

**Einführungsvortrag zum
Proseminar Datenkompression
im Wintersemester 2005/2006**

Dr. Ralf Schlüter

**Lehrstuhl für Informatik VI
RWTH Aachen
52056 Aachen
schlueter@cs.rwth-aachen.de**

- Einführung**
 - Anwendungsbereiche
 - Motivation
 - Beispiele

- Methodik**
 - Verlustlos vs. verlustbehaftet
 - Performanzbewertung
 - Ansätze
 - Vorgehensweise

- Statistik**
 - Grundbegriffe und Konzepte
 - Stochastische Prozesse

- Information**
 - Definition
 - Modellierung
 - Kodierung

Anwendungen für Datenkompression:

- Internet
- Telekommunikation
- Videokommunikation

Undenkbar ohne Datenkompression:

- Bild-, Audio- und Videoinformation im Internet
- Hochqualitative Mobiltelefonie
- Digital-TV

Warum Datenkompression?

- Ergänzung zu Verbesserungen in Speichertechnologie und Datenübertragung
- Informationsbedarf steigt stärker als verfügbare Ressourcen
- Physikalische Grenzen für Speicher- und Übertragungskapazitäten

Jedoch:

- Komprimierbarkeit ist auch begrenzt (Entropie!)

Morse Kodierung:

- **kürzere** Repräsentationen für **häufigere** Zeichen
- z.B.: SOS ... --- ... (9 bit)
 AND .- -. -.. (7 bit)
 DATE -.. .- - . (7 bit)

Braille Kodierung:

- 6 bit pro Zeichen sowie **häufigste Wörter**
- z.B.: AND ⠠⠠⠠ (6 bit)
 DATE ⠠⠠⠠⠠⠠⠠ (24 bit)
- Anwendung: Blindenschrift

Verlustlose Kompression:

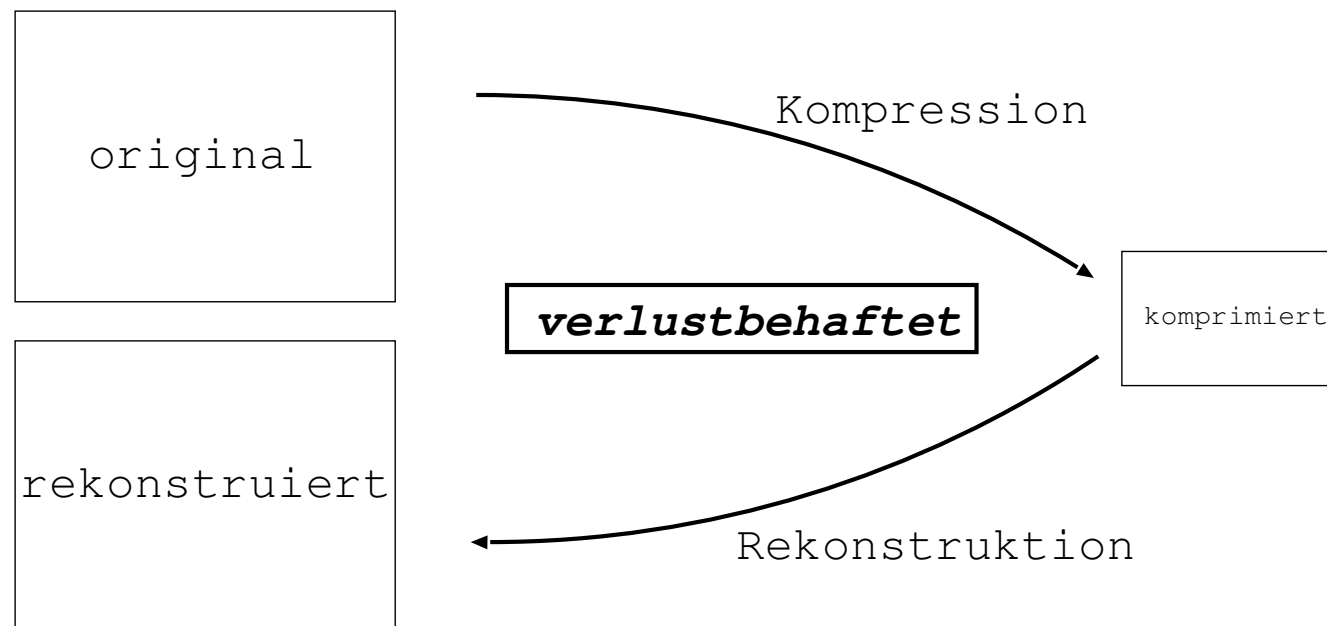
notwendig für z.B.:

- Text-Daten (“Do not jump!” vs. “Do now jump!”)
- System-Daten
- Bank-Daten
- Verhinderung von Artefakten bei Weiterverarbeitung



Verlustbehaftete Kompression:

- höhere Kompressionsraten möglich
- Redundanz der Originaldaten
- Informationsgehalt vs. Perzeptionsgrenzen bzw. Akzeptanz



Performanzbewertung

- Maß der Kompression:
 - Kompressionsrate: $\frac{\text{Anzahl Bits im Original}}{\text{Anzahl Bits nach Kompression}}$
 - Normiert: z.B. für Bilder in Bits pro Pixel
- Verlustbehaftet: zusätzlich Qualitätsmaße
 - Verzerrung: Ähnlichkeit zum Original
 - Sprache, Video: menschliche Perzeption
 - Mathematische Modellierung der menschlichen Perzeption

Ansätze für Kompression:

- **Statistik, z.B.:**
 - Häufigkeit einzelner Symbole (z.B. Huffman, arithmetisch)
 - Häufigkeit von Symbolfolgen/Wörtern (z.B. string-basiert)
 - Berücksichtigung von Kontext (z.B. prädiktiv)

- **Physikalische Strukturierung, z.B.:**
 - Sprache: Vokaltraktparameter statt Abtastwerte

- **Wahrnehmungs-Orientierung, z.B.:**
 - Sprache: Abtastrate angepasst an Verständlichkeit
 - Bilder: Auflösung angepasst an Perzeptionsgrenzen
 - Film: Bildrate angepasst an Fähigkeit, aufeinanderfolgende Bilder explizit zu unterscheiden

Wahl der Kompressionsmethode:

- Finden von Redundanzen
- Modellierung (von Redundanzen), z.B.:
 - Statistisch
 - Gruppierung
 - Prädiktion
 - Funktional
 - Transformation
- Kodierung, z.B.:
 - Statistisch, z.B.:
 - * variable Kodewortlänge
 - * Gruppierung
 - Modellparameter
 - + Abweichung vom Modell (Residuum)

Wahrscheinlichkeit:

- Beschreibung von Ereignissen
- Erwartungsmaß; z.B. Häufigkeit
- Axiomatische Definition: Positivität, Normierung und Additivität
- Unabhängigkeit / Bedingtheit
- Bezeichnungen:
 - a-priori Wahrscheinlichkeit: $p(B)$
 - a-posteriori Wahrscheinlichkeit: $p(A|B)$
 - Verbundverteilung: $p(A, B) = p(A \cup B) = p(A|B) \cdot p(B)$
 - Randverteilung: $p(B) = \sum_A p(A, B)$

Ereignisse:

- mögliche Werte einer Zufallsvariablen, z.B.:
 - Ergebnisse eines Würfelwurfs
 - Zeichen, Wörter oder ganze Sätze
 - Orte eines Meteoriteneinschlags
- diskret oder kontinuierlich
- Gruppierung / Verfeinerung

Wichtige Begriffe und Konzepte:

- **Bayessche Identität:**
$$p(A|B) = \frac{p(A, B)}{p(B)} = \frac{p(A \cup B)}{p(B)}$$
- **kumulative Verteilungsfunktion:**
$$p(x \leq x_0) = \int_{-\infty}^{x_0} p(x) dx$$
- **Erwartungswerte:**
$$E\{f(x)\} = \int_{-\infty}^{\infty} f(x)p(x) dx \quad \text{(kontinuierlich)}$$

$$E\{f(x)\} = \sum_i f(x_i)p(x_i) \quad \text{(diskret)}$$
- **Mittelwert:**
$$\mu = E\{x\}$$
- **Varianz:**
$$\sigma^2 = E\{(x - \mu)^2\} = E\{x^2\} - E\{x\}^2$$

Stochastische Prozesse:

- statistische Modellierung von **Zeitreihen**, z.B.:
 - Niederschlagsmenge
 - Stromverbrauch
 - Radioaktiver Zerfall
 - Sprache
 - Video-Sequenzen

- zeitabhängige Zufallsvariable

- Autokorrelation:

$$R_{xx}(t_1, t_2) = E\{x(t_1) \cdot x(t_2)\}$$

- Stationarität: statt expliziter nur noch relative Zeitabhängigkeit

$$R_{xx}(t_1, t_2) = R_{xx}(t_2 - t_1)$$

Was ist eigentlich Information?

- **Datenmenge:**

- n verschiedene Zeichen z : $i(z) = \log_2 n$ (**Bits pro Zeichen**)
- **Gleichverteilung:** $p(z) = \frac{1}{n} \Rightarrow i(z) = -\log_2 p(z)$
- **Allgemeine Verteilung:** $i(z) = -\log_2 p(z)$
- **C. E. Shannon: “Eigeninformation”**
- **Einheit:**
 - * **bestimmt durch Basis des Logarithmus**
 - * **z.B. Anzahl Bits, Zeichen, Wörter, Seiten, etc.**

- **Informationsgehalt:**

- **Bezug zu Datenmenge?**
- **Intuitiv: minimal mögliche Datenmenge ohne Verluste**
- **Vorsicht: In diesem Sinne enthält eine zufällige Zeichenfolge mehr Information als z.B. eine Seminararbeit gleicher Länge!**

Was ist eigentlich Information?

- Datenmenge:

- n verschiedene Zeichen z : $i(z) = \log_2 n$ (Bits pro Zeichen)
- Gleichverteilung: $p(z) = \frac{1}{n} \Rightarrow i(z) = -\log_2 p(z)$
- Allgemeine Verteilung: $i(z) = -\log_2 p(z)$
- C. E. Shannon: “Eigeninformation”
- Einheit:
 - * bestimmt durch Basis des Logarithmus
 - * z.B. Anzahl Bits, Zeichen, Wörter, Seiten, etc.

- Informationsgehalt:

- Bezug zu Datenmenge?
- Intuitiv: minimal mögliche Datenmenge ohne Verluste
- Vorsicht: In diesem Sinne enthält eine zufällige Zeichenfolge mehr Information als z.B. eine Seminararbeit gleicher Länge!
... d.h.: **Quantität ist nicht gleich Qualität!**

Information von Ereignissen

Ereignisse A und B seien **unabhängig**

- dann gilt $p(A, B) = p(A) \cdot p(B)$
- Information des Verbundereignisses $A \cup B$: $i(A, B) = i(A) + i(B)$
- kein Informationsgewinn durch Gruppierung
- Beispiele:
 - Münzwurf, gleichverteilt: $p(Kopf) = p(Zahl) = 1/2$
 $\Rightarrow i(Kopf) = i(Zahl) = 1 \text{ Bit.}$
 - Münzwurf, nicht gleichverteilt: $p(Kopf) = 7/8, \quad p(Zahl) = 1/8$
 $\Rightarrow i(Kopf) = 0.193 \text{ Bits,} \quad i(Zahl) = 3 \text{ Bits.}$

Information von Ereignissen

Ereignisse A und B seien **abhängig**

- dann gilt $p(A, B) = p(A|B) \cdot p(B)$
- Information des Verbundereignisses $A \cup B$: $i(A, B) = i(A|B) + i(B)$
- Informationsgewinn durch Gruppierung möglich, z.B.:
 - Betrachte Ziffernfolge: $f = 12123333123333123312$
 - Gleiche Kodewortlänge für alle Symbole: $20 \cdot 2 \text{ Bits} = 40 \text{ Bits}$
 - Annahme unabhängiger Einzelsymbole:
 $p(1) = p(2) = 1/4$ und $p(3) = 1/2 \Rightarrow i(f) = 5 \cdot 2 + 5 \cdot 2 + 10 \cdot 1 \text{ Bits} = 30 \text{ Bits}$
 - Abhängigkeit durch Gruppierung zu Blöcken 12 und 33:
 $p(12) = p(33) = 1/2 \Rightarrow i(f) = 5 \cdot 1 + 5 \cdot 1 \text{ Bits} = 10 \text{ Bits}$

Mittlere Information:

- betrachte Zufallsprozess \mathcal{Z} , mögliche Ereignisse: A_i
- Ereignis A_i tritt ein mit Wahrscheinlichkeit $p(A_i)$
- mittlere Information eines Ereignisses dieses Zufallsprozesses:
Erwartungswert des Informationsgehalts

$$\begin{aligned} H(\mathcal{Z}) = E\{i(A)\} &= \sum_i p(A_i) \cdot i(A_i) \\ &= - \sum_i p(A_i) \cdot \log_2 p(A_i) \end{aligned}$$

C. E. Shannon: Entropie H gibt minimale Anzahl von Bits zur verlustlosen Kodierung des Zufallsprozesses an

- Qualität einer Kompressionsmethode: Vergleich mit Entropie
- Vorsicht: Entropie abhängig vom Modell (bzgl. Abhängigkeit)!

Mittlere Information im allgemeinen Fall:

- betrachte stochastischen Prozess \mathcal{S} , der Folge von Ereignissen A_i erzeugt
- Ereignisse seien aus Alphabet $\{A_1, \dots, A_m\}$
- Entropie:

$$H(\mathcal{S}) = - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i_1=1}^m \sum_{i_2=1}^m \cdots \sum_{i_n=1}^m p(A_{i_1}, \dots, A_{i_n}) \cdot \log_2 p(A_{i_1}, \dots, A_{i_n})$$

- Reichweite von Korrelationen/Redundanzen unbekannt:
Betrachtung im Limes unendlich langer Folgen
- Verteilung bzw. Strukturierung der Daten im Allgemeinen nicht (exakt) bekannt
- Notwendigkeit der Modellierung

Herleitung der Entropie

- Informationstheoretische Basis: C. E. Shannon
- Herleitung der mittleren Information allein über Axiome
- Betrachte unabhängige Ereignisse A_i mit Wahrscheinlichkeiten $p_i = p(A_i)$
- Axiome:
 1. Mittlere Information H ist stetige Funktion der Wahrscheinlichkeiten p_i ; kleine Änderungen in den Wahrscheinlichkeiten führen zu kleinen Änderungen in der mittleren Information.
 2. Für gleichverteilte Ereignisse mit $p_i = 1/n$ ist mittlere Information eine monotone Funktion von n , der Anzahl der möglichen Ereignisse.
 3. Konsistenz der mittleren Information unter Gruppierung. Betrachte $A_2 \vee A_3$ als neues Ereignis:

$$H(p_1, p_2, p_3) = H(p_1, p_2 + p_3) + p_1 \cdot H\left(\frac{p_1}{p_1} = 1\right) + (p_2 + p_3) \cdot H\left(\frac{p_2}{p_2 + p_3}, \frac{p_3}{p_2 + p_3}\right)$$

Modellierung:

- **Physikalisch**
 - Wissen über die Strukturierung der Quelle
 - Vorhersage von Werten mittels Modell
 - Kodierung des Residuums: Abweichung vom Modell
- **Statistisch**
 - Steuerung von Kodewortlänge, Gruppierung, etc. anhand der Wahrscheinlichkeiten
 - Beispiele:

- * **Unabhängigkeit:**

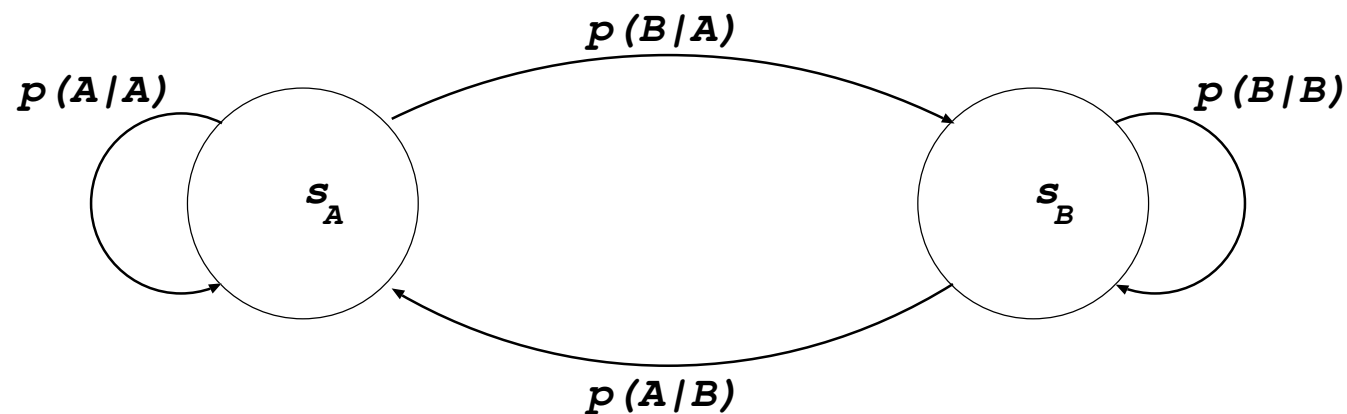
$$p(A_1, \dots, A_n) = \prod_{i=1}^n p(A_i)$$

- * **Markov Annahme: Abhängigkeit endlicher Reichweite m ,**

$$\begin{aligned} p(A_1, \dots, A_n) &= \prod_{i=1}^n p(A_i | A_1, \dots, A_{i-1}) \\ &= \prod_{i=1}^n p(A_i | A_{i-m}, \dots, A_{i-1}) \end{aligned}$$

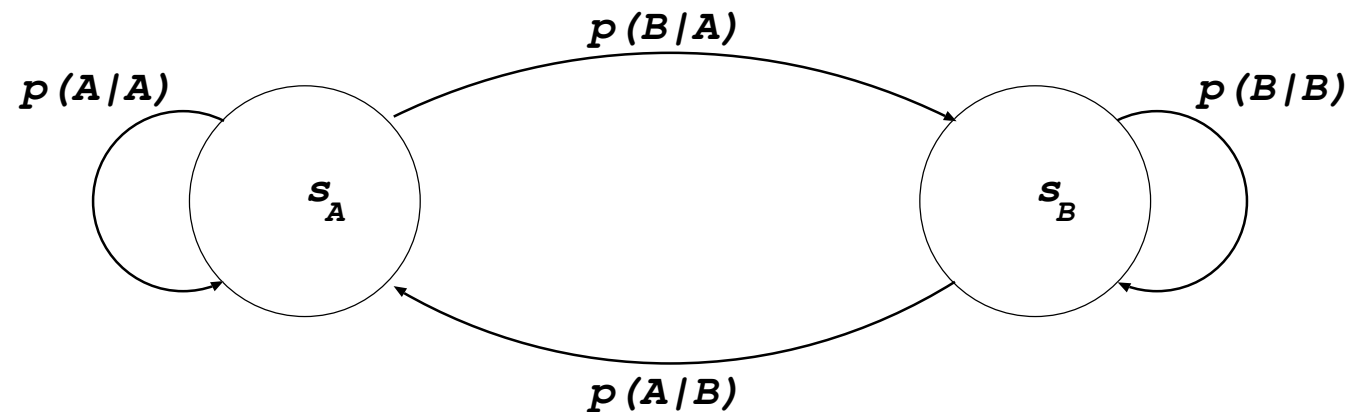
Markov-Prozesse:

- Einfachster nicht-trivialer Fall:
Abhängigkeit allein vom direkt vorhergehenden Ereignis
- $p(A_i | A_1, \dots, A_{i-1}) = p(A_i | A_{i-1})$
- Vgl. stochastischen endlichen Automaten, z.B.:



Markov-Prozesse:

- Vgl. stochastischen endlichen Automaten, z.B.:



- Binärer Zufallsprozess, Ereignisse $X \in \{A, B\}$
 - Zustände des Automaten: s_A, s_B
 - Zustand s_X “emittiert” Ereignis X
 - Übergangswahrscheinlichkeiten: $p(X_i|X_{i-1})$
- Bezüge: Sprachmodell in der Spracherkennung, Hidden Markov Modelle

Kodierung

- Zuweisung binärer Folgen zu Elementen eines Alphabets
- Kode: Menge der binärer Folgen
- Zeichen: Element eines Alphabets
- Kodewörter: Elemente eines Kodes
- Problem: Welche Kodewörter bzw. Kodewortlängen sind den Elementen des Alphabets zuzuordnen, um eine möglichst hohe Kompressionsrate auf den zu erwartenden Datensätzen zu erreichen?

Eindeutige Dekodierbarkeit: Beispiele

Zeichen	Wahrscheinlichkeit	Kode 1	Kode 2	Kode 3	Kode 4
a_1	0.5	0	0	0	0
a_2	0.25	0	1	10	01
a_3	0.125	1	00	110	011
a_4	0.125	10	11	111	0111
mittlere Länge		1.125	1.25	1.75	1.875

- **mittlere Länge:** $\sum_i p(a_i)n(a_i)$, mit **Kodewortlänge** $n(a_i)$
- **Entropie:** 1.75
- **Kode 1:** Identische Kodierung für a_1 und a_2 : **Kode 1 ist nicht eindeutig!**
- **Kode 2:** Dekodierung von 100 liefert a_2a_3 oder $a_2a_1a_1$: **auch Kode 2 ist nicht eindeutig!**
- **Notwendig:** Eindeutige Dekodierbarkeit
- **Kode 3:** Präfix-Kode – Ende des Kodeworts direkt erkennbar!
- **Kode 4:** ähnlich Kode 3, Kodewort-Ende aber erst zu Beginn des Folge-Kodeworts erkennbar!

Eindeutige Dekodierbarkeit: Beispiele (2)

Zeichen	Wahrscheinlichkeit	Kode 5	Kode 6
a_1	0.5	0	0
a_2	0.25	01	01
a_3	0.125	11	10

- **Kode 5: Dekodierung erst am Ende eindeutig**
 - z.B.: Dekodiere 01111111111111111111
 - Beginn mit a_1 mit 8 folgenden a_3 : “baumelndes” Bit am Ende – nicht möglich!
 - Korrekt: a_2 gefolgt von 8 a_3
- **Kode 6: Dekodierung von 010 liefert a_1a_3 oder a_2a_1 : nicht eindeutig!**

Präfix-Kodes

- Definition: Kein Kodewort darf Präfix eines anderen Kodeworts sein.
- Binärbaum-Darstellung:
 - Verzweigung: rechts entspricht 1, links 0
 - mögliche Kodewörter: Knoten
 - Präfix-Kode: Kodewörter nur an Blättern (Endknoten)
- Kraft-McMillan Ungleichung:

Für jeden eindeutig dekodierbaren Code gibt es einen entsprechenden Präfix-Kode mit gleichen Kodewortlängen.
--

Zusammenfassung

Grundlagen

- Motivation
- Anwendungen
- Beispiele

Kompressionskonzepte

- Kompressionsmethoden
- Unterscheidung verlustlos/verlustbehaftet
- Statistische Betrachtungsweise, Grundkonzepte
- Bewertungsmöglichkeiten

Informationstheorie

- Grundkonzepte, Shannon
- Kodierung