

Kodowanie i kompresja danych

Lista nr 1, zadanie 4

Bartosz Chodorowski <166142@student.pwr.wroc.pl>

Wrocław, 1 kwietnia 2009

Zadanie

Niech S oznacza określone źródło symboli, zaś S^k źródło, z którego otrzymujemy bloki k symboli z S , z tym, że poszczególne elementy bloku są generowane niezależnie i każdy element bloku jest generowany zgodnie z prawdopodobieństwem określonym dla S . Pokazać, że $H(S^k) = kH(S)$. Pokazać, że równość ta może być fałszywa, jeśli pominiemy założenie o niezależności.

Rozwiązanie

$$H(S^k) = - \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_k=1}^n P(a_{i_1}, a_{i_2}, \dots, a_{i_k}) \log [P(a_{i_1}, a_{i_2}, \dots, a_{i_k})]$$

Korzystając z założenia o niezależności generowanych bloków otrzymujemy:

$$= - \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_k=1}^n P(a_{i_1})P(a_{i_2}) \cdots P(a_{i_k}) \log [P(a_{i_1})P(a_{i_2}) \cdots P(a_{i_k})]$$

Na podstawie własności logarytmów:

$$= - \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_k=1}^n P(a_{i_1})P(a_{i_2}) \cdots P(a_{i_k}) \sum_{j=1}^k \log [P(a_{i_j})]$$

Grupujemy:

$$\begin{aligned} &= - \sum_{i_1=1}^n P(a_{i_1}) \log [P(a_{i_1})] \left\{ \sum_{i_2=1}^n \cdots \sum_{i_k=1}^n P(a_{i_2}) \cdots P(a_{i_k}) \right\} \\ &\quad - \sum_{i_2=1}^n P(a_{i_2}) \log [P(a_{i_2})] \left\{ \sum_{i_1=1}^n \sum_{i_3=1}^n \cdots \sum_{i_k=1}^n P(a_{i_1})P(a_{i_3}) \cdots P(a_{i_k}) \right\} \\ &\quad \vdots \\ &\quad - \sum_{i_k=1}^n P(a_{i_k}) \log [P(a_{i_k})] \left\{ \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_{k-1}=1}^n P(a_{i_1})P(a_{i_2}) \cdots P(a_{i_{k-1}}) \right\} \end{aligned}$$

Każde z wyreżeń w klamrach sumuje się do jedynki, gdyż sumujemy wszystkie możliwe iloczyny prawdopodobieństw długości $n-1$. Zatem:

$$\begin{aligned}
H(S^k) &= - \sum_{i_1=1}^n P(a_{i_1}) \log [P(a_{i_1})] - \sum_{i_2=1}^n P(a_{i_2}) \log [P(a_{i_2})] - \dots - \sum_{i_k=1}^n P(a_{i_k}) \log [P(a_{i_k})] \\
&= kH(S)
\end{aligned}$$

Aby pokazać, że równość może być fałszywa bez założenia o niezależności, rozpatrzmy następujący przykład:

Niech $S = \{\alpha, \beta\}$ oraz $P(\alpha) = P(\beta) = 0,5$. Rozpatrzmy również dwuelementowe ciągi, w których drugi element będzie wstawiany zależnie od pierwszego (będzie identyczny jak pierwszy). Wówczas:

$$P(\alpha\alpha) = P(\beta\beta) = 0,5$$

$$P(\alpha\beta) = P(\beta\alpha) = 0$$

Licząc $H(S^2)$ oraz $H(S)$ otrzymujemy:

$$H(S^2) = -P(\alpha\alpha) \log [P(\alpha\alpha)] - P(\beta\beta) \log [P(\beta\beta)] - 0 - 0 = 1$$

$$H(S) = -P(\alpha) \log [P(\alpha)] - P(\beta) \log [P(\beta)] = 1$$

Zatem w tym przypadku $H(S^2) \neq 2H(S)$.