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 **3 horas y media**

Problema 1. (3 Puntos)

Se desea diseñar un programa que sea capaz de aprender a jugar al juego de conecta-3. El juego consiste en un tablero de $3 \times 3 = 9$ posiciones y dos jugadores que tienen que poner piezas de forma alterna en el mismo. cada jugador decide en qué columna va a mover su ficha, y esta se coloca encima de la ficha más alta en esa misma columna. El jugador que empieza utiliza fichas blancas, y el otro fichas negras.

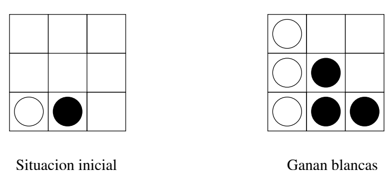


Figura 1: Juego del conecta-3

El juego finaliza cuando:

- En el tablero hay tres fichas del mismo color cubriendo una línea horizontal, una vertical o una diagonal, y ganará el jugador cuyas fichas cumplan la condición
- El tablero está lleno de fichas y no se cumple la condición anterior. En cuyo caso se dice que ha habido un empate

Se desea diseñar un programa que, utilizando alguna de las técnicas vistas en clase, sea capaz de aprender a jugar de forma que, al menos, no pierda nunca.

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, cómo sería capaz de aprender a jugar, etc...

Problema 2. (3 Puntos)

Sea el siguiente problema:

> Dado el perímetro de un rectángulo, que debe medir al menos 81 metros y 92 centímetros y cuyos lados tienen valor entero (en centímetros), construir un rectángulo cuya área sea máxima

Diseñar un sistema, basado en alguna de las técnicas estudiadas a lo largo del curso, capaz de obtener una solución. describir dicho sistema en detalle, incluyendo ejemplos.

Problema 3. (4 Puntos)

Resolver mediante técnicas de computación evolutiva el siguiente sistema de ecuaciones lineales:

$12x + 2y + 4z = 12$

$8x + 13y + z = 6$

$6x + 2y + 2z = 1$

$3x + 4y + 4z = 4$

De manera que se pueda conocer:

1. Si el sistema tiene o no solución, y cuál es.
2. Si el sistema tiene más de una solución y cuales son.

Resolver el sistema consiste en calcular los valores de las incógnitas para que se cumplan TODAS las ecuaciones del sistema simultáneamente.

Para ello es necesario describir con detalle qué técnica de computación evolutiva es la más apropiada para resolver este problema, así como la codificación a utilizar, la función de evaluación, los operadores de selección, cruce y mutación elegidos y los parámetros apropiados para la ejecución de la técnica.

Si se modificaran los valores de los coeficientes (12,2,4, etc..) ¿Cómo afectarían a la técnica descrita, y cúales serían las modificaciones a realizar?. Las respuestas deben estar debidamente justificadas.

Solución al problema 1.

Se pide diseñar un sistema de aprendizaje basado en computación biológica capaz de aprender a jugar al conecta-3. Para jugar se necesita una estrategia, y esta debe poder ser codificada a bajo nivel, para que se pueda utilizar alguna de las técnicas estudiadas en clase. Una manera sería tratar de especificar cada una de las posibles situaciones que un jugador se pueda encontrar y codificar qué es lo que se haría en cada una de ellas. Si se codifican todas las posibles situaciones, tal cual, el número de posibilidades crece enormemente. Por lo tanto es necesario recoger en la codificación información del dominio, en este caso el juego.

Se va a utilizar un Algoritmo Genético, en cuyo caso hay que especificar la codificación y la función de evaluación.

Codificación: Cada individuo es una estrategia de juego, de tal forma que un individuo debe contener reglas capaces de jugar una partida completa. Un individuo estará compuesto por **8** cromosomas. Cada cromosoma se referirá a una columna (3), una fila (3), o una diagonal (2). Cada cromosoma contendrá **6** genes, cada uno especificando las situaciones que se pueden dar en su fila-columna-diagonal correspondiente. El primer gen se refiere a la posibilidad de que esté vacía, el segundo que haya una ficha blanca y ninguna negra, etc... En la tabla siguiente está el significado de todos los genes:

Gen	Significado
1	No hay posibilidad de 3 en raya
2	Hay posibilidad de 3 en raya para blancas y ya hay una ficha blanca colocada
3	Hay posibilidad de 3 en raya para blancas y ya hay dos ficha blancas colocadas
4	Hay posibilidad de 3 en raya para negras y ya hay una ficha negra colocada
5	Hay posibilidad de 3 en raya para negras y ya hay dos ficha negras colocadas
6	Hay posibilidad de 3 en raya para las dos y no hay ninguna ficha colocada

Todos los cromosomas tienen la misma estructura y se refieren a su columna, fila o diagonal asociada. Como la respuesta que debe generar la regla es el movimiento de una pieza y un movimiento viene especificado por una columna a elegir, las posibilidades son tres. Por lo tanto habrá tres alelos por gen, que llamaremos 1, 2 y 3. Por lo tanto el individuo estará compuesto por:

- Un individuo compuesto por ocho cromosomas
- Cada cromosoma tendrá cinco genes
- Cada gen tres posibles valores

Se va a ilustrar la codificación con un ejemplo. Supongamos que el juego está en la situación de la figura 2.

La máquina está jugando con el siguiente individuo:

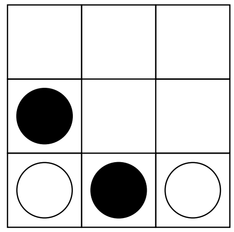


Figura 2: Mueven blancas

Filas		Columnas		Diagonales	
1	21123**1**	1	**3**12313	1	3**3**1233
2	221**3**31	2	213**3**12	2	3**1**1231
3	**2**13123	3	2**3**3311		

En negrita aparecen los valores de los genes que se corresponden con la situación en la que se encuentra el juego. Estos valores viene determinados por la codificación elegida. En la tabla siguiente se indica qué genes están activos en función de la situación del juego observada para cada fila, cada columna y cada diagonal:

	Posibilidad 3-en-raya	Cuantas	Gen	Salida
Fila 1	Los dos		Sexto	Columna 1
Fila 2	Negras	una	Cuarto	Columna 3
Fila 3	Ninguno		Primero	Columna 2
Columna 1	Ninguno		Primero	Columna 3
Columna 2	Negras	una	Cuarto	Columna 3
Columna 3	Blancas	una	Segundo	Columna 3
Diagonal 1	Blancas	una	Segundo	Columna 3
Diagonal 2	Blancas	una	Segundo	Columna 1

Se aprecia que la fila uno está vacía, por lo tanto los dos tienen posibilidad de hacer 3-en-raya, lo cual se corresponde, según la codificación, con el sexto gen del primer cromosoma, que como tiene un valor de 1, está indicando que en esta situación hay que mover en la columna uno. La fila dos contiene una ficha negra y ninguna blanca, por lo que sólo negros podrán hacer 3-en-raya, y además ya tienen una ficha colocada. Esto se corresponde con el cuarto gen del cromosoma 2, que como tiene un valor de 3 recomienda mover en la columna 3. La interpretación sería la misma para el resto de las filas, columnas y diagonales.

Después de que el individuo es evaluado en la situación del juego, nos recomienda 5 veces que movamos en la columna 3, 2 veces que lo hagamos en la 2 y una vez en la 1. Por lo tanto se decide mover en la columna 3.

En el caso de que dicha columna no sea factible, porque ya esté llena, se eligiría la siguiente recomendación, y en caso de que no hubiese ninguna recomendación alternativa (cosa que ocurrirá en un número muy pequeño de casos), se mueve aleatoriamente, y se cambia el valor del gen que ha sido activado, por el que se haya generado aleatoriamente. Esto evita que en una situación similar futura, la regla de una salida incorrecta.

Evaluación: El sistema debe ser capaz de jugar al conecta-3, cada individuo es una estrategia de juego, la forma de evaluarlo es jugando contra otras estrategias. Si la población está compuesta de un número n de individuos, cada individuo se podría evaluar mediante el resultado de partidas realizadas contra los demás. Es decir, para evaluar al primer individuo se le hace jugar contra los $n-1$ restantes, con cada uno dos veces, una vez con blancas y otra con negras, y el valor de evaluación podría ser el número de partidas que ha ganado multiplicado por dos, más las que ha empatado. Así si tenemos 100 individuos, cada uno jugará 198 partidas. Si de entre ellas ha ganado 20, empatado 38 y perdido 140, el valor de evaluación del individuo será $f(I) = 20 \cdot 2 + 38 = 78$.

Ejecución: El sistema se ejecutará de la siguiente manera:

- Se generan n individuos de manera aleatoria
- Se juegan las partidas todos con todos a doble vuelta (una con blancas y otra con negras) y se obtiene un valor de evaluación para cada individuo
- Se genera una nueva población siguiendo el procedimiento general de Algoritmos Genéticos, aplicando selección, cruce y mutación
- Se repite todo el proceso un número m de generaciones

Operadores genéticos: Los operadores genéticos actúan a nivel de cromosoma, es decir para la mutación se elige un gen por cromosoma, con una probabilidad prefijada p_m y se cambia el valor del gen por uno de los otros dos valores. Si el gen tiene un valor de 2 cambia a un valor de 1 o de 3 de manera equiprobable.

El cruce se hace también a nivel de cromosoma, cada cromosoma genera un locus de cruce y se realiza cruce simple cromosoma a cromosoma.

Solución al problema 2.

Si se denominan x e y a los lados del rectángulo, y teniendo en cuenta que el área máxima se corresponderá siempre con un perímetro máximo, la condición del problema podría formularse como:

$$2 \cdot x + 2 \cdot y = 8192$$

Si despejamos una de las variables, por ejemplo la x, quedaría:

$$x = 4096 - y$$

Por lo que sólo es necesarío obtener el valor correspondiente a la varieble x, ya que una vez obtenida esta, el valor de la y se obtendría mediante la expresión anterior.

Por lo tanto se podría utilizar un Algoritmo Genético, en el que se codificaría la variable x, cuyo valor estaría comprendido en el intervalo $[0-4095]$

Por lo tanto la codificación podría ser binaria, utilizando 12 bits ($2^{12} = 4096$). Dado un individuo cualquiera, por ejemplo 100100110011, se correspondería con los valores $x = 2355$ e $y = 1741$.

Como lo que hay que optimizar es el área del rectángulo, el valor de evaluación será precisamente dicha área. Para el individuo anterior la evaluación sería: $x \cdot y = 4100055$.

Un algoritmo diseñado de esta forma tendería a obtener la solución óptima, que como es sabido es un cuadrado, es decir $x = y = 2048$, alcanzando su valor máximo en $x \cdot y = x^2 = 4194304$

Solución al problema 3.

Se trata de un sistema de 4 ecuaciones y 3 incógnitas. Los números reales a_{ij} se denominan coeficientes, x, y, z se denominan incógnitas (serán los números a determinar) y b_j son los términos independientes. Al haber 4 ecuaciones y únicamente 3 incógnitas puede que exista más de una solución que resuelva el sistema.

Es necesario diseñar una técnica de computación evolutiva que resuelva el sistema de 4 ecuaciones y 3 incógnitas planteado. Para ello se opta por una algoritmo de estrategias evolutivas (ES).

Codificación: las ESs utilizan codificaciones reales para representar los parámetros, en este caso escogemos un vector de tamaño fijo de números reales que representan el valor de cada una de las incógnitas:

x	y	z
1.0	2.3	3.2

Tal y como se describió en clase también es necesario trabajar con un vector de varianzas para cada una de las incógnitas inicializado aleatoriamente.

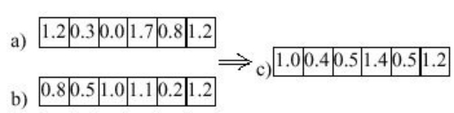
Función de evaluación: Para evaluar se sustituye el valor de las incógnitas en la ecuación, y se obtiene un valor que se compara con los términos independientes. Por ejemplo, si al sustituir las incógnitas se obtienen los valores 8, 3, -5 y 1 respectivamente para cada una de las ecuaciones, como los valores de los términos independientes son 12, 6, 1 y 4, la evaluación sería:

$$f(I) = |8 - 12| + |3 - 6| + |-5 - 1| + |1 - 4| = 16$$

Operadores genéticos: Se utilizará una Estrategia Evolutiva múltiple, donde los operadores se definen de la siguiente manera:

- Operador de cruce: el sobrecruzamiento se realizará sobre los dos padres seleccionados y se generará el nuevo individuo mediante la obtención de los puntos medios de cada uno de los genes que componen el cromosoma, ejemplo:

  Figura 3: Ejemplo de cruzamiento

  También se deberá realizar el proceso análogo para el vector de varianzas.

- Operador de mutación: una vez generado el individuo se le aplicará el operador de mutación a cada uno de los genes y al vector de varianzas siguiendo una distribución Gaussiana.

En este caso será necesario modificar la política de inserción y reemplazo. Como es posible que el número de soluciones sea múltiple, se necesitará utilizar algún método que permita que se generen más de una soluciones. Para ello lo más adecuado sería algún tipo de método de compartición o "crowding". Una solución podría ser utilizar una Estrategia Evolutiva de estado estacionario. En este caso la población constaría de n individuos, pero en cada generación se generaría uno sólo. Para ello se seleccionan dos individuos de forma proporcional a su valor de evaluación, se genera uno nuevo mediante cruce y mutación, y este se introduce en la población mediante:

- Bien sustituyendo a aquel individuo de la población que más se le parezca
- Bien sustituyendo al peor de los padres, siempre que el fitness del descendiente sea mayor que el padre al cual va a sustituir (crowding determinista)

Para asegurarnos de que existe alguna solución bastará con dejar ejecutándose el algoritmo un número suficiente de iteraciones. La condición de parada puede venir determinada por el vector de varianzas, cuando durante m generaciones no haya variaciones significativas podemos dar por finalizada la ejecución del algoritmo. La técnica de crowding determinista asegura la aparición de más de una solución en el caso de que la hubiera.

Si se modifican los coeficientes del sistema de ecuaciones basta con volver a inicializar el algoritmo con los nuevos coeficientes.