



DEPARTAMENTO DE INFORMÁTICA
UNIVERSIDAD CARLOS III DE MADRID

Open Course ware

Ingeniería Informática

Computación Biológica

Prueba de conocimiento

- El tiempo estimado para realizar la prueba es de **2 horas y media**

Problema 1. (4 Puntos)

Una empresa ha diseñado una línea de comunicación privada binaria, para transmitir mensajes etiquetados como de gran importancia.

Esta empresa maneja una serie de códigos de productos, en los que se utilizan tan solo 5 símbolos: $(\alpha, \beta, \gamma, \delta, \epsilon)$. Esto hace que los mensajes tengan que transmitir secuencias de dichos símbolos. Se desea hacer una codificación en binario de dicho juego de símbolos, de forma que, aún habiendo errores de transmisión, se puedan corregir dichos errores, y obtener siempre en el destino, el mismo mensaje que se envió, o lo más parecido posible. Por características de la línea de transmisión, cada símbolo se codificará con 13 bits.

Describir cómo podría resolverse el problema, utilizando alguna de las técnicas estudiadas en clase.

Hay que indicar qué técnica emplear y cómo se utilizaría. Es imprescindible, una vez respondidas las preguntas anteriores, describir detalladamente cómo sería la solución. Esto incluye descripciones de las codificaciones (con un ejemplo), función de evaluación (con un ejemplo), cómo se ejecutaría el sistema, etc.

Problema 2. (2 Puntos)

Responder **brevemente** a las siguientes preguntas:

- En el sistema TIERRA, aparecen en las primeras fases de funcionamiento lo que el autor denomina “parásitos”. ¿Qué son estos parásitos y por qué aparecen?
- En un sistema de optimización mediante colonias de hormigas, ¿cuándo decide el sistema que ha obtenido una solución?, y ¿cómo se decide qué hormiga representa la solución del sistema?
- ¿Qué condiciones debe cumplir un problema para que la utilización de la técnica de Programación Genética sea más ventajosa y por qué?
- Las estrategias evolutivamente estables son técnicas diseñadas para resolver qué tipo de problemas de optimización. Poner tres ejemplos de problemas para los que la utilización de ese tipo de métodos es imprescindible

Problema 3. (4 puntos)

El gobierno ha decidido que es necesario la construcción de cinco centrales nucleares para combatir el déficit energético que padece el país. Por motivos de seguridad y medioambiente, se ha determinado que dichas centrales deben situarse lo más alejadas posibles unas de otras. No todos los lugares del territorio pueden alojar una central nuclear. Los expertos han estudiado el asunto, y le han proporcionado al gobierno una lista de 8.192 posibles posiciones para las centrales.

Se desea diseñar una solución al problema, basada en métodos evolutivos. Justificar el método elegido, así como los parámetros necesarios.

Determine con la mayor claridad posible las siguientes cuestiones:

- Codificación del problema, incluyendo un ejemplo ilustrativo de un individuo
- Función de evaluación diseñada, incluyendo la evaluación del individuo
- Tipos de operadores recomendados

Incluir las consideraciones que se estime oportuno, destinadas a la mejor explicación de la solución propuesta.

Solución al problema 1.

La solución a este problema consiste en asignar una secuencia de 13 bits a cada uno de los símbolos del alfabeto. Para ello se podría utilizar una codificación de $13 \times 5 = 65$ bits. Se pueden utilizar las siguientes representaciones:

1. Cada individuo es un solo cromosoma de 65 bits
2. Cada individuo está constituido por 5 cromosomas de 13 bits cada uno
3. Cada individuo está formado por 5 valores numéricos en el rango de 1 a 8.192

En el primer caso se podría utilizar un algoritmo genético, con cruce monopunto y mutación, siguiendo el esquema de un Algoritmo Genético Canónico. En el segundo caso se puede utilizar, también, un Algoritmo Genético, pero en este caso cada cromosoma representa a un símbolo, por lo que el cruce se realizaría en cada cromosoma de forma independiente. Este último esquema ofrece algunas ventajas que le hace más eficiente.

En el último caso, se podría utilizar la técnica de las Estrategias Evolutivas, con la restricción de que los genes tomen valores enteros, y en el rango $[1, 8.192]$. Después cada valor numérico sería transformado en su correspondiente binario, para obtener la codificación correspondiente a cada símbolo.

Codificación En la figura siguiente aparece un ejemplo de codificación. Se aprecia como cada símbolo se corresponde con una codificación que será lo que será transmitido por la línea de comunicación. El individuo sería la concatenación de los vectores binarios de la tabla, para el caso de un Algoritmo Genético, y su correspondiente número decimal, para el caso de una Estrategia Evolutiva:

Símbolo	Codificación
α	0110110101101
β	1010001010001
γ	1111101011101
δ	0010100001010
ϵ	0010100011111

Así pues, al recibir un mensaje en el destino, solo habrá que dividirlo en bloques de 13 bits, y comprobar a qué símbolo se corresponde, según la tabla adjunta y representada por el individuo ganador. Si no existe correspondencia, será que ha habido un error de transmisión, y se asignará el símbolo cuya codificación más se parezca a la recibida. Para que el sistema funcione de forma óptima, los códigos asignados a los 5 símbolos deben estar los más alejados unos de otros en términos de distancia hamming. Esto podrá ser reflejado en la función de evaluación.

Función de evaluación Un buen conjunto de códigos será aquel cuya distancia entre ellos sea máxima. Si llamamos $d_h(c_1, c_2)$ a la distancia hamming entre c_1 y c_2 , se podrá calcular un valor numérico de distancia total, de la siguiente forma:

$$D = \sum_{i=1}^5 \min_{j=1, \dots, 5, j \neq i} d_h(c_i, c_j)$$

Es decir, para cada uno de los cinco códigos de un individuo, se calcula su distancia hamming a los otros cuatro. Se obtiene la mínima y se suman todas esas distancias mínimas. Cuanto

mayor sea este número, más separados estarán los códigos entre sí. Otras variantes podrían ser utilizadas. Por ejemplo la media de todos los mínimos, en vez de la suma. O el mínimo de todos los mínimos. Esta última condición sería más restrictiva, pero más eficaz, aunque podría llevar a estancamiento por no haber mucha presión selectiva en los primeros ciclos del algoritmo. También podría empezarse por una de las funciones anteriores, y cambiarla al ir obteniéndose mejores resultados.

Para el individuo de la tabla anterior tendríamos la siguiente tabla de distancias hamming:

d_h	0110110101101	1010001010001	1111101011101	0010100001010	0010100011111
0110110101101	-	9	6	6	5
1010001010001	9	-	5	7	6
1111101011101	6	5	-	8	5
0010100001010	6	7	8	-	3
0010100011111	5	6	5	3	-

Si cogemos el mínimo por filas (por columnas sería equivalente), tendremos:

$$D = 5 + 5 + 5 + 3 + 3 = 21$$

Que será el valor de evaluación del individuo en cuestión, si utilizamos el sumatorio. Si utilizamos la media sería $d = \frac{21}{4} = 5,25$, y si se utiliza el mínimo sería 3. En cualquier caso, habría que maximizar dichos valores.

Solución al problema 2.

- *En el sistema TIERRA, aparecen en las primeras fases de funcionamiento lo que el autor denomina “parásitos”. ¿Qué son estos parásitos y por qué aparecen?*

Se denominan parásitos a aquellos programas que, para reproducirse, utilizan el código contenido en otros programas denominados huéspedes. De esta forma consiguen tener un menor tamaño, y una ventaja selectiva al poderse reproducir más rápidamente. Aparecen al producirse una mutación en uno de los replicantes, que hace que busquen el código necesario para replicarse, en otros programas.

- *En un sistema de optimización mediante colonias de hormigas, ¿cuándo decide el sistema que ha obtenido una solución?, y ¿cómo se decide qué hormiga representa la solución del sistema?*

Los sistemas de optimización de colonias de hormigas convergen a una única solución. Después de un número suficiente de ciclos, solo un camino tiene feromona, y el resto de posibilidades no. Ese será el camino que seguirán todas las hormigas. Así pues, la convergencia, y por ende la obtención de la solución, se produce cuando todas las hormigas siguen el mismo camino. Como todas ellas siguen el mismo camino, cualquiera de ellas valdrá como representante de la solución.

- *¿Qué condiciones debe cumplir un problema para que la utilización de la técnica de Programación Genética sea más ventajosa y por qué?*

Todos aquellos problemas en los que la solución exija una semántica compleja, son candidatos a ser solucionados mediante Programación Genética. Si la solución se puede representar de forma numérica, ya sea decimal, binaria, o cualquier otra, existen técnicas alternativas con eficacia similar. Sin embargo si la solución es una fórmula, un programa, o cualquier otra estructura más compleja, sólo podrá utilizarse Programación Genética.

- *Las estrategias evolutivamente estables son técnicas diseñadas para resolver qué tipo de problemas de optimización. Poner tres ejemplos de problemas para los que la utilización de ese tipo de métodos es imprescindible*

Las estrategias evolutivamente estables es un concepto de biología, y no un método de computación. Por lo tanto no resuelven ningún tipo de problemas.

Solución al problema 3.

Cualquier solución válida para el problema 1 de este mismo examen, es también válida para este problema. Ambos problemas son en realidad idénticos. Lo único que habría que cambiar es, en la función de evaluación, en vez de utilizar la distancia hamming, utilizar la distancia euclídea entre las centrales.