



Universidad Simón Bolívar
Departamento de Computación y T.I.
Inteligencia Artificial II
Septiembre-Diciembre 2014

Proyecto 2 - GABIL (15 puntos)

Fecha de entrega: Lunes 3 de Noviembre

Se requiere que, usando un algoritmo genético (GA), para construir un clasificador que permita clasificar las tres clases de la flor Iris (Setosa, Versicolour, Virginica), usando los datos disponibles en:

<http://archive.ics.uci.edu/ml/datasets/Iris>.

Este conjunto de datos está compuesto por 150 ejemplos con 4 atributos continuos (reales) y la clasificación correcta para cada ejemplo.

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”.
- 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).

Probar la variación sobre la función de fitness que penaliza el tamaño de los clasificadores y comparar los resultados obtenidos al incluir o no dicha penalización.

Entrega:

La entrega será hasta el día lunes 3 de noviembre, a la hora de clases (11:30 am).

Deberán entregar:

- Impreso:
Un breve informe que contenga:
 1. Resumen
 2. La descripción de la implementación (o uso de librería).
 3. La descripción del AG (los parámetros base) que usaron.
 4. La descripción y el análisis de los experimentos realizados.
 5. 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) Cuál es el mejor conjunto de reglas hallado por el algoritmo genético? Considerando el número de ejemplos clasificados correcta e incorrectamente.
- c) 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.

- Por correo electrónico :

Un archivo comprimido que contenga: su código, las instrucciones de ejecución (README.txt) y el informe (.zip o .gz) al correo:

`ivettecarolinamartinez@gmail.com` con el asunto “IA2: Proyecto2, Grupo [XX]”.

Referencias

- [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, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.48.6461>. consultado el 22/02/2013.
- [2] Tom M. Mitchell, *Machine Learning*, McGraw-Hill, 1997.
- [3] Melanie Mitchell, *Handbook of genetic algorithms*. Artif. Intell., 100(1-2): 325–330, 1998.