

Una metodología para el reconocimiento de patrones en tareas geólogo-geofísicas

Ricardo Barandela Alonso
Instituto de Geografía
Academia de Ciencias de Cuba, La Habana, Cuba.

Recibido: 25 de febrero, 1993; aceptado: 29 de junio, 1995.

RESUMEN

En el presente trabajo se describen y discuten algunas experiencias obtenidas en la práctica del reconocimiento de patrones en el campo geológico y geofísico, llevadas a cabo con el empleo de un método supervisado (la llamada regla NN) y utilizando el sistema interactivo de computación NNINT. Se enumeran criterios extraídos de esas aplicaciones acerca de la conveniencia de preprocesar la muestra de entrenamiento, y se propone una metodología para ello.

PALABRAS CLAVE: Aprendizaje, imperfectamente supervisado, regla NN, preprocesamiento de la muestra de entrenamiento.

ABSTRACT

Some practical experiences in the application of pattern recognition to geological and geophysical tasks are discussed. The Nearest Neighbor rule and the NNINT interactive computing system are described. Some criteria about preprocessing the training sample are discussed and a methodology for this procedure is proposed.

KEY WORDS: Learning, imperfectly supervised, NN rule, training sample processing.

INTRODUCCION

El propósito de los métodos de Reconocimiento de Patrones es dividir un conjunto de individuos u objetos en varios grupos o clases. Estos métodos han demostrado su utilidad en el marco de las investigaciones geológicas y geofísicas, no sólo por el ahorro en tiempo y recursos que han significado, sino también porque han permitido una mejor calidad de los resultados.

Por su potencia discriminatoria, se destacan los métodos conocidos como supervisados o con aprendizaje (véase, por ejemplo, Chen y Fu, 1986) y que aprovechan la información contenida en la Muestra de Entrenamiento (ME). Esta ME consiste en un conjunto de prototipos o patrones de entrenamiento. Se supone que cada uno de esos patrones ha sido evaluado por especialistas y que ha recibido una identificación que acredita qué clase representa dentro de la ME.

La obtención de esa ME resulta, con frecuencia, costosa y de notable dificultad. Para el caso extremo, no se dispone de ninguna información previa (o sea, no existe la ME), y es necesario recurrir a los métodos no supervisados o con auto-aprendizaje (algoritmos de cluster). En el contexto geólogo-geofísico, es posible mencionar los trabajos de Justice *et al.* (1985), Yuxion *et al.* (1986), Wu y Nyland (1987) y Zhang y Simaan (1987).

Una tercera situación, conocida como "imperfectamente supervisada", no ha recibido mucha atención en la literatura aunque ocurre también en la práctica. Aquí se dispone de una ME, pero no se puede afirmar que es correcta la identificación de todos y cada uno de sus elementos. Se ha reportado, por ejemplo, en la elaboración de mapas pronóstico de depósitos minerales (Alfonso *et al.*, 1987).

Hoffer y Fleming (1978), Hoffer y Swain (1980) y Hixson *et al.* (1980) resaltaron la importancia de una buena selección de la ME, por ser éste un elemento clave para un empleo efectivo de los métodos supervisados de clasificación. Esta preocupación se observa de nuevo en publicaciones más recientes, entre las que es posible mencionar a Buchheim y Lillesand (1989), Warren *et al.* (1990), Foody (1990), Bolstad y Lillesand (1991), Baker *et al.* (1991), Kershaw y Fuller (1992), Foody *et al.* (1992), y Gopal y Woodcock (1994).

En el marco del clasificador de Máxima Verosimilitud para procesar datos de percepción remota, Mather (1987 a, b) propugna un algoritmo iterativo para purificar la ME y demuestra, con datos simulados, que el procedimiento disminuye el efecto negativo de las observaciones atípicas en los estimados de los parámetros, pero no llega a analizar las consecuencias sobre la clasificación. Barandela y García (1994) reportan una incidencia muy pobre en los resultados de la clasificación, que pudiera explicarse por la gran distancia entre las clases que se representan en el modelo escogido por Mather para la simulación.

El presente trabajo tiene como objetivo la exposición de una metodología para preprocesar la ME y ejemplificar sus beneficios en cualquiera de las tres situaciones antes mencionadas (supervisada, no supervisada e imperfectamente supervisada). En la Sección próxima se hace una breve descripción del método de reconocimiento de patrones y del sistema de computación interactivo que fueron empleados en el estudio. Después se enumeran, a modo de ilustración, tres aplicaciones prácticas (una para cada situación) que sirvieron para extraer las enseñanzas y los criterios que dieron base a la mencionada metodología. Esta se resume en una Sección final, donde se exponen también otras consideraciones de interés.

EL SISTEMA INTERACTIVO NNINT

Este sistema de computación (Barandela y Fuentes, 1987) para el trabajo de reconocimiento de patrones, fue creado en 1986 y desde entonces ha sido utilizado en varias tareas de clasificación relacionadas con la prospección gasopetrolífera, investigaciones sismológicas, prospección de minerales sólidos, y estudio de la conductividad eléctrica de los suelos, entre otras. Estas aplicaciones han servido también para, con la experiencia adquirida, introducirle modificaciones al sistema y hacer más cómodo y eficiente su empleo.

El nombre del sistema proviene del hecho de que la mayor parte de sus algoritmos está vinculada a un método supervisado, la regla NN (Nearest Neighbor rule o regla del vecino más cercano) que ha devenido muy popular (Devijver y Kittler, 1982; Hardin, 1994). Una de sus principales ventajas, la de contar con variantes que la facultan para enfrentar con éxito diversas situaciones prácticas es, no obstante, poco reconocida.

Entre las más importantes de esas variantes está la Edición (Wilson, 1972), que tiene el propósito, mediante la eliminación de algunos elementos de la ME, de mejorar la probabilidad de acierto del clasificador. Una modificación, la Edición Generalizada (Koplowitz y Brown, 1978), persigue el mismo objetivo pero busca reducir la cantidad de elementos eliminados.

El resultado de su aplicación es una ME reestructurada, con cambios en la identificación de algunos elementos (traslados de una clase a otra) y con eliminación de otros, lo que recomienda su aplicación para casos imperfectamente supervisados. En experimentos simulados, su empleo en forma repetida (Barandela, 1986) mostró mejoras de hasta 50% en la proporción de decisiones correctas.

Estas dos y otras variantes de la regla NN han sido incorporadas al sistema NNINT, para el trabajo interactivo con microcomputadoras tipo IBM. Una descripción detallada de las características de este sistema escapa al alcance de este trabajo. Basta mencionar aquí algunas de las más interesantes desde el punto de vista de su empleo en la práctica:

- 1.- Se ha concebido buscando una comunicación hombre-máquina que resulte fácil y cómoda, incluso para aquellos con poca experiencia previa en el empleo de computadoras.
- 2.- Se ha estructurado jerárquicamente, en forma de submenús y módulos, lo que permite, entre otras cosas, flexibilidad para añadir con facilidad cualquier nuevo algoritmo deseado por el usuario.
- 3.- Se ha tratado de hacerlo inmune, en la medida de lo posible, a errores por parte del usuario al introducir información.
- 4.- Es posible su empleo tanto con variables cuantitativas como del tipo cualitativo (escala discreta o nominal) y con una combinación de ambos tipos de atributos.

Dos opciones desempeñan un importante papel en la consecución del objetivo principal del sistema:

- Grabar, para conservar en soportes de memoria externa, cada vez que se desee, el estado vigente de la ME, después de haber sufrido modificaciones por algunos de los algoritmos.
- Reiniciar, para comenzar de nuevo con la ME original, anulando todas las modificaciones previamente efectuadas.

Estas dos subrutinas permiten un alto grado de interacción e iteración, cuya importancia se destaca en las secciones próximas.

ALGUNAS APLICACIONES PRACTICAS

De entre las numerosas aplicaciones prácticas que en el campo de la geofísica se han llevado a cabo con el sistema NNINT, se han seleccionado las tres que se exponen a continuación, por razones que resultarán obvias.

Aplicación No. 1 (Rodríguez y Barandela, 1989).

El propósito de esta aplicación fue caracterizar las secuencias estratigráficas que componen el piso estructural Parautóctono en el yacimiento de hidrocarburos Varadero, a partir del comportamiento en conjunto de algunos de los parámetros petrofísicos y geofísicos.

Se seleccionó una ME con 139 patrones agrupados en seis clases:

- 1.- Secuencia Alóctona. Litología: brechas, conglomerados y gravelitas calcáreas de diferentes tipos. Colector: fracturado.
- 2.- Secuencia Paleógena. Litología: argilitas, margas, rocas silíceo arcillosas y areniscas calcáreas. No colector en matriz arcillosa y mixto en olistolitos.
- 3.- Secuencia Varadero. Litología: calizas politomórficas, recristalizadas, arcillosas. Colector: fracturado y fracturado-poroso.
- 4.- Secuencia Sumidero. Litología: calizas arcillosas, bandeadas, recristalizadas, margas y pedernal. Colector: fracturado-poroso.
- 5.- Secuencia Artemisa I. Litología: calizas arcillosas, bandeadas, intercaladas, margas. Colector: fracturado-poroso.
- 6.- Secuencia Artemisa II. Litología: calizas arcillosas, bandeadas, intercaladas. Colector: fracturado.

Como se puede apreciar, la división en clases partió de la diferenciación estratigráfica, pero también se consideraron las características colectoras y de producción. Así es posible observar semejanzas entre las secuencias Alóctona y Artemisa II (clases 1 y 6) que son fracturadas y malas productoras, y entre Sumidero y Artemisa I (clases 4 y 5)

que son fracturado-porosas y, en general, buenas productoras de petróleo.

Para caracterizar los patrones se tuvieron en cuenta 12 variables o atributos, de los que siete son parámetros evaluados de los análisis petrofísicos de los núcleos y cinco son obtenidos de los registros geofísicos que componen el complejo estándar en escala 1:1000. Esos atributos son:

- 1.-Carbonatosidad
- 2.-Capacidad de intercambio catiónico
- 3.-Permeabilidad
- 4.-Porosidad
- 5.-Densidad (peso/volumen)
- 6.-Saturación de agua residual
- 7.-Relación resistividad-porosidad
- 8.-Diámetro de pozo
- 9.-Resistividad
- 10.-Parámetro duplo diferencial gamma
- 11.-Parámetro duplo diferencial neutrón gamma
- 12.-Potencial espontáneo

La zona que se estudió está caracterizada por un corte predominantemente carbonatado y poco diferenciado, lo que hace que la tarea resulte imprecisa y que la experiencia, intuición y conocimiento del área por parte del especialista tengan una muy fuerte influencia en el resultado. Esto conduce, en ocasiones, a caracterizaciones poco fundamentadas o contradictorias.

Estas razones aconsejaron considerar la ME disponible con posibilidades de contener algunos prototipos con identificación incorrecta, o sea, definiendo una situación imperfectamente supervisada. Por ello, y después de proceder a reducir el conjunto de variables mediante eliminación de 5 de ellas que resultaron redundantes y poco informativas (el procedimiento para seleccionar las variables a eliminar se detalla en la última Sección), se sometió la ME a un proceso de "depuración" o restructuración, consistente en varias aplicaciones sucesivas de la Edición Generalizada y, después, de la Edición también repetida. Los resultados se muestran en la Tabla 1, en donde se observa la notable reducción que el proceso produjo en las estimaciones de la probabilidad de clasificación errónea, tanto por el método C (cross-validation) como por el L (leave-one-out).

Para más detalles sobre estos estimadores se recomienda consultar el trabajo de Barandela (1988).

Tabla 1

Resultado de la aplicación No. 1

Secuencia empleada	Cant. de prototipos	Est. del error	
		C	L
ME inicial	139	66.2	42.4
Selección de variables	139	52.5	14.4
Edición Generalizada	136	23.5	5.1
Idem	136	16.9	4.4
Idem	136	13.2	2.9
Edición	120	10.8	0.8
Idem	117	10.3	0.0

En los resultados de esta depuración de la ME se observó que la mayoría de cambios de identificación (traslados de una clase a otra) se producían en patrones ubicados en los límites entre las clases, que no estaban bien precisados, como ya se aclaró. Otros cambios de identificación ocurrieron entre clases (la original representada por el prototipo trasladado y la que al final le fue asignada) en las que las características colectoras y productivas eran semejantes, aunque diferenciables en cuanto a la posición estratigráfica. Algunos ejemplos ilustrativos:

Clase 1. Se le incorporaron tres prototipos que antes representaban a la clase 6 (clases semejantes).

Clase 2. Recibió tres prototipos que pertenecían a un núcleo ubicado en el límite entre esta secuencia y la que le sigue.

Varadero (clase 3). Los especialistas revisaron y consideraron que realmente pertenecían a la clase 2.

Clase 4. Se pasaron 4 prototipos a la clase 5, en un cambio de identificación evaluado como admisible, por estar ubicados en núcleos de la zona límite que dividía a las clases. A la vez, recibió 2 prototipos que eran de la clase 5, y que correspondían a núcleos incluidos en la zona de falla, lo que, en opinión de los especialistas, justificaba el cambio.

Estos análisis condujeron a los especialistas geofísicos a la conclusión de que la ME restructurada era más coherente y de más calidad que la original.

Aplicación No.2 (Barandela, 1990)

Es lógico pensar que, cuando como resultado del empleo de un algoritmo de cluster (cúmulos) a una muestra sin identificaciones, ésta se subdivide en grupos, esa muestra agrupada puede considerarse como una ME en la que algunos prototipos tienen una identificación correcta y otros no. De aquí que resulte natural el extender el proceso de restructuración al caso no supervisado, empleándolo como complemento del algoritmo de cúmulos, sobre la ME que se obtiene como consecuencia de ese algoritmo.

Eso fue lo que ocurrió en esta aplicación. La muestra inicial consistió en 268 prototipos no identificados y correspondientes a secciones de un pozo de la zona Varadero-Cárdenas. Se midieron 6 características o variables:

- 1.-Razón Potencial espontáneo/Intensidad gamma
- 2.-Resistividad aparente (sonda potencial)
- 3.-Resistividad aparente (sonda gradiente)
- 4.-Intensidad gamma
- 5.-Porosidad neutrónica
- 6.-Variación del diámetro del pozo

De inicio, esta muestra fue dividida en 4 clases mediante un algoritmo de cúmulos (Mather, 1976), dando lugar a una ME que, a los efectos prácticos, podía considerarse como imperfectamente supervisada. Por el análisis de los registros de pozo, se pudo comprobar una corresponden-

cia aceptable entre esa clasificación y la litología descrita en el pozo, de manera que se planteó la relación siguiente entre los grupos o clases obtenidos y la litología presente:

grupo 1: arcillas, argilitas, margas

grupo 2: carbonatos (en lo fundamental calizas, brechas calcáreas, etc)

grupo 3: arcosa (areniscas, conglomerados, brechas)

grupo 4: carbonatos que pueden ser silicificados y/o fracturados.

Esta configuración fue tomada como punto inicial para la presente aplicación. Puesto que el propósito era explorar la posibilidad de manifestaciones gaso-petrolíferas, aunque la ME producida por el algoritmo de cluster comprendía 4 clases, los especialistas, atendiendo a los valores de los registros, aconsejaron considerarla de la manera siguiente:

- clases 1 y 3, conformadas por estratos sin perspectivas.
- clases 2 y 4, fuertemente asociadas con los estratos con perspectivas, según se comprobó para una parte de esos prototipos con datos de ensayo.

Los resultados del proceso de reestructuración sobre esta ME se muestran en la Tabla No 2. Aquí se aprecia también la notable reducción en los estimados de la probabilidad de clasificación errónea. En el proceso se detectó que la variable no. 4 no aportaba nada a la precisión del clasificador, y su eliminación significó una disminución del tiempo de cálculo para el trabajo subsiguiente.

La evaluación de la ME después de reestructurada, por parte de los especialistas geofísicos, resultó también satisfactoria. Las clases o grupos quedaron más homogéneas (menos variabilidad interna: en promedio, un 20% de la variabilidad original), caracterizadas con mayor precisión y mejor relacionadas con las características litológicas y colectoras del área estudiada.

Además, se observó que los cambios finales de identificación de prototipos ocurrieron entre clases consideradas como asociadas (la 1 con la 3 o la 2 con la 4).

En esta aplicación fue posible realizar una comprobación adicional pues se disponía de los datos de otras 628 secciones correspondientes a tres pozos del área, distintos al empleado para conformar la muestra inicial. Con el sistema NNINT se procedió a clasificar esa muestra independiente de 628 patrones, tomando como base la ME, antes y después de la reestructuración.

Cuando esos resultados se verificaron con la información geológica disponible (que no se utilizó al clasificar automáticamente), se observó que al emplear la ME depurada se logró una mejoría de cerca de un 40%, en comparación con la ME inicial, en la cantidad de decisiones correctas por parte del sistema de clasificación.

Esa mejoría en el comportamiento del clasificador con los patrones independientes está relacionada con otra consecuencia del proceso de depuración de la ME: se duplicó la eficiencia discriminadora de las variables, según se midió por la "cantidad de información relativa" (Bodrov y Nikitin, 1984).

Tabla 2

Resultado de la aplicación No. 2

Secuencia empleada	Cant. de prototipos	Est. del error	
		C	L
ME original	268	47.4	34.0
Edición Generalizada	130	15.4	6.2
Idem	130	2.3	1.5
Eliminar variable 4	130	2.3	1.5
Edición	122	0.8	0.0

Aplicación No. 3 (Barandela, en prensa)

Los datos utilizados en esta aplicación tuvieron su origen en un trabajo previo (Valladares, 1986) encaminado a estudiar los procesos que condicionan la formación de los colectores en la secuencia ofiolítica en la zona de Cantel-Camarioca, así como la interrelación de las propiedades gaso-petrolíferas con las respuestas a los métodos geofísicos de pozos, para la correcta evaluación de sus propiedades.

La evaluación de estas respuestas determinó un análisis teórico-práctico realizado dentro del complejo estándar conformado por los registros de potencial espontáneo, resistividad con sonda potencial o normal, sonda gradiente (8.5 m), cavernometría, gamma natural, neutrón gamma en escala 1:1000 y, en los casos en que la sonda gradiente no es de 8.5 m, se utilizó la sonda laterolog en escala 1:200. Se procesaron datos para el análisis petrofísico como son la carbonatosidad, Q100 (parámetro vinculado a la arcillosidad), densidad volumétrica, porosidad, agua residual, permeabilidad y análisis espectral.

Utilizando los datos de la centrifugación, capilarimetría y dependencias entre el agua residual y la porosidad, se establecieron los tipos de colectores para la matriz correspondiente a cada litotipo, que resultaron:

1. Serpentinitas brechadas, fracturadas; colector de tipo fracturado.
2. Cataclásitas de serpentinitas, gabrodiabasas o peridotitas (cataclastizadas); colector de tipo poroso o poroso-fracturado.
3. Serpentinita alterada; no colector.
4. Gabros, diabasas, peridotitas; colector fracturado o cavernoso-fracturado.

Todo ese proceso de análisis y evaluación condujo a la identificación de 187 secciones de pozos y a la conforma-

ción de una ME que agrupaba a representantes de las cuatro clases mencionadas. Las variables o atributos que se emplearon para esa ME fueron:

- 1.-Potencial espontáneo
- 2.-G: Sonda gradiente, 8.5 m
- 3.-N: Sonda normal, 1.5 m
- 4.-Razón G/N
- 5.-Diferencia de Caliper
- 6.-Gamma Natural
- 7.-Neutrón gamma

Los especialistas geofísicos consideraban que, como consecuencia de todo el procedimiento para su formación, esa ME podría darse como definitiva y no entendían que fuese necesaria una reestructuración. Les preocupaban, en particular, los cambios de identificación que podrían producirse en algunos prototipos.

Cuando aceptaron participar, con carácter exploratorio, en la evaluación de los resultados de un proceso de depuración, establecieron -a priori- restricciones severas acerca de los tipos y la cantidad de traslados entre clases que podrían ser aceptados.

La Tabla No. 3 recoge el desarrollo del proceso de reestructuración. De nuevo se observa la reducción substancial obtenida en los estimados de la probabilidad de error. La ME resultante fue aceptada sin reservas como mejor estructurada y más precisa que la inicial, por aquellos mismos especialistas que al principio habían expresado sus dudas. El análisis de los cambios registrados permitió observar que una parte importante de las modificaciones de identificación correspondieron a prototipos que representaban estratos analizados en los límites litológicos fuera de los intervalos nucleados de los pozos, y se consideró aceptable su reidentificación, de acuerdo con el complejo de registros geofísicos.

Influyó también en esa aceptación el hecho de que en la ME resultante no se registran traslados no justificables (de la clase 3 a la 4 o viceversa), o se registran en un nivel muy bajo (1 prototipo de la clase 1 a la 4, y 3 en sentido contrario).

Tabla 3

Resultado de la aplicación No. 3

Secuencia empleada	Cant. de prototipos	Est. del error	
		C	L
ME original	187	69.5	52.9
Edición Generalizada	150	32.7	22.0
Idem	138	18.1	14.5
Idem	134	6.0	6.7
Idem	134	5.2	6.7
Idem	134	4.5	6.7
Edición	108	3.7	0.0
Idem	105	1.9	0.0

El éxito de la depuración en este caso, que se suponía perfectamente supervisado, podría explicarse en parte por la existencia de una cierta imprecisión en la definición de las clases, es decir, por la falta de fronteras bien marcadas entre los grupos.

DISCUSION

En las geociencias, razones técnicas y económicas motivan la necesidad de procesar los datos con un bajo riesgo de error y con rapidez. Esto es cierto, en particular, en las tareas de prospección de yacimientos minerales, en las que se requiere discriminar las zonas de interés de aquellas que no lo son, para poder concentrar los esfuerzos en las zonas más favorables. Los métodos tradicionales dependen de dos factores fundamentales: la experiencia y el grado de habilidad del especialista interpretador, y el gran volumen de información con las consiguientes complejidades de cálculo. Lo anterior explica el interés por automatizar el procedimiento de análisis mediante la introducción de métodos matemáticos y la utilización de modernos equipos de computación. Entre los métodos matemáticos de más amplia aceptación se encuentran los del Reconocimiento de Patrones.

Con las aplicaciones aquí expuestas se ha pretendido ejemplificar una metodología que permite hacer el trabajo con más rapidez que el especialista humano, y lograr una mayor precisión y homogeneidad. En consecuencia, se obtiene una mejor modelación del fenómeno que se estudia.

La metodología consiste, en lo fundamental, en la aplicación de la Edición Generalizada en forma reiterada. Ha demostrado sus beneficios en las tres situaciones posibles: no supervisada (como complemento del algoritmo de cluster), imperfectamente supervisada y supervisada. En esta última, la razón para la mejoría está dada, al menos en una buena parte de las tareas geólogo-geofísicas, por la dificultad en lograr una bien definida separación entre las clases, tanto física como conceptualmente. Algunas ideas sobre uso del enfoque difuso (fuzzy) son discutidas por Barandela (1992).

La calidad de la ME resultante ha sido evaluada y confirmada por el análisis y las conclusiones de los especialistas. En la aplicación No. 2 fue posible, además, corroborar la mediante la clasificación de una muestra independiente.

La utilidad de una ME para proceder a clasificar con rapidez y alto grado de seguridad otros elementos del área de estudio, es obvia. El proceso de reestructuración tiene una ventaja adicional pues se logra una reducción significativa del tamaño de la ME, con el consiguiente ahorro en tiempo de cálculo en los trabajos subsiguientes.

Kanal (1972) afirma que al diseñar un clasificador no importa tanto la selección del método, sino escoger adecuadamente las variables. A la luz de este trabajo es posible enfatizar que resulta también de importancia fundamental la selección de los prototipos que integran la ME y la estructura de ésta. El procesamiento constituye una herramienta muy útil para el reajuste o adecuación de la ME.

Los estimadores C y L para la probabilidad de error han sido empleados como guía en la búsqueda de la mejor estrategia para el procesamiento de la ME. En la casi totalidad de las etapas que conformaron las diferentes aplicaciones, el valor estimado por C fue mayor que el de L. En efecto, se cataloga a C como un estimador pesimista, que estima probabilidades de clasificación errónea mayores que la verdadera, mientras que L sería un estimador optimista. La utilización de estos estimadores ha evitado, además, la necesidad de disponer de una muestra independiente de control. Su validez como criterio para determinar la calidad de la ME fue confirmada por las conclusiones de los especialistas.

Para eliminar las variables se empleó también una combinación de dos criterios: Divergencia y Separabilidad no probabilística (Devijver y Kittler, 1982; Tou y González, 1974, Cap. 7). En un caso (aplicación No. 1), la eliminación de variables superfluas se hizo imprescindible desde el inicio. En las demás aplicaciones el número de variables fue relativamente bajo en comparación con el tamaño de la ME y se obtuvo la misma estructura final de la ME cuando la selección de variables se hacía al principio o al final del procesamiento. En todo caso, la detección de las variables a eliminar resultaba más fácil en las últimas etapas, después de eliminar un grupo de prototipos y cuando se había deteriorado la relación inicial entre el tamaño de la ME y la cantidad de variables.

Es difícil establecer a priori una estrategia adecuada para aplicar eficientemente las técnicas incluidas en el proceso.

Factores como, por ejemplo, el número de iteraciones, los valores de los parámetros, etc, dependen del problema concreto que se desea resolver y deben ser determinados interactivamente en la computadora, en forma casuística. La selección de las variables puede realizarse de la misma manera.

En cada aplicación hubo pruebas no ventajosas (repetición excesiva de la Edición, valores inadecuados de los parámetros, etc) que fueron desechadas, hasta que se encontraba la mejor combinación y secuencia de las técnicas empleadas y de los valores de los parámetros. Un sistema de computación como NNINT facilita y hace cómoda la interacción y la iteración que son imprescindibles en esta tarea.

AGRADECIMIENTOS

Agradezco al Dr. José R. Alfonso, Profesor Titular del Instituto Superior Politécnico de La Habana, por su valiosa colaboración en el análisis e interpretación de los resultados de la aplicación No. 2, y por sus comentarios que ayudaron a mejorar la exposición en el presente trabajo.

BIBLIOGRAFIA

- ALFONSO, J. R., R. BARANDELA y M. PACHECO, 1987. Construction of gas-oil deposits prognostic maps: mathematical procedures. *En: Proc. 32th Int. Geophysical Symposium, Dresden*, 121-127.
- BAKER, J. R., S. A. BRIGGS, V. GORDON, A. R. JONES, J. J. SETTLE, J. TOWNSHEND y B. K. WYATT, 1991. Advances in classification for land cover mapping using SPOT HRV imagery. *Int. J. of Remote Sensing*, 12, 5, 1071-1085.
- BARANDELA, R. 1986. Nuevas variantes para el procesamiento de muestras de entrenamiento imperfectamente identificadas. *En: Segundo Congreso Nacional de Matemática, La Habana*.
- BARANDELA, R., 1988. Manual del sistema NNINT. La Habana.
- BARANDELA, R., 1990. Métodos de reconocimiento de patrones en la solución de tareas geólogo-geofísicas. *Ciencias de la Tierra y el Espacio*, 17, 1-7.
- BARANDELA, R., 1992. Consideraciones acerca de la clasificación automática de datos geólogo-geofísicos. *Ciencias de la Tierra y el Espacio*, 19, 1-7.
- BARANDELA, R., en prensa. La práctica de la clasificación con la regla NN. Editorial Ciencia y Técnica, La Habana.
- BARANDELA, R. y N. FUENTES, 1987. Un sistema interactivo para el reconocimiento de patrones con la regla NN. *Investigación Operacional*, VIII, 3, 35-47.
- BARANDELA, R. y E. GARCIA, 1994. El problema de la conformación de una muestra de aprendizaje adecuada, 2do. Taller Iberoamericano Informática y Geociencias, La Habana.
- BODROV, V. y A. NIKITIN, 1984. Complejo de métodos geofísicos para la exploración (en ruso). Editorial Nauka, Moscú.
- BOLSTAD, P.V. y T.M., LILLESAND, 1991. Semi-automated training approaches for spectral class definition. *Int. J. of Remote Sensing*, 13, 16, 3157-3166.
- BUCHHEIM, M.P. y T.M. LILLESAND, 1989. Semi-automated training field extraction and analysis for efficient digital image classification. *Phot. Eng. and Remote Sensing*, 55, 9, 1347-1355.
- CHEN, C.H., y H. FU, 1986. A comparison of decision rules for seismic recognition. *En: Proc. 8th Conf. on Pattern Recognition*. Paris.
- DEVIJVER, P.J. y J. KITTLER, 1982. Pattern Recognition: a statistical approach. Prentice Hall, Londres.
- FOODY, G.M., 1990. Directed ground survey for improved maximum likelihood classification of remotely sensed data. *Int. J. of Remote Sensing*, 11, 10, 1935-1940.
- FOODY, G. M., M. B. McCULLOCH y W. B. YATES, 1992. An assessment of an artificial neural network for

- image classification. Proc. 18th Annual Conf. Remote Sensing Society, Dundee University.
- HARDIN, P. J., 1994. Parametric and Nearest-Neighbor methods for hybrid classification: a comparison of pixel assignment accuracy. *Phot. Eng. and Remote Sensing*, 60, 12, 1439-1448.
- GOPAL, S. y C. WOODCOCK, 1994. Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Phot. Eng. and Remote Sensing*, 60, 2, 181-188.
- HIXXON, M., D. SCHOLZ, H. FUHA y T. AKIYAMA, 1980. Evaluation of several schemes for classification of remotely sensed data. *Phot. Eng. and Remote Sensing*, 46, 12, 1547-1553.
- HOFFER, R.M. y M.D. FLEMMING, 1978. Mapping vegetative cover by computer-aided analysis of satellite data. *Purdue Agricultural Exp. Station Journal*, 7071, 227-237.
- HOFFER, R.M. y P.H. SWAIN, 1980. Computer processing of satellite data for assessing agricultural, forest and rangeland resources. Proc. 14th Congress Int. Soc. Photogrammetry, Hamburg, Alemania.
- JUSTICE, J. H. *et al.*, 1986. Multidimensional attribute analysis and pattern recognition for seismic interpretation. *Pattern Recognition*, 18, 6, 391-407.
- KANAL, L. N., 1972. Interactive pattern analysis and classification systems: a survey and commentary. *Proc. of the IEEE*, 60, 10, 1200-1215.
- KERSHAW, C.D. y R.M. FULLER, 1992. Statistical problems in the discrimination of land cover from satellite images: a case study in Lowland Britain. *Int. J. of Remote Sensing*, 13, 16, 3085-3104.
- KOPLowitz, J. and T. A. BROWN, 1978. On the relation of performance to editing in Nearest Neighbor rules. *En: Proc. 4th Int. Joint Conf. on Pattern Recognition*, Japón.
- MATHER, P.M., 1976. Computational methods of multivariate analysis in Physical Geography. John Wiley and Sons, Londres.
- MATHER, P.M., 1987a. Computer processing of remotely sensed images: an introduction. Wiley and Sons, Chichester.
- MATHER, P.M. 1987b. Preprocessing of training data for multispectral image classification. Proc. 13th Annual Conf. of the Remote Sensing Society. Nottingham.
- RODRIGUEZ, M. y R. BARANDELA, 1989. Aplicación de algunas técnicas de reconocimiento de patrones en la caracterización estratigráfica del yacimiento Varadero. *Serie Geológica*, 2, 29-38.
- TOU, J.T. y R.C. GONZALEZ, 1974. Pattern Recognition Principles. Addison-Wesley, N. Y.
- VALLADARES, S. 1986. Metodología para la evaluación de los colectores y sus propiedades en las rocas pertenecientes al Complejo Alóctono Eugeosinclinal. Tesis Doctoral.
- WARREN, S.D., M.D. JOHNSON, W.D. GORAN y V. E. DURSING, 1990. An automated, objective procedure for selecting representative field samples sites. *Phot. Eng. and Remote Sensing*, 56, 3, 333-335.
- WILSON, D. L., 1972. Asymptotic properties of Nearest Neighbor rules using edited data. *IEEE Trans. Syst., Man and Cybern.*, SMC-2, 408-421.
- WU, X. y E. NYLAND, 1987. Automated stratigraphic interpretation of well-log data. *Geophysics*, 52, 12, 1665-1676.
- YUXION, W. *et al.*, 1986. Dividing seismic active period with Fisher's method. Proc. 8th Conf. on Pattern Recognition, Paris.
- ZHANG, Z. y N. SIMAAN, 1987. A rule-based interpretation system for segmentation of seismic images. *Pattern Recognition*, 20, 1, 45-53.

Ricardo Barandela Alonso
 Academia de Ciencias de Cuba
 Instituto de Geografía, Calle 13 No. 409 esq. con F.
 Vedada, La Habana, Cuba.