Einführung in die Informations- und Lerntheorie

0. Motivation und Ziele

- Wie extrahiert man aus einer großen Menge von Daten die relevante Information?
- Wie erklärt man die gewonnene Information?
- Welche unter den möglichen Erklärungen wählt man als "die richtige Erklärung" aus?

⋮

Die Erforschung von Information in Bezug auf ihre Generierung, ihre Extraktion und ihre Behandlung gehört zu den größten Herausforderungen unserer Zeit. Aufgrund der immer größer werdenden Datenmengen benötigt man Systeme, die Fragen wie die obigen automatisch beantworten. Derartige Systeme heißen <u>induktive Inferenzsysteme</u>.

<u>Ziel</u>:
Lernen der notwendigen Bausteine und Methoden zur Konstruktion von induktiven Inferenzsysteme.

Einfacher als die Extraktion unbekannter Information aus Daten ist die Generierung von Daten zur Speicherung bekannter Information. Mit Problemen dieser Art beschäftigt sich die

sogenannte "klassische" Informationstheorie. Diese hat ihren Ursprung in der Kommunikationstheorie. Fundamental für die Kommunikationstheorie sind folgende zwei Fragen:

- Welche Datenkompression kann ohne Informationsverlust erreicht werden?
- Welche Übertragungsrate kann bestenfalls bei der Datenübermittlung erreicht werden?

Mit beiden Fragestellungen hat sich die klassische Informationstheorie intensiv beschäftigt und diese auch abschließend behandelt. Dabei hat man die Semantik einer Botschaft ignoriert und war ausschließlich an dem Kommunikationsproblem der Übermittlung einer Botschaft von einem Sender zu einem Empfänger interessiert. Dies geschah unter der Annahme, dass beide Beteiligten das Universum aller möglichen Botschaften kennen. Zentral hierbei war die informationstheoretische Quantität Entropie, die diejenige Information misst, die benötigt wird, damit der Empfänger aus dem bekannten Universum aller Botschaften die richtige Botschaft auswählt. Die hierbei entwickelten Methoden und Konzepte finden auch bei der Konstruktion von induktiven Inferenzsystemen ihre Anwendung. Daher werden wir

uns zunächst mit der klassischen Informationstheorie beschäftigen.

Einen anderen Weg geht die sogenannte "algorithmische" Informationstheorie. Diese definiert das Maß an Information nur in Abhängigkeit der individuellen Botschaft, also nicht in Relation zu einem gegeben Universum. Sie nimmt als Maß für die in einem String enthaltenen Information die Länge eines kürzesten Programms, das ohne Eingabe diesen String ausgibt. Die hierbei entwickelten Konzepte sind zentral bei der Konstruktion von induktiven Inferenzsysteme. Daher werden wir uns nach der klassischen mit der algorithmischen Informationstheorie beschäftigen.

Induktive Inferenzsysteme versuchen unter anderen Regelmäßigkeiten in Daten zu entdecken um dann hieraus Information zu gewinnen. Zufällige Daten enthalten keine solche Regelmäßigkeiten. Methoden und Konzepte, die bei der Charakterisierung von binären Zufallsfolgen entwickelt wurden, finden direkt ihre Anwendung bei der Konstruktion von induktiven Inferenzsystemen. Aus diesem Grund werden wir uns nach der algorithmischen Informationstheorie mit der Charakterisierung von binären Zufallsfolgen beschäftigen.

Nun können wir uns endlich den induktiven Inferenzsystemen widmen. Ausgehend von einem sehr abstrakten universellen induktiven Inferenzsystems werden wir auf praktische induktive Inferenzsysteme hinarbeiten.

Literatur:

- Norbert Blum, Einführung in Formale Sprachen, Berechenbarkeit, Informations- und Lerntheorie, Oldenbourg 2007.
- Literatur, die ich bei der Konzeption der Vorlesung verwendet habe, werden zu Beginn der betreffenden Kapitel genannt.
- weitere Literatur im
  - Sonderfach der Vorlesung, Handapparat der Institutsbibliothek.

Hinweis auf:

Stasys Jukna, Crashkurs Mathematik für Informatiker, Teubner 2008.

1. Die klassische Informationstheorie

Literatur:

- Claude E. Shannon, Warren Weaver, The Mathematical Theory of Communication, University of Illinois Press, 1949.
- Norman Abramson, Information Theory and Coding, McGraw-Hill, 1963.
- Silviu Guiaşu, Information Theory with Applications, McGraw-Hill, 1977.
- Thomas M. Cover, Joy A. Thomas, Elements of Information Theory, John Wiley & Sons, 1991.

1.1 Die Entropie

Seien

E ein Ereignis

$p(E) > 0$ die Wahrscheinlichkeit, dass E eintritt.

Die Mitteilung, dass das Ereignis E eingetreten ist, enthält dann

$$I(E) := \log \frac{1}{p(E)}$$

Informationseinheiten.

Frage: Inwieweit entspricht obige Definition unserer Intuition?

- sicheres Ereignis, d.h., $p(E) = 1$:

$$I(E) = \log 1 = 0. \quad ✓$$

- Je kleiner $p(E)$ umso größer ist $I(E)$. ✓

- $p(E) = \frac{1}{2} \Rightarrow I(E) = 1.$
  Ein Bit Information erhalten wir, wenn eine von zwei gleichwahrscheinlichen Alternativen, nämlich E und $\neg E$ spezifiziert wird. ✓

<u>Zufallsexperiment</u>:

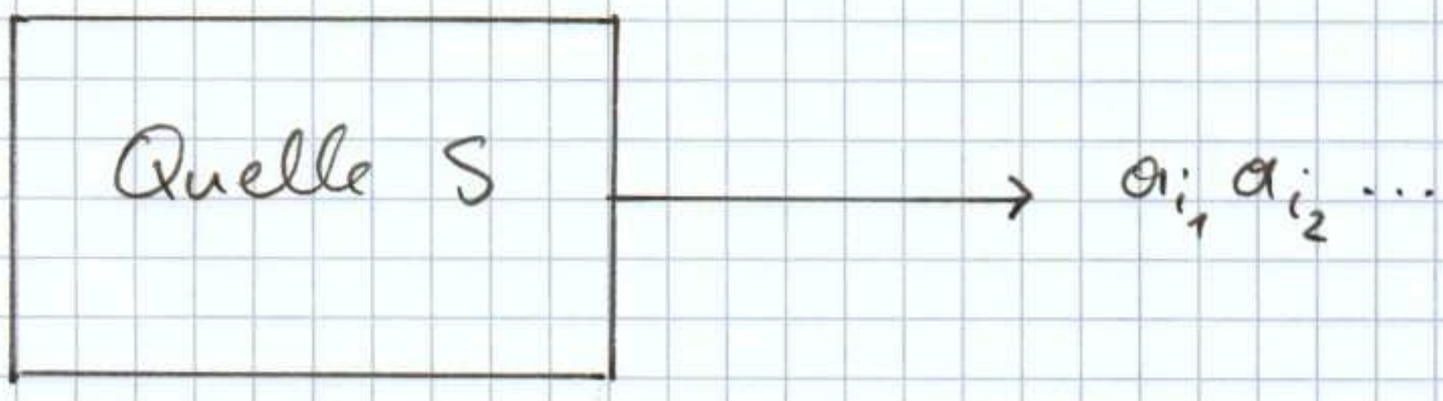
- n elementare Ereignisse $a_1, a_2, \dots, a_n$, die mit den Wahrscheinlichkeiten $p_1, p_2, \dots, p_n$ eintreten. Dabei gilt

$$p_i \geq 0 \quad \text{für } 1 \leq i \leq n \quad \text{und}$$

$$\sum_{i=1}^{n} p_i = 1.$$

$p(a_i)$ Wahrscheinlichkeit, dass a_i eintritt.

<u>Beispiel:</u> (<u>gedächtnislose Informationsquelle</u>)



Sukzessive Generierung von Symbolen
- gemäß Wahrscheinlichkeitsverteilung
$p(a_i) = p_i \quad , 1 \leq i \leq n$

- unabhängig von der bisher generierten Symbolfolge.

Falls das Symbol a_i generiert wird, dann erhalten wir

$$I(a_i) = \log \frac{1}{p_i}$$

Bits an Information.

$\Rightarrow$

<u>mittlere Information / Symbol</u>:

$$\sum_{i=1}^{n} p_i I(a_i) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}.$$

◆

$\rightsquigarrow$

Für eine endliche Wahrscheinlichkeitsverteilung $p = p_1, p_2, \dots, p_n$ definieren wir die <u>Entropie</u> H_n durch

$$H_n := H_n(p_1, p_2, \dots, p_n) := -\sum_{i=1}^{n} p_i \log p_i.$$

Dabei definieren wir $p_i \log p_i = 0$ falls $p_i = 0$.
Wegen $\lim_{x \to 0} x \log x = 0$ ist dies sinnvoll.

<u>Beispiel</u>:

$S = \{a_1, a_2, a_3\}$, $p(a_1) = \frac{1}{2}$, $p(a_2) = p(a_3) = \frac{1}{4}$.

$\rightsquigarrow$

$$H_3 = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\log 4$$
$$= \frac{3}{2}.$$

<u>Interpretation:</u>

- $I(a_i)$ "diejenige Information, die für die Spezifikation, dass das Ereignis a_i eintritt, benötigt wird."

- $H_n(p_1, p_2, \dots, p_n)$
  - <u>mittlere Informationsgröße</u> pro Ausgang eines Zufallsexperimentes oder
  - <u>mittlere Unsicherheit</u> eines Beobachters vor Ausgang des Zufallsexperimentes.

<u>Frage:</u> Besitzt die Entropie diejenige Eigenschaften, die gemäß unserer Intuition mittlere Unsicherheit bzw. mittlere Information pro Ausgang eines Zufallsexperimentes haben sollte?

$\rightsquigarrow$

14.10.

<u>Lemma 1.1</u>

Sei p_i, $1 \le i \le n$ eine Wahrscheinlichkeitsverteilung. Dann gilt: