinical and neuropsychological assessments for predicting the onset and progression of MCI (Mild Cognitive Impairment) and AD. The identification of sensitive and specific markers of very early AD progression will facilitate the diagnosis of early AD and the development, assessment, and monitoring of new treatments.

* Corresponding author at: Department of Computer Science and Engineering, Center for Evolutionary Medicine and Informatics, The Biodesign Institute, Arizona State University, 699 S. Mill Ave, Tempe, AZ 85287, USA.
E-mail address: jieping.ye@asu.edu (J. Ye).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.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.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.03.073

A definitive diagnosis of AD can only be made through an analysis of brain tissue during a brain biopsy or autopsy (Jeffrey et al., 2003). Many clinical/cognitive measures such as Mini Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale cognitive subscale (ADAS-Cog) have been designed to evaluate the cognitive status of the patients and they have been used as important criteria for clinical diagnosis of probable AD (McKhann et al., 1984). Previous studies have shown the correlation between MMSE and the underlying AD pathology and progressive deterioration of functional ability (Jeffrey et al., 2003). ADAS-Cog is the gold standard in AD drug trial for cognitive function assessment (Rosen et al., 1984). Since neurodegeneration of AD proceeds years before the onset of the disease and the therapeutic intervention is more effective in the early stage of the disease, there is thus an urgent need to (1) accurately predict the progression of the disease measured by cognitive scores, e.g., MMSE and ADAS-Cog, and (2) identify a small set of biomarkers (measurements) and risk factors most predictive of the progression. The prime candidate biomarkers and risk factors for tracking disease progression include neuroimages such as MRI, cerebrospinal fluid (CSF), and baseline clinical assessments (Dubois et al., 2007).

Several previous works have studied the relationship between the cognitive scores and possible risk factors such as age, APOE gene, years of education and gender (Ito et al., 2010; Tombaugh, 2005). The relationship between cognitive scores and imaging markers based on MRI such as gray matter volumes, density and loss (Apostolova et al., 2006; Chetelat and Baron, 2003; Frisoni et al., 2002; Frisoni et al., 2010; Stonnington et al., 2010), shape of ventricles (Ferrarini et al., 2008; Thompson et al., 2004) and hippocampal (Thompson et al., 2004) has been explored by correlating these features with baseline MMSE scores. Duchesne et al. showed that the intensity and volume of medial temporal lobe altogether with other risk factors and the gray matter were correlated with the one-year MMSE score (Duchesne et al., 2009), which allowed us to predict near-future clinical scores of patients. Murphy et al. examined the relations between 6-month atrophy patterns in medial temporal region and memory reduction in terms of clinical scores (Murphy et al., 2010). To predict the longitudinal response to AD progression, Ashford and Schmitt built a model with horologic function using “time-index” to measure the rate of dementia progression (Ashford and Schmitt, 2001). In (Davatzikos et al., 2009), the so-called SPARE-AD index was proposed based on spatial patterns of brain atrophy and its linear effect against MMSE was reported. In a more recent study, Ito et al. modeled the progression rate of cognitive scores using power functions (Ito et al., 2010).

There are two types of progression models that have been commonly used in the literature: the regression model (Duchesne et al., 2009; Stonnington et al., 2010) and the survival model (Pearson et al., 2005; Vemuri et al., 2009). The correlation between the ground truth and the prediction, and the squared error between the two are commonly used to evaluate the progression models (Duchesne et al., 2009; Stonnington et al., 2010). Many existing works consider a small number of input features, and the model building involves an iterative process in which the features are added to the model sequentially (Ito et al., 2010; Walhovd et al., 2010); alternatively, univariate analysis is performed individually on all covariates and those who exceed a certain significance threshold are included in the model (Murphy et al., 2010). For high-dimensional data, such as neuroimages (i.e., MRI and/or PET), the methods of sequentially evaluating individual features are suboptimal. In such cases, dimension reduction techniques such as principle component analysis are commonly applied to project the data into a lower-dimensional space (Duchesne et al., 2009). One disadvantage of dimension reduction is that the models are no longer interpretable. A better alternative is to use feature selection in modeling the disease progression (Stonnington et al., 2010). Most existing works focus on the prediction of target at a single time point (baseline Stonnington et al., 2010, or one year Duchesne et al., 2009); however, a joint analysis of the tasks from multiple time points is expected to improve the performance especially when the number of subjects is small and the number of input features is large.

To address the aforementioned challenges, we propose to develop novel multi-task learning formulations to model disease progression. The idea of multi-task learning is to utilize the intrinsic relationships among multiple related tasks in order to improve the prediction performance; it is most effective when the number of samples for each task is small. One of the key issues in multi-task learning is to identify how the tasks are related and build learning models to capture such task relatedness. One way of modeling multi-task relationship is to assume that all tasks are related and the task models are closed to each other (Evgeniou et al., 2006), or the tasks are clustered into groups (Bakker and Heskes, 2003; Jacob et al., 2009; Thrun and O'Sullivan, 1998; Zhou et al., 2011). Alternatively, one can assume that the tasks share a common subspace (Ando and Zhang, 2005; Chen et al., 2009), or a common set of features (Argyriou et al., 2008; Obozinski et al., 2006).

In this paper, we propose novel multi-task learning formulations for predicting disease progression measured by the clinical scores (ADAS-Cog and MMSE). Specifically, we formulate the prediction of clinical scores at a sequence of time points as a multi-task regression problem, where each task concerns the prediction of a clinical score at one time point. For the disease progression considered in this paper, it is reasonable to assume that a small subset of features is predictive of the progression, and the multiple regression models from different time points satisfy the smoothness property, that is, the difference of the cognitive scores between two successive time points is small. To this end, we develop a novel multi-task learning formulation based on a temporal group Lasso regularizer (TGL). The regularizer consists of two components including an $\ell_{2,1}$-norm penalty (Yuan and Lin, 2006) on the regression weight vectors, which ensures that a small subset of features will be selected for the regression models at all time points, and a temporal smoothness term, which ensures a small deviation between two regression models at successive time points. In order to better capture the *temporal patterns* of the biomarkers in disease progression (Caroli et al., 2010; Jack et al., 2010), we further propose a convex fused sparse group Lasso (cFSGL) formulation that allows the simultaneous selection of a common set of biomarkers at all time points and the selection of a specific set of biomarkers at different time points using the sparse group Lasso penalty, and in the meantime incorporates the temporal smoothness using the fused Lasso penalty. The proposed formulation is challenging to solve due to the use of non-smooth penalties including the sparse group Lasso and fused penalties. We show that the proximal operator associated with the optimization problem of cFSGL exhibits a certain decomposition property and can be solved efficiently. Therefore cFSGL can be efficiently solved using the accelerated gradient method (Nemirovski, 2005; Nesterov, 2004).

We have performed extensive experiments to demonstrate the effectiveness of the proposed models using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Specifically, we use the baseline MRI features to predict MMSE/ADAS-Cog scores in the next 4 years. A set of 648 subjects including 191 cognitively normal older individuals (NL), 319 patients with mild cognitive impairment (MCI), and 138 Alzheimer's disease patients (AD), are included in our study. Our experimental results show that the proposed multi-task learning formulations outperform single-task learning algorithms including ridge regression and Lasso for predicting future MMSE/ADAS-Cog scores. We also observe that including demographic and ApoE genotyping information as additional covariates further improves the prediction performance. We apply our models on the subgroup that only consists of MCI converters and AD patients and we observe similar improved performance from the proposed models. We have also performed longitudinal stability selection using our proposed formulations to identify and analyze the temporal patterns of biomarkers selected in our models. We observe that the cortical thickness average of left middle temporal, the cortical thickness