

Actas del V Coloquio de Geografía Cuantitativa
Universidad de Zaragoza
1992, Zaragoza

CLASIFICACION DE OBJETOS MEDIANTE ALGORITMOS GENETICOS

J.A. PEREZ RUY-DIAZ

Facultad de Informática. U.P.M.

Campus de Montegancedo

28660 Boadilla del Monte, Madrid

Tlf: (91) 3367389

E-Mail: ruy@asterix.fi.upm.es

RESUMEN

La creciente exigencia de información elaborada que sirva de manera efectiva para la toma de decisiones, obliga a encontrar nuevas herramientas de análisis.

En este trabajo se describe un método de clasificación basado en técnicas de Inteligencia Artificial. El problema planteado consiste en agrupar n objetos calificados con atributos dicotómicos. La solución propuesta se basa en un Algoritmo Genético.

1. INTRODUCCION

La búsqueda de un orden de las cosas es uno de los problemas más usuales de la ciencia. El reconocimiento de estructuras dentro un universo complejo y la clasificación de sus elementos constituye una de las herramientas básicas de análisis para la comprensión del mismo.

La Geografía no es una excepción. Frecuentemente se plantea la necesidad de cuantificar las diferencias existentes entre distintos objetos, en general zonas, calificados por un cierto número de atributos y, consiguientemente, agrupar los de similares características. Como

fácilmente se puede entender, cuanto mayor es el número de atributos (variables), más compleja resulta la clasificación. Por esta razón, es conveniente tratar de reducir en primer lugar el número de atributos, buscando los más relevantes, para, posteriormente, proceder a la clasificación. Los pasos que se siguen generalmente son:

- 1) Obtención de una matriz de correlación entre los valores de los atributos presentes en la muestra con el fin de obtener una medida de la dependencia entre ellos
- 2) Extracción de los factores principales mediante análisis de los componentes principales o análisis factorial, con el fin de explorar la posible reducción de los datos
- 3) Rotación, para facilitar la interpretación de los factores obtenidos
- 4) Agrupamiento de las observaciones (cluster analysis) por su distancia, definida como función de las variables significativas obtenidas en el paso (2)

Se puede encontrar un cierto número de ejemplos de aplicación de esta metodología en la bibliografía relativa a los Sistemas de Información Geográfica [BURR86] y en publicaciones sobre problemas específicos de clasificación de suelos [WEBS90], análisis de datos sociológicos de ciudades [GAZE84] y planificación territorial [TOLI84].

Dentro de este grupo de problemas existe uno de especiales características que lo hacen particularmente difícil de resolver con los métodos clásicos: es el caso en el que los atributos son variables dicotómicas. Se trata de una clasificación de objetos caracterizados por la presencia o ausencia de determinados atributos. La dificultad es doble, pues afecta tanto al análisis de la dependencia de los datos (son especialmente sensibles al error muestral y no resulta fácil obviar las hipótesis subyacentes en los métodos clásicos sobre la distribución de las variables) como a la propia agrupación (no es evidente la definición de una distancia) [KEND80].

Una forma posible de resolver tales problemas consiste en plantearlos en términos de optimización y aplicar herramientas algorítmicas de efectividad demostrada en este campo. Esta alternativa, ha sido estudiada

por diversos autores [FISH91] [FALK91] y su dificultad estriba en la definición de la función de coste y la elección del método apropiado para optimizar dicha función.

En lo que sigue se describe una solución basada en este planteamiento. Se define una función objetiva y se introducen los Algoritmos Genéticos como técnica de optimización.

2. PLANTEAMIENTO

Supóngase que se quiere agrupar objetos en familias, de forma que objetos similares pertenezcan a la misma familia y objetos diferentes pertenezcan a familias distintas. Los objetos pueden presentar ciertas características o carecer de ellas. Sirva como ejemplo la tabla adjunta con datos referentes a dos tipos de distritos de una ciudad identificables como zonas residenciales y de negocio:

Objetos								Atributos
O ₁	O ₂	O ₃	O ₄	O ₅	O ₆	O ₇	O ₈	
x		x		x	x			A ₁ Hospitales
	x		x	x	x	x	x	A ₂ G.Almacenes
		x					x	A ₃ S.Espectáculo
	x			x	x	x	x	A ₄ Hoteles
x	x	x	x				x	A ₅ Colegios
x	x	x	x					A ₆ I.Deportivas
x			x	x	x	x	x	A ₇ Oficinas
x	x	x	x	x			x	A ₈ Parques

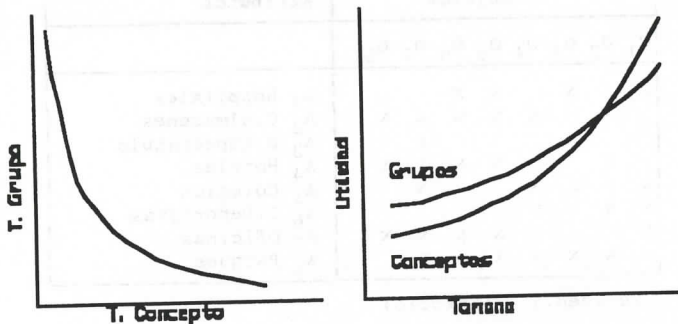
(Residen.) (Negocio)

Una primera aproximación consiste en explorar las *similitudes* entre los objetos y agrupar aquéllos que sean más parecidos (p.e.: O₁ y O₄). Se trata, evidentemente, de similitudes *positivas*, es decir, referidas a la existencia de ciertos atributos y no a la negación de los mismos. Ahora bien, salvo el caso trivial en el que todos los objetos sean iguales, habrá atributos que aparecerán en unos objetos y no en otros, y existirán distintas formas de agruparlos según el subconjunto de atributos (*concepto*) que se elija.

Si el concepto tiene pocos atributos, será muy genérico y probablemente englobará muchos objetos. Por consiguiente, con unos pocos conceptos se formarán otros tantos grupos y se tendrá una clasificación que resultará aceptable siempre que los atributos sean *relevantes*.

Ahora bien, ¿qué garantiza que un atributo es relevante?. Si se examina el problema desde el punto de vista de los atributos, cualquiera de ellos puede ser importante "a priori" y parecerá mejor aquella solución que tenga en cuenta el mayor número posible de ellos. En consecuencia, conceptos detallados, definidos con un gran número de atributos, proporcionarán una clasificación más valiosa que otra genérica.

El número de elementos de un grupo (tamaño del grupo) y el número de atributos que los definen (tamaño del concepto) dan lugar a criterios enfrentados. Si disminuye el número de atributos empleados en la clasificación, se logrará una clasificación clara, de pocos grupos, aunque de poca utilidad. En el otro extremo, si los conceptos se definen con más atributos, la clasificación será más precisa, pero se resentirá por el número excesivo de grupos. En el límite podrán existir tantos grupos como objetos.



La solución está en un punto de equilibrio, pero, para hallarlo, es preciso cuantificar la bondad de una solución. Un criterio aceptable consiste en valorar los conceptos en función del número total de atributos comunes y no comunes al grupo [FALK92]:

- El valor de un concepto es la suma de todos los atributos posibles del grupo a la que se resta el número de atributos del

concepto y el número de atributos de los objetos del grupo no incluidos en el mismo.

- El valor de la solución es la suma de los valores de todos los conceptos que la componen.

Si se consideran en el ejemplo los conceptos $C1=(A_5, A_6, A_8)$, $C2=(A_1, A_2, A_4, A_7)$ y $C3=(A_2, A_4, A_7)$ y sus grupos respectivos $G1=(O_1, O_2, O_3, O_4)$, $G2=(O_5, O_6)$ y $G3=(O_7, O_8)$, de acuerdo con este criterio, sus valoraciones respectivas y la de la solución serían:

$$\text{Valor } C1 = (4 \times 8 - 3 - 8) = 21$$

$$\text{Valor } C2 = (2 \times 8 - 4 - 1) = 11$$

$$\text{Valor } C3 = (2 \times 8 - 3 - 3) = 10$$

$$\text{Valor Total} = 42$$

El valor es superior al que corresponde a los ocho objetos aislados (38) e inferior a la solución indicada (44) en la tabla.

El criterio de valoración cumple con lo dicho anteriormente, pues si aumenta el tamaño de los grupos, aumenta el peso de cada uno pero disminuye el número de sumandos y viceversa. Por lo tanto, refleja bien el planteamiento del problema.

Se llega, de esta manera, a un planteamiento de optimización, que se puede enunciar así:

Dado un conjunto de objetos O_i para los cuales pueden existir una serie de atributos A_j , definir una serie de conceptos (conjunto de atributos) que clasifique el espacio de los objeto y maximice la función de valoración definida

El problema ahora es la elección del método más apropiado para llevar a cabo la búsqueda del óptimo, pues, aunque la función de objetivo se puede calcular en cada punto (solución), *no resulta fácil expresarla analíticamente*. Es en estos caso donde resultan especialmente útiles los Algoritmos Genéticos.

3. ALGORITMOS GENETICOS

Los Algoritmos Genéticos son técnicas de optimización basadas en los sistemas LCS (Learning Classifier Systems) de Inteligencia Artificial. El funcionamiento de tales sistemas consiste en partir de un cierto estado generado aleatoriamente e ir generando series de soluciones (clasificadores) cuyos efectos se contrastan con otras de resultados (mensajes). La efectividad de los mensajes se evalúa y con este dato se producen nuevas soluciones que se adaptan a cada circunstancia.

La originalidad de los Algoritmos Genéticos, estriba en que emulan el comportamiento de colonias de seres vivos siguiendo la regla de Darwin de la supervivencia del más fuerte. Para ello:

1. Cada solución es un ser caracterizado por un cromosoma, cuyos genes son valores de las variables.
2. Dichos valores están codificados, es decir, cada gen representa el valor de una variable mediante caracteres de un alfabeto sencillo. Existe, pues, una distinción entre *genotipo* (valor de los genes) y *fenotipo* (efecto sobre las variables del problema).
3. Existe una función positiva que permite valorar a cada individuo.
4. Partiendo de una población de soluciones generada al azar, se la hace evolucionar mediante operadores de transición derivados de las existentes en la naturaleza.
5. El proceso continúa hasta la obtención de una población prácticamente clónica cuyo representante es la solución.

Los operadores más usuales son:

1. Selección
2. Cruce
3. Mutación

La *Selección* es un proceso mediante el cual se elige para la reproducción un conjunto de soluciones de acuerdo con el valor de la función objetivo (El fuerte se impone al débil). Las mejores soluciones tendrán más probabilidades de contribuir con uno o más representantes (hijos) en la siguiente generación.

El *Cruce* es un operador que se aplica en dos pasos. En primer lugar, los elementos procedentes de la selección se emparejan aleatoriamente. Posteriormente, cada pareja procede a cruzarse intercambiando información genética y dando lugar a dos hijos.

Con ello se obtienen nuevas soluciones que reemplazan a los más débiles (Los fuertes sobreviven).

En cuanto a la *Mutación*, es un operador que actúa sobre los genes de los cromosomas conforme se van transfiriendo a la descendencia en el proceso de cruce. Su acción consiste en introducir cambios al azar en los mismos. La mutación es una protección frente a la pérdida irrecuperable de material genético de una forma prematura, pues los procesos de selección y cruce pueden destruir información valiosa.

Con todo, la mutación juega un papel secundario en los algoritmos genéticos al igual que sucede en la naturaleza. La práctica demuestra que los mejores resultados se obtienen con un número pequeño de mutaciones, pues, de lo contrario, el algoritmo degenera en un paseo aleatorio.

La mecánica de un algoritmo genético es sorprendentemente simple y eficaz. Pero, aunque pueda extrañar a primera vista, se puede entender la lógica subyacente:

- La Selección *acota* el espacio de búsqueda. Si no existiera más que este operador y el algoritmo volviera a seleccionar al azar dentro del nuevo intervalo de soluciones, ello bastaría en funciones unimodales para acercarse a la solución.

- El Cruce al recombinar genes las mejores soluciones, puede significar en ciertos casos una *estrategia de búsqueda* mejor que la aleatoria (p.e: si la función objetivo es de variables separadas).
- En cuanto a la Mutación, puede interpretarse como un *rearranque* limitado del proceso.

Con todo, el buen funcionamiento del Cruce es fundamental para la convergencia del algoritmo y gran parte de la justificación formal de la misma gira alrededor de dicho operador. Si genes individuales o en pequeños grupos influyen decisivamente en el valor de la solución, el cruce no destruirá significativamente el mensaje que los padres transmiten a los hijos y, por consiguiente, se avanzará hacia la solución. Si, por el contrario, es la combinación de un gran número de genes contiguos lo que es relevante, no se puede garantizar la convergencia (Hipótesis de los bloques de Holland)[HOLL75]. En consecuencia, la buena codificación de las soluciones es decisiva para un buen funcionamiento del algoritmo.

4. CLASIFICACION MEDIANTE UN ALGORITMO GENETICO

Como se indicó en el apartado anterior, un Algoritmo Genético se caracteriza por una *codificación* de los datos, una *valoración* de los individuos, un conjunto de *operadores genéticos* y una serie de *parámetros* (tamaño de población, probabilidad de mutación, etc). Una vez definida la función de valoración en el apartado segundo, en éste se revisan las variantes introducidas en los conceptos generales para la aplicación de clasificación propuesta.

4.1. Codificación

En el problema de clasificación se entremezclan dos ideas, grupos/objetos y conceptos/atributos que, aunque están íntimamente relacionadas, pueden prestar a confusión cuando hay que identificar las variables del sistema y proceder a su codificación.

Supóngase en primer lugar que las variables fueran los objetos y sus valores el grupo al que pertenecen. Si así fuera, la codificación de una solución podría hacerse mediante una tira de valores que indicaran a que grupo pertenece cada objeto. Suponiendo ordenados los objetos, la cadena:

AAAABBCC

representaría para el problema de la Tabla 1 que los objetos O_1 , O_2 , O_3 y O_4 pertenecen a un primer grupo, los objetos O_5 y O_6 al segundo y, finalmente, O_7 y O_8 a un tercero. Los valores A, B y C podrían ser los índices de una tabla de conceptos presentes en la población:

A-->(A₅,A₆,A₈)

.....
B-->(A₁,A₂,A₄,A₇)

.....
C-->=(A₂,A₄,A₇)

.....

Supóngase que esta solución se cruza con otra. El intercambio de material genético puede realizarse de muchas formas, pero, en todo caso, las soluciones resultantes del cruce tendrán genes entremezclados y probablemente la secuencia de genes contiguos AAAA quedará rota. Como en el óptimo habrá siempre conjuntos de genes repetidos, el cruce trabaja en contra de él y no se puede asegurar que el algoritmo se estabilice en sus fases intermedias. En términos de la hipótesis de los bloques de Holland, se puede dar una explicación: la hipótesis no se cumple *porque los conjuntos grandes de genes son más relevantes que los pequeños o los genes individuales*. Se tiene que elegir otra codificación.

Supóngase ahora que las variables son los conceptos y sus valores los índices de la tabla anterior. La solución anterior se codifica ahora (ABC), y distingue entre *genotipo* (código genético) y *fenotipo* (resultado). Los criterios de clasificación A, B y C aplicados ordenadamente dan lugar a la clasificación AAAABBCC. El cromosoma es de longitud variable, pues, en el límite puede tener tantos genes como objetos, sin embargo esto no representa una dificultad especial.

El comportamiento de esta codificación frente al cruce, es muy superior a la anterior. Un grupo, con independencia de su tamaño, no se podrá partir porque está codificado por un gen y se transmitirá al menos a uno de los descendientes. Por otra parte, cabe esperar que el óptimo conste de un número pequeño de genes. Por lo tanto se cumplirá la hipótesis de Holland de que *son relevantes los genes individuales o grupos pequeños*.

4.2. Operadores genéticos

Además de los operadores usuales (Selección, Cruce y Mutación), se proponen otros dos (Expansión y Rotación). La Selección se ajusta al esquema general ya descrito, por lo que sólo se comentan las variantes introducidas en el Cruce y la Mutación y los nuevos operadores.

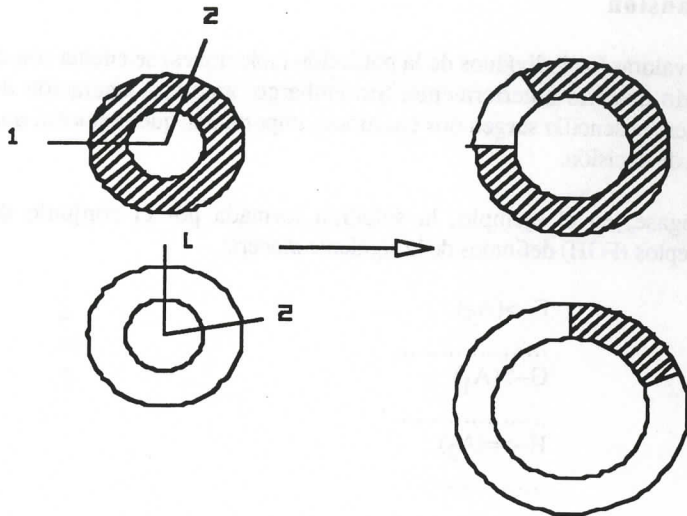
Cruce

Como ya se definió, el cruce consiste en un intercambio de genes. Con el operador más usual, el cruce se lleva a cabo reemplazando genes de un progenitor por los del otro y viceversa. En este caso, esto no es posible por dos razones:

- Los cromosomas son de longitud variable
- La clasificación de los objetos tiene que ser completa

Una solución para resolver el dilema es la citada por Goldberg [GOLD89] y Falkenauer [FALK92], consistente en *inyectar* trozos de longitud aleatoria (no superior a la longitud del cromosoma) de un progenitor en el otro. Los pasos que se siguen son los siguientes:

- Se eligen dos posiciones de corte al azar en ambos progenitores
- Se copia el cromosoma del primero de ellos en el primer hijo intercalando los genes comprendidos entre las posiciones de corte del otro progenitor
- Se repite la misma operación para obtener el segundo hijo intercambiando los papeles



Con el fin de facilitar la operación, se supone que los cromosomas son circulares. De esta forma, la posición del segundo corte puede ser menor que la del primero.

Una vez realizado el cruce, los objetos se clasifican ordenadamente según los conceptos presentes y se eliminan aquellos conceptos que den como resultado grupos vacíos. Así, un concepto duplicado o que sea unión de conceptos aplicados anteriormente, se rechaza.

Mutación

La mutación se ajusta a la forma clásica de los Algoritmos Genéticos, consistente en cambiar aleatoriamente el valor de los genes durante el cruce. Sin embargo, con el fin de garantizar que la combinación de conceptos resultantes en los hijos siempre clasifique a todos los objetos, se aplica únicamente a los genes que se transfieren de un progenitor a otro.

El operador añade o borra atributos del concepto definido en el gen y provoca la aparición de nuevos conceptos en la población.

Expansión

Para valorar los individuos de la población (soluciones) se cuenta con el criterio descrito anteriormente. Sin embargo, ante esta operación de apariencia sencilla surgen dos cuestiones importantes que requieren una toma de decisión.

Supóngase, en el ejemplo, la solución formada por el conjunto de conceptos (FGH) definidos de la siguiente manera:

F-->(A₅)

.....

G-->(A₁)

.....

H-->(A₂)

.....

La agrupación resultante es la misma que la obtenida con la solución (ABC) citada con anterioridad: G₁=(O₁,O₂,O₃,O₄), G₂=(O₅,O₆) y G₃=(O₇,O₈). Quiere esto decir que *para dos genotipos distintos se obtiene el mismo fenotipo*.

Aunque este hecho se da en la naturaleza (existen caracteres recesivos), en lo que al algoritmo respecta trae una consecuencia inmediata: dos clasificaciones iguales tienen valoraciones distintas (31 y 42). La situación es incómoda porque obliga a hacer un replanteamiento sobre qué se valora, si los conceptos o los grupos. Para elegir una alternativa, nada mejor que el método de la *navaja de Occam* debidamente extrapolado: "De dos soluciones igualmente sencillas tomar la que mejor explica el fenómeno". Es preferible, por consiguiente, decidirse por valorar la agrupación.

Si es así, para mantener la definición de la función de valoración, es preciso introducir un nuevo operador que actúa inmediatamente después del cruce: la *Expansión*.

La Expansión explora los grupos resultantes de la aplicación de un conjunto de conceptos con el fin de descubrir otras similitudes existentes no recogidas en los genes. Si se encuentran, se incorporan al concepto.

En cierto modo se trata de un mecanismo de simulación de la dominancia y recesividad: cada grupo contiene atributos almacenados en los objetos que son recesivos frente a los definidos en los conceptos y en ciertas circunstancias aparecen.

No se conocen precedentes de este operador, pero su utilidad en este caso va más allá de la resolución de la ambigüedad en la valoración. Las pruebas realizadas demuestran que es una herramienta muy eficaz para acelerar la convergencia que además permite reducir el tamaño de la población como se verá más adelante.

Rotación

Aparentemente, el proceso de búsqueda del óptimo consiste en combinar conceptos entre todos los posibles, rechazando las soluciones peores y reteniendo las mejores. Los operadores anteriores cumplen perfectamente con este cometido: el Cruce genera combinaciones, la Mutación y la Expansión generan conceptos y la Selección se encarga de elegir los mejores conceptos y combinaciones.

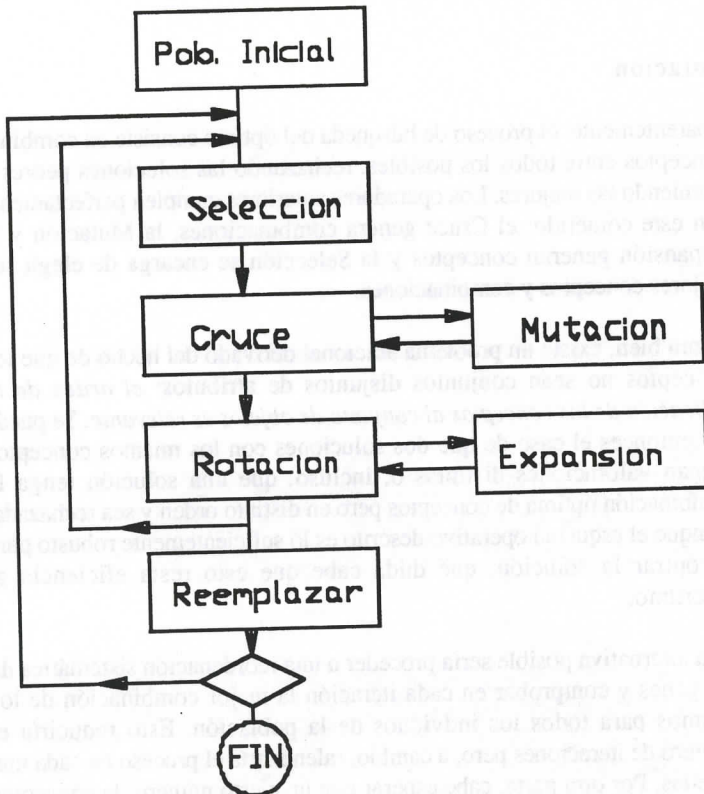
Ahora bien, existe un problema adicional derivado del hecho de que los conceptos no sean conjuntos disjuntos de atributos: *el orden de la aplicación de los conceptos al conjunto de objetos es relevante*. Se puede dar entonces el caso de que dos soluciones con los mismos conceptos tengan valoraciones distintas o, incluso, que una solución tenga la combinación óptima de conceptos pero en distinto orden y sea rechazada. Aunque el esquema operativo descrito es lo suficientemente robusto para encontrar la solución, qué duda cabe que esto resta eficiencia al algoritmo.

Una alternativa posible sería proceder a una reordenación sistemática de los genes y comprobar en cada iteración la mejor combinación de los mismos para todos los individuos de la población. Esto reduciría el número de iteraciones pero, a cambio, ralentizaría el proceso en cada una de ellas. Por otra parte, cabe esperar que un cierto número de conceptos sean efectivamente disjuntos y que el esfuerzo sea en vano.

La solución de compromiso que se propone, coherente con la definición de un cromosoma circular, consiste en limitar la exploración a una

permutación circular mediante un operador de *Rotación* que busca una posición de equilibrio en el cromosoma cada vez que aparece un nuevo individuo en la población. Su implementación no supone un gran gasto de tiempo y se puede mejorar eliminando de la permutación a aquellos conceptos que sean disjuntos, operación relativamente sencilla.

De esta forma, el esquema operativo resultante para el algoritmo es el indicado en la figura:



4.3. Parámetros del algoritmo

Para completar lo expuesto hasta ahora, conviene hacer algún comentario sobre algunos parámetros del algoritmo citados de manera indirecta al hablar del algoritmo y los operadores.

Al hablar del operador de Mutación se decía que era un operador que de forma aleatoria realizaba cambios en los genes. Falta por definir un parámetro de *probabilidad* asociado al mismo o, dicho de otro modo, con qué frecuencia se realizan las mutaciones.

Como ya se dijo, el operador juega un papel secundario. No obstante, guarda una relación muy estrecha con el tamaño de la población: cuanto mayor sea la éste, habrá un mayor número de conceptos presentes para combinar en cada iteración y se alejará el peligro de una convergencia prematura. Por el contrario, cuanto menor sea, habrá que proporcionar un mecanismo más potente de generación de nuevos conceptos, aún a riesgo de impedir la convergencia. Esta última posibilidad obligaría a tomar nuevas medidas cuya descripción se sale del propósito de este trabajo.

La decisión adoptada, que parece más de acuerdo con estudios realizados por otros investigadores [JONG75], consiste en dimensionar convenientemente la población y asignar una probabilidad de mutación baja, entendida como aquella que genera en cada iteración de un 5 a un 10% de nuevos individuos.

En cuanto al *tamaño de población*, el criterio es garantizar la presencia del mayor número de conceptos en la primera generación. Lo ideal sería que todos los conceptos entraran en juego (!más de una vez!) y que el algoritmo explorara sus combinaciones. Esto no es sólo imposible (existen $2^{N_{atr}}$ conceptos posibles en un problema), sino contrario a la concepción de los Algoritmos Genéticos, que se basan en la posibilidad de extraer patrones globales a partir de un conocimiento local. Por otra parte, no todas las combinaciones de atributos son válidas, por lo que parece más razonable considerar conceptos que sean combinaciones de sólo n atributos, siendo n el tamaño medio de los conceptos esperados, y confiar en la eficacia de los operadores.

Es en este punto donde se descubre la eficacia del operador de Expansión, pues, partiendo de conceptos simples, incorpora nuevos atributos si la agrupación lo permite. De esta forma, incluso si los conceptos iniciales son unitarios (de un solo atributo), rápidamente se produce en la

población una adaptación al tamaño medio de los conceptos existentes. Según esto, se puede tomar $n=1$, es decir, considerar como tamaño de población el número de los atributos del problema. En la práctica, se han conseguido muy buenos resultados con tamaños del orden de kn individuos, siendo k un número comprendido entre 1 y 4.

5. CONCLUSIONES

Hasta la fecha el algoritmo descrito se ha probado con éxito en problemas con un máximo de 100 objetos y 50 atributos, trabajando en un ordenador PC 486 a 25 MHz, con una media de 210 segundos de tiempo de ejecución. Los ejemplos probados han sido generados a partir de una solución conocida e introduciendo atributos aleatorios en los objetos. El experimento se ha considerado un éxito cuando la valoración obtenida para la solución ha sido mayor o igual que la esperada. La investigación no se dará por terminada hasta haber pasado una batería significativa de tests.

En todo caso, los resultados obtenidos demuestran las posibilidades de los Algoritmos Genéticos como herramienta de optimización siempre que se tengan en cuenta los principios fundamentales que garantizan su convergencia y se adapten los operadores a los problemas específicos.

6. AGRADECIMIENTOS

Deseo manifestar mi reconocimiento al Profesor Emanuel Falkenauer del Research Centre for Belgian Metalworking Industry (CRIF) por la información proporcionada sobre sus investigaciones en este campo.

7. BIBLIOGRAFIA

- [BURR86] Burroughs, P. (1986). *"Principles of Geographical Information Systems for Land Resources Assessment"*. Oxford Univerity Press.
- [FALK91] Falkenauer, E. (1991). *"A Genetic Algorithm for Grouping"*. Procc. of the 5th Int. Symposium on Applied Stochastic Models and Data Analysis.
- [FALK92] Falkenauer, E. y Delchambre, A. (1992). *"A Genetic Aloritm for Bin Packing and Line Balancing"*. Procc on de IEEE 1992 Int. Conference On Robotic and Automization (en prensa).
- [FISH91] Fisher, D., Panzani, M. y Langley P.(1991). *"Concept Formation: Knowledge and Experience in Unsupervised Learning"*. Morgan Kaufmann Publishers.
- [GAZE84] Gazero, M et al. (1984). *"The Urban Social Structure of Padua: An Approach by Factor Analysis and Automatic Cartography"*. Procc. of the 10th Urban Data Management Symposium.
- [GOLD85] Goldberg, D.E. (1985). *"Optimal Initial Population Size for Binary Coded Genetic Algorithms"*. TCGA Report 85001.
- [GREF84] Grefenstette, J.J. (1984). *"GENESIS: A system for using genetic search procedures"*. Proc. of the 1984 Conference on Intelligent Systems and Machines 161-165.

-
- [HOLL75] Holland, J.H. (1975) "*Adaptation in Natural and Artificial Systems*" University of Michigan Press, Ann Arbor.
- [JONG75] De Jong, K.A. (1975) "*An Analysis of the Behavior of a Class of Genetic Adaptive Systems*", Ph. Thesis.
- [KEND80] Kendall, M. "*Multivariate Analysis*". Charles Griffin & Co. Ltd.
- [TOLI84] Tolini, C. et al. (1984) "*A Decision Support System for Planning and Control in the Public Sector*". Procc. of the 10th Urban Data Management Symposium.
- [WEBS90] Webster, R. y Oliver, M.A. (1990). "*Statistical Methods in Soil and Land Resource Survey*". Oxford University Press.