Informatik IV - Tutorium XVI & XVII

Tut Nr. 8 – Codes, Kanäle

David Münch

Universität Karlsruhe (TH)
Fakultät für Informatik
IBDS Prautzsch

Dank an Yusuf für die Zusammenarbeit.

26. Juni 2009

Universität Karlsruhe (TH)
Forschungsuniversität • gegründet 1825

Inhaltsverzeichnis

1 Auftakt

Inhaltsverzeichnis

1. Auftakt
2. Lernziele

Inhaltsverzeichnis

1 Auftakt

2 Lernziele

3 Themen
   - Hausaufgabenblatt 7
   - Codes
   - Kanäle

Inhaltsverzeichnis

1. Auftakt
2. Lernziele
3. Themen
   - Hausaufgabenblatt 7
   - Codes
   - Kanäle
4. Abspann

Organisatorisches

Email: muenchdavid@gmail.com

https://www.stud.uni-karlsruhe.de/~uhbro/

Tutorium 16: Freitags 8:00 Uhr - Raum -107
Tutorium 17: Freitags 9:45 Uhr - Raum -119

Schein / Übungsblätter

Bearbeitete Hausaufgaben können abgeben werden.
Es gibt vier Scheinklausuren mit je 50 Punkten an folgenden Terminen:

- 19.05.2009 $\surd$ $\varnothing 31$ Punkte
- 09.06.2009 $\surd$ $\varnothing 35$ Punkte
- 02.07.2009 8:00 Uhr $\smile$
- 23.07.2009 8:00 Uhr $\smile$

120 Punkte aus den Scheinklausuren sind notwendig, um einen Schein zu erhalten.

Schein / Übungsblätter

Bearbeitete Hausaufgaben können abgeben werden.
Es gibt vier Scheinklausuren mit je 50 Punkten an folgenden Terminen:

- 19.05.2009 $\surd$ $\varnothing 31$ Punkte
- 09.06.2009 $\surd$ $\varnothing 35$ Punkte
- 02.07.2009 8:00 Uhr $\smile$
- 23.07.2009 8:00 Uhr $\smile$

120 Punkte aus den Scheinklausuren sind notwendig, um einen Schein zu erhalten.

Ab 140 Punkten gibt es einen Notenbonus von ca. 1/3 Note.

Schein / Übungsblätter

Bearbeitete Hausaufgaben können abgeben werden.
Es gibt vier Scheinklausuren mit je 50 Punkten an folgenden Terminen:

- 19.05.2009 $\surd$ $\varnothing 31$ Punkte
- 09.06.2009 $\surd$ $\varnothing 35$ Punkte
- 02.07.2009 8:00 Uhr $\ddot\smile$
- 23.07.2009 8:00 Uhr $\ddot\smile$

120 Punkte aus den Scheinklausuren sind notwendig, um einen Schein zu erhalten.

Ab 140 Punkten gibt es einen Notenbonus von ca. 1/3 Note.

Abgabe in Zweiergruppe erlaubt und ausdrücklich erwünscht!

Schein / Übungsblätter

Bearbeitete Hausaufgaben können abgeben werden.

Es gibt vier Scheinklausuren mit je 50 Punkten an folgenden Terminen:

- 19.05.2009 $\surd$ $\varnothing 31$ Punkte
- 09.06.2009 $\surd$ $\varnothing 35$ Punkte
- 02.07.2009 8:00 Uhr $\smile$
- 23.07.2009 8:00 Uhr $\smile$

120 Punkte aus den Scheinklausuren sind notwendig, um einen Schein zu erhalten.

Ab 140 Punkten gibt es einen Notenbonus von ca. 1/3 Note.

Abgabe in Zweiergruppe erlaubt und ausdrücklich erwünscht!

Bitte folgendes Deckblatt verwenden:

`http://www.stud.uni-karlsruhe.de/~unbdh/deckblatt/`

Literatur

Boehm, Prautzsch: Numerical Methods. AK Peters 1993. ISBN 3-528-06350-5
http://www.ubka.uni-karlsruhe.de/hylib-bin/suche.cgi?opacdb=UBKA_OPAC&fbt=7319953&nd=3204657

Ash: Information Theory. Dover 1990. ISBN 0-486-66521-6
http://www.ubka.uni-karlsruhe.de/hylib-bin/suche.cgi?opacdb=UBKA_OPAC&nd=9866904

Goos: Vorlesungen über Informatik. Bd. 4, Springer 1998. ISBN 3-540-60650-5
http://www.ubka.uni-karlsruhe.de/hylib-bin/suche.cgi?opacdb=UBKA_OPAC&fbt=9316367&nd=6568301

Was wollen wir heute erreichen?

Was wollen wir heute erreichen?

- Hausaufgabenblatt 7 besprechen

Was wollen wir heute erreichen?

- Hausaufgabenblatt 7 besprechen
- verschiedene verlustfreie Codierungsverfahren vorstellen

Was wollen wir heute erreichen?

- Hausaufgabenblatt 7 besprechen
- verschiedene verlustfreie Codierungsverfahren vorstellen
- Wiederholung von Grundbegriffen der Informationstheorie

Was wollen wir heute erreichen?

- Hausaufgabenblatt 7 besprechen
- verschiedene verlustfreie Codierungsverfahren vorstellen
- Wiederholung von Grundbegriffen der Informationstheorie
- diskrete Kanäle

Hausaufgabe 24

Sei das Alphabet $\mathcal{A} = \{a, b, c\}$ mit den unabhängigen Auftrittswahrscheinlichkeiten
$p(a) = 1/2$, $p(b) = 1/3$ und $p(c) = 1/6$
gegeben. Konstruieren Sie einen Präfixcode für $\mathcal{A}^s$ mit $s \in \{1, 2\}$, der die Kraftsche Ungleichung erfüllt und ermitteln Sie die Codelänge L und Entropie H sowie die absolute Redundanz $L - H$.

Hausaufgabe 25

Berechnen Sie die Redundanz $L - H$ eines Huffman-Codes $C : X \to Y^*$ für $X = \{a, b, c\}$ und $Y = \{0, 1\}$ mit den Auftrittswahrscheinlichkeiten $p(a) = p(b) = \beta < 1/2$ und $p(c) = 1 - 2\beta$. Tragen Sie die Redundanz gegen β in einem Diagramm auf.

Hausaufgabe 26

Beweisen Sie, dass bei einem Huffman-Code $C : X \to Y^*$, der die Redundanz Null hat, die Ausgabezeichen $\in Y$ im Mittel exakt gleich oft auftreten werden. Gilt der Umkehrschluss auch?

Burrows-Wheeler-Transformation

Motivation: BWT macht Texte etc. für eine anschliessende Lauflängencodierung geeigneter. Bei einer Lauflängencodierung werden Blöcke beliebiger Länge codiert mit Beachtung der Abhängigkeiten der einzelnen Zeichen.
Die BWT komprimiert noch nicht, sie permutiert die Zeichen.

Burrows-Wheeler-Transformation: $w \to v$

Gegeben: Blöcke w der Länge n: $w = x_1 ... x_n$
Gesucht: Eine Permutation der Indizes $i \in \{1, ..., n\}$

Aufgabe

Führe die BWT auf der Zeichenkette $w = BLUBBLUB$ durch. Gib als Ergebnis alle Informationen an, die die Rücktransformation benötigt um w wieder herzustellen.

Lösung

$$\left[\begin{array}{c|cccccc|c} B & B & L & U & B & B & L & U \\ B & B & L & U & B & B & L & U \\ B & L & U & B & B & L & U & B \\ B & L & U & B & B & L & U & B \\ L & U & B & B & L & U & B & B \\ L & U & B & B & L & U & B & B \\ U & B & B & L & U & B & B & L \\ U & B & B & L & U & B & B & L \end{array}\right] \quad \begin{array}{l} (1) \\ (2) \\ (3) \Leftarrow \text{Original} \\ (4) \Leftarrow \text{Original} \\ (5) \\ (6) \\ (7) \\ (8) \end{array}$$

$L = UUBBBBLL$ und $k = 3$ oder $k = 4$

Move-to-Front

Algorithm 1 MOVE-TO-FRONT KODIERUNG

Schreibe das komplette Alphabet in eine Zeichenkette a.

for *jedes Zeichen* z *der Eingabe* **do**

  Gib die Position von z in a aus.

  Entferne z aus a und füge es vorne wieder an.

end

Move-to-Front

Algorithm 3 MOVE-TO-FRONT KODIERUNG

Schreibe das komplette Alphabet in eine Zeichenkette a.
for *jedes Zeichen* z *der Eingabe* **do**
  Gib die Position von z in a aus.
  Entferne z aus a und füge es vorne wieder an.
end

Algorithm 4 MOVE-TO-FRONT DEKODIERUNG

Schreibe das komplette Alphabet in eine Zeichenkette a.
for *jede Zahl* z *der Eingabe* **do**
  Gib das Zeichen an der Position z von a aus.
  Entferne dieses Zeichen aus a und füge es vorne wieder an.
end

Aufgabe

Kodiere $Mississippi$ mit der Move-to-Front Kodierung.

Aufgabe

Kodiere $Mississippi$ mit der Move-to-Front Kodierung.

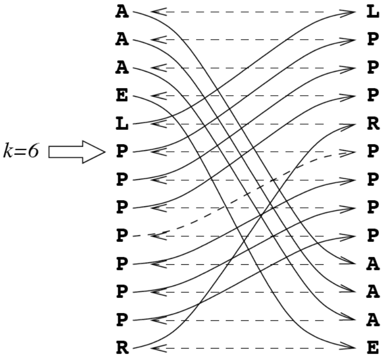
Lösung:
Alphabet: IMPSimps
Ausgabe: 14701101701

Aufgabe

Dekodiere die lauflängencodierte Folge
$(L, 1), (P, 3), (R, 1), (P, 4), (A, 3), (E, 1)$.
Führe für die Ergebniszeichenkette v die inverste BWT mit Startindex $k = 6$ durch und gib das Ergebnis w an.

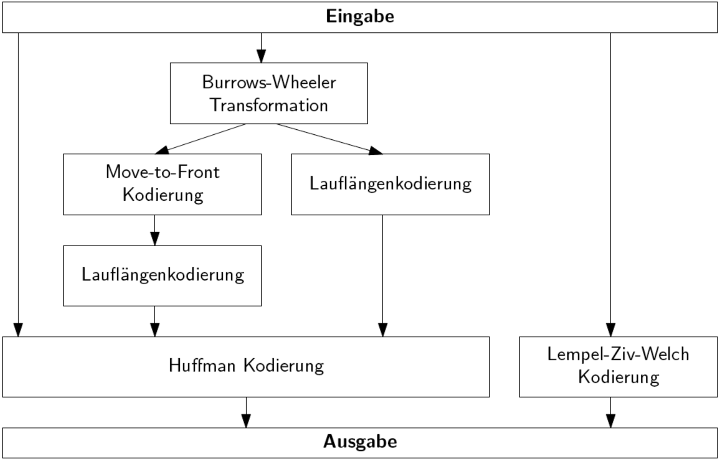
Lösung

$v = LPPPRPPPPAAAE$



$w = PAPPERLAPPAPP$

Verlustfreie Codierungen



Gemeinsame Entropie

Seien X und Y zwei Zufallsvariablen mit den Werten $x_1, ..., x_m$ bzw. $y_1, ..., y_n$.
Sei $p_{ij} = p(x_i y_j) = P(X = x_i$ und $Y = y_j)$.
Wenn x_i und y_i (wie oben) unabhängig sind,
dann gilt: $p_{ij} = p(x_i) \cdot p(y_j)$
sonst $p_{ij} = p(x_i) \cdot p(y_j|x_i) = p(y_j) \cdot p(x_i|y_j)$.

Definition: Gemeinsame Entropie

Die Gemeinsame Entropie von X und Y ist
$H(X, Y) = H(p_{11}, .., p_{mn}) = -\sum_{i,j} p_{ij} \log p_{ij}$

Gemeinsame Entropie

Seien X und Y zwei Zufallsvariablen mit den Werten $x_1, ..., x_m$ bzw. $y_1, ..., y_n$.
Sei $p_{ij} = p(x_i y_j) = P(X = x_i \text{ und } Y = y_j)$.
Wenn x_i und y_i (wie oben) unabhängig sind,
dann gilt: $p_{ij} = p(x_i) \cdot p(y_j)$
sonst $p_{ij} = p(x_i) \cdot p(y_j | x_i) = p(y_j) \cdot p(x_i | y_j)$.

Definition: Gemeinsame Entropie

Die Gemeinsame Entropie von X und Y ist
$H(X, Y) = H(p_{11}, .., p_{mn}) = -\sum_{i,j} p_{ij} \log p_{ij}$

Satz

$H(X_1, ..., X_n) \leq H(X_1) + ... + H(X_n)$, mit Gleichheit $\Leftrightarrow X_i$ unabhängig.

Bedingte Entropie

Definition: Bedingte Entropie

$H(X|Y) = -\sum_{i,j} p_{ij} \cdot \log p(x_i|y_j)$

Bedingte Entropie

Definition: Bedingte Entropie

$H(X|Y) = -\sum_{i,j} p_{ij} \cdot \log p(x_i|y_j)$

Satz

$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$

Information

Definition: Information

Information ist die beseitigte Unsicherheit. D.h. die durch Kenntnis von Y über X erzielbare Information ist:
$I(X|Y) = H(X) - H(X|Y) = I(Y|X) \geq 0$

Kullback-Leibler-Distanz

Definition: Kullback-Leibler-Distanz

Die Kullback-Leiber-Distanz, oder auch relative Entropie genannt, ist folg. definiert:

$$D(p||q) = \sum_i p_i \log \frac{p_i}{q_i}$$

Kullback-Leibler-Distanz

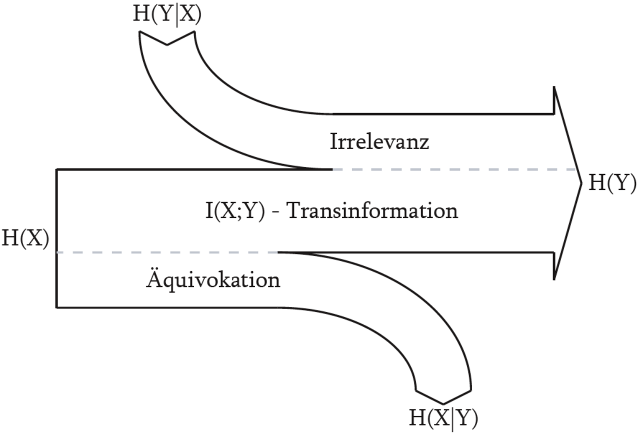
Definition: Kullback-Leibler-Distanz

Die Kullback-Leiber-Distanz, oder auch relative Entropie genannt, ist folg. definiert:

$$D(p||q) = \sum_i p_i \log \frac{p_i}{q_i}$$

Sie gibt aus informationstheoretischer Sicht an, wieviele Bits durchschnittlich verschwendet werden, wenn eine eigentlich auf q basierende Kodierung auf Ereignisse angewendet wird, die p folgen.

Übertragungskanal



Zusammenfassung

Transinformation: $I(X|Y) = H(X) - H(X|Y) = I(Y|X)$

Zusammenfassung

Transinformation: $I(X|Y) = H(X) - H(X|Y) = I(Y|X)$
Äquivokation: $H(X|Y) = H(X) - I(X|Y)$

Zusammenfassung

Transinformation: $I(X|Y) = H(X) - H(X|Y) = I(Y|X)$
Äquivokation: $H(X|Y) = H(X) - I(X|Y)$
Irrelevanz: $H(Y|X) = H(Y) - I(X|Y)$

Zusammenfassung

Transinformation: $I(X|Y) = H(X) - H(X|Y) = I(Y|X)$
Äquivokation: $H(X|Y) = H(X) - I(X|Y)$
Irrelevanz: $H(Y|X) = H(Y) - I(X|Y)$
Totalinformation: $H(Y|X) + I(X|Y) + H(X|Y)$

Kanaleigenschaften

- deterministisch: $H(Y|X) = 0$

Kanaleigenschaften

- deterministisch: $H(Y|X) = 0$
- verlustfrei: $H(X|Y) = 0$

Kanaleigenschaften

- deterministisch: $H(Y|X) = 0$
- verlustfrei: $H(X|Y) = 0$
- störungsfrei: $H(X|Y) = H(Y|X) = 0$

Kanaleigenschaften

- deterministisch: $H(Y|X) = 0$
- verlustfrei: $H(X|Y) = 0$
- störungsfrei: $H(X|Y) = H(Y|X) = 0$
- nutzlos: $I(X|Y) = 0$

Kanaleigenschaften

- deterministisch: $H(Y|X) = 0$
- verlustfrei: $H(X|Y) = 0$
- störungsfrei: $H(X|Y) = H(Y|X) = 0$
- nutzlos: $I(X|Y) = 0$

Die Zufallsvariablen X und Y sind genau dann stochastisch unabhängig, wenn der Kanal nutzlos ist.
$H(X,Y) = H(X) + H(Y)$

Kanalkapazität

Definition: Kanalkapazität

$C = max_{P(X)}\{I(X|Y)\}$

C ist die höchste Informationsmenge, die unter allen möglichen Quellenverteilungen über den Kanal übertragen werden kann.

Aufgabe

Gegeben sei folgender binärer, asymetrischer Kanal über den Alphabeten $\mathcal{X} = \mathcal{Y} = \{0, 1\}$:

$$\begin{array}{cc} P(0|0) = 1 - \beta & P(1|0) = \beta \\ P(0|1) = 0 & P(1|1) = 1 \end{array}$$

a) Für welches β hat der Kanal maximale Kapazität?

Aufgabe

Gegeben sei folgender binärer, asymetrischer Kanal über den Alphabeten $\mathcal{X} = \mathcal{Y} = \{0, 1\}$:

$$\begin{array}{cc} P(0|0) = 1 - \beta & P(1|0) = \beta \\ P(0|1) = 0 & P(1|1) = 1 \end{array}$$

a) Für welches β hat der Kanal maximale Kapazität?

b) Für welches $H(X)$ wird bei gegebener Wahrscheinlichkeitsverteilung β der Wert von $H(Y)$ maximal?

Aufgabe

Gegeben sei folgender binärer, asymetrischer Kanal über den Alphabeten $\mathcal{X} = \mathcal{Y} = \{0, 1\}$:

$$\begin{array}{cc} P(0|0) = 1 - \beta & P(1|0) = \beta \\ P(0|1) = 0 & P(1|1) = 1 \end{array}$$

a) Für welches β hat der Kanal maximale Kapazität?

b) Für welches $H(X)$ wird bei gegebener Wahrscheinlichkeitsverteilung β der Wert von $H(Y)$ maximal?

c) Berechne Irrelevanz, Äquivokation und Transinformation für $\beta = 0.9$ und Gleichverteilung auf X.

Quellen

Pajor - Informatik 4 Tutorium SS2007

Prautzsch - Skript Informatik 4 SS2008

Ash: Information Theory. Dover 1990. ISBN 0-486-66521-6
http://www.ubka.uni-karlsruhe.de/hylib-bin/suche.cgi?opacdb=UBKA_OPAC&nd=9866904

Wikipedia

Reflexion

Was haben wir heute gelernt?

Reflexion

Was haben wir heute gelernt?

- Hausaufgabenblatt 7 besprochen

Reflexion

Was haben wir heute gelernt?

- Hausaufgabenblatt 7 besprochen
- Überblick über ver. verlustfreie Codierungen

Reflexion

Was haben wir heute gelernt?

- Hausaufgabenblatt 7 besprochen
- Überblick über ver. verlustfreie Codierungen
- diskrete Kanäle verstanden

Noch Fragen?

Vorschau

Vorschau

- Fehlerkorrigierende Codes

Bis zum nächsten Mal

Aufgabe

Über dem Alphabet $X = \{A, B, C\}$ sei eine Zeichenkette w gegeben durch die LZW-Codierung

$$0, 1, 2, 3, 5, 4, 6$$

Dabei sind die Wörterbucheinträge anfänglich gegeben durch $0 = A, 1 = B, 2 = C$

a) Wie lang ist w?

Aufgabe

Über dem Alphabet $X = \{A, B, C\}$ sei eine Zeichenkette w gegeben durch die LZW-Codierung

$$0, 1, 2, 3, 5, 4, 6$$

Dabei sind die Wörterbucheinträge anfänglich gegeben durch $0 = A, 1 = B, 2 = C$

a) Wie lang ist w?

b) Um wie viele Bits differiert die Länge der LZW-Codierung von w von der Länge einer Huffman-Codierung von w?

Lösung

Eingabe	Ausgabe	w
0	A	3=AB
1	B	4=BC
2	C	5=CA
3	AB	6=ABC
4	CA	7=CAB
5	BC	8=BCA
6	ABC	

$$w = ABCABCABCABC$$

Lösung

Eingabe	Ausgabe	w
0	A	3=AB
1	B	4=BC
2	C	5=CA
3	AB	6=ABC
4	CA	7=CAB
5	BC	8=BCA
6	ABC	

$w = ABCABCABCABC$
Huffman-Codierung: $code(w) = 4 \cdot (1bit + 2bit + 2bit) = 20bit$
LZW-Codierung: $0..6 \leq 8 = 2^3 \rightarrow 3bit/Zeichen \rightarrow 7 \cdot 3bit = 21bit$
Differenz ist dann $1bit$.