Analizis avanzado de algoritmos

Por: David Gomez 20359596



Algoritmos geneticos

tecnica de busqueda por algorimos evolucionarios

ASIGNACIÓN NÚMERO UNO

El problema

Se desea encontrar la secuencia de bits, de 16 dígitos de longitud, que contenga la mayor cantidad de subsecuencias 1 0.

Si bien este problema pudiera parecer de solución trivial sirve de mucho para ilustrar las cualidades de los algoritmos genéticos, los cuales se orientan a resolver problemas para los cuales sabemos evaluar una solución propuesta pero no sabemos encontrar la mejor solución.

La dinámica del algoritmo genético es la de inducir una solución a partir un conjunto inicial de posibles soluciones factibles; esto es, generar una nueva solución tomando las propiedades de otras soluciones, antes que resolver el problema en sí mismo. Por ello los algoritmos genéticos se clasifican como algoritmos de búsqueda heurística.

La [selección](https://es.wikipedia.org/wiki/Selecci%C3%B3n_(computaci%C3%B3n_evolutiva)) de las soluciones más deseables sobre las cuales formar una nueva solución viene dada de acuerdo con algún criterio el cual, en este caso, resulta dado por el número de pares **1 0** contenidos en dicha secuencia.

El que conozcamos cual es el máximo de pares para cualquier solución factible (entendiéndose solución factible a toda secuencia de valores binarios, con una longitud 16 dígitos) implica que conocemos de antemano cual es la solución óptima al problema (Esto es, una secuencia que contiene 8 subsecuencias 10). Este hecho a de usarse para nada más que conocer lo que quisiéramos que fuera la solución obtenida en cada aplicación puesto que la utilidad de este tipo de algoritmo radica en su habilidad para encontrar la mejor solución, o una muy cercana, aunque no se conozca con certeza cuál es esta.

Descripción de la solución propuesta

Desarrollar un programa escrito en código java que implemente la técnica de algoritmos genéticos para dar con la solución óptima a partir de un conjunto de soluciones generadas aleatoriamente.

El primer concepto clave en los algoritmos genéticos es el de **conjunto de soluciones factibles**. Este conjunto contendría todas y cada una de las posibles soluciones a nuestro problema. Una búsqueda convencional implicaría, en el peor de los casos, recorrer la totalidad del conjunto para hallar la solución deseada; en lugar de ello, los algoritmos genéticos exploran una pequeña porción del conjunto de soluciones factibles dividiéndolo en subconjuntos, a los que llamamos **poblaciones** para mantener una analogía completa a los postulados de la evolución biológica.

Llamamos **generación** a cada repetición de un ciclo iterativo que consiste en producir una nueva población conformada por los “descendientes” de los **individuos** (Soluciones) de la población actual, los cuales se dice que “mueren” al pasar a la próxima generación.

Llamamos a los tres métodos principales involucrados en la producción de nuevas población Selección, Cruce y Mutación.

**Selección:** dada una población cuyos individuos se encuentran ordenados descendentemente según cuan deseables sean, esto es; en este caso, ordenados según la cantidad de subsecuencias **10** que contengan. Se ha de dotar con más probabilidad de ser seleccionados para un cruce a aquellos individuos que demuestren las características más deseables. La probabilidad de selección para cada individuo viene dada por la función:

Donde **n** es el tamaño de la población y **Pi** se refiere a la posición del individuo en el ordenamiento.

**Cruce:** Una vez seleccionados dos individuos, procedemos a desarrollar un nuevo individuo que heredara las características de sus padres. El proceso de cruce para este caso consiste en que para cada posición de la secuencia hija se decide aleatoriamente si tomar los valores del primer padre o si hacerlo del segundo.

**Mutación:** En promedio, uno de cada mil individuos producidos por la función de cruce será seleccionado para sufrir una mutación. La mutación cambiara un número aleatorio de dígitos en la secuencia por su negación lógica.

Tras producir una nueva población, el individuo más deseable de ella (aquel ubicado en la cima de la lista) es tomado como solución general al problema en esta generación. Para cada generación se ha de evaluar que la población actual tenga suficiente diversidad como para producir cambios significativos en la próxima generación con respecto a la presente. Esto se efectúa mediante el cálculo de la **varianza**, que viene dada por la función:

s_n^2 = \frac 1n \sum_{i=1}^n \left(X_i - \overline{X} \right)^ 2 = \left(\frac{1}{n} \sum_{i=1}^{n}X_i^2\right) - \overline{X}^2

Donde n es el tamaño de la población, X’ es la media y Xi es el valor de un individuo en particular. Para facilitar el cálculo de la varianza, se toma que el valor de cada individuo viene dado por el valor de la representación decimal del número binario que formara dicha cadena, habiendo entonces un total de 216 (65536) posibles valores.

El programa solo se detiene hasta que la varianza alcanza un mínimo permitido o se alcanza el máximo de iteraciones.

Experimentos y resultados

Se trabajó con dos tamaños de población, 100 y 200.

El número máximo de generaciones se mantuvo a un total de 200.

La probabilidad de mutación se mantuvo siempre a un total de 1/1000.

Para poblaciones de tamaños 100 se trabajó con un mínimo de varianza de 7,5% y para las de 200, 7,5% y 5%.

Se ha tomado una muestra de 12 aplicaciones para cada una de las dos configuraciones. Los resultados permiten hacer ciertas especulaciones, como que mientras más grande sean las poblaciones, mayor será la posibilidad de obtener una mejor solución. Por otro lado, parece que valores el número de generaciones esta en relación inversa con la el mínimo permitido a la varianza.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| N = 100 var min = 7,5% | | N = 200 var min = 5% | | N = 200 var min = 7,5% | |
| Fitness | Generaciones | Fitness | Generaciones | Fitness | Generaciones |
| 7 | 191 | **8** | 22 | **8** | 22 |
| 7 | 48 | **8** | 18 | **8** | 13 |
| 8 | 17 | **8** | 22 | **8** | 32 |
| 7 | 144 | 7 | 200 | **8** | 18 |
| 7 | 59 | 7 | 200 | **8** | 14 |
| 7 | 32 | **8** | 23 | **8** | 19 |
| 8 | 22 | 7 | 200 | 7 | 200 |
| 7 | 36 | 7 | 60 | 7 | 148 |
| 7 | 74 | 7 | 85 | 7 | 71 |
| 7 | 86 | 7 | 72 | **8** | 21 |
| 7 | 100 | 7 | 137 | 7 | 200 |
| 8 | 22 | **8** | 17 | 7 | 200 |
| Media=7,25 | Media=**69,25** | Media=**7,41** | Media=**106,58** | Media**=7,58** | Media**=79,83** |

Conclusiones

Se concluye que los algoritmos genéticos son dinámicas útiles para explorar un conjunto de soluciones muy amplio y realizar búsquedas eficientes. Son, además, herramientas interesantes capaces de inducir soluciones, soluciones para las que no se tenía indicios claros de como hallarlas.