2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)
Venice, Italy, April 8-11, 2019

Sparse Low-rank Constrained Adaptive Structure Learning using Multi-template for Autism Spectrum Disorder Diagnosis

Fanglin Huang, *Ahmed Elazab, Le OuYang, Joseph Tan, Tianfu Wang*, *Baiying Lei*[*]
School of Biomedical Engineering, Health Science Center, Shenzhen University, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen, China, 518060
(*Corresponding author: Baiying Lei (leiby@szu.edu.cn))

ABSTRACT

Autism spectrum disorder (ASD) is a developmental disability that causes severe social, communication and behavioral challenges. Up to now, many imaging-based approaches for ASD diagnosis have been proposed. However most of them limited to single template. In this paper, we propose a novel sparse low-rank constrained multi-templates data based method for ASD diagnosis, which performs feature selection and adaptive local structure learning simultaneously. Specifically, we encode modularity prior while constructing functional connectivity (FC) brain networks from different templates for each subject. After extracting features from FC networks, feature selection is applied. Meanwhile, the local structure is learnt via an adaptive process. Extensive experiments are conducted to demonstrate the effectiveness of our proposed method on the Autism Brain Imaging Data Exchange (ABIDE) database. Experimental results verify our proposed method can enhance the diagnosis performances and outperform the commonly used and state-of-the-art methods.

Index Terms—Autism spectrum disorder, sparse low-rank, adaptive structure learning, multi-template

1. INTRODUCTION

Autism spectrum disorder (ASD) is a prevalent and highly heterogeneous developmental disability, defined by deficits in social communication and interaction, restricted interest and repetitive behaviors [1]. These conditions begin during early childhood and typically last throughout an individual's life. Although there is currently no cure for ASD, opportunely intervention services can help children learn important skills and greatly improve their quality of life.

As ASD is considered as a disability with disrupted brain connectivity patterns [2], resting state functional magnetic resonance imaging (rs-fMRI) has been recognized as a versatile tool to reveal its pathological mechanism in recent years. In particular, function connectivity (FC), represented by the temporal correlation of blood oxygen-level dependent (BOLD) signals in different brain regions, reflects the close interaction of multiple structurally separated brain regions. Since the FC network may be damaged by pathology, we can extract important features from it to effectively diagnose ASD. Pearson's correlation (PC) coefficient is often estimated as FC between brain regions, due to its capacity of fitting data and computational simplicity. Unfortunately, PC succeeds in estimating simple linear relationship, but it fails to find biological mechanisms of brain diseases. There is FC network constructed by incorporating sparsity prior [3]. However, considering the complexity of the brain network, encoding sparsity alone is insufficient. Inspired by the work of Qiao *et al*. [4], we take modularity structure into the FC network, so as to effectively fit the data and encode prior information reasonably.

In fact, features obtained from the FC network are often with high dimension, which may result in overfitting issue with limited sample size. In this regard, feature selection is necessary to remove the redundant or unrelated features. Among them, the spectral based feature selection methods have been widely used in machine learning. However, most of them generate similarity matrix and select features separately. Thus, the similarity matrix is often derived from the raw data with lots of noise, which may disrupt the local manifold structure. Moreover, the similarity matrix obtained by conventional methods such as k-nearest neighbors only has one connected component. In the ideal state, the number of connected component in the similarity matrix should be the same as the number of cluster. For this reason, an optimal graph matrix should be learnt while performing feature selection.

To date, most feature selection methods only take a single template into account. However, it has been demonstrated that the multi-template based approaches are more promising to capture disease status related discriminative information [5]. Therefore, we utilize multiple diverse templates for ASD diagnosis.

To address the above problems, we propose a novel sparse low-rank constrained multi-templates data based method for ASD diagnosis, which does feature selection and adaptive local structure learning concurrently. Specifically, we construct FC networks by introducing the sparse low-rank representation (SLR). Second, we introduce sparse constraint for feature selection. Meanwhile, a similarity matrix can be acquired via an adaptive process. By adding a rank constraint to the associated Laplace matrix, the similarity matrix has the exactly connected components. We

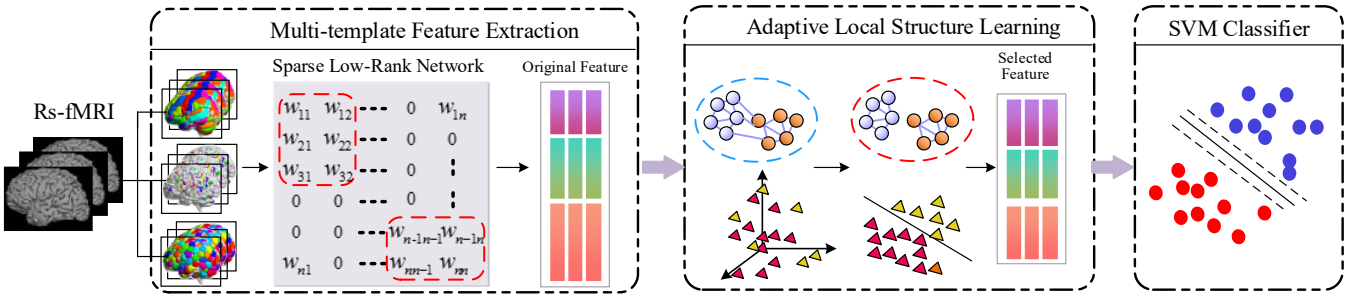

Fig.1. Illustration of the proposed method.

thus can learn accurate manifold structure for diagnosis.

2. METHODOLOGY

2.1. System overview

The steps for ASD identification are illustrated in Fig. 1. First, we construct FC networks from different templates for each subject using sparse and low-rank constraints. Second, we extract multiple sets of features from the multiple sets of networks corresponding to different templates. Third, we conduct feature selection for each set of features based on the proposed method. Finally, the selected relevant features are together fed into a linear support vector machine (SVM) classifier for classification.

2.2. Spare low-rank FC network

Let $\mathbf{T} \in \mathbb{R}^{m\times r}$ denotes the data we input where m and r are the number of time series and regions of interest (ROIs), respectively. Then we construct a FC network by incorporating modularity prior, which is defined as

$$\min_{\mathbf{M}} \|\mathbf{T}-\mathbf{T}\mathbf{M}\|_{\mathrm{F}}^2+\lambda_1\|\mathbf{M}\|_1+\lambda_2\|\mathbf{M}\|_*. \quad (1)$$

where $\|\mathbf{T}-\mathbf{T}\mathbf{M}\|_{\mathrm{F}}^2$ is used to fit the data, $\|\mathbf{M}\|_1$ is used for measuring the sparsity, and $\|\mathbf{M}\|_*$ is the rank constraint. The combination of sparse and low-rank regularizers can model the modularity prior using λ_1 and λ_2 as turning parameters.

From the above processing, we get multiple sets of FC matrix from different templates for each subject. After that, the upper triangles of the matrixes are used according to symmetry. Then we reshape these measures into the vectors for each subject to generate feature matrixes.

2.3. Spare feature selection

Let $\mathbf{X} \in \mathbb{R}^{n\times d}$ and $\mathbf{y} \in \mathbb{R}^{n}$ denote the FC feature matrix and its corresponding response vector of n subjects and d features, respectively. We assume there is a linear relationship between $\mathbf{X}$ and $\mathbf{y}$. We thus have

$$\min_{\mathbf{w}} \|\mathbf{y}-\mathbf{X}\mathbf{w}\|_{\mathrm{F}}^2, \quad (2)$$

Where weight coefficient vector $\mathbf{w} \in \mathbb{R}^{d}$ measures the relative importance of the input features in predicting the disease status. As is well known, $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ is the approximate solution of Eq. (2), but the inverse operation is often ill-posed when there are redundant features. Therefore, we impose a l_1-norm regularizer to tackle this issue and encourage the sparsity among features, which is given by

$$\min_{\mathbf{w}} \|\mathbf{y}-\mathbf{X}\mathbf{w}\|_{\mathrm{F}}^2+\gamma\|\mathbf{w}\|_1, \quad (3)$$

where γ denotes the nonnegative sparsity control parameter.

2.4. Structured optimal graph feature selection

Let $\mathbf{S} \in \mathbb{R}^{n\times n}$ and s_{ij} denote similarity matrix and the $(i\ j)$-th element of $\mathbf{S}$. We assume $\mathbf{x}_i \in \mathbb{R}^{d}$ is linked to all the other data with the probability s_{ij}, and such probability can be identified as the similarity between them. Since the distance between two samples is inversely related to their similarity, we can obtain the similarity matrix by solving

$$\min_{\mathbf{s}_i^{\mathrm{T}}\mathbf{1}=\mathbf{1},0\le s_{ij}\le 1} \sum_{\mathrm{i,j}}\|\mathbf{x}_i-\mathbf{x}_j\|_2^2 s_{ij}+\alpha\|\mathbf{S}\|_{\mathrm{F}}^2, \quad (4)$$

where α is the turning parameter. The second term is imposed to avoid trivial solution. However, $\mathbf{S}$ obtained from Eq. (4) only has one connected component in most instance.

In spectral analysis, $\mathbf{L}_{\mathrm{s}}=\mathbf{D}-(\mathbf{S}^{\mathrm{T}}+\mathbf{S})/2$ is called Laplacian matrix, where $\mathbf{D}\ (d_{ii}=\sum_j(s_{ij}+s_{ji})/2)$ is the corresponding degree matrix. According to the property of the Laplacian matrix, the number of occurrences of 0 in the eigenvalue is the number of connected components in the graph. Thus, we add a rank constraint in Eq. (4) as follows

$$\min \sum_{i,j}\left(\|\mathbf{x}_i-\mathbf{x}_j\|_2^2 s_{ij}+\alpha s_{ij}^2\right) \quad (5)$$
$$s.t.\ \forall i, \mathbf{s}_i^T\mathbf{1}=\mathbf{1}, 0\le s_{ij}\le 1, \mathrm{rank}(\mathbf{L}_{\mathrm{s}})=n-c.$$

Where c is the number of label category.

To handle the above problem, let $\sigma_i(\mathbf{L}_{\mathrm{s}})$ denotes the i-th smallest eigenvalue of $\mathbf{L}_{\mathrm{s}}$. Due to the positive semi-definite nature of $\mathbf{L}_{\mathrm{s}}$, we have $\sigma_i(\mathbf{L}_{\mathrm{s}})\ge 0$. So $\mathrm{rank}(\mathbf{L}_{\mathrm{s}})=n-c$ is equivalent to $\sum_{i=1}^{c}\sigma_i(\mathbf{L}_s)=0$. According to Ky Fan's Theorem, we have

$$\sum_{i=1}^{c}\sigma_i(\mathbf{L}_{\mathrm{s}})=\min_{\mathbf{F}\in\mathbb{R}^{n\times c},\mathbf{F}^{\mathrm{T}}\mathbf{F}=\mathbf{I}}\mathrm{Tr}(\mathbf{F}^{\mathrm{T}}\mathbf{L}_{\mathrm{S}}\mathbf{F}), \quad (6)$$

Therefore, the Eq.(6) is same as the following formula

$$\min \sum_{i,j}\left(\|\mathbf{x}_i-\mathbf{x}_j\|_2^2 s_{ij}+\alpha s_{ij}^2\right)+2\lambda\mathrm{Tr}(\mathbf{F}^{\mathrm{T}}\mathbf{L}_{\mathrm{S}}\mathbf{F}) \quad (7)$$
$$s.t.\ \forall i, \mathbf{s}_i^T\mathbf{1}=\mathbf{1}, 0\le s_{ij}\le 1, \mathbf{F}\in\mathbb{R}^{n\times c}, \mathbf{F}^{\mathrm{T}}\mathbf{F}=\mathbf{I}.$$

Obviously, when λ is large enough, $\mathrm{Tr}(\mathbf{F}^{\mathrm{T}}\mathbf{L}_{\mathrm{S}}\mathbf{F})$ will approach to zero and thus $\sum_{i=1}^{c}\sigma_i(\mathbf{L}_{\mathrm{s}})=0$ holds. It allots adaptive neighbors to each data point adaptively, which means the similarity between the data points changes. Accordingly, $\mathbf{S}$ is improved until it contains the exactly connected component.

The main goal of manifold learning is to find low-dimensional smooth manifolds embedded in high-dimensional data spaces. We can map high-dimensional features to low-dimension manifold by $\mathbf{X}\mathbf{w}$, where $\mathbf{w} \in \mathbb{R}^{d}$ is the projection vector. Finally, combining sparse feature

selection mentioned above, we arrive to a new feature selection method, i.e.,

$$\min \sum_{i,j} \left(\left\| \mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{x}_j \right\|_2^2 s_{ij} + \alpha s_{ij}^2 \right) + \gamma \|\mathbf{w}\|_1 + 2\lambda \mathrm{Tr}(\mathbf{F}^{\mathrm{T}} \mathbf{L}_S \mathbf{F}) + \beta \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_F^2, \quad (8)$$
$$s.t. \forall i, \mathbf{s}_i^T \mathbf{1} = \mathbf{1}, 0 \le s_{ij} \le 1, , \mathbf{F} \in \mathbf{R}^{n \times c}, \mathbf{F}^{\mathrm{T}} \mathbf{F} = \mathbf{I}.$$

3. EXPERIMENTAL SETTING

3.1. Image preprocessing

All the subjects in our experiment are collected from the ABIDE database, and we download rs-fMRI scans from the New York University (NYU) Langone Medical Center and the University of Michigan: sample 1 (UM_1). Among them, NYU dataset contains 171 subjects (73 ASD and 98 normal control (NC)), while UM_1 dataset includes 82 subjects (36 ASD and 46 NC), respectively. Our experiment includes male and female subjects aged from 6 to 40.

We extract features from rs-fMRI using the Data Processing Assistant for Resting-State fMRI (DPARSF). Specifically, we first discard the first 10 obtained rs-fMRI volumes of each subject. Then, all rs-fMRI images are normalized to the MNI atlas space, and Nuisance variable regression is further conducted. After that, the resulting rs-fMRI images are partitioned into 116, 160, and 200 ROIs according to the different templates (i.e., Automated Anatomical Labeling, Dosenbach 160, Craddock 200). Band-pass filtering is applied to rs-fMRI time series for each ROI.

3.2. Experimental setting

In our study, we compare our proposed method with other methods, i.e., Least absolute shrinkage and selection operator method (LASSO) [6], Laplacian score (LS) [7], and unsupervised feature selection with structured graph optimization (SOGFS) [8].

We adopt nested 10-fold cross-validation to evaluate the performances of the competitive methods. Aiming to obtain the optimal parameters of each method, a 5-fold inner loop is used. Moreover, to avoid the biased result caused by the fold selection, we repeat the processing five times and report the average of the results. To make the experiments fair enough, we set the parameters of all methods using the same strategy, i.e., $\{10^{-3}, \dots, 10^{3}\}$.To quantitatively estimate the diagnosis performances, we utilize the metrics of accuracy (ACC), sensitivity (SEN), precision (PREC), youden index (You), F-scores (F1), and area under the receiver operating characteristic (ROC) curve (AUC).

4. RESULTS AND DISCUSSIONS

In our study, we have three segmentation templates. Table 1 shows that, 116 ROIs reaches to the highest performance on both datasets of NYU and UM_1 compared to other two single templates. It indicates that 116 ROIs provides the suitable number of brain regions for classification as a single template. Meanwhile, the 200 ROIs achieves the lowest performance. The reason is that more segmentation is beneficial to provide more features but fails to provide representative features. We observe that some columns of the feature values in the 200 ROIs template turn out to be zero, which means there is no feature found in these ROIs. We also observe that our multi-template based method can generally achieve significant higher results compared to these single-template based methods. For example, the highest accuracy achieved by the single-template based method is only 75.30% and 70.93%, while our multi-template based method can reach to 78.36 and 81.74%, respectively.

To evaluate the benefits of the SLR, we also apply our method on the features derived from different FC (i.e., PC, sparse representation (SR), low-rank representation (LR)). As shown in Table 1, the SLR network achieves the highest classification performances. It indicates that the SLR network incorporating the modularity prior by adding sparse and low-rank regularizers can construct functional brain networks more accurate.

We compare our method with LASSO, LS, SOGFS in Fig. 3, and Fig. 4, and other state-of-the-art approaches in Table 2. Specifically, we find that the proposed method performs generally better than competing methods in terms of various metrics. Obviously, our method outperforms its counterparts and eventually proves that combining feature selection and local structure learning can effectively boost ASD diagnosis.

We report the most discriminative features selected from rs-fMRI to recognize the ASD patients from NC. The thickness of lines stands for the corresponding weight. From Fig.5, we find the connections facilitating accurate ASD diagnosis not only exist within the same hemisphere, but across both hemispheres. Besides, the connections weights exist in left hemispheres are larger than right hemispheres. Furthermore, the selected connections are involved multiple cortical regions and subcortical structures, such as the temporal lobes, calcarine sulcus and occipital lobes, which have been related to ASD in the previous literatures [9].

5. CONCLUSION

In this paper, we propose a novel supervised feature selection approach using local structure learning. To obtain the optimal local structure, we introduce a constraint to the Laplacian matrix. Our extensive experiment results based on two datasets, NYU and UM_1, show that the performance of the proposed method is superior to other existing methods.

6. ACKNOWLEDGEMENT

This work was supported partly by National Natural Science Foundation of China (Nos. 61871274, 61801305 and 81571758), National Natural Science Foundation of Guangdong Province (No. 2017A030313377), Shenzhen Peacock Plan (No.KQTD2016053112051497), and Shenzhen Key Basic Research Project (Nos. JCYJ20170818142347251 and JCYJ20170818094109846).

Table 1 ASD diagnosis performance comparison with different FC construction methods. Boldface denotes best accuracy.

Network	NYU				UM_1			
	116 ROIs	160 ROIs	200 ROIs	Multi-ROIs	116 ROIs	160 ROIs	200 ROIs	Multi-ROIs
PC	71.23±1.20	69.29±4.83	75.30±2.09	76.38±2.34	70.72±6.87	64.49±4.02	62.48±4.28	78.09±3.44
SR	64.89±3.30	64.41±3.40	62.22±3.92	66.46±4.88	66.18±2.33	64.84±5.27	63.66±4.26	69.91±1.25
LR	67.59±2.06	65.84±4.08	60.21±3.29	66.69±3.86	66.57±5.28	64.62±4.84	64.60±6.68	75.52±2.21
SLR	71.33±2.05	68.16±0.34	66.41±2.63	**78.36±2.41**	70.93±1.21	70.77±3.67	68.86±3.36	**81.74±3.84**

Table 2 Algorithm comparison with existing studies. Boldface denotes the best performance.

Method	Network	Classifier	Template	NYU				UM_1			
				Subject	ACC	SEN	SPE	Subject	ACC	SEN	SPE
Wang *et al.* [10]	PC	M3CC	Single	112	76.51	**78.46**	74.69	65	68.40	67.94	68.89
Kam *et al.* [11]	PC	DRBM	Single	184	75.24	61.33	**85.71**	149	80.82	**75.48**	85.00
Dvornek *et al.* [12]	-	RNN+LSTM	Single	1100 (17 Centers)	68.50	-	-	-	-	-	-
Proposed	SLR	Linear SVM	mutiple	171	**78.36**	68.89	85.44	82	**81.74**	71.83	**89.50**

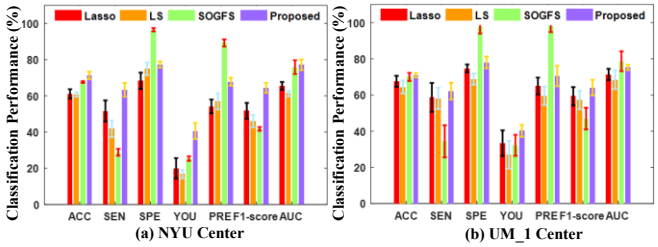

Fig.2. Algorithm comparison via 116 ROIs template.

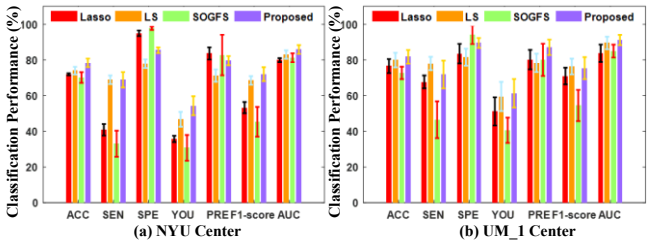

Fig.3. Algorithm comparison via multi-template.

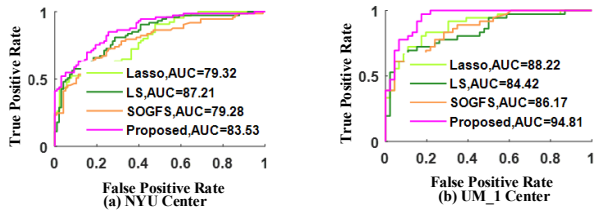

Fig.4. ROC comparison of competing methods using multi-ROIs fused features.

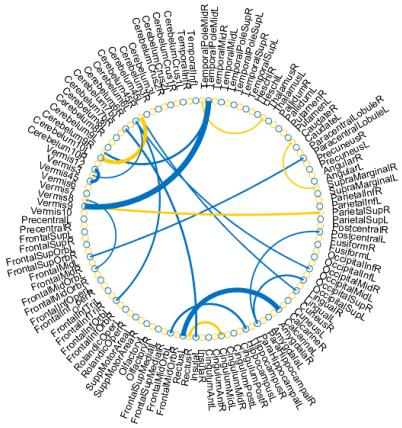

Fig.5. Discriminative features selected by our method on UM_1 dataset.

6. REFERENCES

[1] D. G. Amaral, C. M. Schumann, and C. W. Nordahl, "Neuroanatomy of autism," *Trends in neurosciences,* vol. 31, pp. 137-145, 2008.

[2] F. Shi, L. Wang, Z. Peng, C.-Y. Wee, and D. Shen, "Altered modular organization of structural cortical networks in children with autism," *PLoS One,* vol. 8, p. e63131, 2013.

[3] O. Sporns, *Networks of the Brain*: MIT press, 2010.

[4] L. Qiao, H. Zhang, M. Kim, S. Teng, L. Zhang, and D. Shen, "Estimating functional brain networks by incorporating a modularity prior," *NeuroImage,* vol. 141, pp. 399-407, 2016.

[5] Y. Jin, C. Y. Wee, F. Shi, K. H. Thung, D. Ni, P. T. Yap, *et al.*, "Identification of infants at high‐risk for autism spectrum disorder using multiparameter multiscale white matter connectivity networks," *Human brain mapping,* vol. 36, pp. 4880-4896, 2015.

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological),* pp. 267-288, 1996.

[7] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Advances in neural information processing systems*, 2006, pp. 507-514.

[8] F. Nie, W. Zhu, and X. Li, "Unsupervised Feature Selection with Structured Graph Optimization," In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pp. 1302-1308, 2016.

[9] M. H. Johnson, R. Griffin, G. Csibra, H. Halit, T. Farroni, M. De Haan, *et al.*, "The emergence of the social brain network: Evidence from typical and atypical development," *Development and psychopathology,* vol. 17, pp. 599-619, 2005.

[10] J. Wang, Q. Wang, J. Peng, D. Nie, F. Zhao, M. Kim, *et al.*, "Multi‐task diagnosis for autism spectrum disorders using multi‐modality features: A multi‐center study," *Human brain mapping,* vol. 38, pp. 3081-3097, 2017.

[11] T. E. Kam, H. I. Suk, and S. W. Lee, "Multiple functional networks modeling for autism spectrum disorder diagnosis," *Human brain mapping,* vol. 38, pp. 5804-5821, 2017.

[12] N. C. Dvornek, P. Ventola, K. A. Pelphrey, and J. S. Duncan, "Identifying autism from resting-state fMRI using long short-term memory networks," in *International Workshop on Machine Learning in Medical Imaging*, 2017, pp. 362-370.