Resumen Árboles Decisión Bayesiana

Índice

1 Introducción

En problemas de decisión bayesiana intervienen una tabla de probabilidades y otra de valoraciones, donde E_i son los estados de la naturaleza, X_j son las distintas situaciones-decisiones, A_k son las distintas alternativas entre las que elegir, y $R_{k,i}$ son las valoraciones de las alternativas frente a las decisiones:

P	X_j	
E_i	$P(E_i, X_j)$	$P(E_i)$
	$P(X_j)$	

R	E_i	
A_k	$R_{k,i}$	

Generalmente la información conocida es:

$$P(E_i) \qquad \textbf{y} \qquad P(X_j|Ei)$$

A partir de estos datos podremos calcular las probabilidades conjuntas:

$$P(E_i, X_j) = P(E_i) \cdot P(X_j|E_i)$$

También podremos obtener sumando las columnas de la matriz de probabilidades:

$$P(X_j) = \sum_i P(E_i, X_j)$$

Y también las probabilidades a posteriori:

$$P(E_i|X_j) = \frac{P(E_i, X_j)}{P(X_j)}$$

De donde, para cada A_k y para cada X_j, podremos obtener las valoraciones medias:

$$E(A_k|X_j) = \sum_i R_{k,i} \cdot P(E_i|X_j)$$

2 Análisis a priori

En un primer paso llevamos a cabo un análisis que denominamos a priori, puesto que se efectúa sin ninguna información adicional:

$$A_i \to VME_i = E(A_i) \textbf{ optando por } A_p^* \to mejor_i\, E(A_i)$$

donde A_p^* es la notación que utilizamos para designar a la alternativa óptima a priori y $E(A_p^*)$ representa el resultado a priori que conseguimos con ella.

Denominamos **RER**, es decir, **resultado esperado en riesgo**, a la cantidad monetaria que el decisor espera conseguir sin información adicional, esto es, la cantidad que conseguimos con la alternativa óptima a priori.

$$\boxed{RER = mejor_i\, E(A_i) = E(A_p^*)}$$

3 Análisis a posteriori

En el análisis a posteriori trabajamos con cada una de las concreciones de la información adicional. Para cada concreción de la información X_s determinamos la distribución a posteriori $E(\cdot|X_s)$ y aplicamos el criterio de decisión del valor monetario esperado:

$$A_i \to E(A_i/X_s) \qquad A^*/X_s \to mejor_i E(A_i/X_s)$$

donde A^*/X_s es la notación que utilizamos para designar a la alternativa óptima en la situación X_s y $E(A^*/X_s)$ representa el resultado a posteriori que conseguimos con ella. Siguiendo esta notación, $E(A_p^*/X_s)$ **representa el resultado a posteriori (si se da X_s) que conseguimos con la alternativa óptima a priori**.

Denominaremos valor esperado de la información X_s, denotado $V(X_s) = VEX_s$, al valor que para el decisor tiene la concreción de la información adicional X_s. Se calcula como:

$$V(X_s) = VEX_s = E(A^*/X_s) - E(A_p^*/X_s) \tag{1}$$

y representa la mejora en el resultado óptimo que obtiene el decisor cuando cuenta con la información X_s

4 Análisis pre-a-posteriori

En esta tercera parte realizamos, propiamente, la valoración de la información adicional.

Denominamos **REII** al **resultado esperado con información perfecta**. Es el resultado que el decisor espera conseguir con la información adicional, sea cual sea la concreción de la misma que se presente. Se calcula, por tanto, como:

$$REII = \sum_s E(A^*/X_s)\; P(X_s)$$

El valor esperado de la información X, denotado por $V(X) = VEX$, representa **el valor que para el decisor tiene la información adicional** X, sea cual sea la concreción de la misma que se presente. Se puede determinar de dos maneras distintas, que ofrecen el mismo resultado:

$$V(X) = VEX = E(VEX_s) = \sum_s VEX_s\; P(X_s) \tag{2}$$

o bien

$$V(X) = VEX = REII - RER = \sum_s E(A^*/X_s)\; P(X_s) - E(A_p^*) \tag{3}$$

y **representa la mejora en el resultado óptimo que obtiene el decisor cuando cuenta con la información X cualquiera que sea su concreción**.

A este valor también le denominamos valor otorgado a la regla de decisión que se sigue del análisis pre-a-posteriori.

Denotando por **CIX el coste de la información adicional**:

- Si $CIX \leq VEX$, el decisor adquirirá la información perfecta X.
  Así, la regla de decisión óptima será la que se sigue del análisis pre-a-posteriori, es decir, si se produce lo que establece X_s, la alternativa óptima será A^*/X_s. Y en este caso, el resultado que el decisor espera conseguir será $E(A^*/X_s)$, y el valor otorgado a esa regla de decisión vendrá dado por VEX.
- Si $CIX > VEX$, el decisor no adquirirá la información perfecta X. En este caso, la regla de decisión óptima será la que se sigue del análisis a priori, es decir, sea cual sea la concreción de la información que se presente, la alternativa óptima será A_p^* y el resultado que el decisor espera conseguir vendrá dado por $E(A_p^*) = RER$.

5 Problema de decisión secuencial.

Un problema de decisión bayesiana siempre supone la toma de al menos dos decisiones secuenciales. **La primera** de ellas consiste en saber si el decisor adquirirá o no la información adicional, y **la segunda** se corresponde con la decisión propia del problema.

Por ello, resultado muy útil representar estos problemas a través de un árbol de decisión. **Puesto que los problemas de decisión bayesiana son problemas de decisión secuenciales, la solución de los mismos ya no será la alternativa óptima, sino una regla de decisión. La regla de decisión óptima es la que se deduce del análisis coste-beneficio.**

6 Eficiencia de la información

6.1 Introducción

En el análisis según la metodología bayesiana hemos valorado la información según su capacidad para cambiar la elección a priori, de forma que el decisor siempre le interesa la información cuando su coste no supere el valor otorgado a esa información (análisis coste-beneficio).

Sin embargo, también *podemos intentar cuantificar la información a través del grado de conocimiento que aporta.*

Aún cuando el coste de la información sea superior a su valor puede ocurrir que los resultados reales que se presenten coincidan casi siempre con las previsiones efectuadas por la consultora o entidad asesora que facilita la información. Se trata, entonces, de comparar lo que sabemos sobre la distribución de probabilidad de los estados de la naturaleza antes de tener más información (a priori) y lo que sabemos después (a posteriori). Esto no es otra cosa que valorar la eficiencia de la información. Para valorar dicha eficiencia necesitamos del concepto de **función de entropía**.

6.2 Función de entropía

Dado un fenómeno aleatorio o sistema X cuyas concreciones son $X_1, \ldots, X_i, \ldots, X_n$ y con probabilidades respectivas $P(X = X_i) = p_i \; \forall i = 1, 2, \ldots, n$, tales que $p_i \in [0, 1]$ y $\sum_i p_i = 1$.

Definición 6.1 Definimos la función de entropía del fenómeno aleatorio X como:

$$H(X) = \sum_{i=1}^{n} I_{X_i} P(X_i) = \sum_i -\log P(X_i)\; P(X_i) = -\sum_i P(X_i)\; \log P(X_i)$$

La función de entropía representa la cantidad media de información que aporta la realización de un fenómeno cuando se presenta, dicho de otra forma, mide el grado de desorden del sistema (grado de información).

Proposición 6.1 Las propiedades de la función de entropía son:

1. La entropía de un fenómeno es siempre mayor o igual que cero: $H(X) \geq 0$.
2. La cota mínima para la entropía es cero, es decir, existe un fenómeno cuya entropía es nula (es el fenómeno cierto): $\exists X$ tal que $H(X) = 0$.
3. La cota máxima para la entropía de un fenómeno aleatorio X con n concreciones es $\log n$. Ese valor se alcanza para el fenómeno aleatorio de n concreciones con máximo desorden, es decir, el equiprobable.
   Sea X: $X_1, \ldots, X_i, \ldots, X_n$, entonces $H(X) \leq \log n$.

Además si $p_i = \frac{1}{n}, \forall i = 1,2,\ldots,n$ se tiene que $H(X) = \log n$.

4. Si la base del logaritmo es 2, la unidad de medida para la entropía es el **BIT**, o unidad de desorden del fenómeno aleatorio más sencillo (dos concreciones) con máxima entropía (equiprobable). Si la base del logaritmo es distinto de 2 se utilizan otras unidades de medida.
   Sea X: X_1, X_2 con $p_1 = p_2 = \frac{1}{2}$. Entonces $H(X) = 1$ BIT.
5. La entropía de un sistema aumenta al aumentar su análisis (número de concreciones).
   Sea X: $X_1,\ldots,X_i,\ldots,X_n$ y $P(X = X_i) = p_i, \forall i = 1,2,\ldots,n$ tales que $p_i \in [0,1]$ y $\sum_i p_i = 1$.
   Sea X': $X_1,\ldots,X_j',X_j'',\ldots,X_n$ y $P(X = X_i) = p_i, \forall i \neq j$ y $p_j = p_j' + p_j''$. Entonces, $H(X') \geq H(X)$.

6.3 Función de entropía bidimensional

La función de entropía bidimensional no es otra cosa que la extensión de la definición de la función de entropía a un fenómeno o sistema bidimensional.

Así, sea el fenómeno bidimensional $(E,X) = \{(E_i,X_j),\ p_{ij}\}$, donde $p_{ij} = P(E = E_i, X = X_j)$ con $p_{ij} \in [0,1]$ y $\sum_i \sum_j p_{ij} = 1$.

Designamos por $p_{i\cdot}$ la probabilidad marginal del suceso E_i, es decir, $p_{i\cdot} = P(E = E_i) = \sum_j p_{ij}$ y por $p_{\cdot j}$ la probabilidad marginal del suceso X_j, es decir, $p_{\cdot j} = P(X = X_j) = \sum_i p_{ij}$.

Entropía conjunta, entropías marginales y condicionales

Definición 6.2 Definimos la función de entropía conjunta del fenómeno $(E,X) = \{(E_i,X_j),\ p_{ij}\}$ como

$$\boxed{H(E,X) = -\sum_i \sum_j p_{ij}\ \log p_{ij}} \tag{4}$$

y definimos las entropías marginales de los fenómenos E y X respectivamente como

$$\boxed{H(E) = -\sum_i p_{i\cdot} \log p_{i\cdot}, \qquad H(X) = -\sum_j p_{\cdot j} \log p_{\cdot j}} \tag{5}$$

Además, podemos definir las entropías de un fenómeno condicionadas a que el otro fenómeno se haya manifestado en una concreción determinada, de la forma:

$$\boxed{H(E|X = X_j) = -\sum_i p_{i|j} \log p_{i|j} \text{ donde } p_{i|j} = P(E = E_i | X = X_j) = \frac{p_{ij}}{p_{\cdot j}}} \tag{6}$$

Esta entropía representa la cantidad media de información que aporta E cuando X toma el valor concreto X_j. Será, por tanto, una entropía a posteriori condicionada a un valor concreto de X.

Análogamente, definimos

$$\boxed{H(X|E = E_i) = -\sum_i p_{j|i} \log p_{j|i} \text{ donde } p_{j|i} = P(X = X_j | E = E_i) = \frac{p_{ij}}{p_{i\cdot}}} \tag{7}$$

Para calcular el valor medio de esta entropía cuando se dan las distintas concreciones de los fenómenos, definimos las entropías condicionadas, de la siguiente manera:

Entropías a posteriori

Definición 6.3 Definimos la entropia del E a posteriori a X como:

$$\boxed{H(E|X) = E(H(E|X = X_j)) = \sum_j H(E|X = X_j) p_{\cdot j} = -\sum_i \sum_j (p_{i|j} \log p_{i|j})\ p_{\cdot j}} \tag{8}$$

Esta entropía representa la cantidad media de información que aporta E cualquiera que sea la concreción de X.

Definimos la entropía del X a posteriori a E como:

$$\boxed{H(X|E) = E(H(X|E=E_i)) = \sum_i H(X|E=E_i)p_{i\cdot} = -\sum_i \sum_j (p_{j|i} \log p_{j|i})\ p_{i\cdot}} \tag{9}$$

De forma análoga a la entropía anterior, ésta representa la cantidad media de información que aporta X cualquiera que sea la concreción de E.

Proposición 6.2 Se verifica:

1. $H(E|X) \leq H(E)$ y $H(X|E) \leq H(X)$. La igualdad se da cuando E y X son independientes.
2. $H(E,X) = H(E) + H(X|E)$.
3. $H(E,X) = H(X) + H(E|X)$.
4. $H(E,X) \leq H(X) + H(E)$. Si E y X son independientes se tiene que

$$H(E,X) = H(E) + H(X)$$

6.4 Información debida al canal

Hemos visto que $H(E/X) \leq H(E)$, por lo que la información que aporta X para el conocimiento de E será la diferencia entre entropía a priori y entropía a posteriori. A esta información la denominamos **información debida al canal**. Así definimos:

Definición 6.4 La **información debida al canal** viene dada por la expresión

$$I_c(E,X) = H(E) - H(E|X) \tag{10}$$

- Valor mínimo: $I_c(E,X) = 0$. X no aporta información para el conocimiento de E, luego serán **independientes**.
- Valor máximo: $I_c(E,X) = \log n$. X aporta toda la información necesaria para el conocimiento de E, del que **pasamos a tener conocimiento cierto**.

6.5 Coeficiente de redundancia debida al canal

Aunque el coeficiente de información debida al canal resulta muy útil para determinar la información que la realización de un fenómeno aporta para el conocimiento de otro, resulta inapropiado porque:

1. Ofrece valores absolutos, no relativos, que no permiten la comparación.
2. No es independiente de las unidades de medida, por lo que tiene difícil interpretación.

Para solventar estos problemas, introducimos el siguiente concepto:

Definición 6.5 Definimos el coeficiente de redundancia del fenómeno X para el conocimiento de E como:

$$\boxed{R_c(E,X) = \frac{I_c(E,X)}{H(E)} = \frac{H(E) - H(E|X)}{H(E)} = 1 - \frac{H(E|X)}{H(E)}} \tag{11}$$

Se trata de una medida adimensional, que mide la eficiencia de la información, y $R_c \in [0,1]$, de manera que:

- Si $R_c = 0$, la información aportada por X no es eficiente y el desorden de E permanece igual.
- Si $R_c = 1$, la información aportada por X es completamente eficiente y el desorden de E se reduce a cero.

Por tanto, **el coeficiente de redundancia puede ser interpretado como el porcentaje con el que la información aportada por X reduce el desorden de E.**