ETH Zürich
Dept. Informatik

Prof. Markus Gross
WS 06/07

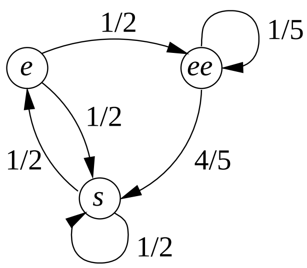
Informationstheorie

Lösung 6

6.1 Informationsquelle

a) Wir haben die drei Zustände "die letzte Prüfung war einfach (und die vorletzte schwierig)", "die beiden letzten Prüfungen waren einfach" und "die letzte Prüfung war schwierig", die wir kurz mit "e", "ee" und "s" bezeichnen.

Das Zustandsübergangsdiagramm ist



und die entsprechende Zustandsübergangmatrix

$$\begin{bmatrix} 0.5 & 0.5 & 0.8 \\ 0.5 & 0 & 0 \\ 0 & 0.5 & 0.2 \end{bmatrix}$$

b) Die stationäre Verteilung kann gemäss des im Unterkapitel 3.6 beschriebenen Verfahrens berechnet werden. Für unsere Quelle erhalten wir

$$P_X(s) = \frac{16}{29} \approx 0.552, \; P_X(e) = \frac{8}{29} \approx 0.276, \; P_X(ee) = \frac{5}{29} \approx 0.172$$

Somit erhalten wir für die Entropierate

$$\begin{aligned} \overline{H}(\{X_i\}) &= \frac{16}{29} \cdot H([0.5, 0.5, 0]) + \frac{8}{29} \cdot H([0.5, 0, 0.5]) + \frac{5}{29} \cdot H([0.8, 0.2, 0]) \\ &= \frac{24}{29} + \frac{5}{29} \cdot h(0.2) \;\approx\; 0.952 \end{aligned}$$

6.2 Codes und Bäume

a) Die Definition für die mittlere Baumtiefe ist

$$t_T \;=\; \sum_{b\in\mathcal{B}} P(b)t(b).$$

Ersetzt man nun ein Blatt b_c der Tiefe $t(b_c)$ durch zwei Blätter $b_{c\|0}$ und $b_{c\|1}$ der Tiefe $t(b_{c\|0}) = t(b_{c\|1}) = t(b_c)+1$, dann erhält man

$$\begin{aligned}
t_{T'} &= \underbrace{\sum_{b\in\mathcal{B}} P(b)t(b)}_{t_T} - \overbrace{P(b_c)t(b_c)}^{\text{entferne } b_c} + \overbrace{P(b_{c\|0})\,\underbrace{t(b_{c\|0})}_{t(b_c)+1}}^{\text{füge } b_{c\|0} \text{ hinzu}} + \overbrace{P(b_{c\|1})\,\underbrace{t(b_{c\|1})}_{t(b_c)+1}}^{\text{füge } b_{c\|1} \text{ hinzu}} \\
&= t_T - P(b_c)t(b_c) + \big(t(b_c)+1\big)\underbrace{\big(P(b_{c\|0}) + P(b_{c\|1})\big)}_{P(b_c)} \\
&= t_T - P(b_c)t(b_c) + (t(b_c)+1)P(b_c) \\
&= t_T + P(b_c)
\end{aligned}$$

b) Wir beweisen dies mittels Induktion:

Aussage: Sei $S(n)$ das Prädikat: „$t_T = \sum_{z\in Z} P(z)$ gilt für alle Bäume T mit n inneren Knoten und allen Wahrscheinlichkeitsverteilungen an den Blättern.“ Wir müssen zeigen, dass $S(n)$ für alle $n > 0$ gilt.

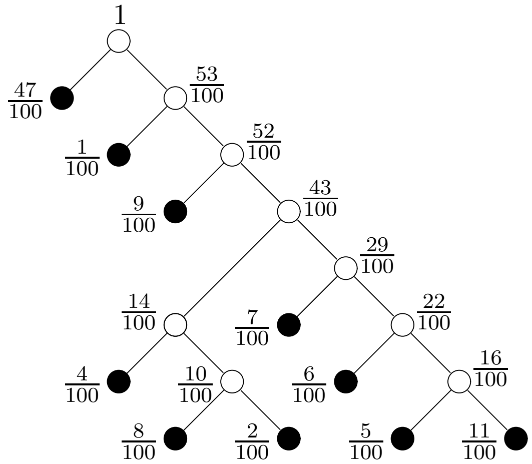
Verankerung $\mathbf{S(1)}$: Es gibt zwei Bäume mit einem inneren Knoten: Eine Wurzel mit nur einem Blatt oder mit zwei Blättern. Es ist leicht zu sehen, das beide Bäume dass Prädikat erfüllen.

Induktionsschritt $\mathbf{S(n) \to S(n+1)}$: Sei T' ein Baum mit $n+1$ innneren Knoten und sei $P_{T'}$ eine Wahrscheinlichkeitsverteilung auf den Blättern von T'. In jedem endlichen ausgefüllten Baum existiert immer ein innerer Knoten z_c, so dass seine beiden Kinder Blätter sind, sagen wir $b_{c\|0}$ und $b_{c\|1}$. (Ist T' nicht ausgefüllt, so nehmen wir einen inneren Knoten, dessen einzelnes Kind ein Blatt ist und verfahren analog.)

Sei T der entsprechende Baum, wenn wir die Blätter $b_{c\|0}$ und $b_{c\|1}$ entfernen (dafür wird z_c zum Blatt b_c). Sei P_T diejenige Wahrschenlichkeitsverteilung auf den Blättern von T, die aus $P_{T'}$ entsteht, wenn man $b_{c\|0}$ und $b_{c\|1}$ entfernt und stattdessen $P_T(b_c) = P_{T'}(b_{c\|0}) + P_{T'}(b_{c\|1})$ setzt. Ferner sei Z_T die Menge der inneren Knoten von T (es gilt $Z_{T'} = Z_T \cup \{b_c\}$). Nun gilt:

$$\begin{aligned}
t_{T'} &= t_T + P_{T'}(z_c) && \text{aus Aufgabe a)} \\
&= \sum_{z\in Z_T} P_T(z) + P_{T'}(z_c) && \text{Induktionsannahme} \\
&= \sum_{z\in Z_T} P_{T'}(z) + P_{T'}(z_c) = \sum_{z\in Z_{T'}} P_{T'}(z)
\end{aligned}$$

c) Von unten nach oben berechnen wir die Wahrscheinlichkeiten $P(z)$ für alle inneren Knoten z und addieren sie auf:



$$\frac{10}{100} + \frac{14}{100} + \frac{16}{100} + \frac{22}{100} + \frac{29}{100} + \frac{43}{100} + \frac{52}{100} + \frac{53}{100} + 1 \; = \frac{339}{100} = 3.39.$$