

Übung 10

Tutorübung zu Grundlagen: Rechnernetze und Verteilte Systeme (Gruppen MI-T7 / DO-T5 SS 2015)

Michael Schwarz

Technische Universität München
Fakultät für Informatik

01.07.2015 / 02.07.2015

Datenkompression

Bei Kompressionsverfahren muss unterscheiden werden:

1. Verlustfreie Komprimierung (engl. lossless compression)

- ▶ Komprimierte Daten können verlustfrei, d. h. exakt und ohne Informationsverlust, wiederhergestellt werden.
- ▶ Verlustfrei komprimierte Dateiformate sind beispielsweise ZIP, PNG² (Bilder), FLAC³ (Musik), ...

2. Verlustbehaftete Komprimierung (engl. lossy compression):

- ▶ Komprimierte Daten können im Allgemeinen nicht wieder exakt rekonstruiert werden.
- ▶ Es tritt also ein Verlust von Information bei der Komprimierung auf.
- ▶ Dafür ermöglichen diese Verfahren meist höhere und in Abhängigkeit des Verlustfaktors variable Kompressionsraten.
- ▶ Verlustbehaftet komprimierte Dateiformate sind beispielsweise MP3, MPEG, JPEG, ...

² Portable Network Graphics

³ Free Lossless Audio Codec

Beispiel 1: Huffman-Code

- ▶ Viele Protokolle komprimieren Daten vor dem Senden (Quellenkodierung).
- ▶ TLS beispielsweise bietet optional Kompressionsmethoden. Diese werden vor der Verschlüsselung angewandt. (Warum davor?)
- ▶ Ein häufig (u. a. von TLS) verwendetes Kompressionsverfahren für Texte ist der Huffman-Code.

Grundlegende Idee der Huffman-Kodierung:

- ▶ Nicht alle Textzeichen treten mit derselben Häufigkeit auf, z. B. tritt der Buchstabe „E“ in der deutschen Sprache mit einer Häufigkeit von 17,4 % gefolgt von „N“ mit 9,8 % auf.
- ▶ Anstelle Zeichen mit uniformer Codewortlänge zu kodieren (z. B. ASCII-Code), werden häufigen Zeichen kürzere Codewörter zugewiesen.
- ▶ Die Abbildung zwischen Zeichen und Codewörtern bleibt dabei eindeutig und umkehrbar, weswegen es sich um ein verlustloses Kompressionsverfahren handelt.

Konstruktion eines Huffman-Codes

- ▶ Gegeben Sei das Alphabet $\mathcal{A} = \{A, B, C, D, E, F, G, H\}$ sowie Auftretswahrscheinlichkeiten $\Pr[X = z]$ für alle Zeichen $z \in \mathcal{A}$.
- ▶ Es sei außerdem vorausgesetzt, dass die einzelnen Zeichen unabhängig voneinander auftreten.¹

z	$\Pr[X = z]$
A	0,02
B	0,03
C	0,05
D	0,08
E	0,12
F	0,15
G	0,25
H	0,30

¹ Andernfalls würden die Aussagen zur Optimalität des Huffman-Codes im Allgemeinen nicht mehr zutreffen.

Konstruktion eines Huffman-Codes

- ▶ Gegeben Sei das Alphabet $\mathcal{A} = \{A, B, C, D, E, F, G, H\}$ sowie Auftretswahrscheinlichkeiten $\Pr[X = z]$ für alle Zeichen $z \in \mathcal{A}$.
- ▶ Es sei außerdem vorausgesetzt, dass die einzelnen Zeichen unabhängig voneinander auftreten.¹

z	$\Pr[X = z]$
A	0,02
B	0,03
C	0,05
D	0,08
E	0,12
F	0,15
G	0,25
H	0,30

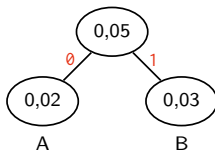


¹ Andernfalls würden die Aussagen zur Optimalität des Huffman-Codes im Allgemeinen nicht mehr zutreffen.

Konstruktion eines Huffman-Codes

- ▶ Gegeben Sei das Alphabet $\mathcal{A} = \{A, B, C, D, E, F, G, H\}$ sowie Auftretswahrscheinlichkeiten $\Pr[X = z]$ für alle Zeichen $z \in \mathcal{A}$.
- ▶ Es sei außerdem vorausgesetzt, dass die einzelnen Zeichen unabhängig voneinander auftreten.¹

z	$\Pr[X = z]$
A	0,02
B	0,03
C	0,05
D	0,08
E	0,12
F	0,15
G	0,25
H	0,30

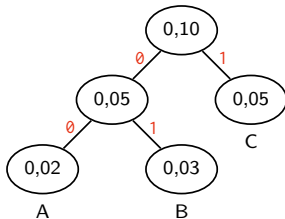


¹ Andernfalls würden die Aussagen zur Optimalität des Huffman-Codes im Allgemeinen nicht mehr zutreffen.

Konstruktion eines Huffman-Codes

- ▶ Gegeben Sei das Alphabet $\mathcal{A} = \{A, B, C, D, E, F, G, H\}$ sowie Auftretswahrscheinlichkeiten $\Pr[X = z]$ für alle Zeichen $z \in \mathcal{A}$.
- ▶ Es sei außerdem vorausgesetzt, dass die einzelnen Zeichen unabhängig voneinander auftreten.¹

z	$\Pr[X = z]$
A	0,02
B	0,03
C	0,05
D	0,08
E	0,12
F	0,15
G	0,25
H	0,30

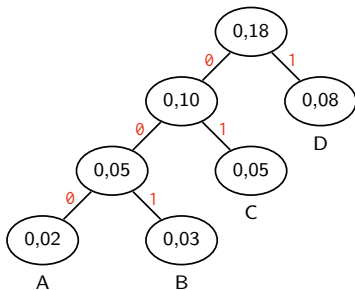


¹ Andernfalls würden die Aussagen zur Optimalität des Huffman-Codes im Allgemeinen nicht mehr zutreffen.

Konstruktion eines Huffman-Codes

- ▶ Gegeben Sei das Alphabet $\mathcal{A} = \{A,B,C,D,E,F,G,H\}$ sowie Auftretswahrscheinlichkeiten $\Pr[X = z]$ für alle Zeichen $z \in \mathcal{A}$.
- ▶ Es sei außerdem vorausgesetzt, dass die einzelnen Zeichen unabhängig voneinander auftreten.¹

z	$\Pr[X = z]$
A	0,02
B	0,03
C	0,05
D	0,08
E	0,12
F	0,15
G	0,25
H	0,30

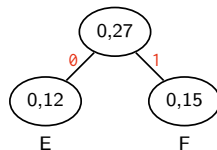
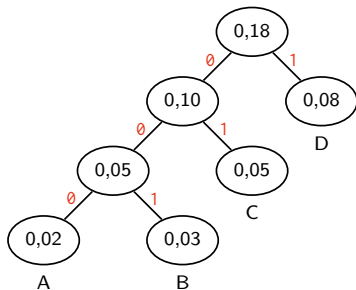


¹ Andernfalls würden die Aussagen zur Optimalität des Huffman-Codes im Allgemeinen nicht mehr zutreffen.

Konstruktion eines Huffman-Codes

- ▶ Gegeben Sei das Alphabet $\mathcal{A} = \{A,B,C,D,E,F,G,H\}$ sowie Auftretswahrscheinlichkeiten $\Pr[X = z]$ für alle Zeichen $z \in \mathcal{A}$.
- ▶ Es sei außerdem vorausgesetzt, dass die einzelnen Zeichen unabhängig voneinander auftreten.¹

z	$\Pr[X = z]$
A	0,02
B	0,03
C	0,05
D	0,08
E	0,12
F	0,15
G	0,25
H	0,30

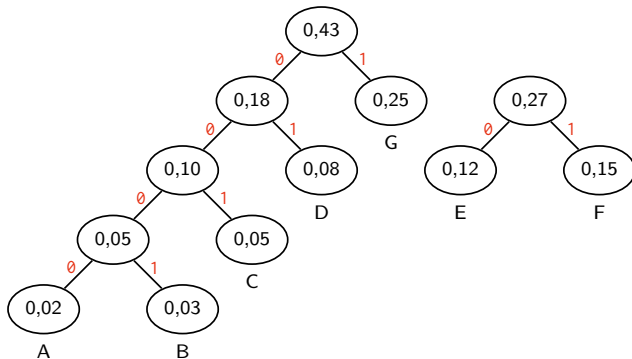


¹ Andernfalls würden die Aussagen zur Optimalität des Huffman-Codes im Allgemeinen nicht mehr zutreffen.

Konstruktion eines Huffman-Codes

- ▶ Gegeben Sei das Alphabet $\mathcal{A} = \{A, B, C, D, E, F, G, H\}$ sowie Auftretswahrscheinlichkeiten $\Pr[X = z]$ für alle Zeichen $z \in \mathcal{A}$.
- ▶ Es sei außerdem vorausgesetzt, dass die einzelnen Zeichen unabhängig voneinander auftreten.¹

z	$\Pr[X = z]$
A	0,02
B	0,03
C	0,05
D	0,08
E	0,12
F	0,15
G	0,25
H	0,30

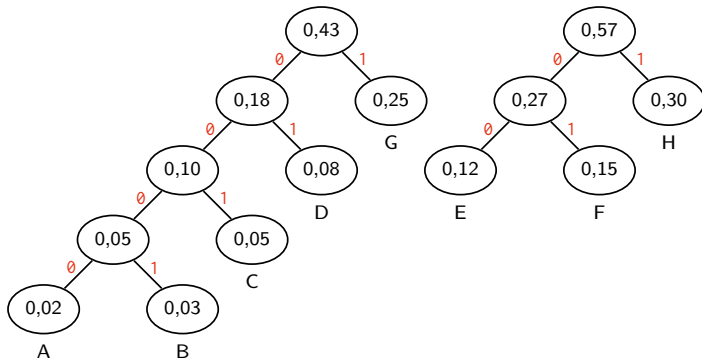


¹ Andernfalls würden die Aussagen zur Optimalität des Huffman-Codes im Allgemeinen nicht mehr zutreffen.

Konstruktion eines Huffman-Codes

- ▶ Gegeben Sei das Alphabet $\mathcal{A} = \{A, B, C, D, E, F, G, H\}$ sowie Auftretswahrscheinlichkeiten $\Pr[X = z]$ für alle Zeichen $z \in \mathcal{A}$.
- ▶ Es sei außerdem vorausgesetzt, dass die einzelnen Zeichen unabhängig voneinander auftreten.¹

z	$\Pr[X = z]$
A	0,02
B	0,03
C	0,05
D	0,08
E	0,12
F	0,15
G	0,25
H	0,30

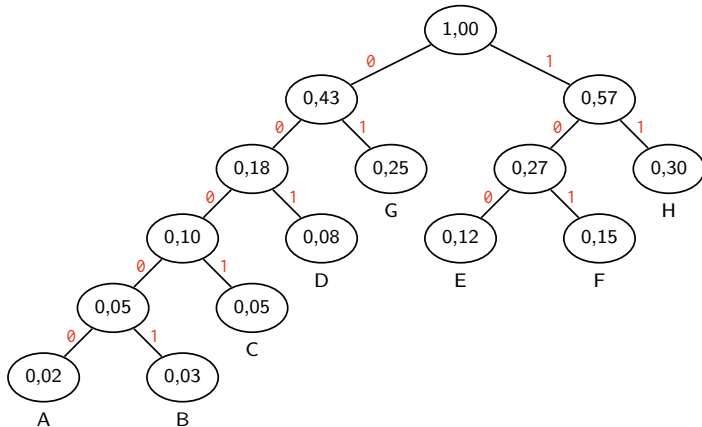


¹ Andernfalls würden die Aussagen zur Optimalität des Huffman-Codes im Allgemeinen nicht mehr zutreffen.

Konstruktion eines Huffman-Codes

- ▶ Gegeben Sei das Alphabet $\mathcal{A} = \{A,B,C,D,E,F,G,H\}$ sowie Auftretswahrscheinlichkeiten $\Pr[X = z]$ für alle Zeichen $z \in \mathcal{A}$.
- ▶ Es sei außerdem vorausgesetzt, dass die einzelnen Zeichen unabhängig voneinander auftreten.¹

z	$\Pr[X = z]$
A	0,02
B	0,03
C	0,05
D	0,08
E	0,12
F	0,15
G	0,25
H	0,30

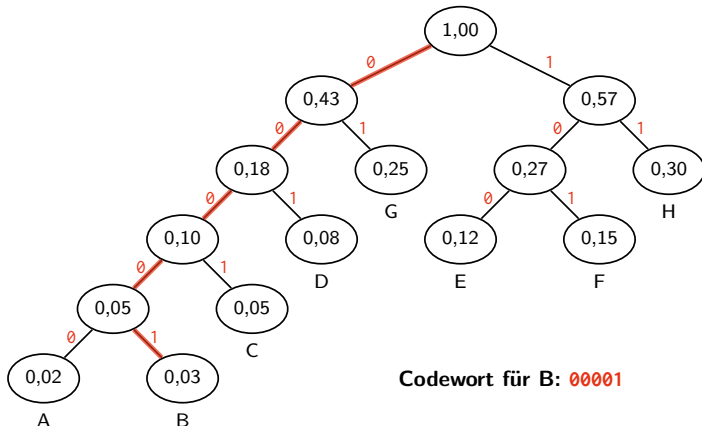


¹ Andernfalls würden die Aussagen zur Optimalität des Huffman-Codes im Allgemeinen nicht mehr zutreffen.

Konstruktion eines Huffman-Codes

- ▶ Gegeben Sei das Alphabet $\mathcal{A} = \{A, B, C, D, E, F, G, H\}$ sowie Auftretswahrscheinlichkeiten $\Pr[X = z]$ für alle Zeichen $z \in \mathcal{A}$.
- ▶ Es sei außerdem vorausgesetzt, dass die einzelnen Zeichen unabhängig voneinander auftreten.¹

z	$\Pr[X = z]$
A	0,02
B	0,03
C	0,05
D	0,08
E	0,12
F	0,15
G	0,25
H	0,30



¹ Andernfalls würden die Aussagen zur Optimalität des Huffman-Codes im Allgemeinen nicht mehr zutreffen.

Durchschnittliche Codewortlänge

z	$\Pr[X = z]$	Huffman-Code	Länge $l_H(z)$	Uniformer Code
A	0,02	00000	5	000
B	0,03	00001	5	001
C	0,05	0001	4	010
D	0,08	001	3	011
E	0,12	010	3	100
F	0,15	011	3	101
G	0,25	10	2	110
H	0,30	11	2	111

► Uniformer Code:

$E[l(z)] = 3,0$, da alle Codewörter gleich lang sind

► Huffman-Code:

$$E[l_H(z)] = \sum_{z \in \mathcal{A}} \Pr[X = z] l_H(z) = 2,6$$

$$\Rightarrow \text{Die Einsparung beträgt } 1 - \frac{E[l_H(z)]}{E[l(z)]} \approx 13\%$$