Datenkompression (Rote), SS 2003, 5. Übungsblatt
Abgabe: bis Freitag, 23. 5. 2003, in der Vorlesung

20. (15 Punkte) Kodierung einer geordneten Auswahl ohne Wiederholung

    Man möchte eine Folge von k verschiedenen Symbolen aus einem vorgegebenen Grundalphabet mit n Symbolen kodieren (zum Beispiel die Blätter eines Huffman-Baumes). Nehmen wir an, dass k und n bekannt sind und nicht mehr kodiert werden müssen.

    (a) (5 Punkte) Überlegen Sie sich einen Algorithmus zur platzsparenden Darstellung dieser Information. Geben Sie näherungsweise die benötigte Anzahl von Bits Ihres Verfahrens an, für große Werte von n, k, und $n - k$.

    (b) (5 Punkte) Aus dem binären Logarithmus für die Anzahl der Möglichkeiten ergibt sich die *informationstheoretische untere Schranke* I: Jedes Kodierungsverfahren für diese Aufgabe benötigt im schlimmsten Fall mindestens I Bits. Wie groß ist die informationstheoretische untere Schranke für große Werte von n, k, und $n-k$, angenähert? (Hinweis: Verwenden Sie die Stirling'sche Formel $n! \approx \left(\frac{n}{e}\right)^n \cdot \sqrt{2\pi n}$.)

    (c) (5 Punkte) Bestimmen Sie die genaue benötigte Anzahl von Bits für (a) und (b) für die Fälle $n = 256$ und $k = 100$, sowie für $n = 10000$ und $k = 50$.

    (d) (0 Punkte) Durch welches Verfahren kann man (im Prinzip) die informationstheoretische untere Schranke exakt erreichen?

21. (15 Punkte) Versuchen Sie, einen längenbeschränkten Kode für die Gewichte p_i zu erzeugen, indem Sie die Gewichte p_i durch $p'_i := \max\{p_i, s\}$ für einen geeignet gewählten Parameter s ersetzen und auf die neuen Gewichte den Huffman-Algorithmus anwenden. Der Parameter s soll durch Probieren ermittelt werden.

    (a) (0 Punkte) $(p_1, \ldots, p_n) = (1, 1, 3, 5, 6, 11, 13)$ und die Längenschranke $L = 3$.

    (b) (5 Punkte) $(p_i) = (1, 1, 3, 5, 6, 11, 13)$ und $L = 4$.

    (c) (5 Punkte) $(p_i) = (1, 3, 4, 7, 11, 18, 29, 47, 76)$ und $L = 5$.

    (d) (0 Punkte) $(p_i) = (1, 1, 2, 3, 5, 8, 13, 21, 34)$ und $L = 4$.

    (e) (0 Punkte) $(p_i) = (1, 1, 1, 2, 3, 15, 18, 130, 200)$ und $L = 4$.

    (f) (5 Punkte) Wie ändert sich die Länge des längsten Kodewortes, wenn man s vergrößert/verkleinert? (Beweisen Sie Ihre Antwort.)

    (g) (0 Punkte) Wenn es mehrere verschiedene Werte von s gibt, die zur gleichen maximalen Kodewortlänge führen, ist es dann günstiger, ein größeres oder eine kleineres s zu wählen? (Diese Frage können Sie auch experimentell untersuchen.)

    (h) Zusatzfrage (15 Punkte). Kann man durch geeignete Wahl von s auf diese Weise immer einen *optimalen* längenbeschränkten Kode erzeugen? Finden Sie einen Beweis oder ein Gegenbeispiel.

    (i) Zusatzfrage (5 Punkte). Kann es passieren, dass man durch keine Wahl von s einen Kode bekommt, bei dem die größte Kodewortlänge genau gleich L ist? Finden Sie einen Beweis oder ein Beispiel.

22. (0 Punkte) Gibt es zu dem Kode aus Aufgabe 3c einen *präfixfreien* Kode über dem Alphabet $\{a, b\}$, bei dem jedes der 16 Quellsymbole durch ein Kodeword mit genau gleich vielen a's und b's dargestellt wird wie im Kode aus Aufgabe 3c: b, ba, ba^7, ba^{13}, ba^{14}; a^3b, a^3ba^2, a^3ba^4, a^3ba^6; a^8b, a^8ba^2, a^8ba^4, a^8ba^6; $a^{11}b$, $a^{11}ba$, $a^{11}ba^2$.