Lehrveranstaltung

Informationstheorie

— Sommersemester 2023 —

Martin Mittelbach (Vorlesung, Tutorium), Anne Wolf (Übung, Tutorium)
{martin.mittelbach, anne.wolf}@tu-dresden.de

Professur für Informationstheorie und maschinelles Lernen, TU Dresden

Vorlesung 1
5. April 2023

Team

Dr. Martin Mittelbach
Vorlesung, Skript,
Tutorium

Dr. Anne Wolf
Übung, Tutorium

Inhalt Vorlesung 1

- Organisatorisches

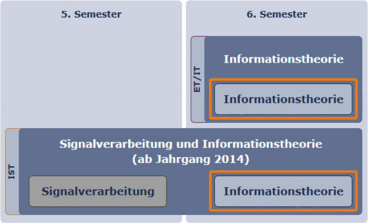
- Einführung, Motivation
- Themenübersicht

- Erste mathematische Grundlagen, Notation
- Online-Umfrage zur Komprimierbarkeit

Organisatorisches

Einordnung

- Einordnung in die Pflichtmodule im Hauptstudium:



Export: Bestandteil des Moduls Nachrichtentechnik und Informationstheorie im Studiengang Wirtschaftsingenieurwesen und Bestandteil des Moduls INF-D-920, Vertiefung im Nebenfach Nachrichtentechnik im Studiengang Informatik.

- Umfang: 2 SWS Vorlesung und 2 SWS Übung / Tutorium
- Vorkenntnisse:
  - Mathematische Grundlagen, insbesondere (elementare) Wahrscheinlichkeitstheorie
  - Nachrichtentechnisches Grundverständnis

Webseite(n) zur Lehrveranstaltung

- Zentrale Webseite:

  https://tu-dresden.de/et/itml/studium/lv/informationstheorie

  (Lehrstuhlwebseite: Studium $\rightarrow$ Lehrveranstaltungen $\rightarrow$ Informationstheorie)

  Hier werden Materialien sowie Informationen (organisatorisch, inhaltlich) zur Lehrveranstaltung bereitgestellt.

- OPAL-Kurs:

  https://bildungsportal.sachsen.de/opal/auth/RepositoryEntry/23048454161

  Bitte schreiben Sie sich in den OPAL-Kurs zur Lehrveranstaltung ein!

  $\Longrightarrow$ Der OPAL-Kurs dient der allgemeinen Kommunikation mit den Studierenden.

- Bitte informieren Sie sich regelmäßig auf diesen Webseiten.
- Bitte rufen Sie regelmäßig das für die OPAL-Einschreibung verwendete E-Mail-Konto ab.

Termine

Vorlesungen: mittwochs, 2. DS, zunächst APB E23 (ggf. später HÜL S386)
Übungen / Tutorien: donnerstags, 2. DS, TOE 317

Woche	Datum	DS	V / Ü / T	Bemerkung / Material
14	05.04.2023	2.	V	
	06.04.2023	2.	Ü	
15	12.04.2023	2.	V	
	13.04.2023	2.	T	
16	19.04.2023	2.	V	
	20.04.2023	2.	V	
17	26.04.2023	2.	V	
	27.04.2023	2.	Ü	
18	03.05.2023	2.	V	
	04.05.2023	2.	T	
19	10.05.2023	2.	–	*dies academicus*
	11.05.2023	2.	Ü	
20	17.05.2023	2.	V	
	18.05.2023	2.	–	*Himmelfahrt*
21	24.05.2023	2.	V	
	25.05.2023	2.	T	
22	31.05.2023	2.	–	*Pfingstferien*
	01.06.2023	2.	–	*Pfingstferien*
23	07.06.2023	2.	V	
	08.06.2023	2.	Ü	
24	14.06.2023	2.	V	
	15.06.2023	2.	T	
25	21.06.2023	2.	V	
	22.06.2023	2.	Ü	
26	28.06.2023	2.	V	
	29.06.2023	2.	T	
27	05.07.2023	2.	V	
	06.07.2023	2.	Ü	
28	12.07.2023	2.	V	
	13.07.2023	2.	T	

Bausteine der Lehrveranstaltung

Stoffvermittlung

- **Vorlesung:** Wöchentlich in Präsenz, systematische Vermittlung und Herleitung der Inhalte der Lehrveranstaltung mit Präsentationsfolien, Tafel, Demos, ...

Anwendung / Vertiefung

- **Übung:** Zweiwöchentlich in Präsenz, Vorführung der Lösung typischer (Rechen-) Übungsaufgaben, Vermittlung von Lösungstechniken

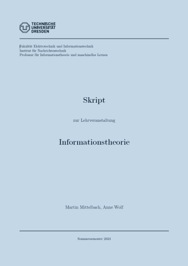
Selbststudium / Intensivierung / Festigung

- **Hausaufgaben:** Eigenständige Bearbeitung von Aufgaben im Selbststudium bis zum nächsten Tutorium, angepasst an vorhergehende Übungen
- **Tutorien:** Zweiwöchentlich in Präsenz im Wechsel mit Übungen, Besprechung von Fragen zu den Hausaufgaben, von Hausaufgabenlösungen und offenen Fragen aller Art zum Stoff von Übung und Vorlesung

Bausteine der Lehrveranstaltung

Wissensspeicher

- **Skript:** Begleitend zur Lehrveranstaltung als Referenz, Formelsammlung, zum Nachschlagen, mit mathematischen Grundlagen



Das Skript können Sie unter folgendem Link herunterladen. [Download]

Eine handliche, gedruckte Ringbuchversion können Sie bei uns erwerben (Selbstkostenpreis: 6.00 €).
Bitte in OPAL-Bestellliste eintragen.

Machen Sie sich frühzeitig mit der Verwendung des Skriptes vertraut!

Abschluss

- **Prüfung:** Schriftlich über 120 Minuten. Zugelassene Hilfsmittel sind druckbare Materialien aus Vorlesung, Übung und Tutorium, eigene handschriftliche Aufzeichnungen, Fachbücher, Taschenrechner, sowie insbesondere das Skript zur Lehrveranstaltung.

Ihr Start in die Lehrveranstaltung

Vorgesehener Ablauf zu Beginn der Lehrveranstaltung:

(1) 1. Vorlesung

$\Longrightarrow$ Mini-Online-Umfrage am Ende der Vorlesung, Abstimmung bis 10.04.2023

(2) Beschäftigung mit Aufgaben der 1. Übung

$\Longrightarrow$ Aufgaben sind online

$\Longrightarrow$ 1. Übung am 06.04.2023, 2. DS

(3) Bearbeitung 1. Hausaufgabe / Selbsttest

$\Longrightarrow$ Bereitstellung Aufgaben nach 1. Übung

$\Longrightarrow$ Selbsttest (siehe Webseite) ist zur eigenständigen Kontrolle des benötigten Vorwissens

$\Longrightarrow$ 1. Tutorium am 13.04.2023, 2. DS

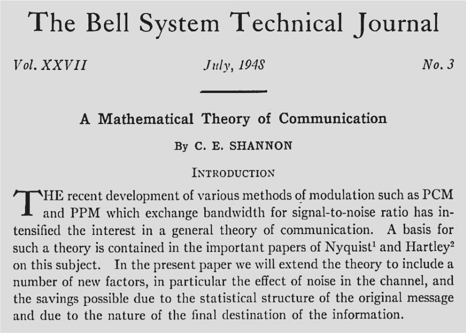
$\Longrightarrow$ **Bitte legen Sie direkt los!**

Kommunikation: Bei Fragen nutzen Sie bitte vorzugsweise die Präsenztermine. Individuelle Fragen, Anregungen und Feedback können Sie gern per E-Mail an uns richten.

Einführung, Motivation, Themenübersicht

Informationstheorie im Sinne Shannons

- In dieser Lehrveranstaltung wollen wir uns mit den Grundlagen der sogenannten Shannonschen Informationstheorie befassen.
- Sie wurde weitestgehend von **Claude Elwood Shannon (1916 – 2001)** mit seiner Arbeit "*A Mathematical Theory of Communication*"



The Bell System Technical Journal

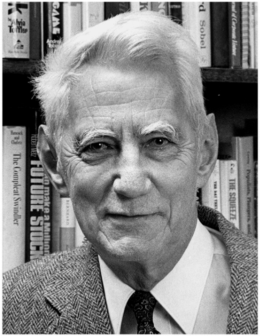
Vol. XXVII *July, 1948* *No. 3*

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

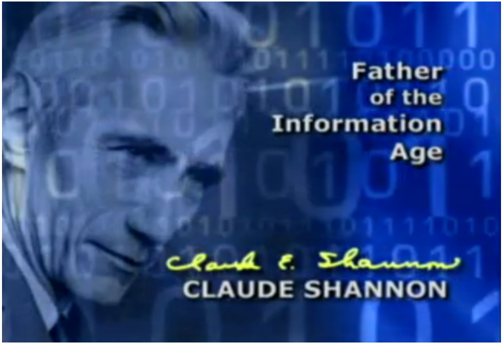


begründet und seitdem von vielen Wissenschaftlern weiterentwickelt.

Zur Bedeutung von Shannons Arbeit

Für interessierte Studierende:

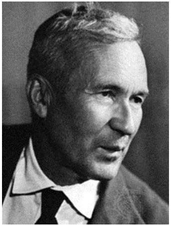
- Video (ca. 30 min) zur Bedeutung Shannons unter folgendem Link: www.youtube.com/watch?v=z2Whj_nL-x8



- Journal-Artikel: R. G. Gallager: Claude E. Shannon: A Retrospective on His Life, Work, and Impact, IEEE Trans. Inf. Theory, vol. 47, no. 7, 2001. [Link]

  Artikel-Download innerhalb des Campusnetzes möglich. Verbindung zum Campusnetz über (Open)VPN-Client.

Kolmogorow über Shannon



Andrei Nikolajewitsch Kolmogorow (1903 – 1987)

Andrei Nikolajewitsch Kolmogorow, der selbst mit einer Reihe herausragender russischer Mathematiker grundlegende Beiträge zur Informationstheorie geleistet hat, sagte folgendes über Shannon:

"In our age, when human knowledge is becoming more and more specialized, Claude Shannon is an exceptional example of a scientist who combines deep abstract mathematical thought with a broad and at the same time very concrete understanding of vital problems of technology. He can be considered equally well as one of the greatest mathematicians and as one of the greatest engineers of the last few decades."

aus: R. G. Gallager: *Claude E. Shannon: A Retrospective on His Life, Work, and Impact*, IEEE Trans. Inf. Theory, vol. 47, no. 7, 2001.

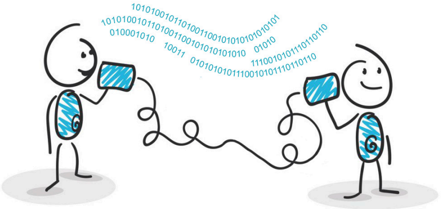
Zentrale Fragen der Shannonschen Informationstheorie

Datenkompression

Wie weit lassen sich Daten durch Umcodieren – verlustlos oder kontrolliert verlustbehaftet – komprimieren?

Datenübertragung

Was sind die theoretischen Grenzen einer zuverlässigen, codierten Datenübertragung, wenn Störungen bei der Übertragung unvermeidbar sind?



Informationsbegriff

- Information ist ein vielfältig verwendeter Begriff (mathematisch, physikalisch, philosophisch, sprachwissenschaftlich, empirisch, . . .).

- In dieser Vorlesung wird ein (nachrichten-) technischer Informationsbegriff verwendet, der nützlich ist für die Analyse effizienter Datenkompression sowie für die Analyse effizienter, zuverlässiger Datenübertragung.
- Dieser Informationsbegriff charakterisiert statistische Aspekte der verwendeten Zeichen in einem Code.
- Es geht stets um Daten und **nicht** um deren Inhalt oder Bedeutung. Daher sollte man eigentlich besser von Shannonscher Datentheorie anstatt von Shannonscher Informationstheorie sprechen.

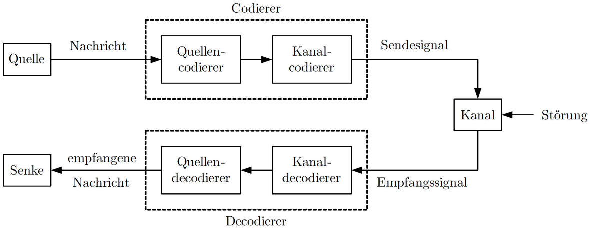
Wesen der Informationstheorie

- Programmatischer Titel von Shannons grundlegender Arbeit:

$$\textit{Eine } \underbrace{\textit{mathematische}}_{\text{Stichwort 1}} \; \underbrace{\textit{Theorie}}_{\text{Stichwort 2}} \textit{ der } \underbrace{\textit{Kommunikation}}_{\text{Stichwort 3}}.$$

  (Shannon ersetzte später "Eine" durch "Die".)

- In der Informationstheorie werden vereinfachte, abstrakte stochastische Modelle von nachrichtentechnischen Systemen bzgl. theoretischer Grenzen untersucht, um allgemeingültige, grundlegende Konzepte zu verstehen und daraus Kriterien für ein leistungsfähiges Systemdesign abzuleiten.
- **Abstraktes Kommunikationsmodell nach Shannon:**



Professur für Informationstheorie und maschinelles Lernen

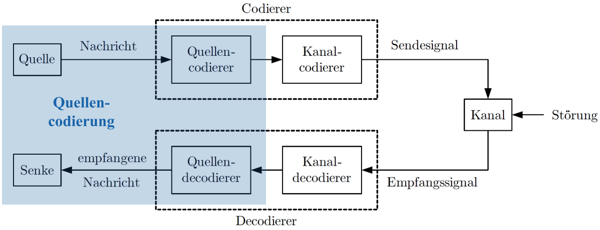
Wesen der Informationstheorie

- Programmatischer Titel von Shannons grundlegender Arbeit:

  Eine $\underbrace{\textbf{\textit{mathematische}}}_{\text{Stichwort 1}}$ $\underbrace{\textbf{\textit{Theorie}}}_{\text{Stichwort 2}}$ ***der*** $\underbrace{\textbf{\textit{Kommunikation}}}_{\text{Stichwort 3}}$.

  (Shannon ersetzte später "Eine" durch "Die".)

- In der Informationstheorie werden vereinfachte, abstrakte stochastische Modelle von nachrichtentechnischen Systemen bzgl. theoretischer Grenzen untersucht, um allgemeingültige, grundlegende Konzepte zu verstehen und daraus Kriterien für ein leistungsfähiges Systemdesign abzuleiten.
- **Abstraktes Kommunikationsmodell nach Shannon:**



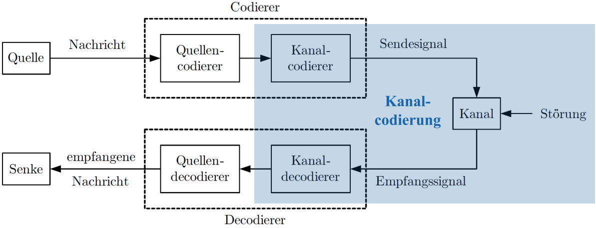
Wesen der Informationstheorie

- Programmatischer Titel von Shannons grundlegender Arbeit:

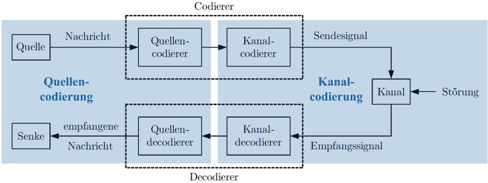
  ***Eine** $\underbrace{\textbf{\textit{mathematische}}}_{\text{Stichwort 1}}$ $\underbrace{\textbf{\textit{Theorie}}}_{\text{Stichwort 2}}$ **der** $\underbrace{\textbf{\textit{Kommunikation}}}_{\text{Stichwort 3}}$.*

  (Shannon ersetzte später "Eine" durch "Die".)

- In der Informationstheorie werden vereinfachte, abstrakte stochastische Modelle von nachrichtentechnischen Systemen bzgl. theoretischer Grenzen untersucht, um allgemeingültige, grundlegende Konzepte zu verstehen und daraus Kriterien für ein leistungsfähiges Systemdesign abzuleiten.
- **Abstraktes Kommunikationsmodell nach Shannon:**



Abstraktes Kommunikationsmodell



Quelle =
(stochastisches) Datenmodell

Kanal =
(stochastisches) Übertragungsmodell

Codierer-Decodierer-Paar =
(deterministisches) Verfahren der Datenrepräsentation und -rekonstruktion

Quellencodierung =
Repräsentation der Daten für eine effiziente, platzoptimierte Speicherung

Kanalcodierung =
Repräsentation der Daten für eine effiziente, störungsrobuste Übertragung

Ziel ist die systematische Analyse und der Entwurf geeigneter Codier-Decodier-Verfahren, die bzgl. gewisser Leistungsparameter (bspw. Kompressionsrate, Datenübertragungsrate, Decodierfehlerwahrscheinlichkeit) optimal sind.

Anwendungsszenario: Homeoffice und Online-Material

Bereitstellung von Lehre-Material

Beispiel 1: Präsentationsfolien (enthalten Textdaten, Bilddaten, Farb- und Formatierungsinformationen, . . .)

(1) **Speicherung** als PDF-Datei auf Heim-PC des Dozenten

(2) **Übertragung** der PDF-Datei vom Heim-PC des Dozenten zum Server der TU Dresden (Upload)

(3) **Speicherung** der PDF-Datei auf Server der TU Dresden

(4) **Übertragung** der PDF-Datei vom Server der TU Dresden zum PC / Mobilgerät des Studierenden (Download)

(5) **Speicherung** der PDF-Datei auf PC / Mobilgerät des Studierenden

(vereinfachte Darstellung: weitere Speicher- und Übertragungsvorgänge u. a. zwischen PC-Komponenten, Servern, Netzknoten, Basisstationen, . . .)

Speicherung $\longrightarrow$ möglichst effizient und platzsparend
Übertragung $\longrightarrow$ möglichst effizient und ohne Fehler

Beispiel 2: Vorlesungs-Videos (enthalten Videodaten, Audiodaten, . . .)
$\Longrightarrow$ analoges Szenario

Relevanz der Informationstheorie

Einige wenige praktische Anwendungen / Implementierungen:

(aus der Nachrichtentechnik / Informatik, weitere Anwendung u. a. auch in Wirtschaft, Statistik, Physik, Neurowissenschaften, . . .)

- Komprimierungsalgorithmen wie bspw. Deflate nutzen Methoden und Resultate aus der Informationstheorie. Ebenso Formate wie MP3, JPG, MPEG,. . .

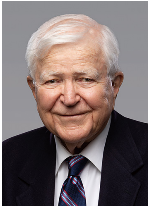
  (Verwendung Deflate u. a. in Packsoftware (ZIP), Protokollen (HTTP, PPP), Bilddateiformaten (PNG), Dokumentdateiformaten (PDF))

- Fehlerschutzcodes nahe am theoretischen Optimum wie bspw. LDPC-Codes kommen bei modernster Datenspeichertechnik und drahtloser / drahtgebundener Datenübertragungtechnik zum Einsatz und haben ebenfalls informations- bzw. codierungstheoretischen Ursprung.

  (Verwendung LDPC-Codes u. a. in digitalem Satellitenfernsehen (DVB-S2), Gigabit Ethernet (IEEE802.3an), drahtlosem Breitbandinternet (WiMAX IEEE802.16e), Mobilfunknetzen (5G Standard), HDD- und SSD-Speichermedien)

Relevanz der Informationstheorie

Japan-Preis 2020 ($\sim$ 50 Mio. Yen) für Robert G. Gallager



Robert G. Gallager (*1931)

[Link Japan-Preis Wikipdia]

Aus der Begründung der Japan Prize Foundation [Link zur Begründung]:

"... for his contributions to information theory, including the development of error-correcting codes that are used in 5G systems ... Gallager invented low-density parity-check (LDPC) codes, a coding scheme that can achieve coding efficiency very close to its theoretical limit ... led to the realization of today's highly reliable high-speed and large-capacity communication ... his ideas [published in the 1960s] were not adopted for the next 30 years ... "

$\Longrightarrow$ Die Beschäftigung mit den informationstheoretischen Grundlagen lohnt sich!

$\Longrightarrow$ Mitunter dauert die praktische Umsetzung theoretischer Erkenntnisse eine Weile!

Vorgehensweise / Themenübersicht

(1) Praktische Problemstellung

systematischer Entwurf effizienter Datenkompression	systematischer Entwurf effizienter zuverlässiger Datenübertragung

(2) Abstraktion / Mathematische Modellbildung

stochastisches Datenmodell Quelle = Folge von Zufallsgrößen **deterministisches Kompressionsmodell** Quellen-Codier-Decodier-Schemata **operationelle Bewertungskenngrößen** (minimale) mittlere Codewortlänge / Coderate, Distortionmaß, Rate-Distortion-Funktion	**stochastisches Übertragungsmodell** Kanal = bedingte Wahrscheinlichkeitsverteilung **deterministisches Kommunikationsmodell** Kanal-Codier-Decodier-Schemata **operationelle Bewertungskenngrößen** (maximale) Coderate, Decodierfehlerwahrscheinlichkeit, Codierkapazität

(3) Entwicklung mathematischer Werkzeuge für Modellanalyse

Informationsmaße

Entropie (bedingte, relative, differentielle, empirische ∼), Transinformation (bedingte, empirische ∼)

Eigenschaften, Rechenregeln

(Symmetrie, Nichtnegativität), Kettenregeln, Konvexität / Konkavität

Unabhängigkeit / (Gleichverteilung (Normalverteilung)) maximiert Entropie (differentielle ∼)

Ungleichungen, grundlegende Theoreme

Kraft-McMillan-Ungleichung, fundamentale Ungleichung, Datenverarbeitungsungleichung, Fano-Ungleichung

Shannon-McMillan-Theorem, Feinstein-Lemma

Informationsmaßebasierte Kenngrößen

Redundanz, Entropierate, Informations-Rate-Distortion-Funktion, Informationskapazität

• • •

Vorgehensweise / Themenübersicht

...

(4) Herleitung fundamentaler Resultate für abstrahiertes Modell mit math. Werkzeugen

Quellencodierungstheoreme

Aussagen über theoretische Grenzen effizienter Datenkompression (bei definierter Güte)

Sinngemäß: Verlustlose (asymptotisch $\sim$) Datenkompression ist begrenzt durch Entropie(-rate)

(Entropie(-rate) $\leq$ min. mittl. Cwl (pro Sym.) / Coderate)

Sinngemäß: Verlustbehaftete Datenkompression ist begrenzt durch Informations-Rate-Distortion-Funktion

(Rate-Distortion-Funktion $=$ Informations-$\sim$)

Kanalcodierungstheoreme

Aussagen über theoretische Grenzen effizienter Datenübertragung bei definierter Zuverlässigkeit

Sinngemäß: Maximale Coderate bei zuverlässiger Datenübertragung ist begrenzt durch Informationskapazität

(Codierkapazität $=$ Informationskapazität)

(5) Anwendung der Resultate auf praktische Problemstellung

Entwurfskriterien für optimale Datenkompression / -übertragung, Systembewertung

Vorgehensweise / Themenübersicht

(1) Praktische Problemstellung

systematischer Entwurf
effizienter Datenkompression

systematischer Entwurf
effizienter zuverlässiger Datenübertragung

(2) Abstraktion / Mathematische Modellbildung

stochastisches Datenmodell
Quelle = Folge von Zufallsgrößen
deterministisches Kompressionsmodell
Quellen-Codier-Decodier-Schemata
operationelle Bewertungskenngrößen
(minimale) mittlere Codewortlänge / Coderate, Distortionmaß, Rate-Distortion-Funktion

stochastisches Übertragungsmodell
Kanal = bedingte Wahrscheinlichkeitsverteilung
deterministisches Kommunikationsmodell
Kanal-Codier-Decodier-Schemata
operationelle Bewertungskenngrößen
(maximale) Coderate, Decodierfehlerwahrscheinlichkeit, Codierkapazität

(3) Entwicklung mathematischer Werkzeuge für Modellanalyse

Informationsmaße
Entropie (bedingte, relative, differentielle, empirische ∼), Transinformation (bedingte, empirische ∼)
Eigenschaften, Rechenregeln
(Symmetrie, Nichtnegativität), Kettenregeln, Konvexität / Konkavität
Unabhängigkeit / (Gleichverteilung (Normalverteilung)) maximiert Entropie (differentielle ∼)
Ungleichungen, grundlegende Theoreme
Kraft-McMillan-Ungleichung, fundamentale Ungleichung, Datenverarbeitungsungleichung, Fano-Ungleichung
Shannon-McMillan-Theorem, Feinstein-Lemma
Informationsmaßebasierte Kenngrößen
Redundanz, Entropierate, Informations-Rate-Distortion-Funktion, Informationskapazität

(4) Herleitung fundamentaler Resultate für abstrahiertes Modell mit math. Werkzeugen

Quellencodierungstheoreme
Aussagen über theoretische Grenzen effizienter Datenkompression (bei definierter Güte)
Sinngemäß: Verlustlose (asymptotisch ∼) Datenkompression ist begrenzt durch Entropie(-rate)
(Entropie(-rate) $\leq$ min. mittl. Cwl (pro Sym.) / Coderate)
Sinngemäß: Verlustbehaftete Datenkompression ist begrenzt durch Informations-Rate-Distortion-Funktion
(Rate-Distortion-Funktion = Informations-∼)

Kanalcodierungstheoreme
Aussagen über theoretische Grenzen effizienter Datenübertragung bei definierter Zuverlässigkeit
Sinngemäß: Maximale Coderate bei zuverlässiger Datenübertragung ist begrenzt durch Informationskapazität
(Codierkapazität = Informationskapazität)

(5) Anwendung der Resultate auf praktische Problemstellung

Entwurfskriterien für optimale Datenkompression / -übertragung, Systembewertung

Literatur

Empfehlenswerte Lehrbücher zur Thematik der Vorlesung für interessierte Studierende:

(keine Pflichtlektüre)

- R. W. Yeung: Information Theory and Network Coding, Part I, Springer, 2008.

  Kapitel zu diskreten Modellen identisch zu:
  R. W. Yeung: A First Course in Information Theory, Springer, 2002.

- R. G. Gallager: Information Theory and Reliable Communication, Wiley, 1968.

- R. G. Gallager: Principles of Digital Communication, Cambridge University Press, 2008.

Erste mathematische Grundlagen, Notation

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

$\mathbb{W}$:= Abkürzung für "Wahrscheinlichkeit(s)"

Vorbemerkungen:

- Das Skript enthält ein etwas umfangreicheres Kapitel 0 zu $\mathbb{W}$-Theorie-Grundlagen, welches wir als Referenz verwenden und aus dem wir uns je nach Bedarf "bedienen" werden.
- Die grundlegenden Begriffe aus der $\mathbb{W}$-Theorie setzen wir als bekannt voraus.
- Wir wiederholen zu Beginn zunächst einige Begriffe zu diskreten $\mathbb{W}$-Modellen, um mit der Informationstheorie "loslegen" zu können und um eine einheitliche, konsistente, in der gesamten Lehrveranstaltung verwendete Notation zu vereinbaren.

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

Übersicht:

- Diskrete Zufallsgröße, $\mathbb{W}$-Funktion, Alphabet, Träger, transformierte Zufallsgröße, Erwartungswert, Varianz
- Schreibweisen
- Gleichverteilung, identisch verteilte Zufallsgrößen
- Diskreter Zufallsvektor, gemeinsame $\mathbb{W}$-Funktion, Rand-$\mathbb{W}$-Funktion, Unabhängigkeit, Kovarianz, Unkorreliertheit
- Bedingte $\mathbb{W}$-Funktion, bedingte Rand-$\mathbb{W}$-Funktion, bedingte Unabhängigkeit, Markowkette der Länge $3, 4, \ldots, n$

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

- Diskrete Zufallsgröße, Alphabet, Träger: Eine Zufallsgröße X, die nur Werte aus der endlichen oder abzählbaren Menge

$$\mathcal{X} = \{a_1, a_2, \ldots, a_m\} \qquad \text{oder} \qquad \mathcal{X} = \{a_1, a_2, a_3, \ldots\}$$

  annehmen kann, nennen wir diskrete Zufallsgröße mit (endlichem) Alphabet $\mathcal{X}$. Die nichtleere Menge

$$\mathcal{S}_X = \{x \in \mathcal{X} : \mathbb{P}(X = x) > 0\} \subset \mathcal{X}$$

  nennen wir Träger von X.

- Es gilt

$$\sum_{x \in \mathcal{X}} \mathbb{P}(X = x) = \sum_{x \in \mathcal{S}_X} \mathbb{P}(X = x) = 1.$$

- $\mathbb{W}$-Funktion: Die Funktion p_X mit

$$p_X(x) = \mathbb{P}(X = x), \qquad x \in \mathcal{X},$$

  die den Werten x der Zufallsgröße X ihre $\mathbb{W}$en zuordnet, nennen wir $\mathbb{W}$-Funktion.

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

- Transformation einer diskreten Zufallsgröße: Sei X eine diskrete Zufallsgröße mit Alphabet $\mathcal{X}$, Träger $\mathcal{S}_X$ und $\mathbb{W}$-Funktion p_X. Sei $\mathcal{Y}$ eine endliche / abzählbare Menge und $g : \mathcal{X} \to \mathcal{Y}$ eine Abbildung. Dann ist

  $$Y = g(X)$$

  eine diskrete Zufallsgröße mit dem Alphabet $\mathcal{Y}$, dem Träger $\mathcal{S}_Y = \{g(x) : x \in \mathcal{S}_X\} \subset \mathcal{Y}$ und der $\mathbb{W}$-Funktion p_Y, für die gilt

  $$p_Y(y) = \sum_{x \in \mathcal{X} : g(x) = y} p_X(x), \qquad y \in \mathcal{Y}.$$

- Erwartungswert einer reellen, diskreten Zufallsgröße:
  Der Erwartungswert der diskreten Zufallsgröße X mit Alphabet $\mathcal{X} \subset \mathbb{R}$ und $\mathbb{W}$-Funktion p_X ist wie folgt definiert.

  $$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} x \cdot p_X(x)$$

  Für die transformierte Zufallsgröße $Y = g(X)$ mit Alphabet $\mathcal{Y} \subset \mathbb{R}$ und $\mathbb{W}$-Funktion p_Y gilt

  $$\mathbb{E}(Y) = \sum_{y \in \mathcal{Y}} y \cdot p_Y(y) = \mathbb{E}(g(X)) = \sum_{x \in \mathcal{X}} g(x) \cdot p_X(x).$$

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

- Varianz einer reellen, diskreten Zufallsgröße:
  Die Varianz der diskreten Zufallsgröße X mit Alphabet $\mathcal{X} \subset \mathbb{R}$ und $\mathbb{W}$-Funktion p_X ist wie folgt definiert.

$$\operatorname{var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$$

  Bemerkung: Die Varianz der Zufallsgröße X ist also gleich dem Erwartungswert der transformierten Zufallsgröße $g(X)$ mit

$$g(x) = (x - \mathbb{E}(X))^2.$$

  Für die Varianz gilt

$$\operatorname{var}(X) = \sum_{x \in \mathcal{X}} (x - \mathbb{E}(X))^2 \cdot p_X(x)$$

  und

$$\operatorname{var}(X) = \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2.$$

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

- Grundsätzliche Schreibweisen – Übersicht:

doppeltgestrichener Großbuchstabe:	$\mathbb{P}$	Wahrscheinlichkeitsmaß
	$\mathbb{E}$	Erwartungswert
normaler Großbuchstabe:	$X, Y, Z, \ldots$	Zufallsgrößen
normaler Kleinbuchstabe:	$x, y, z, \ldots$	Werte der Zufallsgrößen
kalligraphischer Großbuchstabe:	$\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \ldots$	Alphabete der Zufallsgrößen
	$\mathcal{S}_X, \mathcal{S}_Y, \mathcal{S}_Z, \ldots$	Träger der Zufallsgrößen
normaler Kleinbuchstabe p mit normalem Großbuchstaben X als Index:	p_X	$\mathbb{W}$-Funktion der Zufallsgröße X

Hat die Zufallsgröße X das Alphabet $\mathcal{X} = \{a_1, a_2, \ldots, a_m\}$, so wird die $\mathbb{W}$-Funktion p_X mitunter auch als Vektor ihrer Funktionswerte angegeben, d. h. in der Form $p = (p_1, p_2, \ldots, p_m)$, wobei $p_k = p_X(a_k)$ für $k = 1, 2, \ldots, m$.

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

- Gleichverteilung: Eine diskrete Zufallsgröße X mit endlichem Alphabet $\mathcal{X} = \{a_1, a_2, \ldots, a_m\}$ heißt gleichverteilt, falls

$$p_X(x) = \frac{1}{m}, \qquad x \in \mathcal{X}$$

für die $\mathbb{W}$-Funktion von X gilt.

- Identisch verteilte Zufallsgrößen: Zwei diskrete Zufallsgrößen X und Y heißen identisch verteilt, falls

$$\mathcal{S}_X = \mathcal{S}_Y$$

für die zugehörigen Träger und

$$p_X(x) = p_Y(x), \qquad x \in \mathcal{S}_X$$

für die zugehörigen $\mathbb{W}$-Funktionen gilt. Notation: $X \sim Y$.

- Die diskreten Zufallsgrößen $X_1, X_2, \ldots, X_n$ heißen identisch verteilt, falls

$$X_1 \sim X_2 \quad \text{und} \quad X_2 \sim X_3 \quad \text{und} \quad \ldots \quad \text{und} \quad X_{n-1} \sim X_n.$$

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

Beispiel:

- Die Zufallsgröße X beschreibt das Würfelergebnis mit einem fairen Würfel.
  - Alphabet $\mathcal{X}$, Träger $\mathcal{S}_X$:

$$\mathcal{X} = \{1, 2, 3, 4, 5, 6\} = \mathcal{S}_X$$

  - $\mathbb{W}$-Funktion p_X: (Gleichverteilung)

$$p_X(x) = \frac{1}{6}, \quad x \in \mathcal{X}$$

  - $\mathbb{E}$-Wert $\mathbb{E}(X)$:

$$\mathbb{E}(X) = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{7}{2}$$

  - Varianz $\mathrm{var}(X)$:

$$\begin{aligned}\mathrm{var}(X) &= \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2 \\ &= 1^2\frac{1}{6} + 2^2\frac{1}{6} + 3^2\frac{1}{6} + 4^2\frac{1}{6} + 5^2\frac{1}{6} + 6^2\frac{1}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}\end{aligned}$$

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

Beispiel (Fortsetzung):

- Mit $g : \mathcal{X} \to \mathcal{Y}$ definieren wir die Zufallsgröße $Y = g(X)$, wobei gilt:

$$g(x) = 0 \quad \text{falls} \quad x \in \{1,3,5\} \qquad \text{und} \qquad g(x) = 1 \quad \text{falls} \quad x \in \{2,4,6\}$$

  - Alphabet $\mathcal{Y}$, Träger $\mathcal{S}_Y$:

$$\mathcal{Y} = \{0,1\} = \mathcal{S}_Y$$

  - $\mathbb{W}$-Funktion p_Y: (Gleichverteilung, Bernoulli-Verteilung)

$$p_Y(0) = \sum_{x \in \{1,3,5\}} p_X(x) = \frac{1}{2}, \qquad p_Y(1) = \sum_{x \in \{2,4,6\}} p_X(x) = \frac{1}{2}$$

  - $\mathbb{E}$-Wert $\mathbb{E}(Y) = \mathrm{E}(g(X))$:

$$\mathbb{E}(Y) = 0\frac{1}{2} + 1\frac{1}{2} = 0\frac{1}{6} + 0\frac{1}{6} + 0\frac{1}{6} + 1\frac{1}{6} + 1\frac{1}{6} + 1\frac{1}{6} = \frac{1}{2}$$

  - Varianz $\mathrm{var}(Y)$:

$$\begin{aligned} \mathrm{var}(Y) &= \mathbb{E}\left(Y^2\right) - (\mathbb{E}(Y))^2 \\ &= 0^2\frac{1}{2} + 1^2\frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4} \end{aligned}$$

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

- Seien X und Y zwei diskrete Zufallsgrößen mit den Alphabeten $\mathcal{X}$ und $\mathcal{Y}$.
- Gemeinsame $\mathbb{W}$-Funktion: Die Funktion $p_{X,Y}$ mit

  $$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y), \qquad x \in \mathcal{X},\, y \in \mathcal{Y}$$

  heißt gemeinsame $\mathbb{W}$-Funktion der Zufallsgrößen X und Y.
- Diskreter Zufallsvektor: Das 2-Tupel (X,Y), bestehend aus den diskreten Zufallsgrößen X und Y heißt (2-dimensionaler) diskreter Zufallsvektor.

  (Gemäß Definition auf Folie 25 gilt: Der diskrete Zufallsvektor (X,Y) ist zugleich eine diskrete Zufallsgröße mit dem Alphabet $\mathcal{X} \times \mathcal{Y}$ und der $\mathbb{W}$-Funktion $p_{X,Y}$.)
- Rand-$\mathbb{W}$-Funktionen: Die $\mathbb{W}$-Funktion p_X von X und die $\mathbb{W}$-Funktion p_Y von Y heißen auch Rand-$\mathbb{W}$-Funktionen, denn es gilt:

  $$p_X(x) = \sum_{y \in \mathcal{Y}} p_{X,Y}(x,y),\ x \in \mathcal{X}, \quad p_Y(y) = \sum_{x \in \mathcal{X}} p_{X,Y}(x,y),\ y \in \mathcal{Y}.$$
- Unabhängigkeit: Die diskreten Zufallsgrößen X und Y heißen (stochastisch) unabhängig, falls für alle $(x,y) \in \mathcal{X} \times \mathcal{Y}$ gilt:

  $$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y).$$
- n-dimensionale Zufallsvektoren: $X = (X_1, X_2, \ldots, X_n) \to$ Skript $\S\, 0.3$.

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

- Kovarianz reeller, diskreter Zufallsgrößen:
  Die Kovarianz der diskreten Zufallsgrößen X und Y mit Alphabeten $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}$ und gemeinsamer $\mathbb{W}$-Funktion $p_{X,Y}$ ist wie folgt definiert.

  $$\operatorname{cov}(X,Y) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big) \cdot \big(Y - \mathbb{E}(Y)\big)\Big)$$

  Für die Kovarianz gilt

  $$\operatorname{cov}(X,Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \big(x - \mathbb{E}(X)\big) \cdot \big(y - \mathbb{E}(Y)\big) \cdot p_{X,Y}(x,y)$$

  und

  $$\operatorname{cov}(X,Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

- Gilt $\operatorname{cov}(X,Y) = 0$, so nennt man X und Y unkorreliert.
- Unabhängigkeit/Unkorreliertheit: Sind X und Y unabhängig, dann sind sie auch unkorreliert. Aus der Unkorreliertheit folgt im Allgemeinen jedoch nicht die Unabhängigkeit.

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

- Seien X und Y zwei diskrete Zufallsgrößen mit den Alphabeten $\mathcal{X}$ und $\mathcal{Y}$ sowie den Trägern $\mathcal{S}_X$ und $\mathcal{S}_Y$.
- Bedingte $\mathbb{W}$-Funktion: Ist $p_{X,Y}$ die gemeinsame $\mathbb{W}$-Funktion von X und Y sowie p_X die $\mathbb{W}$-Funktion von X, dann heißt für jedes $x \in \mathcal{S}_X$ die $\mathbb{W}$-Funktion $p_{Y|X}(\cdot|x)$ mit

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}, \qquad y \in \mathcal{Y}$$

die bedingte $\mathbb{W}$-Funktion von Y unter der Bedingung $X = x$. Es gilt für alle $x \in \mathcal{S}_X$

$$p_{Y|X}(y|x) = 0, \quad y \in \mathcal{Y} \setminus \mathcal{S}_Y \qquad \text{und} \qquad \sum_{y \in \mathcal{Y}} p_{Y|X}(y|x) = 1.$$

Erweitert man die Definition auch für alle $x \in \mathcal{X} \setminus \mathcal{S}_X$ geeignet (z. B. derart, dass $p_{Y|X}(\cdot|x)$ stets eine $\mathbb{W}$-Funktion ist), dann nennt man $p_{Y|X}$ als Funktion auf $\mathcal{X} \times \mathcal{Y}$ die bedingte $\mathbb{W}$-Funktion von Y unter der Bedingung X.

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

- Seien X, Y und Z drei diskrete Zufallsgrößen mit den Alphabeten $\mathcal{X}$, $\mathcal{Y}$ und $\mathcal{Z}$ sowie den Trägern $\mathcal{S}_X$, $\mathcal{S}_Y$ und $\mathcal{S}_Z$. Weiterhin seien:

  $p_{(X,Y)|Z}$ bedingte $\mathbb{W}$-Funktion von (X,Y) unter der Bedingung Z,
  $p_{X|Z}$ bedingte $\mathbb{W}$-Funktion von X unter der Bedingung Z,
  $p_{Y|Z}$ bedingte $\mathbb{W}$-Funktion von Y unter der Bedingung Z.

- Bedingte Rand-$\mathbb{W}$-Funktionen: Für alle $z \in \mathcal{S}_Z$ und $x \in \mathcal{X}$ gilt

$$p_{X|Z}(x|z) = \sum_{y \in \mathcal{Y}} p_{(X,Y)|Z}\big((x,y)|z\big)$$

  sowie für alle $z \in \mathcal{S}_Z$ und $y \in \mathcal{Y}$ gilt

$$p_{Y|Z}(y|z) = \sum_{x \in \mathcal{X}} p_{(X,Y)|Z}\big((x,y)|z\big).$$

- Bedingte Unabhängigkeit: Die diskreten Zufallsgrößen X und Y heißen bedingt (stochastisch) unabhängig gegeben Z, falls für alle $z \in \mathcal{S}_Z$, $x \in \mathcal{X}$ und $y \in \mathcal{Y}$ gilt

$$p_{(X,Y)|Z}\big((x,y)|z\big) = p_{X|Z}(x|z) \cdot p_{Y|Z}(y|z).$$

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

- Markowkette der Länge 3: Die drei diskreten Zufallsgrößen X, Z und Y bilden eine Markowkette in dieser Reihenfolge, falls für alle $(z, x) \in \mathcal{S}_{(Z,X)}$ und $y \in \mathcal{Y}$ gilt

$$p_{Y|(Z,X)}\big(y|(z,x)\big) = p_{Y|Z}(y|z),$$

wobei $\mathcal{S}_{(Z,X)}$ den Träger des diskreten Zufallsvektors (Z, X) bezeichnet. Notation: $X \to Z \to Y$.

- Es gelten folgende Äquivalenzen:

$$X \to Z \to Y \quad \Longleftrightarrow \quad X \text{ und } Y \text{ sind bedingt unabhängig gegeben } Z$$

$$X \to Z \to Y \quad \Longleftrightarrow \quad Y \to Z \to X$$

Wiederholung Grundbegriffe $\mathbb{W}$-Theorie

- Markowkette der Länge 4: Die vier diskreten Zufallsgrößen X_1, X_2, X_3 und X_4 bilden eine Markowkette in dieser Reihenfolge, falls gilt

$$X_1 \to X_2 \to X_3 \quad \text{und} \quad (X_1, X_2) \to X_3 \to X_4.$$

  Notation: $X_1 \to X_2 \to X_3 \to X_4$

- Markowkette der Länge n: Die diskreten Zufallsgrößen $X_1, X_2, \ldots, X_n$ bilden eine Markowkette in dieser Reihenfolge, falls gilt

$$(X_1, X_2, \ldots, X_{k-2}) \to X_{k-1} \to X_k$$

  für alle $k = 3, 4, \ldots, n$ gilt.

  Notation: $X_1 \to X_2 \to \ldots \to X_{n-1} \to X_n$

Umfrage zur Komprimierbarkeit

Umfrage zur Komprimierbarkeit von Bilddaten

Situation:

- Wir betrachten zwei Bilddateien mit identischer Pixelanzahl im RGB-Farbformat mit einer Farbtiefe von jeweils 8 Bit (= 1 Byte) pro Farbkanal.
- Die beiden Bilddateien haben folgende Abmessungen:

$$1125 \times 1040 = 1.170.000 \text{ Pixel} \qquad \text{und} \qquad 1300 \times 900 = 1.170.000 \text{ Pixel}.$$

- Daraus ergibt sich für jede Bilddatei ein identischer Speicherbedarf von

$$1.170.000 \times 3 \times 8 \text{ Bit} = 28.080.000 \text{ Bit} = 3.510.000 \text{ Byte}$$

  bei unkomprimierter Speicherung (Vernachlässigung von Headerdaten).

Online-Umfrage:

Was denken Sie, welche der beiden nachfolgend eingeblendeten Bilddateien lässt sich mit verlustloser Datenkompression stärker komprimieren?

A) Das Bild mit dem Titel: "Strip (927-9)"

B) Das Bild mit dem Titel: "Abstraktes Bild (946-3)"

$\Longrightarrow$ Bitte stimmen Sie nach Betrachtung der Bilder bis **Montag, 10.04.2023,** unter folgendem Link ab (Link im PDF anklickbar).

https://bildungsportal.sachsen.de/onyxeditor/published/BrKoBG0qNZOr

Gerhard Richter: Strip (927-9), 2012
(Galerie Neue Meister, Staatliche Kunstsammlungen Dresden)

Gerhard Richter: Abstraktes Bild (946-3), 2016

(gezeigt im Albertinum, Staatliche Kunstsammlungen Dresden, in der Ausstellung "Gerhard Richter: Neue Bilder")