

PRÁCTICA 2

ÁRBOL DE

DECISIÓN ID3



Rubén García Mateos – Ingeniería del conocimiento

Objetivo, Lenguaje utilizado y detalles de implementación.

El objetivo principal de esta práctica es implementar el árbol de decisión ID3 en un determinado lenguaje de programación elegido por el alumno, donde el resultado final debe mostrar el árbol general con sus nodos construido a través del algoritmo ID3 explicado en clase.

En mi caso, el lenguaje de programación que he utilizado es C++ junto con el IDE Visual Studio.

Mi algoritmo consiste principalmente en cargar los dos archivos de texto proporcionados al programa. El primero de ellos llamado atributos.txt, contendría los diferentes atributos que el algoritmo debe aprender para posteriormente calcular el mérito y en el segundo, llamado juego.txt, se mostraría los valores diferentes que contienen estos atributos. (Ejemplo: del atributo TiempoExterior, sus diferentes valores serían Soleado, Nublado o Lluvioso).

El programa implementado cargaría inicialmente todos los datos de la tabla en dos vectores, atributos[] donde se guarda el valor de los atributos que tenemos en el primer .txt, y ejemplos[] donde se guardan todos los datos que contiene el segundo .txt.

A partir de este momento, es cuando se ejecutaría el algoritmo ID3, donde principalmente contaríamos con otro vector denominado "columna" en el que se guardan los datos de la primera columna, segunda... de la tabla proporcionada en el "juego.txt" en cada iteración. Con estos datos guardamos en el vector columna, a través de un bucle obtenemos cuantos valores diferentes en total aparecen en la columna, y comparando con la última columna de la tabla, que es el denominado vector decisión[], obtenemos el número de casos positivos y negativos de cada valor.

Junto con estos valores obtenidos calcularíamos el porcentaje de casos positivos y negativos obtenidos, y llamaríamos al método creado llamado calculaMérito que nos proporcionaría el mérito de los diferentes valores que aparecen en la columna, y para generar el mérito final del atributo de la columna, se sumarían estos méritos individuales.

El programa cuando calcule todos los méritos, seleccionaría el mas pequeño que se corresponderá con el primer nodo del árbol del ID3.

Partes Opcionales

En lo relativo a las partes opcionales, únicamente he dispuesto de tiempo para implementar la parte en la que el algoritmo se puede utilizar para cualquier archivo de texto que queramos pasarle como argumentos en el programa. De tal manera que podemos tener 100 archivos de texto y te calculará el primer nodo del algoritmo para todos los archivos.

Manual de Usuario

Para poder ejecutar el programa hay que descomprimir el proyecto en la carpeta raíz de proyectos de V.S. Dentro de Visual Studio pincharíamos archivo->abrir proyecto y seleccionamos el proyecto que hemos guardado. (Dentro del proyecto se incluyen varios ficheros de texto llamados "atributosX.txt", donde la X indica el archivo para el cual queremos calcular su nodo inicial para el ID3. "atributosX.txt" contiene los atributos del juego y "juegoX.txt" contiene los diferentes ejemplos que completan las columnas de los atributos.

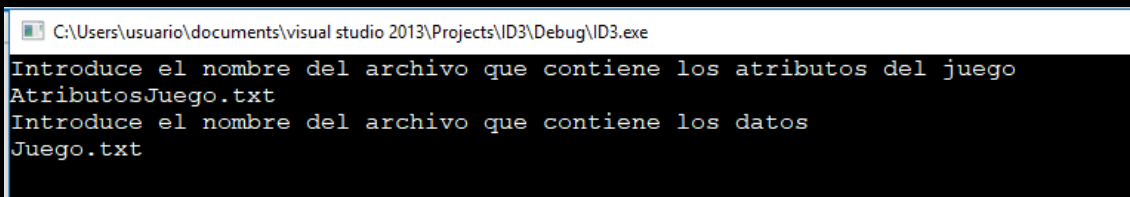
Nota: El proyecto completo contiene varios ficheros de texto para que se pueda probar por parte del profesor más ejemplos añadidos por el alumno para comprobar que funciona para todos los archivos de texto insertados. Si se quiere realizar pruebas de los otros archivos basta con volver a ejecutar y escribir el nombre dentro del programa.

Los nombres de los archivos que hay que introducir dentro del programa para ejecutar los distintos juegos se corresponden con los siguientes:

Juego1: "AtributosJuego.txt" y "Juego.txt"

Juego2: "AtributosJuego2.txt" y "Juego2.txt"

Juego3: "AtributosJuego3.txt" y "Juego3.txt"



```
C:\Users\usuario\documents\visual studio 2013\Projects\ID3\Debug\ID3.exe
Introduce el nombre del archivo que contiene los atributos del juego
AtributosJuego.txt
Introduce el nombre del archivo que contiene los datos
Juego.txt
```

El simulador se ejecutaría y nos mostraría paso por paso lo que ha ido realizando:

Primero antes de mostrar el paso 1, el programa pedirá que se inserte el nombre del archivo de los atributos, y el nombre del archivo de los ejemplos del juego para posteriormente ejecutar los pasos siguientes:

1: En el paso 1, mostraría la tabla inicial a la que se ha aplicado el algoritmo ID3, atributos junto con sus datos diferentes en cada columna. (Este paso se corresponde con la carga de los datos en los vectores y la salida por pantalla de los mismos)

```
C:\Users\usuario\documents\visual studio 2013\Projects\ID3\Debug\ID3.exe

1: MOSTRAMOS LA TABLA QUE SE DA EN LOS ARCHIVOS TXT CON LOS ATRIBUTOS

-----
TiempoExterior  Temperatura  Humedad  Viento  Jugar
-----

soleado         caluroso      alta      falso   no
soleado         caluroso      alta      verdad  no
nublado         caluroso      alta      falso   si
lluvioso       templado      alta      falso   si
lluvioso       frio         normal    falso   si
lluvioso       frio         normal    verdad  no
nublado         frio         normal    verdad  si
soleado         templado      alta      falso   no
soleado         frio         normal    falso   si
lluvioso       templado      normal    falso   si
soleado         templado      normal    verdad  si
nublado         templado      alta      verdad  si
nublado         caluroso      normal    falso   si
lluvioso       templado      alta      verdad  no
-----

columnadeMeritoMasPequeno = contadorColumna:
```

2: En el siguiente paso se procede a ir calculando el mérito de todos los atributos de los que consta el archivo, sin incluir la última columna denominada vector decisión. Podemos comprobar como el algoritmo nos muestra el porcentaje de casos positivos en los que aparece un determinado parámetro del atributo, en el caso de la captura de pantalla, en el atributo tiempo exterior con sus diferentes parámetros “soleado, nublado y lluvioso” , los respectivos porcentajes, y al final de esto, el mérito del atributo mostrado. Este paso se ejecuta hasta llegar como he dicho hasta la penúltima columna.

```
C:\Users\usuario\documents\visual studio 2013\Projects\ID3\Debug\ID3.exe

2: VAMOS CALCULANDO EL MERITO DE LOS DISTINTOS ATRIBUTOS PARA OBTENER EL PRIMER NODO

-----
                ATRIBUTO 1: TiempoExterior
-----

El numero de casos positivos de 'soleado' es 2 y negativos 3
El porcentaje de los positivos: 0.4
El porcentaje de los negativos: 0.6

El numero de casos positivos de 'nublado' es 4 y negativos 0
El porcentaje de los positivos: 1
El porcentaje de los negativos: 0

El numero de casos positivos de 'lluvioso' es 3 y negativos 2
El porcentaje de los positivos: 0.6
El porcentaje de los negativos: 0.4

Merito del Nodo TiempoExterior: 0.693536
```

3: Finalmente y para concluir, se muestra el dato del mérito mas pequeño que se ha encontrado entre todos los mostrados anteriormente, y la columna a la que pertenece este mérito y así, se muestra finalmente el Nodo que se corresponde para iniciar el árbol.

```
el merito mas pequeno es : 0.693536 y su columna es: 1
PRIMER NODO DEL ID3: ----> TiempoExterior <----

Presione una tecla para continuar . . . █
```

Como comentario final añadir que el algoritmo me ha parecido algo difícil de implementar, y sobre todo a partir de obtener el primer nodo del árbol, me parecía demasiado complicado la llamada a la recursividad ya que me daba problemas constantes al intentar generar la nueva tabla sin los atributos seleccionados anteriormente para poder continuar con el siguiente nodo, aunque si que he intentado realizar mas ramas del dicho árbol.

Pienso que debido a la dificultad del algoritmo y el tiempo que he tenido que emplear en la práctica junto con los detalles de implementación realizados mi nota debería estar en torno al (6.5-7.0).

Rubén García Mateos, 02551077D, Ingeniería del conocimiento. 4E