Spatially Regularized SVM for the Detection of Brain Areas Associated with Stroke Outcome

Rémi Cuingnet[1,2], Charlotte Rosso[1,3], Stéphane Lehéricy[1,4], Didier Dormont[1,4], Habib Benali[2], Yves Samson[1,3], and Olivier Colliot[1]

[1] Université Pierre et Marie Curie-Paris 6, CNRS UMR 7225, Inserm UMR_S 975, Centre de Recherche de l'Institut Cerveau-Moelle (CRICM), Paris, France
[2] Inserm, UMR_S 678, LIF, Paris, France
[3] AP-HP, Urgences Cérébro-Vasculaires, Pitié-Salpêtrière, Paris, France
[4] Neuroradiology department & CENIR, Pitié-Salpêtrière, Paris, France

Abstract. This paper introduces a new method to detect group differences in brain images based on spatially regularized support vector machines (SVM). First, we propose to spatially regularize the SVM using a graph encoding the voxels' proximity. Two examples of regularization graphs are provided. Significant differences between two populations are detected using statistical tests on the margins of the SVM. We first tested our method on synthetic examples. We then applied it to 72 stroke patients to detect brain areas associated with motor outcome at 90 days, based on diffusion-weighted images acquired at the acute stage (one day delay). The proposed method showed that poor motor outcome is associated to changes in the corticospinal bundle and white matter tracts originating from the premotor cortex. Standard mass univariate analyses failed to detect any difference.

1 Introduction

Diffusion-weighted imaging (DWI) is of considerable interest to the clinical evaluation of acute stroke patients [1]. The location of the lesions has been suggested to represent a better predictor than their global volume [2]. At the subacute or chronic phases, previous studies have shown that damages to the corticospinal tract (CST) [3] and lesions to the primary sensorimotor cortex [2,4] correlated with poor motor outcome. At the acute stage, regional changes in the apparent diffusion coefficients (ADC) were suggested as early quantitative indices of regional irreversible ischemic damage [5]. However, at the acute stage, the spatial pattern of ADC changes associated with motor outcome remains unclear.

Group analyses of differences between populations in brain imaging have widely relied on univariate voxel-wise analyses, such as voxel-based morphometry (VBM) for structural MRI [6] or their equivalent for diffusion imaging (VB-DWI). In such analyses, brain images are first spatially registered to a common stereotaxic space, and then mass univariate statistical tests are performed in each voxel to detect significant group differences. However, the sensitivity of theses approaches is limited when the differences are spatially complex and involve a

T. Jiang et al. (Eds.): MICCAI 2010, Part I, LNCS 6361, pp. 316–323, 2010.

combination of different voxels or brain structures [7]. Recently, there has been a growing interest in support vector machines (SVM) methods [8,9] to overcome the limits of these univariate analyses. Theses approaches allow capturing complex multivariate relationships in the data and have been successfully applied to the individual classification of a variety of neurological conditions [10,11,12,13]. Moreover, the output of the SVM can also be analyzed to localize spatial patterns of discrimination, for example by drawing the coefficients of the optimal margin hyperplane (OMH) – which, in the case of a linear SVM, live in the same space as the MRI data [12,13,14]. However, one of the problems with analyzing directly the OMH coefficients is that the corresponding maps are noisy and lack spatial coherence. Moreover only a few of these approaches perform a statistical analysis of the OMH coefficients [14].

In this paper, we propose a new method to detect group differences in brain images based on spatially regularized SVM. In particular, we show how spatial consistency can be directly enforced into the SVM by using Laplacian regularization. We then propose a statistical analysis based on the spatially regularized SVM to detect brain regions which are significantly different between two groups of subjects. The proposed framework is tested on 2D synthetic test images and then applied to the detection of differences between stroke patients with good and poor outcome based on DWI acquired at the acute stage.

2 Spatially Regularized SVM Using the Graph Laplacian

In this section, after some background on SVM, we propose to spatially regularize the SVM using a graph encoding the voxels' proximity. We then give two examples of regularization graphs.

2.1 Background

In this contribution, we consider the case of brain images which are spatially normalized to a common stereotaxic space. These images can be any characteristics extracted from the MRI, such as gray matter concentration maps (in VBM) or ADC maps (in diffusion MRI). Let $(\mathbf{x}_s)_{s\in[1,N]} \in (\mathbb{R}^d)^N$ be the images of N subjects and $(y_s)_{s\in[1,N]} \in \{\pm 1\}^N$ their group labels (e.g. diagnosis). SVMs search for the hyperplane for which the margin between groups is maximal, the OMH. The standard linear SVM solves the following optimization problem [8]:

$$\min_{\mathbf{w}\in\mathbb{R}^d, b\in\mathbb{R}} \underbrace{\sum_{s=1}^{N} \phi\left(y_s\left[\langle \mathbf{w}, \mathbf{x}_s\rangle + b\right]\right)}_{\text{Empirical Loss}} + \underbrace{\lambda \parallel \mathbf{w} \parallel^2}_{\text{Classical Tikhonov Regularization}} \quad (1)$$

where ϕ is the *hinge loss function* ($\phi : u \mapsto (1-u)^+$), b the bias and $\lambda \in \mathbb{R}^+$.

With a linear SVM, the *feature space* is the same as the *input space*. Thus, when the input features are the voxels of the image, each component of $\mathbf{w}$ also corresponds to a voxel. One can therefore represent the values of $\mathbf{w}$ in the image

space, and use this map to localize differences. However, the map $\mathbf{w}$ can be noisy and scattered (as for example in [12]). This is due to the fact that the regularization term of the standard linear SVM *is not a spatial regularization*. Voxel-based comparisons are subject to registration errors and interindividual variability. Gaussian smoothing is therefore often used as a preprocessing step. However, some image information is lost with the smoothing which, for example, mixes white matter with gray matter voxels. Tissue probability maps could be used to overcome this limitation. More generally, if voxels are connected, meaning for example spatially, anatomically or functionally close, we would like the SVM to consider them as similar. In the next section, we propose to encode this proximity in a graph and to use its Laplacian to spatially regularize the SVM.

2.2 Regularization Based on Diffusion on Graph

Graphs are a natural framework to take spatial information into consideration. Voxels of a brain image can be considered as nodes of a graph which models the voxels' proximity. This graph can be the voxel connectivity (6, 18 or 26) or a more sophisticated graph.

By changing the regularization term of the standard linear SVM equation (1), one can force $\mathbf{w}$ to be smooth with respect to the graph, hence spatially smooth. That is to say, if two voxels are close, the classifier will consider them as similar. The minimization problem becomes:

$$\min_{\mathbf{w}\in\mathbb{R}^d, b\in\mathbb{R}} \underbrace{\sum_{s=1}^{N} \phi\left(y_s\left[\langle \mathbf{w}, \mathbf{x}_s\rangle + b\right]\right)}_{\text{Empirical Loss}} + \underbrace{\lambda \parallel \exp\left\{\frac{1}{2}\beta L\right\} \mathbf{w} \parallel^2}_{\text{Spatial Regularization}} \tag{2}$$

where L is the graph Laplacian [15] and $\beta \in \mathbb{R}^+$ and $\exp\{\cdot\}$ is the matrix exponential[1] [16]. This new minimization problem (2) is equivalent to an SVM optimization problem [9,17]. The new kernel K is given by: $K(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^t e^{-\beta L}\mathbf{x}_2$. Note that $K(\mathbf{x}_1, \mathbf{x}_2) = \langle e^{-\frac{\beta}{2}L}\mathbf{x}_1, e^{-\frac{\beta}{2}L}\mathbf{x}_2\rangle$.

Our approach differs from the diffusion kernels introduced by Kondor et al. [18]. In our case, the nodes of the graph are the features, here the voxels, whereas in [18], the nodes were the objects to classify. Laplacian regularization was also used in satellite imaging [19] but, again, the nodes were the objects to classify. Our approach can also be considered as a spectral regularization on the graph [20]. To our knowledge, such spectral regularization has not been applied to brain images but only to the classification of microarray data [21].

2.3 Examples of Regularization Graphs

One has now to define the graph depending on type of spatial proximity one wants to enforce. The simplest option is to use the image connectivity (6, 18

[1] For any square matrix M, $\exp\{M\} = \sum_{k=0}^{\infty} \frac{1}{k!} M^k$

or 26). In this case, the regularized SVM would be equivalent to smoothing the data with a Gaussian kernel with standard deviation $\sigma = \sqrt{\beta}$ [20]. But this would mix gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Instead, we propose a graph which takes into consideration both the spatial localization and the tissue types. Based on tissue probability maps, in each voxel v, we have the set of probabilities p_v that this voxel belongs to GM, WM or CSF. We considered the following graph. Two voxels are connected if and only if they are neighbors in the image (6-connectivity). The weight $a_{u,v}$ of the edge between two connected voxels u and v is $a_{u,v} = e^{-d_{\chi^2}(p_u,p_v)^2/(2\sigma^2)}$, where d_{χ^2} is the χ^2-distance between two distributions. We chose beforehand σ equal to the standard deviation of $d_{\chi^2}(p_u, p_v)$.

3 Statistical Analysis of the Margins

In this section, we propose a statistical analysis to detect brain regions which are significantly different between two groups of subjects, based on the results of the spatially regularized SVM.

The classification function obtained with a linear SVM is the sign of the inner product of the features with $\mathbf{w}$, a vector orthogonal to the OMH [8,9]. Therefore, if the absolute value of the i^{th} component of the vector $\mathbf{w}$, $|w_i|$, is small compared to the other components $(|w_j|)_{j\neq i}$, the i^{th} feature will have a small influence on the classification. Conversely, if $|w_i|$ is relatively large, the i^{th} feature will play an important role in the classifier. However, one cannot compare directly the weights from two different comparisons. More precisely, let A, A', B, B' be four groups of subjects and let $\mathbf{w}^{(A)}$ and $\mathbf{w}^{(B)}$ be the optimal weights (SVM outputs) for the comparison A versus A' and the comparison B versus B' respectively. If the separation between A and A' is larger than the separation between B and B' then $\| \mathbf{w}^{(A)} \|$ will be smaller than $\| \mathbf{w}^{(B)} \|$. Hence, one cannot directly compare the components $|w^{(A)}{}_i|$ and $|w^{(B)}{}_i|$ for significance tests.

SVMs search for the hyperplane for which the margin [2] between groups is maximal (Fig. 1). The margin m is large when there is a large separation between two groups. By combining m and $\frac{|w_i|}{\|\mathbf{w}\|}$, one can simultaneously quantify the separation between groups and the relative influence of the different features. Therefore, we propose to analyze the statistic of $\frac{m|w_i|}{\|\mathbf{w}\|}$. We performed permutation tests on $\frac{m|w_i|}{\|\mathbf{w}\|}$ under the null hypothesis $\mathcal{H}_0$ of no relationship between the class labels and the global structure of the MR scan. By randomly permuting the subjects labels 20,000 times and training the SVM with this permutation of labels, we estimated for each voxel i the probability distribution of $\frac{m|w_i|}{\|\mathbf{w}\|}$ under $\mathcal{H}_0$. Based on these distributions, it is possible to test $\mathcal{H}_0$ at the voxel level. The false discovery rate (FDR) was used to correct for multiple comparisons [22]. To the best of our knowledge, other statistical analyses of the OMH did not take the margin into account (e.g. [14]).

[2] For the standard linear SVM (1), the margin m is given by: $m = 2 \| \mathbf{w} \|^{-1}$. As for the spatially regularized version (2), $m = 2 \| \exp\left(\frac{1}{2}\beta L\right) \mathbf{w} \|^{-1}$ [9].

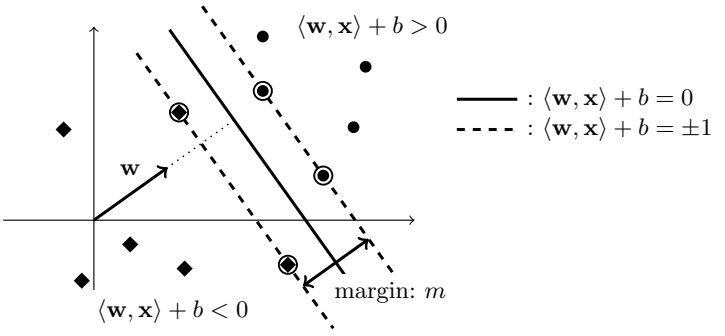


Fig. 1. Illustration of an optimal margin hyperplane with a linear SVM. The support vectors are circled.

4 Experiments and Results

We first tested our method on 2D synthetic test images. We then applied it on real data to the detection of brain areas associated with stroke outcome based on diffusion-weighted MRI acquired at the acute stage.

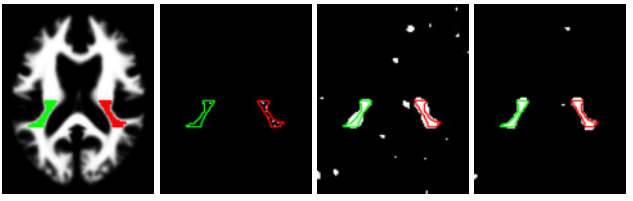
4.1 Synthetic Images

We first evaluated the ability of the method to detect artificial differences between two groups of 20 2D synthetic images (116×92 with 1.5 mm isotropic voxels) which were constructed as follows. We considered a slice of a WM template. For each of the 40 images, the voxels of the WM were assigned a random number between zero and one, the intensity of the other voxels being null. In each image of the first group, we constructed a hyperintensity h_{green} in the green region of Fig. 2 and h_{red} in the red region such as $(h_{\text{green}} + h_{\text{red}}) \sim \mathcal{N}(2, 0.2)$. Gaussian white noise ($\mathcal{N}(0, 1)$) was added to all images.

We tested six methods: three univariate methods and three SVM methods. We performed three univariate analyses on the voxel intensities: on the raw images, on the images smoothed with a Gaussian kernel and on the images preprocessed by $e^{-\frac{\beta}{2}L}$ (where L was the Laplacian of the graph used in the spatially regularized SVM). We tested three SVM methods: the standard linear SVM on the raw images, the standard linear SVM on the smoothed images and the spatially regularized SVM on the raw images.

All tests were corrected for multiple comparisons with a 5% FDR. The C parameter of the SVM was fixed to one ($\lambda = \frac{1}{2NC}$ [9]). The full width at half maximum (FWHM) of the Gaussian smoothing kernel was three voxels (the red and green regions' widths). β controls the size of the spatial regularization and was chosen to be equivalent to the FWHM of the Gaussian smoothing.

The results are shown on Fig. 2. The univariate analyses did not detect any difference for any images (raw, smoothed and preprocessed with $e^{-\frac{\beta}{2}L}$).The SVM detected only a fraction of the green and red regions on raw images. On smoothed



(a) (b) (c) (d)

Fig. 2. Synthetic example: (a) WM template and regions to detect; detection with a linear SVM (b) on raw images; (c) on smoothed images; (d) detection with a spatially regularized SVM on raw images. All other analyses detected no difference.

images both regions were detected. The spatially regularized SVM also detected both regions and decreased the number of scattered clusters.

4.2 Brain Areas Associated with Stroke Outcome

Subjects and MRI Acquisition. We included 72 consecutive acute stroke patients (mean age: 60±14 years[24-81]). Exclusion criteria were symptomatic hemorrhagic transformation on follow-up MRI and death during the follow-up period (90 days). The modified Rankin Scale (mRS) was used to assess outcome at 90 days. Good outcome was defined as independency (mRS 0 to 2; 39 subjects) and poor outcome as severe disability (mRS 3 to 5; 33 subjects).

The median delay between stroke onset and MRI aquisition was 1.2 day[3]. MR imaging was performed on a 1.5 Tesla MR General Electric Signa. Axial isotropic DWI spin echo EPI included 24 slices of 5 mm thickness, with an interslice gap of 0.5 mm, a 280x210 mm FOV, a 96x64 matrix, TE = 98.9 ms, and TR = 2825 ms. A baseline T2 acquisition and a diffusion-weighted acquisition using a diffusion gradient of 1000 s.mm^{-2} were both acquired within 40 seconds. ADC maps were generated using dedicated commercially-available software (Functool 2, General Electric). ADC maps were then normalized to the Montreal Neurological Institute (MNI) space using SPM5 software[4]. To avoid any effect from lateralization, the template was symmetrized. To put all the lesions on the same side, ADC maps with the infarct lesion in the left hemisphere were flipped to the right for the analysis.

Statistical Analysis and Results. Univariate analyses were done with both a permutation test and a parametric Student's T-test on smoothed images (8-mm FWHM Gaussian filter). We also performed analyses with the spatially regularized SVM on the raw images. As in the previous section, the β parameter was chosen to correspond to the FWHM of the univariate analyses. All tests were corrected with a 5% FDR.

[3] All imaging and clinical data were generated during routine clinical workup of the patients in our stroke center. The study was approved by the La Pitié-Salpêtrière Hospital Ethics Committee.

[4] Statistical Parametric Mapping, Institute of Neurology, London, UK

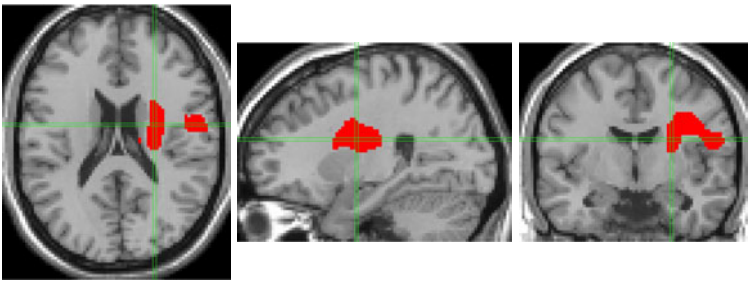


Fig. 3. Group differences of the ADC maps between poor and good motor outcome with a spatially regularized SVM (z=20 mm, x=28mm and y=-8mm in the MNI-space)

No regions were detected by the univariate analyses. The spatially regularized SVM detected significant changes predominantly localized in the WM (Fig. 3). The larger part of the detected cluster included the corticospinal tract at the level of the internal capsule. The smaller part was more superficial and was originating from the lower part of the motor and premotor cortex.

5 Conclusion

We proposed a new method based on spatially regularized SVM to detect group differences in brain images. Spatial consistency was directly enforced into the SVM by using the graph Laplacian. This provides a flexible approach to model different types of proximity between voxels. We then proposed to detect differences between groups using a statistical analysis which takes the margin of the SVM into account.

The proposed approach was applied to the detection of brain areas associated with stroke outcome based on DWI acquired at the acute stages. It allowed detecting changes localized in a large WM tract including the corticospinal tract and the lower part of the motor and premotor bundles. Univariate analyses failed to detect any differences. These results, obtained at the acute stage, are in line with previous studies carried out at the subacute or chronic phases [3,2,4]. Our results suggest that spatially regularized SVM might be useful to analyze MR images acquired in clinical routine as soon as 24 hours post stroke onset.

The proposed approach is not specific to diffusion MRI or stroke patients, and can be applied to other pathologies and other types of data (e.g. anatomical MRI). It has the potential to overcome the limits of traditional mass univariate voxel-wise analyses by detecting complex spatial patterns of alterations.

References

1. Chalela, J., et al.: Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. The Lancet 369(9558), 293–298 (2007)
2. Crafton, K., et al.: Improved understanding of cortical injury by incorporating measures of functional anatomy. Brain 126(7), 1650–1659 (2003)

3. Domi, T., et al.: Corticospinal tract pre-wallerian degeneration: A novel outcome predictor for pediatric stroke on acute MRI. Stroke 40(3), 780–787 (2009)
4. Lo, R., et al.: Identification of critical areas for motor function recovery in chronic stroke subjects using voxel-based lesion symptom mapping. NeuroImage 49(1), 9–18 (2010)
5. Rosso, C., et al.: Prediction of Infarct Growth Based on Apparent Diffusion Coefficients: Penumbral Assessment without Intravenous Contrast Material. Radiology 250(1), 184–192 (2009)
6. Ashburner, J., Friston, K.J.: Voxel-based morphometry–the methods. NeuroImage 11(6), 805–821 (2000)
7. Davatzikos, C.: Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. NeuroImage 23(1), 17–20 (2004)
8. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, Heidelberg (1995)
9. Schölkopf, B., Smola, A.J.: Learning with Kernels. MIT Press, Cambridge (2001)
10. Lao, Z., et al.: Morphological classification of brains via high-dimensional shape transformations and machine learning methods. NeuroImage 21(1), 46–57 (2004)
11. Fan, Y., et al.: COMPARE: classification of morphological patterns using adaptive regional elements. IEEE TMI 26(1), 93–105 (2007)
12. Klöppel, S., et al.: Automatic classification of MR scans in Alzheimer's disease. Brain 131(3), 681–689 (2008)
13. Vemuri, P., et al.: Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. NeuroImage 39(3), 1186–1197 (2008)
14. Mourão-Miranda, J., et al.: Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. NeuroImage 28(4), 980–995 (2005); Special Section: Social Cognitive Neuroscience
15. Chung, F.R.K.: Spectral Graph Theory, vol. 92. AMS (1992)
16. Golub, G.H., Van Loan, C.F.: Matrix computations. J. Hopkins Univ. Press, Baltimore (1996)
17. Smola, A.J., Schölkopf, B.: On a kernel-based method for pattern recognition, regression, approximation, and operator inversion. Algorithmica 22, 211–231 (1998)
18. Kondor, R.I., Lafferty, J.D.: Diffusion kernels on graphs and other discrete input spaces. In: Proc. International Conference on Machine Learning, pp. 315–322 (2002)
19. Gómez-Chova, L., et al.: Semi-supervised image classification with Laplacian support vector machines. IEEE Geo. Rem. Sens. Let. 5(3), 336–340 (2008)
20. Smola, A., Kondor, R.: Kernels and regularization on graphs. In: Schölkopf, B., Warmuth, M.K. (eds.) COLT/Kernel 2003. LNCS (LNAI), vol. 2777, p. 144. Springer, Heidelberg (2003)
21. Rapaport, F., et al.: Classification of microarray data using gene networks. BMC bioinformatics 8(1), 35 (2007)
22. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B. Met., 289–300 (1995)