

SOBRE LA APLICACION DE LA CORRECCION CONTEXTUAL A LA CLASIFICACION OBTENIDA MEDIANTE EL VECINO MAS PROXIMO UTILIZANDO UN CONJUNTO DE ENTRENAMIENTO REDUCIDO

F.J. Cortijo Bon¹ y N. Pérez de la Blanca²

Dpto. Ciencias de la Computación e I.A. E.T.S. Ingeniería Informática. Universidad de Granada. Avda. Andalucía, 38. 18071 Granada. Fax: (958) 243317.

¹ e-mail: *cb@robinson.ugr.es* Tfno: (958) 244078

² e-mail: *nicolas@robinson.ugr.es* Tfno: (958) 243301

RESUMEN.- En la clasificación de imágenes debe considerarse el coste computacional global asociado a la clasificación junto a la bondad del resultado obtenido. El coste asociado al proceso de aprendizaje está determinado, además, por el tamaño del conjunto de aprendizaje. En este trabajo mostraremos la aplicación de algunas técnicas de aprendizaje y clasificación no paramétricas que permiten obtener una alta tasa de acierto en la clasificación de imágenes de Teledetección utilizando un conjunto de entrenamiento reducido. Mostraremos cómo puede aplicarse un método de corrección contextual sobre las clasificaciones obtenidas por métodos de clasificación no paramétricos y cómo los resultados obtenidos mejoran los conseguidos por la metodología clásica de corrección contextual sobre la clasificación de máxima verosimilitud.

ABSTRACT.- The computational cost associated to image classification must be considered together with the accuracy of the classification in order to measure the global performance of the classification. It is well known that the learning accuracy is determined by the learning set size. In this work we will show some non-parametric classification rules that achieve a high accuracy classifications while managing low-sized learning sets. We will show that we can use those classifications as starting points for contextual classifiers and we will show that the accuracy of the so obtained classifications will improve significantly.

1.- INTRODUCCION

Las técnicas supervisadas de clasificación se engloban en dos grandes familias: los métodos paramétricos y los no paramétricos. Los métodos paramétricos son apropiados cuando se conoce la forma de la distribución de probabilidad subyacente de las clases a discriminar. Sin embargo, presentan serios problemas si el solapamiento entre las clases es importante, dando lugar a tasas de error inaceptables (Cortijo, 1995). Los métodos no paramétricos son aplicables si no se realiza ninguna suposición acerca de la distribución de probabilidad, consiguiendo, en general, un mejor resultado en problemas con solapamiento entre clases, ya que la forma de las fronteras de decisión son más flexibles y pueden adaptarse mejor a las muestras. Entre las técnicas no paramétricas han alcanzado gran popularidad las denominadas "del vecino más próximo" por su simplicidad conceptual, facilidad de implementación y propiedades teóricas que, bajo ciertas condiciones, tienden al clasificador de Bayes (Devijver y Kittler, 1982).

La regla de los k vecinos más próximos (k -NNR) asigna a una observación la clase más votada o representada entre las k muestras más cercanas del conjunto de entrenamiento. Se

demuestra que esta regla es óptima si k es grande y el tamaño del conjunto de entrenamiento es virtualmente infinito (Devijver y Kittler, 1982). Sin embargo, en la mayoría de las aplicaciones prácticas esta condición no se cumple. Existen serios problemas de eficiencia para la aplicación de esta regla si el conjunto de entrenamiento es muy numeroso ya que el cálculo de los k vecinos más próximos está en el orden de $O(n^2)$ o mayor, donde n es el número de muestras del conjunto de entrenamiento. Por último, durante la fase de entrenamiento es fácil seleccionar muestras erróneas como representantes de una clase; estas muestras pueden inducir a errores en la clasificación y es necesario eliminarlas del conjunto de entrenamiento.

Para resolver estos problemas asociados a las reglas k -NNR se han propuesto diferentes soluciones en la literatura tales como algoritmos rápidos de búsqueda (Vidal, 1986), métricas modificadas (Short y Fukunaga, 1981) y métodos para obtener un conjunto *reducido y representativo* de muestras del conjunto de entrenamiento. El objetivo de la última aproximación -adoptada en este trabajo- es realizar la búsqueda de los vecinos en un nuevo conjunto de aprendizaje con un equilibrio razonable entre la bondad de la clasificación obtenida y el costo computacional asociado ((Devijver y Kittler, 1982), (Kohonen, 1989) y (Geva y Sitte, 1991) entre otros). Esto último puede abordarse desde dos distintas aproximaciones: a) mediante técnicas de **edición-condensado** y b) mediante técnicas de **aprendizaje adaptativo**.

Este artículo se estructura como sigue. En la sección 2 mostraremos las técnicas más importantes para obtener un conjunto de aprendizaje reducido y representativo. Una vez seleccionado el nuevo conjunto de aprendizaje se realiza la clasificación 1-NNR utilizando este conjunto de aprendizaje. En la sección 3 mostraremos cómo esta clasificación puede mejorarse notablemente mediante la aplicación de alguna técnica de corrección contextual. En la sección 4 describiremos los datos sobre los que se ha aplicado la metodología propuesta y los resultados obtenidos. Las conclusiones más importantes que podemos extraer de este trabajo los mostramos en la sección 5.

2.- UN CONJUNTO DE ENTRENAMIENTO REDUCIDO Y REPRESENTATIVO

El objetivo de las *técnicas de edición-condensado* es doble: aumentar la bondad de la clasificación eliminando del conjunto de aprendizaje las muestras situadas en regiones con baja probabilidad a posteriori (regiones de solapamiento espectral) -*técnicas de edición-* y reducir el coste computacional asociado a la búsqueda de los vecinos más próximos -*técnicas de condensado-* ya que ésta se aplica a conjuntos muy reducidos.

La reducción del tamaño del conjunto de entrenamiento conseguida tras la aplicación de estas técnicas es importante, permitiendo además, la aplicación de la regla más simple y rápida del vecino más próximo ($k = 1$) utilizando este conjunto reducido como conjunto de aprendizaje. Sin embargo, hemos encontrado ((Cortijo et al., 1995), (Cortijo, 1995)) que si el grado de solapamiento es elevado, es posible que algunas clases pierdan todos sus representantes durante la multiedición, lo que produce un descenso en la tasa de acierto de la clasificación, y lo que es más importante, una pérdida de información. Las técnicas de edición-condensado proporcionan un conjunto de aprendizaje que es un subconjunto del conjunto de entrenamiento original. Es más, las técnicas de condensado requieren que el conjunto de aprendizaje sobre el que actúan esté previamente editado. En su formulación original no es posible especificar la cardinalidad del conjunto (editado o condensado) resultante.

Las *técnicas de aprendizaje adaptativo* comparten el mismo objetivo que las de edición condensado pero su planteamiento es radicalmente opuesto. El conjunto de entrenamiento resultante no tiene porqué ser (de hecho, muy raramente lo es) un subconjunto del conjunto

de entrenamiento original. Estas técnicas actúan sobre el conjunto de entrenamiento original, extrayendo un número fijo de muestras *-prototipos-* y ajustando los valores de estos prototipos a las muestras de aprendizaje presentadas. El método de actualización de los prototipos (semajante a la actualización de los pesos en una red neuronal) determina el método de aprendizaje. Entre los métodos más relevantes citaremos las actualizaciones mediante las técnicas **LVQ** (learning vector quantization) (Kohonen, 1990) y **DSM** (Decision Surface Mapping) (Geva y Sitte, 1991).

En este trabajo mostramos la aplicación conjunta de los métodos de edición-condensado a la clasificación de imágenes de Teledetección. La aplicación de los métodos de aprendizaje adaptativo sobre estas imágenes puede encontrarse en (Cortijo, 1995) y (Cortijo y Pérez de la Blanca, 1995). Entre la técnicas de edición presentes en la literatura hemos adoptado en este trabajo el algoritmo **Multiedit** (Devijver y Kittler, 1982) que se ha mostrado como el más robusto por su convergencia. Entre los algoritmos de condensado hemos adoptado, por su sencillez y convergencia, el algoritmo de **condensado de Hart** (Hart, 1968).

El algoritmo Multiedit elimina las muestras de una clase que se encuentran inmersas en los agrupamientos de otra clase, reduciendo el error de clasificación. El condensado de Hart selecciona del conjunto de muestras multieditado aquellas que determinan la frontera de decisión entre las clases.

3.- CORRECCION CONTEXTUAL DE LA CLASIFICACION

El resultado de la clasificación por un método espectral como los descritos anteriormente produce imágenes "ruidosas" ya que en la clasificación de un píxel no se considera la información que puede aportar un entorno del píxel considerado en cada momento, lo cual no es natural. Es necesario, por tanto, incorporar información del contexto de cada píxel en la clasificación. La incorporación de esta información contextual puede realizarse, básicamente desde dos enfoques muy diferentes (Cortijo, 1995): a) mediante **técnicas de alisamiento**, que pueden aplicarse previamente a la clasificación (métodos de pre-alisamiento) o bien posteriormente a la clasificación (métodos de post-alisamiento) y b) mediante **técnicas basadas en Campos Aleatorios Markovianos** (Dubes y Jain, 1989).

La suposición de un campo aleatorio de Markov (CAM) subyacente entre las etiquetas es el modelo más adecuado (Cortijo, 1995) ya que permite modelar de una forma muy precisa la distribución a priori de las etiquetas. En la literatura se han propuesto diversos modelos entre los cuales citaremos los modelos de Welch y Salter, Haslett y de Owen y Switzer (Saebo et al., 1985) y especialmente el **algoritmo ICM** propuesto por Besag (Besag, 1986).

El algoritmo ICM parte de la suposición de que la distribución a priori de las etiquetas es un campo aleatorio de Markov. Resulta una alternativa computacionalmente muy atractiva al cálculo directo del máximo a posteriori o de la moda de la distribución a posteriori (Dubes y Jain, 1989) y salva la tendencia de un CAM a degenerar en una imagen plana de etiquetas. El algoritmo puede escribirse como sigue (Dubes y Jain, 1989):

- 1.- Elegir un modelo de CAM para las etiquetas verdaderas.
- 2.- Inicializar el mapa de etiquetas inicial.
- 3.- Para cada píxel de la imagen (*i*) actualizar la etiqueta actual por aquella etiqueta l_i que maximiza:

$$P (X_i=x_i | L_i=l_i) P (L_i=l_i | L_{\rho_i}=l_{\rho_i})$$

- 4.- Repetir el paso 3 N_{iter} veces.

Sobre este algoritmo podemos realizar las siguientes consideraciones:

1. Sobre el modelo de CAM (paso 1) Besag sugiere utilizar un modelo de interacción por parejas en una vecindad de segundo orden y de un sólo parámetro. En este caso hemos adoptado un parámetro variable (creciente) con la iteración, lo que permite un ajuste más acentuado a las etiquetas conforme crece éste (véase (Cortijo, 1995) para más detalles).
2. Sobre la solución inicial (paso 2) hemos utilizado la clasificación 1-NNR utilizando los conjuntos inicial, multieditado y multieditado y condensado además de la clasificación de máxima verosimilitud.
3. Sobre los términos de la maximización (paso 3) hemos utilizado un modelo paramétrico (estimación por máxima verosimilitud) para el primer término de la maximización. Sería posible, sin embargo realizar una estimación no paramétrica para este término. La formulación exacta del segundo término, que determina el modelo escogido puede encontrarse en la literatura (véase, p.e. (Cortijo,1995)).
4. Respecto al número de iteraciones, hemos utilizado $N_{iter}=8$, lo que asegura sobradamente la convergencia del algoritmo.

4.- EXPERIMENTOS: DATOS Y RESULTADOS

Los datos usados para comprobar la eficacia de la metodología propuesta son dos imágenes LANDSAT de tamaño 512 x 512 de dos regiones de Groenlandia, Igaliko (LANDSAT-2 MSS) e Ymer O (LANDSAT-5 TM). En Igaliko se seleccionaron 5 clases informacionales y 42796 muestras de entrenamiento y en Ymer O se seleccionaron 20 clases y 12574 muestras de entrenamiento. Los dos casos estudiados responden a diferentes problemáticas, ya que independientemente de las diferencias implícitas relacionadas con los diferentes sensores utilizados, presentan diferentes comportamientos espectrales, siendo el solapamiento espectral presente en el caso de Ymer O muy acentuado (Cortijo, 1995).

La medición de la tasa de acierto de los métodos se realizó mediante un conjunto de test independiente del utilizado para construir el clasificador; el conjunto de test contiene 1/3 de las muestras y el utilizado para construir el clasificador los 2/3 restantes. La aplicación de los procedimientos de edición y condensado sobre estos datos se resume en las tablas I y II. En la tabla I mostramos la cardinalidad de los diferentes conjuntos de muestras. Las claves deben interpretarse como sigue: *Inicial* (conjunto de entrenamiento original), *Aprendizaje* (subconjunto del anterior utilizado para construir el clasificador), *Test* (subconjunto de *Inicial* utilizado para medir la bondad de las clasificaciones), *Multieditado* (subconjunto de *Aprendizaje* resultado de la multiedición y, por último, *Condensado* (subconjunto de *Multieditado* resultado del condensado de Hart).

	Inicial	Aprendizaje	Test	Multieditado	Condensado
Igaliko	42796	28441	14355	1741	39
Ymer O	12574	8560	4014	3398	68

Tabla I.- Tamaños de los conjuntos de datos.

En la tabla II mostramos las bondades de las clasificaciones de máxima verosimilitud y de las clasificaciones 1-NNR que utilizan estos conjuntos como conjuntos de aprendizaje. Para cada ejemplo se muestran dos líneas: en la superior se proporciona la bondad del método espectral y en la inferior, la bondad conseguida mediante la corrección contextual por el algoritmo ICM utilizando las correspondientes clasificaciones como punto de partida. Las

claves deben interpretarse como sigue: *MV* (máxima verosimilitud), *1-NNR* (vecino más próximo utilizando el conjunto de aprendizaje *Inicial*), *1-NNR (M)* (Vecino más próximo utilizando *Multieditado* y, por último, *1-NNR (MC)* (Vecino más próximo utilizando *Condensado*).

	MV	1-NNR	1-NNR (M)	1-NNR (MC)
Igaliko	73,51	74,61	77,76	77,08
	81,33	86,94	83,02	82,83
Ymer O	61,92	78,50	65,67	63,23
	91,37	97,98	90,07	81,09

Tabla II.- Bondades de las clasificaciones.

5.- CONCLUSIONES

La elección de la clasificación inicial sobre la que se realiza la corrección contextual determina la bondad de la clasificación contextual final. La utilización de un método no paramétrico en problemas con alto solapamiento espectral proporciona unos resultados superiores a los obtenidos con la clasificación de máxima verosimilitud. Entre los métodos de corrección contextual, aquellos que consideran un modelo de CAM para la distribución a priori de las etiquetas resultan los más adecuados, y en particular el algoritmo ICM proporciona los mejores resultados con un coste computacional reducido frente a los otros métodos estudiados.

6.- REFERENCIAS

- Besag, J. 1986. On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society, Ser. B.* 48:259-302.
- Cortijo, F.J. 1995. *Un Estudio Comparativo de Métodos de Clasificación de Imágenes Multibanda*. Tesis Doctoral. Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada.
- Cortijo, F.J. y Pérez de la Blanca. 1995. Image Classification Using Adaptive Learning Algorithms. Automatic Estimation of the LVQ-1 Parameters. Sometido a *Pattern Recognition*.
- Cortijo, F.J., Pérez de la Blanca, N., Molina, R. y Abad, J. 1995. On the Combination of Nonparametric Nearest Neighbor and Contextual Correction. En *Pattern Recognition and Image Analysis. VI Spanish Symposium on Pattern Recognition and Image Analysis*. Córdoba. pp. 503-510. Asociación Española de Reconocimiento de Formas y Análisis de Imágenes (AERFAI).
- Devijver, P.A. y Kittler, J.V. 1982. *Pattern Recognition. A Statistical Approach*. Prentice Hall, Englewood Cliffs.
- Dubes, R.C. y Jain, A. 1989. Random Field Models in Image Analysis. *Journal of Applied Statistics.* 16:131-164.
- Geva, S. y Sitte, J. 1991. Adaptive nearest Neighbor Pattern Classification. *IEEE Transactions on Neural Networks.* 2:318-322.
- Hart, P. 1968. The Condensed Nearest Neighbor Rule. *IEEE Transactions on Information Theory.* 14:515-516.
- Kohonen, T. 1989. *Self-Organization and Associative Memory*. Springer-Verlag.
- Kohonen, T. 1990. The Self-Organizing Map. En *Proc. of the IEEE.* 78:1464-1480.
- Saebo, H., Braten, K., Hjort, N, Llewellyn y Mohn, E. 1985. *Contextual Classification of Remotely Sensed Data: Statistical Methods and Development of a System*. Technical Report 768, Norwegian Computer Center.
- Short, R.D. y Fukunaga, K. 1981. The Optimal Distance Measure for Nearest Neighbor Classification. *IEEE Transactions on Information Theory.* 27:622-627.
- Vidal, E. 1986. An Algorithm for Finding Nearest Neighbors in (approximately) Constant Average Time. *Pattern Recognition Letters.* 4:145-157.