

Detección de áreas de interés bajo la hipótesis de relación espacial de voxels activados en fMRI

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. Las imágenes de resonancia magnética funcional (fMRI) utilizan una serie de imágenes de resonancia magnética para mapear de forma no invasiva las áreas de actividad neuronal aumentada del cerebro humano. La baja relación señal a ruido (SNR) de las imágenes funcionales, hace necesario el uso de técnicas de procesamiento de imágenes específicas, para la detección regiones correlacionadas con la respuesta a un estímulo determinado. En este artículo se presenta un método alternativo para segmentar regiones activadas en imágenes de fMRI. Se propone abordar el problema en dos etapas de clasificación, una no supervisada y una segunda etapa supervisada. El método propuesto utiliza máquinas de soporte vectorial (SVM) y difusión anisotrópica (DA) para la generación de patrones de entrenamiento, y SVM para la clasificación de regiones activadas. La aplicación del método propuesto permite incluir valiosa información con respecto a la interrelación entre las series temporales correspondientes a cada elemento de volumen (vóxel) en un espacio 3-D.

Palabras clave: Reconocimiento de patrones, Resonancia magnética funcional, Máquinas de soporte vectorial, Difusión anisotrópica, Procesamiento de señales

1 Introducción

Las imágenes de Resonancia Magnética Funcional (fMRI) utilizan imágenes de resonancia magnética (MRI) para mapear las áreas de actividad neural aumentada del cerebro humano. fMRI es una técnica muy útil para investigar respuestas específicas del cerebro y ha sido empleada para estudiar, entre otras actividades, experimentos basados en percepción, estímulos visuales, auditivos, aprendizaje de idiomas y memoria. En 1990 el primer experimento fue realizado con un estímulo visual usando como contraste un agente externo, no mucho tiempo después la mayoría de los experimentos estuvieron basados en un nuevo descubrimiento relacionado con el nivel de oxígeno en la sangre [1][2]. La hemoglobina es un contraste natural, tiene propiedades magnéticas diferentes dependiendo del nivel de oxigenación y por lo tanto afecta la señal obtenida a través de fMRI.

Esto hace posible disponer de una técnica no invasiva. Mediante la adquisición de fMRI cuando el paciente realiza una tarea específica, y cuando esta en reposo o realizando una tarea alternativa, es posible obtener una imagen funcional del cerebro. Los experimentos pueden ser clasificados en dos paradigmas principales denominados en bloques, o relacionados a eventos [3]. Una imagen fMRI puede ser vista como un conjunto de series temporales donde cada serie corresponde a un voxel particular. La clasificación de series temporales constituye el principal tópico de trabajo para el análisis de fMRI del cerebro humano. Las principales técnicas propuestas para análisis de datos en fMRI pueden clasificarse en dos categorías principales, orientadas, centradas en un determinado modelo [4][5], o las orientadas a los datos [6][7]. Los métodos guiados por los datos utilizan un método de aprendizaje automático o estadístico para analizar las series temporales, los guiados por un modelo asumen un modelo relacionado con la estructura y función del cerebro. Las máquinas de soporte vectorial son métodos basados en los datos y han sido aplicados a la clasificación de estados cognitivos [8], optimizando el margen entre clases y utilizando kernels en casos no lineales [9]. Las máquinas de soporte vectorial de una clase “One class SVM (OCSVM)” [10][11] han sido también utilizadas en MRI y fMRI [12]. 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 [13]. 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 [14][15][16]. En este artículo se presenta un método alternativo que tiene en consideración los dos factores característicos antes mencionados utilizando máquinas de soporte vectorial y difusión anisotrópica robusta.

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 las conclusiones.

2 Máquinas de soporte vectorial

Las máquinas de soporte vectorial son métodos originalmente propuestos para clasificación supervisada y posteriormente extendidos para operar en forma no supervisada o semi-supervisada. 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 patrón, 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

$(w \cdot x) + b = 0$ queda determinado por b y 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

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \\ \text{sujeto a } & y_i[(w - x_i) + b] \geq 1 - \xi_i \end{aligned} \quad (1)$$

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:

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

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” [10] y “one-class SVM (OCSVM)” [11]. Esto nos permite abordar problemas de clasificación de clases minoritarias, objetos no característicos en un conjunto de datos. 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.

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

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 (2) 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.

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w} \cdot \Phi(\mathbf{x})) - \rho \quad (3)$$

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:

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

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

$$\rho = (\mathbf{w} \cdot \Phi(\mathbf{x}_i)) = \sum_j \alpha_j k(\mathbf{x}_i, \mathbf{x}_j). \quad (4)$$

3 Difusión Anisotrópica Robusta

Perona y Malik [17] definieron difusión anisotrópica como

$$\frac{\partial I(x, y, t)}{\partial t} = \text{div} [g(\|\nabla I(x, y, t)\|) \nabla I(x, y, t)], \quad (5)$$

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 [17]. Black et al. [18] utilizaron teoría de estimación robusta para determinar una mejor función de parada en los bordes denominada “Tukey’s biweight”.

$$g(x) = \begin{cases} \left[1 - \frac{x^2}{5\sigma^2}\right]^2, & \frac{x^2}{5} \leq \sigma^2 \\ 0, & \text{en otro caso} \end{cases} \quad (6)$$

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 [17] propusieron una solución discreta para la ecuación de difusión anisotrópica robusta (5) dada por:

$$I(s, t + 1) = I(s, t) + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} g(|\nabla I_{s,p}(t)|) \nabla I_{s,p}(t), \quad (7)$$

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:

$$\nabla I_{s,p}(t) = I(p, t) - I(s, t), \quad p \in \eta_s. \quad (8)$$

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. [18] mediante elementos de teoría de estimación robusta, obtienen un valor estimado de la escala robusta σ_e dado por:

$$\sigma_e = 1.4826 \text{MAD}(\nabla I) = 1.4826 \text{median}_I [|\|\nabla I\| - \text{median}_I(\|\nabla I\|)|], \quad (9)$$

donde MAD es la desviación absoluta media.

4 Método Propuesto

Sea I' una imagen de resonancia magnética funcional. El primer paso de pre-procesamiento simple consiste en remover el valor medio de I' , obteniendo la imagen fMRI I :

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

Este primer paso es muy importante porque solo la información correspondiente a áreas activadas debe ser difundida, y no 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 (2). La solución optima dada por \mathbf{w} y ρ se puede obtener resolviendo el problema dual y (4). 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 de fMRI en la iteración $t \geq 0$ del proceso de difusión, $I(s, n, t)$, donde $I(s, n, 0)$ es la imagen fMRI 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 fMRI 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 (3).
7. Considerar 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 (6), σ . La escala óptima en el método propuesto es entre dos y tres veces mayor que la escala robusta estimada. Tsotsios y Petrou ([19]) analizan la elección adecuada de parámetros para imágenes en general.

5 Resultados Experimentales

La adquisición de imágenes fMRI es costosa, sujeta a condiciones de baja relación de señal a ruido y requiere técnicas de pre-procesamiento específicas. La necesidad de disponer de imágenes de referencia (“gold standard”) para comparar y

evaluar métodos en condiciones controladas sugiere, además de imágenes reales, también utilizar dos modelos sintéticos de imágenes fMRI. El primer modelo consiste en un conjunto de datos fMRI totalmente artificiales que reproduzcan lo mas fielmente posible las condiciones reales pero con regiones activadas conocidas. Otro modelo posible consiste en regiones activadas artificiales sobre una imagen fMRI real. En este artículo presentamos resultados preliminares utilizando imágenes sintéticas y semi-sintéticas que permiten evaluar el método propuesto con respecto a las hipótesis bajo las cuales fue desarrollado y compararlo con respecto a métodos alternativos. Se generaron dos imágenes sintéticas basadas en la imagen artificial propuesta en [20], 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 formada 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.1 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.

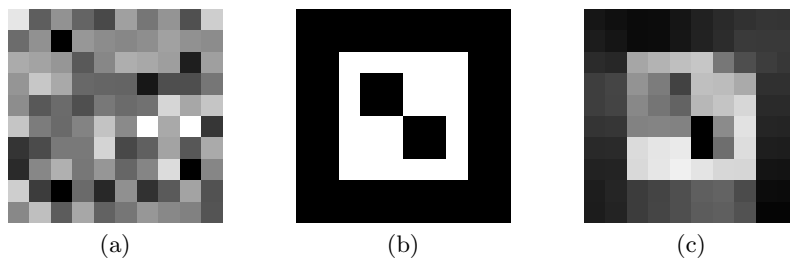


Fig. 1. (a) IA-activada, (b) Imagen de referencia, (c) Mapa de activación

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.2 presenta tres imágenes, un corte de la imagen MRI estructural de referencia (157x189x136) 2(a), denominada ISSe, un corte de la imagen funcional 2(b) y una representación del estímulo generado artificialmente sobre la imagen estructural 2(c).

La Fig.3 muestra curvas ROC (“Receiving operating Characteristics”) [21][22]. Las curvas permiten comparar el método propuesto (MP), OCSVM y DOCSVM

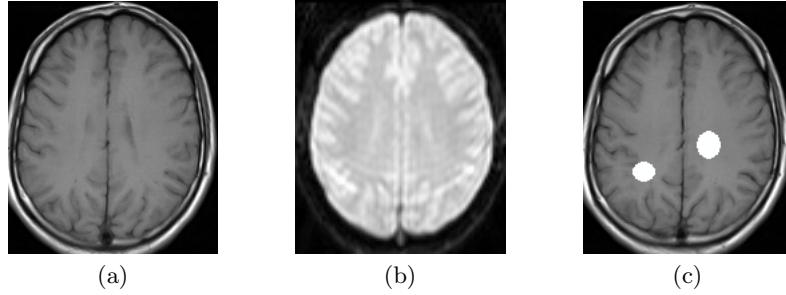


Fig. 2. fMRI con activación controlada

en los modelos IA2 e IA4 en las figuras 3(a) y 3(b). En la figura 3(c) se comprara OCSVM y el MP con mayor SNR que en 3(a) y 3(b) para IA7. En la fig. 3(d) las pruebas son realizadas sobre la imagen ISSf mostrando las curvas ROC correspondientes a MP y OCSVM. La tabla 1 presenta para las pruebas realizadas sobre IA7 e ISSf, los valores numéricos de las principales métricas para el punto de operación óptimo de cada curva ROC.

fMRI	Método	Area	d_{po}	TPF_{po}	FPF_{po}
IA7	$OCSVM_{\nu=0.37}$	0.9645	0.5687	0.9524	0.1481
	$MP_{\sigma=1.8,t=3,\nu=0.37}$	0.9801	0.6277	0.9571	0.0694
ISSf	$OCSVM_{\nu=0.4}$	0.9659	0.5812	0.8914	0.0694
	$MP_{\sigma=2,t=3,\nu=0.4}$	0.9835	0.6343	0.9452	0.0481

Tabla 1. Métricas de rendimiento

6 Conclusiones

En este artículo se ha presentado un método alternativo de post-procesamiento en fMRI utilizando máquinas de soporte vectorial y difusión anisotrópica. Este método, basado en dos hipótesis fundamentales sobre las características particulares de los datos, muestra resultados preliminares favorables. 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. A pesar de los resultados alentadores que el método presenta, han sido planificadas dos etapas futuras. En primer término analizar sobre

datos sintéticos el ajuste automático o semiautomático de los parámetros del método propuesto y especificar los intervalos de confianza experimentales. Una segunda etapa mas compleja contempla validar en datos fMRI reales paradigmas en bloques y relacionado a eventos.

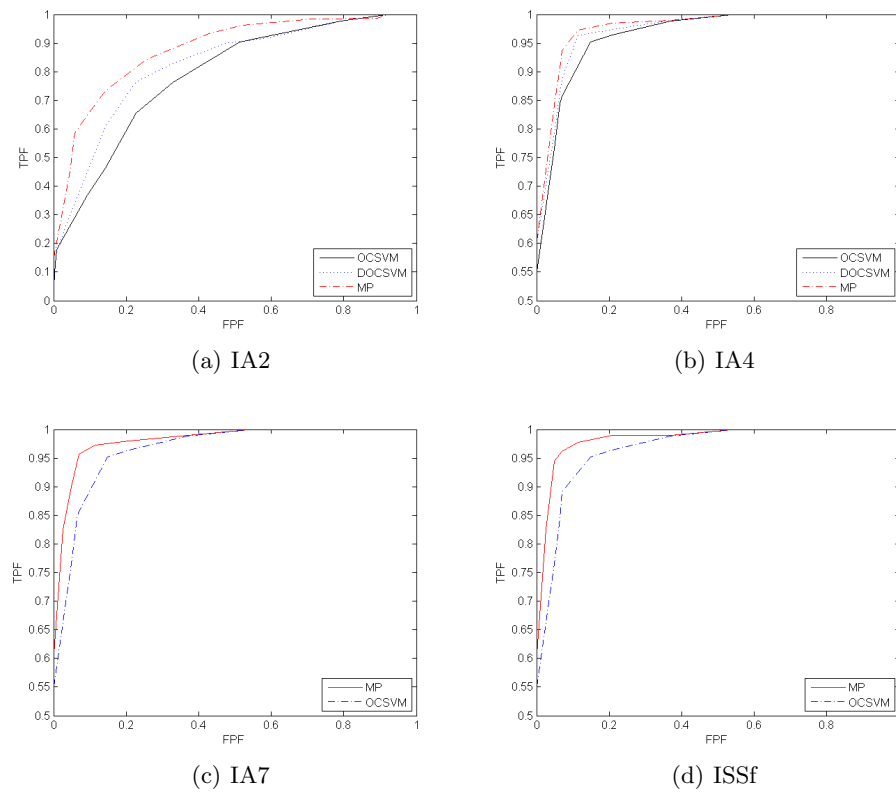


Fig. 3. Curvas ROC

Referencias

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. N. Lange et. al. Plurality and Resemblance in fMRI Data Analysis. *NeuroImage* 10:282-303, 1999.
3. B. Ardekani, I. Kanno. Statistical Methods for Detecting Activated Regions in Functional MRI of the Brain. *Magnetic Resonance Imaging* 16(10):1217-1225, 1998.
4. 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.
5. 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.
 6. C. Goutte, P. Toft, E. Rostrup, F. Nielsen, L. Hansen. On clustering fMRI time series. *NeuroImage*, 9:298-310, 1999.
 7. 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.
 8. 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.
 9. V. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988-999, 1999.
 10. D. M. J. Tax, R. P. W. Duin. Support vector data description. *Machine Learning*, 54:45-66, 2004.
 11. 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.
 12. 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
 13. X. Song, A. M. Wyrwicz. Unsupervised spatiotemporal fMRI data analysis using support vector machines. *NeuroImage*, 47:204-212, 2009.
 14. 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.
 15. 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.
 16. 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.
 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. C. Tsotsios, M. Petrou. On the choice of the parameters for anisotropic diffusion in image processing. *Pattern Recognition*, 46:1369-1381, 2013.
 20. 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.
 21. P. Skudlarski, T. Constable, J. C. Gore. ROC Analysis of Statistical Methods Used in Functional MRI: Individual Subjects. *Neuroimage*, 9:311-329, 1999.
 22. A. Bradley. ROC curve equivalence using the Kolmogorov-Smirnov test. *Pattern Recognition Letters*, 34:470-475, 2013.