

Clasificación binaria, desbalanceada y contextual de voxels asociados a series temporales

Javier Giacomantone, Armando De Giusti

Instituto de Investigación en Informática (III-LIDI),
Facultad de Informática - Universidad Nacional de La Plata - Argentina.
La Plata, Buenos Aires, Argentina.
{jog,degiusti}@lidi.info.unlp.edu.ar

Resumen. En este artículo se presenta un método computacional para clasificación de regiones 3D en función de sus características dinámicas. La clasificación de voxels atípicos se implementa en función de las series temporales asociadas a los mismos. El método opera en clasificación binaria, clases desbalanceadas y correlación espacial de las series asociadas a cada clase. El método propuesto utiliza máquinas de soporte vectorial y difusión anisotrópica robusta para detectar la estructura subyacente en los datos y clasificar los voxels correspondientes en cada clase. Se presentan resultados experimentales del método propuesto para datos de resonancia magnética funcional e imágenes de rango.

Palabras clave: Reconocimiento de patrones, Series temporales, Máquinas de soporte vectorial, Difusión anisotrópica, fMRI, TOF.

1. Introducción

El problema de segmentar una región en un volumen discreto está presente en diversas áreas de aplicación entre las que podemos mencionar imágenes de tomografía computada, imágenes de resonancia magnética (MRI), imágenes de resonancia magnética funcional (fMRI), imágenes de tiempo de vuelo (TOF), imágenes de ultrasonido, fusión de modalidades y procesamiento de datos multidimensionales discretos en general [1][2]. Las coordenadas de cada elemento de volumen, denominado voxel, normalmente un cubo, son discretas. Así como se puede generalizar un pixel en 2D a un voxel en 3D, también se puede generalizar el término voxel a un elemento de volumen dinámico o doxel en 4D. La cuarta variable independiente normalmente es el tiempo o la distancia. La Fig.1(a) muestra una estructura de datos 3D y un voxel en particular. La Fig.1(b) nos muestra una estructura 4D y un doxel en particular. En este artículo consideramos que todas las etapas de pre-procesamiento ya fueron realizadas tales como co-registro, correcciones de escala, normalización y deriva. Si bien facilita y resulta intuitivo plantear el problema como imágenes 3D o 4D, no necesariamente los datos considerados tienen que corresponder a imágenes. En general

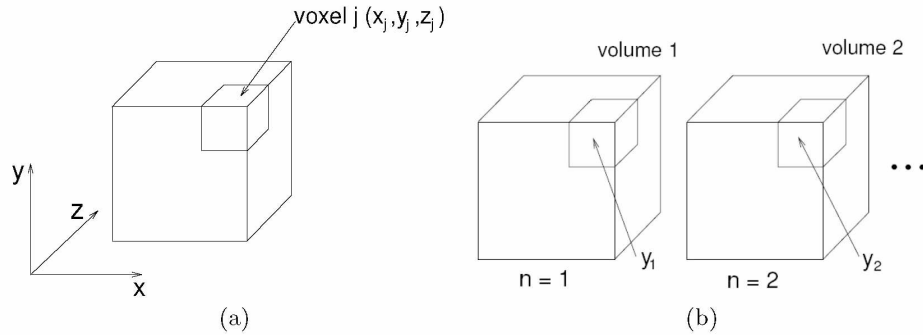


Figura 1. (a) voxel 3D, (b) doxel 4D

no son imágenes naturales, sino datos de distinto tipo visualizados como imágenes o series temporales para facilitar su interpretación, como datos de rango, temperatura o presión entre otros.

El método propuesto (MP) tiene como objetivo la clasificación de voxels en dos clases desbalanceadas. Los voxels de la clase minoritaria pueden considerarse patrones atípicos de acuerdo al nivel de desbalance. La segunda hipótesis del método es que existe correlación espacial entre voxels de cada clase. En particular el objetivo es detectar voxels correlacionados en la clase minoritaria. En este artículo se presentan resultados experimentales con imágenes funcionales TOF y fMRI. La modalidad de imágenes médicas denominada fMRI involucra considerar una imagen estructural 3D y una imagen funcional 4D [3][4]. La cámara TOF MESA SR4000 nos permite obtener secuencias de intensidad y rango [5]. Una imagen fMRI puede ser vista como un conjunto de series temporales donde cada serie, cada doxel, corresponde a un voxel particular en la imagen estructural [6][7]. Las máquinas de soporte vectorial de una clase “One class SVM (OCSVM)” [8][9] han sido también utilizadas en MRI y fMRI [10]. En fMRI del cerebro las series temporales de la mayoría de los voxels son independientes del estímulo experimental y solo pocas series temporales están relacionadas con el paradigma experimental utilizado. Estas últimas pueden ser consideradas como atípicas y por lo tanto satisfacen las hipótesis necesarias para aplicar OCSVM sobre el conjunto de datos fMRI [11]. La hipótesis de correlación espacial sostiene que las series temporales de voxels próximos tienen un estado similar de activación, correlacionado o no correlacionado con el estímulo experimental. Distintos métodos han sido propuestos para reflejar esta última característica de las señales fMRI [12][13][14]. El método propuesto, basado en asumir clases desbalanceadas y correlación espacial, es aplicable en fMRI y en imágenes de tiempo de vuelo. El artículo está organizado del siguiente modo: las secciones 2 y 3 presentan una breve revisión de los conceptos fundamentales utilizados en el método propuesto. La sección 4 describe el método propuesto. En la sección 5 se muestran resultados experimentales. Finalmente, en la sección 6 se presentan conclusiones.

2. Difusión Anisotrópica Robusta

Perona y Malik [15] definieron difusión anisotrópica como

$$\frac{\partial I(x, y, t)}{\partial t} = \text{div} [g(\|\nabla I(x, y, t)\|) \nabla I(x, y, t)], \quad (1)$$

utilizando la imagen original $I(x, y, 0) : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ como la condición inicial, donde t es un parámetro de tiempo artificial y g es una función que se denomina de parada en los bordes (“edge-stopping function”). La elección adecuada de g determina en que medida, al difundir, las discontinuidades son preservadas. Perona y Malik sugirieron utilizar dos posibles funciones con este objetivo [15]. Black et al. [16] utilizaron teoría de estimación robusta para determinar una mejor función de parada en los bordes denominada “Tukey’s biweight”.

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

La función anterior, g , es una versión dilatada y escalada de la función de Tukey, donde $g(0) = 1$ y el máximo local de su función de influencia $\psi(x) = xg(x)$ se localiza en $x = \sigma$. Cuando se utiliza la función de Tukey el proceso de difusión se denomina Difusión Anisotrópica Robusta (RAD).

Perona y Malik [15] propusieron una solución discreta para la ecuación de difusión anisotrópica robusta (1) dada por:

$$I(s, t + 1) = I(s, t) + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} g(|\nabla I_{s,p}(t)|) \nabla I_{s,p}(t), \quad (3)$$

donde $I(s, t)$ es una imagen muestreada, s es la posición del pixel en una grilla discreta 2-D o 3-D, $t \geq 0$ son los pasos temporales, la constante λ determina la difusión (normalmente, $\lambda = 1$), y η_s representa el conjunto de vecinos de s . Para imágenes 2-D son los cuatro vecinos y para 3-D los seis vecinos del voxel correspondiente, excepto en las fronteras. La magnitud del gradiente en un voxel en la iteración t es aproximada por:

$$\nabla I_{s,p}(t) = I(p, t) - I(s, t), \quad p \in \eta_s. \quad (4)$$

Utilizando la función de Tukey, el proceso de difusión converge mas rápido y genera bordes mas agudos que usando las funciones de Perona y Malik. Black et al. [16] mediante elementos de teoría de estimación robusta, obtienen un valor estimado de la escala robusta σ_e dado por:

$$\sigma_e = 1,4826 \text{ MAD}(\nabla I) = 1,4826 \text{ median}_I [|\|\nabla I\| - \text{median}_I(\|\nabla I\|)|], \quad (5)$$

donde MAD es la desviación absoluta media.

3. Máquinas de soporte vectorial

Dados l patrones de entrenamiento para dos clases $\{(x_i, y_i), i = 1, \dots, l\}$, donde $x_i \in \mathbb{R}^N$ es un patron, un vector de características de dimensión N e $y_i \in \{1, -1\}$ es la etiqueta de la clase a la que pertenece x_i . Decimos que $\mathbf{x}_1, \dots, \mathbf{x}_l \in \mathcal{X}$ donde $l \in \mathbb{N}$ es el número de observaciones y \mathcal{X} es un subconjunto compacto de \mathbb{R}^N . Una SVM determina el hiperplano que maximiza la separación entre ambas clases y es matemáticamente equivalente a maximizar el margen dado por $\frac{2}{\|\mathbf{w}\|}$ sujeto a $y_i(x_i \cdot \mathbf{w}) + b \geq 1$. El hiperplano $(\mathbf{w} \cdot \mathbf{x}) + b = 0$ queda determinado por b y \mathbf{w} . Se puede relajar la condición de muestras de entrenamiento separables incluyendo variables artificiales que permitan mejorar también la generalización. Para obtener el hiperplano minimizamos

$$\begin{aligned} \text{mín} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \\ \text{sujeto a} \quad & y_i[(\mathbf{w} \cdot \mathbf{x}_i) + b] \geq 1 - \xi_i \end{aligned} \tag{6}$$

donde C controla el compromiso entre complejidad del hiperplano y el error sobre el conjunto de entrenamiento. Resolviendo el Lagrangiano en la forma dual obtenemos la función de decisión como:

$$f(\mathbf{x}) = \text{sgn}\left[\sum_i \alpha_i y_i (\mathbf{x} \cdot \mathbf{x}_i) + b\right]$$

donde α_i son los multiplicadores de lagrange. Si las clases no son linealmente separables es posible extender el método proyectando los patrones en un espacio de mayor dimensión donde si sea posible clasificarlos linealmente, para luego obtener la función de decisión no lineal en el espacio original. Pueden ser utilizados distintos tipos de kernels que corresponden a estimadores no lineales en el espacio de entrada. En particular en este trabajo es utilizado un kernel de base radial $k(x, x_i) = e^{-\gamma \|x - x_i\|^2}$, γ es el parámetro que determina el ancho del kernel.

Han sido propuestos dos algoritmos basados en SVM para operar con una clase, “support vector data description” [8] y “one-class SVM (OCSVM)” [9]. El algoritmo estima una función de decisión f que es positiva para un subconjunto del espacio de características y negativa para el complemento. El mapeo de los datos al espacio de características corresponde al kernel y separa los datos del origen con máximo margen. Sea $\Phi : \mathcal{X} \rightarrow \mathcal{F}$ un mapeo de características en \mathcal{F} de tal modo que el producto interno en la imagen de Φ puede ser calculada evaluando un kernel simple.

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

Se puede reescribir como un problema de optimización.

$$\begin{aligned} \text{mín}_{\mathbf{w} \in \mathcal{F}, \xi \in \mathbb{R}^l, \rho \in \mathbb{R}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\nu l} \sum_i \xi_i - \rho \\ \text{sujeto a} \quad & (\mathbf{w} \cdot \Phi(\mathbf{x}_i)) \geq \rho - \xi_i, \xi_i \geq 0, \end{aligned} \tag{7}$$

donde $\nu \in (0, 1]$ es un parámetro que controla el término de penalización y ξ_i son variables de holgura. Resolviendo el problema de optimización (7) obtenemos \mathbf{w} y ρ y la función de decisión es -1 para valores atípicos en el conjunto de datos y +1 para el resto de los patrones en el conjunto de datos.

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w} \cdot \Phi(\mathbf{x})) - \rho \quad (8)$$

Introduciendo multiplicadores de Lagrange $\alpha_i, \beta_i \geq 0$, obtenemos

$$L(\mathbf{w}, \xi, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\nu l} \sum_i \xi_i - \sum_i \beta_i \xi_i - \rho - \sum_i \alpha_i (\mathbf{w} \cdot \Phi(\mathbf{x}_i) - \rho + \xi_i)$$

Fijando las derivadas con respecto a las variables primarias \mathbf{w}, ξ, ρ igual a cero obtenemos

$$\begin{aligned} \mathbf{w} &= \sum_i \alpha_i \Phi(\mathbf{x}_i), \\ \alpha_i &= \frac{1}{\nu l} - \beta_i \leq \frac{1}{\nu l}, \\ \sum_i \alpha_i &= 1. \end{aligned}$$

La función de decisión puede ser escrita como

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

Los multiplicadores α_i se puede resolver en el problema dual:

$$\begin{aligned} \underset{\alpha}{\text{mín}} \quad & \frac{1}{2} \sum_{ij} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{sujeto a} \quad & 0 \leq \alpha_i \leq \frac{1}{\nu l}, \quad \sum_i \alpha_i = 1. \end{aligned}$$

El parámetro ρ puede ser obtenido notando que para cualquier α_i y el correspondiente patron \mathbf{x}_i se satisface que:

$$\rho = (\mathbf{w} \cdot \Phi(\mathbf{x}_i)) = \sum_j \alpha_j k(\mathbf{x}_i, \mathbf{x}_j). \quad (9)$$

4. Método Propuesto

Sea I' una imagen funcional. El primer paso consiste en remover el valor medio de I' , obteniendo así:

$$I = I' - \bar{I} \quad (10)$$

y de este modo se evita difundir información estructural. Las series temporales correspondientes a cada voxel son consideradas como un patron. Los $\mathbf{x}_{i,j,k}$ son los patrones que corresponden la fila i , la columna j y la imagen k , identificando

una serie temporal particular o un punto del espacio de características. Los $\mathbf{x}_{i,j,k}$ son directamente los valores de entrada del problema de optimización (7). La solución óptima dada por \mathbf{w} y ρ se puede obtener resolviendo el problema dual y (9). Luego para cada $\mathbf{x}_{i,j,k}$ un valor de decisión primario $\mathbf{y}_{i,j,k}$ es obtenido,

$$\mathbf{y}_{i,j,k} = (\mathbf{w} \cdot \Phi(\mathbf{x}_{i,j,k})) - \rho \quad (11)$$

que representa la distancia entre un punto $\Phi(\mathbf{x}_{i,j,k})$ y un hiperplano en el espacio de mayor dimensión que el kernel mapea $(\mathbf{w} \cdot \Phi(\mathbf{x})) - \rho = 0$.

Sea la imagen funcional en la iteración $t \geq 0$ del proceso de difusión, $I(s, n, t)$, donde $I(s, n, 0)$ es la imagen inicial con la corrección de valor medio ya aplicada en la posición correspondiente al voxel s y volumen n , y $T(s, t)$ el mapa de activación formado a partir de $\mathbf{y}_{i,j,k} \forall i, j, k$ donde s es una posición particular i, j, k .

1. Sea $t \leftarrow 0$.
2. Calcular el mapa de activación $T(s, 0)$ utilizando OCSVM (11).
3. Calcular los coeficientes de difusión. El coeficiente de difusión entre un voxel s y su voxel vecino p en el instante t es:

$$g(|\nabla T_{s,p}(t)|), \text{ donde } \nabla T_{s,p}(t) = T(p, t) - T(s, t). \quad (12)$$

4. Con estos coeficientes aplicar difusión en $I(s, n, t)$, generando la imagen funcional difundida, $I(s, n, t + 1)$, en la iteración $t + 1$:

$$I(s, n, t + 1) \leftarrow I(s, n, t) + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} g(|\nabla T_{s,p}(t)|) \nabla I_{s,p}(t), \quad (13)$$

donde $\nabla I_{s,p}(n, t) = I(p, n, t) - I(s, n, t)$.

5. Asignar $t \leftarrow t + 1$ y repetir los pasos 2 a 5 un número predefinido de veces o hasta que el valor medio de los valores difundidos (segundo término de la ecuación (13)) se encuentre por debajo de un determinado umbral.
6. Clasificar cada voxel aplicando la función de decisión dada por la ecuación (8).
7. Considerar los resultados los resultados anteriores como prototipos, conjunto de patrones de entrenamiento para un clasificador SVM de dos clases.
8. Re-clasificar todos los patrones del experimento con la máquina de soporte vectorial final obtenida.

La difusión anisotrópica es controlada por el número de iteraciones y el parámetro de escala de la función de parada (2), σ . La escala óptima en el método propuesto es entre dos y tres veces mayor que la escala robusta estimada. Tsiot-sios y Petrou ([17]) analizan la elección adecuada de parámetros para imágenes en general.

5. Resultados Experimentales

Para evaluar el método propuesto utilizamos imágenes fMRI e imágenes TOF. Se generaron dos imágenes sintéticas basadas en la imagen artificial (IA) propuesta en [18], con diferentes niveles de activación. El modelo fMRI utilizado tiene las siguientes características 10x10x3 voxels por volumen y 84 volúmenes. Todos los voxels tienen sumado ruido gaussiano de media 16000 y desvío estándar 4000. El valor de los voxels activados ha sido incrementado entre 600 y 1500 con incrementos de 100 generando así 10 imágenes sintéticas fMRI denominadas IA1 a IA10. Otras 10 imágenes sintéticas fueron generadas con ruido Rayleigh denominadas IA11 a IA20. El experimento fMRI que la imagen sintética utilizada representa está formado por bloques alternados de seis volúmenes activados y seis no activados. Los volúmenes activados tienen una región de 6x6x3 voxels activada y una de 2x2x3 voxels no activada. La Fig. 2 muestra una imagen 2-D activada correspondiente a la IA, la imagen de activación de referencia y el mapa de activación sin segmentar aplicando en método propuesto.

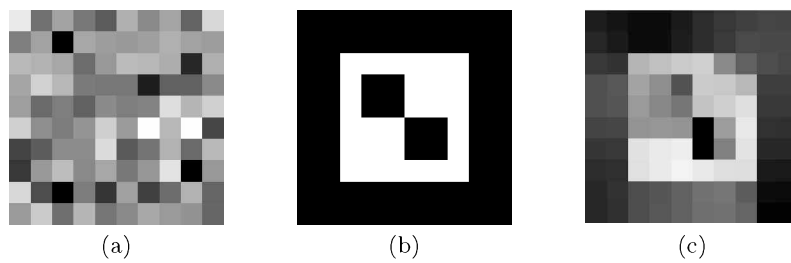


Figura 2. (a) IA-activada, (b) Imagen de referencia, (c) Mapa de activación

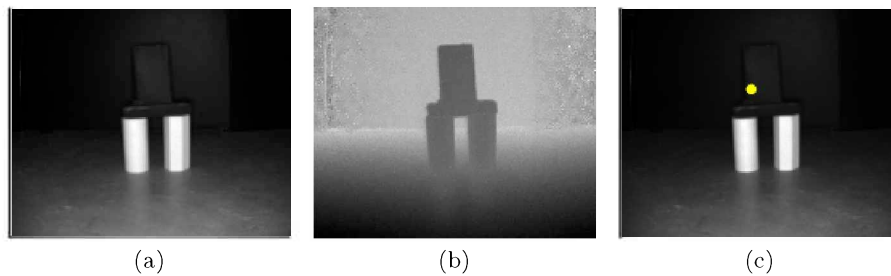


Figura 3. TOF

Se generaron 15 imágenes TOF semi-sintéticas sobre imágenes de intensidad y secuencias de 60 imágenes de rango correspondientes. Se utilizó una cámara ME-

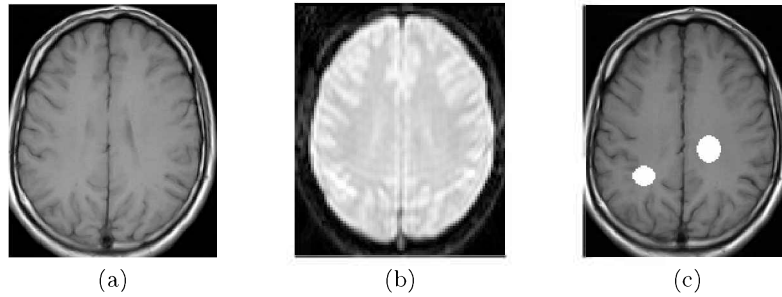


Figura 4. fMRI con activación controlada

SA SwissRanger SR 4000, generando un paradigma experimental en bloques de 3 imágenes de 144x176 pixels. La Fig. 3 muestra una imagen de intensidad, una imagen de rango y el área generada artificialmente sobre la imagen de intensidad. En esta primer etapa de evaluación también fue generada una imagen semi-sintética, denominada ISSf, con dos estímulos esféricos artificiales sobre una imagen fMRI que después de ser interpolada, re-alineada y normalizada tiene dimensiones 79x95x68 voxels y una estructura en bloques como la de las imágenes totalmente sintéticas propuestas. La Fig. 4 presenta tres imágenes, un corte de la imagen MRI estructural de referencia (157x189x136) 4(a), denominada ISSe, un corte de la imagen funcional 4(b) y una representación del estímulo generado artificialmente sobre la imagen estructural 4(c).

La Fig. 5 muestra curvas ROC (“Receiving operating Characteristics”). Las curvas permiten comparar el método propuesto para el modelo fMRI artificial en tres métodos para IA4 y en dos para IA8, con menor relación señal a ruido, en las figuras 5(a) y 5(b). La curva ROC de la figura 5(c) corresponde al experimento TOF dinámico correspondiente a la Fig. 2. En la Fig. 5(d) las pruebas son realizadas sobre la imagen ISSf mostrando las curvas ROC correspondientes a MP y OCSVM. La tabla 1 presenta los parámetros correspondientes al punto óptimo de operación para las pruebas realizadas sobre IA8, ISSf y TOF.

6. Conclusiones

En este artículo se ha presentado un método alternativo de post-procesamiento de imágenes funcionales utilizando máquinas de soporte vectorial y difusión anisotrópica. La detección de regiones activadas mejora como consecuencia de dos mecanismos: la capacidad de detectar valores atípicos y la capacidad de reconocer la presencia espacio-temporal no aislada de los mismos. El análisis presentado muestra resultados que permiten considerar el método propuesto como viable y comparable con métodos alternativos. Han sido planificadas dos etapas futuras. En primer término analizar el comportamiento en distinto tipo de imágenes funcionales y paradigmas experimentales. Una segunda etapa involucra evaluar las

Modalidad	Método	Area	d_{po}	TPF_{po}	FPF_{po}
fMRI IA8	$OC SVM_{\nu=0,37}$	0.9615	0.5685	0.9521	0.1483
	$MP_{\sigma=1,8,t=3,\nu=0,37}$	0.9789	0.6275	0.95679	0.0691
fMRI ISSf	$OC SVM_{\nu=0,4}$	0.9659	0.5812	0.8914	0.0694
	$MP_{\sigma=2,t=3,\nu=0,4}$	0.9835	0.6343	0.9452	0.0481
TOF	$OC SVM_{\nu=0,38}$	0.9665	0.5821	0.8984	0.0684
	$MP_{\sigma=1,9,t=3,\nu=0,38}$	0.9866	0.6545	0.9552	0.0475

Tabla 1. Métricas de rendimiento

condiciones de ruido y variaciones entre los distintos tipos de imágenes funcionales.

Referencias

1. J.A Noble. Ultrasound image segmentation and tissue characterization. Journal of Engineering in Medicine, vol. 223, pp. 1-10, 2009.
2. A. Salomon et. al. Simultaneous Reconstruction of Activity and Attenuation for PET/MR. IEEE Transactions on Medical Imaging, 30(3):804-813, 2011.
3. S. Ogawa et. al. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. Biophysics Journal, 14(3):803-812, 1993.
4. N. Lange et. al. Plurality and Resemblance in fMRI Data Analysis. NeuroImage 10:282-303, 1999.
5. F. Chiabrando, D. Piatti, F. Rinaudo. R-4000 TOF Camera: Further Experimental Tests and First Applications to Metric Surveys. Proceedings of V Symposium on Remote Sensing and Spatial Information Sciences, 149-154, 2010.
6. B. Ardekani, I. Kanno. Statistical Methods for Detecting Activated Regions in Functional MRI of the Brain. Magnetic Resonance Imaging 16(10):1217-1225, 1998.
7. 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.
8. D. M. J. Tax, R. P. W. Duin. Support vector data description. Machine Learning, 54:45-66, 2004.
9. 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.
10. J. Zhang, K. Ma, M. H. Er, V. Chong. Tumor segmentation from magnetic resonance imaging by learning via one-class support vector machine. Proc. 2004 Int. Conf. Intell. Mechatronics Automation, 207-211, 2004
11. X. Song, A. M. Wyrwicz. Unsupervised spatiotemporal fMRI data analysis using support vector machines. NeuroImage, 47:204-212, 2009.
12. 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.

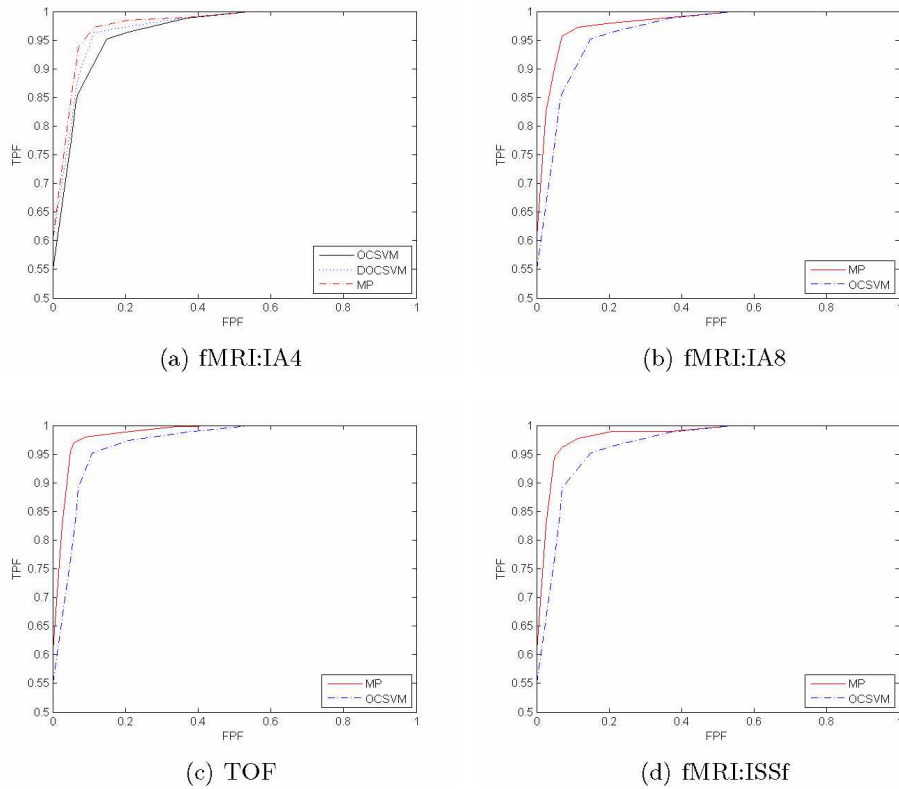


Figure 5. Curvas ROC

13. 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.
14. J. Yang, N. Zhong, P. Liang, J. Wang, Y. Yao, S. Lu. Brain activation detection by neighborhood one-class SVM. *Cognitive Systems Research*, 11:16-24, 2010.
15. 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.
16. M. J. Black, G. Shapiro, D. H. Marimont, D. Hegger. Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3):421-432, 1998.
17. C. Tsitsios, M. Petrou. On the choice of the parameters for anisotropic diffusion in image processing. *Pattern Recognition*, 46:1369-1381, 2013.
18. 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.