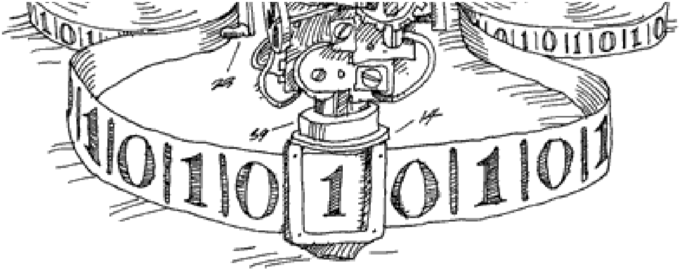
Karlsruher Institut für Technologie

Theoretische Grundlagen der Informatik

Tutorium 12

Institut für Kryptographie und Sicherheit



Best of Übungsblatt 5

- Abgaben: 10 von 21 (47.6%)
- Mindestens 50% Punkte: 9 von 10 (90%)
- Durchschnittliche erreichte Punktzahl: 10.5

Aufgabenverteilung

Aufgabe 1: Gesamt: 35.5P (durchscnittlich 3.55)

Aufgabe 2: Gesamt: 19.5P (durchschnittlich 1.95)

Aufgabe 3: Gesamt: 18.5P (durchschnittlich 1.85)

Aufgabe 4: Gesamt: 31.5P (durchschnittlich 3.15)

Shannonscher Informationsbegriff

- Jede Information bzw. Nachricht besitzt eine Quelle
  - Oft randomisiert a.k.a. Zufallsquellen
  - Wenn alle gesendete Nachrichten unabhängig voneinander sind, ist die Quelle gedächtnislos
- Es gibt immer einen Empfänger, der die Nachrichten beobachtet
- Je unvorhersehbarer die Nachricht, desto mehr Informationsgehalt
  - Wird deshalb auch manchmal Überraschungswert genannt
- Entropie ist ein Begriff für die Dichte der Informationen

Etwas genauer

- Informationsgehalt soll nicht negativ sein
- Ein sicheres Ergebnis (p = 1) enthält keine Information
- Informationen von unabhängigen Nachrichten sollen sich addieren
- Kleine Änderungen der Wahrscheinlichkeit $\Rightarrow$ kleine Änderung des Informationsgehalts
- $I(x) = -log_b(p(x)) = log_b(\frac{1}{p(x)})$ erfüllt diese Bedingungen
  - Meist wird als Basis b = 2 verwendet

- Entropie ist entsprechend definiert
  $H(X) = \sum_{x \in X}(p(x) \cdot log_2(\frac{1}{p(x)})) = \sum_{x \in X}(p(x) \cdot I(x))$

Beispiele

Karlsruher Institut für Technologie

- Zufallsquelle 1: p(A) = $\frac{1}{2}$, p(B) = $\frac{1}{2}$
- I(A) = $log_2(\frac{1}{0.5}) = log_2(2) = 1$ = I(B)
- H(X) = $(p(A) \cdot I(A)) + (p(B) \cdot I(B)) = 0.5 + 0.5 = 1$

- Zufallsquelle 2: p(A) = $\frac{1}{16}$, p(B) = $\frac{15}{16}$
- I(A) = $log_2(\frac{1}{0.0625}) = log_2(16) = 4$
- I(B) = $log_2(\frac{1}{0.9375}) = log_2(\frac{16}{15}) = 0.0931\ldots$
- H(X) = $(p(A) \cdot I(A)) + (p(B) \cdot I(B))$
  $= (\frac{1}{16} \cdot 4) + (\frac{15}{16} \cdot 0.0931) = \frac{1}{4} + 0.873 = 0.337$

Aufgabe B11 A1

1. Wie groß sind der Informationsgehalt und die Entropie, wenn eine Quelle mit dem Alphabet $\{0, 1\}$ nur aus dem Zeichen 0 bestehende Folgen sendet?
2. An einer Quelle mit n Zeichen tritt jedes Zeichen gleichverteilt auf. Wie groß sind der Informationsgehalt und die Entropie eines einzelnen Zeichens?
3. Berechnen Sie die Entropie des Wurfes eines idealen Würfels mit 8 Seiten, dessen Wahrscheinlichkeit für jede Seite $p = \frac{1}{8}$ ist!
4. Was ist der Unterschied zwischen den beiden Folgen, die aus verschiedenen gedächtnislosen Quellen mit der gleichen Wahrscheinlichkeit für 0 und 1 gesendet werden, wenn man sie unter dem Aspekt Entropie und Ordnung betrachtet?
   4.1 ...1010101010101010...
   4.2 ...0110110011011100010...

Übertragung von Information

Daten können über einen Kanal von der Quelle zu einem Empfänger gesendet werden. Dieser Kanal ist in der Regel nicht störungsfrei, dass heißt die gesendeten Daten können ungleich zu den empfangenen sein. Ein gestörter Kanal kann durch seine Übertragungswahrscheinlichkeiten $P(r|q)$, welche angeben wie wahrscheinlich es ist, dass r beim Empfänger ankommt, wenn q aus der Quelle gesendet wurde, charakterisiert werden.
Damit ergibt sich der Zusammenhang

$$P(R = r) = \sum_{q \in Q} P(Q = q) P(R = r | Q = q)$$

und für die Wahrscheinlichkeit das r und q gleichzeitig auftreten:

$$P(Q = q, R = r) = P(Q = q) P(R = r | Q = q).$$

Einige Definitionen

Totalinformation oder auch Verbundentropie H(Quelle Q, Empfänger R) ist die gesamte von Quelle und Empfänger erzeugte Entropie

$$H(Q,R) = -\sum_{q\in Q}\sum_{r\in R} P(Q=q,R=r)\log(P(Q=q,R=r))$$

Äquivokation H(Quelle $Q|$ Empfänger R) gibt dem Entropieverlust durch die Übertragung an.

$$H(Q|R) = H(Q,R) - H(R)$$

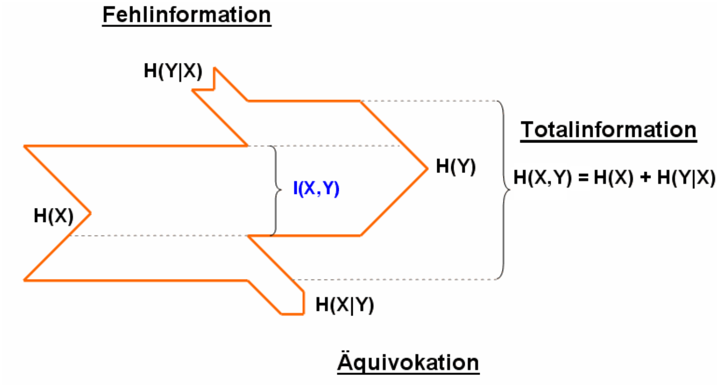
Fehlinformation H(Empfänger $R|$ Quelle Q) entspricht dem anscheinenden Entropiegewinn durch die Übertragung.

$$H(R|Q) = H(Q,R) - H(Q)$$

Transinformation I(Quelle Q, Empfänger R) ist die richtig empfangene Informationsmenge.

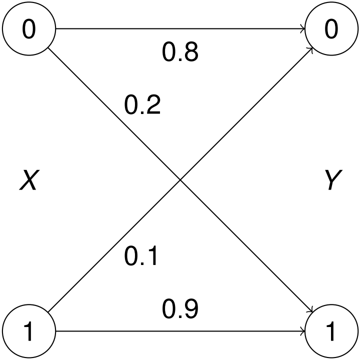
$$I(Q,R) = H(Q) - H(Q|R) = H(R) - H(R|Q) = H(Q) + H(R) - H(Q,R)$$

Bild zur Veranschaulichung

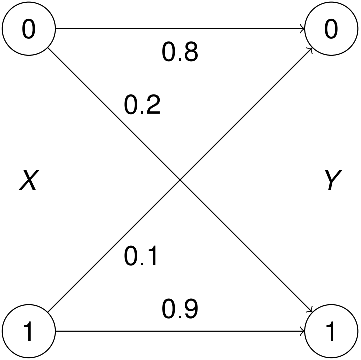


Aufgabe B11 A2

Studieren Sie den Fall eines asymmetrischen binären Kanals mit Quelle X und Empfänger Y. Die Übertragungswahrscheinlichkeiten $P(Y|X)$ seien durch das folgende Diagramm gegeben:

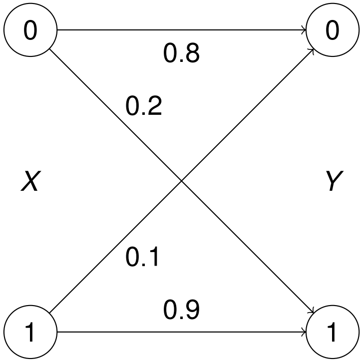


Aufgabe B11 A2



1. Wie groß ist die Wahrscheinlichkeit dafür, dass die Bitkette “1100” als “1001” übertragen wird?
2. Wenn die Entropie der Quelle $H(X) = 1$ bit ist, wie groß ist dann $H(Y)$?
3. Wie groß muss $H(X)$ sein, damit $H(Y) = 1$ bit gilt?

Aufgabe B11 A2



1. Wie groß ist die Verbundentropie $H(X, Y)$ des Übertragungssystems? Gehen Sie ab dieser Teilaufgabe von der Situation der 2. Teilaufgabe aus!
2. Wie groß ist die sog. Irrelevanz $H(Y|X)$? Und wie groß ist die sog. Äquivokation $H(X|Y)$?
3. Wie groß ist schließlich die Transinformation $I(X; Y)$?

Aufgabe B11 A4

Gegeben sei eine gedächtnislose Quelle Q, die mit Wahrscheinlichkeit $p_0 = \frac{1}{4}$ eine 0 und mit Wahrscheinlichkeit $p_1 = \frac{3}{4}$ eine 1 sendet. Gegeben sei zudem ein Empfänger R, der die Zeichen von Q zu empfangen versucht. Dieser Empfänger empfängt eine 0 immer richtig. Sendet die Quelle Q jedoch eine 1, so empfängt R mit Wahrscheinlichkeit $\frac{1}{2}$ eine 1 und mit Wahrscheinlichkeit $\frac{1}{2}$ eine 0.

1. Berechnen Sie die Information $I(0)$ und $I(1)$ bezüglich der Quelle Q!
2. Berechnen Sie die Entropie der Quelle Q!
3. Die Quelle Q sendet die Zeichenfolge 0110. Wie hoch ist der Informationsgehalt dieser Zeichenfolge?
4. Berechnen Sie die Totalinformation $H(Q, R)$, die Fehlinformation $H(R|Q)$, die Äquivokation $H(Q|R)$ und die Transinformation $I(Q; R)$!

Huffman-Codierung

Die Huffman-Codierung ist ein Algorithmus zur verlustfreien Datenkompression.

Problemdefinition

- Gegeben
  - Ein Alphabet $A = \{a_0, a_1, ..., a_n\}$ der Größe n
  - Gewichte $W = \{w_0, w_1, ..., w_n\}$ für alle $a \in A$.
    Meist die Wahrscheinlichkeit, dass ein Zeichen auftritt.
- Gesucht
  - Eine binäre Codierung für alle Zeichen aus A, sodass die erwartete Code-Wortlänge in Bezug auf die Gewichte minimal ist.

Huffman-Codierung

Die Huffman-Codierung ist ein Algorithmus zur verlustfreien Datenkompression.

Problemdefinition

- Gegeben
  - Ein Alphabet $A = \{a_0, a_1, ..., a_n\}$ der Größe n
  - Gewichte $W = \{w_0, w_1, ..., w_n\}$ für alle $a \in A$.
    Meist die Wahrscheinlichkeit, dass ein Zeichen auftritt.
- Gesucht
  - Eine binäre Codierung für alle Zeichen aus A, sodass die erwartete Code-Wortlänge in Bezug auf die Gewichte minimal ist.

Lässt sich sowohl auf konkrete Wörter anwenden als auch auf Quellen, von denen man weiß, wie wahrscheinlich sie welches Zeichen sendet.

Huffman-Codierung Beispiel

Karlsruher Institut für Technologie

Gegeben sei das Wort **abacabadabacaba**. Wie lautet eine Huffman-Codierung?

Huffman-Codierung Beispiel

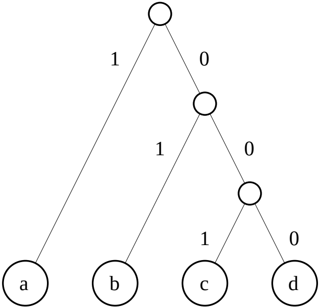
Gegeben sei das Wort **abacabadabacaba**. Wie lautet eine Huffman-Codierung?

- #a = 8
- #b = 4
- #c = 2
- #d = 1

Huffman-Codierung Beispiel

Gegeben sei das Wort **abacabadabacaba**. Wie lautet eine Huffman-Codierung?

- #a = 8
- #b = 4
- #c = 2
- #d = 1



Aufgabe B11 A3

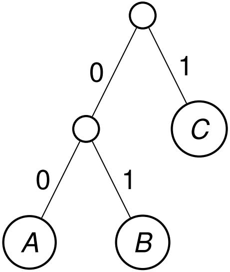
Gegeben sei eine Quelle mit Alphabet $\{A, B, C, D\}$ und mit den folgenden Wahrscheinlichkeiten:
$P(A) = \frac{1}{2}, P(B) = \frac{1}{4}, P(C) = \frac{1}{8}, P(D) = \frac{1}{8}$

- Berechnen Sie die Entropie der Quelle!
- Erstellen Sie eine entsprechende Huffman-Codierung!
- Was ist die mittlere Codewortlänge? Gibt es einen Zusammenhang zur Entropie?

Aufgabe B11 A3

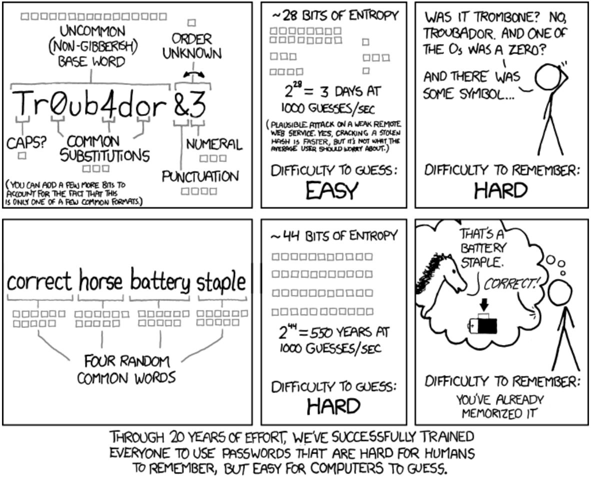
Gegeben sei eine Quelle mit Alphabet $\{A, B, C, D\}$ und mit den folgenden Wahrscheinlichkeiten:
$P(A) = \frac{1}{2}, P(B) = \frac{1}{4}, P(C) = \frac{1}{8}, P(D) = \frac{1}{8}$

- Gegeben sei der folgende Huffman-Baum:



Dekodieren Sie 01101110110010101 1! Ist der Huffman-Code geeignet?

Bis zum nächsten Mal!



Lizenzen