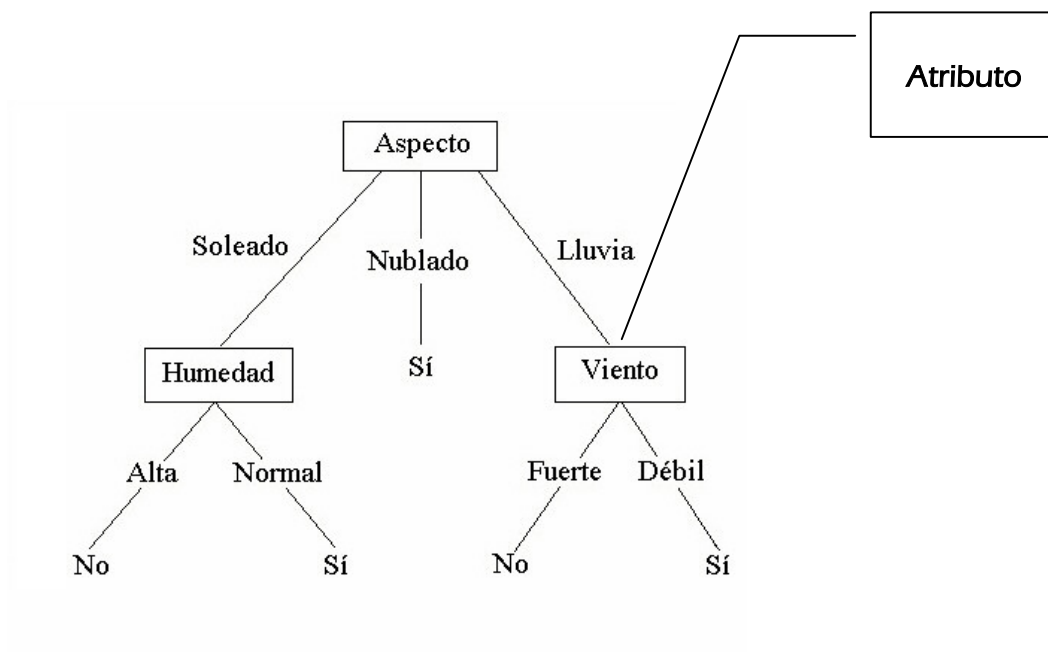


I. CARACTERISTICAS DEL ALGORITMO ID3

- ♦ El investigador J. Ross Quinlan desarrolló el algoritmo conocido como ID3 (*Induction Decision Trees*) en el año de 1983.
 - ♦ Pertenece a la familia TDIDT (Top-Down Induction of Decision Trees).
 - ♦ **Objetivo:** construir un árbol de decisión que explique cada instancia de la secuencia de entrada de la manera más compacta posible, según los criterios de coste y bondad. En cada momento elige el mejor atributo dependiendo de una determinada heurística.
 - ♦ **Inconveniente:** favorece indirectamente a aquellos atributos con muchos valores, los cuales no tienen que ser los más útiles.
 - ♦ Genera árboles de decisión a partir de ejemplos de partida.
-
- ♦ Intenta encontrar el árbol más sencillo que separa mejor los ejemplos.
 - ♦ Recursividad.
 - ♦ No se realiza "*backtracking*".
 - ♦ Utiliza la entropía.

II. CONCEPTOS IMPORTANTES A TENER EN CUENTA EN EL ALGORITMO ID3

Atributos: La selección de atributos debe basarse en el conocimiento acumulado por la experiencia. Los atributos son los factores que influyen en la clasificación o decisión. En este algoritmo cada atributo forma un nodo intermedio en un árbol cuyas hojas o nodos terminales son las clases o decisiones. Dado el conjunto de ejemplos, el ID3 selecciona el atributo que subdivide los ejemplos *de la mejor manera*. La estructura del ID3 es iterativa.



- ♦ **Entropía:** Es la medida de la incertidumbre que hay en un sistema. Es decir, ante una determinada situación, la probabilidad de que ocurra cada uno de los posibles resultados.

La función de entropía más usada es la denominada binaria. Su expresión es con logaritmos base 2:

$$I(p,n) = -(p/(p + n)) * \text{LOG}(p/(p + n)) - (n/(p + n)) * \text{LOG}(n/(p + n))$$

Formula de logaritmo base 2:

$$\text{Log } (p/p+n) / \text{Log } (2)$$

$$\text{Log } (n/p+n) / \text{Log } (2)$$

Formula para sacar el total de la entropía de los atributos:

$$E(A) = ((p_1 + n_1) * I(p_1, n_1) + (p_2 + n_2) * I(p_2, n_2) + \dots + (p_v + n_v) * I(p_v, n_v)) / (p + n)$$

Un ejemplo de la entropía binaria podría ser sacar una bola de color blanco o negro de una bolsa. Si en la bolsa hay 3 bolas blancas y 3 negras el resultado es completamente desconocido, es decir la incertidumbre es máxima, es decir la entropía es 1. Si, yéndonos al otro extremo, en la bolsa hay 6 bolas negras el resultado es conocido de antemano, luego la incertidumbre no existe, y la entropía es 0.

- ♦ **Ganancia:** Es la diferencia entre la entropía de un nodo y la de uno de sus descendientes. En el fondo no es más que una heurística, que como veremos nos servirá para la elección del mejor atributo en cada nodo.

$$\text{Ganancia}(A) = I(p,n) - E(A)$$

Un buen criterio parece ser escoger el atributo que *gana* la mayor información. ID3 examina todos los atributos y escoge el de máxima ganancia, forma la ramificación y usa el mismo proceso recursivamente para formar sub-árboles a partir de los v nodos generados

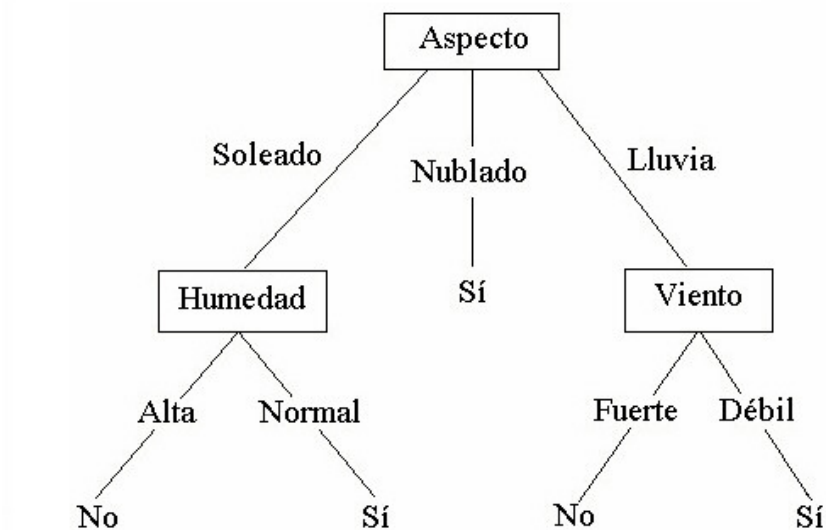
III. ESTRUCTURAS UTILIZADAS

Árbol de decisión es un método para aproximar una función objetivo de valores discretos, que es resistente al ruido en los datos y que es capaz de hallar o aprender una disyunción de expresiones. El resultado puede, de esta manera, expresarse como un conjunto de reglas *Si-entonces*. Por otra parte, los árboles de decisión pueden entenderse como una representación de los procesos involucrados en las tareas de clasificación.

Están formados por:

- ♦ **Nodos:** Nombres o identificadores de los atributos.
- ♦ **Ramas:** Posibles valores del atributo asociado al nodo.
- ♦ **Hojas:** Conjuntos ya clasificados de ejemplos y etiquetados con el nombre de una clase.

Un ejemplo de árbol de decisión es el siguiente:



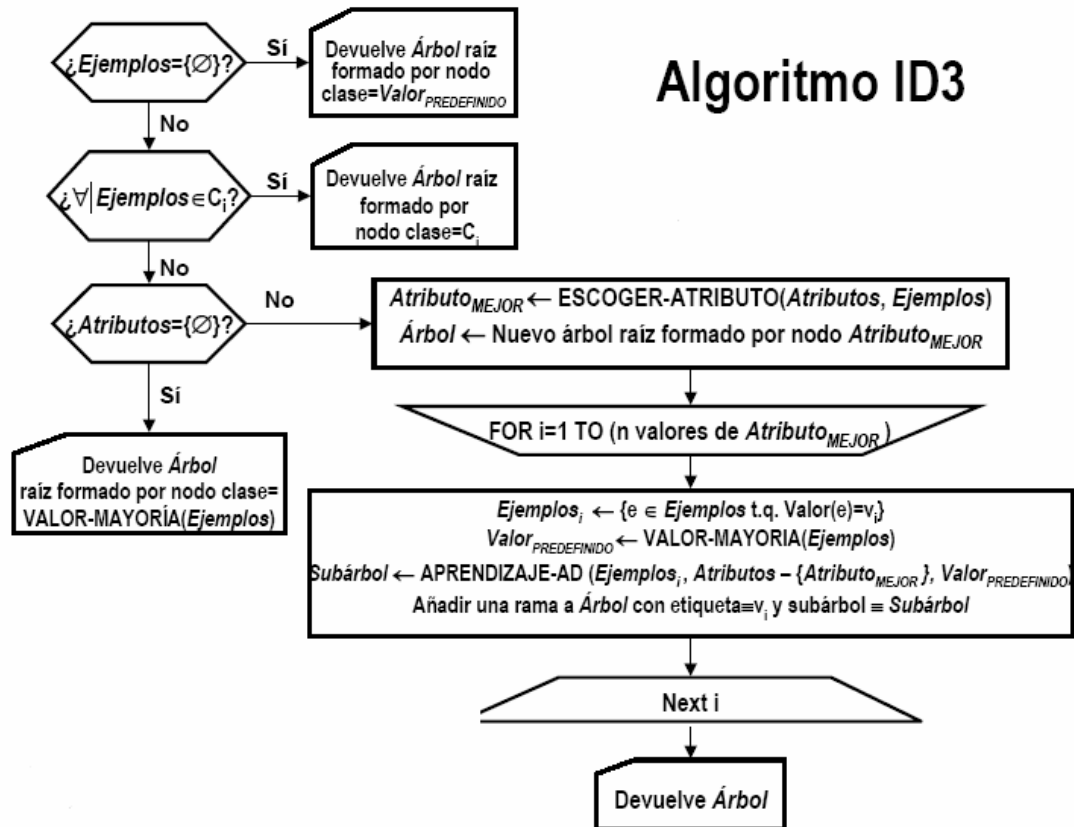
El árbol de decisión se recorre desde la raíz, y tanto en ella como en cada uno de los demás nodos se decide cual rama tomar basándonos en el valor de algún

atributo del ejemplar que se esté clasificando, hasta llegar a un nodo terminal (hoja), que, corresponde a la clase en que queda clasificado el ejemplar.

Los árboles de decisión se adaptan especialmente bien a ciertos tipos de problemas. Básicamente, los casos para los que son apropiados son aquellos en los que:

- Los ejemplos pueden ser descritos como pares valor-atributo.
- La función objetivo toma valores discretos.
- Podemos tomar hipótesis con disyunciones.
- Posible existencia de ruido en el conjunto de entrenamiento.
- Los valores de algunos atributos en los ejemplos del conjunto de entrenamiento pueden ser desconocidos.

IV. DIAGRAMA DE FLUJO



V. EJEMPLO

Caso #	General	Temper.	Humedad	Viento	Clase
1	asoleado	caliente	alta	no	N
2	asoleado	caliente	alta	si	N
3	nublado	caliente	alta	no	P
4	lluvioso	templada	alta	no	P
5	lluvioso	fría	normal	no	P
6	lluvioso	fría	normal	si	N
7	nublado	fría	normal	si	P
8	asoleado	templada	alta	no	N
9	asoleado	fría	normal	no	P
10	lluvioso	templada	normal	no	P
11	asoleado	templada	normal	si	P
12	nublado	templada	alta	si	P
13	nublado	caliente	normal	no	P
14	lluvioso	templada	alta	si	N

Nueve objetos son clase P y cinco son clase N, entonces la información requerida para la clasificación es:

$$I(p,n) = - (9/14) * \text{LOG}_2(9/14) - (5/14) * \text{LOG}_2(5/14) = 0.940 \text{ bits}$$

Considerando el atributo General , con sus tres valores (v=3):

Para el primer valor, hay 5 objetos que lo tienen, 2 clase P y 3 clase N, entonces:

$$p1 = 2, n1 = 3, I(p1,n1) = 0.971$$

Análogamente, para el segundo valor posible de A:

$$p2 = 4, n2 = 0, I(p2,n2) = 0$$

Y para el tercer valor de A:

$$p3 = 3, n3 = 2, I(p3,n3) = 0.971$$

Por lo tanto el requisito de información esperada, después de chequear este atributo es:

$$E(\text{General}) = (5 * I(p1,n1) + 4 * I(p2,n2) + 5 * I(p3,n3)) / 14$$

$$E(\text{General}) = 0.694$$

Y la Ganancia de este atributo es:

$$\text{Ganancia (General)} = 0.940 - E(\text{General}) = 0.246$$

Y el mismo procedimiento aplicado a los otros tres atributos da:

$$\text{Ganancia (Temperatura)} = 0.029$$

$$\text{Ganancia (Humedad)} = 0.151$$

$$\text{Ganancia (Viento)} = 0.048$$

VI. APLICACIONES

GASOIL(1986): Diseño de sistemas de separación gas-petróleo en plataformas petrolíferas marinas de BP. Más de 2.500 reglas, 100 días/persona (10 años/persona).

Ahorró a BP millones de dólares.

BMT (1990): Configuración de equipo de protección de incendios en edificios. Más de 30.000 reglas.

Aprendiendo a volar (1992): En lugar de construir un modelo de la dinámica del sistema (muy complejo), se aprendió un mapeo entre el estado actual y la decisión de control correcta para volar un Cessna en un simulador de vuelo. Resultados: aprendió a volar e incluso mejoraba algunas decisiones de sus “maestros”.

VII. BIBLIOGRAFÍA

<http://www.cs.us.es/~delia/sia/html98-99/pag-alumnos/web2/seudo.html>

<http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t11arboles.pdf>

<http://www.cs.us.es/cursos/ia2/temas/tema-05.pdf>

<http://www.dsic.upv.es/asignaturas/facultad/apr/decision.pdf>

<http://www.dsic.upv.es/asignaturas/facultad/apr/decision.pdf>

http://www.cs.us.es/~delia/sia/html98-99/pag-alumnos/web2/con_teor.html