

REVISTA PUBLICADA ANUAL por la
 Universidad de La Habana en los campos de la
 Matemática y la Física. Desde su creación en 1950
 ha publicado trabajos de investigadores cubanos y
 extranjeros.

Artículos de carácter pedagógico en los que se
 discuten cuestiones técnicas y metodológicas de
 temas de Matemática y Física.

Son se publican artículos resúmenes sobre el
 estado actual de algún aspecto de estas ciencias por
 solicitud expresa del Consejo de Redacción.

CONSEJO DE REDACCION

PRESIDENTE

Dr. José Martí

SECRETARIO

Dr. Roberto

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Dr. Miguel A.

Vol. 14, No. 2-3, 1993

CIENCIAS MATEMÁTICAS

DEPARTAMENTO DE EDICIONES

1. EDICIÓN

Cervantes, G. López

REDACCIÓN

Miguel G. Vázquez Pérez

DISEÑO

Pérez

COMPOSICIÓN

Roberto Suárez López

PUBLICACIÓN

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

Cervantes, G. López

MOS CANJE EN GENERAL:

a la correspondencia a

Universidad de La Habana

Departamento de Ediciones

Calle 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

Código Postal 10400, Cuba

UNIVERSIDAD DE LA HABANA



PROGNOSIS Y SUS APLICACIONES A LAS GEOCIENCIAS

José Ruiz Schulcloper, Ramón Pico Peña, Instituto de Cibernética, Matemática y Física, ACC

Ramón López País, César Alaminos Ibarria, Centro de Estudios de Materiales Básicos, SEAN

Manuel Lazo Cortés, Marta B. Boggiano Castillo, Universidad Central de Las Villas, MES

Eligio Barreto Fiú, Alfredo Santana Machado, Instituto Superior de Ciencias Médicas - VC, MINSAP

Leonardo Álvarez Gómez, Tomás Chuy Rodríguez, Instituto de Geofísica y Astronomía, ACC

RESUMEN

PROGNOSIS es un sistema herramienta para expertos en Geociencias que emplea una técnica avanzada para proporcionarle a los especialistas de estas ramas los instrumentos indispensables para la solución de problemas de clasificación con aprendizaje; con aprendizaje parcial y sin aprendizaje. El sistema está subordinado a una metodología de la modelación matemática de los fenómenos geocientíficos para cuya solución está elaborado, que permite profundizar en los modelos geólogo-geofísicos, mejorarlos, desarrollarlos e incluso puede ayudar en la creación científica; permite el trabajo simultáneo de variables cualitativas y cuantitativas a la manera que lo hace el geocientífico; mejora la calidad del pronóstico; permite la elaboración de mapas y gráficos entre muchas otras ventajas y beneficios. **PROGNOSIS** (versión

1.2) está avalado por más de cuatro años de trabajo de un grupo interdisciplinario de profesionales de centros de investigación y de la producción y por su utilización práctica satisfactoria en pronóstico de perspectividad de rocas fosfóricas de génesis sedimentarias; de mineralización de materiales de interés para la esfera nuclear y de magnitudes máximas de terremotos en las regiones del Caribe y Cuba.

ABSTRACT

PROGNOSIS is a tool-system for Geoscience experts. An advanced technique is employed which provides the necessary tools to solve classification problems with knowledge, partial knowledge or without knowledge. The system is subordinated to a methodology of mathematical modeling of geoscientific phenomena. The system allows us to get into the Geophysical-Geologic model, improving it, developing it and even contributing to scientific research. It can work simultaneously with both qualitative and quantitative variables in the way a geoscientist does, improves prognosis quality, and makes graphics and maps among other advantages. PROGNOSIS version 1.2 is backed by the work of an interdisciplinary group of professionals from different Research Institutions or production centers, by the satisfactory practical use in detecting phosphorus rocks of sedimentary genesis, searching minerals of interest to nuclear branch, and determining the maximum magnitudes of earthquakes in the Caribbean and Cuba.

INTRODUCCION

La aplicación de modelos matemáticos y técnicas de computación a las Geociencias tiene más de cuatro décadas y sus resultados podríamos catalogarlos de discretos. En general es la situación que también caracteriza la incursión de estas disciplinas en las llamadas *ciencias poco formalizadas* (Geociencias, Medicina, Pedagogía, Psicología, Criminología y otras). En trabajos anteriores [1-4] hemos afirmado la convicción que las restricciones de los éxitos obedecen, en una buena medida, a la metodología de la modelación matemática empleada, en particular, en los problemas de reconocimiento de patrones: clasificación, diagnóstico, pronóstico, entre otros, que aparecen con mucha frecuencia en las citadas áreas de trabajo científico.

Hay que decir que los problemas a los que hacemos referencia son de los más complejos a los que se enfrenta la ciencia contemporánea. Por ejemplo, la determinación de las magnitudes máximas (M_{max}) de los terremotos es una de las tareas más difíciles de la Sismología que ha sido abordada por diversos autores mediante diferentes modelos

y enfoques [8]. Los problemas relacionados con el pronóstico de la perspectividad de materias primas minerales no es menos complejo que el anterior y también ha sido y es objeto de atención de muchos especialistas en casi todo el mundo.

Nosotros nos hemos propuesto la elaboración de una herramienta para el trabajo y la investigación de sismólogos, geólogos y geofísicos que resulte tan familiar como confiable. Obviamente esa herramienta no podía ser otra cosa que el fruto de un trabajo interdisciplinario serio, paciente, profundo, en los que cada especialista tuviera sus funciones bien definidas pero bajo una óptica común, un mismo lenguaje.

Lo que hoy exponemos es el resultado de una primera etapa del trabajo de nuestro Proyecto PROGNOSIS: un sistema automatizado avalado por la práctica de investigaciones tales como el pronóstico de M_{max} de terremotos en la zona del Caribe [8]; discriminación de anomalías aero - gamma - espectrométricas (AGE) perspectivas para rocas fosfóricas de tipo sedimentarias [6]; pronóstico de la perspectividad de materiales de interés para la esfera nuclear [7]. Este proyecto está respaldado por una serie de profesionales de una amplia red de centros de investigación y de producción. Esta composición nos garantiza tanto su actualidad científica como su utilidad práctica.

ASPECTOS METODOLOGICOS

Algunas de las características más sobresalientes de la metodología de la modelación matemática de problemas de reconocimiento de patrones que aparecen en las Geociencias son los siguientes:

- No se puede modelar lo que no se conoce.
- Los especialistas de las Geociencias no tienen que dominar el lenguaje de la Matemática.
- El problema geocientífico se formula estrictamente en el lenguaje de las Geociencias.
- No vamos a *experimentar* con sus datos para ver si la respuesta le conviene.
- La frecuencia es una de las bases del conocimiento geológico.
- La analogía forma parte de la base metodológica del pronóstico geológico.

En modo alguno se parte de un modelo matemático, sino que a ese modelo se llega, por lo que no pretendemos ajustar la realidad a modelo preconcebido alguno, sino que se busca el modelo matemático que mejor se adecue a la realidad. Adecuación quiere decir que responda a los requisitos del problema, como pudiera ser por ejemplo: trabajar simultáneamente con variables cualitativas y cuantitativas, que admite ausencia de información, etc.

No se trata tampoco que, digamos, el sismólogo le dé datos al matemático, este le aplique un modelo matemático y un computista lo realice en computadoras. Tampoco se trata que el matemático se convierta en sismólogo y este en matemático. Se trata de tener todos un nivel esencial del conocimiento de las áreas en las que no son especialistas, para poder hablar un mismo lenguaje y pensar bajo un mismo enfoque [2,4].

En esta versión (1.2) de PROGNOSIS se han incluido los modelos de algoritmos de reconocimiento de patrones que resultaron de la aplicación de la citada metodología a la solución de los problemas mencionados en la Introducción. Aunque pensamos que estos son de utilidad a una inmensa cantidad de tareas similares, nuestros pasos en estos momentos se encaminan a la ampliación de las posibilidades del mismo.

PROBLEMAS QUE RESUELVE PROGNOSIS

PROGNOSIS es un sistema herramienta para expertos en Geociencias, que emplea la tecnología más avanzada para proporcionarle a los especialistas de estas ramas los instrumentos indispensables para la solución de problemas de clasificación con aprendizaje, con aprendizaje parcial y sin aprendizaje, es decir, respectivamente problemas de clasificación a partir de muestras de dos o más clases; problemas de clasificación en dos clases pero con muestras de objetos clasificados de sólo una de ellas; o sin muestra alguna, ya que lo que se desea conocer es justamente cómo se agrupan los objetos de la población en estudio.

Estos problemas en las Geociencias son tales como: la determinación del pronóstico de M_{max} de los terremotos; el pronóstico de perspectiva de cualquier materia prima mineral a partir del muestreo de zonas de ca a uno de los tipos de perspectiva concebidos en el MGG formulado por dichos especialistas; el pronóstico de perspectiva o no de cualquier materia prima mineral a partir del muestreo de zonas perspectivas (no se incluyen en la muestra descripciones de objetos de zonas no perspectivas); agrupamiento de zonas geológicas; determi-

nación de la importancia informacional de los rasgos que intervengan en la descripción de los objetos geológicos para los tres primeros problemas mencionados anteriormente; de manera análoga de cada uno de los objetos geológicos y un conjunto de posibilidades para el análisis de los datos disponibles sin tener necesidad de hacer suposiciones sobre los mismos, ajenas a los MGG, entre muchas otras aplicaciones.

MODELOS DE ALGORITMOS DE RECONOCIMIENTO [4]

Las mencionadas tareas que tuvimos que abordar y las características concretas de las mismas nos llevaron, consecuentemente con nuestra metodología, a que las siguientes características debían ser satisfechas por los algoritmos que empleásemos para la solución de las mismas:

- admitir ausencia de información en algunas de las descripciones de los objetos geológicos;
- permitir el trabajo simultáneo de variables cualitativas y cuantitativas;
- uso de la analogía entre descripciones de objetos como eje central de los análisis;
- admitir valoraciones de expertos (subjetivas) en el análisis de los datos tanto cualitativos como cuantitativos;
- permitir el tratamiento diferenciado de los atributos empleados en las descripciones, así como de grupos de ellos;
- permitir un análisis comparativo entre los objetos geológicos y poner de relieve las relaciones existentes entre los mismos.

Para lograr estos propósitos tuvimos que formalizar estas ideas mediante los *criterios de comparación de valores de una variable*, las *funciones de semejanza*, los *algoritmos de votación* y los *algoritmos de tipo Holotipo*.

Expondremos sucintamente estos conceptos:

Sean dadas las descripciones $I(O_1), \dots, I(O_m)$ de objetos en términos de las variables $x_1, \dots, x_n$. Sea M_i el dominio de definición de la variable x_i , $i = 1, \dots, n$, completados con un signo adicional "?" que denota la ausencia de información acerca del valor de esa variable en una descripción dada.

Uno de los aspectos novedosos del sistema consiste en la forma en que se pueden definir los criterios de comparación de valores de una variable. Estos criterios ya no sólo se definen para variables cualitativas o cuantitativas, sino, que además aparecen criterios que toman

en cuenta la naturaleza de la cual se toma dicha variable, i. e., si hay errores de medición, si la variable presenta valores cíclicos, si cada intervalo de valores significa algo distinto en términos de la naturaleza del rasgo que se mide, etc. Estas cualidades de los criterios de comparación permiten que la formalización matemática sea más adecuada y representativa del MGG definido en la formulación.

Un *criterio de comparación de valores de una variable* x_s es una función definida del producto cartesiano $M_s \times M_s$ en uno de los siguientes conjuntos:

a) $\{0,1\}$. Es el caso en que expresamos que los valores son o no coincidentes.

$$1. C_s(x_s(O_i), x_s(O_j)) = \begin{cases} 1 & \text{si } x_s(O_i) = x_s(O_j) \vee v_s(O_i) = ? \vee \\ & x_s(O_j) = ?, \\ 0 & \text{en otro caso,} \end{cases}$$

siendo $x_s(O)$ el valor de la variable x_s en el objeto O ; 1 significa que los valores son coincidentes y 0 que son diferentes;

$$2. C_s(x_s(O_i), x_s(O_j)) = \begin{cases} 1 & \text{si } x_s(O_i), x_s(O_j) \in [a_p, a_{p+1}[\vee \\ & x_s(O_i) = ? \vee x_s(O_j) = ? \\ 0 & \text{en otro caso,} \end{cases}$$

siendo el conjunto donde la variable toma valores igual a la unión de intervalos disjuntos del tipo señalado.

$$3. C_s(x_s(O_i), x_s(O_j)) = \begin{cases} 1 & \text{si } x_s(O_i) - x_s(O_j) \leq \epsilon_s \vee \\ & x_s(O_i) = ? \vee x_s(O_j) = ? \\ 0 & \text{en otro caso,} \end{cases}$$

donde $\epsilon_s > 0$ es un umbral asociado a cada rasgo x_s . Por supuesto que para poder usar este criterio de comparación es necesario que los valores de los rasgos que estemos considerando admitan las operaciones señaladas, es decir, sean números.

$$4. C_s(x_s(O_i), x_s(O_j)) = \begin{cases} 1 & \text{si } x_s(O_i), x_s(O_j) \in A_p \vee \\ & x_s(O_i) = ? \vee x_s(O_j) = ? \\ 0 & \text{en otro caso,} \end{cases}$$

siendo el conjunto donde la variable x_s toma valores igual a la unión disjunta de los conjuntos A_p .

A estos criterios de comparación le denominaremos genéricamente booleanos.

b) $[0,1]$. Es el caso en que queremos expresar una cierta medida del grado de coincidencia entre los valores.

$$5. C_s(x_s(O_i), x_s(O_j)) = 1 - \frac{|x_s(O_i) - x_s(O_j)|}{\nabla x_s}$$

siendo $\nabla x_s = \max\{x_s\} - \min\{x_s\}$. Este criterio también exige que la variable sea cuantitativa y que no presente ausencia de información. En contraste con los criterios anteriores a este tipo de criterio le denominaremos genéricamente *reales*.

Es importante subrayar que en la experiencia que hemos acumulado, estos son los criterios que han surgido de manera natural y que los especialistas han podido reflejar sus criterios técnicos a través de los mismos sin renunciar a los preceptos esbozados en su MGG. Esto no quiere decir que no sea necesaria la construcción de nuevas formas de expresión de estos criterios, por el contrario somos partidarios de que proliferen, pero siempre que lo requiera el caso.

Una vez definidos los criterios de comparación de valores de una variable, es obvio que el criterio de semejanza entre objetos geológicos se ponga en dependencia de los mismos. De hecho las cosas son a la inversa, empezamos por introducir estos criterios en las variables para tratar de hacer menos compleja la labor de definir la semejanza entre dos objetos geológicos. De manera análoga a los criterios antes ejemplificados, existen una serie de expresiones que nos permiten formalizar el concepto de analogía del MGG. También estas son producto de la aplicación consecuente de la metodología antes citada.

Toda *función de semejanza* tiene como dominio el producto cartesiano del conjunto de las descripciones de los objetos por sí mismo. Por lo dicho anteriormente ese conjunto no es el producto cartesiano de conjuntos numéricos en general, sino de los dominios de definición de las respectivas variables.

Al igual que los criterios de comparación, las funciones de semejanza también pueden ser booleanas o reales. Así la imagen de estas funciones pueden ser:

a) $\{0,1\}$. En este caso expresamos que las descripciones de los objetos comparados son o no semejantes.

$$1. B(I(O_i), I(O_j)) = \begin{cases} 1 & \text{si } |(x_s : C_s(x_s(O_i), x_s(O_j)) = 0)| < \epsilon \\ 0 & \text{en otro caso,} \end{cases}$$

siendo C_g un criterio de comparación booleano, de tipo a), y ϵ un umbral dado por el usuario.

Dos descripciones de objetos geológicos son semejantes si la cantidad de variables cuyos valores en dichos objetos no son semejantes, no excede un número prefijado.

Es importante que hagamos notar algo que es de mucho valor práctico:

a) se puede definir un criterio de comparación de valores para cada variable independientemente de las restantes; b) por consiguiente en esta función de semejanza, cada variable puede tener una forma particular de identificar los valores semejantes, es decir, recibir un tratamiento diferenciado;

c) la cantidad de variables no semejantes que podemos admitir es un parámetro que podemos ajustar según la información que tengamos. En ocasiones nos interesa hacer el análisis entre dos descripciones de objetos pero atendiendo sólo a las comparaciones de los valores de una parte de dichas descripciones (subdescripciones; subconjuntos de rasgos que presentan un cierto interés, como pudiera ser un conjunto de rasgos geofísicos o geomorfológicos, etc.). A esos subconjuntos de rasgos los denominaremos *w-partes*, siendo $\bar{w} = (w_1, \dots, w_n)$ un vector booleano donde $w_i = 1$ significa que la variable x_i pertenece al subconjunto de variables que se desea considerar ($\bar{w}$ caracteriza a dicho subconjunto). En esos casos podremos usar todas las funciones de semejanza que se definan pero restringiendo las comparaciones a las variables que aparezcan indicadas en la *w*-parte y en esos casos usaremos la notación $B_{\bar{w}}(I(O_i), I(O_j))$ o también $B(wI(O_i), wI(O_j))$.

$$2. B(I(O_i), I(O_j)) = \begin{cases} 1 & \text{si } |\{x_s : C_s(x_s(O_i), x_s(O_j)) = 0\}| \leq \epsilon_2 \\ & \text{si } |\{x_s : C_s(x_s(O_i), x_s(O_j)) = 1\}| \geq \epsilon_1 \\ 0 & \text{en otro caso.} \end{cases}$$

siendo ϵ_1 y ϵ_2 parámetros que regulan la cantidad máxima admisible de rasgos diferentes y la cantidad mínima de rasgos coincidentes.

$$3. B(I(O_i), I(O_j)) = \begin{cases} 1 & \text{si el \% de rasgos coincidentes es } > \lambda \% \\ 0 & \text{en otro caso,} \end{cases}$$

siendo λ un parámetro de umbral dado por el usuario.

b) $[0,1]$. En este caso se expresa una medida del grado de semejanza entre las descripciones de los objetos.

$$4. B(I(O_i), I(O_j)) = \begin{cases} \sum P(x_i) / \sum P(x_i) & \text{si el \% de rasgos} \\ x_i \in S & x_i \in O & \text{coincidentes es } > \lambda \% \\ 0 & \text{en otro caso.} \end{cases}$$

donde $P(x_i)$ es el peso informacional del rasgo x_i ; S es el conjunto de rasgos coincidentes entre los dos objetos que se comparan; O es la $\bar{w}$ -parte que se considera y λ es un parámetro como antes.

$$5. B(I(O_i), I(O_j)) = \left(\frac{\sum_{x_i \in S} P(x_i)}{\sum_{x_i \in O} P(x_i)} \right)$$

$$6. B(I(O_i), I(O_j)) = \frac{1}{n} \sum_{p=1}^n C_p(x_p(O_i), x_p(O_j))$$

siendo n el número de variables que se están considerando y C_p un criterio de comparación cualquiera de los anteriormente definidos.

Una vez definidos los criterios de comparación de valores para cada una de las variables y la función de semejanza a la que debemos haber llegado producto de nuestro proceso de modelación matemática de problemas particulares (existen decenas de expresiones diferentes para dichas funciones en la literatura [4]) ya estamos en condiciones de empezar el trabajo con los algoritmos de clasificación.

Uno de los modelos de algoritmos de reconocimiento de patrones que de manera óptima responde a los requisitos que mencionamos anteriormente es el de los *algoritmos de votación* [4].

Estos algoritmos se caracterizan en general por seis parámetros globales denominados respectivamente: sistema de conjuntos de apoyo; función de semejanza; evaluación por fila para un conjunto de apoyo fijo; evaluación por clase para un conjunto de apoyo fijo; evaluación por clase para todo el sistema de conjuntos de apoyo y la regla de solución.

° El sistema de conjuntos de apoyo no es otra cosa que uno de los posibles conjuntos de $\bar{w}$ -partes que pueden formarse. En PROGNOSIS estos son: el conjunto de todos los conjuntos de k elementos (siendo k un parámetro que define el usuario); el conjunto de todos los conjuntos de $k_1, \dots, k_s$ elementos (donde los parámetros s y k_i $i = 1, \dots, s$ son definidos por el usuario); el conjunto potencia de los rasgos; el conjunto de todos los testores típicos de la matriz de aprendizaje (testor típico es un conjunto de rasgos que cumple las propiedades i) que no hay subdescripciones coincidentes en clases diferentes y ii)

no es posible eliminar rasgo alguno sin que deje de cumplirse i)).

- Las funciones de semejanza y los criterios de comparación de valores de una variable que aparecen en el sistema ya fueron expuestos anteriormente.

- Los parámetros globales del tercero al quinto nos darán la forma en que se desea realizar las comparaciones entre las descripciones de los objetos, qué parámetros valorar, etc.

En la evaluación por fila para un conjunto de apoyo fijo, el modelo permite ponderar el valor de la semejanza entre las subdescripciones con el peso informacional de los rasgos que están presentes en dicho conjunto de apoyo y/o peso informacional de los objetos de la matriz de aprendizaje con los que se compara el objeto a clasificar. Estas magnitudes (pesos informacionales) pueden ser también calculadas por el sistema sobre la base del concepto de testor típico [4].

En el sistema este concepto dispone de un nivel considerable de flexibilidad ya que puede usar cuatro criterios de comparación de valores de una variable, incluso de manera simultánea. En este sentido la única restricción que presenta es el uso de un concepto booleano de semejanza y eso obedece al hecho de que las matrices resultantes deben ser no difusas (por razones fundamentalmente relacionadas con la interpretación de estos en los MGG).

- Finalmente el parámetro regla de solución en el sistema se realiza de modo tal que el usuario pueda hacer uso de variantes flexibles que han respondido a las necesidades planteadas en los MGG hasta el momento.

El otro modelo de algoritmos de reconocimiento del sistema es el denominado *HOLOTIPO* [4.10]. La idea inicial de estos algoritmos, de génesis paleontológica, fue la de hallar los objetos a los que más se parecieran un grupo de objetos, es decir, buscar los *prototipos*, los *biotipos*, los *holotipos* de un grupo de objetos. Por supuesto que determinar el parecido entre objetos puede realizarse de diferentes formas y por ahí se puede mejorar mucho el concepto y por tanto la familia de estos algoritmos, en particular, dotándolos de mayor aplicabilidad. Estos algoritmos obviamente dependen mucho del concepto función de semejanza y de la definición que usemos de objetos semejantes.

A partir de la idea de Karataeva [10] de holotipo, en nuestro grupo se han obtenido una serie de resultados [4.12] en torno a la misma, que han sido incluidos en el sistema.

En cuanto a las funciones de semejanza HOLOTIPO hace uso de las funciones y criterios introducidos anteriormente. En cuanto al criterio de objetos semejantes, estos algoritmos trabajan de la siguiente manera:

Sea dada una función de semejanza y un conjunto de descripciones de objetos MA. Construimos una matriz de semejanza MS en la que está contenida el resultado de comparar cada objeto con todos los de MA. Obviamente MS es una matriz simétrica, en la diagonal alcanza el máximo.

Dos descripciones de objetos se denominan *semejantes* si $B(I(O_1), I(O_1)) > B_0$ donde B_0 se calcula por una de las siguientes variantes:

$$a) B_0 = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m B(I(O_i), I(O_j))$$

$$b) B_0 = \frac{1}{m} \sum_{i=1}^m \max_{j=1, \dots, m} B(I(O_i), I(O_j))$$

Diremos que dos objetos están en la misma *componente conexa* si son semejantes o si existe una cadena de objetos semejantes que los une. Un concepto más restrictivo que este es el introducido por Sirotinskaia [14] (en la que nuestro grupo también trabajó [4]) y que da paso a los algoritmos tipo CLASS:

Diremos que un subconjunto B de objetos es *compacto* si 1) dado un elemento cualquiera de B el objeto que más se le parece está también en B; 2) no existe fuera de B un objeto cuyo más parecido esté en B y 3) B es el conjunto más pequeño que cumple 1) y 2).

Note que no toda componente conexa es un conjunto compacto. Este último es por lo general más restrictivo que el primero.

Esto es, con las ideas de componentes conexas y conjuntos compactos, en general tenemos dos formas de agrupar los objetos, susceptibles de diferentes interpretaciones y usos. A partir de estos agrupamientos, se puede continuar con la idea de Karataeva [10] de buscar los holotipos.

Luego, dado los agrupamientos obtenidos sobre la base de la función de semejanza escogida y el valor de umbral de semejanza B_0 calculado, se procede a determinar lo que se denomina el *coeficiente de tipicidad* de cada objeto, es decir,

$$t(O_1) = t_1 = \frac{\bar{B}_1}{\frac{1}{G-1} - \sum_{\substack{j \neq 1 \\ j=1, \dots, G}} (\bar{B}_1 - B_{1j})^2}$$

donde

$$\bar{B}_1 = \frac{1}{G-1} \sum_{\substack{j=1 \\ I(O_1) \in G(O_1) \\ i \neq j}} B_{1j}, \quad G(O_1) = \{I(O_p) : B_{1p} > B_0\} \text{ es la}$$

componente conexa que tiene como raíz al objeto O_1 y $|G(O_1)| = G > 1$.

Se define como *holotipo* de $G(O_1)$ al objeto O_p tal que

$$t_p = \max_{O_c \in G(O_1)} \{t_c\}$$

es decir, su coeficiente de tipicidad es máximo.

Cada holotipo así determinado posee un *radio de influencia*, el de la hiperesfera con centro en $I(O_1)$ y radio R^1

$$R^1 = 1 - \min_{I(O) \in G(O_1)} \{B(I(O_1), I(O))\}$$

es decir, el radio de influencia es una magnitud invariante, proporcional a la medida de semejanza entre el holotipo y el menos semejante a este dentro de G_1 .

Así, dada una descripción $I(O)$ se dice que se asocia a la clase K del holotipo $I(O_1)$ si

$$B(I(O_1), I(O_j)) \geq 1 - R^1$$

lo que significa que $I(O)$ se encuentra en el interior (o la frontera) de la hiperesfera con centro en $I(O_1)$ y radio R^1 .

Las ideas del HOLOTIPO han sido utilizadas en el sistema para la solución del problema de clasificación sin aprendizaje y con aprendizaje parcial así como las ideas vinculadas con CLASS. Ambos se usan también en el sistema en los problemas de clasificación con aprendizaje para el análisis de las clases, es decir, para el estudio de las relaciones espaciales entre las descripciones de los objetos. Todas esas ideas han sufrido modificaciones y han sido enriquecidas con otras y con la experiencia de geólogos y geofísicos [12].

APLICABILIDAD DE PROGNOSIS

Por las razones metodológicas que desde el inicio hicimos referencia, este sistema no es un experimento, una sugerencia para ver si alcanza los objetivos propuestos, PROGNOSIS es una realidad avalada por su explotación en una serie de investigaciones realizadas por un equipo interdisciplinario de especialistas de diferentes centros de investigación y de producción. La investigación acerca del pronóstico de M_{max} de terremotos es una serie de trabajos, algunos de ellos ya concluidos [8], encaminados a la elaboración de mapas pronósticos en la zona que abarca la región del Caribe, América Central y una parte importante de América del Sur. En un período de cinco años hemos estudiado las características de los fenómenos involucrados. Aquí las cuestiones fundamentales son: el uso simultáneo de variables cualitativas y cuantitativas; la ausencia de información de muchas de las áreas estudiadas; la necesidad de usar valoraciones de los expertos (no mediciones) en variables de mucha importancia informacional para el estudio del fenómeno; la necesidad de evaluaciones de expertos acerca de las mediciones que se realizan. A las conclusiones que hemos arribado en estos trabajos son las siguientes:

- Se han obtenido una serie de mapas pronósticos de M_{max} de terremotos de la citada región que en términos generales corresponden al cuadro de ocurrencias de terremotos fuertes en la región y que han recibido una evaluación satisfactoria de los especialistas [1,13].
- El trabajo en esta dirección debe continuarse, buscando mejorar aún más los resultados obtenidos, así como realizar estudios a escalas más detalladas en regiones más pequeñas.

Estas investigaciones han sido llevadas a cabo con el Instituto de Geofísica y Astronomía de la Academia de Ciencias de Cuba.

Los trabajos del pronóstico de perspectiva de rocas fosfóricas de génesis sedimentarias a partir de las anomalías AGE y los de perspectiva de mineralización de materiales de interés en la esfera nuclear fueron desarrollados con el Centro de Estudios de Materiales Básicos de la SEAN.

En estas investigaciones lo característico resultó ser la necesidad también de utilizar simultáneamente variables cualitativas y cuantitativas, junto a criterios de expertos en cuanto a las magnitudes y conceptos involucrados en el estudio de las diferentes zonas. De especial significación resultó el hecho de la ausencia de descripciones de

patrones de zonas no perspectivas junto a la necesidad de pronosticar zonas de este tipo, que en muchos casos desafortunadamente son la mayoría. Esta situación nos llevó de inmediato a desarrollar modelos de algoritmos de reconocimiento de patrones con aprendizaje parcial a partir de un algoritmo propuesto por Karataeva [10] en 1965, como ya habíamos mencionado. Las conclusiones a las que hemos llegado hasta el momento en esa dirección de trabajo son las siguientes:

- La mayoría de los pesos informacionales de los rasgos obtenidos han resultado compatibles con los MGG elaborados, lo que nos ha permitido obtener un ordenamiento de los mismos en cuanto a su importancia informacional.
- Algunos resultados no esperados nos han permitido revelar regularidades no tenidas en cuenta al conformar el modelo inicial.
- La aplicación práctica de los resultados obtenidos permitió seleccionar áreas en la zona de estudio que serán sometidas posteriormente a investigaciones detalladas de campo sobre la base del nivel de perspectividad determinado. Es decir, los resultados permitieron organizar y planificar los trabajos de campo detallados y preliminares (para las de menor grado de perspectividad calculado). La reducción obtenida fue sustancial, de un área de 504 km² se decidió el estudio de un área de 50 km² correspondientes a las zonas de alto nivel de perspectividad calculado.
- En cuanto a los pronósticos de perspectividad, la mayoría de las zonas o ya había sido establecida la presencia de mineralización o existían resultados de otros levantamientos que no habían sido valorados hasta el momento y la coincidencia con estos hechos hizo pensar que efectivamente podía existir mineralización en el área; o existían razones teóricas que confirman la imposibilidad de mineralización. En resumen, los resultados obtenidos y la comparación de los mismos con los trabajos de campo y gabinete anteriores permite considerar que la calidad del pronóstico realizado es satisfactoria. Esta comparación permite además concluir que la metodología para la ejecución del pronóstico fundamentada en la definición inicial de un modelo geológico - conceptual y en la posterior modelación geólogo - matemática del mismo, resulta adecuada para la resolución de tareas de pronóstico geológico.

En general las evaluaciones de los especialistas acerca de los resultados son satisfactorias [1,5].

ESTRUCTURA OPERATIVA DE PROGNOSIS

Para su trabajo el sistema PROGNOSIS ofrece una sucesión de bloques de acciones que barren todas las necesidades de una gran cantidad de problemas que surgen en las diferentes ramas de las Geociencias.

OPCIONES DE TRABAJO. El sistema le permite abordar, como ya hemos dicho problemas de clasificación con aprendizaje, con aprendizaje parcial y sin aprendizaje. Esta decisión restringe el acceso a algunas acciones del sistema y se limitan a aquellas que son necesarias para la solución del tipo de problema seleccionado. Esta limitación de acceso, el sistema la mostrará mediante un evidente cambio de coloración en los indicativos de las mismas y obviamente por la imposibilidad real de acceder a ellas.

DATOS. Le brinda las dos únicas entradas que posee el sistema: LECTURA de problemas que ya han sido formulados anteriormente y se encuentran en un fichero en la memoria de la computadora y CREAR que permite formularlo dando las variables que describen a los objetos, sus características, valores, etc. En los casos de problemas que requieran el manejo de mapas, es el momento de advertir al sistema de esa necesidad para que el mismo se disponga a pedirle en el momento oportuno los datos que serán necesarios para la elaboración de mapas.

En este bloque se ofrecen también un conjunto de facilidades para comunicarse con el sistema operativo, cambiar de torre, directorio, salvar datos, etc.

EDITOR. Aunque no es el único en el sistema, en este se concentran las principales opciones para realizar modificaciones a los datos y criterios formulados: cambiar valores, eliminar o agregar objetos y/o variables, formar manual o automáticamente las matrices de aprendizaje y de control. (Si es necesario en el problema)

ANALISIS. Aquí ya se pueden realizar algunos cálculos tendentes a esclaircernos la composición de nuestros datos, la estructura del conjunto de descripciones en su espacio de representación (que ni con mucho es uno de esos nobles espacios con estructuras bien estudiadas en la Matemática). Así, se han introducido opciones para el cálculo de máximos, mínimos, media aritmética y otros que tienen la característica de no exigir prácticamente nada a los componentes del problema ajeno a los postulados de los MGG. También para los problemas de clasificación con aprendizaje se dan las opciones de calcular el peso informacional de rasgos y/o objetos (por la vía de la Teoría de Test [3,4]) y también el análisis de las clases (haciendo uso de los algo-

ritmos de agrupamiento de objetos antes mencionados). De igual manera el usuario podrá visualizar en gráficos y tablas la distribución de los valores de las variables, la relación espacial entre los objetos, etc.

En estas opciones hemos preferido dar estrictamente las posibilidades de cálculo que estamos seguros no permitirán la comisión de errores metodológicos de consideración. Este es un tópico abierto a la discusión en nuestro equipo.

SOLUCION. Es el bloque donde realizamos las acciones fundamentales del sistema, a saber, la selección del algoritmo de solución del problema planteado, sus peculiaridades, exigencias, etc., y la explotación del mismo en un conjunto de objetos a clasificar.

Para esta primera etapa, obviamente de alta significación, decisiva en buena medida para la calidad de la solución, hemos dotado al sistema de una gran variedad de posibilidades (de decisión del experto o automáticas), todas ellas producto de nuestro trabajo de modelación matemática. Las familias de algoritmos de reconocimiento que en la actualidad integran el potencial científico del PROGNOSIS son las que por el momento responden adecuadamente (en ocasiones de forma ideal) a los requerimientos de los problemas abordados en la práctica por el equipo de investigadores que hemos hecho realidad el sistema [1,5,6,7,8,13]. Para los problemas de clasificación con aprendizaje se utilizan los de la familia de algoritmos de VOTACION (extendidos) [4], para los otros dos tipos de problemas se usarán los de la familia de algoritmos del tipo HOLOTIPO [12,15] que también serán usados en el análisis de los datos en el caso de problemas del primer tipo.

El usuario tiene la posibilidad de formar su algoritmo en específico, ya que nuestra experiencia nos ha permitido elaborar un conjunto de variantes que satisfacen las necesidades hasta ahora enfrentadas (la programación modular del sistema nos permitirá de manera casi que operativa, realizar las ampliaciones que sean necesarias en un futuro). Como además hemos realizado el trabajo con el objetivo de elaborar un SISTEMA HERRAMIENTA [4,9,11], el usuario encontrará toda la información requerida para llevar a efectos este paso. Obviamente el proceso de elaboración de un algoritmo que responda a las exigencias metodológicas de un problema real de las Geociencias no es simple y requerirá de una serie de etapas, en el caso que no se decida dejar enteramente al sistema esa responsabilidad.

PROGNOSIS cuenta con una adecuada información en línea, para las

decisiones importantes, un extenso Manual del usuario, en el que se ha ejemplificado con los trabajos reales realizados por el equipo y que sin dudas será de una inapreciable ayuda para el trabajo no sólo con el PROGNOSIS. Además se elabora una monografía contentiva de todo lo necesario desde el punto de vista matemático, computacional y geocientífico. Es decir, el usuario dispone de una información escalonada que le permite ir profundizando en la medida que lo requiera (y pensamos que puede ir todo lo lejos que desee) y por ello el lenguaje utilizado hace un uso gradual, comedido, de tecnicismos pero sin perder rigor.

LISTADOS. Aquí podrá revisar por pantalla o imprimir en un formato previamente establecido en parte por el usuario, toda la información acerca del problema.

MAPAS. El trabajo opcional con mapas reviste especial interés para algunos problemas que aborda PROGNOSIS. Por ejemplo, la elaboración de mapas pronósticos de M_{max} de terremotos [8,13], o mapas de perspectiva de alguna materia prima mineral [5,6,7]. PROGNOSIS ofrece una amplia gama de posibilidades que van desde el uso de coordenadas geográficas u opcionalmente geodésicas hasta la posibilidad de utilizar cualquier contorno de cualquier parte del mundo. (Siempre que se disponga del fichero correspondiente de datos)

SITUACION. Es una opción pensada para el manejo de diferentes problemas ya formulados. De manera inmediata le brindará toda la información acerca de la tarea. (Lo que se ha hecho, lo que le falta, etc.)

Finalmente, de las características técnicas de PROGNOSIS diremos que el mismo ha sido programado en lenguaje Turbo C versión 2.0 y Turbo ASM versión 2.0, con diseño modular que garantiza un futuro desarrollo de sus posibilidades y facilidades. El sistema lo conforman tres ficheros, compactados en un solo disco flexible.

Para su explotación PROGNOSIS requiere de una IBM PM, modelo XT, AT o COMPATIBLE con 256 KB mínimo de memoria operativa y con tarjeta gráfica CGA, EGA o VGA. A excepción de los mapas y gráficos, todas las restantes operaciones del sistema están adaptadas para tarjetas monocromáticas. No es imprescindible un procesador 8087, 80287, 80387, etc., aunque si está presente alguno de ellos muchos procesos de cálculo mejorarán su velocidad.

Es necesaria una versión 3.0 o posterior de sistema operativo DOS o MS-DOS.

CONCLUSIONES

- * PROGNOISIS garantiza una metodología para la realización de investigaciones de clasificación en las Geociencias que en muchos casos ha evitado la comisión de errores conceptuales importantes en los MGG y que ha permitido profundizar en los conceptos geocientíficos que se emplean.
- * PROGNOISIS es un sistema herramienta para expertos en Geociencias que le permite el uso simultáneo de variables cualitativas y cuantitativas a la manera que lo hace el geocientífico; el tratamiento de objetos con ausencia de información; la clasificación se realiza sobre la base de concepciones geólogo - geofísicos del usuario y no de presupuestos ajenos a sus modelos; mejora la calidad de sus pronósticos; permite hacer estudios con diferentes MGG; no es una *caja negra* ya que cuenta con un sistema de AYUDAS en línea y a través del Manual del usuario y de una monografía redactados en un lenguaje totalmente accesible al geocientífico, que le permiten conocer con toda la profundidad lo que desee saber del sistema; permite conocer de manera extensiva, las relaciones existentes entre los objetos, los atributos y simplificar sus estudios sin perder información, reduciendo los costos de la investigación y finalmente y no por menos importante, la entrada, manipulación y salida de mapas y gráficos donde se visualizan los resultados alcanzados.
- * PROGNOISIS es uno de los pocos sistemas capacitados para abordar el problema de la clasificación con aprendizaje parcial de sus clases, con lo cual se coloca entre los menos que permiten toda la gama de problemas de clasificación y tareas afines.
- * Los resultados prácticos alcanzados ya por la explotación de PROGNOISIS son importantes aportes a las investigaciones que se llevan a cabo en el país, en el terreno de las geociencias y constituyen un punto de partida de investigaciones geomatemáticas bajo un nuevo enfoque de la solución de problemas de clasificación en esta importante zona del conocimiento.
- * PROGNOISIS no es un sistema perfecto (¿existen?), lo bueno que tiene en su posibilidad de mejorar en manos del usuario y en interacción con nosotros ya que el mismo está respaldado por una amplia red de profesionales de diferentes centros de investigación y de producción que garantizan su actualidad científica y su utilidad práctica.

Las versiones futuras de PROGNOISIS, en las que ya se trabaja, serán

el complemento necesario, la continuidad progresiva que todo sistema debía de poseer.

REFERENCIAS

- [1] RUIZ SCHULCLOPER, JOSE; et al. (en prensa): "Modelación matemática de fenómenos geocientíficos". Proceedings del I Simposio acerca del desarrollo de la Matemática 2-5 de abril de 1990. Cuba.
- [2] CHEREMESINA, EUGENIA V.; JOSE RUIZ SCHULCLOPER (en prensa): "Cuestiones metodológicas de la aplicación de modelos matemáticos de reconocimiento de patrones en zonas del conocimiento poco formalizado". Revista Ciencias Matemáticas.
- [3] RUIZ SCHULCLOPER, JOSE (1987): "Modelos matemáticos para el reconocimiento de patrones". (Notas de curso homólogo). México.
- [4] RUIZ SCHULCLOPER, JOSE (1990): "Modelos matemáticos para el reconocimiento de patrones". UCLV. Cuba.
- [5] ALAMINOS IBARRIA, CESAR, et al. (en prensa): "Aplicación de los métodos de reconocimiento de patrones lógico-combinatorios a la resolución de tareas geológicas". Proceedings de GEOLOGIA'89. Cuba.
- [6] RUIZ SCHULCLOPER, JOSE, et al. (en prensa): "Modelación matemática del problema de discriminación de anomalías AGE perspectivas para rocas fosfóricas de génesis sedimentarias". Revista Ciencias Matemáticas.
- [7] RUIZ SCHULCLOPER, JOSE, et al. (en prensa): "Modelación matemática de pronóstico de perspectiva de minerales de interés para la esfera nuclear". Revista Ciencias Matemáticas.
- [8] RUIZ SCHULCLOPER, JOSE, et al. (en prensa): "Modelación matemática del pronóstico de magnitudes máximas de los terremotos en la región del Caribe". Revista Ciencias Matemáticas.
- [9] RUIZ SCHULCLOPER, JOSE; RAMON PICO PEÑA (en prensa): "Un nuevo enfoque en la construcción de sistemas de reconocimiento: los sistemas herramientas". Revista Ciencias Matemáticas.
- [10] VORONIN, YU.A.; G.N. KARATAEVA; E.N. EPSHTEIN (1968): "Programas HOLOTIPO Para la solución de problemas de reconocimiento de patrones" (en ruso). Alma Atá.

- [11] RUIZ SCHULCLOPER, JOSE: "El problema de la selección de rasgos y los sistemas herramientas". En: Reconocimiento de elementos de estructuras espaciales. Ed. V.G. Guitis; J. Ruiz Schulcloper; L. Alvarez Gómez, ACC.
- [12] RUIZ SCHULCLOPER, JOSE: "Algoritmos de estructuración espacial de datos". En: Reconocimiento de elementos de estructuras espaciales. Ed. V.G. Guitis; J. Ruiz Schulcloper; L. Alvarez Gómez, ACC.
- [13] ALVAREZ GOMEZ, L. et al. (en prensa): "Estimación de las magnitudes máximas de terremotos para la región del Caribe mediante técnicas de reconocimiento de patrones". II Congreso Internacional sobre Desastres", 4-7 de julio de 1989. Cuba.
- [14] SIROTINSKAIA, S.V. (1986): "Métodos lógicos de análisis de la información geológica" (en ruso). Nedra.
- [15] RUIZ SCHULCLOPER, JOSE (inédito): "MARAP: modelo de algoritmos de reconocimiento con aprendizaje parcial". IV Congreso Nacional de Matemática y Computación. COMPUMAT'91, Holguín, 11-16 de noviembre, 1991. Cuba.