Tensorial Kernel Based on Spatial Structure Information for Neuroimaging Classification

YingJiang WU††, Student Member and BenYong LIU††, Nonmember

SUMMARY Recently, a high dimensional classification framework has been proposed to introduce spatial structure information in classical single kernel support vector machine optimization scheme for brain image analysis. However, during the construction of spatial kernel in this framework, a huge adjacency matrix is adopted to determine the adjacency relation between each pair of voxels and thus it leads to very high computational complexity in the spatial kernel calculation. The method is improved in this manuscript by a new construction of tensorial kernel wherein a 3-order tensor is adopted to preserve the adjacency relation so that calculation of the above huge matrix is avoided, and hence the computational complexity is significantly reduced. The improvement is verified by experimental results on classification of Alzheimer patients and cognitively normal controls.

1. Introduction

Alzheimer disease (AD) is the most common forms of neurodegenerative dementia, and thus early detection and accurate diagnosis of AD is increasingly crucial for patients, clinicians, and researchers. In the last two decades, many machine learning methods, such as support vector machine (SVM) [1] and supervised tensor learning (STL) framework, have been developed and applied to classification of magnetic resonance imaging (MRI) data involving AD and cognitively normal (CN) controls [5]–[8].

As an example, Cuingnet proposes a framework to include spatial structure information in SVM by using regularization operator [5], and it indicates a flexible way to model various types of proximity. For the purpose of introducing spatial structure information during the construction of spatial kernel, a huge adjacency matrix is used to determine the adjacency relation between each pair of voxels [5], [6]. Unfortunately, calculation of the spatial kernel involving the huge matrix is very computationally intensive. In order to overcome this shortcoming, a new tensorial kernel is constructed in this manuscript. This kernel preserves the adjacency relation in the form of a 3-order tensor instead of the above huge matrix so that the complicated calculation is avoided. In this manner, the optimization problem of tensor-based kernel learning is converted to a multiple kernel learning (MKL) problem and thus it may be solved by the sequential minimal optimization (SMO) algorithm [9].

To illustrate the performance of the proposed method, the MRI data in [7] are used and compared for AD patient and CN control classification, and the results show that the proposed method outperforms previous ones in computational efficiency and also maintains high classification accuracy.

2. Spatial Structure Information in SVM

SVM is a supervised learning algorithm for nonlinear classification of data by mapping into a high dimensional feature space [1]. We denote scalars by italic lowercase letters (a, b, c, . . .), vectors as boldface italic lowercase letters (a, b, c, . . .), matrices as italic capital letters (A, B, C, . . .), sets or spaces as capital letters (A, B, C, . . .), and tensors as boldface italic capital letters (A, B, C, . . .). Let (x, y), s ∈ 1, 2, . . . , n be a training set of instance-labeled pairs, x ∈ Ω and y ∈ [−1, 1] with input space Ω, SVM solves the following optimization problem [5]:

\[
\begin{align*}
(w^opt, b^opt) & = \arg \min_{w \in R^n, b \in R} \frac{1}{2} \sum_{s=1}^{n} l_{hinge}(y_s[<w, x_s> + b]) \\
& \quad + \lambda \left[ \|\hat{w}\|_2^2 \right],
\end{align*}
\]

where \( \lambda \in R^+ \) denotes the regularization parameter, \( l_{hinge} \) is a hinge loss function, \( L \) is a Laplacian matrix of a graph and \( \beta \) is a parameter to control the size of the regularization.

In the case of 3D discrete brain MRI images, image connectivity (6, 18, or 26) is used as a simple way to define spatial proximity. Let \( L_s \) be the Laplacian matrix of a graph including the spatial proximity, it may be transformed to a diagonal matrix \( S \) by a symmetric orthogonal matrix \( Q \). Therefore, the spatial exponential matrix is computed as:

\[
e^{-\beta L_s} = e^{-\beta SQ} = Q e^{-\beta S} Q.
\]

Meanwhile, a kernel is obtained as \( k_p(x_1, x_2) = x_1^T e^{-\beta L_s} x_2 \), and this is a heat or diffusion kernel on the graph [5].

Let the size of a 3D discrete brain MRI image be \( n_1 \times n_2 \times n_3 \), \( t = n_1 \times n_2 \times n_3 \), the matrices \( Q \) and \( S \) are constructed as:

\[
Q(u, v) = \sqrt{2 \pi} \sin \left( \frac{(2u+1)(2v+1)\pi}{4t} \right), \quad 0 \leq u, v \leq t-1,
\]

\[
S = Q \times L_s \times Q,
\]
where $L_n$, $Q$, $S$ are $t \times t$ matrices.

Unfortunately, the size of the matrices $L_n$, $Q$, and $S$ is so huge that it is generally very difficult to store them in the memory of a general computer, and what is worse, the computational complexity of the heat kernel is $O(n_1 \times n_2 \times n_3)^3)$. To circumvent this problem, we construct a new tensorial kernel wherein a 3-order tensor is adopted to improve the computational efficiency of the algorithm in [5], [6] and at the same time, to preserve the adjacency relation in 3D MRI data so that high classification results may be obtained.

3. Tensorial Kernel Function

Let a H Hilbert space be a class of functions defined on a set $X$, and the function $k(x, y)$ of $x$ and $y$ in $X$ is a reproducing kernel of H. Kloft devises a new connection between several MKL formulations and extends these traditional MKL algorithms to arbitrary norms using two efficient interleaved optimization strategies to allow for robust kernel mixtures [2]. This approach contains an important precondition that we are given $J$ different feature mappings $\phi_j$: $X \rightarrow H_j$, $j = 1, 2, \ldots, J$, each giving rise to a reproducing kernel $k_j$ of $H_j$. That is to say, all the kernels must be defined on the same set $X$. Vishwanathan demonstrates that linear MKL regularized with certain Bregman divergence may be trained using SMO algorithm [9]. It can be proven that SMO-MKL algorithm is a special case of Kloft approach under the same precondition.

Let $(X_i, y_i), i = 1, 2, \ldots, m$ be a training set of instance-labeled pairs, $X_i \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, and $y_i \in \{-1, 1\}$. Given a tensor $X_i \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we assume that it may be mapped into a high dimensional tensor product feature space by a map function $\phi$: $X \rightarrow \phi(X) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, and then the optimization problem of tensor-based kernel learning may be written as [3]:

$$\min_{W, b} \frac{1}{2} ||W||^2_F + C \sum_{i=1}^{m} \xi_i$$

s.t. $y_i(\langle W, \phi(X_i) \rangle + b) \geq 1 - \xi_i,$

$$\xi_i \geq 0, \forall i = 1, \ldots, m,$$

where $W \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a weight tensor.

Using the Lagrangian relaxation method, the resulting decision function may be obtained as [3]:

$$f(X) = \text{sign} \left( \sum_{i=1}^{m} a_i y_i k(X_i, X) + b \right).$$

In (6), it is necessary to predetermine the tensorial kernel function $k(X_i, X)$ and estimate the parameter set $a_i$ with $i = 1, 2, \ldots, m$, wherein the former is the key step.

In the case of 3D discrete brain images, $X_i \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $y_i \in \{-1, 1\}$. In order to satisfy the precondition of the SMO-MKL algorithm, all the kernel functions must be defined on the same set. Let $n_{\text{max}} = \max\{n_1, n_2, n_3\}$, 3D discrete brain images may be extended from $\mathbb{R}^{n_1 \times n_2 \times n_3}$ to $\mathbb{R}^{n_{\text{max}} \times n_{\text{max}} \times n_{\text{max}}}$ by defining gray value $f(x, y, z) = 0$ if neither $x$ or $y$ or $z$ is in $\mathbb{R}^{n_1 \times n_2 \times n_3}$, that is:

$$f(x, y, z) = \begin{cases} f(x, y, z), & x, y, z \in \mathbb{R}^{n_1 \times n_2 \times n_3} \\ 0, & x \text{ or } y \text{ or } z \notin \mathbb{R}^{n_1 \times n_2 \times n_3} \text{ and } x, y, z \in \mathbb{R}^{n_{\text{max}} \times n_{\text{max}} \times n_{\text{max}}}. \end{cases}$$

We call this the zero extension of 3D discrete brain images. Therefore, all the sizes of frontal slices, horizontal slices and lateral slices of 3D discrete brain images are equal to $n_{\text{max}} \times n_{\text{max}}$. As a result, all the heat kernels generated from these slices are defined on the same space $\mathbb{R}^{n_{\text{max}} \times n_{\text{max}} \times n_{\text{max}}}$ and satisfy the precondition of the SMO-MKL algorithm. It is important to note that such heat kernels do not increase the amount of computation because the gray values of extension part are zeros and not used for actual calculation of heat kernels.

With the definition of the third-order tensor inner product in [12], the corresponding linear kernel function may be defined as:

$$k(X, Y) = \langle X, Y \rangle = \sum_{i_1=1}^{n_{\text{max}}} \sum_{i_2=1}^{n_{\text{max}}} \sum_{i_3=1}^{n_{\text{max}}} X_{i_1i_2i_3} Y_{i_1i_2i_3} = \sum_{i_1=1}^{n_{\text{max}}} \sum_{i_2=1}^{n_{\text{max}}} \sum_{i_3=1}^{n_{\text{max}}} X_{i_1i_2i_3} Y_{i_1i_2i_3} \xi_{i_1i_2i_3}.$$

The above formula shows that the third-order tensor linear kernel function is the sum of the multiplication of vectorization results derived from the corresponding frontal slices of tensors $X$ and $Y$. Since vectorization of a frontal slice breaks adjacency relation within the slice, a heat kernel generated from Laplacian matrix is introduced to alleviate the problem.

Therefore, a new kernel function of the third-order tensor may be constructed as follows:

$$k_{\text{frontal}}(X, Y) = \sum_{i=1}^{n_{\text{max}}} d_{(i)} k_{\text{heat}}(X_{(i)}, Y_{(i)})$$

$$= \sum_{i=1}^{n_{\text{max}}} d_{(i)} [\text{vec } X_{(i)}]^T e^{-R t_{(i)}} [\text{vec } Y_{(i)}],$$

where $X_{(i)}$ and $Y_{(i)}$ are respectively the $i_3$-th frontal slices of tensors $X$ and $Y$, $d_{(i)}$ is kernel weight, $\text{vec}$ is a vectorization operation of a slice, and $L_{(i)}^f$ is the Laplacian matrix of a graph constructed from all the voxels of the $i_3$-th frontal slice. A geometric interpretation of this tensorial kernel function is shown in Fig. 1.

For the horizontal and lateral slices of a tensor data, the above process may be repeated and then a tensorial kernel function considering the spatial structure information is obtained as follows:

$$k(X, Y) = \sum_{i=1}^{n_{\text{max}}} d_{(i)} [\text{vec } X_{(i)}]^T e^{-R t_{(i)}} [\text{vec } Y_{(i)}]$$
where $d_{i1}$ and $d_{i2}$ are kernel weights, and $L_{h(i)}$ and $L_{l(i)}$ are the Laplacian matrices of graphs constructed from all the voxels of the $i_1$-th horizontal and the $i_2$-th lateral slices, respectively.

If the above kernel function is adopted in (6), then the optimization problem of tensor-based kernel learning is converted to a multiple kernel learning problem and it is named Spatial-Tensor-Kernel in this manuscript, and it may be solved by the SMO-MKL algorithm. And what is more, it is named Spatial-Tensor-Kernel in this manuscript, and it may be solved by the SMO-MKL algorithm. And what is more, it is named Spatial-Tensor-Kernel in this manuscript, and it may be solved by the SMO-MKL algorithm.

$$+ \sum_{i_1=1}^{n_{max}} d_{i1} [\text{vec} X_{(i1)}]^T e^{-\beta L_{h(i)}} [\text{vec} Y_{(i1)}],$$

$$+ \sum_{i_2=1}^{n_{max}} d_{i2} [\text{vec} X_{(i2)}]^T e^{-\beta L_{l(i)}} [\text{vec} Y_{(i2)}],$$

(10)

where $d_{i1}$ and $d_{i2}$ are kernel weights, and $L_{h(i)}$ and $L_{l(i)}$ are the Laplacian matrices of graphs constructed from all the voxels of the $i_1$-th horizontal and the $i_2$-th lateral slices, respectively.

If the above kernel function is adopted in (6), then the optimization problem of tensor-based kernel learning is converted to a multiple kernel learning problem and it is named Spatial-Tensor-Kernel in this manuscript, and it may be solved by the SMO-MKL algorithm. And what is more, it is named Spatial-Tensor-Kernel in this manuscript, and it may be solved by the SMO-MKL algorithm.

4. Experiment

The performance of the proposed method is evaluated by experimental results on classification of 299 subjects, 137 patients with AD and 162 cognitively normal elderly controls, as the same study population in [5]–[8]. T1-weighted image is one of the basic pulse sequences in MRI and demonstrates differences in the T1 relaxation times of tissues. We only select T1-weighted, acquired at 1.5 Tesla, and pre-processed MRI images. All these images are obtained from Alzheimer Disease Neuroimaging Initiative (ADNI) public database (www.loni.ucla.edu/ADNI).

The selected images are segmented into gray matter (GM), white matter (WM) and Cerebrospinal Fluid (CSF) using the Statistical Parametric Mapping (SPM) [10], and then spatially normalized using the Diffeomorphic Anatomical Registration through Exponentiated Lie Algebra (DARTEL) algorithm with the default parameters in [11]. According to the SPM8 Manual, the New Segment procedure in SPM8 has been found to be generally more robust than the Segment procedure in SPM5. Therefore, in contrast to SPM5, which is adopted in Cuingnet study [5], [7], the key process of the proposed method applies SPM8 package to structural brain image segmentation, utilizes DARTEL algorithm to create GM/WM average-shaped template and normalizes the segmented image to Montreal Neurological Institute (MNI) standard space respectively. The features are

the GM probability maps in the MNI standard space and it is worth noticing that all maps should be modulated to ensure that the overall tissue amount does not change.

In the cortical thickness extraction or the SPHARM (spherical harmonics) computation as described in detail in [7], 7 subjects, 4 AD patients and 3 CN controls, are failed. However, 2 AD patients, 005_S_0929 and 013_S_0699, are also failed in the cortical thickness extraction in our experiment. These 9 subjects are excluded in [6]. In order to obtain a better evaluation of the proposed approach, all experimental data and the whole process should be as close as possible to the previous method. Therefore, the set of participants is randomly split, for 10 times, into two groups of the same size as Cuingnet study [5], [7], 150 training samples and 149 testing samples, based on the Mersenne Twister algorithm. Furthermore, those 9 subjects are also excluded if they belong to the testing set, and 50 percent misclassified if they belong to the testing set.

A grid-search and 10 cross validation on the training set is carried out to obtain optimal parameter values. Diffusion parameter $\beta$ would be set at 1.25, 1.5, 1.75, 2, 4, and 6 because the optimal value is around 1.75 according to our experiment. Table 1 shows the average results of 10 times random experiment with different $\beta$. In both Voxel-Regular-Spacial [5] and Spatial-Anatomical-MKL [6] methods, after the pre-processing procedure of MRI brain image, the size of a brain image is $121 \times 145 \times 121$. Thus, $t = 121 \times 145 \times 121 = 2122945$ and $L_s$, $Q$, $S$ are 2122945 and 2122945 matrices. These matrices are so huge that it needs nearly 70 hours to compute the heat kernel $x_1^T e^{-\beta L_s} x_2$. However, using the proposed Spatial-Tensor-Kernel algorithm, it needs only 15 minutes, about 300 times faster, while classification accuracy (Acc.) still maintains a high level. All the experiments are conducted on a workstation with Intel(R) 3.4 GHz processor and 32 GB RAM memory running Microsoft Windows XP.

As shown in Table 2, a single spatial regularization kernel (Voxel-Regular-spacial) reaches a peak accuracy of 89% and the Spatial-Anatomical-MKL algorithm yields a peak accuracy of 96%. Finally, it can be observed that the

| Table 1 | Classification performances of different parameter |
|---------|-----------------------------------------------|
| Diffusion parameter | Running time (seconds) | Spatial-Anatomical-MKL | Spatial-Tensor-Kernel |
| $\beta=1.25$ | 247812 | 966 | 95 | 94 |
| $\beta=1.5$ | 246590 | 925 | 96 | 94 |
| $\beta=1.75$ | 246782 | 919 | 96 | 95 |
| $\beta=2$ | 238781 | 983 | 96 | 95 |
| $\beta=4$ | 241666 | 979 | 96 | 94 |
| $\beta=6$ | 245338 | 996 | 96 | 94 |

| Table 2 | Performance comparison of different methods |
|---------|-----------------------------------------------|
| Method | Acc. (%) | Sens. (%) | Spec. (%) |
| 1. Voxel-Regular-Spacial | 89 | 85 | 93 |
| 2. Spatial-Anatomical-MKL | 95 | 96 | 99 |
| 3. Spatial-Tensor-Kernel | 95 | 95 | 98 |
proposed Spatial-Tensor-Kernel algorithm reaches a peak accuracy of 95%, and the sensitivity (Sens.) related to AD patients and specificity (Spec.) related to control subjects are also maintained at a high level.

5. Conclusion

A new tensorial kernel based on the spatial structure information is proposed for probable AD subject classification. In order to preserve the adjacency relation, Laplacian matrices of graphs constructed from all the voxels of every slice are introduced in the construction of this tensorial kernel. Numerical results show that the resultant classifier can greatly reduce the computational complexity.

In our work, the traditional inner product of two tensors, defined by sum of multiplication of the corresponding element, is adopted, and thus the natural structure in a tensorial data may be destroyed. Therefore it is part of our future work to find a new definition of tensorial inner product so that the intrinsic structural information contained in a tensor is retained.

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] V.N. Vapnik, The Nature of Statistical Learning Theory, 2nd ed., Springer-Verlag, New York, USA, 2000.
[2] M. Kloft, U. Brefeld, S. Sonnenburg, and A. Zien, “lp-Norm Multiple Kernel Learning,” Journal of Machine Learning Research, vol.12, no.12, pp.953–997, 2011.
[3] L. He, X. Kong, P.S. Yu, A.B. Ragin, Z. Hao, and X. Yang, “DuSK: A dual structure-preserving kernel for supervised tensor learning with applications to neuroimaging,” In SDM, SIAM, 2014.
[4] Z. Hao, L. He, B. Chen, and X. Yang, “A linear support higher-order tensor machine for classification,” IEEE Trans. Image Process., vol.22, no.7, pp.2911–2920, July 2013.
[5] R. Cuingnet, J.A. Glaunes, M. Chupin, 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.
[6] Y. Wu and B. Liu, “Spatial and anatomical regularization based on multiple kernel learning for neuroimaging classification,” IEICE Trans. Inf. & Syst., vol.E99-D, no.4, pp.1272–1274, April 2016.
[7] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot, “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.
[8] Y. Wu and B. Liu, “Spatial regularization based on support tensor machine for neuroimaging classification,” JIMET, Chongqing, China, Dec. 2015.
[9] S. Vishwanathan, Z. Sun, and N. Theera-Ampornpunt, “Multiple kernel learning and the SMO algorithm,” Adv. Neural. Inform. Process Syst, Dec. 2010.
[10] J. Ashburner and K.J. Friston, “Unified Segmentation,” NeuroImage, vol.26, no.3, pp.839–851, 2005.
[11] J. Ashburner, “A fast diffeomorphic image registration algorithm,” NeuroImage, vol.38, no.1, pp.95–113, 2007.
[12] T.G. Kolda and B.W. Bader, “Tensor decompositions and applications,” SIAM Rev., vol.51, no.3, pp.455–500, 2009.