Prediction of Memory Impairment with MRI Data: A Longitudinal Study of Alzheimer's Disease

Xiaoqian Wang[1], **Dinggang Shen**[2], and **Heng Huang**[1]

[1] Computer Science and Engineering, University of Texas at Arlington, Arlington, USA

[2] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, USA

Abstract

Alzheimer's Disease (AD), a severe type of neurodegenerative disorder with progressive impairment of learning and memory, has threatened the health of millions of people. How to recognize AD at early stage is crucial. Multiple models have been presented to predict cognitive impairments by means of neuroimaging data. However, traditional models did not employ the valuable longitudinal information along the progression of the disease. In this paper, we proposed a novel longitudinal feature learning model to simultaneously uncover the interrelations among different cognitive measures at different time points and utilize such interrelated structures to enhance the learning of associations between imaging features and prediction tasks. Moreover, we adopted Schatten p-norm to identify the interrelation structures existing in the low-rank subspace. Empirical results on the ADNI cohort demonstrated promising performance of our model.

1 Introduction

Alzheimer's Disease (AD), the most common form of dementia, is a neurodegenerative disorder which severely impacts patients' thinking, memory and behavior. Current consensus has emphasized the demand of early recognition of this disease, with which the goal of stoping or slowing down the disease progression can be achieved [8]. The effectiveness of neuroimaging in predicting the progression of AD or cognitive performance has been studied and reported in plentiful research [4,12]. However, many previous research merely paid attention to the prediction using the baseline data, which neglected correlation among longitudinal cognitive performance. AD is a progressive neurodegenerative disorder, thus it is significant to discover neuroimaging measures that impact the progression of this disease along the time axis.

In the association study of predicting cognitive scores from imaging features, the input data usually consists of two matrices: the imaging feature matrix $\mathscr{X}$ and the cognitive score matrix $\mathscr{Y}$. If we denote the number or samples as n, the number of features as d while the number of different measures of a certain cognitive performance test as m, then $\mathscr{X}$ and $\mathscr{Y}$ can be formed in the following format: $\mathscr{X} = [X_1, \cdots, X_T] \in {}^{d \times nT}$ corresponds to the

heng@uta.edu.

imaging features at T consecutive time points where $X_t \in {}^{d \times n}$ is the imaging marker matrix at the t-th time point; $\mathscr{Y} = [Y_1, \cdots, Y_T] \in {}^{n \times mT}$ corresponds to the cognitive scores at T consecutive time points with $Y_t \in {}^{n \times m}$ denoting the measurement at the t-th time point.

Let's consider the prediction of one cognitive measure at one time point to be one task, then the association study between cognitive scores and imaging features can be regarded as a multi-task problem. Apparently, in our setting of the longitudinal association study, the number of tasks is mT. The goal of the association study is to find a weight matrix $\mathscr{W} = [W_1, \cdots, W_T] \in {}^{d \times mT}$, which captures the relevant features for predicting the cognitive scores.

A forthright method is to perform linear regression at each time point and determine W_t separately. However, the linear regression treats all tasks independently and ignores the useful information reserved in the change along the time continuum. Since AD is a progressive neurodegenerative disorder and cognitive performance is an intuitive indication of the disease status, we can reasonably regard the various tasks to be possibly related. In one cognitive experiment, the result of a certain measure at different time points may be correlated and also different cognitive measures at a certain time point may have mutual influence. To excavate the correlations among the cognitive scores, several multi-task models are put forward.

One possible method is the longitudinal $\ell_{2,1}$-norm regression model [6,11]. In this model, the introduced $\ell_{2,1}$-norm regularization enforces structured sparsity, which helps to detect features related to all the cognitive measures along the whole time axis. Moreover, with the assumption that imaging features may be correlated with each other thus gain an overlap in their effects on brain structure or disease progression, we can use the trace norm (also known as nuclear norm) regularization to impose a low-rank restriction. Also, there are models combining these two regularization terms to enforce the structured sparsity as well as low-rank constraint [13,14].

Indeed, these models impose trace norm regularization to the whole parameter matrix, such that the common subspace globally shared by different prediction tasks can be extracted. However, the longitudinal prediction tasks can be interrelated as different groups. The straightforward way to discover such interrelated groups is to conduct the clustering analysis first and extract the group structures. However, such a heuristic step is independent to the entire longitudinal learning model, thus the detected group structures are not optimal for the longitudinal learning process.

To address this challenging problem, we propose a novel longitudinal structured low-rank learning model to uncover the interrelations among different cognitive measures and utilize the learned interrelated structures to enhance cognitive function prediction tasks.

2 Longitudinal Structured Low-Rank Regression Model

In our multi-task problem, suppose these mT tasks come from c groups, where tasks in each group are correlated. We can introduce and optimize a group index matrix set $Q = \{Q_1,$

$Q_2, \ldots Q_c\}$ to discover this group structure. Each Q_i is a diagonal matrix with $Q_i \in \{0,1\}^{mT \times mT}$ showing the assignment of tasks to the i-th group. For the (k, k)-th element of Q_i, $(Q_i)_{kk} = 1$ means that the k-th task belongs to the i-th group while $(Q_i)_{kk} = 0$ means not.

To avoid overlap of groups, we constrain $\sum_{i=1}^{c} Q_i = I$.

Since each group of tasks share correlative dependence, we can reasonably assume the latent subspace of each group maintains a low-rank structure. We impose Schatten-p norm as a low-rank constraint to uncover the common subspace shared by different tasks. According to the discussion below, Schatten p-norm makes a better approximation of the low-rank constraint than the popular trace norm regularization [7].

For a matrix $A \in \mathbb{R}^{d \times n}$, suppose σ_i is its i-th singular value, then the rank of A can be written as $rank(A) = \sum_{i=1}^{min\{d,n\}} \sigma_i^0$, where $0^0 = 0$. And the definition of p-th power Schatten p-norm $(0 < p < \infty)$ of A is: $\|A\|_{S_p}^p = Tr\left(\left(A^T A\right)^{\frac{p}{2}}\right) = \sum_{i=1}^{min\{d,n\}} \sigma_i^p$. Specially, when $p = 1$, we find the Schatten p-norm of A is exactly its trace norm:

$$\|A\|_{S_1} = \left(Tr\left(\left(A^T A\right)^{\frac{1}{2}}\right)\right) = \sum_{i=1}^{min\{d,n\}} \sigma_i = \|A\|_*.$$

So when $0 < p < 1$, Schatten p-norm is a better low-rank regularization than trace norm. Accordingly, our longitudinal structured low-rank regression model is:

$$\min_{\mathscr{W}, Q_i|_{i=1}^c \in \{0,1\}^{mT \times mT}, \sum_{i=1}^{c} Q_i = I} \sum_{t=1}^{T} \|W_t^T X_t - Y_t\|_F^2 + \gamma \sum_{i=1}^{c} \left(\|\mathscr{W} Q_i\|_{S_p}^p\right)^l. \tag{1}$$

In Problem (1), the grouping structure tends to be unstable when p is small, so we add a power parameter l to the regularization term and make our model robust. It is diffcult to solve this new non-convex and non-smooth objective function. In next section, we will propose a novel alternating optimization method for Problem (1).

3 Optimization Algorithm for Solving Problem (1)

According to the property of Q_i that $Q_i^2 = Q_i$, Problem (1) can be rewritten as:

$$\min_{\mathscr{W}, Q_i|_{i=1}^c \in \{0,1\}^{mT \times mT}, \sum_{i=1}^{c} Q_i = I} \sum_{t=1}^{T} \|W_t^T X_t - Y_t\|_F^2 + \gamma \sum_{i=1}^{c} Tr\left(\mathscr{W}^T D_i \mathscr{W} Q_i\right), \tag{2}$$

where D_i is defined as:

$$D_i = \frac{lp}{2}\left(\|\mathscr{W} Q_i\|_{S_p}^p\right)^{l-1}\left(\mathscr{W} Q_i \mathscr{W}^T\right)^{\frac{p-2}{2}}. \tag{3}$$

We can solve Problem (2) via alternating optimization method.

The first step is fixing $\mathcal{W}$ and solving Q, and then Problem (2) becomes:

$$\min_{Q_i|_{i=1}^c \in \{0,1\}^{mT \times mT}, \sum_{i=1}^c Q_i = I} \sum_{i=1}^c Tr\left(\left(\mathcal{W}^T D_i \mathcal{W}\right) Q_i\right). \tag{4}$$

Letting $A_i = \mathcal{W}^T D_i \mathcal{W}$, then the solution of Q_i is:

$$(Q_i)_{kk} = \begin{cases} 1 & i = arg\min_j (A_j)_{kk} \\ 0 & otherwise \end{cases} \tag{5}$$

The second step is fixing Q and solving $\mathcal{W}$, and then Problem (2) becomes:

$$\min_{\mathcal{W}} \sum_{t=1}^T \|W_t^T X_t - Y_t\|_F^2 + \gamma \sum_{i=1}^c Tr\left(\mathcal{W}^T D_i \mathcal{W} Q_i\right). \tag{6}$$

Denote Q_i in the format that $Q_i = diag(Q_{i1}, Q_{i2}, \ldots, Q_{iT})$. Since $Tr\left(\mathcal{W}^T D_i \mathcal{W} Q_i\right) = \sum_{t=1}^T Tr\left(W_t^T D_i W_t Q_{it}\right)$, we can decouple Problem (6) for each t:

$$\min_{W_t} \|W_t^T X_t - Y_t\|_F^2 + \gamma \sum_{i=1}^c Tr\left(W_t^T D_i W_t Q_{it}\right). \tag{7}$$

Problem (7) can be further decoupled for each column of W_t as follows:

$$\min_{(\mathbf{w_t})_k} \|\left(\mathbf{w_t}^T\right)_k X_t - (\mathbf{y_t})^k\|_2^2 + \gamma\, Tr\left(\left(\mathbf{w_t}^T\right)_k \left(\sum_{i=1}^c (Q_{it})_{kk} D_i\right) (\mathbf{w_t})_k\right). \tag{8}$$

Taking derivative w.r.t. $(\mathbf{w_t})_k$ in Problem (8) and setting it to zero, then we get:

$$(\mathbf{w_t})_k = \left(X_t X_t^T + \gamma \left(\sum_{i=1}^c (Q_{it})_{kk} D_i\right)\right)^{-1} X_t \left((\mathbf{y_t})^k\right)^T. \tag{9}$$

We can iteratively update Q, $\mathcal{W}$ and D with the alternating steps mentioned above and the algorithm of Problem (2) is summarized in Algorithm 1.

Convergence Analysis

Our algorithm uses alternating optimization method, whose convergence has already been proved in [1]. In Algorithm 1, variables in each iteration has a closed form solution and can be computed fairly fast. In the following experiments on the ADNI data, the running time of each iteration is about 0.005 s and our method usually converges within one second.

4 Experimental Results

In this section, we evaluate the prediction performance of our proposed method by applying it to the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

4.1 Data Description

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). Each MRI T1-weighted image was first anterior commissure (AC) posterior commissure (PC) corrected using MIPAV2, intensity inhomogeneity corrected using the N3 algorithm [10], skull stripped [16] with manual editing, and cerebellum-removed [15]. We then used FAST [17] in the FSL package3 to segment the image into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), and used HAMMER [9] to register the images to a common space. GM volumes obtained from 93 ROIs defined in [5], normalized by the total intracranial volume, were extracted as features. Longitudinal scores were downloaded from three independent cognitive assessments including Fluency Test, Rey's Auditory Verbal Learning Test (RAVLT) and Trail making test (TRAILS). The details of these cognitive assessments can be found in the ADNI procedure manuals. The time points examined in this study for both imaging markers and cognitive assessments included baseline (BL), Month 6 (M6), Month 12 (M12) and Month 24 (M24). All the participants with no missing BL/M6/M12/M24 MRI measurements and cognitive measures were included in this study. A total of 385 sample subjects are involved in our study, among which we have 56 AD samples, and 181 MCI samples and 148 health control (HC) samples. Seven cognitive scores were included: (1) RAVLT TOTAL, RAVLT TOT6 and RAVLT RECOG scores from RAVLT cognitive assessment; (2) FLU ANIM and FLU VEG scores from Fluency cognitive assessment; (3) Trails A and Trails B scores from Trail making test.

4.2 Performance Comparison on the ADNI Cohort

We first evaluate the ability of our method to predict a certain set of cognitive scores via neuroimaging marker. We tracked the process along the time axis and intended to find the set of markers which could influence the cognitive score over the time points. As the evaluation metric, we reported the Root Mean Square Error (RMSE) as well as the Correlation Coefficient (CorCoe) between the predicted score and the ground truth.

We compared our method with all the counterparts discussed in the introduction, which are: Multivariate Linear Regression (MLR), Multivariate Ridge Regression (MRR), Longitudinal Trace-norm Regression (LTR), Longitudinal $\ell_{2,1}$ norm Regression (L21R) and their combination (L21R + LTR). To illustrate the advantage of simultaneously conducting task correlation and longitudinal feature learning, we also compared with the method of using *K*-

means to cluster the tasks first and then implementing LTR in each group (*K*-means + LTR) as the baseline.

We utilized the 10-fold cross validation technique and ran 50 times for each method. The average RMSE and CorCoe on these 500 trials are reported. For MLR and MRR, since they were not designed for the longitudinal tasks, we computed the weight matrix for each time point separately and then merged them to the final weight matrix according to the definition $\mathscr{W} = [W_1, \cdots, W_T]$. Here in this experiment, the number of time points *T* is 4. Our initial analyses indicated that our model performs fairly stable when choosing parameter *l* from {2, 2.5, . . . , 5} and choosing parameter *p* from {0.1, 0.2, . . . , 0.8} (data not shown). In our experiments, we fixed $p = 0.1$ and $l = 3$.

The experimental results are summarized in Table 1. From all the results, we can notice that our method outperforms all other methods consistently on all data sets. The reasons go as follows: MLR and MRR assumed the cognitive measures at different time points to be independent, thus didn't consider the correlations along the time. Their neglects of the longitudinal correlation within the data was detrimental to their prediction ability. As for L21R, LTR and their combination LTR + L21R, even though they take into account the longitudinal information, they cannot handle the possible group structure within the cognitive scores. That is why they overweigh the standard methods like MLR and MRR in most cases, but are inferior to our proposed method. For *K*-means + LTR, the clustering step is detached from the longitudinal association study, thus the learned interrelation structure is not optimal for the following longitudinal learning process. As for our proposed method, we not only captured longitudinal correlations among imaging features, but also detected group structure within cognitive scores. As was discussed in the theoretical sections, our model is able to find features which impact on the cognitive result at different stages and meanwhile cluster the cognitive results into groups. Thus, our model can capture features responsible for some, but not necessarily all, cognitive measures along the time continuum, which saves more effective information in the prediction.

4.3 Identification of Longitudinal Imaging Markers

We further take a special case, the RAVLT assessment, as an example to analyze results of our model. RAVLT is composed of three cognitive measures, which are: (1) the total number of words kept in mind by the testee in the first five trials, RAVLT TOTAL; (2) the number of words recalled during the 6th trial, RAVLT TOT6; and (3) the number of words recognized after a gap of 30 min, RAVLT RECOG. According to the common sense, these three measures should be interrelated with each other, thus clustered into the same group in our model. The result of our model shows a consistent obedience of this rule, *i.e.*, no matter what the *c* value (number of groups) is, our model invariably put all these three measures to the same group, which is in line with reality. Specially, when c is larger than the real number of groups, the extra groups become empty.

Figure 1 shows the heat maps of the weight matrices learned by our method. The figures demonstrate the capture of a small set of features that are consistently associated to a certain group of cognitive measures (here the group includes all measures). Among the selected features, we found the top two are the hippocampal formation and thalamus, whose impacts

on AD have already been proved in the previous papers [2,3]. In summary, our model is competent to select a small set of features that consistently correlate with a certain group of cognitive measures along the time axis. And the effectiveness of the selected features can be confirmed by previous reports in the literature.

5 Conclusion

In this paper, we proposed a novel longitudinal structured low-rank regression model to study the longitudinal cognitive score prediction. Our model can simultaneously uncover the interrelation structures existing in different prediction tasks and utilize such learned interrelated structures to enhance the longitudinal learning model. Moreover, we utilized Schatten p-norm to extract the common subspace shared by the prediction tasks. Our new model is applied to ADNI cohort for cognitive impairment prediction using MRI data. Empirical results validate the effectiveness of our model, showing a potential to provide reference for current clinical research.

Acknowledgments

References

1. Bezdek JC, Hathaway RJ. Convergence of alternating optimization. Neural Parallel Sci. Comput. 2003; 11(4):351–368.

2. De Jong L, Van der Hiele K, Veer I, Houwing J, Westendorp R, Bollen E, De Bruin P, Middelkoop H, Van Buchem M, Van Der Grond J. Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. Brain. 2008; 131(12):3277–3285. [PubMed: 19022861]

3. De Leon M, George A, Golomb J, Tarshish C, Convit A, Kluger A, De Santi S, Mc Rae T, Ferris S, Reisberg B, et al. Frequency of hippocampal formation atrophy in normal aging and Alzheimer's disease. Neurobiol. Aging. 1997; 18(1):1–11. [PubMed: 8983027]

4. Ewers M, Sperling RA, Klunk WE, Weiner MW, Hampel H. Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia. Trends Neurosci. 2011; 34(8):430–442. [PubMed: 21696834]

5. Kabani NJ. 3D anatomical atlas of the human brain. Neuroimage. 1998; 7:P–0717.

6. Nie F, Huang H, Cai X, Ding CH. Efficient and robust feature selection via joint $l_{2,1}$-norms minimization. Advances in Neural Information Processing Systems. 2010:1813–1821.

7. Nie, F., Huang, H., Ding, CH. Low-rank matrix recovery via efficient schatten p-norm minimization. AAAI; 2012.

8. Petrella JR, Coleman RE, Doraiswamy PM. Neuroimaging and early diagnosis of Alzheimer disease: a look to the future 1. Radiology. 2003; 226(2):315–336. [PubMed: 12563122]

9. Shen D, Davatzikos C. Hammer: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imaging. 2002; 21(11):1421–1439. [PubMed: 12575879]

10. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging. 1998; 17(1):87–97. [PubMed: 9617910]

11. Wang H, Nie F, Huang H, Kim S, Nho K, Risacher SL, Saykin AJ, Shen L. Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. Bioinformatics. 2012; 28(2):229–237. [PubMed: 22155867]

12. Wang, H., Nie, F., Huang, H., Risacher, S., Saykin, AJ., Shen, L. Identifying AD-sensitive and cognition-relevant imaging biomarkers via joint classification and regression.. In: Fichtinger,

G.Martel, A., Peters, T., editors. MICCAI 2011, Part III. LNCS. Vol. 6893. Springer; Heidelberg: 2011. p. 115-123.

13. Wang H, Nie F, Huang H, Yan J, Kim S, Nho K, Risacher SL, Saykin AJ, Shen L, et al. From phenotype to genotype: an association study of longitudinal phenotypic markers to Alzheimer's disease relevant SNPs. Bioinformatics. 2012; 28(18):i619–i625. [PubMed: 22962490]

14. Wang H, Nie F, Huang H, Yan J, Kim S, Risacher S, Saykin A, Shen L. High-order multi-task feature learning to identify longitudinal phenotypic markers for Alzheimer's disease progression prediction. Advances in Neural Information Processing Systems. 2012:1277–1285.

15. Wang Y, Nie J, Yap PT, Li G, Shi F, Geng X, Guo L, Shen D. Alzheimer's Disease Neuroimaging Initiative: Knowledge-guided robust MRI brain extraction for diverse large-scale neuroimaging studies on humans and non-human primates. PloS One. 2014; 9(1):e77810. [PubMed: 24489639]

16. 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., editors. MICCAI 2011. LNCS. Vol. 6893. Springer; Heidelberg: 2011. p. 635-642.doi:10.1007/978-3-642-23626-6 78

17. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging. 2001; 20(1): 45–57. [PubMed: 11293691]
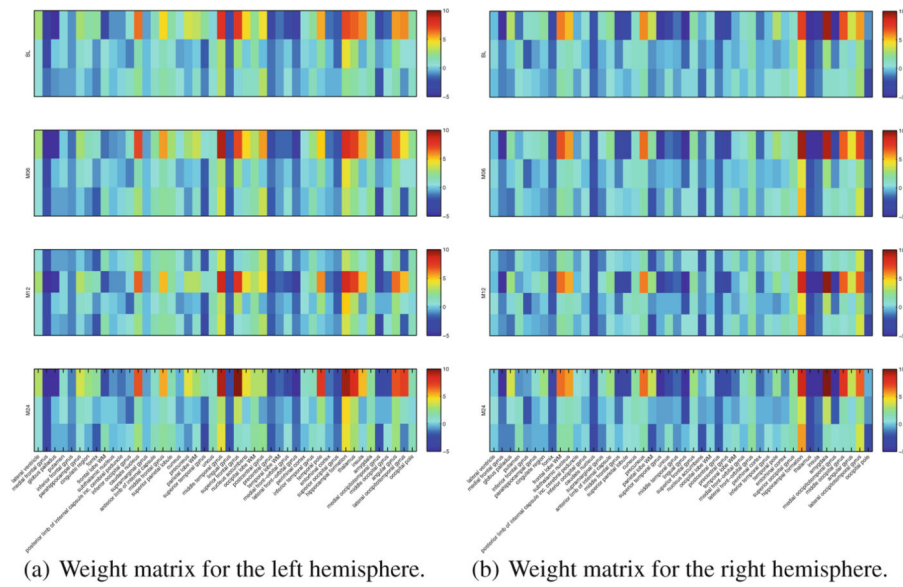
(a) Weight matrix for the left hemisphere.   (b) Weight matrix for the right hemisphere.

Fig. 1.
Heat maps of our learned weight matrices on the RAVLT cognitive assessment via MRI data. The weight matrices at four time points, BL, M6, M12 and M24, are plotted. We draw two matrices for each time point, where the left figure is for the left hemisphere and the right figure for the right hemisphere. For each weight matrix, columns denote neuroimaging features while rows represent three different RAVLT scores, which are RAVLT TOTAL, RAVLT TOT6 and RAVLT RECOG, respectively. Imaging features (columns) with larger weights possess higher correlation with the corresponding cognitive measure.

Table 1

Cognitive assessment FLUENCY, RAVLT and TRAILS prediction comparison via RMSE and CorCoe. Better performance corresponds to lower RMSE or higher CorCoe value.

		MLR	MRR	LTR	L21R	L21R + LTR	*K*-means + LTR	OURS
RMSE	FLUENCY	0.352	0.350	0.343	0.339	0.345	0.351	**0.316**
	RAVLT	0.469	0.447	0.458	0.445	0.448	0.459	**0.417**
	TRAILS	0.571	0.554	0.564	0.551	0.567	0.557	**0.511**
CorCoe	FLUENCY	0.504	0.499	0.516	0.528	0.513	0.503	**0.579**
	RAVLT	0.872	0.880	0.877	0.879	0.879	0.874	**0.891**
	TRAILS	0.541	0.551	0.548	0.558	0.547	0.562	**0.600**

Algorithm 1

Algorithm to solve problem (2).

Input:

Longitudinal imaging feature matrix $\mathcal{X} = \left[X_1, X_2, \cdots, X_T\right] \in \mathbb{R}^{d \times nT}$, longitudinal cognitive score matrix

$\mathcal{Y} = \left[Y_1, Y_2, \cdots, Y_T\right] \in \mathbb{R}^{n \times mT}$, parameter γ, and number of groups c.

Output:

Weight matrix $\mathcal{W} = \left[W_1, W_2, \cdots, W_T\right]$ where $W_t \in \mathbb{R}^{d \times m}$ and c different group matrix $Q_i \in \mathbb{R}^{mT \times mT}$ which groups the tasks into exactly c groups.

Initialize $\mathcal{W}$ by the optimal solution to ridge regression problem

while not converge **do**

    1. Update $D_i \mid_{i=1}^{c}$ according to the definition in Eq. (3).

    2. Update $Q_i \mid_{i=1}^{c}$ according to the solution in Eq. (5)

    3. Update $\mathcal{W}$, where the solution to the k-th column of W_t is displayed in Eq. (9).

end while