Capítulo 2.- Clasificación supervisada

Introducción al segundo capítulo

El problema de la clasificación supervisada (con aprendizaje) ha sido abordado desde ópticas diferentes, a partir de enfoques distintos. Así, suponiendo que los objetos están descritos en forma de n-uplos de un cierto espacio, se han desarrollado modelos basados en *superficies de separación*, cuya idea esencial consiste en que un conjunto de objetos representados en un espacio normado, digamos $\mathfrak{R}^n$, (denotando por $\mathfrak{R}$ al conjunto de los números reales), puede ser separado por una cierta hipersuperficie, (superficies de dimensión mayor que 3) si éstos pertenecen a clases diferentes. Presuponiendo además que los objetos, al agruparse en clases diferentes, se deben ubicar en dicho espacio de modo tal que los que pertenecen a la misma clase están más cercanos entre sí que con respecto a clases diferentes. Esto en la práctica no siempre es así.

Teniendo esta misma idea como base metodológica, se ha desarrollado otro modelo denominado *vecinos más cercanos* cuya idea esencial es la distancia entre los objetos. Al ubicar los objetos en el espacio, que también se supondrá normado, decidiremos por una u otra clase en dependencia de cuán cercanos estén éstos respecto a ciertos objetos de los ya ubicados en las diferentes clases definidas en el problema. La selección de esos objetos “distinguidos” entre los ya clasificados se realiza de varias formas, todas basadas en la idea fundamental de este modelo: la cercanía entre las representaciones de objetos en el espacio seleccionado con las características antes mencionadas.

Una exposición de estos modelos la podrá hallar en el libro “Pattern Recognition Principles” de los autores J.T.Tou y Rafael C. González.

En los modelos anteriores, las ideas esenciales de los mismos han estado relacionadas con ciertos aspectos “geométricos”, en cierto sentido “topológico”, que pueden aparecer en el espacio de representación de los objetos admisibles. Bien la superficie de separación, bien la menor distancia a un n-uplo, o a un conjunto de ellos, requieren del agrupamiento de los n-uplos en el espacio, requieren de una cierta disposición espacial para que los modelos se aproximen con una aceptable eficiencia a la realidad. Más aún, requieren de determinadas características de ese espacio de representación inicial (ver pag. 28 ERI), como por ejemplo la de permitir definir una distancia al menos.

En muchas disciplinas, como la Medicina, las Geociencias, la Criminalística, la Sociología, etcétera, los objetos se describen en términos de variables cuantitativas, cualitativas, pudiendo simultáneamente presentar ausencia de información. Por esta razón es útil el **enfoque lógico-combinatorio para la solución de problemas de clasificación** y su aplicación a la solución práctica de problemas en las áreas mencionadas.

El primer modelo que vamos a estudiar descansa sobre la idea de las *precedencias parciales*. Se trata del modelo de **algoritmos para el cálculo de las evaluaciones**, también denominado **algoritmo de votación.** Las precedencias parciales constituyen una significativa forma de pensar en las ciencias naturales, de donde la tomó uno de sus iniciadores, el ruso Yuri I. Zhuravliov. En un inicio, como veremos, el modelo se concibió en forma muy simplificada (en comparación con las exigencias de los problemas prácticos en las citadas disciplinas), como modelo Booleano. Estos estudios se iniciaron alrededor de los años 1965 en el Centro de Cálculo de la entonces Academia de Ciencias de la URSS (AC URSS) y fueron las Geociencias (en particular la prospección geológica) la motivación del algoritmo. Estas ideas se fueron extendiendo a otros países, notablemente en Polonia, Hungría, Checoeslovaquia, Bulgaria, Vietnam, Korea, Alemania y Cuba. No en todos estos países las aplicaciones han alcanzado el mismo nivel de realización práctica, de hecho no todos compartimos los mismos criterios metodológicos. En la actualidad estos estudios abarcan otros países iberoamericanos.

El algoritmo **KORA-3**, el segundo de los modelos basados en las precedencias parciales que estudiaremos, fue elaborado por un grupo de colaboradores de Instituto de Problemas de Trasmisión de la Información de la AC URSS, dirigido por M. M. Bongard, con el fin de utilizar el programa para la solución de problemas de Geofísica y Geología.

El KORA-3 ocupa un lugar especial en la historia del desarrollo de los métodos de Reconocimiento de Patrones en la Geología. En primer lugar este fue el primer programa de clasificación supervisada que se utilizó en la solución de problemas geológicos. En segundo lugar es el programa de más larga vida, activo aún después de 30 años.

Este modelo es también de corte lógico-combinatorio y, como en el caso de los algoritmos de votación, ha sido inspirador de una gran cantidad de investigaciones. Se han hecho y continúan haciendo modificaciones y generalizaciones. Kora-3 se relaciona directamente con el modelo de algoritmos de votación.

Otro modelo cercano en cuanto a las ideas que lo sustenta es el de **algoritmos basados en los conjuntos representantes**. Desarrollado también en el Centro de Cálculo de la AC URSS, el modelo propuesto por L. V. Baskakova y Yuri I. Zhuravliov, desarrolla ideas cercanas a los anteriores modelos citados y es también generador de investigaciones en la actualidad en el sentido de extensiones de sus posibilidades.

El último conjunto de algoritmos de clasificación con aprendizaje en el enfoque lógico-combinatorio que estudiaremos en este texto está basado esencialmente sobre la idea de la importancia informacional de los objetos, es aún un modelo por desarrollar que sin embargo ha sido empleado en la modelación de algunos problemas en el área de la salud, a estos les hemos denominado **algoritmos basados en el peso de los objetos**.

§2.1.-Una aproximación conjuntual

Es natural considerar el problema de la clasificación supervisada desde el punto de vista de la Teoría de Conjuntos, ya que la solución de un problema cualquiera de clasificación es, en esencia, estudiar las relaciones de pertenencia de ciertos elementos respecto a ciertos conjuntos.

En la teoría de conjuntos existen **tres conceptos primarios**: conjunto, elemento, pertenencia. Esto significa que son conceptos que no se definen, sólo se determinan.

Los **conjuntos** no se definen, no existe un concepto en término del cual se pueda dar la definición de conjunto. Colección, agrupamiento, grupo, y otros, no son más que sinónimos de conjunto y ninguno de ellos es conceptualmente, semánticamente anterior al de conjunto. Así pues, no los podemos definir pero si se pueden determinan. Existen dos formas de determinar un conjunto:

A) **por extensión** (característico de conjuntos finitos) listando los elementos del conjunto,
B) **por intención** (característico de conjuntos más grandes o infinitos) indicando la propiedad que deben cumplir los elementos del conjunto.

Existen dos formas de saber si un **elemento** ***pertenece o no*** a un conjunto:

1. si se tiene el listado de todos sus elementos componentes, se comprueba si el elemento aparece en el mismo,
2. si se conoce la propiedad que lo caracteriza, que lo determina, se verifica si la cumple o no.

Sin embargo, en muchas ocasiones no se tiene el listado completo de los elementos de un conjunto ni se conoce la propiedad que caracteriza al conjunto de manera unívoca o se conoce la propiedad de una manera poco precisa como por ejemplo "las personas muy altas", "las personas sanas". Puede ocurrir que se conoce una formulación que presupone determinados conocimientos previos de la persona que tiene que decidir en cuanto a la pertenencia de un elemento a un conjunto dado, como por ejemplo: "el conjunto de versos de Gabriela Mistral", "el conjunto de pacientes de infarto al miocardio en el hospital La Raza el día 9 de octubre de 1997".

Ejemplo 2.1

Sean los conjuntos: A={1,2,3,4,5} y B={6,7,8,9,10}. Dado un número, digamos el 3, el problema se resuelve revisando los listados que determinan los conjuntos A y B respectivamente. Obviamente este no es un problema de reconocimiento de patrones. En este caso $3 \in A$.

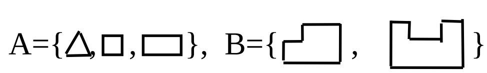
Ejemplo 2.2

Sean los conjuntos: A={números pares}, B={números impares}.

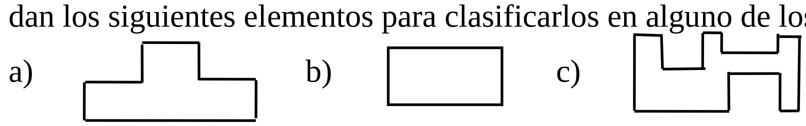
Dado un número (por ejemplo el 25) se debe ver la propiedad que lo caracteriza para ver a cuál pertenece. En este caso 25 es impar entonces $25 \in B$.

Ejemplo 2.3

Sean los conjuntos de figuras geométricas (G. Birkhoff Matemática y Psicología, Moscú 1977):



Se dan los siguientes elementos para clasificarlos en alguno de los conjuntos:



en este caso no se tiene un listado con todos los elementos de cada conjunto ni se dice cuál es la propiedad que caracteriza a cada conjunto, por lo que se impone un análisis de la información dada y se une a otros conocimientos (en este caso de geometría) para concluir que los elementos del conjunto A tienen la propiedad: dados dos puntos cualesquiera, los puntos de la recta que los une quedan dentro de la figura, esto es "figura plana convexa". Es claro que la propiedad enunciada no la cumple ninguna figura del conjunto B.

Para clasificar a los objetos dados se debe analizar su propiedad característica y luego asociar el elemento con algún conjunto. En este caso a), c) $\in$B; b) $\in$A.

Ahora, qué sucedería si por alguna razón no conociéramos la propiedad de la convexidad de las figuras planas y de todas formas tuviésemos que decidir la misma cuestión. En ese caso no cabría otra alternativa que tratar de asociar las figuras por el "parecido".

Vuelva sobre el ejemplo de Birkhoff, olvídese de la convexidad y trate de clasificar las figuras anteriores.

Creemos que todos coincidiremos en que las cosas ya no resultan tan claras como cuando se contaba con la propiedad de convexidad. No faltará quien encuentre figuras "parecidas" a las que queremos clasificar tanto en K_1 como en K_2 y quizás se pueda, en algunos casos, coincidir con las respuestas anteriores, en otros errar y en algunos abstenernos por las dudas. No vamos a tener en cuenta a los que redescubrirían el concepto de convexidad.

Ejemplo 2.4

Consideremos el problema del diagnóstico médico. Sobre la base de la información almacenada en las historias clínicas se tiene una sucesión de pacientes $P_1,...,P_{m_1}$ que padecen la enfermedad E_1 y los siguientes $P_{m_1+1},...,P_{m_2}$ padecen la enfermedad E_2 y así sucesivamente se tiene información de las enfermedades $E_1,...,E_r$; sin descartar la posibilidad de que un mismo paciente padezca más de una de las enfermedades.

No se puede dar un listado completo de las personas que padecen y padecerán cada una de las enfermedades (caso extensional). En muchos casos, tampoco, se tiene una buena propiedad (o una sucesión de ellas) que caracterice a cada conjunto como en el ejemplo 3. En este caso del diagnóstico médico, existen muchas situaciones en las que tampoco resulta fácil, del análisis de los datos, descubrir tales propiedades.

En estas condiciones aparece el problema de decidir la o las enfermedades que padece un nuevo paciente (que antes no se había tratado). Este ya es un problema de clasificación.

De manera similar al ejemplo 2.4, en las Geociencias aparecen casos análogos para determinar la posible existencia de un yacimiento de recursos minerales para su explotación industrial o no; hacer mapas de pronósticos de magnitudes máximas de terremotos; encontrar conjuntos más pequeños de propiedades que permitan hacer los pronósticos a un costo más bajo; y otras muchas investigaciones del mismo tipo. Estos ya son problemas de clasificación. Situación análoga aparece en otras zonas poco formalizadas del conocimiento y de la práctica profesional.

Por lo tanto el planteamiento del problema de Clasificación Supervisada (Con Aprendizaje) en el seno de la Teoría de Conjuntos, es un primer paso de abstracción en la búsqueda de un modelo matemático adecuado para su solución efectiva.

En casos como el ejemplo 2.4 se han utilizado técnicas estadísticas, probabilísticas, la teoría de conjuntos difusos, funciones potenciales, lógica matemática clásica y polivalente, lingüística matemática, combinatoria, teoría de grafos, ecuaciones diferenciales y otros modelos de la Matemática. Cada uno de ellos ha tenido ventajas y desventajas, éxitos y restricciones.

Como se dijo anteriormente, no hay un modelo omnipotente para la solución de un problema que esencialmente es el mismo: **decidir la pertenencia de un objeto a un conjunto dado.**

El problema también puede verse desde el punto de vista de formular un modelo que reduzca la información a representar o guardar. En el ejemplo de los pacientes enfermos, una lista de todos los pacientes y sus enfermedades respectivas es una representación completa pero extensa del conjunto. Si halláramos las propiedades que cada enfermedad presenta o posee, entonces el modelo sería más compacto (menos información), al decirnos Tuberculosis = {tos seca, fiebre alta, sangre en el esputo, ...}; sífilis = { ...}, etc.

§2.2.-Algunos conceptos básicos

Definición Se llama *r-uplo de pertenencia de* O a $\overline{\alpha}$ (O)=(α_1(O),...,α_r(O)), donde α_i(O)≡"O$\in$K$_i$" siendo α_i(O)$\in${0,1,$*$}. Si α_i(O)$\neq *$ para i=1,...,l se llama r-uplo de pertenencia *completo*. Si α_i(O)$\neq *$ implica que α_i(O)=P_i(O) siendo P_i(O) el predicado que describe correctamente la pertenencia de O a K_i, se denomina r-uplo de pertenencia *correcto*. Se denomina r-uplo de pertenencia *verdadero* si es completo y correcto.

α_i(O)	significa que
0	$O \notin K_i$
1	$O \in K_i$

$*$	no se sabe si el objeto O pertenece o no a la clase K_i

Definición Por *información estándar de las clases* $K_1,...,K_r$ se entiende un conjunto de descripciones de objetos de los cuales se conocen sus relaciones con las clases $K_1,...,K_r$, la cual será denotada por $I_0(K'_1,...,K'_r)=\{I(O_1), \overline{\alpha}(O_1),...,I(O_m), \overline{\alpha}(O_m)\}$ donde

1. $I(O_i)$ es la descripción de O_i y
2. $\overline{\alpha}(O_i)$ su r-uplo de pertenencia.

$I_0(K'_1,...,K'_r)$ es una información estándar *correcta* (*completa, verdadera*) en dependencia de que todos sus r-uplos de pertenencia sean correctos (completos, verdaderos); $K'_r \subset K_i$, i=1,...,r.

<u>Nota aclaratoria</u>.- En lo sucesivo se supondrá que se trabaja con información estándar verdadera de las clases $K_1,...,K_r$.

El **planteamiento formal del problema de clasificación supervisada** se formula como sigue: Sea M un conjunto de objetos admisibles; $x_1,...,x_n$ rasgos que caracterizan a los elementos de M y cuyos conjuntos de valores admisibles son $M_1,...,M_n$ sobre los que se han definido criterios de comparación $C_1,...,C_n$ respectivamente; $K_1,...,K_r$ es un cubrimiento finito de subconjuntos propios de M; dada una información estándar verdadera $I_0(K'_1,...,K'_r)$ de las clases $K_1,...,K_r$ donde $K'_r \subset K_i$, i=1,...,r; dada la sucesión $I(O'_1),...,I(O'_q)$ de descripciones de objetos de M que pudieran no ser completas de objetos admisibles $O'_1,...,O'_q$ (se llama *muestra* o *matriz de control*). El problema consiste en hallar un algoritmo A tal que:

$$A(I_0(K'_1,...,K'_r), I(O'_1),...,I(O'_q))= \left\| \alpha_j^A(O'_i) \right\|_{q\times r}$$

siendo $\alpha_j^A(O'_i) \in \{0, 1, *\}$ la respuesta del algoritmo A en cuanto a la pertenencia de O'_i a la clase K_j, $\left\| a_{ij} \right\|_{q\times r}$ denota una matriz de q filas y r columnas de valores a_{ij} en nuestro caso los r-plos de pertenencia, $*$ denota la abstención del algoritmo A a clasificar a O.

<u>Nota aclaratoria</u>.- Cuando no sea imprescindible especificar el nombre de las clases con las que estamos trabajndo emplearemos $I_0(K'_1,...,K'_r)$=MA

Se puede suponer que el algoritmo puede equivocarse en cuanto a la correcta asignación de un objeto a una clase dada, es decir, $\alpha_j^A(O'_i) \neq P_j(O'_i)$, donde $P_j(O'_i)$ está denotando la propiedad que correctamente caracteriza la pertenencia de O'_i a la clase K_j. En otras palabras los r-uplos que produce A para cada objeto a clasificar O'_i pudieran no ser completos (abstenciones) e incluso no ser correctos (malas clasificaciones). Esta suposición que resulta muy plausible, conlleva el que a cada algoritmo que se elabore se le asigne cierta medida de su "eficiencia relativa", es decir, una medida de sus aciertos, errores y abstenciones respecto a la información de aprendizaje y la de control. Para ello se introduce el concepto de *funcional de calidad*. Este concepto obviamente tiene que responder a los criterios que con relación a dicho aspecto dió el especialista del área de aplicación en la etapa de la formulación del problema.

Definición Sea dada una familia finita $\{f_i\}$ de funciones de i variables definidas para valores $x \geq 0$, no crecientes en cada variable y que alcanzan su máximo en el i-uplo nulo. Sea D un criterio de comparación entre tuplos (usamos este término cuando se específica la cantidad de coordenadas, no necesariamente numéricas, del mismo) cualesquiera (una distancia, por ejemplo):

$$\Phi(A) = f_q(D(\overline{\alpha}_1, \overline{\alpha}_1^A),...,D(\overline{\alpha}_q, \overline{\alpha}_q^A))$$

denota al *funcional de calidad* del algoritmo A para la sucesión de objetos $O'_1,...,O'_q$, siendo $\overline{\alpha}_i$ los r-uplos verdaderos de los objetos O'_i y $\overline{\alpha}_i^A$ los r-uplos de pertenencia construidos por el algoritmo A para dichos objetos, i=1,...,q.

Ejemplos de funcionales de calidad

1. Sea X el número de objetos bien clasificados por un algoritmo A, Y la cantidad de objetos mal clasificados y Z la cantidad en la que el algoritmo se abstiene de clasificar.

$$\Phi(A) = \frac{X}{X+Y+Z}$$

En este caso se busca maximizar el resultado dado que la calidad es directamente proporcional a los aciertos, independientemente de cómo se comporten los errores y abstenciones. A este funcional le da igual un error que una abstención.

2. $\Phi(A) = \dfrac{\alpha Y + \gamma Z}{X + Y + Z}$

donde α y γ son parámetros que evalúan los errores y las abstenciones de modo diferente. En este caso se busca minimizar la función. No se distinguen los errores entre sí, ni las abstenciones entre sí. Hay casos en que conviene evaluar diferentemente distintos tipos de errores, cosa que el funcional anterior no hace. Ejemplo: no es lo mismo clasificar una situación de la clase de emergencia sísmica en la clase de calma que dar la falsa alarma de una emergencia. En el último caso las consecuencias, gastos, etc. pueden ser muchísimo menores que en el primero. No siempre es lo mismo abstenerse en un caso de emergencia sísmica que hacerlo en una situación de calma.

Esto significa que los ejemplos de funcionales de calidad no deben ser vistos en abstracto, fuera del problema que modelan.

3. Una generalización del primer ejemplo es:

$$\Phi(A) = \frac{1}{N}\sum_{i=1}^{q}\sum_{j=1}^{r}\gamma_{ij} \qquad \textbf{(9)}$$

donde N es un factor de normalización y $\gamma_{ij} = \Phi_{ij}^{\alpha_{ij},\alpha_{ij}^{A}}$ es una evaluación numérica del suceso que dependa del valor verdadero en cuanto a la pertenencia del objeto O'_i a K_j y del valor que da el algoritmo A al respecto.

Observe que este funcional pudiera utilizarse para problemas de clasificación con clases no disjuntas y admitiendo la multiclasificación, pero sin distinguir entre errores y abstenciones.

4. Un caso particular de (9) está dado por la tabla siguiente. En este caso sólo contribuyen a la "calidad" del algoritmo A las coincidencias con el valor verdadero, mientras que las abstenciones y los errores están a un mismo nivel, no aportan valor a la calidad del algoritmo.

α^{A}_{ij} \ α^{A}_{ij}	0	1	*
0	1	0	0
1	0	1	0

5. El siguiente es un funcional en el que la calidad se mide en términos de sus errores; diferencia el error de la abstención, diferencia el error entre las clases (no hay la misma gravedad relacionando un objeto que pertenece a K_i con K_j que a la inversa) y diferencia las abstenciones entre las clases (no es lo mismo abstenerse en una clase que en otra). En este caso se busca que el algoritmo minimice su valor:

$$\Phi(A) = \frac{1}{q}\left(\sum_{i=1}^{r}\sum_{j=1}^{r}\varepsilon_{ij}x_{ij} + \sum_{i=1}^{r}w_{i0}y_{i0}\right) \qquad \textbf{(10)}$$

   donde q=cantidad de objetos a clasificar, x_{ij}=cantidad de objetos de K_i clasificados en K_j, ε_{ij}=costo de cometer tal error, y_{i0}=cantidad de objetos en K_i en los que hubo abstención del algoritmo, W_{i0}=costo de la abstención.

La expresión (10) presupone un problema de clasificación supervisada con clases disjuntas y se aplica a algoritmos tales que para cada objeto a clasificar, lo ubica en a lo sumo una de las clases. Por tal razón, en este caso sólo existen dos tipos de errores:

a.- ubicar el objeto en otra clase,

b.- no ubicar el objeto en clase alguna.

6. Finalmente, consideremos un funcional de calidad que nos permitirá evaluar la eficiencia de los algoritmos que trabajan sobre clases no disjuntas y/o permiten multiclasificación, es decir, clasificar a un objeto en más de una clase pero estableciendo, como en (10), una diferenciación entre tipos de errores y abstenciones.

   Comoquiera que estos algoritmos responderán en general: "si" (1); "no" (0); "no sé" ($*$) ante la pertenencia de un objeto a una clase dada, podrán cometer los siguientes tipos de errores:

   1) decir "si" cuando realmente es "no";

   2) decir "no" cuando realmente es "si";

   3) abstenerse.

   Estas tres situaciones son valoradas en forma diferenciada por el siguiente funcional:

   Consideremos las funciones auxiliares,

$$X_j(O_i') = \begin{cases} 1 & \text{si } a_j^A(O_i') \\ 0 & \text{en otro caso} \end{cases}$$

   (esta función se asocia a las respuestas positivas de A).

$$Y_j(O'_i) = \begin{cases} 1 & \text{si } a_j^A(O'_i) = 0 \\ 0 & \text{en otro caso} \end{cases}$$

(esta función se asocia a las respuestas negativas de A).

$$Z_j(O'_i) = \begin{cases} 1 & \text{si } a_j^A(O'_i) = * \\ 0 & \text{en otro caso} \end{cases}$$

(esta función se asocia a las abstenciones de A).

Llamaremos *error relativo del algoritmo* A ante la clasificación de O'_i y lo denotaremos $\Phi'(A, O'_i)$ a la magnitud

$$\Phi'(A, O'_i) = \frac{1}{r}\left\{\sum_{j=1}^{r}\left[\gamma_j X_j(O'_i)(1-\alpha_j(O'_i) + \varepsilon_j Y_j(O'_i)\alpha_j(O'_i) + \omega_j Z_j(O'_i)\right]\right. \qquad \textbf{(11)}$$

donde γ_j es el costo de cometer un error de tipo 1) en la clase K_j; ε_j es el costo de cometer un error de tipo 2) en la clase K_j; y ω_j es el costo de abstenerse en la clase K_j; j=1,...,r.

Y finalmente podemos asociar al algoritmo A la magnitud

$$\Phi(A) = \frac{1}{q}\sum_{i=1}^{q}\Phi'(A, O'_i) \qquad \textbf{(12)}$$

que evalúa el error total relativo y que por tanto es deseable que alcance valores pequeños, próximos a cero.

§2.3.-Modelo de algoritmos de votación

Como ya sabemos las características espaciales que fueron referidas anteriormente en la introducción y el epígrfe 2.1, con frecuencia no aparecen en los problemas de clasificación supervisada que surgen en las ciencias poco formalizadas y eso nos hace plantearnos la necesidad de elaborar algoritmos de clasificación que se adecuen de una mejor manera a las características ya mencionadas de estas disciplinas.

Por otra parte, si analizamos el proceder de diferentes especialistas de diversos campos de las ciencias naturales pudiéramos apreciar, sin la pretención de modelar el pensamiento humano pero sí teniendo en cuenta algunas de sus manifestaciones externas, que casi en ningún caso se tienen en cuenta todos los rasgos que describen un objeto o fenómeno dado, sino más bien hay una tendencia al análisis de subfamilias de rasgos (como proyecciones a subespacios de dimensiones más pequeñas que las del espacio inicial; como teniendo sólo un cierto "punto de vista"). Hay una tendencia a, sobre la base de esas subfamilias de rasgos, hacer conclusiones *parciales*, temporales, que permitan posteriormente, con la consecución de otras conclusiones de este tipo con otras subfamilias de rasgos, llegar a una conclusión final. Es decir, se presupone que estas familias de problemas tienen implícita una cierta precedencia parcial que permite el análisis por partes y posteriormente totalizarlos en una decisión final. Esta es la idea básica sobre la que descansa la familia de algoritmos *para el cálculo de las evaluaciones*, como primeramente se denominó este modelo, o *algoritmos de votación* como le llamaremos en lo sucesivo.

El modelo de Algoritmos de Votación tuvo su origen en los años 1965 aproximadamente Dmitriev et al.[18], y sus desarrollos se deben al especialista ruso Yu. I. Zhuravliov y a su grupo. El que aquí empleamos tiene añadidos algunos resultados obtenidos posteriormente por el grupo de Reconocimiento de Patrones del Instituto de Cibernética, Matemática y Física de la Ministerio de Ciencias, Tecnología y Medio Ambiente de Cuba.

La idea esencial del modelo, como ya habíamos mencionado, es la de la precedencia parcial es decir, *analogías parciales*. Se trata de que un objeto puede parecerse a otro pero no en su totalidad y las partes en que sí se parecen pueden dar información acerca de posibles regularidades. Por supuesto, no todas en la misma magnitud.

Supongamos que tenemos planteado un problema de clasificación con aprendizaje. Sea dado un subconjunto de rasgos $\Omega=\{ x_{i_1} ,..., x_{i_s} \}$.

Definición Llamaremos Ω-*parte* de la descripción I(O) de un objeto, y la denotaremos ΩI(O), o también ΩO, a la subdescripción de O en términos sólo de los rasgos $x_i \in \Omega$. Es decir,

$$\Omega I(O) = (x_{i_1} (O) ,..., x_{i_s} (O)).$$

Modelo de algoritmos de votación

Veamos ahora en qué consiste el modelo de algoritmos de votación. Este modelo se describe mediante 6 etapas:

1. SISTEMA DE CONJUNTOS DE APOYO.
2. FUNCIÓN DE SEMEJANZA.
3. EVALUACIÓN POR FILA DADO UN CONJUNTO DE APOYO FIJO.
4. EVALUACIÓN POR CLASE DADO UN CONJUNTO DE APOYO FIJO.

5. EVALUACIÓN POR CLASE PARA TODO EL SISTEMA DE CONJUNTOS DE APOYO.
6. REGLA DE SOLUCIÓN.

Es decir, dar un algoritmo de votación **A**, significa dar un conjunto de parámetros en cada una de las 6 etapas enumeradas, como veremos a continuación.

Sistemas de conjuntos de apoyo

Por conjunto de apoyo se entiende un subconjunto no vacío de rasgos en términos de los cuales se analizarán los objetos. Un sistema de conjuntos de apoyo son varios conjuntos de apoyo.

Ejemplos.-

1.- *Conjunto potencia*. Se trata del conjunto formado por todos los subconjuntos posibles del conjunto de rasgos definidos por el usuario en el problema.

2.- *Cardinal fijo* k. Se trata del conjunto formado por todos los subconjuntos de k elementos del conjunto de rasgos.

3.- *Combinaciones de cardinal fijo*. Es lo mismo que en el caso an terior pero con más de un k.

4.- *Testores típicos*. Son todos los testores típicos de la matriz de aprendizaje.

5.- Existe también una opción para que el usuario seleccione el sistema de conjuntos de apoyo que le sea necesario, es decir, un sistema de conjuntos de apoyo es cualquier conjunto de subconjuntos de rasgos, incluso pudiera ser el propio *R*.

Funciones de semejanza

Como se vio en el §1.1.2, estas funciones definen la forma en que serán comparadas las descripciones (subdescripciones) de los objetos en cuestión. Para ello es imprescindible empezar por determinar la comparación entre los valores de los rasgos en los mismos.

Ejemplos.- (Ver también los del §1.1.2).

1.- $\beta\ (\Omega O_i, \Omega O_j) = \left|\{x_p | x_p(O_i) = x_p(O_j)\}\right|$. El número de rasgos en que los objetos O_i y O_j coinciden (igualdad en los valores de tales rasgos).

2.- $\beta(\Omega O_i, \Omega O_j) = \left| \{x_p | C_p(x_p(O_i), x_p(O_j)) \leq \varepsilon_p\} \right|$, donde C_p es un criterio de comparación de valores de la variable x_p. Nos da el número de rasgos casi coincidentes (el rasgo x_p de O_i difiere del rasgo x_p de O_j cuando mucho en ε_p) entre dos objetos.

3.- $\beta(\Omega O_i, \Omega O_j) = \sum_{\substack{\omega_p=1 \\ x_p(O_i)=x_p(O_j) \\ p=1,\ldots,n}} P(x_p)$. Nos da la suma ponderada de los rasgos en que O_i y O_j coinciden (unas variables son más importantes que otras, y P nos da ese peso o importancia).

4.- $\beta(\Omega O_i, \Omega O_j) = \sum_{\substack{\omega_p=1 \\ p=1,\ldots,n}} \frac{P(x_p)}{1 + \left| x_p(O_i) - x_p(O_j) \right|}$. Esta β se hace más pequeña conforme aumentan las diferencias entre valores de un rasgo en O_i y O_j. Supone rasgos numéricos.

<u>Evaluación por fila dada una Ω-parte fija</u>

Una vez definidos el sistema de conjuntos de apoyo y la función de semejanza se inicia un proceso de "contabilización" de votación, en cuanto a la medida de semejanza entre las diferentes partes de las descripciones de los objetos ya clasificados y los que se desean clasificar. Cada fila O_b de MA se compara con el objeto O por medio de la función Γ_Ω dada a continuación. Esta evaluación es una función, por tanto, de los valores de las semejanzas entre las diferentes Ω-partes comparadas.

Ejemplos.-

1.- $\Gamma_\Omega(O_p,O) = \gamma(O_p).P(\Omega).\beta(\Omega I(O_p), \Omega I(O))$

donde $\gamma(O_p)$ es un parámetro de ponderación asociado a cada objeto de MA, $P(\Omega)$ es el *peso*, la importancia, del conjunto de apoyo Ω y $\beta(\Omega I(O_p),\Omega I(O))$ es el valor de la semejanza entre los objetos comparados.

2.- $\Gamma_\Omega(O_p,O) = \gamma(O_p).\ (P(x_{i_1})+P(x_{i_2}),\ldots,P(x_{i_s}))\beta(\Omega I(O_p),\Omega I(O))$

donde, $\Omega=\{x_{i_1}, x_{i_2}, \ldots, x_{i_s}\}$ y pudiéramos tener

$$\gamma(O_p) = \frac{1}{\sum_{i=1}^{n} P(x_i)} \sum_{i \in \{p | x_p(O) \neq *\}} a_i^v(O)\, P(x_i)$$

siendo $P(x)=\frac{T(x)}{T}$, donde T es la cantidad total de testores típicos de la matriz de aprendizaje y T(x) el número de ellos que contienen el atributo x; $*$ denota la ausencia de información y los coeficientes $a_i^v(O)$ se calculan por una de las expresiones siguientes:

$a_i^v(O)$=cantidad de valores de la variable x_i coincidentes con $x_i(O)$ en K'_v dividido por el cardinal (la cantidad de elementos) de K_v, v=1,…,r>1; siendo K_1, digamos por caso, las descripciones de zonas perspectivas y K_2 las no perspectivas, etc. El número de clases puede ser cualquiera.

$a_i^v(O)$=cantidad de valores de la variable x_i coincidentes con $x_i(O)$ en K'_v dividido por el cardinal de K_v menos el total de $*$ en dicha columna.

Evaluación por clase dado una Ω-parte fija

De lo que se trata es de totalizar las evaluaciones obtenidas para cada uno de los objetos de MA respecto al objeto O que se quiere clasificar. Esta totalización es evidentemente función de las evaluaciones por filas anteriormente obtenidas para la Ω-parte fija. Es decir, le preguntamos a cada objeto O_p de MA su "opinión" sobre si el objeto O a clasificar pertenece a la clase del objeto interrogado. Esta opinión es el valor de Γ_Ω. Nótese que cada objeto O_p de MA sólo opina sobre la pertenencia de O a **su** clase; O_p no puede opinar sobre clases a las que O_p no pertenece.

Ejemplos.-

1.- $\Gamma_\Omega^j(O)=\frac{1}{m_j}\sum_{t=1}^{m_j}\Gamma_\Omega(O,O_t)$

donde t varía en los elementos de la clase K_j y m_j es la cantidad de elementos de la misma.

2.- $\Gamma_\Omega^j(O)=\begin{cases}1 & \text{si } \Gamma_\Omega^j(O)\geq\gamma\\ \\ 0 & \text{si } \Gamma_\Omega^j(O)\leq\gamma\text{, con } \gamma\geq 0\end{cases}$

Evaluación por clase para todo el sistema de conjuntos de apoyo

Hasta el paso anterior todos los cálculos se habían hecho para un conjunto de apoyo fijo, ahora totalizaremos los mismos para todo el sistema seleccionado.

Ejemplos.-

1.- $\Gamma_j(O) = \frac{1}{|\Omega_A|} \sum_{\Omega \in \Omega_A} \Gamma^j_\Omega(O)$

siendo Ω_A el sistema de conjuntos de apoyo seleccionado para el algoritmo A.

2.- $\Gamma_j(O) = \frac{1}{|\Omega_A|} \sum_{\Omega \in \Omega_A} \Delta_\Omega \Gamma^j_\Omega(O)$

que a diferencia del anterior pondera con Δ_Ω los conjuntos de apoyo de manera diferente.

<u>Regla de solución</u>

Se trata ahora de establecer un criterio, una regla, para que sobre la base de cada una de las votaciones obtenidas en la etapa precedente, se pueda tomar una decisión en cuanto a las relaciones del objeto a clasificar con cada una de las clases del problema.

En general se tiene que

$$r_A(\Gamma_1(O),...,\Gamma_r(O)) = (\alpha^A_1(O),...,\alpha^A_r(O))$$

siendo $\alpha^A_i(O)$ el valor, la respuesta del algoritmo de clasificación respecto a la relación entre el objeto O y la clase K_i, i=1,...,r.

Ejemplos.-

1.- $\alpha^A_i(O) = \begin{cases} 1 & \text{si } \Gamma_j(O) > \Gamma_i(O) \text{ para } i \neq j \\ * & \Gamma_p = \Gamma_j(O) > \Gamma_i(O) \text{ para } p \neq j \\ 0 & \text{en otro caso} \end{cases}$

Se asigna el objeto A a la clase K_i que minimiza Γ_i

2.- $\alpha^A_i(O) = \begin{cases} 1 & \text{si } \Gamma_j(O) - \lambda > \Gamma_i(O) \text{ para } i \neq j \\ * & \text{si existen al menos dos j que cumplan lo anterior} \\ 0 & \text{en otro caso} \end{cases}$

Hasta aquí la descripción general de cómo obtener, cómo determinar un algoritmo de votación. Hemos mencionado algunas de las familias de algoritmos de reconocimiento de patrones existentes, en particular, la de los algoritmos de votación. Uno de los problemas que surge ante la abundancia de opciones es el seleccionar el mejor de los algoritmos posibles para la solución del problema en cuestión. Para ello es imprescindible establecer un criterio evaluativo de la calidad de los mismos. Esto es lo que viene a resolver el concepto de *funcional de calidad*, que introdujimos en §2.2.

Con el propósito de tener una idea más clara de la esencia del modelo de algoritmos de votación les proponemos "correr" **un ejemplo numérico**. Sea

$$MA=\left|\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right|$$

e I(O)=(0,1,1) la descripción del objeto a clasificar. Tomemos como sistema de parámetros para determinar un algoritmo del modelo, el siguiente:

Ω_A.- el conjunto de todos los subconjuntos de $\{x_1,x_2,x_3\}$ de cardinal 2. Es decir, $\Omega_A=\{\{x_1,x_2\},\{x_1,x_3\},\{x_2,x_3\}\}$.

β.- como función de semejanza, usaremos un caso particular de unos de los ejemplos dado anteriormente:

$$\beta(\Omega O_i,\Omega O_j)=\begin{cases} 1 & \text{si } x_p(O_i)=x_p(O_j)\ p=1,2; \\ 0 & \text{en otro caso} \end{cases}$$

f.- Supongamos calculados los pesos informacionales de los rasgos y los objetos y que estos son: $P(x_1)=0.5$; $P(x_2)=P(x_3)=1$; y $P(O_1)=P(O_3)=0.5$; $P(O_2)=P(O_4)=1$.

Usaremos como variante de evaluación por fila para un conjunto de apoyo fijo :

$\Gamma_\Omega(O,O_j)=P(O_j)(P_{i_1}+P_{i_2})\beta(\Omega O,\Omega O_j)$.

ρ.- Para totalizar las evaluaciones por clase dado un conjunto de apoyo fijo usaremos:

$$\Gamma_\Omega^j(O)=\frac{1}{2}\sum_{O_{i_t}\in K'_j}\Gamma_\Omega(O,O_{i_t})$$

ψ.- como quinto parámetro, para realizar la suma de las votaciones por clase para todo el sistema de conjuntos de apoyo dado, emplearemos:

$$\Gamma_j(O)= \sum_{\Omega\in\Omega_A} \Gamma_{\Omega}^{j}(O)$$

r_A.- y finalmente, como regla de solución usaremos aquella que nos permite colocar al objeto en la clase para la cual obtuvo su máxima evaluación. Esto es:

$$\alpha_{ji}^{A}(O)= (r_A(\Gamma_1(O), \Gamma_2(O)))_j=\begin{cases} 1 & \text{si } \Gamma_j(O) > \Gamma_i(O), \quad i\neq j, \quad i=1,2 \\ * & \text{si } \Gamma_j(O) = \Gamma_i(O), \quad i\neq j, \quad i=1,2 \\ 0 & \text{si } \Gamma_j(O) < \Gamma_i(O), \quad i\neq j, \quad i=1,2 \end{cases}$$

Ejecutemos el ejemplo sobre la base de los parámetros seleccionados.

1) Sea

$\Omega_1=\{x_1,x_2\}$	$\Omega_2=\{x_1,x_3\}$	$\Omega_3=\{x_2,x_3\}$

Calculemos las respectivas Ω-partes de los objetos de MA y del objeto a clasificar:

$\Omega_1O_1=(1,1)$	$\Omega_2O_1=(1,0)$	$\Omega_3O_1=(1,0)$
$\Omega_1O_2=(1,0)$	$\Omega_2O_2=(1,0)$	$\Omega_3O_2=(0,0)$
$\Omega_1O_3=(0,1)$	$\Omega_2O_3=(0,0)$	$\Omega_3O_3=(1,0)$
$\Omega_1O_4=(0,0)$	$\Omega_2O_4=(0,1)$	$\Omega_3O_4=(0,1)$

$\Omega_1O=(0,1)$	$\Omega_2O=(0,1)$	$\Omega_3O=(1,1)$

2) Calculemos el valor de semejanza entre las respectivas Ω–partes:

$\beta(\Omega_1O_1,\Omega_1O=0$	$\beta(\Omega_2O_1,\Omega_2O)=0$	$\beta(\Omega_3O_1,\Omega_3O)=0$
$\beta(\Omega_1O_2,\Omega_1O)=0$	$\beta(\Omega_2O_2,\Omega_2O)=0$	$\beta(\Omega_3O_2,\Omega_3O)=0$
$\beta(\Omega_1O_3,\Omega_1O)=1$	$\beta(\Omega_2O_3,\Omega_2O)=0$	$\beta(\Omega_3O_3,\Omega_3O)=0$
$\beta(\Omega_1O_4,\Omega_1O)=0$	$\beta(\Omega_2O_4,\Omega_2O)=1$	$\beta(\Omega_3O_4,\Omega_3O)=0$

3) $\Gamma_{\Omega}(O,O_j)=P(O_j)(P_{i_1}+P_{i_2})\beta(\Omega O,\Omega O_j)$. Por tanto:

$$\Gamma_{\Omega_1}(O,O_1)=0.5\times(0.5+1)\times 0=0.$$

Es obvio que sólo tenemos que hacer el cálculo en aquellos valores para los cuales la función de semejanza es 1.

$$\Gamma_{\Omega_1}(O,O_3)=\tfrac{3}{2}; \qquad \Gamma_{\Omega_1}(O,O_4)=\tfrac{3}{2}$$

4) $\Gamma_{\Omega}^{j}(O)=\frac{1}{2}\sum_{O_{i_t}\in K'_j}\Gamma_{\Omega}(O,O_{i_t})$, luego:

$\Gamma_{\Omega_1}^{1}(O)=0;$ $\quad\Gamma_{\Omega_1}^{2}(O)=3/8;$

$\Gamma_{\Omega_2}^{1}(O)=0;$ $\quad\Gamma_{\Omega_2}^{2}(O)=3/4;$

$\Gamma_{\Omega_3}^{1}(O)=0;$ $\quad\Gamma_{\Omega_3}^{3}(O)=0.$

5) $\Gamma_j(O)=\sum_{\Omega\in\Omega_A}\Gamma_{\Omega}^{j}(O)$. Esto es $\Gamma_1(O)=0$; $\Gamma_2(O)=9/8$.

Finalmente:

6) $r_A(\Gamma_1(O),\Gamma_2(O))=(0,1)$ ya que $\Gamma_2(O)>\Gamma_1(O)$. De donde concluimos que según el algoritmo A seleccionado anteriormente, el objeto O = (0,1,1) pertenece a la clase K_2. Podemos pensar que esta decisión es razonable si observamos que O tiene con cada objeto de K'_1 dos coordenadas diferentes, mientras que sólo presenta una coordenada diferente con cada elemento de la otra clase. Obviamente esto es tan sólo un ejemplo, en la práctica las cosas no siempre salen tan bien.