Dr. Jürg M. Stettbacher
Neugutstrasse 54
CH-8600 Dübendorf

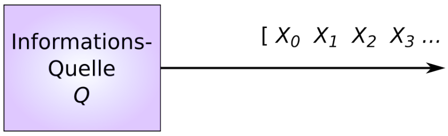
Telefon: +41 43 299 57 23
Email: dsp@stettbacher.ch

Quiz

Quellencodierungstheorem

Sie sollten in der Lage sein, die folgenden Fragen ohne langes Nachdenken beantworten zu können.

Die Fragen beziehen sich auf eine stochastische Quelle Q:



1. Was ist die *Redundanz* R in Worten?

2. Was ist der Unterschied zwischen *Redundanz* und *Irrelevanz*?

3. Was ist die *mittlere Codewortlänge* in Worten?

4. Wie berechnet man die *mittlere Codewortlänge* L?

5. Was bedeutet es, wenn $R < 0$ ist?

6. Wie lautet das Quellencodierungstheorem?

Antworten

Generell: Wir gehen davon aus, dass die Symbole x_n mit $n = 0 \ldots N-1$, die als Resultate der Zufallsvariablen X_k mit $k \geq 0$ aus der Quelle Q kommen, mit einem bestimmten Code C dargestellt sind. Die Codeworte c_n haben die Codewortlängen $\lambda(c_n)$. Die Codeworte können demnach alle gleich lang oder unterschiedlich lang sein. Beispiel:

$$c_1 = (10110) \qquad \longrightarrow \qquad \lambda(c_1) = 5\,\text{Bit}$$

1. Die *Redundanz*[1] ist die durchschnittliche Anzahl Bits in einem Codewort c_n, die keine Information tragen. In der Quellencodierung gilt: Weil sie keine Information tragen, haben sie im Grunde keinen Nutzen. Man würde sie mit Vorteil weglassen.

2. Die *Irrelevanz* bezeichnet die durchschnittliche Anzahl Bits in einem Codewort c_n, die *verzichtbare* Information tragen. Gemeint ist Information, die für den vorgesehenen Verwendungszweck der Daten nicht wichtig ist. Beispiele dafür sind physiologiesche Kompressionsverfahren wie MP3, JPEG, etc. Dabei werden Eigenschaften von Gehör oder Auge berücksichtigt, in dem Sinne, dass unhörbare oder unsichtbare Elemente der Daten entfernt werden.

   Die Redundanz wurde oben beschrieben. Ein Code, resp. ein Codewort kann Redundanz und Irrelevanz enthalten. Bei der verlustlosen Datenkompression geht es um die Redundanz, bei der verlustbehafteten in der Regel und Redundanz und Irrelevanz.

3. Die *mittlere Codewortlänge* ist dies: Liefert eine Quelle Q sehr viele Ereignisse x_n, also $K \to \infty$ Stück, so umfassen alle Codewort zusammen beispielsweise Λ Bit. Dann ist die mittlere Codewortlänge Λ/K Bit.

4. Da in der mittleren Codewortlänge die relative Häufigkeit berücksichtigt wird, verwenden wir den gewichteten Mittelwert:

   $$L(C) = \sum_n P(c_n) \cdot \lambda(c_n)$$

   Dabei ist $P(c_n) = P(x_n)$ die Auftretenswahrscheinlichkeit, resp. die relative Häufigkeit des Symbols x_n und des zugehörigen Codeworts c_n.

5. Wenn die Redundanz $R < 0$ ist, so heisst das, dass die mittlere Codewortlänge des Codes C kürzer ist, als die Entropie der Quelle Q, resp. kürzer als die mittlere Information eines Codeworts c_n. Der Code ist folglich nicht in der Lage, die gesamte Information der Quelle Q zu transportieren. Es geht Information verloren.

6. Quellencodierungstheorem:

   - Falls $R > 0$: Es ist möglich die Daten aus der Quelle Q verlustlos zu komprimieren.
   - Falls $R \leq 0$: Es ist unmöglich die Daten aus der Quelle Q verlustlos zu komprimieren.

[1] Es existieren zwei unterschiedliche Definitionen der Redundanz. $R = L - H$ bezieht sich auf einen gegebenen Code C mit der mittleren Codewortlänge $L(C)$. Für diesen anschaulichen Fall trifft die angegebene Lösung zu. Daneben wird die Redundanz gelegentlich als $R = \log_2(N) - H$ definiert, wobei N die Anzahl Symbole der Quelle ist. Dies ist jedoch eine eher theoretische Grösse und hat keinen Bezug zum Code C.