Published in final edited form as:
Proc IEEE Int Symp Biomed Imaging. 2018 April ; 2018: 1048–1051. doi:10.1109/ISBI.2018.8363751.

INFANT BRAIN DEVELOPMENT PREDICTION WITH LATENT PARTIAL MULTI-VIEW REPRESENTATION LEARNING

Changqing Zhang[1,2], **Ehsan Adeli**[3], **Zhengwang Wu**[1], **Gang Li**[1], **Weili Lin**[1], and **Dinggang Shen**[1]

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

[2]School of Computer Science and Technology, Tianjin University, Tianjin, China

[3]Department of Psychiatry and Behavioral Sciences, Stanford University, California, USA

Abstract

The early postnatal period witnesses rapid and dynamic brain development. Understanding the cognitive development patterns can help identify various disorders at early ages of life and is essential for the health and well-being of children. This inspires us to investigate the relation between cognitive ability and the cerebral cortex by exploiting brain images in a longitudinal study. Specifically, we aim to predict the infant brain development status based on the morphological features of the cerebral cortex. For this goal, we introduce a *multi-view multi-task* learning approach to dexterously explore complementary information from different time points and handle the missing data simultaneously. Specifically, we establish a novel model termed as *Latent Partial Multi-view Representation Learning*. The approach regards data of different time points as different views, and constructs a latent representation to capture the complementary underlying information from different and even incomplete time points. It uncovers the latent representation that can be jointly used to learn the prediction model. This formulation elegantly explores the complementarity, effectively reduces the redundancy of different views, and improves the accuracy of prediction. The minimization problem is solved by the Alternating Direction Method of Multipliers (ADMM). Experimental results on real data validate the proposed method.

Index Terms

Infant brain development; Longitudinal analysis; Cognitive ability; Multi-view learning

1. INTRODUCTION

Exploring the quantitative relationship between the cognitive behavior scores and the brain development status during childhood period is of great importance for understanding the cognitive ability development at early ages. Although it is very important for health and well-being of children, the related works for this task are scarce, and only recently researchers started to associate brain development measures with risks of autism [1]. In this paper, we build quantitative mappings between five essential cognitive scores [2] (i.e., visual reception scale (VRS), fine motor scale (FMS), receptive language scale (RLS), expressive language scale (ELS), and early learning composite (ELC)) and the longitudinal morphological features of the cerebral cortex. The challenges of such longitudinal infant study would lie in the following factors: (1) the samples are often very limited (i.e., Small-Sample-Size (SSS) problem), due to novelty of such studies; (2) missing data at certain time points are unavoidable and even very serious in longitudinal studies due to various reasons (e.g., no show-up at certain time points or dropouts) [3]; (3) unlike single output regression, our problem comprises multiple outputs that are naturally correlated.

The missing data issue is itself a key challenge, and, to address this, one straightforward approach is to learn one model based on the available data at each time point and then integrate the outputs of these models. To exploit multiple data sources, some improved methods [4, 5] further manually group samples according to the availability of data source, and then learn one model for each group. However, in both of these types of approaches, the missing data contributes to less effective usage of the partly available data, making SSS problem even more serious. In addition, the latter approach [4] will be rather complex with the increase in the number of data sources. Matrix completion methods [6, 7] usually devote themselves to recover missing values with a low-rank constraint. To be able to utilize the low-rank assumption, these approaches assume that the data is uniformly and randomly missing, which is not the case for our application of longitudinal early brain development study.

The longitudinal MRI data comprises multiple data sources from multiple time points that describe subjects from multiple views. Note that, for each time point, the data corresponding to a subset of subjects is missing, as shown in Fig. 1. To build the relation between the incomplete multi-view data and multiple cognitive scores, we propose a novel partial multi-view multi-task regression method, termed as *Latent Partial Multi-view Representation Learning*. Our model seeks to reconstruct a comprehensive and compact latent representation for each subject, from the observed data at multiple time points. Then, a prediction model is learned based on the inferred latent representation, as shown in Fig. 2. The main advantages of our model include: 1) Unlike most existing multi-view methods [8, 9] that learn models directly on the original noisy features, our model elegantly exploits the complementarity among different time points and reduces effectively the redundancy of the learned latent representations. 2) Our regression model is learned based on all N subjects, while existing methods [4, 5] learn multiple regression models based on different subsets of subjects, which is not applicable for dealing with SSS problem. The optimization of the proposed method is conducted by the Alternating Direction Method of Multipliers (ADMM) [10].

2. MATERIAL AND PREPROCESSING

Material—In our study, the T1- and T2-weighted MR images from 23 infant subjects at 9 different time points (i.e., 0, 3, 6, 9, 12, 18, 24, 36 and 48 months) are collected. Some subjects have not shown-up for scans at certain time points, which causes the missing data issue, as illustrated in Fig. 1. Five of the Mullen behavior scores [2], i.e., Visual Reception Scale (VRS), Fine Motor Scale (FMS), Receptive Language Scale (RLS), Expressive Language Scale (ELS), and Early Learning Composite (ELC) are measured for each subject at 48 months. Note that the fifth score (i.e., ELC) can be interpreted as the composite of the first four scores.

Preprocessing & Features—As mentioned before, our goal is to predict five behavior scores as early as possible in the early years of life. As studied in the literature [11, 1], morphological features from the cerebral cortex correlate with such behavioural measurements. To get the morphological features of cerebral cortex, the cerebral cortical surface is built following the pipeline in [12]. Then, at each vertex of the cerebral cortical surface, 7 measurements of the morphological properties are computed, i.e., cortical thickness, local gyrification index, mean curvature, vertex volume, sulcal depth measured in Euclidean distance, sulcal depth measured in string distance, and normalized area of the local vertex [13, 14]. Next, we register the FreeSurfer parcellation atlas [12] onto each surface to get the individual parcellation, after which for each Region- Of-Interest (ROI) of the FreeSurfer parcellation, the average of the above 7 measurements can be obtained. For the entire cerebral cortex, the FreeSurfer parcellation includes 70 anatomical meaningful ROIs [12], and for each ROI we can compute 7 mean morphological measurements from all the vertices within the ROIs. Therefore, totally, for each subject at one time point with available MRI data, a 490-dimensional feature vector can be obtained.

3. MODEL

In this section, we introduce a novel multi-view multi-task learning method, which overcomes the limitation of discarding or completing incomplete data in advance, and thus could fully take advantages of all the observed data based on the latent multi-view representation for each subject.

3.1. Latent Partial Multi-view Representation Learning

We denote the multiple-time-point data as $\{\mathbf{X}_1, \ldots, \mathbf{X}_T; \mathbf{Y}\}$, where $\mathbf{X}_t \in \mathbb{R}^{D \times N}$ is the data matrix at the t^{th} time point and $\mathbf{Y} \in \mathbb{R}^{C \times N}$ is the score matrix. In our model, we formulate the learning task as a multi-task (C scores) multi-view learning problem with each view comprised by data from one of the T time points. We aim to uncover the multi-view latent representation which holds the reconstructive ability for the data at different time points. Accordingly, the formulation is as follows:

$$\min_{\mathbf{H}} \sum_{t=1}^{T} \mathscr{V}(f_t(\mathbf{H}), \mathbf{X}_t), \quad (1)$$

where $\mathscr{V}(\cdot, \cdot)$ measures the reconstruction loss and $f_t(\cdot)$ indicates the underlying mapping from the latent representation $\mathbf{H}$ to the observations at the t^{th} time point, i.e., $\mathbf{X}_t$, $\forall t \in \{1, \ldots, T\}$, defined as:

$$\mathscr{V}(f_t(\mathbf{H}), \mathbf{X}_t) = \|(f_t(\mathbf{H}) - \mathbf{X}_t)\|_{2,1}, \quad (2)$$

where $\|\cdot\|_{2,1}$ is the $\ell_{2,1}$-norm of the residual encouraging the columns of the matrix to be zero. The underlying assumption is that the corruptions are sample-specific. Hence, this loss leads to a level of robustness against sample outliers. For the mapping $f_t(\cdot)$, we employ linear projection in our model, which is a simple but effective technique especially for high-dimensional data. Accordingly, we have:

$$\mathscr{V}(f_t(\mathbf{H}), \mathbf{X}_t) = \|\mathbf{P}_t\mathbf{H} - \mathbf{X}_t\|_{2,1}. \quad (3)$$

Based on the learned latent representation **H** from the multiple views, we can define the following multi-task regression term to predict our five scores:

$$\min_{\mathbf{W}} \mathscr{L}(\mathbf{W}, \mathbf{H}, \mathbf{Y}) = \|\mathbf{W}\mathbf{H} - \mathbf{Y}\|_1. \quad (4)$$

This ℓ_1 loss function leads to a robust loss [15]. Note that, **W** is learned based on all the N samples regardless of the missing status.

Since the tasks of interest themselves are highly interrelated, we introduce a well-known low-rank regularization [16] for the multi-task prediction model **W** as:

$$\mathscr{R}(\mathbf{W}) = \|\mathbf{W}\|_*, \quad (5)$$

where $\|\cdot\|_*$ is the matrix nuclear-norm. Putting these terms in a unified optimization problem, our objective function is induced as:

$$\begin{aligned} &\min_{\Omega} \|\mathbf{W}\mathbf{H} - \mathbf{Y}\|_1 + \alpha \sum_{t=1}^{T} \omega_t^r \|\mathscr{P}_{\mathbf{O}_t}(\mathbf{P}_t\mathbf{H} - \mathbf{X}_t)\|_{2,1} + \beta\|\mathbf{W}\|_* \quad (6) \\ &s.t. \sum_{t=1}^{T} \omega_t = 1, \omega_t \geq 0; \ \mathbf{P}_t^\top \mathbf{P}_t = \mathbf{I}, \ t = 1, \cdots, T. \end{aligned}$$

For convenience, we denote $\Omega = \{\mathbf{W}, \mathbf{H}, \{\mathbf{P}_t\}_{t=1}^T, \{\omega_t\}_{t=1}^T\}$ as the variable set to be optimized, and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_T)$ is the weight vector for different time points. $\alpha > 0$ and $\beta > 0$ encode the beliefs for reconstruction and task correlation, respectively. $r > 1$ for ω_t is used to avoid a trivial solution that only considers one of the T time points and adjusts the complementarity of multiple time points [17]. The constraint $\mathbf{P}_t^\top \mathbf{P}_t = \mathbf{I}$ is introduced, since without this constraint $\mathbf{P}_t$ can be pushed arbitrarily close to zero only by re-scaling $\mathbf{P}_t/s$ and $\mathbf{H}s$ $(s > 0)$ while preserving the same loss. Moreover, our model can be efficiently solved with the constraint (see $\mathbf{P}_t$-subproblem in optimization part). $\mathscr{P}_{\mathbf{O}_t}(\cdot)$ is a filter function to take care of the incomplete data for the t^{th} time point. Let o_t^s be an indicator variable showing the existence of data for subject s in time point t, i.e., $o_t^s = 1$ if we have the data available, and a very small scalar $\varepsilon > 0$ otherwise. $\mathbf{o}_t$ will then be defined as the indicator vector from all indicator variables of training samples. Accordingly, we can define a diagonal matrix $\mathbf{O}_t = \text{diag}(\mathbf{o}_t)$, denoted as the filter matrix of the t^{th} view, and hence $\mathcal{P}_{\mathbf{O}_t}(\mathbf{P}_t\mathbf{H} - \mathbf{X}_t) = (\mathbf{P}_t\mathbf{H} - \mathbf{X}_t)\mathbf{O}_t$. Note that $\varepsilon > 0$ is a small value to strictly guarantee the unique solution of the optimization problem (see **H**-subproblem in next subsection).

Optimization—Our objective function in Eq. (6) simultaneously seeks the latent representations from multiple views and learns the multi-task prediction model with respect to the latent representations. Since the objective function is not jointly convex with respect to all the variables $\mathbf{P}_t$, $\mathbf{H}$, $\mathbf{W}$, we employ Augmented Lagrange Multiplier (ALM) with Alternating Direction Minimizing (ADM) strategy [10] and omit the detail of optimization due to space limitation.

4. EXPERIMENTS

We conduct experiments the real infant brain data to evaluate our method. The performance is measured with Root Mean Squared Error (RMSE). All the parameters are tuned in the set $\{10^{-3}, 10^{-2}, 0.1, 1, 10, 10^2, 10^3\}$ through a nested leave-one-out cross-validation.

Performance with different number of time points—For the real infant brain data, we first run our method with the data from different number of time points. According to Table 1, our model could well leverage the data of different time points for promising performance. The improvements are not significant by incorporating the data at the 18th, 24th and 36th month. One reason is that data at these time points are much more severely incomplete (for example, the missing rate at the 36th month is $14/23 \approx 61\%$). The second reason can be related to the law of diminishing marginal return property.

Performance Comparison—We adopt two strategies to process the data to make them suitable for the existing multitask methods. Specifically, the first way is to complete the missing values simply with zero, and the second way is to fill the incomplete values with the averaged values of the observed ones (results for both strategies are included in the table, for all compared methods). We compare our model with the following methods: 1) NN (nearest neighbour); 2) MtJFS (Multi-Task Learning with Joint Feature Selection) [18]; 3) RMTL (Robust Multi-Task Feature Learning) [19]; 4) TrMTL (Trace-Norm Regularized Multi-Task Learning) [16]. From Table 2, it can be observed that simply filling the missing values with zero is not reasonable since the performance is usually relatively low.

Our method outperforms both TrMTL and RMTL that also constrain the prediction model to be low-rank, thus validating the power of learning the latent multi-view representation.

5. CONCLUSION

We have proposed to explore the relation between the cognitive ability and the cerebral cortex, and developed a novel multi-task multi-view regression model for the challenging problem. Based on the latent representation, our model dexterously addresses the challenge of learning with incomplete longitudinal data. We also introduced an optimization algorithm for the proposed method and validate the effectiveness on real data. Since the problem is new, there are still valuable directions for the future research such as: (1) Nonlinear mappings (e.g., kernel technique) from latent representation to observations will be considered; (2) More data should be acquired for better performance.

Acknowledgments

This work was supported in part by NIH grants (MH100217, MH108914, MH107815, MH110274), and National Natural Science Foundation of China (Grand No: 61602337).

References

1. Hazlett Heather CodyGu HongbinMunsell Brent C, Kim Sun HyungStyner MartinWolff Jason J, Elison Jed T, Swanson Meghan R, Zhu HongtuBotteron Kelly N. , et al. Early brain development in infants at high risk for autism spectrum disorder. Nature. 2017; 542(7641):348–351. [PubMed: 28202961]
2. Mullen Eileen M, , et al. Mullen scales of early learning. AGS Circle Pines; MN: 1995.
3. Meng YuLi GangGao YaozongLin WeiliShen Dinggang. Learning-based subject-specific estimation of dynamic maps of cortical morphology at missing time points in longitudinal infant studies. Human Brain Mapping. 2016; 37(11):4129–4147. [PubMed: 27380969]
4. Yuan LeiWang YalinThompson Paul M, Narayan Vaibhav A, Ye Jieping, et al. Alzheimer’s Disease Neuroimaging Initiative. Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. Neuro Image. 2012; 61(3):622–632. [PubMed: 22498655]
5. Li Shao-YuanJiang YuanZhou Zhi-Hua. Partial multiview clustering. AAAI. 2014
6. Cai Jian-FengCandès Emmanuel J, Shen Zuowei. A singular value thresholding algorithm for matrix completion. SIAM J Optimization. 2010; 20(4):1956–1982.
7. Candes EmmanuelRecht Benjamin. Exact matrix completion via convex optimization. Communications of the ACM. 2012; 55(6):111–119.
8. Zheng ShuaiCai XiaoDing ChrisNie FeipingHuang Heng. A closed form solution to multi-view low-rank regression. AAAI. 2015:1973–1979.
9. Sharma AbhishekKumar AbhishekDaume HalJacobs David W. Generalized multiview analysis: A discriminative latent space. CVPR. 2012:2160–2167.
10. Boyd StephenParikh NealChu EricPeleato BorjaEckstein Jonathan. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Machine Learning. 2011; 3(1):1–122.
11. Harris Kenneth D, Mrsic-Flogel Thomas D. Cortical connectivity and sensory coding. Nature. 2013; 503(7474):51–58. [PubMed: 24201278]
12. Li GangWang LiShi FengLin WeiliShen Dinggang. Constructing 4d infant cortical surface atlases based on dynamic developmental trajectories of the cortex. MICCAI. 2014:89–96.
13. Li GangWang LiShi FengLyall Amanda E, Lin WeiliGilmore John H, Shen Dinggang. Mapping longitudinal development of local cortical gyrification in infants from birth to 2 years of age. Journal of Neuroscience. 2014; 34(12):4228–4238. [PubMed: 24647943]
14. Fischl Bruce. Freesurfer. Neuroimage. 2012; 62(2):774–781. [PubMed: 22248573]
15. Adeli-Mosabbeb EhsanThung Kim-HanAn LeShi FengShen Dinggang. Robust feature-sample linear discriminant analysis for brain disorders diagnosis. NIPS. 2015:658–666.
16. Ji ShuiwangYe Jieping. An accelerated gradient method for trace norm minimization. ICML. 2009:457–464.
17. Wang MengHua Xian-ShengYuan XunSong YanDai LiRong. Optimizing multi-graph learning: towards a unified video annotation scheme. ACM MM. 2007:862–871.
18. Argyriou AndreasEvgeniou TheodorosPontil Massimiliano. Multi-task feature learning. NIPS. 2007; 19:41.

19. Chen JianhuiZhou JiayuYe Jieping. Integrating low-rank and group-sparse structures for robust multi-task learning. ACM KDD. 2011:42–50.

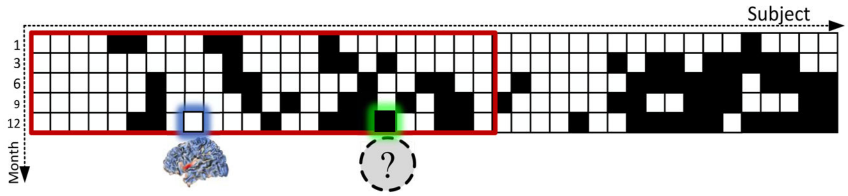

Fig. 1.
Illustration of our infant brain image data set: The white blocks indicate the data available while the black blocks indicate missing data. We have 23 subjects with behavior scores, denoted by the red rectangle.

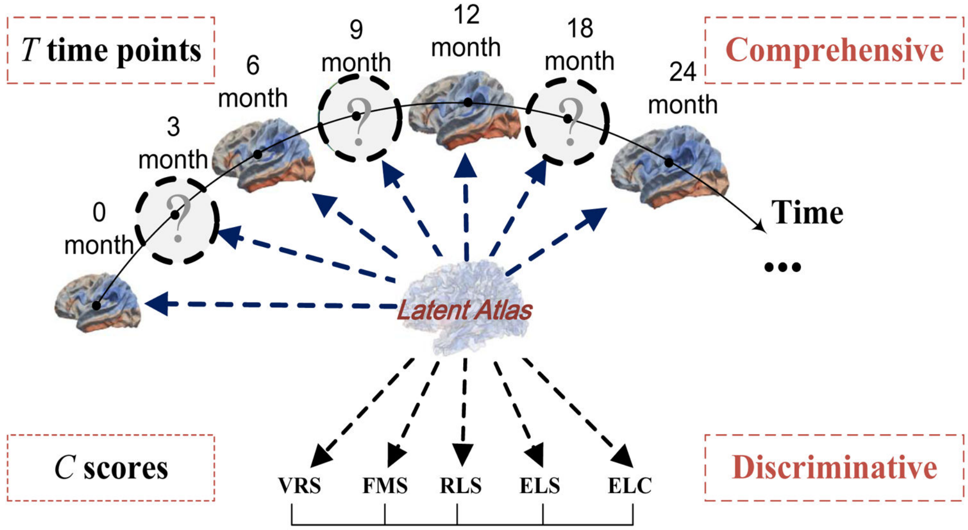

Fig. 2.

Illustration of Latent Partial Multi-view Representation Learning. Our model uncovers the comprehensive and discriminative latent representation (termed as latent atlas derived from medical image field) jointly from incomplete observations, based on which the multi-task (C scores) multiview (T time points) prediction model is learned.

Table 1

Results with different number of time points.

Time Points	VRS	FMS	RLS	ELS	ELC	Mean	MRate
0–03 Month	0.164	0.206	0.159	0.183	0.159	0.174	20.8%
0–06 Month	0.162	0.205	0.151	0.176	0.155	0.170	23.6%
0–09 Month	0.162	0.189	0.143	0.163	0.143	0.160	22.9%
0–12 Month	0.158	0.190	0.137	0.164	0.138	0.158	25.8%
0–18 Month	0.157	0.190	0.137	0.164	0.136	0.157	25.6%
0–24 Month	0.159	0.191	0.138	0.162	0.137	0.157	29.7%
0–36 Month	0.162	0.194	0.140	0.165	0.141	0.160	37.2%
0–48 Month	0.162	0.189	0.139	0.165	0.138	0.158	35.7%

Table 2

Performance comparison (RMSE).

Method	VRS	FMS	RLS	ELS	ELC	Mean
NN	0.200	0.220	0.259	0.291	0.209	0.236
	0.219	0.259	0.165	0.196	0.182	0.204
MtJFS	0.284	0.296	0.279	0.278	0.286	0.285
	0.276	0.273	0.189	0.214	**0.134**	0.217
RMTL	0.313	0.321	0.242	0.229	0.273	0.276
	0.146	0.200	0.178	0.188	0.137	0.170
TrMTL	0.349	0.373	0.280	0.256	0.319	0.315
	0.279	0.276	0.192	0.217	0.136	0.220
Ours	**0.162**	**0.189**	**0.139**	**0.165**	0.138	**0.158**