Association Study of Alzheimer's Disease with Tree-Guided Sparse Canonical Correlation Analysis

Shangchen Zhou[1,2], Shuai Yuan[1], Zhizhuo Zhang[3], and Zenglin Xu[1(✉)]

[1] SMILE Lab, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, Sichuan, China
shangchenzhou@gmail.com, shuaiyuan0209@gmail.com, zenglin@gmail.com
[2] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, Heilongjiang, China
[3] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
zhizhuo@mit.edu

Abstract. We consider the problem of finding the sparse associations between two sources of data, for example the sparse association between genetic variations (e.g., single nucleotide polymorphisms, SNPs) and phenotypical features (e.g., magnetic resonance imaging, MRI) in the study of Alzheimer's disease (AD). Despite the success of Canonical Correlation Analysis (CCA) based its sparse variants in a number of applications, they usually neglect the underlying natural tree structures SNPs and MRI data. Specifically, the whole candidate set, genes, SNPs of gene form a path of tree structure in SNPs data, and the whole image, regions of image, features of region form a path of tree structure in the MRI data. In order to model the tree structure of features in both sources of data, in this paper, we propose a Tree-guided Sparse Canonical Correlation Analysis (TSCCA). The proposed model equips CCA with special mixed-norm regularization terms in order to model the underlying multilevel tree structures among both the inputs and outputs. To solve the resulted complicated optimization problem, we introduce an efficient iterative algorithm for TSCCA by rewriting tree-structured regularization into the common form of overlapping group lasso. To evaluate the proposed model, we have designed the simulation study and real world study respectively on Alzheimer's disease. Experimental results on the simulation study have shown that the proposed method outperforms CCA with Lasso and group Lasso. The real world study on Alzheimer's disease has shown that our model can find biologically meaningful associations between SNPs and MRI features.

Keywords: Tree-guided Sparse · Canonical correlation analysis
Association study · Alzheimer's disease

L. Cheng et al. (Eds.): ICONIP 2018, LNCS 11307, pp. 585–597, 2018.
https://doi.org/10.1007/978-3-030-04239-4_53

1 Introduction

Many real world problems in machine learning and science discovery amount to finding a sparse and consistent mapping between one source of high dimensional features to another source of output signals. For example, in social media study, one wants to find the association between the objects in images and the labels raised by users [5]. In bioinformatics, one wants to find the association between a selected set of single nucleotide polymorphism and the output genetic expressions, which is known as expression quantitative trait loci (eQTL) mapping [4]. In the study of many complex diseases such as Alzheimer's disease, identifying associations between genetic variations and intermediate phenotypes is crucial [6,7]. In other words, a key step in this task is to discover cross linkages between genetic risk factors based on genomic data−−such as SNPs, and indicative intermediate phenotypes−−such as brain regions abnormalities measured by MRI. This result can help us find out a subset of SNPs which may have functional consequences potentially on brain region structures.

Many approaches have been proposed to solve this problem including canonical correlation analysis [8] and its variants with various sparse priors [2,3,6,7,18]. For example, Parkhomenko et al. [18] applied sparse CCA (SCCA) to find relationships between genetic loci and gene expression levels in Utah families; Witten and Tibshirani [3] used SCCA to reveal associations between gene expression and DNA copy variation; and Chen et al. [2] used structured CCA for pathway selection. In general, by assuming the independence of the output variables, sparse priors such as Lasso [21], elastic net [36], group Lasso or equivalently multiple kernel learning [11,24–28,30–32], overlapping group Lasso [1,9,10], and Bayesian automatic relevance determination [15,17] can be applied, where the non-zero coefficients can be interpreted as the markers truly associated with the output variables. Despite the success in some applications, these sparse priors suffer from various problems, for example, the features selected by Lasso may randomly distribute throughout the whole feature set, but neglect structured information in features. Group Lasso aiming modeling the group structure of features either does not consider overlaps among groups or cause an imbalance among different features due to over penalization on large groups.

However, the structure information underneath data in real world problems can be more complex. For Example, in study of person re-identification, Zhou et al. [35] divide the person image into three parts (head, torso and legs), which can be represented by a tree-structured feature. And in the study of Alzheimer's disease, both the SNPs and the MRI features naturally form a tree structure (as shown in Fig. 1): the whole set of corresponding genes related to the AD study forms the root node whose children are the genes and grandchildren are the SNPs; the whole MRI image forms the root of a tree with different regions (e.g. Left Hippocampus) of the image being children and each feature (e.g., volume of Left Hippocampus) being the grandchildren. In order to model the tree structure of features in both sources of data, in this paper, we propose a Tree-guided Sparse Canonical Component Analysis (TSCCA). The proposed model equips CCA with special mixed-norm regularization terms in order to model the

underlying multilevel tree (i.e., hierarchical) structures among both the inputs and outputs motivated from literature on tree-structured regularization [13,14]. To solve the resulted complicated optimization problem, we introduce an efficient iterative algorithm for TSCCA by rewriting tree-structured regularization into the common form of overlapping group lasso. To evaluate the proposed model, we have designed a simulation study and a real world study on Alzheimer's disease. Experimental results on the simulation study have shown that the proposed method outperforms CCA with Lasso and CCA with group Lasso. The real world study on Alzheimer's disease has shown that our model can discover meaningful correlations between SNPs and MRI features.

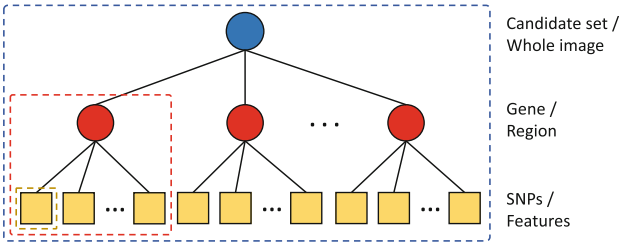

Fig. 1. The illustration the tree structures of SNPs and MRI data. The whole feature set of genes and MRI features are considered as foot nodes of the two trees, respectively; the internal nodes are groups of features, which are grouped based on corresponding genes and regions, respectively; The individual features of SNPs and MRI are considered as leaf nodes.

2 Methods

Before presenting the model, we first introduce some basic notations. In our paper, there are two sources of data (e.g., SNPs and MRI), denoted by $X = [x_1, \ldots, x_n]$ and $Y = [y_1, \ldots, y_n]$, respectively. Here both X and Y contain n subjects, and described by p-dimensional and q-dimensional feature vectors, respectively.

2.1 Model

Sparse Canonical Correlation Analysis. Given the two data sources X and Y, the variance matrices of data X, Y are denoted by Σ_{XX} and Σ_{YY}, respectively. And the covariance matrix between X and Y is denoted by Σ_{XY} or Σ_{YX}. The CCA method aims to find two projection vectors α and β, i.e., the linear combinations of two data variables, which maximize the correlation

between these two linear combinations (i.e., $\alpha^\top X$ and $\beta^\top Y$), as shown in the Eq. (1):

$$\begin{aligned} &\max_{\alpha,\beta} \ \alpha^\top \Sigma_{XY} \beta \\ &\text{s. t. } \ \alpha^\top \Sigma_{XY} \alpha = 1, \ \beta^\top \Sigma_{XY} \beta = 1. \end{aligned} \tag{1}$$

In practical applications, the amount of samples may be much smaller than the high dimensions of data (i.e., $p, q \gg n$). To address this issue, sparse methods such as Lasso [21], elastic net [36], and group Lasso [32], can be imposed on the projection vectors as the penalty function, as shown in the Eq. (2).

$$\begin{aligned} &\min_{\alpha,\beta} \ -\alpha^\top \Sigma_{XY} \beta + \lambda_1 \Psi(\alpha) + \lambda_2 \Phi(\beta) \\ &\text{s. t. } \ \alpha^\top \Sigma_{XX} \alpha \leq 1, \ \beta^\top \Sigma_{YY} \beta \leq 1, \end{aligned} \tag{2}$$

where $\Psi(\cdot)$ and $\Phi(\cdot)$ denote the penalty function on α and β, respectively.

Tree-Structure Modelling. As shown in Fig. 1, the features of SNPs and MRI data form a tree structure, respectively. In order to select important features in the hierarchical tree structures, we introduce the tree-structured lasso into the CCA model. In tree structure, we observe that the overlapping between nodes may exist: any child node overlaps with its parent nodes, actually, the child node's feature set belongs to that of its parent node; the nodes from the same depth in the tree do not overlap.

Let T be an index tree with a depth d, and let $T_i = \{G_1^i, G_2^i, \ldots, G_{n_i}^i\}$ denote the set of nodes corresponding to the depth i, where $n_0 = 1$, $G_1^0 = \{1, 2, \ldots, p\}$ and $n_i \geq 1$, $i = 1, 2, \ldots, d$. Then this set of nodes satisfies the following conditions: (1) the nodes from the same tree depth have non-overlapping features, i.e., $G_j^i \cap G_k^i = \emptyset$, $\forall_i \in \{1, 2, \ldots, d\}$, $j \neq k, 1 \leq j, k \leq n_i$; and (2) If $G_{j_0}^{i-1}$ is the parent node of a non-root node G_j^i, then $G_j^i \subseteq G_{j_0}^{i-1}$.

Given the tree structure, the tree-structured regularization can be defined as:

$$\Phi(x) = \sum_{i=0}^{d} \sum_{j=1}^{n_i} \omega_j^i \| x_{G_j^i} \|_2 + \frac{h}{2} x^\top x, \tag{3}$$

where $x \in \mathbb{R}^p$, $\omega_j^i \geq 0$ $(i = 0, 1, \ldots, d, j = 1, 2, \ldots, n_i)$ denotes the pre-defined weight for the node G_j^i, $\|\cdot\|$ denotes the Euclidean norm, h is a tuning parameter, and $x_{G_j^i}$ is a vector composed of the feature of x with indices in G_j^i. To address the problem of collinearity, the ridge penalty $\frac{h}{2} x^\top x$ is introduced in Eq. (3). Tree-guided group Lasso can be seen as a special case of the overlapping group Lasso [13], which models the overlapping groups through a tree structure.

Each group of coefficients $x_{G_j^i}$ in Eq. (3) is weighted with ω_j^i. For convenience of subsequent discussion, we denote $\omega_v = \omega_j^i$, where v is the jth node of the depth i. Existing overlapping group Lasso methods may bring an imbalanced penalty among different features. To address this downside, we follow a weight-balanced

scheme. We define ω_v in terms of two quantities g_v and s_v, where $g_v + s_v = 1$. The s_v is the weight for selecting the variables according to each of the children node v separately, and g_v is the weight to select these nodes jointly. Now, for a given tree T, we apply the Eq. (4) operation recursively, from the root node towards the leaf nodes.

$$\sum_{i=0}^{d_2} \sum_{j}^{m_i} \omega_j^i \|\beta_{G_j^i}\|_2 = W(v_{root}), \tag{4}$$

where

$$W(v) = \begin{cases} s_v \cdot \sum\limits_{c \in \text{Children}(v)} |W(c)| + g_v \cdot \|\beta_{G_v}\|_2, & (a) \\ |\beta_{G_v}|, & (b) \end{cases}$$

In the above equation, (a) represents that v is an internal node, and (b) represents that v represents the leaf node.

It is shown the relationship in Eq. (5) holds between ω_v's and (s_v, g_v)'s [13].

$$\omega_v = \begin{cases} g_v \prod\limits_{m \in \text{Ancestors}(v)} s_m, \\ \prod\limits_{m \in \text{Ancestors}(v)} s_m, \end{cases} \tag{5}$$

Actually, the above weighting scheme extends the elastic-net-like penalty hierarchically, which result in a balance of our model. It is easy to verify that the following proposition holds [13]:

Proposition 1. *For feature k, the sum of the weights ω_v for the nodes $v \in T$ whose feature group G_v contains the feature k equals one, i.e., as following holds:*

$$\sum_{v:k \in G_v} \omega_v = \prod_{m \in \text{Ancestors}(v_{leaf})} s_m + \sum_{l \in Ancestors(v_{leaf})} g_l \prod_{m \in \text{Ancestors}(l)} s_m = 1.$$

Tree-Guided Sparse CCA. In this study, since the features can be naturally be represented using certain tree structures as shown in Fig. 1, we incorporate the tree-guided sparse penalty on the both source of features into the design of CCA models. This results in a tree-guided sparse CCA (TSCCA), as shown in the following equation:

$$\begin{aligned} \min_{\alpha,\beta} \ & -\alpha^\top \Sigma_{XY} \beta + \lambda_1 \Phi_1(\alpha) + \lambda_2 \Phi_2(\beta) \\ \text{s. t.} \ & \alpha^\top \Sigma_{XY} \alpha \le 1, \ \beta^\top \Sigma_{XY} \beta \le 1, \end{aligned} \tag{6}$$

where $\Phi_1(\alpha) = \sum\limits_{i=0}^{d_1} \sum\limits_{j}^{n_i} \psi_j^i \|\alpha_{G_j^i}\|_2 + \frac{h_1}{2} \alpha^\top \alpha$ and $\Phi_2(\beta) = \sum\limits_{i=0}^{d_2} \sum\limits_{j}^{m_i} \omega_j^i \|\beta_{G_j^i}\|_2 + \frac{h_2}{2} \beta^\top \beta$.

2.2 Optimization Algorithm

The main difficulty in solving Eq. (6) grows from the tree-structured regularization. Motivated by [2], we transform the tree-structured regularization into the common form of the overlapping group lasso. Let the nodes of the index tree be numbered from left to right and from top to bottom (e.g. $G_1^0 = 1, G_1^1 = 2, ..., G_j^i = k, ..., G_{n_d}^d = l$). Then the tree-structured regularization can be rewritten as:

$$\Phi(x) = \sum_{k=1}^{l} \omega_k \|x_k\|_2 + \frac{h}{2} x^\top x. \tag{7}$$

For illustrating easily, we assume that $\Phi_1(\alpha) = \|\alpha\|_1$. Thus, we get the following equation:

$$\begin{aligned} &\min_{\alpha,\beta} \ -\alpha^\top \Sigma_{XY} \beta + \lambda \sum_{k=1}^{l} \omega_k \|\beta_k\|_2 + \frac{h}{2} \beta^\top \beta \\ &\text{s. t. } \ \alpha^\top \Sigma_{XY} \alpha \le 1, \ \beta^\top \Sigma_{XY} \beta \le 1, \ \|\alpha\|_1 \le 1. \end{aligned} \tag{8}$$

Let the β domain be denoted as $Q_1 = \{\beta | \|\beta\|_2 \le 1\}$, $v = \frac{1}{\tau} Y^\top X \alpha$ and $\gamma = \frac{\lambda}{\tau}$. Then the optimization of Eq. (6) respecting β can be written as:

$$\min_{\beta \in Q_1} f(\beta) \equiv l(\beta) + \Phi(\beta), \tag{9}$$

where $l(\beta) = \frac{1}{2}\|\beta - v\|_2$ is the Euclidean distance loss function and $\Phi(\beta)$ is the rewritten tree-structured regularization. According to the Theorems 1–3 in [2] and the Fenchel duality theorem, the optimization problems in the form like Eq. (9) can be solved by the Excessive Gap Method. The details of proof and calculation process can be found in [2].

3 Results

3.1 Simulation Study

At first, we design a simulation study to examine TSCCA to estimate the accuracy on finding cross linkages between the two source of data, to investigate whether TSCCA can improve the detection power compared to the other sparse model.

Simulation Data. For generating the ground truth, we simulate two data sets X and Y, consisting of $p = 400$ and $q = 300$ variables (features), which the index tree structures in α and β are given as a priori. Data X and Y are divided into $G_X = 20$ and $G_Y = 10$ groups respectively as internal nodes. For simplicity, the feature size of internal nodes are the same in both X and Y. We set the depth to be 3, i.e., $d = 3$. Each individual feature in X and Y is a leaf node, and the whole feature set in both X and Y is considered as the foot node. We set the sample size n as 200. To append the potential correlation between

variables in X and variables in Y, a latent variable is set $\Upsilon = \{\gamma_i | i = 1, \ldots, n\}$ with a Gaussian distribution $N(0, \sigma_\gamma^2)$ and normalize all the variables to the unit length, which have the similar effect on the correlated variables from two source of data. We generate X and Y with each sample $x_i \sim N(\gamma_i \alpha, \sigma_e^2 \Sigma_{XX})$, and $y_i \sim N(\gamma_i \beta, \sigma_e^2 \Sigma_{YY})$, where $\alpha = [\alpha_1, \ldots, \alpha_p]$, $\beta = [\beta_1, \ldots, \beta_q]$ are the projection vectors of X and Y, respectively, where $\alpha_j \neq 0$, $\beta_k \neq 0$, if x_j, y_k are the correlated variables, σ_e^2 is the variance of noise variable, and Σ_{XX} and Σ_{YY} are the variance-covariance matrices of X, Y data which are used to simulate the tree-structured group effect within each dataset. We set $\Sigma_{ij} = \rho^{|i-j|}$, while variables i and j are correlated variables, i.e., they are included in the same group, where ρ is preferred to be 0.5 referring to [16,32].

For performance evaluation on the sparsity, we adopt the F1-measure to evaluating different models. We compare the proposed model with current state-of-the-art sparse association study methods, including CCA with lasso sparse penalty (denoted by CCA-Lasso) and CCA with Group Lasso penalty (denoted by CCA-Group Lasso).

In the following, we first compare the recovering results of correlated variables from two source of date among different sparse CCA methods.

Figures 2 and 3 show the results of recovered projection vectors α and β by CCA-Lasso, CCA-Group Lasso and TSCCA methods, respectively. It shows that TSCCA can estimate the α and β more accurate than the other two sparse models. The CCA-Lasso misses out some true variables, and CCA-Group Lasso selects more noise variables. The TSCCA not only can recover the tree structure between two source of date accurately but also can distinguish noise variables.

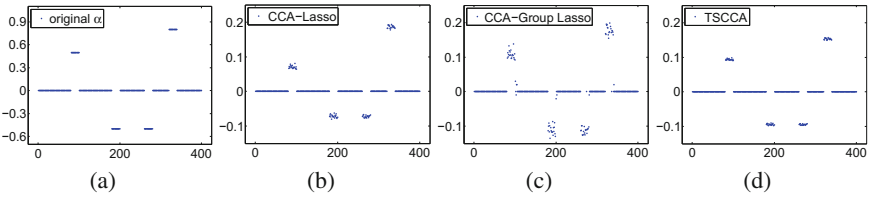

Fig. 2. Performance comparison in recovering α of three sparse CCA models. (a) Ground Truth α; (b–d) α recovered by CCA-Lasso, CCA-Group Lasso and TSCCA.

Secondly, we discuss the how the sample size of data effects the recovery performance. We increase the sample size n from 50 to 250, with step of 50, comparing the performance of these sparse models. Figure 4(a) presents the F1 score with different sample size. It shows that, with the sample size increasing, TSCCA keeps the best F1-measure among these methods.

Finally, we discuss how noise in data affects the model performance. To compare these models under different noise levels, we fixed other conditions but adjust the standard deviation σ_e, starting from 0.1 to 1 with step 0.1, manipulating the relation coefficient between two datasets. From Fig. 4(b), we can see that TSCCA is more robust than other methods whenever the noise level increase.

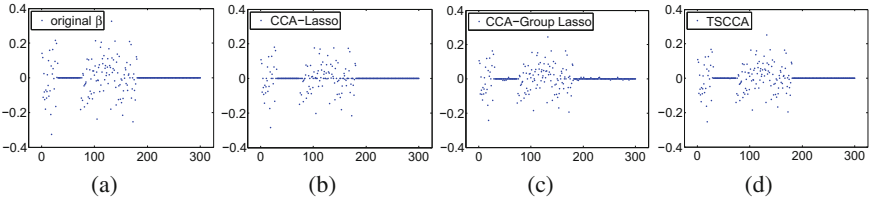

Fig. 3. Performance comparison in recovering β of three sparse CCA models. (a) Ground Truth β; (b–d) β recovered by CCA-Lasso, CCA- Group Lasso and TSCCA.

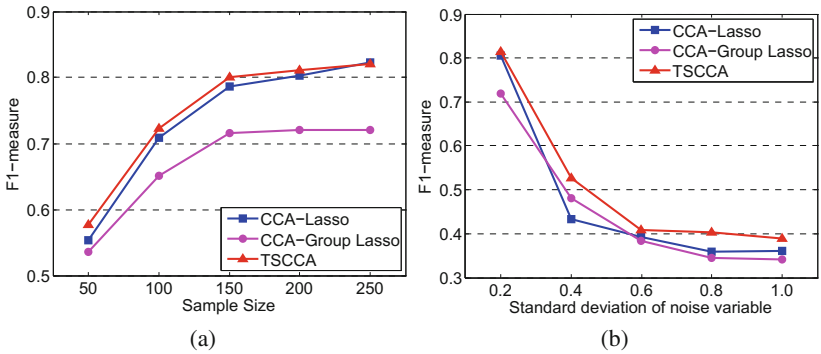

Fig. 4. A comparison of three SCCA models in different situations. (a) F1-measures for three methods when varying sample sizes from 50 to 250. (b) F1-measures for three methods when varying the standard deviation of noises.

3.2 AD Study

We conduct association analysis study based on a dataset called Alzheimer's Disease Neuroimaging Initiative (ADNI). The ADNI study is a longitudinal multi-site observational study of the elderly individuals with different cognition conditions (i.e., normal cognition, mild cognitive impairment (MCI) or AD). We applied our model TSCCA to discover the correlation between genetic variations and brain regions atrophy measured by MRI.

The ADNI dataset is available on adni.loni.ucla.edu. After removing samples with missing values, it contains 618 samples (182 normal, 302 MCI and 134 AD) and each sample includes 924 SNPs (selected as the top SNPs with high discriminative power to separate normal subjects from MCI and AD) and 328 MRI features (measuring the brain atrophies in different regions about surface area, volume or cortical thickness with FreeSurfer software). Based on the feature belongs to different brain regions, we divided the 328 MRI features into 118 groups, which are considered as internal nodes. We first obtain the gene that each SNP belongs to by analyzing the SNPs information downloaded from www.ncbi.nlm.nih.gov/projects/SNP/dbSNP.cgi?list=rslist, then we divided the 937 SNPs into 529 groups based on the SNP belongs to different genes, which are also considered as internal nodes. Each individual feature of SNPs and MRI is

considered as leaf node, and the whole feature set of both SNPs and MRI is considered as foot node, hence SNPs and MRI data naturally be represented using two certain tree structures with depth three.

We apply our proposed method TSCCA on association discovery of Alzheimer's Disease using SNPs and MRI features. Lastly, the heatmap Fig. 5 shows the biclustering of the SNPs-MRI cross linkages, witch reveals meaningful correlations between genetic variations and brain atrophy. For example, the top ranked 19 SNPs are included in a few genes, such as PVRL2 (rs8105340), TOMM40 (rs2075650), HK2 (rs3771773), MAGI2 (rs508990), of which some have been studied more carefully in AD (www.alzgene.org); In selected 22 MRI features, Fusiform, Middle Temporal, and Hippocampus play a important role in formating long-term memory.

4 Discussion

Since the dimensions of SNPs and MRI features are far larger than sample size in ADNI dataset, the standard CCA models have the problem of overfitting, which cannot be applied directly. To handle this problem, many methods impose an L1−norm on α and β as a penalty function to shrink the coefficients of the irrelevant variables toward zero [19,22,23]. Corresponding models are referred to sparse CCA. However, sparse learning with L1−norm is limited which neglects the latent rich structural information among data. Actually, data structure as prior information is crucial to improve model performance and interpret ability, especially when leaning from high-dimensional data.

Recently, To take advantage of the structure prior knowledge of data, various extensions of $L1-$norm penalty have been proposed, such as the elastic net [36], the Group Lasso penalty [32], overlapping Group Lasso [2], as well as the mixed-norm tree-structured penalty [12–14]. However, the tree-structured penalty has not been incorporated into the CCA framework for AD study. The main challenge arises from the computational side. For solving the group-sparsity regularized optimization problem, [14] showed the Moreau-Yosida regularization associated with the tree structured group Lasso admits an analytical solution, and design efficient algorithm for solving the problem for smooth convex loss functions; and Kim [13] adopted the smoothing proximal gradient approach to solve the problem.

For associations discovery study, [29,33,34] presented Bayesian multiview learning models for joint associations discovery and disease prediction in the AD study. Different from the proposed unsupervised association study, those methods also require the given disease status as an input. It is also important to note that there are some approaches aim to make diagnosis based on MRI imaging data [20], while the focus of this paper is to find the associations.

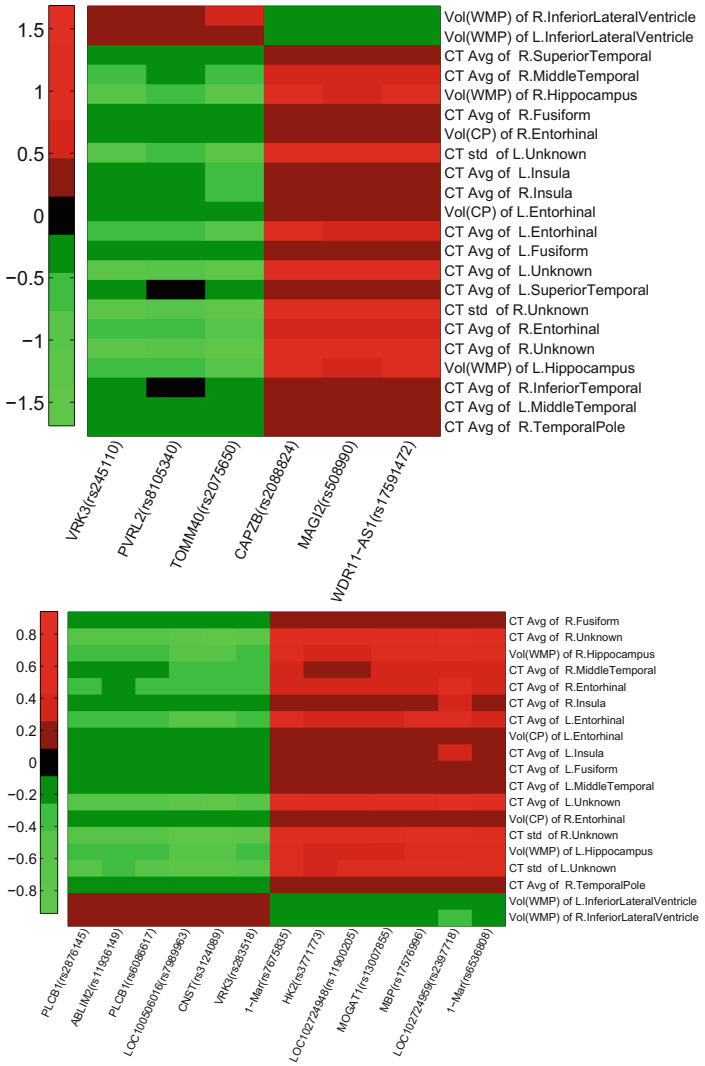

Fig. 5. The heatmap estimated by TSCCA, which shows the associations between SNPs and MRI features. Where, the SNP names are given at the bottom and the MRI features are listed on the right.

5 Conclusions

In this paper, we presented, TSCCA, a tree-guided sparse CCA model for associations discovery in the Alzheimer's disease study. We also develop an efficient optimization algorithm for the proposed model. Our experimental results on the simulation study have demonstrated that the proposed method can discover more rich structure information. Experimental results on the real world AD study has

shown that the proposed method can output biologically meaningful results and we plan to apply our model to a wide range of applications in bioinformatics and computer vision involving the tree structure.

Acknowledgements. Datasets used in this paper are obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (ADNI official website: adni.loni.ucla.edu). The investigators who contributed to the design and implementation of ADNI and/or collected data can be found on ADNI official website.

This work was in part supported by grants of NSF China (No. 61572111), a 985 Project of UESTC (No. A1098531023601041) and a Fundamental Research Project of China Central Universities (No. ZYGX2016Z003).

References

1. Chen, J., Bushman, F.D., Lewis, J.D., Wu, G.D., Li, H.: Structure-constrained sparse canonical correlation analysis with an application to microbiome data analysis. Biostatistics **14**(2), 244–258 (2013)
2. Chen, X., Liu, H., Carbonell, J.G.: Structured sparse canonical correlation analysis. In: International Conference on Artificial Intelligence and Statistics, pp. 199–207 (2012)
3. Daniela, M., Tibshirani, R.: Extensions of sparse canonical correlation analysis, with applications to genomic data. Stat. Appl. Genet. Mol. Biol. **383**(1), 1–27 (2009)
4. Du, L., et al.: Pattern discovery in brain imaging genetics via scca modeling with a generic non-convex penalty. Sci. Rep. **7**(1), 14052 (2017)
5. Eisenschtat, A., Wolf, L.: Linking image and text with 2-way nets. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
6. Hao, X., et al.: Mining outcome-relevant brain imaging genetic associations via three-way sparse canonical correlation analysis in alzheimer's disease. Sci. Rep. **7**, 44272 (2017)
7. Hao, X., et al.: Identification of associations between genotypes and longitudinal phenotypes via temporally-constrained group sparse canonical correlation analysis. Bioinformatics **33**(14), i341–i349 (2017)
8. Hotelling, H.: Relations between two sets of variates. Biometrika **28**, 321–377 (1936)
9. Jacob, L., Obozinski, G., Vert, J.P.: Group lasso with overlap and graph lasso. In: Proceedings of the 26th Annual international Conference on Machine Learning, pp. 433–440. ACM (2009)
10. Jenatton, R., Audibert, J.Y., Bach, F.: Structured variable selection with sparsity-inducing norms. J. Mach. Learn. Res. **12**, 2777–2824 (2011)
11. Kang, Z., Lu, X., Yi, J., Xu, Z.: Self-weighted multiple kernel learning for graph-based clustering and semi-supervised classification. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI), pp. 2312–2318 (2018)
12. Kim, S., Xing, E.P.: Tree-guided group lasso for multi-task regression with structured sparsity. In: Proceedings of the 27th International Conference on Machine Learning (ICML 2010), pp. 543–550 (2010)
13. Kim, S., Xing, E.P., et al.: Tree-guided group lasso for multi-response regression with structured sparsity, with an application to eqtl mapping. Ann. Appl. Stat. **6**(3), 1095–1117 (2012)

14. Liu, J., Ye, J.: Moreau-yosida regularization for grouped tree structure learning. In: Advances in Neural Information Processing Systems, pp. 1459–1467 (2010)
15. MacKay, D.J.: Bayesian interpolation. Neural Comput. **4**(3), 415–447 (1991)
16. Meier, L., Van De Geer, S., Buhlmann, P.: The group lasso for logistic regression. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **70**(1), 53–71 (2008)
17. Neal, R.M.: Bayesian Learning for Neural Networks, vol. 118, p. 118. Springer Science & Business Media, New York (1996)
18. Parkhomenko, E., Tritchler, D., Beyene, J.: Genome-wide sparse canonical correlation of gene expression with genotypes. BMC Proc. **1**(Suppl. 1), S119 (2007)
19. Parkhomenko, E., Tritchler, D., Beyene, J.: Sparse canonical correlation analysis with application to genomic data integration. Stat. Appl. Genet. Mol. Biol. **8**(1), 1–34 (2009)
20. Que, X., Ren, Y., Zhou, J., Xu, Z.: Regularized multi-source matrix factorization for diagnosis of Alzheimer's disease. In: Neural Information Processing - 24th International Conference, ICONIP, pp. 463–473 (2017)
21. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B (Methodol.) **58**, 267–288 (1994)
22. Witten, D.M., Tibshirani, R., Hastie, T.: A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. Biostatistics p. kxp008 (2009)
23. Witten, D.M., Tibshirani, R.J.: Extensions of sparse canonical correlation analysis with applications to genomic data. Stat. Appl. Genet. Mol. Biol. **8**(1), 1–27 (2009)
24. Xu, Z., Jin, R., King, I., Lyu, M.R.: An extended level method for efficient multiple kernel learning. In: Advances in Neural Information Processing Systems 21, Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems (NIPS), pp. 1825–1832 (2008)
25. Xu, Z., Jin, R., Yang, H., King, I., Lyu, M.R.: Simple and efficient multiple kernel learning by group lasso. In: Proceedings of the 27th International Conference on Machine Learning (ICML 2010), pp. 1175–1182 (2010)
26. Xu, Z., Jin, R., Ye, J., Lyu, M.R., King, I.: Non-monotonic feature selection. In: Proceedings of the 26th Annual International Conference on Machine Learning (ICML), pp. 1145–1152 (2009)
27. Xu, Z., Jin, R., Zhu, S., Lyu, M.R., King, I.: Smooth optimization for effective multiple kernel learning. In: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (2010)
28. Xu, Z., King, I., Lyu, M.R., Jin, R.: Discriminative semi-supervised feature selection via manifold regularization. IEEE Trans. Neural Networks **21**(7), 1033–1047 (2010)
29. Xu, Z., Zhe, S., Qi, Y., Yu, P.: Association discovery and diagnosis of alzheimer's disease with bayesian multiview learning. J. Artif. Intell. Res. **56**, 247–268 (2016)
30. Yang, H., Xu, Z., King, I., Lyu, M.R.: Online learning for group lasso. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 1191–1198 (2010)
31. Yang, H., Xu, Z., Lyu, M.R., King, I.: Budget constrained non-monotonic feature selection. Neural Networks **71**, 214–224 (2015)
32. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **68**(1), 49–67 (2006)
33. Zhe, S., Xu, Z., Qi, Y., Yu, P.: Sparse bayesian multiview learning for simultaneous association discovery and diagnosis of alzheimer's disease. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, pp. 1966–1972 (2015)

34. Zhe, S., Xu, Z., Qi, Y., Yu, P., et al.: Joint association discovery and diagnosis of alzheimer's disease by supervised heterogeneous multiview learning. In: Pacific Symposium on Biocomputing, vol. 19. World Scientific (2014)
35. Zhou, S., Yao, H., Yu, W., Wang, Y.: Tree-guided group sparse based representation for person re-identification. In: Proceedings of the International Conference on Internet Multimedia Computing and Service, pp. 14–17. ACM (2016)
36. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **67**(2), 301–320 (2005)