

Informativitás számítása

Minden tulajdonságra kiszámítható egy **informatívási jelzőszám**, és ebből számítható az informatívási nyereség. A generáló algoritmus az összes tulajdonság összes lehetséges szétosztásai közül azt választja ki, amelyeknek a legnagyobb az informatívási nyeresége.

A b tulajdonság informatívásának (I_b) kiszámítása a következő:

Legyen a csomópontához tartozó esetek halmaza C , a minősítő tulajdonság a , értékei $a_1 \dots a_n$, és ezek előfordulási arányai a C halmazban $w_{a1} \dots w_{an}$ ($\sum_i w_{ai} = 1$). Ekkor a C halmaz minősítő tulajdonságának entrópiája így írható: $E_C = - \sum_i w_{ai} \log_n w_{ai}$.

Legyenek a b tulajdonság értékei $b_1 \dots b_m$, ezek halmaza β . Bontsuk fel β -t $\beta_1 \dots \beta_m$ nem üres részhalmazokra! Ekkor $\cup_i \beta_i = \beta$.

Bontsuk fel C -t $C_1 \dots C_m$ részhalmazokra úgy, hogy C_i valamennyi elemének b tulajdonsága β_i -be essen minden i -re. Jelölje w_i a C_i súlyát C -ben. ($\sum_i w_i = 1$). Ekkor $I_b = E_C - \sum_i w_i E_{C_i}$, szóval az informatívás a $\beta_1 \dots \beta_m$ felbontásból adódó entrópianyereség. A számítás tényleges kimenete az optimális felbontáshoz tartozó I_{bmax} .

Az ehhez tartozó informatívási nyereség «Density» pedig: $D_b = w_C I_{bmax} / E_C$, ahol w_C a C halmaz elemeinek száma.