

Spatial regularization based on support tensor machine for neuroimaging classification

Yingjiang Wu^{1,2}, Benyong Liu^{1,*}

¹Institute of Intelligent Information Processing, Guizhou University, Guiyang, China

²School of Information Engineering, Guangdong Medical University, Dongguan, China

*csc.byliu@gzu.edu.cn

Keywords: spatial regularization; support tensor machine; Alzheimer disease; neuroimaging

Abstract. Recently, a high dimensional classification framework has been proposed to introduce spatial and anatomical priors in support vector machine (SVM) optimization scheme for brain image analysis. However, classical SVM has to convert 3D discrete brain images naturally represented by higher-order tensors to one-dimensional vectors in order to meet the input requirements. This traditional method destroys the natural structure and correlation in the original data, and generates high dimensional vectors. In this manuscript, the method is improved by a modified support tensor machine (STM) algorithm to make full use of spatial prior and the inherent information of tensor. The new approach reduces memory requirements and computational complexity significantly, and it is comparably demonstrated by experimental results on classification of Alzheimer patients and elderly controls.

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, for the purpose of satisfying the input requirements of classical SVM, 3D discrete brain images naturally represented by higher-order tensors have to convert to vectors in advance. This conversion breaks the inherent structure and correlation in the original data, and easily produces high dimensional vectors and leads to a poor performance in time complexity and space complexity. In order to improve above method including spatial prior of 3D discrete brain images in SVM, a new method is proposed in this manuscript to maintain the natural structure in the original tensor data and include spatial prior of all frontal slices of the tensors in STM [2]. 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 [3]. It can be observed that the proposed method reduces size of memory requested and the time complexity significantly and produces better classification performances.

Spatial prior in SVM

SVM is a supervised learning algorithm introduced by Vapnik and performs classification by mapping the data into higher dimensional space [4]. Let (\mathbf{x}_s, y_s) , $s \in 1, \dots, S$ be a training set of instance-labeled pairs, $\mathbf{x}_s \in \Omega$, $y_s \in \{-1, 1\}$ and Ω is the input space, then SVM optimization problem can be written as:

$$(\mathbf{w}^{opt}, b^{opt}) = \arg \min_{\mathbf{w} \in \Omega, b \in \mathbb{R}} \frac{1}{S} \sum_{s=1}^S l_{hinge}(y_s[\langle \mathbf{w}, \mathbf{x}_s \rangle_{\Omega} + b]) + \lambda \|\mathbf{w}\|_{\Omega}^2 \quad (1),$$

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 β is a parameter to control the size of the regularization, the optimization problem may be rewritten as:

$$(\mathbf{w}^{opt}, b^{opt}) = \arg \min_{\mathbf{w} \in \Omega, b \in R} \frac{1}{S} \sum_{s=1}^S l_{hinge}(y_s [\langle \mathbf{w}, \mathbf{x}_s \rangle + b]) + \lambda \| e^{\frac{1}{2}\beta L} \mathbf{w} \|^2 \quad (2).$$

Meanwhile, the heat kernel is obtained as $k_\beta(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^T e^{-L\beta} \mathbf{x}_2$.

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_s be the Laplacian matrix of the graph including the spatial proximity. It is not difficult to verify that L_s is very close to a cyclic matrix, so it is diagonalizable to diagonal matrix S by a symmetric orthogonal matrix Q which can be constructed from discrete sine transform (DST) matrix [5]. Hence, the matrix Q and the matrix S can be calculated respectively as:

$$Q(m, n) = \sqrt{\frac{2}{N}} \sin \frac{(2m+1)(2n+1)\pi}{4N} \quad 0 \leq m, n \leq T-1 \quad (3),$$

$$S = Q^* L_s Q \quad (4),$$

where L_s , Q , S are $T \times T$ matrices, $T = n_1 * n_2 * n_3$ and the size of a 3D discrete brain image is $n_1 \times n_2 \times n_3$.

Therefore, the exponential matrix is computed as:

$$e^{-\beta_s L_s} = e^{-\beta_s Q S Q} = Q e^{-\beta_s S} Q \quad (5).$$

Finally, the computational complexity of the exponential matrix is $O((n_1 * n_2 * n_3)^6)$.

Spatial prior in STM

Notation and Basic Definitions. In this section, we first introduce some traditional notation and basic definitions in the area of multilinear algebra [6-7].

Definition 1 (tensor): A tensor, as known as multi-dimensional array, can be represented as $\chi \in R^{I_1 \times I_2 \times \dots \times I_N}$, where N is the order of the tensor. The element of χ is denoted by $a_{i_1 i_2 \dots i_N}$, $1 \leq i_n \leq I_n$, $1 \leq n \leq N$.

Definition 2 (outer product): The outer product $\chi \circ y$ of a tensor $\chi \in R^{I_1 \times I_2 \times \dots \times I_N}$ and another tensor $y \in R^{I_1 \times I_2 \times \dots \times I_M}$ is defined by $(\chi \circ y)_{i_1, i_2, \dots, i_N, i'_1, i'_2, \dots, i'_M} = \chi_{i_1, i_2, \dots, i_N} y_{i'_1, i'_2, \dots, i'_M}$ for all values of the indices.

Definition 3 (inner product): The inner product $\langle \chi, y \rangle$ of two same-sized tensors $\chi, y \in R^{I_1 \times I_2 \times \dots \times I_N}$ is defined as the sum of the products of their entries, that is: $\langle \chi, y \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} \chi_{i_1, i_2, \dots, i_N} y_{i_1, i_2, \dots, i_N}$.

Definition 4 (n -mode (matrix) product): The n -mode (matrix) product of a tensor $\chi \in R^{I_1 \times I_2 \times \dots \times I_N}$ with a matrix $U \in R^{J \times I_n}$ is denoted by $\chi \times_n U$ and is a tensor in $R^{I_1 \times I_2 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N}$. Elementwise, we have $(\chi \times_n U)_{i_1, \dots, i_{n-1}, j, i_{n+1}, \dots, i_N} = \sum_{i_n=1}^{I_n} \chi_{i_1, i_2, \dots, i_N} u_{j, i_n}$.

Definition 5 (n -mode (vector) product): The n -mode (vector) product can be seen as a special case of the n -mode (matrix) product, and the n -mode (vector) product of a tensor $\chi \in R^{I_1 \times I_2 \times \dots \times I_N}$ with a vector $V \in R^{I_n}$ is denoted by $\chi \times_n V$ and is a tensor in $R^{I_1 \times I_2 \times \dots \times I_{n-1} \times I_{n+1} \times \dots \times I_N}$. Elementwise, we have $(\chi \times_n V)_{i_1, \dots, i_{n-1}, i_{n+1}, \dots, i_N} = \sum_{i_n=1}^{I_n} \chi_{i_1, i_2, \dots, i_N} v_{i_n}$.

Definition 6 (rank-one tensor): $\chi \in R^{I_1 \times I_2 \times \dots \times I_N}$ is a rank-one tensor if it can be represented as the outer product of N vectors, that is, $\chi = a^{(1)} \circ \dots \circ a^{(N)}$, where the symbol “ \circ ” represents the vector outer product.

Definition 7 (CP decomposition): CP decomposes a tensor into a sum of component rank-one tensors. For example, a third-order tensor $\chi \in R^{I \times J \times K}$ can be decomposed to $\chi \in \sum_{l=1}^L a_l \circ b_l \circ c_l$, where L is a positive integer, $a_l \in R^I$, $b_l \in R^J$, $c_l \in R^K$ and $l = 1, \dots, L$.

Definition 8 (Frobenius norm): The Frobenius norm of a tensor $\chi \in R^{I_1 \times I_2 \times \dots \times I_N}$ is the square root of the sum of the squares of all its elements, that is: $\|\chi\|_F = \sqrt{\langle \chi, \chi \rangle} = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} \chi_{i_1, i_2, \dots, i_N}^2}$.

Classic STM. Let (x_m, y_m) , $m \in 1, \dots, M$ be a training set of instance-labeled pairs, $x_m \in R^{I_1 \times I_2 \times \dots \times I_N}$ and $y_m \in \{-1, 1\}$. Classic STM restricts tensor weight parameter to rank-one tensor as:

$$\mathbf{W} = \mathbf{w}^{(1)} \otimes \mathbf{w}^{(2)} \otimes \dots \otimes \mathbf{w}^{(N)} = \bigotimes_{n=1}^N \mathbf{w}^{(n)} \quad (6),$$

and consists of N quadratic programming (QP) problems with inequality constraints, wherein the n th QP problem can be represented as:

$$\begin{aligned} \min_{\mathbf{w}^{(n)}, b^{(n)}, \xi^{(n)}} J(\mathbf{w}^{(n)}, b^{(n)}, \xi^{(n)}) &= \frac{1}{2} \prod_{i=1}^N \|\mathbf{w}^{(i)}\|_F^2 + c \sum_{m=1}^M \xi_m^{(n)} \\ \text{s.t. } y_m ((\mathbf{w}^{(n)})^T (x_m \prod_{1 \leq i \leq N} \times_i \mathbf{w}^{(i)})) + b^{(n)} &\geq 1 - \xi_m^{(n)}, \\ \xi_m^{(n)} &\geq 0, \quad m = 1, \dots, M. \end{aligned} \quad (7).$$

These N optimization models have no closed-form solution, so the alternating projection algorithm is used to solve them and updates one weight parameter using LIBSVM [8] at one time.

Finally, the class label of a testing example χ can be predicted as follow: $y(\chi) = \text{sign}(\chi \prod_{n=1}^N \times_n \mathbf{w}^{(n)} + b)$.

Spatial Prior in STM. As described in the text, L_s is a Laplacian matrix of the graph including the spatial proximity and the size of L_s is $(n_1 * n_2 * n_3) \times (n_1 * n_2 * n_3)$, nearly $10^6 \times 10^6$ in practice. Unfortunately, this size of matrix is so large that cannot be stored in the memory, and the same situation also appears in the process of calculating the matrix Q , the matrix S and the exponential matrix. Although these matrices are barely able to calculate by computation skills, the time complexity and the space complexity is rather poor.

In the following, this problem is reconsidered from the point of view of tensor which constitutes efficient representation of 3D discrete brain images.

Classic STM updates one weight parameter at one time. In order to include spatial prior in STM, We modify the original algorithm and divide the iterative process into two steps. The first step calculates n -mode product of $\chi \times_3 \mathbf{w}_3$ and updates two weight parameters, \mathbf{w}_1 and \mathbf{w}_2 , using LIBSVM and CP decomposition, where $\chi \in R^{I_1 \times I_2 \times I_3}$ denotes a 3D discrete brain image. The second step calculates n -mode product of $\chi \times_1 \mathbf{w}_1 \times_2 \mathbf{w}_2$ and updates one weight parameter \mathbf{w}_3 using LIBSVM. The detailed procedures of modified STM algorithm are described in Table 1.

Let R be the result of n -mode (vector) product of a tensor χ with \mathbf{w}_3 , i.e., $R = \chi \times_3 \mathbf{w}_3$, and R is an $n_1 \times n_2$ matrix as the same size as each frontal slice of the tensor. The Laplacian matrix of the graph including the spatial proximity of matrix R is named $newL_s$. It keeps the same degree and adjacency relations as the Laplacian matrix of the graph including the spatial proximity of each frontal slice. Matrix R is a linear combination of all frontal slices according to the definition of n -mode (vector) product, hence, introducing spatial prior of matrix R in STM is equivalent to introducing spatial prior of all frontal slices in STM. In other words, different from including the spatial prior of 3D discrete brain images in SVM, the proposed method maintains the inherent structure and correlation in the original tensor data and introduces the spatial prior of all frontal slices of the tensors in STM.

Table 1: Including spatial prior in STM

Input: A Set of Tensor Samples $\{\mathbf{x}_m \in R^{I_1 \times I_2 \times I_3}, \mathbf{y}_m \in \{-1,1\}\}_{m=1}^M$, the Threshold Parameter ε .
Output: $y(x)$: the class label of a testing example x
Set \mathbf{w}_n equal to random unit vector in R^n , $n=1, 2, 3$. for $i=1$ to <i>number_of_iteraton</i> or until convergence % update parameter \mathbf{w}_1 and \mathbf{w}_2 for $m=1$ to M $x_iter = \mathbf{x}_m \times_3 \mathbf{w}_3$; $traindata(m,:) = x_iter$; next m $ktrain = traindata^T * exponential_matrix * traindata$ $Model = svm(trainlabel, ktrain, 't -4')$ $(\mathbf{w}_1, \mathbf{w}_2) = CP(\mathbf{w})$ % update parameter \mathbf{w}_3 for $m=1$ to M $x_iter = \mathbf{x}_m \times_1 \mathbf{w}_1 \times_2 \mathbf{w}_2$; $traindata(m,:) = x_iter$; next m $Model = svm(trainlabel, traindata, 't -0')$ if $abs[\sum_{n=1}^3 (\mathbf{w}_n^{(i-1)} ^T \mathbf{w}_n^{(i)} \ \mathbf{w}_n^{(i)}\ _F^2 - 1)] < \varepsilon$ then break where $i=2$ to <i>number_of_iteraton</i> next i $y(x) = sign(x \times_1 \mathbf{w}_1 \times_2 \mathbf{w}_2 \times_3 \mathbf{w}_3 + b)$

Similarly, the matrix Q , the matrix S and the exponential matrix $e^{-\beta_s, newL_s}$ can be calculated according to (3), (4) and (5) respectively, with different dimensions. Obviously, the size of $newL_s$ is $(n_1 * n_2) \times (n_1 * n_2)$. Therefore, the computational complexity of the exponential matrix is greatly reduces from $O((n_1 * n_2 * n_3)^6)$ to $O((n_1 * n_2)^6)$.

In the first step of iterative process of modified STM algorithm, the optimal weight parameter \mathbf{w} of training result of LIBSVM is an $(n_1 * n_2) \times 1$ matrix. The weight parameter \mathbf{w} can be decomposed into \mathbf{w}_1 and \mathbf{w}_2 using the CP decomposition software package (www.sandia.gov/~tgkolda/TensorToolbox) according to the assumption of (6). Thus, LIVSM and CP decomposition makes it possible to update the two weight parameters, \mathbf{w}_1 and \mathbf{w}_2 , at the same time.

Experiment

After the proposed method is applied to classification of 299 subjects, 137 AD patients and 162 CN controls, as the same study population in [3]. 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) [9], and the Diffeomorphic Anatomical Registration through Exponentiated Lie Algebra (DARTEL) algorithm is used for spatial normalization [10]. 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. Wherein 9 subjects, 6 AD patients and 3 CN controls, are failed in pre-processing as described in detail in [3]. 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 20 times, into two groups of the same size, 145 training and 145 testing, based on the Mersenne Twister algorithm.

Diffusion parameter β should be set at 0.001, 0.003, 0.005, 0.007 and 0.009, because the optimal value is around 0.005 according to our experiment. Table 2 shows the average results of 20 times random

experiment with different β , wherein the number of iteration is set at 100 and the threshold parameter ε is set at $1e-4$. Furthermore, contrast to classic SVM costing about 70 hours, the proposed method only spends nearly one percent of running time and classification accuracy (Acc.) still maintains a high level.

We propose two other ways of using STM for AD subject classification. The first method, named Direct-STM, applies classic STM for training 3D discrete brain images directly and yields the lowest accuracy of 89.53%. The second method, named ROI (Regions of Interest)-STM, applies classic STM for training ROI dataset extracting from 3D discrete brain images and produces the lowest running time because the ROI dataset is a large sparse tensor.

Table 3 shows the classification performances of three proposed methods. It is observed that including spatial prior in STM can significantly improve classification accuracy.

As shown in Table 4, 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]. It can be observed that the proposed method based on including spatial prior in STM ranks first for all classification types. The sensitivity (Sens.) related to AD patients and specificity (Spec.) related to control subjects is also quantified.

Table 2: Classification performances of different parameter

Diffusion parameter	Running time (seconds)	Acc. (%)
$\beta=0.001$	1877	94.61
$\beta=0.003$	2280	94.71
$\beta=0.005$	2469	94.85
$\beta=0.007$	2479	94.68
$\beta=0.009$	2823	94.58

Table 3: Classification performances of proposed methods

Methods	Running time (seconds)	Acc. (%)
Direct-STM	1766	89.53
ROI-STM	1286	92.71
Spatial-prior-in-STM	2469	94.85

Table 4: Performance comparison of different methods

Method	Acc. (%)	Sens. (%)	Spec. (%)
1、 Voxel-Regual-spacial	89	85	93
2、 Voxel-Regual-atlas	90	82	96
3、 Voxel-Regual-combinegraph	87	82	90
4、 Ours	95	94	98

Conclusions

A novel method including spatial prior in STM algorithm is proposed for classification of Alzheimer patients and elderly controls. Theoretical analysis and experiment show that the proposed new method not only reduces the time complexity and space complexity significantly, but also produces better classification accuracy.

Acknowledgment

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. 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 Analysis and Machine Intelligence, 2013, vol.35, no.3, pp.682-696.
- [2] D. Tao, X. Li, X. Wu, W. Hu, and S. Maybank, Supervised tensor learning, Knowl. Inf. Syst., 2007, vol. 13, no. 1, pp. 1–42.
- [3] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M. Habert, M. Chupin, H. Benali d, 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, 2011, vol.56, vol.2, pp.766-781.
- [4] V.N.Vapnik, The nature of statistical learning theory, Springer-Verlag, New York, USA, 2nd edition, 2000.
- [5] P. Davis, Circular matrices, John Wiley, New York, 1979.
- [6] T. Kolda and B. Bader, Tensor decompositions and applications, SIAM Rev., 2009, vol. 51, no. 3, pp. 455–500.
- [7] Z. Hao, L. He, B. Chen, and X. Yang, A linear support higher-order tensor machine for classification, IEEE Trans. Image Process., Jul. 2013, vol. 22, no. 7, pp. 2911–2920.
- [8] C. Chang and C. Lin, LIBSVM: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.
- [9] J. Ashburner and k. Friston, Unified Segmentation, NeuroImage, 2005, vol.26, no.3, pp.839-851.
- [10] J. Ashburner, A fast diffeomorphic image registration algorithm, Neuroimage, 2007, vol.38, no.1, pp.95-113.