2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

PREDICTING PROGRESSIONS OF COGNITIVE OUTCOMES VIA HIGH-ORDER MULTI-MODAL MULTI-TASK FEATURE LEARNING

Lyujian Lu[1], *Hua Wang*[1†], *Xiaohui Yao*[2], *Shannon Risacher*[2], *Andrew Saykin*[2], *Li Shen*[2,3], *for the ADNI*[‡]

[1]Department of Computer Science, Colorado School of Mines, Golden, CO 80401
[2]Radiology and Imaging Sciences, BioHealth Informatics, Indiana University, IN 46202
[3]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, PA 19104

ABSTRACT

Many existing studies on complex brain disorders, such as Alzheimer's Disease, usually employed regression analysis to associate the neuroimaging measures to cognitive status. However, whether these measures in multiple modalities have the predictive power to infer the trajectory of cognitive performance over time still remain under-explored. In this paper, we propose a high-order multi-modal multi-mask feature learning model to uncover temporal relationship between the longitudinal neuroimaging measures and progressive cognitive output scores. The regularizations through sparsity-induced norms implemented in the proposed learning model enable the selection of only a small number of imaging features over time and capture modality structures for multi-modal imaging markers. The promising experimental results in extensive empirical studies performed on the ADNI cohort have validated the effectiveness of the proposed method.

Index Terms— Alzheimer's Disease, Feature Learning, Multi-Modal, Longitudinal Regression

1. INTRODUCTION

Neuroimaging measures have been widely studied to predict cognitive outcomes [1, 2, 3, 4, 5]. However, whether these measures have further predictive power to infer a trajectory of cognitive performance over time still remains an under-explored, yet important, topic in Alzheimer's Disease (AD) research. A simple strategy typically used in longitudinal studies (*e.g.*, [6]) is to analyze a single summarized value such as average change, rate of change, or slope. This approach may be inadequate to distinguish the complete dynamics of cognitive trajectories and thus become unable to identify the underlying neurodegenerative mechanisms [7, 8]. However, directly associating the temporal imaging features to longitudinal cognitive outcome scores is very challenging for the following reasons. Different to traditional regression models, the input data and the response cognitive scores are high-order tensors. Both the input neuroimaging measures ($samples \times features \times time$) and output cognitive scores ($samples \times scores \times time$) are three-dimensional tensors, which inevitably complicates the learning problems.

Our previous work [7] has made an attempt to tackle the above challenges, which, however, performed studies on only one single modality of neuroimaging measures. Recent studies [4] have explored multi-modal neuroimaging measures, such as the biomarkers extracted from magnetic resonance imaging (MRI), positron emission tomography (PET), or cerebrospinal fluid (CS) data, and reported improved performance in association studies to predict cognitive outcomes. Thus, it is of apparent necessity to perform longitudinal association studies using multiple modalities of neuroimaging measures. With this recognition, in this paper we propose a novel high-order multi-modal multi-task feature learning model to identify longitudinal neuroimaging markers that can accurately predict cognitive scores over all the time pointsp.

2. THE METHOD

For AD progression prediction using longitudinal phenotypic biomarkers, the input imaging features are a set of matrices: $\mathcal{X} = \{X_1, X_2, \cdots, X_T\} \in \mathbb{R}^{d\times n\times T}$ corresponding to the measurements at T consecutive time points. $X_t \in \mathbb{R}^{d\times n}$ $(1 \leq t \leq T)$ is the concatenation of different types of phenotype measurements of the entire cohort at certain time point t, such as voxel-based morphometry (VBM), modified voxel-based morphometry (mVBM) and FreeSurfer (FS): $X_t = [X_{t1}; X_{t2}; \cdots; X_{tK}]$ where K represents the number of neuroimaging feature modalities. Obviously, the input neuroimaging data $\mathcal{X}$ is a tensor with d imaging measures, n subject samples, and T time points. Similarly, the cogni-

†Correspondence to Hua Wang (huawangcs@gmail.com). This research was partially supported by NSF-IIS 1423591 and NSF-IIS 1652943; NIH R01 EB022574, R01 LM011360, R01 AG19771, U19 AG024904, and P30 AG10133.

‡Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.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.usc.edu/wp-content/uploads/how to apply/ADNI Acknowledgement List.pdf.

978-1-5386-3636-7/18/$31.00 ©2018 IEEE

tive outcomes are a set of matrices: $\mathcal{Y} = \{Y_1, Y_2, \cdots, Y_T\} \in \mathbb{R}^{n\times c\times T}$ for certain type of cognitive outcomes, such as the Rey's Auditory Verbal Learning Test (RAVLT) scores at the same T consecutive time points as input. Again, $\mathcal{Y}$ is a tensor with n subject samples, c cognitive scores, and T time points.

Traditional multivariate regression models minimize the following objective:

$$\begin{aligned}\min_{\mathcal{W}} \quad J_0 &= ||\mathcal{W}\otimes\mathcal{X}-\mathcal{Y}||+\gamma_1||\mathcal{W}||_2^2 \\ &= \sum_{t=1}^{T}||W_t^T X_t - Y_t||_F^2+\gamma_1\sum_{t=1}^{T}\sum_{j=i}^{d}||\mathbf{w}_t^j||_2^2,\end{aligned} \tag{1}$$

where w_t^j indicates the j-th row of coefficient matrix W_t at time point t and $\mathcal{W} = \{W_1, \ldots, W_T\} \in \mathbb{R}^{d\times c\times T}$ is a tensor of regression weights. The objective J_0 in Eq. (1) does not take into account the temporal correlations over time, because the optimization problem can be decoupled for each individual time point separately. To uncover the temporal relationship, we introduce the sparse regularization as follows [7, 8]:

$$\begin{aligned}\min_{\mathcal{W}} \quad J_1 &= ||\mathcal{W}\otimes\mathcal{X}-\mathcal{Y}||+\gamma_1\sum_{i=1}^{d}\sqrt{\sum_{t=1}^{T}||\mathbf{w}_t^i||_2^2} \\ &= \sum_{t=1}^{T}||W_t^T X_t - Y_t||_F^2+\gamma_1||W^{(1)}||_{2,1},\end{aligned} \tag{2}$$

where $W^{(1)} = \text{unfold}_1(\mathcal{W}) = [W_1, \ldots, W_T]$ denotes the unfolding operation of the tensor $\mathcal{W}$ along its first mode.

In addition, to integrate the neuroimaging measures in different modalities, we further impose into our objective the regularization using a group ℓ_1-norm introduced in [5, 9, 10] as $||W^{(1)}||_{G_1} = \sum_{g=1}^{G}\sum_{i=1}^{g_i}||w_g^i||_2 = \sum_{g=1}^{G}||\mathbf{w}_g||_2$, where G represents the number of modalities, and g_i represents the number of imaging measures in g-th modality:

$$\begin{aligned}\min_{\mathcal{W}} \; J_2 &= \sum_{t=1}^{T}||W_t^T X_t - Y_t||_F^2 \\ &+ \gamma_1||W^{(1)}||_{2,1}+\gamma_2||W^{(1)}||_{G_1}\end{aligned} \tag{3}$$

Finally, to capture the correlations among different learning tasks at different time points [7, 8], we impose low-rank regularization to discover the the common subspace shared by predictive tasks as follows:

$$\begin{aligned}\min_{\mathcal{W}} \quad J_3 &= \sum_{t=1}^{T}||W_t^T X_t - Y_t||_F^2+\gamma_1||W^{(1)}||_{2,1} \\ &+ \gamma_2||W^{(1)}||_{G_1}+\gamma_3(||W^{(1)}||_*+||W^{(2)}||_*)\end{aligned} \tag{4}$$

where $W^{(2)} = \text{unfold}_2(\mathcal{W}) = \left[W_1^T, \ldots, W_T^T\right]$ denotes the unfolding operation of the tensor $\mathcal{W}$ along its second mode and $||\cdot||_*$ denotes the trace norm of a matrix. Given a matrix $M \in \mathbb{R}^{n\times m}$ and its singular values σ_i $(1 \le i \le \min(n, m))$, the trace norm of M is defined as $||M||_* = \sum_{i=1}^{\min(n,m)}\sigma_i = Tr(MM^T)^{\frac{1}{2}}$.

The objective in Eq. 4 is our proposed high-order multi-modal multi-task longitudinal feature learning model, which is advantageous in that it is able to integrate multi-modal neuroimaging measures and capture the correlations between cognitive outcomes over time. Its solution algorithm is provided in Algorithm. 1. Due to space limit, the derivation of this algorithm and the proof of its convergence will be supplied in the extended journal version of this paper.

Algorithm 1: The solution algorithm to the proposed objective J_3 in Eq.(4).

Data: $\mathcal{X} = [X_1, X_2, \cdots, X_T] \in \mathbb{R}^{d\times n\times T}$,
$\mathcal{Y} = [Y_1, Y_2, \cdots, Y_T] \in \mathbb{R}^{n\times c\times T}$.
Result: $\mathcal{W} = [W_1, W_2, \cdots, W_T] \in \mathbb{R}^{d\times c\times T}$
initialization: $W_t \in \mathcal{R}^{d\times c}(1 \le t \le T)$ using the linear regression results at each individual time point.
while *not converge* **do**
1. Calculate the diagonal matrix D_1, where the i-th diagonal element is $D_1(i,i) = \frac{1}{2\sqrt{\sum_{t=1}^{T}||w_t^j||_2^2}}$;
2. Calculate the diagonal matrix D_2, where the g-th diagonal block is $\frac{1}{2||w_g||_2}I_g$, where I_g is an identity matrix with size of g_i;
3. Calculate $\hat{D_3} = \frac{1}{2}(W^{(1)}W^{(1)^T})^{-\frac{1}{2}}$ and $\tilde{D_3} = \frac{1}{2}(W^{(2)}W^{(2)^T})^{-\frac{1}{2}}$;
3.Update $W_t(1 \le t \le T)$ by solving the Styler equation of
$(X_tX_t^T+\gamma_1 D_1+\gamma_2 D_2+\gamma_3\hat{D_3})W_t+\gamma_3 W_t\tilde{D_3} = X_tY_t$.
end

3. EXPERIMENTS AND RESULTS

We evaluate the proposed method by applying it to the Alzheimers Disease Neuroimaging Initiative (ADNI) cohort. Our goal is to test whether MRI, PET, or other biological markers can be combined to predict the progression of mild cognitive impairment (MCI) and early AD over a certain period. We downloaded the 1.5T MRI scans and the demographic information for the 821 ADNI-1 participants. We performed VBM on the MRI data by following [6] and extracted mean modulated gray matter measures for 90 target regions of interest. These measures were adjusted for the baseline intracranial volume using regression weights derived from the healthy control (HC) participants at the baseline. We also downloaded the longitudinal scores of the participants in two independent cognitive assessments including Fluency test (FLU) and RAVLT. The time points examined in this study for both image markers and cognitive assessments includes baseline (BL), Month 6 (M6), Month 12 (M12), Month 24 (M24) and Month 36 (M36). All the participants with com-

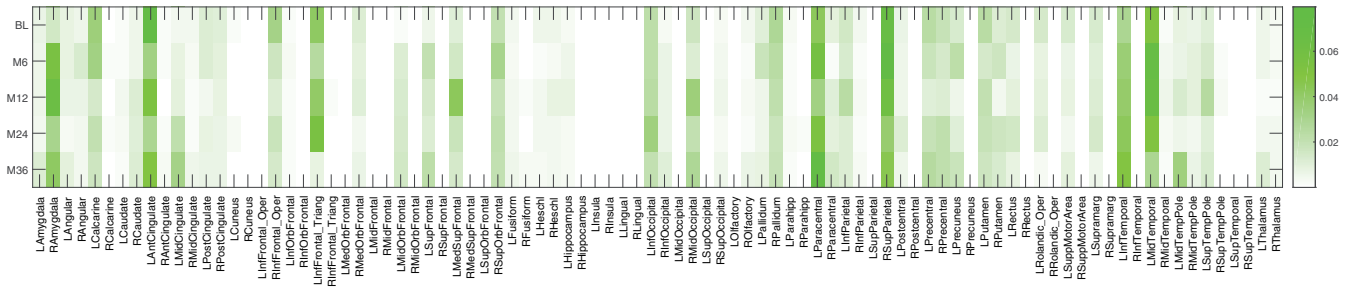

Fig. 1. Visualization of the learned weights that associate VBM imaging markers and RAVLT cognitive outcomes.

plete information of BL/M6/M12/M24/M36 MRI measurements and cognitive measures were selected in our studies, which leads to a total of 147 subjects, including 32 AD, 63 MCI, and 52 HC participants. We examined 3 RAVLT scores RAVLT_TOTAL, RAVLT_TOT6 and RAVLT_RECOG, and 2 Fluency scores FLU_ANIM and FLU_VEG.

We first evaluate the proposed method by applying it to the ADNI cohort for predicting the two types of cognitive scores using VMB, mVBM and FS markers, tracked over five time points. We compare the proposed the method against three baseline regression methods, including ridge regression (RR), Lasso regression, support vector regression (SVR). We also compare our new method against one very recent competing method in [11] and we denote it as TSA. RR is a regularized version of of linear regression to avoid over-fitting. Lasso regression performs both variable selection and regularization. SVR is the regression version of support vector machine, which has demonstrated state-of-the-art regression performance. Because RR, Lasso and SVR are designed natively to deal with static data, these methods are tested for every cognitive measures at every time point separately. Thus they can not make use of the temporal correlations over time and cross-task interrelations. TSA [11] uses longitudinal model to study the projection from temporal neuroimaging measures to the temporal cognitive scores, in which Schatten p-norm is employed to analyze the task associations. Our new method aims to leverage the temporal consistency and the correlations across different cognitive outcomes over time.

To measure prediction performance, we use standard 5-fold cross-validation method by computing the root mean square error (RMSE) between the predicted values and ground-truth values of the cognitive scores on the testing data only. In the standard 5-fold cross-validation, the data is equally and randomly divide into 5 groups. In every trail, one group is as testing data and the other four groups are used as training data. This process repeats five times in turn so that all the data can be treated as testing data by one time and the average RMSE values are reported in Table. 1.

From Table. 1, we can see that the proposed method is consistently better than RR, Lasso, since theoretically the they are the degenerated versions of our new method. In addition,

Method	RAVLT ↓	FLU ↓
RR	28.1260	17.7056
Lasso	29.0075	20.2097
SVR	10.9045	7.1885
TSA	11.1747	7.3471
Our method	**7.1305**	**6.9172**

Table 1. Performance comparison for memory score prediction measured by RMSE (↓ means that smaller is better).

because the first three methods in Table. 1 do not utilize the correlations along any mode of the tensors, they do not perform as good as our new method. Although the TSA [11] takes the advantage of temporal consistency and the relationship among tasks, it can only deal with single-modal data, while the data structures across multiple modalities of the tensor data are not explored.

Now we examine the imaging markers identified by the proposed method which take into account the longitudinal variations encoded by the cognitive scores recorded at five consecutive time points. Shown in Fig. 1 are the learned weights that associate VBM neuroimaging measures and RAVLT cognitive outcomes (magnitudes of the average weights for three RAVLT scores over five time points). The top 10 selected imaging markers in Fig. 1 are visualized on the brain volume map in Fig. 2. Due to space limit, the learned weights between other combinations of the input data modalities against the two types of cognitive outcomes, such as VBM *v.s.* FLU and FS *v.s.* RAVLT, cannot be shown in this paper. These results will be provide in the extended journal version of this paper, from which the following observations can be seen as well. In Fig. 1 we can observe clear patterns that span across all five studied time points, which prove that the markers discovered by our new method are longitudinally stable and thereby could serve as screening targets over the progress of AD.

4. CONCLUSION

In this paper, we proposed a novel high-order multi-modal multi-task feature learning model to associate the longitudi-

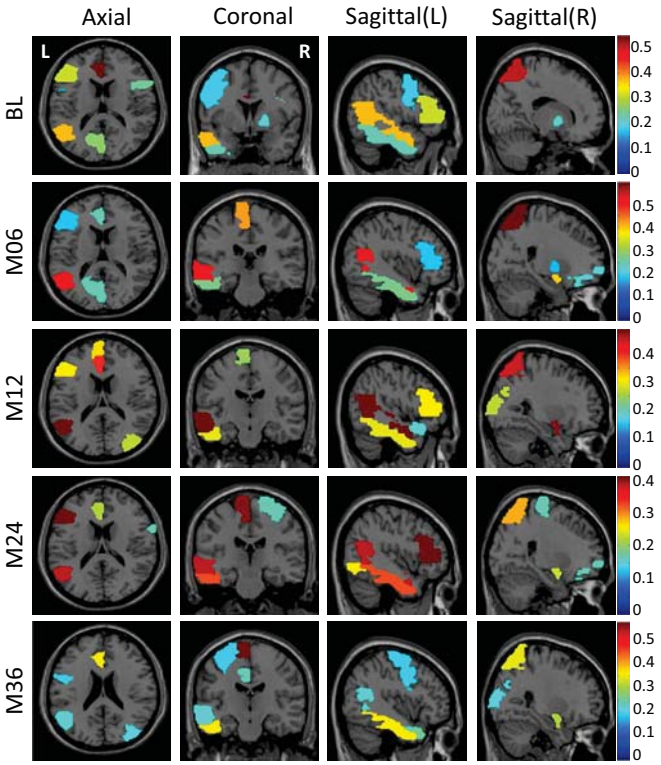

Fig. 2. Visualization of the top 10 VBM imaging markers mapped onto the brain volumes with the learned weights.

nal imaging markers to the cognitive outcomes over time. Our tensor regression model can uncover the interrelationships over different time points and utilize the temporal consistencies to enhance the learning model. Besides, through trace norm regularizations the correlations across cognitive outcomes can be captured. In addition, the group ℓ_1-norm is imposed to integrate imaging markers in multiple different modalities. The validation using ADNI imaging and cognitive data have demonstrated the effectiveness of our new method.

5. REFERENCES

[1] Hua Wang, Feiping Nie, Heng Huang, Shannon Risacher, Chris Ding, Andrew J Saykin, and Li Shen, "Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance," in *ICCV*, 2011.

[2] Hua Wang, Feiping Nie, Heng Huang, Shannon Risacher, Andrew J Saykin, Li Shen, et al., "Identifying ad-sensitive and cognition-relevant imaging biomarkers via joint classification and regression," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2011, pp. 115–123.

[3] Hua Wang, Feiping Nie, Heng Huang, Sungeun Kim, Kwangsik Nho, Shannon L Risacher, Andrew J Saykin, Li Shen, and Alzheimer's Disease Neuroimaging Initiative, "Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the adni cohort," *Bioinformatics*, vol. 28, no. 2, pp. 229–237, 2011.

[4] Hua Wang, Feiping Nie, Heng Huang, Shannon L Risacher, Andrew J Saykin, and Li Shen, "Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning," *Bioinformatics*, vol. 28, no. 12, pp. i127–i136, 2012.

[5] Hua Wang, Feiping Nie, Heng Huang, Shannon L Risacher, Andrew J Saykin, Li Shen, and Alzheimer's Disease Neuroimaging Initiative, "Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning," *Bioinformatics*, vol. 28, no. 12, pp. i127–i136, 2012.

[6] Shannon L Risacher, Li Shen, John D West, Sungeun Kim, Brenna C McDonald, Laurel A Beckett, Danielle J Harvey, Clifford R Jack, Michael W Weiner, Andrew J Saykin, et al., "Longitudinal mri atrophy biomarkers: relationship to conversion in the adni cohort," *Neurobiology of aging*, vol. 31, no. 8, pp. 1401–1418, 2010.

[7] Hua Wang, Feiping Nie, Heng Huang, Jingwen Yan, Sungeun Kim, Shannon Risacher, Andrew Saykin, and Li Shen, "High-order multi-task feature learning to identify longitudinal phenotypic markers for alzheimer's disease progression prediction," in *NIPS*, 2012.

[8] Hua Wang, Feiping Nie, Heng Huang, Jingwen Yan, Sungeun Kim, Kwangsik Nho, Shannon L Risacher, Andrew J Saykin, Li Shen, and Alzheimer's Disease Neuroimaging Initiative, "From phenotype to genotype: an association study of longitudinal phenotypic markers to alzheimer's disease relevant snps," *Bioinformatics*, vol. 28, no. 18, pp. i619–i625, 2012.

[9] Hua Wang, Feiping Nie, Heng Huang, and Chris Ding, "Heterogeneous visual features fusion via sparse multimodal machine," in *CVPR*, 2013.

[10] Hua Wang, Feiping Nie, and Heng Huang, "Multi-view clustering and feature learning via structured sparsity," in *ICML*, 2013.

[11] Xiaoqian Wang, Jingwen Yan, Xiaohui Yao, Sungeun Kim, Kwangsik Nho, Shannon L Risacher, Andrew J Saykin, Li Shen, Heng Huang, et al., "Longitudinal genotype-phenotype association study via temporal structure auto-learning predictive model," in *RECOMB*, 2017.