

Ingeniería del Conocimiento

Titulación: Grado en Ingeniería del Software

Entrega Memoria Práctica 2 ID3

Curso 2014-2015

Profesor: Gonzalo Pajares Martinsanz

Participante:

Pérez Valbuena, Juan Luis

Indice

[1. Describir el sistema mediante una tarea de Clasificación de CommonKADS 3](#_Toc406322556)

[2. Describir los detalles de implementación propios 3](#_Toc406322557)

[3. Breve manual de usuario para poder ejecutar la práctica. 3](#_Toc406322558)

# Describir el sistema mediante una tarea de Clasificación de CommonKADS

**Planteamiento inicial:** *se trata de diseñar un clasificador de forma que dado un conjunto de atributos con sus correspondientes valores, se tome una decisión (clasificación). Se asume que el clasificador contendrá distintos métodos de razonamiento, tales como un conjunto de reglas producido por el algoritmo ID3, un segundo conjunto por el C4.5, un tercer conjunto introducido directamente por el usuario, etc.* El esquema de la plantilla es el que se muestra a continuación.



1. **Objeto (rol dinámico):** el objeto a clasificar contiene cuatro valores, cada uno de los cuales asociado a los atributos correspondientes, tal y como se muestra en la siguiente tabla.

|  |  |  |  |
| --- | --- | --- | --- |
| TiempoExterior | Temperatura | Humedad | Viento |
| soleado | caluroso | alta | falso |

1. **Generar (inferencia):**

* *Operación:* dado el objeto, seleccionar uno de los conjuntos de reglas o métodos de razonamiento disponibles.
* *Conocimiento estático:* se conocen las reglas generadas por los árboles o introducidas por el usuario.
* *Conducta de control:* seleccionar uno de los conjuntos de reglas aleatoriamente. Si no se encuentra solución o la solución no es satisfactoria se selecciona el siguiente. En lugar de aleatoriamente, podría preferirse seleccionar el conjunto proveniente del ID3.
* *Métodos computacionales:* para la selección no existe ningún método predefinido. Simplemente podría elegirse mediante un generador de números aleatorios o por aplicación de algún conocimiento estático (rol estático no contemplado en el esquema), tal como que el ID3 siempre ha dado buenos resultados.

1. **Candidato (rol dinámico):** el conjunto de reglas seleccionado.
2. **Especificar (inferencia):**

*Operación:* seleccionar por orden cada una de las reglas disponibles

* Conocimiento estático: conjuntos de reglas generadas por los métodos o introducidas por el usuario.
* Conducta de control: disparo de cada una de las reglas disponibles en el conjunto seleccionado hasta encontrar, mediante alguna de ellas, una decisión o solución final.
* Métodos computacionales: razonamiento basado en reglas

1. **Atributo (rol dinámico):** tomando en consideración cada una de las reglas, los atributos que se identifican en este rol dinámico se corresponden con los atributos del objeto a clasificar. Es decir: TiempoExterior, Temperatura, Humedad, Viento.
2. **Obtener (función de transferencia):** asocia un atributo a un valor de ese atributo. Siempre siguiendo una regla.
3. **Característica (rol dinámico):** uno de los valores de los atributos.

**Aclaración**, dada una regla del tipo:

si TiempoExterior = soleado y Humedad = normal 🡪 +TiempoExterior obtiene como característica soleado y Humedad obtiene normal

1. **Corresponder (inferencia):**

* *Operación:* dado el objeto de entrada, comprobar cada una de las asociaciones realizadas mediante: atributo obtener característica. En el caso anterior serían dos: TiempoExterior = soleado y Humedad = normal
* *Conocimiento estático:* N/A.
* *Conducta de control:* compara cada una de las asignaciones realizadas mediante la función de transferencia obtener.
* *Métodos computacionales*: comparación de caracteres, para chequear concordancias de cadena para atributos y características. Funciones propias de los lenguajes de programación, tales como comparar\_strings(str1,str2).

1. **Valor-verdad (rol dinámico):** coincidencia o no entre las premisas de las reglas y los valores de los atributos y las características del objeto dado a la entrada.

# Describir los detalles de implementación propios

En esta segunda práctica de Ingeniería del Conocimiento, se plantea la realización de un algoritmo para toma de decisiones, en concreto el **algoritmo ID3**.

El algoritmo ID3 es un algoritmo de aprendizaje mediante inducción basada en ejemplos, que es un proceso de búsqueda en un modelo de clasificación lo más sencillo posible.

En el caso del ID3, se utilizan árboles para la toma de decisiones, basado en una construcción iterativa desde la raíz del mismo hasta las hojas.

* Los nodos intermedios son atributos de los ejemplos presentados.
* Las ramas representan los valores de los atributos y los nodos finales son los valores de la clase.

## Pasos para la implementación del ID3:

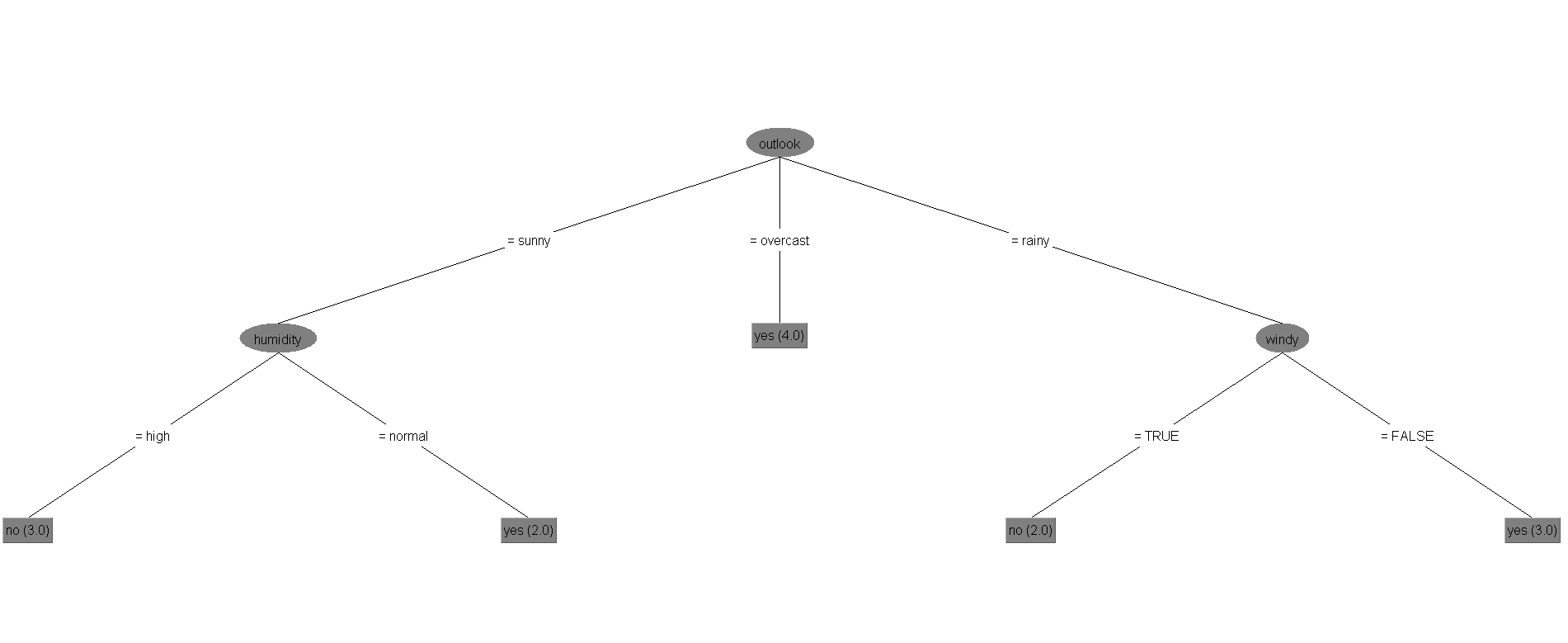
1. Si lista-ejemplos está vacía, "regresar"; en caso contrario, seguir.
2. Si todos los ejemplos en lista-ejemplos son +, devolver "+"; de otro modo seguir
3. Si todos los ejemplos en lista-ejemplos son , devolver ""; de otro modo seguir
4. Si lista-atributos está vacía, devolver "error"; en caso contrario:
5. llamar mejor al elemento a de lista-atributos que minimice mérito
6. iniciar un árbol cuya raíz sea mejor:
   1. para cada valor vi de mejor
      1. incluir en ejemplos-restantes los elementos de lista-ejemplos que tengan valor vi del atributo mejor.
      2. dejar en atributos-restantes todos los elementos de lista-atributos excepto mejor.
      3. devolver el valor de: ID3 (ejemplos-restantes, atributos-restantes) (llamada recursiva al algoritmo)

## Interpretar un árbol de decisión

Para poder determinar la clasificación de un objeto se pregunta sucesivamente en cada nodo del árbol, empezando desde la raíz, el valor indicado en el nodo del objeto, así hasta llegar a un nodo hoja.



Se proporciona un ejemplo basado en el resultado de la propia práctica utilizando Weka 3.7+



En el árbol anterior , se puede observar , por ejemplo , que el primer atributo es *Outlook* ( Tiempo Exterior ) y dependiendo del valor que tenga ( sunny , overcast , rainy ) ( soleado , nublado o lluvioso ) se determina el cual es el siguiente atributo a comprobar el valor.

## Detalles propios de la implementación

Se ha utilizado Python 2.7+ íntegramente para la realización de la práctica

Se ha utilizado una metodología incremental e iterativa para el desarrollo de la práctica.

Como las anteriores prácticas, se encuentra el código realizado colgado en github.com.

Indicar para la valoración numérica de la práctica que ha sido realizada por un único alumno.

Se han realizado las siguientes ampliaciones de la misma:

* Se incluye todos los niveles de recursividad.
* Se comprueba el correcto funcionamiento del algoritmo para ejemplos pasados por fichero , determinados como TestXJuego.txt
* El algoritmo está diseñado de forma general , es decir puede resolver más problemas del que está actualmente indicado.

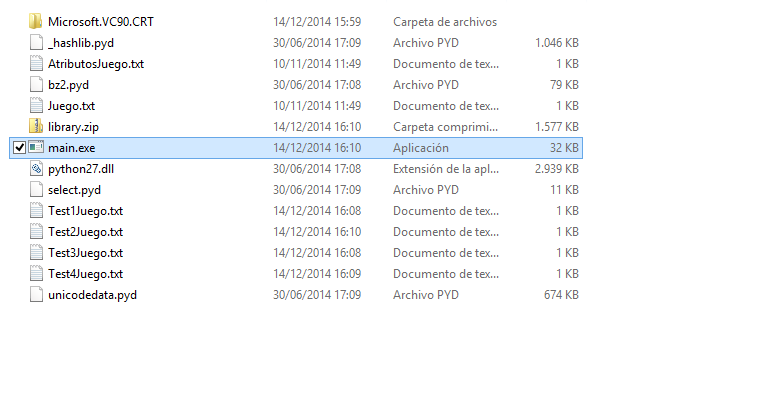
# Breve manual de usuario para poder ejecutar la práctica.

Se incluye, junto a esta documentación, un archivo ejecutable para Windows. Para su correcto funcionamiento, debe ser extraída junto al resto de ficheros que le acompañan, ya que se incluyen diversas librerías que no están por defecto instaladas en el sistema operativo.

Al no ser requerida interfaz gráfica, se ha utilizado íntegramente salida por consola, en modo texto.

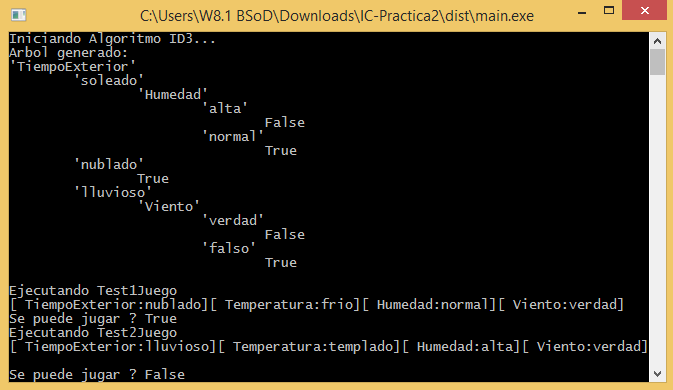
Para comenzar a ejecutar la práctica, ejecutar el archivo adjunto “main.exe”

Se ha incluido un test extra realizado manualmente para poder comprobar la corrección.



Aparecerá la consola y mostrará los siguientes datos:

* Árbol generado de ID3
* Los tests pasados a la aplicación, el valor que tienen en el fichero y la salida obtenida.



Iniciando Algoritmo ID3...

Arbol generado:

'TiempoExterior'

'soleado'

'Humedad'

'alta'

False

'normal'

True

'nublado'

True

'lluvioso'

'Viento'

'verdad'

False

'falso'

True

Ejecutando Test1Juego

[ TiempoExterior:nublado][ Temperatura:frio][ Humedad:normal][ Viento:verdad]

Se puede jugar ? True

Ejecutando Test2Juego

[ TiempoExterior:lluvioso][ Temperatura:templado][ Humedad:alta][ Viento:verdad]

Se puede jugar ? False

Ejecutando Test3Juego

[ TiempoExterior:nublado][ Temperatura:caluroso][ Humedad:alta][ Viento:falso]

Se puede jugar ? True

Ejecutando Test4Juego

[ TiempoExterior:soleado][ Temperatura:templado][ Humedad:alta][ Viento:verdad]

Se puede jugar ? False

Pulse cualquier tecla para salir