Spatial and Anatomical Regularization Based on Multiple Kernel Learning for Neuroimaging Classification

Yingliang Wu, Student Member and Benyong Liu, Nonmember

SUMMARY Recently, a high dimensional classification framework has been proposed to introduce spatial and anatomical priors in classical single kernel support vector machine optimization scheme, wherein the sequential minimal optimization (SMO) training algorithm is adopted, for brain image analysis. However, to satisfy the optimization conditions required in the single kernel case, it is unreasonably assumed that the spatial regularization parameter is equal to the anatomical one. In this letter, this approach is improved by combining SMO algorithm with multiple kernel learning to avoid that assumption and optimally estimate two parameters. The improvement is comparably demonstrated by experimental results on classification of Alzheimer patients and elderly controls.

key words: neuroimaging, spatial regularization, anatomical regularization, multiple kernel learning

1. Introduction

Alzheimer disease (AD) is the most prevalent neurodegenerative dementia in worldwide, and thus prevention and early accurate diagnosis of AD is increasingly crucial. In the last few years, support vector machine (SVM) methods for AD subject classification have become an incredibly active research topic. In the related work, SVM approach, as same as feature selection and extraction, has been taking the specificity of neuroimaging data into account. As an example, a frame work is proposed in [1] to include spatial and anatomical priors into SVM by using regularization operators, and it indicates a flexible way to model various types of proximity. However, the classical kernel-based learning algorithms are based on a single kernel, and for the purpose of satisfying the solving conditions of SVM optimization problem in the 3D discrete case, in the above method, parameter of spatial regularization and that of anatomical regularization is unreasonably assumed to be equal so that the two regularization items can be integrated into a single framework. To improve the method, we proposed in this manuscript to combine sequential minimal optimization (SMO) algorithm with multiple kernel learning (MKL) to solve the optimization problem of regularization parameter estimation. To illustrate the improvement, we apply the proposed method to classification of 299 subjects, 137 AD patients and 162 cognitively normal (CN) controls, as the same study population in [2]. It is observed that the new classifier can significantly improve classification accuracy.

2. Priors in SVM

SVM is a supervised learning algorithm introduced by Vapnik and performs classification by mapping the data into higher dimensional space [3]. Let \((x_s, y_s), s \in 1, \ldots, N\) be a training set of instance-labeled pairs, \(x_s \in \Omega, y_s \in \{-1, 1\}\) and \(\Omega\) is the input space, then SVM optimization problem can be written as:

\[
(w_{opt}, b_{opt}) = \arg \min_{w \in \mathcal{H}, b \in \mathbb{R}} \frac{1}{N} \sum_{s=1}^{N} l_{hinge}(y_s (w, x_s) + b) + \lambda \|w\|_2^2.
\]

where \(\lambda \in \mathbb{R}^+\) denotes the regularization parameter and \(l_{hinge}\) denotes the hinge loss function.

If \(L\) is a graph Laplacian regularization and \(\beta\) is a parameter to control the size of the regularization, the optimization problem may be rewritten as:

\[
(w_{opt}, b_{opt}) = \arg \min_{w \in \mathcal{H}, b \in \mathbb{R}} \frac{1}{N} \sum_{s=1}^{N} l_{hinge}(y_s (w, x_s) + b) + \lambda \|e_{\beta L} w\|_2^2
\]

Meanwhile, the heat kernel is obtained as \(k_\beta(x_1, x_2) = e^{-\frac{d(x_1, x_2)}{2\beta}}\).

In the case of 3D discrete brain images, using image connectivity (6, 18, or 26) is the simplest way to define spatial proximity. Let \(L_a\) be the Laplacian matrix of the graph including the spatial proximity, it is diagonalizable to diagonal matrix \(S\) by a symmetric orthogonal matrix \(Q\) which can be constructed from discrete sine transform (DST) matrix. Therefore, the exponential matrix is computed as:

\[
e^{-\beta L_a} = e^{-\beta Q S Q} = Q e^{-\beta S} Q
\]

Similarly, anatomical proximity can be defined by the probability that each voxel of a 3D brain image is connected to another, and two voxels are regarded close if they belong to the same structural or functional region. Let \(L_m\) be a Laplacian matrix of the graph including the anatomical proximity, then the normalized Laplacian matrix \(L_m\) is obtained as \(L_m = I_d - D^{-\frac{1}{2}} E C E^T D^{-\frac{1}{2}}\), where \(d\) is the number of voxels in 3D brain image, \(C\) is probability matrix which two Regions of Interest (ROI) are connected, \(E\) is adjacency matrix, and \(D\) is diagonal matrix which each diagonal element is equal to the sum of corresponding row of matrix \(E\).
A real symmetric matrix $\tilde{E}^T \tilde{E}$, where $\tilde{E} = D^{-\frac{1}{2}} EC^2$, can be diagonalized to diagonal matrix $\Lambda$ by a symmetric orthogonal matrix $X$. Therefore, the exponential matrix is computed as follows:

$$e^{-\beta_s L_s} = X e^{-\beta_s \Lambda} X^T + e^{-\beta_s \{I_d - \tilde{X} X^T\}},$$

where $\tilde{X}_r = \frac{1}{\Lambda_{rr}} \tilde{E} X$ is the $r$-th column of the matrix $\tilde{X}$, $\Lambda_{rr} = 1 - \Lambda_{rr}$, $r = 1, \ldots, k$ and the sole nonzero entries of $\Lambda$, without loss of generality, are its first $k$ diagonal entries.

For the purpose of introducing anatomical and spatial priors into SVM, we should combine the two regularization terms and solve the following optimization problem:

$$(w^{opt}, b^{opt}) = \arg \min_{w \in \Omega, b \in \mathbb{R}} \frac{1}{N} \sum_{i=1}^{N} \max\{y_i \langle w, x_i \rangle + b \} + \lambda_s \|e^{\frac{\beta_s}{2} L_s w}\|^2 + \lambda_d \|e^{\frac{\beta_d}{2} L_d w}\|^2,$$

where $\lambda_s$ is the spatial regularization parameter and $\beta_s$ is the anatomical regularization parameter, $\beta_s$ and $\beta_d$ are corresponding diffusion parameters.

3. SMO-MKL Based Method

The classical kernel-based learning algorithms are based on a single kernel. Therefore, during the solution of formula (5), in order to integrate spatial regularization and anatomical regularization to meet the solving conditions of standard SVM optimization problem, the previous method assumed that regularization parameter of spatial regularization is equal to that of anatomical regularization, that is, $\Lambda_s = \Lambda_d$. However, this assumption is unreasonable because $\Lambda_s$ is not necessarily the optimal regularization parameter values.

The aim of this paper is to find a method, based on SMO-MKL algorithm, to solve this problem [4].

Based on the SMO algorithm, our SMO-MKL algorithm focuses on training $p$-norm multi-kernel learning. Let $\{K_i\}$ be a set of base kernels, and $\{\phi_k\}$ is corresponding feature maps, then the primal can be written as follows:

$$\min_{w, b, \xi \geq 0, d \geq 0} \frac{1}{2} \sum_k w_k^T w_k + C \sum_i \xi_i + \frac{1}{2} \sum_k \left(\frac{d_k}{\xi_i}\right)^\gamma$$

s.t. $y_i \left(\sum_k \sqrt{d_k} w_k \phi_k(x_i) + b\right) \geq 1 - \xi_i$

Using Lagrangian multiplier method repeatedly and differentiating with respect to $w$, $b$, $\xi$ and $d$ respectively can result in the $l_p$-MKL dual problem as $D = \max_{\alpha \in \Lambda} 1^T \alpha - \frac{1}{8\lambda} \left(\sum_i (\alpha^T H_{k} \alpha)^\gamma\right)^\frac{1}{\gamma}$, where $p^{-1} + q^{-1} = 1$, $A = \{a | 0 \leq a \leq C, 1^T Y \alpha = 0\}$, $H_k = Y K_k Y$, and $Y$ is a diagonal matrix which each diagonal element is a label. Therefore, the kernel weights can be computed as following:

$$d_k = \frac{1}{2\lambda} (\alpha^T H_{k} \alpha)^{\frac{1}{\gamma}} \left(\sum_k (\alpha^T H_{k} \alpha)^{\gamma}\right)^{-\frac{1}{\gamma}}$$

Finally, spatial regularization kernel matrix and anatomical regularization kernel matrix can be computed using (3) and (4). Consequently, optimizing the $l_\infty$-MKL dual problem using SMO-MKL algorithm, the optimal regularization parameters, $\lambda_s$ and $\lambda_d$, can be solved according to (7) by such kernel matrices. Whether in the process of computing kernel matrices or solving optimal regularization parameters, searching various optimal parameters is the most critical point. Therefore, in consideration of these kernel matrices are pre-computed matrices, the kernel matrix-related code in LIBSVM [5] and SMO-MKL should be adapted.

4. Experiment

The proposed method is applied to classification of 299 subjects, 137 AD patients and 162 CN controls, as the same study population in [2]. Only T1-weighted, acquired at 1.5 T and pre-processed magnetic resonance (MR) images are selected and entire study on the Alzheimer Disease Neuroimaging Initiative (ADNI) public database (www.loni.ucla.edu/ADNI).

The Statistical Parametric Mapping (SPM) is used to segment the selected images into gray matter (GM), white matter (WM) and Cerebrospinal Fluid (CSF) [6], and the Diffeomorphic Anatomical Registration through Exponentiated Lie Algebra (DARTEL) algorithm is used for spatial normalization [7]. The key process are applying SPM8 package to structural brain image segmentation, utilizing DARTEL algorithm to create GM/WM average-shaped template and normalize the segmented image to Montreal Neurological Institute (MNI) Space respectively. It is worth noticing that all GM probability maps in the MNI space should be modulated to ensure that the overall tissue amount maintains constant.

For better evaluation of the proposed approach, all experimental data and the whole process should be as close to the previous method as possible, except the solution of optimal regularization parameters in the last step. Whereas 9 subjects, 6 AD patients and 3 CN controls, are failed in preprocessing as described in detail in [2]. Those subjects are excluded if they belong to the training set, and 50 percent misclassified if they belong to the testing set. As a consequence, the set of participants is randomly split, for 10 times, into two groups of the same size, 145 training and 145 testing, based on the Mersenne Twister algorithm.

A grid-search and 10 cross validation on the training set is carried out to obtain optimal parameter values. According to the evaluation of [1], diffusion parameter $\beta_s = \beta_d = \beta$. Diffusion parameter $\beta$ would be set at 1.25, 1.5, 1.75, 2, 4 and 6 because the optimal value is around 1.5 according to our experiment. Table 1 shows the results of one random experiment with different $\beta$. Contrast to a single anatomical regularization kernel taking a nosedive accuracy (Acc.) after diffusion parameter $\beta > 2$, a single spatial regularization kernel maintains a high level with the same parameter. Therefore, combining the two regularization terms based on SMO-MKL still produce better performance by assigning a higher weight to spatial regularization kernel. The sensitivity (Sens.) related to AD patients and specificity (Spec.)
related to control subjects is also quantified.

As shown in Table 2, a single spatial regularization kernel (Voxel-Regual-spacial) with a peak accuracy of 89%, a single anatomical regularization kernel (Voxel-Regual-atlas) yielding a peak accuracy of 90%, but combining the two regularization terms (Voxel-Regual-combinegraph) as the previous method only reaches a peak accuracy of 87% [1]. This result indicates that it is unreasonable to assume $\lambda_s = \lambda_a$ and they are not the optimal regularization parameters values. Finally, it can be observed that the proposed method based on SMO-MKL ranks first for all classification types.

### 5. Conclusion

A novel method based on SMO-MKL algorithm is proposed for AD subject classification. It combines anatomical regularization and spatial regularization based on SMO-MKL to include priors in SVM. Numerical results show that the new classifier can significantly generate high performance.

SVM have to convert 3D discrete brain images naturally represented by higher-order tensors to vector in order to meet their input requirements. This destroys the natural structure and correlation in the original data and results in the curse of dimensionality. In our experiment, huge Laplacian matrix $L_s$ leads to the computation of spatial kernel matrix $e^{-\beta_s L_s}$ requiring nearly 70 hours. How to avoiding the loss of information inherent in matricization of the tensor and reducing computational complexity will be part of our future work.

### Acknowledgments

This work was supported in part by grants of the Natural Science Foundation of China (60862003), Project of Int’l S&T Cooperation Program of China (2009DFR10530), and The Ph.D. Programs Foundation of Ministry of Education of China (20095201110002).

### References

[1] R. Cuingnet, J.A. Glaunés, M. Chapin, H. Benali, O. Colliot, and The Alzheimer’s Disease Neuroimaging Initiative, “Spatial and Anatomical Regularization of SVM, A General Framework for Neuroimaging Data,” IEEE Trans. Pattern Anal. Mach. Intell., vol.35, no.3, pp.682–696, 2013.

[2] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M. Habert, M. Chapin, H. Benalid, O. Colliot, and The Alzheimer’s Disease Neuroimaging Initiative, “Automatic Classification of Patients with Alzheimer’s Disease from Structural MRI: A Comparison of ten Methods Using the ADNI Database,” Neuroimage, vol.56, no.2, pp.766–781, 2011.

[3] V.N. Vapnik, The Nature of Statistical Learning Theory, 2nd ed., Springer-Verlag, New York, USA, 2000.

[4] S. Vishwanathan, Z. Sun, and N. Theera-Ampornpunt, “Multiple kernel learning and the SMO algorithm,” Adv. Neural. Inform. Process. Syst, Dec. 2010.

[5] C.-C. Chang and C.-J. Lin, “LIBSVM: a library for support vector machines,” ACM Transactions on Intelligent Systems and Technology, vol.2, no.3, pp.1–27, 2011.

[6] J. Ashburner and K.J. Friston, “Unified Segmentation,” Neuroimage, vol.26, no.3, pp.839–851, 2005.

[7] J. Ashburner, “A fast diffeomorphic image registration algorithm,” Neuroimage, vol.38, no.1, pp.95–113, 2007.