

PRACTICA Nº 4

EL SISTEMA FOIL (V. 6)

1º) **FOIL** es un sistema de aprendizaje de reglas con variables, en forma de cláusulas Horn que admiten literales negativos pero no funciones. Su autor Ross Quinlan, quien en algunas versiones fue ayudado por Cameron-Jones, no lo considera un sistema definitivo, al menos en la versión 6 que se manejará en esta práctica. El software es el original de sus autores.

2º) El sistema se encuentra ubicado en el ordenador *centauro.aulario.uniovi.es*. Existen unos ficheros con conjuntos de ejemplos que acompañan al propio código del sistema y que, junto con otros añadidos, se puede copiar de la dirección */profesores/alguero/foil/ejemplos* en el citado ordenador. La sintaxis de los ficheros de entrada tiene su dificultad y se hace imprescindible leer el manual.

3º) Para una primera ejecución se puede utilizar el fichero *golf.d* creado automáticamente de la versión homónima para C4.5 mediante la utilidad *c4tofoil* que traduce del formato de entrada de C4.5 al formato de entrada de FOIL. Edítese el fichero *golf.d* y analícese su sintaxis. Para realizar el proceso de aprendizaje sobre este conjunto de ejemplos, ejecútese el comando:

foil6 < golf.d

que en esta sencilla formato toma todos los parámetros por defecto. Estúdiense el resultado y compárese con el obtenido por los sistemas de las prácticas anteriores. El resultado pueden ser reglas del tipo *hechos*; es decir, sin antecedente, o reglas en el pleno sentido de la palabra y con la estructura:

consecuente :- antecedente

que se lee:

consecuente si antecedente

El antecedente puede constar de una o más relaciones (predicados) separadas por comas que serán interpretadas como conjunciones. Por ejemplo, si:

antecedente = *relación_1(-, -, ... , -), relación_2(-, -, ... , -), ...*

se leerá:

relación_1(-, -, ... , -) y relación_2(-, -, ... , -) y ...

4º) El archivo *member.d* contiene ejemplos positivos y ejemplos negativos del concepto *número entero miembro de una lista de enteros*. Se recomienda observar la estructura del fichero y familiarizarse con la distribución de las partes más importantes: *declaración de tipos*, *relación objetivo* del aprendizaje y los *ejemplos positivos* y *ejemplos negativos* de ella; a continuación le siguen las *ejemplos positivos de las relaciones intermedias* que se utilizan para definir la relación objetivo y por último los *ejemplos de la relación objetivo para realizar el test*, positivos y negativos, si los hubiera.

Una segunda ejecución de FOIL con los parámetros por defecto puede realizarse sobre este archivo *member.d* para lo que se ejecutará el comando anterior sobre el archivo *member.d*:

foil6 < member.d

Para que no proporcione soluciones con literales negativos ni con la relación $\langle \rangle$ (*ser diferente*) se excluyen esas posibilidades mediante el parámetro *-n*; con el añadido del parámetro *-v3* se le obligará a proporcionar información con un grado de detalle 3. Por tanto, si se ejecuta el comando:

foil6 -n -v3 < member.d

obtendremos un resultado sin literales negativos, si la relación $\langle \rangle$ y con cierta abundancia de información. Concretamente el resultado de esta ejecución aparece con explicaciones detalladas en el Manual Breve del Sistema FOIL¹ que complementa a este guión.

5º) La utilidad que permite la traducción de ficheros de ejemplos para el sistema C4.5 al formato de FOIL necesita previamente la ejecución del siguiente comando para generar un ejecutable:

cc -0 cf c5tofoil.c

¹ Traducción libre y adaptación del manual de Quinlan que y Cameron-Jones que acompaña al software de FOIL

Para realizar una traducción de formatos será necesaria la existencia de al menos dos archivos relativos a un *filestem*, aquellos con nombres *filestem.names* y *filestem.data*; eventualmente puede también existir el archivo *filestem.test*. El archivo original *filestem.names* necesitará también un pequeño añadido en la declaración de todos los atributos; concretamente se especificará el tipo de cada atributo a continuación de su declaración en el archivo *filestem.names* original, añadido como comentario para C4.5. Tal como se indica a continuación:

"Declaración del atributo según sintaxis de C4.5, terminando en un ." | **type:** nombre_tipo
donde se respetarán escrupulosamente los espacios blancos previo y posterior a la barra vertical. que señala el comienzo de un comentario en el formato de C4.5. Ejemplo:

Aspecto_del_cielo: soleado, cubierto, lluvioso. | type: Aspecto-

donde el nombre del tipo es elección del usuario siempre que no sea continuo (para más información véase el Manual Breve del Sistema FOIL). Un nombre de un tipo no puede ser un prefijo del nombre de otro tipo. Además, hay que tener presente que la relación objetivo del aprendizaje será la relativa a la primera clase que aparezca declarada en el fichero *filestem.names*, por lo que habrá que cerciorarse que aparece la que interesa. Después de tener adaptado el archivo *filestem.data* se ejecutará el siguiente comando:

cf -f filestem

El resultado será un archivo con nombre *filestem.d* que se adecuará el formato de FOIL, e integrará los ejemplos para el test, si es que existía el fichero *filestem.test*. Se recomienda editarlo y estudiar su sintaxis para facilitar la comprensión de resultado del aprendizaje que se realice posteriormente.

6º) Para los conjuntos de ejemplos con *filestem monk1* y *monk2* existen ya los archivos *monk1.d* y *monk2.d*, obtenidos mediante la traducción automática de formatos, aunque siempre es posible realizar una nueva traducción con nombres más sugerentes para los tipos. Edítense, inspecciónense, realícense sobre ellos procesos de aprendizaje y compárense los resultados con los obtenidos por C4.5; en esta comparación ténganse presente la definición de la clase 1 en esos conjuntos de ejemplos (aparece en los archivos de extensión *names* para cada *filestem*).

7º) Tradúzcanse los archivos del *filestem monk3*, para lo que previamente habrá que adaptar el archivo *monk3.names*, y no olvidarse de situar la clase 1 en primer lugar. Puesto que todos los problemas *monk* vienen definidos en términos de la clase 1, ésta será la clase objetivo en el aprendizaje de FOIL.

8º) Trabájese con otros conjuntos de ejemplos, como los relativos a los *filestem crx*, *member*, etc. y manéjense diferentes parámetros para observar como transcurre su manipulación en el resultado final.

GIJÓN, MAYO DE 2001