

## Clasificación de imágenes multispectrales mediante redes neuronales

J. A. Triñanes<sup>1</sup>, J. Torres<sup>3</sup>; A. Tobar<sup>2</sup>; C. Hernández<sup>1</sup>

(1) Departamento de Física Aplicada

(2) Departamento de Electrónica y Ciencias de la Computación  
Universidad de Santiago

(3) Departamento de Física Fundamental y Experimental  
Universidad de La Laguna

### RESUMEN

En este artículo proponemos una red neuronal para llevar a cabo una clasificación supervisada de imágenes multispectrales. Adicionalmente, como una extensión de la arquitectura básica veremos que la información sobre textura puede ser integrada dentro del clasificador sin una definición explícita de la misma. Variantes de la regla delta generalizada tendentes a aumentar la rapidez de la fase de entrenamiento son también presentados.

### ABSTRACT

In this paper, we propose a neural network to perform a supervised classification of multispectral images. In addition, in an extension of the basic network architecture we show that textural information can be integrated into the neural network classifier without the explicit definition of a texture measure. We modify the generalized delta rule to reduce the training phase.

### Introducción

Avances recientes en el campo de las redes neuronales han demostrado la utilidad de las mismas en una amplia variedad de áreas. La aplicación de redes neuronales en tareas de reconocimiento de patrones está basada en varias consideraciones. Primero, las redes neuronales llevan a cabo su tarea tan bien o mejor que otras técnicas, y no requieren hipótesis sobre la naturaleza explícita del conjunto de datos que descamos clasificar (por ejemplo, distribuciones normales). En este aspecto, son similares a los algoritmos denominados *K-nearest neighbor* (Lee et al., 1990). Sin embargo, una red neuronal, una vez ha sido entrenada, es computacionalmente más eficiente. La propiedad de aprendizaje de una red neuronal proporciona una alternativa interesante al clasificador estadístico. Además, cualquier clasificador Gaussiano puede ser implementado por medio de una red neuronal (Bischof et al., 1992). Nosotros las hemos aplicado al análisis multispectral de imágenes Landsat-TM.

**Perceptrón multicapa**

El tipo de red neuronal que hemos elegido para llevar a cabo la tarea de clasificación es la denominada perceptrón multicapa (Hush and Horne, 1993). Esta estructura consta de diversas capas de nodos o neuronas que interactúan a través de conexiones ponderadas. No existen interconexiones dentro de una capa mientras que todas las neuronas en un nivel están conectadas a las neuronas situadas en capas adyacentes. Los pesos asignados a cada conexión miden el grado de correlación entre los niveles de actividad de las neuronas que conectan. Esta disposición en cascada de neuronas permite implementar superficies de decisión complejas y expresiones Booleanas arbitrarias. Para problemas de clasificación, un perceptrón multicapa de 2 capas (esto es, una capa de salida y otra denominada oculta, situada a continuación de la de entrada) puede formar una aproximación arbitrariamente cercana a cualquier superficie no lineal de decisión. Sin embargo, estos resultados no implican necesariamente que de redes de más de dos capas no se puedan extraer beneficios considerables. Para algunos problemas, como veremos más adelante, una pequeña red de 3-capas puede ser empleada en lugar de otra de 2-capas que requiriera un número muy elevado de nodos. De hecho, hay problemas que requieren un número exponencial de nodos en una red de 2-capas que pueden ser implementados en una red de 3-capas con un número polinómico de neuronas.

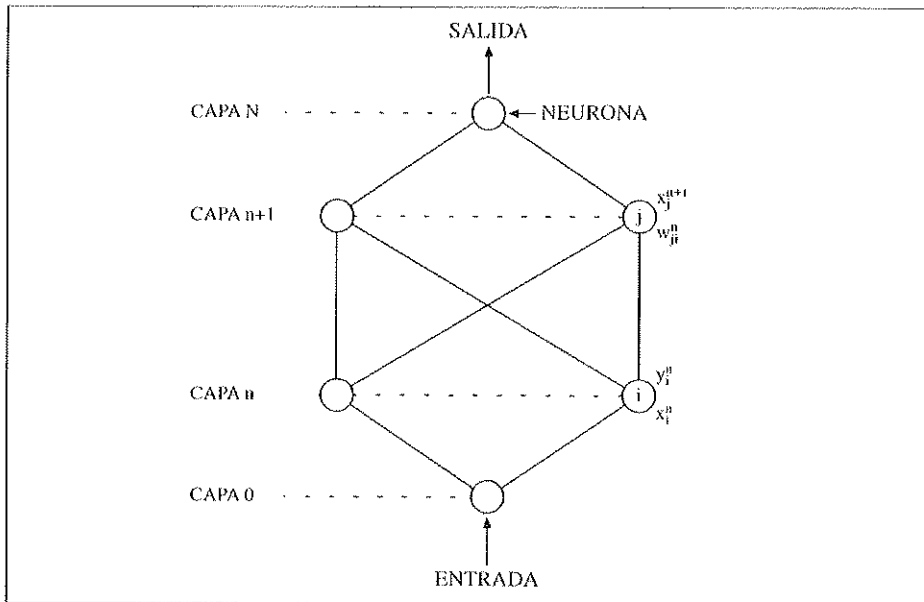


FIGURA 1  
Red Neuronal con 2 capas ocultas. Las neuronas se interconectan por medio de pesos variables de conexión

Dada la red de la figura 1, la entrada total,  $x_j^{n+1}$ , recibida por la neurona  $j$  en la capa  $n+1$  se define como (Benediktsson et al., 1990):

$$x_j^{n+1} = \sum_i y_i^n w_{ji}^n - \theta_j^{n+1}$$

donde  $y_i^n$  es el estado de la neurona  $i$  en la capa  $n$ ,  $w_{ji}^n$  es el coeficiente de peso en la conexión de la neurona  $i$  de la capa  $n$  a la neurona  $j$  de la capa  $n+1$ , y  $\theta_j^{n+1}$  es el umbral de la neurona  $j$  en la capa  $n+1$ .

En este estudio, la salida de una neurona de cualquier capa distinta de la de entrada es una función no lineal de su entrada. Dicha función se denomina función sigmoide y viene dada por:

$$y_j^n = \frac{1}{1 + e^{-x_j^n}}$$

Uno de los algoritmos de aprendizaje más populares aplicados a estructuras de perceptrones multicapa es el denominado *back-propagation* o regla delta generalizada (Gorman and Sejnowski, 1988). Está diseñado para reducir el error entre la salida actual y la salida deseada. La medida usual de error es el error cuadrático total (E) dado por:

$$E = \sum_p \sum_k (o_{pk} - t_{pk})^2$$

donde  $p$  abarca todos los vectores de los conjuntos de entrenamiento y  $k$  denota el nodo de salida.  $o_{pk}$  indica la salida de la unidad  $k$  cuando el vector de entrada  $p$  es aplicado a la red, y  $t_{pk}$  es la salida deseada correspondiente.

Una forma de minimizar el error  $E$  es aplicar el método del gradiente descendente a partir de un conjunto arbitrario de coeficientes de peso para los nodos de la red que vamos repetidamente actualizando a través de las siguientes expresiones:

$$\Delta w_{ji}^n(t) = -\epsilon \frac{\partial E}{\partial w_{ji}^n} + \alpha \Delta w_{ji}^n(t-1) - h \times w_{ji}^n(t-1)$$

$$\frac{\partial E}{\partial w_{ij}^n} = \frac{\partial E}{\partial y_j} \frac{dy_j}{dx_j} \frac{\partial x_j}{\partial w_{ij}^n} = \frac{\partial E}{\partial y_j} y_j^n (1 - y_j^n) y_i^{n-1}$$

Para la capa de salida ( $n=N$ ):

$$\frac{\partial E}{\partial y_j} = y_j^N - t_{pj}$$

donde  $\epsilon$  es el denominado factor de aprendizaje que controla el descenso de  $E$ ,  $\alpha$  es el momento de la red,  $h$  es el coeficiente de % de decaimiento y  $t$  denota el número de la iteración que estamos llevando a cabo.

El proceso de computar el gradiente y ajustar los pesos se repite hasta que encontremos un mínimo (o un punto suficientemente cercano al mínimo).

### **Metodología**

#### **Conjuntos de entrenamiento**

Para el estudio de la red neuronal, se seleccionaron conjuntos de entrenamiento pertenecientes a una imagen de 1024x1024 pixels del sensor Thematic Mapper del Landsat-5 correspondientes a una zona comprendida entre las rías de Arosa y Pontevedra, en la fachada atlántica de la Comunidad Autónoma de Galicia. Cada pixel de la imagen consta de siete bandas de 8-bits.

Para la elección de los conjuntos de entrenamiento y posteriormente los de test de resultados, empleamos mapas forestales y fotogramétricos elaborados por la Xunta de Galicia en 1988.

En cuanto a la clasificación de los distintos grupos de pixels, distinguiremos un total de 9 clases:

- Zonas urbanas densas
- Zonas urbanas no densas
- Matorral-pastizal
- Agua
- Pinos
- Eucaliptos
- Vegetal-noforestal

La peculiar idiosincrasia en la disposición de las distintas cubiertas en el territorio objeto de estudio nos ha llevado a la elección de dichas clases. En nuestro ánimo está distinguir entre las fuertes concentraciones urbanas que se encuentran en la imagen y aquellas concentraciones, que sin ser de la entidad de las primeras, presenten una presencia humana no desdeñable y que en Galicia alcanza sus mayores cotas de aparición. También hemos intentado distinguir entre zonas pobladas por EUCALIPTOS y PINOS, en un intento por encontrar alguna relación entre ambos tipos de cubiertas. Para ello hemos seleccionado como conjuntos de entrenamiento zonas pobladas mayoritariamente por una u otra especie. Las zonas de bosque mixto por el momento las hemos mantenido aparte, debido también a la imprecisión mostrada por el mapa en la definición de estas zonas, y, en un futuro, se podrían tener en cuenta considerando las salidas de la red neuronal correspondientes a la clase PINOS y a la clase EUCALIPTOS, o bien asignando un conjunto de entrenamiento aparte para dicha clase. Otras clases forestales distintas de las de PINOS y EUCALIPTOS presentan muy poca presencia en este área y además en su mayor parte se

encuentran fuertemente mezcladas con las dos especies anteriores. Debido a ello hemos decidido obviar su presencia.

Dada la no existencia de mapas detallados acerca de las especies de cultivo existentes, en parte debido también al minifundismo patente en Galicia y a la gran variabilidad temporal y espacial entre cultivos, nos hemos visto obligados a condensar todas estas zonas dentro de una sola clase temática que hemos denominado VEGETAL-NOFORESTAL. Estudios realizados con un mayor aporte de medios nos permitirán incluir estos datos y otros en nuestra clasificación.

En total tenemos un total de 4231 puntos entre los conjuntos de entrenamiento y los de test, de los cuales empleamos 2100 en la fase de entrenamiento de la red neuronal y los restantes en la fase de testeo de resultados.

#### ***Datos de entrada a la red***

Antes de establecer la estructura de la red neuronal, debemos construir los patrones de entrada a partir de los conjuntos extraídos de la imagen a clasificar. Decididos a no desaprovechar ninguna de las bandas disponibles, introduciremos en la red los niveles digitales en cada banda para cada patrón de entrada. Se da el caso de que hay pixels que no pueden ser asignados a una determinada clase basándose únicamente en su firma espectral. En estos casos, los valores de los pixels vecinos juegan un papel relevante en la tarea clasificatoria. Decididos a no pasar por alto este hecho, introduciremos en la red los niveles digitales de los pixels situados dentro de una ventana de tamaño 5x5 centrada en el punto de estudio (Bischof et al., 1992). En esta aproximación realizamos un análisis de Componentes Principales, tras el cual, empleamos los valores de la primera componente para este análisis textural. La razón de ello estriba en que esta banda condensa una mayor varianza que el resto de las bandas, por lo cual, resulta la más apropiada para llevar a cabo esta labor.

#### ***Representación de los datos***

Cuando empleamos redes neuronales para clasificar, los niveles digitales deben ser convertidos en un conjunto de valores que puedan ser introducidos apropiadamente dentro de la red neuronal correspondiente. Esto es lo que se conoce como representación de los datos o codificación. De manera similar, debemos definir una representación de salida para las distintas clases.

En nuestro estudio hemos optado por emplear en la capa de entrada de la red neuronal el denominado *coarse coding*, el cual nos permite encontrar suficientes diferencias entre códigos para separar valores semejantes en caso necesario, pero, al mismo tiempo, manteniendo la cualidad de asignar un código similar cuando los valores también lo son (Bischof et al., 1992). Este aspecto favorece las propiedades de

generalización de la red. Podemos decir que estamos haciendo una forma de interpolación. Nosotros usamos 13 unidades para representar los niveles digitales de entrenamiento en el intervalo comprendido entre el nivel mínimo y el máximo entre los que varía dicho valor. Estas unidades las distribuimos uniformemente en este intervalo. Cuando codificamos un valor numérico  $x$ , la salida de la unidad  $i$  en la posición  $z_j$  viene dada por:

$$o_i = e^{-\frac{(x - z_i)^2}{\sigma^2}}$$

donde  $\sigma$  caracteriza el ancho de la función Gaussiana.

En cuanto a la salida de la red neuronal, hemos optado por asignar a cada clase un nodo de la última capa en lugar de emplear un esquema de codificación binario. Esto provoca que la red neuronal crezca en tamaño y, por tanto, sea computacionalmente más compleja, pero, por el otro lado, el número de ciclos de entrenamiento se reduce debido a que la distancia Hamming entre las representaciones de salida de clases diferentes es mayor (Benediktsson et al., 1990). Para cada entrada de la red asignaremos como clase afín la correspondiente a la neurona de salida con mayor nivel de activación.

#### *Arquitectura de la red neuronal*

El número de nodos de entrada viene especificado por la dimensión de los patrones de entrada. A su vez, los nodos de salida están determinados por el número de categorías o clases. Sin embargo, las neuronas de las capas ocultas no guardan una relación directa con el problema. Si escogemos un número limitado de nodos, la red puede que no contenga suficientes grados de libertad como para formar una clasificación. Por otro lado, un número elevado de neuronas puede implicar que la red simplemente memorice los conjuntos de entrenamiento, lo que implica una generalización muy pobre, así como el incremento del tiempo de computación requerido (Kwan and Lec, 1993). Diversas pruebas nos inclinan por la elección de una estructura con 2 capas ocultas de 20 y 10 nodos respectivamente.

Al objeto de acelerar el algoritmo de convergencia, sin poner en peligro la estabilidad del proceso de aprendizaje, esto es, evitando que la red entre en oscilación o inestabilidad, debemos realizar una elección adecuada del factor de aprendizaje  $\epsilon$ , el parámetro más importante de todo este proceso. Un valor mínimo inversamente proporcional al número de patrones de entrenamiento y al número de nodos de la red se ha revelado muy eficaz en este cometido. Aun así, mantener este factor constante durante todo el proceso puede alargar de manera considerable el aprendizaje de la

arquitectura perceptrón multicapa. Para ello, vamos a ir adaptándolo en función de la curvatura local de la superficie, incrementándolo si el último paso implica una disminución del error  $E$ . Para el caso contrario, el momento  $\alpha$  es desactivado y el factor de aprendizaje se reduce (Heermann and Khazenie, 1992). El momento hace que la dirección actual de búsqueda sea una media exponencialmente ponderada de direcciones pasadas, y ayuda a que los pesos se muevan a través de las porciones llanas de la superficie de error después de que hayan descendido a través de zonas con mayor pendiente. Por su parte, el factor de decaimiento  $h$  activa sólo aquellos coeficientes de peso que reducen el error y por lo tanto incrementa las capacidades de generalización de la red.

### **Resultados**

Después de realizar el entrenamiento de la red, realizamos una estimación de los resultados proporcionados por la misma mediante la introducción de los patrones de test. Después de finalizar la fase de aprendizaje, el 95.95% de los pixels pertenecientes al conjunto de entrenamiento y el 91.32% de los asignados al conjunto de test fueron correctamente clasificados. El error en la estimación de las clases forestales se reparte entre ellas casi en su totalidad, esto es, la red distingue a la perfección esta cubierta con relación a las restantes, hallando problemas a la hora de distinguir entre las clases PINOS y EUCALIPTOS. Aun así, los resultados obtenidos son esperanzadores. Lo mismo ocurre con las clases URBANO DENSO y URBANO NODENSO, las cuales se reparten los errores mutuos, aspecto este que es relativamente lógico debido a la, en ocasiones subjetiva, asignación en una clase u otra realizada por el operador. Con respecto al resto de las clases los niveles de error se mantienen relativamente bajos en la fase de test, lo que indica una buena capacidad de generalización de la red neuronal. Como es normal una clase tan fácilmente discernible como la que corresponde a AGUA presenta una correspondencia del 100% desde las primeras fases del entrenamiento, manteniéndose este alto nivel de exactitud en el proceso de testeo. Los pixels pertenecientes a VEGETAL-NOFORESTAL que son incorrectamente clasificados son asignados en su mayoría a MATORRAL, mientras que los pertenecientes a esta clase que son asignados a otras se reparten equitativamente entre las demás clases (excepto la clase AGUA). La Imagen 1 muestra la imagen resultante de la aplicación de la red neuronal sobre cada pixel de la imagen original.

La tabla siguiente resume los resultados obtenidos:

CLASE	FASE DE ENTRENAMIENTO % de acierto	FASE DE TEST
PINOS	93.67%	75.67%
EUCALIPTOS	86.33%	79.21%
URBANO DENSO	99.33%	97.10%
URBANO NO DENSO	95.00%	93.58%
MATORRAL-PASTIZAL	99.67%	97.51%
AGUA	100%	100%
VEGETAL-NO FORESTAL	97.67%	92.06%
TOTAL	95.95%	91.32%

TABLA 1  
Porcentaje de reconocimiento de patrones e imagen test

### Conclusiones

Los resultados expuestos en la tabla anterior demuestran claramente que con redes neuronales como la que hemos empleado se pueden lograr clasificaciones con un alto grado de exactitud. Además, modificaciones del algoritmo de *back-propagation* pueden ser implementadas para acelerar el proceso de aprendizaje de la red, de manera que el coste temporal que implica esta tarea sea considerablemente reducido. Asimismo, la información de textura puede ser fácilmente introducida en la arquitectura neuronal, lo cual posibilita que podamos llegar a resultados más exactos. El modelo propuesto permite la adición de datos procedentes de otros sensores, lo cual, con miras a realizar una clasificación más completa de las distintas cubiertas abre todo un camino de posibilidades.

### Agradecimientos

Agradecemos a la Xunta de Galicia el apoyo financiero prestado a través de la Dirección Xeral de Política Científica sin el cual no hubiera sido posible la realización de este trabajo.

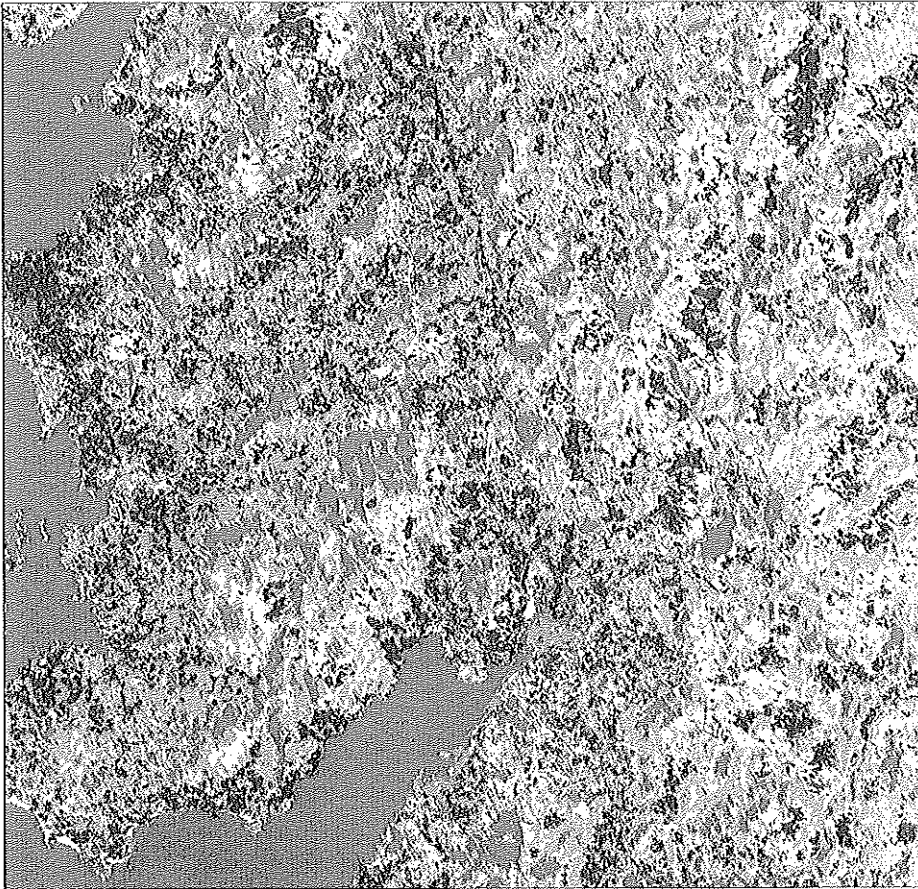


IMAGEN 1

Imagen clasificada con el siguiente código de colores  
PINOS- Verde oscuro. EUCALIPTOS- Verde Claro. URBANO DENSO- Rojo.  
URBANO NODENSO- Granate. MATORRAL-PASTIZAL- Amarillo.  
AGUA- Azul. VEGETAL-NOFORESTAL- Rosa. (A. 81)

### ***Bibliografía***

- Benediktsson, J.A., Swain, P.H. and Ersoy, O.K.:** 1990. Neural Network Approaches Versus Statistical Methods in Classification of Multisource Remote Sensing Data. *IEEE Transactions on Geoscience and Remote Sensing*, 28: 540-551.
- Bischof, H., Schneider, W. and Pinz, A.J.:** 1992. Multispectral Classification of Landsat-Images Using Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, 30: 482-490.
- Gorman, R.P. and Squowski, T.J.:** 1988. Learned Classification of Sonar Targets Using a Massively Parallel Network. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36: 1135-1140.

**Heermann, P.D. and Khazenie, N.:** 1992. Classification of Multispectral Remote Sensing Data Using a Back-Propagation Neural Network. *IEEE Transactions on Geoscience and Remote Sensing*. 30: 81-88.

**Hush, D.R. and Horne, B.G.:** 1993. Progress in Supervised Neural Networks. *IEEE Signal Processing Magazine*. January: 8-39.

**Kwan, H.K. and Lee, C.K.:** 1993. A Neural Network Approach to Pulse Radar Detection. *IEEE Transactions on Aerospace and Electronic Systems*. 29. 9-21.

**Lee, J. et al.:** 1990. A Neural Approach to Cloud Classification. *IEEE Transactions on Geoscience and Remote Sensing*. 28: 846-855.