Computer Methods and Programs in Biomedicine 162 (2018) 19–45

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Generalized fused group lasso regularized multi-task feature learning for predicting cognitive outcomes in Alzheimers disease

Peng Cao [a,*], Xiaoli Liu [a], Hezi Liu [c], Jinzhu Yang [a], Dazhe Zhao [a], Min Huang [b], Osmar Zaiane [d]

[a] *Computer Science and Engineering, Northeastern University, Shenyang, China*
[b] *College of Information Science and Engineering, Northeastern University, Shenyang, China*
[c] *The Third People's Hospital of Fushun, Fushun, China*
[d] *Computing Science, University of Alberta, Edmonton, Alberta, Canada*

A R T I C L E I N F O

Article history:
Received 28 September 2017
Revised 19 March 2018
Accepted 30 April 2018

Keywords:
Alzheimer's disease
Regression
Multi-task learning
Sparse learning
Group lasso
Fused lasso
Biomarker identification

A B S T R A C T

Objective: Alzheimers disease (AD) is characterized by gradual neurodegeneration and loss of brain function, especially for memory during early stages. Regression analysis has been widely applied to AD research to relate clinical and biomarker data such as predicting cognitive outcomes from Magnetic Resonance Imaging (MRI) measures. Recently, the multi-task feature learning (MTFL) methods have been widely studied to predict cognitive outcomes and select the discriminative feature subset from MRI features by incorporating inherent correlations among multiple clinical cognitive measures. However, the existing MTFL assumes the correlation among all the tasks is uniform, and the task relatedness is modeled by encouraging a common subset of features with neglecting the inherent structure of tasks and MRI features.

Methods: In this paper, we proposed a generalized fused group lasso (GFGL) regularization to model the underlying structures, involving (1) a graph structure within tasks and (2) a group structure among the image features. Then, we present a multi-task learning framework (called GFGL–MTFL), combining the $\ell_{2,1}$-norm with the GFGL regularization, to model the flexible structures.

Results: Through empirical evaluation and comparison with different baseline methods and the state-of-the-art MTL methods on data from Alzheimer's Disease Neuroimaging Initiative (ADNI) database, we illustrate that the proposed GFGL–MTFL method outperforms other methods in terms of both Mean Squared Error (nMSE) and weighted correlation coefficient (wR). Improvements are statistically significant for most scores (tasks).

Conclusions: The experimental results with real and synthetic data demonstrate that incorporating the two prior structures by the generalized fused group lasso norm into the multi task feature learning can improve the prediction performance over several state-of-the-art competing methods, and the estimated correlation of the cognitive functions and the identification of cognition relevant imaging markers are clinically and biologically meaningful.

1. Introduction

Alzheimer's disease (AD) is the most common cause of dementia, which mainly affects memory function, ultimately culminating in a dementia state where all cognitive functions are affected. The disease poses a serious challenge to the aging society. The worldwide prevalence of AD is predicted to quadruple from 46.8 million in 2016 [1] to 131.5 million by 2050 according to ADI's World Alzheimer Report [3]. Dementia also has a huge economic impact. Today, the total estimated worldwide cost of dementia is US $818 billion, and it will achieve a trillion dollar disease by 2018.

Magnetic resonance imaging (MRI) provides a chance to directly observe brain changes such as cerebral atrophy or ventricular expansion [8]. It has been proved that brain atrophy detected by MRI is correlated with neuropsychological deficits [13]. Many cognitive measures have been designed to evaluate the cognitive status of the patients and used as important criteria for clinical diagnosis of probable AD. Modeling the cognitive scores has recently received a significant amount of attention due to its importance for early AD diagnosis [6,40,47,54]. The relationship between commonly used cognitive measures and structural changes with MRI has been previously studied by regression models and the results demonstrated there exist a relationship between baseline MRI features and cognitive measures [13,39,41]. The most commonly used cognitive measures are Alzheimer's Disease Assessment Scale cog-

* Corresponding author.
E-mail address: caopeng@cse.neu.edu.cn (P. Cao).

https://doi.org/10.1016/j.cmpb.2018.04.028
0169-2607/

nitive total score (ADAS), Mini Mental State Exam score (MMSE) and Rey Auditory Verbal Learning Test (RAVLT). ADAS is the gold standard in AD drug trial for cognitive function assessment, which is the most popular cognitive testing instrument to measure the severity of the most important symptoms of AD. MMSE measures cognitive impairment, including orientation to time and place, attention and calculation, immediate and delayed recall of words, language and visuo-constructional functions. RAVLT is a measure of episodic memory and used for the diagnosis of memory disturbances, which consists of eight recall trials and a recognition test.

Early studies focused on traditional regression models to predict cognitive outcomes one at a time. To achieve more appropriate predictive models performance and identify relevant imaging biomarkers, many previous works formulated the prediction of multiple cognitive outcomes as a multi-task learning problem and developed regularized multi-task learning methods to model disease cognitive outcomes [39,41,43,54]. Here, the assumption of both methods is that these clinical cognitive score predictors are inherently related and essentially determined by the same underlying pathology (the diseased brain regions). Multi-task learning [7] is a learning paradigm which seeks to improve the generalization performance of a learning task with the help of some other related tasks. To solve the multi-task learning problems, regularization has been introduced to produce better performance than traditional solution using single-task learning. The most commonly used regularization is $\ell_{2,1}$ norm [26], which is employed to extract features that have impact on all or most clinical scores, since the assumption is that a given imaging marker can affect multiple cognitive scores and only a subset of the brain regions (region-of-interest, ROI) are relevant. The $\ell_{2,1}$ norm regularized MTL is also called multi-task feature learning (MTFL). Wang et al. [42] and Zhang and Yeung [52] employed multi-task feature learning strategies for selecting biomarkers that could predict multiple clinical scores. Specifically, Wang et al. [42] employed an ℓ_1-norm regularizer to impose the sparsity among all elements and propose to use the combined $\ell_{2,1}$-norm and ℓ_1-norm regularizations to select features; Zhang proposed a multi-task learning with $\ell_{2,1}$-norm to select the common subset of relevant features for multiple variables from each modality.

The major limitations in the existing MTFL methods are that the underlying structure within the imaging markers and the complex relationship among the cognitive outcomes are often ignored. When correlating the multiple prediction models, it assumes that all the tasks shared the same feature subset, which is not realistic since it equally treats all the cognitive outcomes (response) and MRI features (predictors) with neglecting the underlying correlation between the cognitive tasks and structure within MRI features. Specifically, (1) for the cognitive outcomes, each assessment typically yields multiple evaluation scores from a set of relevant cognitive tasks, and thus these scores are inherently correlated, such as the score of TOTAL and TOT6 in the RAVLT. Different assessments evaluate the subjects' different cognitive functions and prefer the different brain regions, resulting in a low correlation between them. For example, the measures in TRAILS aimed to test a combination of visual, motor and executive functions, while the measures in RAVLT aimed to test verbal learning memory. It is reasonable to assume that the correlation among the tasks are not uniform, some tasks may be more closely related than others in the assessment tests of cognitive outcomes. (2) On the other hand, many MRI features interrelate with each other and work together to reveal the brain cognitive functions [48]. In our data, multiple shape measures (volume, area and thickness) from the same region provide a comprehensively quantitative evaluation of cortical atrophy, and tend to be selected together as joint predictors. Our previous study proposed a prior knowledge guided multi-task feature learning model, using the group information to enforce the intra-group similarity, has been demonstrated the effectiveness of the group guided learning processing [27,28]. Therefore, it is desired to explore and utilize such interrelation structure and select these important and structurally correlated features together.

To solve these limitations of traditional MTFL, we introduce a new regularization term, named generalized fused group lasso (GFGL), to flexibly model the relationship among the tasks (responses) and features (predictors) based on the prior knowledge. The structured sparsity norm is based on the natural assumption that if some tasks are correlated, they should have similar weight vectors and similar selected brain regions. To discover such the underlying correlation structure among the cognitive outcomes, we employed the Pearson correlation coefficient to uncover the interrelations among cognitive measures and estimated the correlation matrix of all the tasks. In our work, all the cognitive measures in the ADNI dataset are used to exploit the relationship. To the best of our knowledge, our approach is the first work on analysis and exploitation of all the cognitive measures and their complex correlation in the ADNI dataset. With the estimated task correlation, we extend the 1D generalized fused lasso to the 2D generalized fused lasso (GFL) to capture the dependence of task response variables. On the other hand, the group structure among MRI features is taken into account and incorporated into the generalized fused lasso, leading to enforce the intra-group similarity with group sparse. By incorporating the proposed GFGL regularization into the MTFL model (called GFGL–MTFL), we can better understand the underlying associations of prediction tasks of cognitive measures and obtain more stable identification of cognition-relevant imaging markers. An overview of the proposed multi-task learning framework is illustrated in Fig. 1. The proposed formulation is challenging to solve due to the use of non-smooth penalties including the generalized fused group lasso and the $\ell_{2,1}$ norm. An effective ADMM algorithm is proposed to tackle the complicated non-smoothness. Through empirical evaluation and comparison with different baseline methods and the state-of-the-art MTL methods on data from ADNI, we illustrate that the proposed method outperforms other methods. We evaluated the effectiveness of our proposed general framework on inferring the cognitive outcomes on single time-point of data (cross-sectional analysis), tracking disease progression (longitudinal analysis), multi-modality data fusing and multi-label learning. Improvements are statistically significant for most scores. The results demonstrate that the plugging the GFGL regularization into the traditional MTFL formulation has shown improvements in predictive performance relative to traditional machine learning methods. We also present a discussion on the top ROIs and the task correlations identified by GFGL–MTFL.

The rest of the paper is organized as follows. In Section 2, we provide a description of the preliminary used in our work: multi-task learning (MTL), $\ell_{2,1}$ norm, group lasso norm and 1D generalized fused lasso norm. A detailed mathematical formulation and optimization of GFGL–MTFL is provided in Section 3. In Section 4, we present the experimental results and evaluate the performance of GFGL–MTFL on the ADNI-1 dataset and the multi-labeled dataset. The conclusion is drawn in Section 5.

2. Preliminary and dataset

2.1. Multi-task learning

Consider a multi-task learning (MTL) setting with k tasks. Let p be the number of covariates, shared across all the tasks, n be the number of samples. Let $X \in \mathbb{R}^{n\times p}$ denote the matrix of covariates, $Y \in \mathbb{R}^{n\times k}$ be the matrix of responses with each row corresponding to a sample, and $\Theta \in \mathbb{R}^{p\times k}$ denote the parameter matrix, with column $\theta_{.m} \in \mathbb{R}^p$ corresponding to task m, $m = 1,\dots,k$, and row $\theta_{j.} \in \mathbb{R}^k$ corresponding to feature j, $j = 1,\dots,p$. The MTL problem