Diffuse Outlier Time Series Detection Technique for Functional Magnetic Resonance Imaging

Javier Giacomantone[1], Tatiana Tarutina[2,3], and Armando De Giusti[1,3]

[1] Instituto de Investigación en Informática (III-LIDI),
Facultad de Informática - Universidad Nacional de La Plata
{jog,degiusti}@lidi.info.unlp.edu.ar
[2] Instituto de Física, Facultad de Ciencias Exactas
Universidad Nacional de La Plata
tarutina@fisica.unlp.edu.ar
[3] Consejo Nacional de Investigaciones Científicas y Tecnológicas

Abstract. We propose a new support vector machine (SVM) based method that improves the time series classification in magnetic resonance imaging (fMRI). We exploit the robust anisotropic diffusion (RAD) technique to increase the classification performance of the one class support vector machine by taking into account the hypothesis of spatial relationship between active voxels. The proposed method was called Diffuse One Class Support Vector Machine (DOCSVM). DOCSVM method treats activated voxels as outliers and applies one class support vector machine to generate an activation map and RAD to include the neighborhood hypothesis, improving the classification and reducing the iteration steps with respect to RADSPM. We give a brief review of the main methods, present receiver operating characteristic (ROC) results and conclude suggesting further research alternatives.

Keywords: Time Series, Functional Magnetic Resonance Imaging, classification, Support Vector Machines, Robust Anisotropic Diffusion

1 Introduction

The purpose of fMRI is to map areas of increased neuronal activity of the human brain associated with cognitive or motor tasks. The hemoglobin in the blood is a natural contrast agent, because it has different magnetic properties depending of its state of oxygenation. These differences affect the voxel intensity in the magnetic resonance images [1]. Baseline images are scanned periodically while the subject is at rest (or in other baseline condition) and activation images are acquired when the subject is performing a specific task or receiving a stimulus. A fMRI image can be seen as a set of time series where each time series corresponds to one voxel in the structural image. Classification of time series is the main subject of brain fMRI data analysis. A number of different techniques have been developed for fMRI data analysis, and can be classified in two main categories, model driven [2][3][4] and data driven methods [5][6]. Data

driven methods use a method in machine learning or statistics to analyze fMRI time series while model driven methods assume a model related to the structure and function of the brain. Support vector machine (SVM) being a data driven method has been applied to the supervised classification of cognitive states [7] by optimizing a margin and using the kernel trick [8]. One class SVM (OCSVM) [9][10] has been applied to fMRI unsupervised classification [11][12]. Brain fMRI time series on most voxels are independent of the experimental stimulus and time series on only few voxels are related to the experimental stimulus. Time series related to the stimulus can be considered as outliers and time series not related to the experimental stimulus as normal data points, satisfying the hypothesis necessary to apply the OCSVM methods. In order to include the spatial relationship between activated voxels, that assumes that time series on close voxels have similar state activation correlative or irrelative to the experimental stimulus, new alternatives have been proposed [13][14][15][16]. In this work we present preliminary results of a new technique, DOCSVM, that combines the one class support vector machine and the RAD to improve the classification of fMRI time series by considering the neighborhood spatial relationship.
The paper is organized as follows, section 2 to 4 cover the fundamental ideas behind OCSVM, RAD and DOCSVM. Section 5 and 6 show experimental results, conclusions and propose future research work.

2 One-class SVM

There are two main one class classification algorithms based on SVM, support vector data description [9] and one-class SVM [10]. A typical example of interest of one class classification is the outlier detection that attempts to detect uncharacteristic objects from a data set. The one-class SVM estimates a function f that is positive for a subset of the sample space and negative for the complement. The algorithm maps the data into a feature space corresponding to the kernel and separates them from the origin with maximum margin. Different types of kernels can be used corresponding to nonlinear estimators in the input space.

Consider a given data set

$$\mathbf{x}_1, \ldots, \mathbf{x}_l \in \mathcal{X},$$

where $l \in \mathbb{N}$ is the number of observations and $\mathcal{X}$ is a compact subset of $\mathbb{R}^N$. Let $\Phi : \mathcal{X} \to \mathcal{F}$ be a feature map, that is, a map into an inner product space $\mathcal{F}$ such that the inner product in the image of Φ can be computed by evaluating a simple kernel.

$$k(\mathbf{x}, \mathbf{z}) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{z}))$$

It can be formulated as an optimization problem.

$$\begin{aligned} \min_{\mathbf{w}\in\mathcal{F},\xi\in\mathbb{R}^l,\rho\in\mathbb{R}} \quad & \tfrac{1}{2}||\mathbf{w}||^2 + \tfrac{1}{\nu l}\textstyle\sum_i \xi_i - \rho \\ \text{s.t.} \quad & (\mathbf{w} \cdot \Phi(\mathbf{x}_i)) \geq \rho - \xi_i, \xi_i \geq 0, \end{aligned} \tag{1}$$

where $\nu \in (0, 1]$ is a parameter controlling the penalized term and ξ_i are slack variables. By solving the optimization problem (1) we obtain $\mathbf{w}$ and ρ and the decision function is -1 for outliers in the data set and +1 for the rest of the samples in the data set.

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w} \cdot \Phi(\mathbf{x})) - \rho) \tag{2}$$

Introducing Lagrangian multipliers $\alpha_i, \beta_i \geq 0$, we obtain

$$L(\mathbf{w}, \xi, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \tfrac{1}{2}||\mathbf{w}||^2 + \tfrac{1}{\nu l}\textstyle\sum_i \xi_i - \sum_i \beta_i \xi_i - \rho$$

$$- \textstyle\sum_i \alpha_i (\mathbf{w} \cdot \Phi(\mathbf{x}) - \rho + \xi_i)$$

Setting the derivatives with respect to the primal variables $\mathbf{w}, \xi, \rho$ equal to zero yields

$$\mathbf{w} = \textstyle\sum_i \alpha_i \Phi(\mathbf{x}_i),$$

$$\alpha_i = \tfrac{1}{\nu l} - \beta_i \leq \tfrac{1}{\nu l},$$

$$\textstyle\sum_i \alpha_i = 1.$$

The decision function can be written as

$$f(\mathbf{x}) = \text{sgn}(\sum_i \alpha_i k(\mathbf{x}_i, \mathbf{x}) - \rho)$$

The multipliers α_i can be solved from the dual problem:

$$\min_{\alpha} \quad \tfrac{1}{2}\textstyle\sum_{ij} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j)$$

$$\text{s.t. } \; 0 \geq \alpha_i \geq \tfrac{1}{\nu l}, \;\; \textstyle\sum_i \alpha_i = 1.$$

The parameter ρ can be recovered by exploiting that for any such α_i and the corresponding pattern $\mathbf{x}_i$ satisfies

$$\rho = (\mathbf{w} \cdot \Phi(\mathbf{x}_i)) = \sum_j \alpha_i k(\mathbf{x}_i, \mathbf{x}_j). \tag{3}$$

3 Robust Anisotropic Diffusion

Perona and Malik [17] defined the anisotropic diffusion as

$$\frac{\partial I(x, y, t)}{\partial t} = \text{div}\left[g\left(\|\nabla I(x, y, t)\|\right) \nabla I(x, y, t)\right], \tag{4}$$

using the original image $I(x, y, 0) : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ as the initial condition, where t is an artificial time parameter and g is an "edge-stopping" function. The right choice of g can greatly affect the extent to which discontinuities are preserved.

Perona and Malik suggested two possible edge-stopping functions in their paper [17]. Black et al. [18] used the robust estimation theory to choose a better edge-stopping function, called Tukey's biweight:

$$g(x) = \begin{cases} \left[1 - \frac{x^2}{5\sigma^2}\right]^2, & \frac{x^2}{5} \leq \sigma^2 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

The function g above is the dilated and scaled version of the original Tukey's function, where $g(0) = 1$ and the local maxima of its "influence function" $\psi(x) = xg(x)$ is situated at $x = \sigma$. The diffusion that uses the Tukey's function is called robust anisotropic diffusion (RAD) and this is the edge-stopping function adopted in this paper.
Perona and Malik [17] discretized spatio-temporally their anisotropic diffusion equation (4) as:

$$I(s, t+1) = I(s,t) + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} g(|\nabla I_{s,p}(t)|) \nabla I_{s,p}(t), \tag{6}$$

where $I(s,t)$ is a discretely sampled image, s denotes the pixel position in a discrete 2-D or 3-D grid, $t \geq 0$ now denotes discrete time steps, the constant λ determines the rate of diffusion (usually, $\lambda = 1$), and η_s represents the set of spatial neighbors of pixel s. For 2-D images, usually four neighbors are considered: *north*, *south*, *west* and *east*, except at the image boundaries. For 3-D images, six voxels are usually considered: the above-mentioned four plus "up" and "down" voxels. The gradient magnitude of a voxel in a particular direction at iteration t is approximated by:

$$\nabla I_{s,p}(t) = I(p,t) - I(s,t),\ p \in \eta_s. \tag{7}$$

Black et al. [18] suggested to use the "robust scale" defined by:

$$\sigma_e = 1.4826\,\mathrm{MAD}(\boldsymbol{\nabla} I) = 1.4826\,\mathrm{median}_I\left[\|\|\boldsymbol{\nabla} I\| - \mathrm{median}_I(\|\boldsymbol{\nabla} I\|)\|\right], \tag{8}$$

where MAD is the Median Absolute Deviation.

4 Diffuse One-class SVM

By combining OCSVM and RAD we proposed a new technique that improves the classification of fMRI temporal series under the validity of the spatial neighborhood hypothesis.

Let I' be an fMRI data. First of all, the mean value is removed from I', yielding the mean-removed fMRI I:

$$I = I' - \overline{I'} \tag{9}$$

This pre-processing is very important, because structural and functional regions of the brain do not necessarily match. No structural information should be diffused, but only the activation information. Note that the activation information is not affected at all by the mean-correction.

Time series on each voxel is taken as a data point. The $\mathbf{x}_{i,j,k}$ is the data point corresponding to the ith row, jth file and kth slice identifying one particular time series or data point. The $\mathbf{x}_{i,j,k}$ are directly the input into the optimization problem (1). The optimal solutions $\mathbf{w}$ and ρ can be obtained by solving the dual problem and (3). Then for each $\mathbf{x}_{i,j,k}$ a primal decision value $\mathbf{y}_{i,j,k}$ is obtained,

$$\mathbf{y}_{i,j,k} = (\mathbf{w} \cdot \Phi(\mathbf{x}_{i,j,k})) - \rho \tag{10}$$

that represents the distance between a point $\Phi(\mathbf{x}_{i,j,k})$ and a hyperplane in the high-dimensional kernel space $(\mathbf{w} \cdot \Phi(\mathbf{x})) - \rho = 0$.

Let us denote the fMRI data at iteration $t \geq 0$ of the diffusion process as $I(s, n, t)$, where $I(s, n, 0)$ is the initial mean-corrected fMRI at spatial voxel position s and volume n, and $T(s, t)$ the activation map form by $\mathbf{y}_{i,j,k} \forall i, j, k$ where s is particular position i, j, k.

1. Let $t \leftarrow 0$.
2. Calculate the activation map $T(s, 0)$ by using OCSVM (10).
3. Compute the diffusion coefficients. The diffusion coefficient between a voxel s and its neighboring voxel p at instant t is:

$$g(|\nabla T_{s,p}(t)|), \text{ where } \nabla T_{s,p}(t) = T(p, t) - T(s, t). \tag{11}$$

4. Use these coefficients to perform the diffusion in $I(s, n, t)$, yielding the diffused fMRI, $I(s, n, t + 1)$, at iteration $t + 1$:

$$I(s, n, t + 1) \leftarrow I(s, n, t) + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} g(|\nabla T_{s,p}(t)|) \nabla I_{s,p}(t), \tag{12}$$

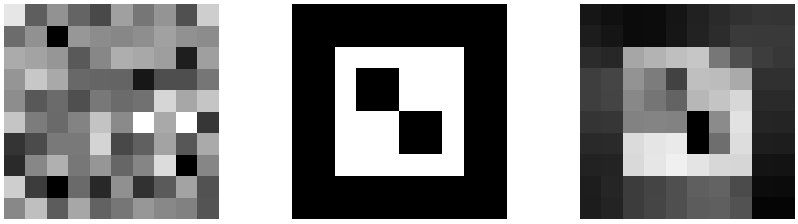
   where $\nabla I_{s,p}(n, t) = I(p, n, t) - I(s, n, t)$.
5. Let $t \leftarrow t + 1$ and repeat steps 2 to 5 some predefined number of times or until the average of diffused values (second term of equation (12)) is below some predefined threshold.
6. Classify each voxel applying the decision function of equation (2).

The anisotropic diffusion is controlled by the number of iterations and the scale parameter of the edge stopping function (5), σ.

5 Experimental Results

In order to test and develop classification models in fMRI three main data sets are commonly used. The first model is a completely synthetic one, carefully designed to reproduce real fMRI conditions like signal to noise ratio (SNR), type

of noise and spatial distribution of activated voxels. The second alternative is to generate artificial foci of activated voxels in real fMRI data. The third possible data set involves working with real fMRI time series. We present comparative results on synthetic data sets known also as synthetic time series, artificial images or phantoms. These types of experiments provide controlled conditions and knowledge of the exact activated region location, namely a gold standard. In order to test and compare results of the proposed method we generate two artificial images, based on the phantom proposed in [19], with different activation levels. The 4D fMRI model is formed by $10 \times 10 \times 3$ voxels per volume and 84 volumes. Voxels values were 16000 corrupted by zero-mean Gaussian noise with standard deviation $\sigma = 4000$. Active voxels had their values increased by 1000 for phantom I and 1500 for phantom II. The fMRI experiment had alternating blocks of 6 non-active and 6 active volumes, beginning with non-active volumes. Activated volumes had a $6 \times 6 \times 3$ activated region in the center of the volume, with two non-activated regions of $2 \times 2 \times 3$ voxels each. Fig. 1 depicts one activated slice of phantom II, the gold standard and the DOCSVM activation map.



(a) (b) (c)

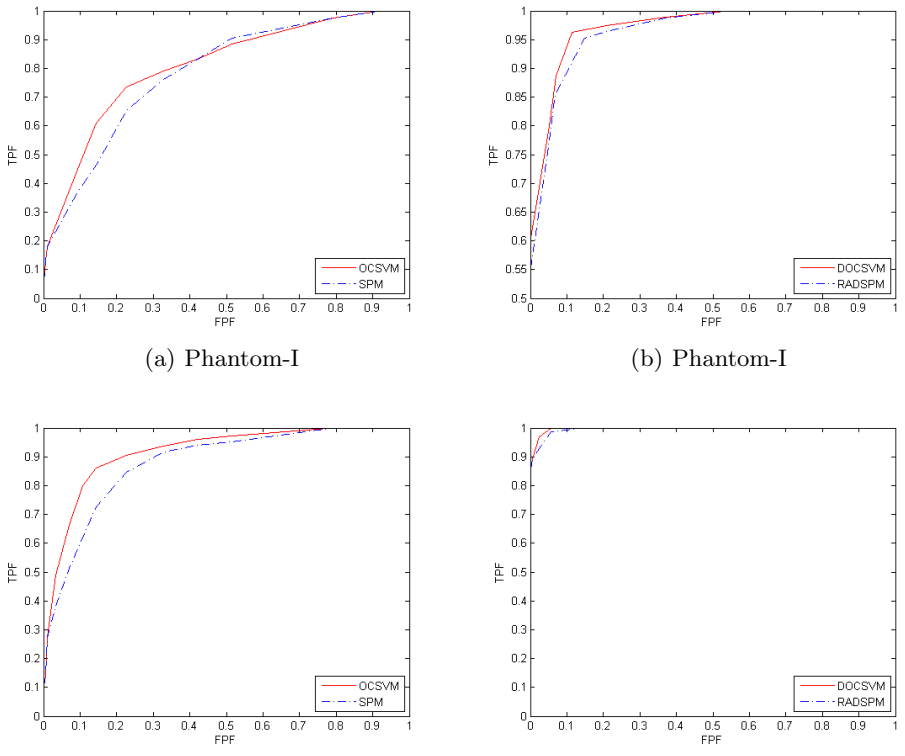
Fig. 1. (a) Simulated fMRI slice, (b) Reference image (gold standard), (c) Activation map produced by DOCSVM ($\nu = 0.7, \sigma = 2, t = 6$)

In the experiments on the two synthetic datasets, the principal parameters of the algorithms were set as follows. The radial basis function was chosen as the kernel function for the OCSVM and $\nu = 0.7$. RADSPM and DOCSVM are sensible to the scale parameter selection [20], beginning with $\sigma = \sigma_e$ we adjust σ using ROC curves as a gauging procedure as suggested in [21].

We have obtained comparative results of four different methods by using the the well-known Receiver Operating Characteristics (ROC) analysis [21][22][23]. Let TP, FN, FP and TN be respectively the number of true positives, false negatives, false positives and true negatives obtained by comparing the ideal classification (gold standard) and the results obtained by each of the evaluated methods. Then, the True Positive Fraction (TPF) and the False Positive Fraction (FPF) are defined as:

$$\mathrm{TPF} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \quad \mathrm{FPF} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{13}$$

Figure 2 depicts correlation's SPM, OCSVM's, RADSPM's and DOCSVM's ROC curves. Each point of a ROC curve is obtained by solving equation (13) for a specific threshold value. Table 1 presents some performance metrics of the four ROC curves, all of them demonstrating the improved performance of DOCSVM and RADSPM compared to the non-spatial oriented methods considered in the experiments. The area under the curve and the distance d_{oop} from the principal diagonal to the optimal operating point (OOP)(the point of the curve most distant from the principal diagonal), are superior for DOCSVM with less diffusion iteration steps with respect to the results of RADSPM.



(a) Phantom-I (b) Phantom-I

(c) Phantom-II (d) Phantom-II

Fig. 2. ROC curves

	Method	Area	d_{po}	TPF_{po}	FPF_{po}
Phantom-I	Correlation-SPM	0.7863	0.3063	0.7619	0.3287
	$RADSPM_{\sigma=1.8,t=10}$	0.9645	0.5687	0.9524	0.1481
	$OCSVM_{\nu=0.7}$	0.8081	0.3591	0.7348	0.2269
	$DOCSVM_{\sigma=1.8,t=8,\nu=0.7}$	0.9716	0.5993	0.9624	0.1148
Phantom-II	Correlation-SPM	0.8798	0.4373	0.8452	0.2269
	$RADSPM_{\sigma=2,t=10}$	0.9958	0.6594	0.9881	0.0556
	$OCSVM_{\nu=0.7}$	0.9166	0.5080	0.8619	0.1415
	$DOCSVM_{\sigma=2,t=6,\nu=0.7}$	0.9975	0.6685	0.9686	0.0231

Table 1. Performance metrics

6 Conclusions and Future Work

In this paper we have presented a new SVM based technique named DOCSVM taking into account the spatial relationship activation hypothesis. This technique improves fMRI time series classification. We compared this method to OCSVM, correlation analysis and RADSPM. Experimental results using ROC curves on synthetic data sets have shown promising results for DOCSVM. The proposed method treats activated voxels as outliers and applies OCSVM to generate an activation map and RAD to include the neighborhood hypothesis, improving the classification and reducing the iteration steps with respect to RADSPM. The obtained results of DOCSVM are preliminary. Further research involves improving the probabilistic model used to create the artificial images considering the noise distribution and the signal to noise ratio in order to approximate the complex and noisy fMRI signal structure. Extensive tests on real fMRI and artificial data must be done in order to adjust parameters and extend our results to different real experiment paradigms.

References

1. S. Ogawa et. al. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. Biophysics Journal, 14(3):803-812, 1993.
2. K. J. Friston, A. P. Holmes, K. J. Worsley, Poline J. P., C. D. Frith, R. S. Frackowiak. Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping, 2:189-210, 1995.
3. S. Faisan, L. Throava, J. Armspach, M. Metz-Lutz, F. Heith. Unsupervised learning and mapping of active brain functional MRI signals based on hidden semi-Markov event sequence models. IEEE. Transactions on Medical Imaging, 24(2):263-276, 2005.

4. J. Tian, L. Yang, J. Hu. Recent advances in the data analysis method of functional magnetic resonance imaging and its applications in neuroimaging. Progress in Natural Science, 16(8):785-795, 2006.
5. C. Goutte, P. Toft, E. Rostrup, F. Nielsen, L. Hansen. On clustering fMRI time series. NeuroImage, 9:298-310, 1999.
6. O. Friman, J. Carlsson, P. Lundberg, M. Borga, H. Knutsson. Detection of neural activity in functional MRI using canonical correlation analysis. Magnetic Resonance in Medicine, 45(2):323-330, 2001.
7. D. Cox, R. L. Savoy. Functional magnetic resonance imaging (fMRI) brain reading: detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage, 19:261-270, 2003.
8. V. Vapnik. An overview of statistical learning theory. IEEE Transactions on Neural Networks, 10(5):988-999, 1999.
9. D. M. J. Tax, R. P. W. Duin. Support vector data description. Machine Learning, 54:45-66, 2004.
10. B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Samola, R. C. Williamson. Estimating the support of a high dimensional distribution. Neural Computation, 13(7):1443-1471, 2001.
11. D. F. Wang, D. S. Yeung, E. C. Tsang. Ellipsoidal support vector clustering for functional MRI analysis. Pattern Recognition, 40(10):2685-2695, 2007.
12. X. Song, A. M. Wyrwicz. Unsupervised spatiotemporal fMRI data analysis using support vector machines. NeuroImage, 47:204-212, 2009.
13. H. F. Chen, D. Z. Yao, S. Becker, Y. Zhou, M. Zeng, L. Chen. A new method for fMRI data processing: Neighborhood independent component correlation algorithm and its preliminary application. Science in China Series, 45(5):373-382, 2002.
14. A. F. Sole, S. C. Ngan, G. Shapiro, X. P. Hu, A. Lopez. Anisotropic 2D and 3D averaging of fMRI signals. IEEE Transactions on Medical Imaging, 20(2):86-93, 2001.
15. H. Y. Kim, J. Giacomantone, Z. H. Cho. Robust Anisotropic Diffusion to Produce Enhanced Statistical Parametric Map. Computer Vision and Image Understanding, 99:435-452, 2005.
16. J. Yang, N. Zhong, P. Liang, J. Wang, Y. Yao, S. Lu. Brain activation detection by neighborhood one-class SVM. Cognitive Systems Reasearch, 11:16-24, 2010.
17. P. Perona, J. Malik. Scale space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(7):629-639, 1990.
18. M. J. Black, G. Shapiro, D. H. Marimont, D. Hegger. Robust anisotropic diffusion. IEEE Transactions on Image Processing, 7(3):421-432, 1998.
19. H. Y. Kim, J. Giacomantone. A New Technique to Obtain Clear Statistical Parametric Map By Applying Anisotropic Diffusion to fMRI. IEEE International Conference on Image Processing,724-727, 2005.
20. F. Voci, S. Eiho, N. Sugimoto, H. Sekiguchi. Estimating the gradient threshold in the perona-malik equation. IEEE Signal Processing Magazine,39-46, 2004.
21. J. Giacomantone, A. De Giusti. ROC performance evaluation of RADSPM technique. Argentinian Congress on Computer Science, 2008.
22. J. A. Sorenson, X. Wang. ROC Method for Evaluation of fMRI techniques. Magnetic Resonance in Medicine, 36:737-744, 1996.
23. P. Skudlarski, T. Constable, J. C. Gore. ROC Analysis of Statistical Methods Used in Functional MRI: Individual Subjects. Neuroimage, 9:311-329, 1999.