

Sesión 7-Tema 5: Árboles de decisión y Redes bayesianas

Tema 5: Árboles de precisión y Redes bayesianas- Resumen

Un árbol de decisiones se crea a partir del conocimiento extraído a partir de ejemplos particulares, su estructura se basa en la clasificación de vectores de atributos.

Estos atributos vienen dados por el problema en cuestión, y para establecer el orden de exploración de esos atributos debemos clasificarlos de manera adecuada, para ello, los ordenamos según tengan mayor ganancia de información con respecto al problema a resolver.

A continuación, debemos definir qué es la entropía de esos atributos y lo que conlleva. La entropía es el grado de incertidumbre asociado a una distribución de probabilidad, esa distribución viene dada por los casos que se ponen a prueba en el problema, y gracias a ellos obtenemos datos muy interesantes.

Existen varias maneras de distribuir la entropía entre varios elementos, una de ellas es cuando hay una distribución uniforme, en este caso la entropía es máxima, lo que conlleva a que tenemos máxima incertidumbre.

Lo que esto viene a decir es, si la entropía es máxima, implica que todas las opciones disponibles de la variable tienen la misma probabilidad, por lo que es imposible predecir cual va a resultar cierta.

El caso opuesto es cuando la entropía es mínima, por lo que tenemos una absoluta certeza de cual va a ser el resultado. Esto se traduce a que el 100% (aprox.) de la probabilidad recae sobre una única opción, siempre resultará la misma cierta. Cuando menor es la entropía, más información tenemos del sistema.



La entropía puede estar condicionada por otras entropías, por ejemplo, tenemos la probabilidad de que roben en un banco, y la probabilidad de que salte la alarma. La prob. de robar el banco es independiente, aleatoria, depende de factores externos, pero la prob. de que salte la alarma depende de que roben el banco, a esto se le llama entropía condicionada.

Cuando aparecen las entropías condicionadas, entra en juego un nuevo término, la ganancia. Este valor viene a detallar cuanto ayuda obtener una variable X para conocer el verdadero valor de otra Y.

Si para obtener Y, puede haber distintas variables predecesoras no relacionadas entre sí, cuál es la mejor entre ellas, la que más ganancia tenga, que nos ayude a conocer el valor real de poder obtener esa variable Y, es eso se basa la ganancia, en conocer cuál es la mejor predecesora. Cuanto mayor sea la ganancia de una variable, menor será la incertidumbre de la variable que queremos obtener.

Puede darse el caso de tener valores que no sean realistas con el conjunto de variables, casos que desequilibren la balanza, a veces dados por valores continuos, otras por valores muy elevados o pequeños. Para afrontar esos valores, se forman rangos para reducir el impacto de estos valores en la métrica general del problema.

El algoritmo recursivo ID3 se encarga de calcular la ganancia de las diferentes variables, ordenarlas según esta, filtrar por los rangos dados según el tipo de datos, obtener la entropía de la variable que se busca obtener la máxima información posible con los datos dados.



Para entender que es una Red bayesiana debemos definir y aclarar ciertos conceptos. Para obtener la probabilidad de un evento 'a', debemos conocer la frecuencia con la que se realiza ese evento en casos anteriores.

Si este evento está condicionado por otro llamado 'b', debemos aplicar el proceso de Marginalización, suma de probabilidades de 'a' para todo 'b' disponible, de este modo sabremos con certeza según el estado de 'b', la probabilidad de que ocurra 'a'.

A continuación, la que es la definición de una Red Bayesiana:

Un grafo acíclico dirigido para representar dependencias entre variables y mostrar una descripción escueta de cualquier distribución de probabilidad conjunta completa

Cada nodo de una red es una variable aleatoria con una distribución asignada, la unión de estos determina las dependencias entre los distintos nodos de la red, formando un conjunto de nodos enlazados entre sí con una jerarquía. Su finalidad principal es calcular la distribución conjunta de las variables nodo.

Dado un nodo de la red, podemos afirmar mediante la 'cobertura de Markov', que el nodo es condicionalmente independiente de todos los nodos de la red (Padre, hijos, padres de padre, etc.).



Una vez que sabemos ya todo esto, ¿para qué queremos que en la red haya una distribución conjunta? Pues para poder responder a cualquier pregunta relacionada con la red con mayor o menor precisión, pero siempre con una respuesta, y esto es lo que se denomina la inferencia en una red bayesiana, la capacidad de responder cualquier pregunta con un método u otro.

Existen varios tipos de inferencias en redes bayesianas, la inferencia exacta general, la cual sigue la siguiente fórmula:

$$P(B|C) = \alpha \text{ SUM}(P(B,D,C))$$

El problema de este tipo de inferencia es que puede llegar a ser muy compleja, al final, el árbol de decisiones se divide en 2 ramas cada vez que se depende del estado de 1 variable, por lo que en redes muy grandes puede tener una alta complejidad.

Otro tipo de inferencias son las aproximadas, comentaré 2 de ellas, las directas y las de rechazo. Las directas parten de la red bayesiana, genera el vector de variables, aquellas que desea conocer o llegar hasta un estado concreto, las marca con el estado objetivo, el resto, sino tienen un nodo padre, las genera de manera aleatoria.

Al final devuelve un vector X de sucesos, donde indica que valor deben tener las variables. Para obtener ese resultado, hace una comprobación con una batería de muestras de casos similares, escoge el nº de casos favorables y devuelve la probabilidad aproximada, de este modo evitamos cálculos complejos.

La inferencia por rechazo se basa en el muestreo directo, pero descarta previamente de esa batería de muestras aquellas que generan inconsistencias con el objetivo marcado, es decir, el resultado que queremos obtener. De este modo, si queremos que llueva y el aspersor esté apagado, descartamos aquellas muestras que generen inconsistencia con eso, después calculamos la probabilidad y listo.

[Redes bayesianas](#)

[Índice Wiki individual](#)