

DISCRIMINACIÓN AUTOMÁTICA DE VERTIDOS DE HIDROCARBUROS A PARTIR DE IMÁGENES ASAR UTILIZANDO UNA MÁQUINA DE VECTORES SOPORTE (SVM)

L. González, J. M. Torres, N. Yarovenko y J. Martín.

*Laboratorio de Teledetección y SIG, Departamento de Física Aplicada, Universidad de Vigo.
Facultad de Ciencias del Mar, Campus Lagoas-Marcosende. 36310. Vigo. Pontevedra. luisgv@uvigo.es*

RESUMEN

La detección de vertidos de hidrocarburos en áreas oceánicas en imágenes de apertura sintética (SAR) es una operación complicada debido a la presencia de otros fenómenos que dan lugar a firmas similares a la de los vertidos, por lo que se han propuesto diferentes sistemas automáticos o semiautomáticos de detección que permitan distinguir los vertidos reales. En este trabajo se presenta un clasificador basado en un método de aprendizaje supervisado denominado Máquina de Vectores Soporte (SVM). Para el desarrollo del algoritmo se han utilizado 26 imágenes ENVISAT ASAR adquiridas durante la marea negra causada por el petrolero Prestige a finales del 2002 y que afectó a la costa noroeste de España. Estas imágenes presentan no solamente gran número de vertidos sino también gran cantidad de falsas alarmas. Para poder entrenar y validar el clasificador fue necesario llevar a cabo una categorización *a priori* de las firmas que aparecían en las imágenes, para lo cual se utilizaron otras fuentes de datos, incluyendo observaciones directas de las manchas que nos permitieron verificar muchos de los vertidos presentes. Los resultados mostraron un alto grado de precisión (98.1%) y significancia (93.5%) en la validación, así como una prometedora capacidad de generalización del algoritmo.

ABSTRACT

Oil spill detection over ocean areas using synthetic aperture radar (SAR) images is a complicated operation due to the presence of other phenomena with signatures similar to those of oil slicks, and hence, different automatic or semi-automatic detection systems have been proposed in order to distinguish the real oil spills. In this work it is proposed a classifier based on a supervised learning method named Support Vector Machine (SVM). The algorithm was developed using 26 ENVISAT ASAR images acquired during the Prestige oil spill at the end of 2002 on the north-west coast of Spain. These images show not only a great number of oil slicks but also a lot of false alarms (or look-alikes). With the aim of training and validating the classifier, it was necessary *a priori* categorization of the signatures using other data sources, including direct observations that allow us to verify several slicks as oil. Results show a high degree of accuracy (98.1%) and significance (93.5%) in the validation, and also promising generalization capabilities of the algorithm.

Palabras clave: radar de apertura sintética (SAR), vertidos de hidrocarburos, Máquina de Vectores Soporte (SVM), sistema de detección, clasificador.

INTRODUCCIÓN

La contaminación por hidrocarburos es una seria amenaza para los ecosistemas marinos y puede afectar de forma grave a todas las actividades económicas dependientes de mar. El problema se agrava especialmente en zonas costeras o con un tráfico marítimo muy intenso, como es el caso de la plataforma continental de Galicia. Además del riesgo evidente que suponen las grandes catástrofes, como la marea negra causada por el petrolero Prestige a finales de 2002 en esta zona, la mayoría de los vertidos son causados por los derrames debidos a operaciones rutinarias de los buques o a pequeños accidentes.

La detección de los vertidos es por tanto muy importante a la hora de planificar las medidas encaminadas a reducir los impactos ocasionados por la contaminación, usándose con este fin diferentes sensores remotos, tanto a bordo de satélites, siendo las imágenes más ampliamente utilizadas las de radar de apertura sintética (SAR), como aerotransportados (Brekke y Solberg 2005).

La identificación de vertidos en las imágenes SAR es posible ya que aparecen como áreas oscuras en contraste con el fondo más brillante. Este método cuenta con grandes ventajas, como pueden ser la gran cobertura espacial o la posibilidad de obtener información de noche o bajo

malas condiciones meteorológicas, y se han obtenido muy buenos resultados en diferentes trabajos. Sin embargo, la presencia de otros fenómenos, tanto oceánicos como atmosféricos, que originan firmas muy similares a la de los vertidos, provoca que el riesgo de aparición de falsas alarmas sea muy alto y que se requiera a un interpretador experto para poder discriminar los vertidos auténticos (Torres *et al.* 2006).

En este trabajo se presenta un algoritmo clasificador basado en una máquina de vectores soporte (SVM) que permite discriminar los vertidos de otros fenómenos con una firma similar. Dicho algoritmo forma parte de un sistema automático de detección de vertidos a partir de imágenes ENVISAT ASAR (Advanced Synthetic Aperture Radar) desarrollado para la plataforma continental de Galicia.

FUENTES DE DATOS

Para poder llevar a cabo el diseño, entrenamiento y validación del algoritmo de clasificación se utilizaron 26 imágenes ASAR adquiridas durante los siguientes meses a la catástrofe del Prestige en el noroeste peninsular. ASAR opera en banda-C (5.331 GHz) y las imágenes fueron adquiridas en el Modo de Barrido Ancho (WS), con una resolución espacial media de 150 m y una amplia cobertura (400 x 400 km).

Estas imágenes se caracterizan por una gran cantidad de manchas de hidrocarburos, las cuales en su mayoría tienen su origen en las 77000 toneladas de petróleo crudo (tipo M-100) que transportaba el petrolero Prestige a finales del 2002 (Figura 1), aunque tampoco se pueden descartar vertidos provocados por otros buques en operaciones rutinarias como la limpieza de las sentinas. Sin embargo, a pesar de proceder principalmente de la misma fuente, existe heterogeneidad en las manchas debido a la amplia zona geográfica (incluyendo la costa Atlántica portuguesa, la costa gallega y la costa cantábrica) y al largo período de tiempo (de noviembre a abril) cubiertos por las imágenes, de forma que el petróleo permaneció en el agua sujeto a condiciones meteorológicas diversas y sufriendo diferentes procesos de meteorización.

Las imágenes muestran también abundantes firmas similares a la de los vertidos, es decir, áreas que se ven como zonas oscuras debido a su baja retrodispersión. Las falsas alarmas son principalmente zonas de bajo viento en la costa causadas por el efecto pantalla del relieve. Sin

embargo, también se detectaron otros fenómenos atmosféricos u oceanográficos, como afloramientos o estuarios con elevada descarga fluvial.

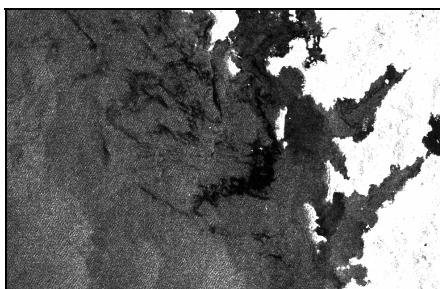


Figura 1: Fragmento de la imagen ASAR del 3 de diciembre del 2002, con vertidos de hidrocarburos amenazando las rías gallegas.

Además de las imágenes ASAR también se utilizaron otras fuentes de datos en la categorización *a priori* de las posibles manchas (ver Metodología). Así, se utilizaron los datos de observaciones directas de los vertidos obtenidos desde barco, avión o helicóptero y recopilados por la Unidad de Observación Próxima (UOP) del gobierno gallego. También se obtuvieron los campos de viento derivados del escaterómetro SeaWinds a bordo del QuickScat y otras imágenes de satélite (MERIS, MODIS o AVHRR).

METODOLOGÍA

Sistemas de detección de vertidos

La mayoría de los sistemas automáticos de detección de vertidos a partir de imágenes SAR constan de tres etapas (González y Torres 2007): segmentación, es decir, identificación de las firmas en la imagen susceptibles de ser hidrocarburos, caracterización de las mismas y aplicación de un clasificador para distinguir los vertidos reales de otros fenómenos. Nuestro sistema sigue el mismo esquema, aunque en este trabajo nos centramos en la última fase.

La segmentación la llevamos a cabo mediante una umbralización adaptativa y la caracterización mediante un algoritmo de etiquetado, que permitió extraer características geométricas (forma, tamaño y localización), radiométricas y adicionales (asociadas a fuentes de datos externas). Su puede consultar una información más detallada sobre la metodología utilizada en otros trabajos (González *et al.* 2006, González y Torres 2007).

Clasificación

La clasificación se plantea como un problema de dos clases, vertidos y falsas alarmas, y se utiliza un algoritmo que a partir de las características obtenidas en la etapa de caracterización asigne a cada signatura extraída en una imagen la clase a la cual se aproxima más. Por lo tanto, para poder entrenar y validar el clasificador necesitamos un conjunto de signaturas de las cuales podamos conocer de antemano y con seguridad si son vertidos o no.

Así, se llevo a cabo una categorización *a priori* mediante la cual a cada una de las 1438 signaturas inicialmente identificadas se le asignó una categoría en función de su probabilidad de ser un vertido (Tabla 1). La utilización de los datos de observaciones directas permitió verificar 54 signaturas como hidrocarburos, mientras que 270 fueron constatados como falsas alarmas en base a información contextual y a fuentes de datos adicionales, como los campos de viento. Utilizando la misma información se les asignó a las restantes signaturas categorías intermedias entre ambas clases, incluyendo una de indeterminados que supone más de la mitad de las signaturas.

Tabla 1: Número de signaturas y porcentaje del total para cada categoría establecida *a priori*.

Categoría	N	%
Falsa Alarma Verificada (FAV)	270	18.8%
Falsa Alarma Probable (FAP)	259	18.0%
Indeterminado (IND)	810	56.3%
Vertidos probables (VEP)	45	3.1%
Vertidos verificados (VEV)	54	3.8%
	1418	100%

Para la implementación del algoritmo se usaron características geométricas y radiométricas (Tabla 2). Dentro de las primeras se incluyo la distancia a la costa y varios parámetros de forma. Las radiométricas incluyen la intensidad y la dispersión de la retrodispersión en el interior y en el exterior (dentro del rectángulo mínimo envolvente) de la mancha, y la media y desviación del gradiente calculado en el borde de la signatura. Se usaron estadísticas robustas, es decir, mediana y desviación absoluta de la mediana (MAD), ya que éstos estadísticos no se ven tan afectados por valores extremos debidos al ruido de la imagen.

Tabla 2: Características radiométricas y geométricas utilizadas en la clasificación.

Radiométricas	Geométricas
Mediana (dentro)	Distancia a la costa (km)
MAD (fuera)	Momento 2º eje principal
Mediana (dentro)	Ángulo eje principal
MAD (fuera)	Radio área (dentro/fuera)
Mediana (gradiente)	Radio de aspecto
MAD (gradiente)	Compacidad ($100 P^2 / A$)

Máquina de vectores soporte

La máquina de vectores soporte (SVM) fue el algoritmo elegido para desarrollar el clasificador, en vista de los buenos resultados obtenidos en otras aplicaciones relacionadas con la teledetección (Melgani y Bruzzone 2004). Los SVMs son algoritmos de aprendizaje que se engloban dentro de los métodos de kernel, y que se basan en el clasificador lineal, el cual busca un hiperplano óptimo que nos permita separar dos clases (vertidos y falsas alarmas) de forma que la distancia entre un dato dado y la superficie de separación sea máxima. Las SVMs operan en un espacio de características, es decir, en un espacio de mayor dimensionalidad que el conjunto de entrada en el cual existe una mayor probabilidad de que los datos sean linealmente separables. La función kernel permite transformar los datos de entrada en ese espacio de características. En este trabajo se utilizó como kernel una función guasiana de base radial (RBF), resolviendo la clasificación como un problema de optimización cuadrática introduciendo operadores de Lagrange.

Entrenamiento y validación

Como el número de signaturas verificadas (54 vertidos y 270 falsas alarmas) no era lo bastante grande como para obtener conjuntos de entrenamiento y validación significativos, se optó por llevar a cabo una validación cruzada de primer orden (o dejando uno fuera). Así, el entrenamiento se realiza con todos los elementos menos uno, y posteriormente se clasifica dicho elemento y se almacena el resultado, y así consecutivamente tantas veces como elementos tenga el conjunto de entrada. El proceso se llevó a cabo utilizando tres conjuntos de parámetros distintos: todas las características, sólo características geométricas y sólo características radiométricas.

Para evaluar la eficacia de los tres clasificadores se aplicaron los algoritmos ya entrenados al conjunto de entrenamiento, y se calculó la exactitud global (EG) como el porcentaje de firmas bien clasificadas, así como el estadístico kappa (κ), medida de la exactitud de la clasificación que prescinde del efecto del azar (Tabla 3). Además, para valorar la capacidad de generalización de la SVM se llevó a cabo una clasificación del conjunto de datos completo, y se determinó ver el porcentaje de firmas dentro de cada una de las categorías asignadas *a priori* que el algoritmo clasifica como vertido (Figura 2).

Tabla 3: Exactitud global (EG) y estadístico kappa (k) de los diferentes clasificadores

Características	EG	k
Geométricas	93.2%	74.0%
Radiométricas	98.1%	93.5%
Todas las características	98.1%	93.2%

RESULTADOS

Los resultados muestran la gran eficacia del algoritmo cuando se aplica sobre el conjunto de entrenamiento, que tan solo incluye firmas verificadas, clasificando correctamente más del 90% del total. Sin embargo, vemos como utilizando conjuntamente todas las características o bien sólo las radiométricas la clasificación es más confiable, con un mayor valor del estadístico kappa.

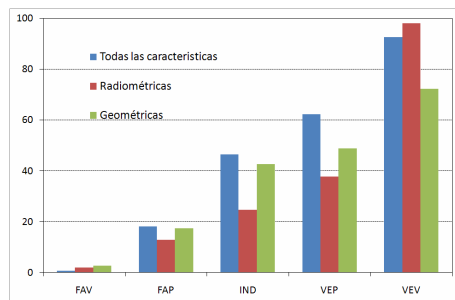


Figura 2: Porcentaje de firmas clasificadas como vertidos dentro de cada categorías.

El porcentaje de firmas clasificadas como vertidos (Figura 2) aumenta hacia aquellas categorías con una mayor probabilidad de ser hidrocarburos, por lo que el algoritmo presenta en principio una buena capacidad de generalización. Vemos también que cuando se utilizan solo características geométricas la SVM falla bastante con los vertidos verificados, ya que apenas clasifica correctamente algo más del 70%, lo cual explica su

menor fiabilidad ($k = 0.74$). Finalmente, también es llamativo como en las categorías intermedias se clasifican como vertidos un mayor porcentaje de firmas utilizando todas las características que empleando sólo las radiométricas. Necesitaríamos más información sobre estas manchas para saber con más certeza que aproximación es más fiable.

CONCLUSIONES

Los resultados de este trabajo muestran el gran potencial de las SVMs para la correcta clasificación de las firmas detectadas en imágenes SAR. El algoritmo puede además ser mejorado mediante la utilización de datos adicionales (como el viento) o mediante un entrenamiento más completo incluyendo otros tipos de contaminantes u otras zonas.

BIBLIOGRAFÍA

- Brekke, C. and Solberg, A. H. S. 2005. Oil spill detection by satellite remote sensing. *Remote Sensing of Environment* 95: 1–13.
- González Vilas, L. and Torres Palenzuela, J.M. 2007. Improving oil slick detection by SAR imagery using ancillary data. *Proceedings of 2007 IEEE International Symposium on Industrial Electronics*: 1657-1665.
- González-Vilas, L., Torres J.M., Martínez-Iglesias, G., Mosquera, A. and Corchado, J.M. 2006. Utilización de datos de ENVISAT para la detección de vertidos de hidrocarburos. *Revista de la Asociación Española de Teledetección* 25: 55-59.
- Melgani, F. and Bruzzone, L. 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42, no. 8: 1778–1790.
- Torres Palenzuela, J.M., González Vilas, L. and Sacau Cuadrado, M. 2006. Use of ASAR images to study the evolution of the Prestige oil spill off the Galician coast. *International Journal of Remote Sensing* 27, no. 10: 1931-1950.

AGRADECIMIENTOS

Las imágenes ASAR utilizadas en este trabajo han sido proporcionadas por la ESA dentro del proyecto AO623. La investigación se ha desarrollado en el marco de los proyectos CONTINMAR (2003-2006), y DRIFTER (2009-2010).