Universidad Simón Bolívar
Departamento de Computación y T.I.
Inteligencia Artificial II
Septiembre-Diciembre 2011

Tarea 3

Fecha de entrega: Lunes 21 de Noviembre.

Se requiere que, usando un algoritmo genético (GA), se realice la clasificación un conjunto de datos correspondiente a una encuesta realizada en el año 1987 en Indonesia para determinar el método anticonceptivo más utilizado por parejas en el país.

Los ejemplos se corresponden con características socio-económicas de mujeres casadas que o bien no estaban embarazadas al momento de la encuesta o no lo sabían.

El problema es predecir el método anticonceptivo que utilizan (ninguno, métodos a largo plazo, métodos a corto plazo).

La descripción de los datos, así como los conjuntos ejemplos (que deben separar en los conjuntos de entrenamiento y prueba se encuentran disponibles en:
`http://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice.`

Su implementación del GA debe partir del sistema GABIL [1]. Una breve descripción de este sistema también la pueden conseguir en el capítulo 9 del Mitchell [2]. Pueden usar cualquier librería de GAs en el lenguaje de su preferencia.

Se espera lo siguiente de su proyecto:

- Deberán codificarse y probarse 2 versiones de la función de selección de padres y 2 versiones de la función de selección sobrevivientes. Una de las funciones de selección deberá ser la de "rueda de ruleta".
- Agregar las dos extensiones propuestas al algoritmo básico de GABIL: el operador "AddAlternative" y el operador "Drop Condition".
- Incorporar a su función de fitness un mecanismo para controlar el tamaño de los clasificadores.

Deben realizar experimentos para conseguir la mejor configuración (combinación de los operadores de selección de padres y sobrevivientes). Una vez determinada la mejor configuración, para la mejor de estas configuraciones realizar variaciones sobre las tasas de mutación y crossover (al menos 3 valores para cada una). Sobre la mejor configuración y parámetros hallados, probar con cada uno de los operadores de extensión ("AddAlternative" y 'Drop Condition") por separado y con los dos operadores juntos. Probar variaciones sobre la penalización al tamaño de los clasificadores y comparar los resultados obtenidos al incluir o no dicha penalización.

Entrega: La fecha de entrega Lunes 21 de noviembre a la 1:30 pm en clases. Deberán entregar:

- Impreso:
  Un breve informe que contenga:
  1. La descripción de la implementación (o uso de librería).
  2. La descripción del AG (los parámetros base) que usaron.
  3. La descripción y el análisis de los experimentos realizados.
  4. En el informe deben dar respuesta a las siguientes preguntas:
     a) ¿Cuál es la mejor configuración de su algoritmo genético para clasificar los datos estudiados?

b) ¿ Son útiles los operadores de la extensión?

c) ¿ Cuál es el mejor conjunto de reglas hallado por el algoritmo genético? Considerando el número de ejemplos clasificados correcta e incorrectamente.

d) Describa la función de fitness utilizada. ¿Es útil incluir en la función de fitness un factor de penalización a clasificadores muy grandes?

Recuerde que por ser los algoritmos genéticos algoritmos estocásticos deben reportar el promedio de varias corridas (al menos 10) para cada configuración.

- En el aula virtual:
  Deberán subir un archivo tar.gz que contenga el código del proyecto, el informe en PDF, e intrucciones de ejecución del programa (README.txt).

Bibliografía

[1] Kenneth A. De Jong and William M. Spears and Diana F. Gordon. *Using Genetic Algorithms for Concept Learning*, Machine Learning, 13, pp. 161-188, 1993, `citeseer.ist.psu.edu/dejong93using.html`. consultado el 15/11/2007.

[2] Tom M. Mitchell, *Machine Learning*, McGraw-Hill, 1997.

[3] Melanie Mitchell, *Handbook of genetic algorithms*. Artif. Intell., 100(1-2): 325–330, 1998.