Grundlagen der Schaltungstechnik

Eine Reihe,
herausgegeben von Prof. Dr.-Ing. Wolfgang Hilberg

Die Buchreihe umfaßt Themen aus dem Gebiet des Entwurfs, der technologischen Realisierung und der Anwendung von Schaltungen. Vorzugsweise sind dies integrierte Halbleiterschaltungen, die heute die gemeinsame „hardware"-Basis für viele Anwendungen, z. B. in der Nachrichtentechnik, Meßtechnik, Digitaltechnik, Datentechnik bzw. der Elektronik bilden. Die Darstellungen sollen dem heutigen Stand der Technik entsprechend die grundlegenden Kenntnisse vermitteln.

Bisher erschienen:

Großintegration
herausgegeben von
Bernd Höffinger

Wolfgang Hilberg
Impulse auf Leitungen

Frank-Thomas Mellert
Rechnergestützter Entwurf
elektrischer Schaltungen

Wolfgang Hilberg/Robert Piloty
Grundlagen elektronischer
Digitalschaltungen

Adolf Finger
Digitale Signalstrukturen
in der Informationstechnik

Wolfgang Hilberg
Grundprobleme der
Mikroelektronik

Günter Zimmer
CMOS-Technologie

Manfred Lobjinski
Meßtechnik mit Mikrocomputern
2. Auflage

Wolfgang Hilberg
Assoziative Gedächtnisstrukturen.
Funktionale Komplexität

Hans Spiro
Simulation integrierter
Schaltungen
2. Auflage

Samuel D. Stearns
Digitale Verarbeitung analoger
Signale, 5. Auflage

Klaus Schumacher
Integrationsgerechter Entwurf
analoger MOS-Schaltungen

Steffen Graf/Michael Gössel
Fehlererkennungsschaltungen

Wolfgang Hilberg
Digitale Speicher 1

Hochintegrierte analoge Schaltungen
herausgegeben von Bernd Höfflinger
und Günter Zimmer

Friedberth Riedel
MOS-Analogtechnik

Wolfgang Hilberg
Grundlagen elektronischer
Schaltungen, 2. Auflage

Robert Schwarz
Analyse nichtlinearer Netzwerke

Albrecht Rothermel
Digitale BiCMOS-Schaltungen

Manfred Gerner / Bruno Müller /
Gerd Sandweg
Selbsttest digitaler Schaltungen

Oppenheim / Schafer
Zeitdiskrete Signalverarbeitung

Hans Tzschach / Gerhard Haßlinger
Codes für den störungssicheren
Datentransfer

Codes für den störungssicheren Datentransfer

von
Professor Dr. rer. nat. Hans Tzschach
und
Dr. rer. nat. Gerhard Haßlinger

Mit 41 Bildern, 70 Beispielen und 67 Übungen

R. Oldenbourg Verlag München Wien 1993

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Tzschach, Hans:
Codes für störungssicheren Datentransfer : mit 70 Beispielen und 67 Übungen / von Hans Tzschach und Gerhard Haßlinger. – München : Oldenbourg, 1993
(Grundlagen der Schaltungstechnik)
ISBN 3-486-22569-3

NE: Haßlinger, Gerhard:

Gesamtherstellung: Grafik + Druck, München

ISBN 3-486-22569-3

Inhaltsverzeichnis

Vorwort

Dieses Buch entstand aus dem Skript zu einer stets rege besuchten Vorlesung über Codierungstheorie, welche von den Autoren seit mehr als zehn Jahren an der Technischen Hochschule Darmstadt im Informatik-Hauptstudium angeboten wird.

Es soll dem Informatiker und Ingenieur einen schnellen Einstieg in das Gebiet der fehlererkennenden und -korrigierenden Codes ermöglichen und einen Einblick in die dazu grundlegenden Ideen von Shannon vermitteln.
Neben einer Zusammenstellung der zum Verständnis notwendigen Begriffe aus der Algebra finden Modelle des Übertragungskanals, Codierungs- und Decodierungsverfahren sowie die Bestimmung ihrer Restfehlerraten Beachtung. In Anpassung an heutige Anwendungen wurde z.B. auf Reed-Muller-Codes verzichtet, die nur noch von historischem Interesse sind, während Faltungs- und Odd-Weight-Codes ausführlich dargestellt sind.
Auch wenn technische Fortschritte zur Verringerung der Fehleranfälligkeit in der Datenverarbeitung beitragen, wie etwa Glasfaser als Übertragungsmedium, so gewinnt das Thema des Buches dennoch durch die Vielfalt und den rasant wachsenden Umfang des Informationsaustauschs an Bedeutung.

Der Dank der Autoren gilt insbesondere Herrn Dr. C. Kröll für das von ihm verfaßte Kapitel über gedächtnisbehaftete Übertragungskanäle und Herrn Dipl.-Math. M. König. Für ihre Unterstützung bei der Ausarbeitung von Text und Abbildungen danken wir Frau M. Jayme, Frau M. Skrobic und Frau U. Schott. Darüber hinaus haben Mitarbeiter und Studierende am Fachbereich Informatik der THD durch zahlreiche Hinweise auf Fehler und Verbesserungsvorschläge eine wichtige Hilfestellung gegeben.

Hans Tzschach, Gerhard Haßlinger im Dezember 1992

Kapitel 1

Informationstheorie

1.1 Information und Entropie

Die Erzeugung und Verbreitung von Nachrichten bzw. Informationen zielt stets darauf ab, dem Adressaten einen neuen und mehr oder weniger unvorhersehbaren Tatbestand zu vermitteln. Informationsbedarf setzt andererseits eine Ungewißheit über den Inhalt einer eintreffenden Nachricht voraus. Daher kann der Nachrichtenempfang als vom Zufall beeinflußter Vorgang betrachtet werden, zu dessen Beschreibung stochastische Modellbildungen herangezogen werden.
Unter den vielfältigen Möglichkeiten, die als Träger von Information in Frage kommen, beschränken wir uns auf zeichenorientierte Darstellungsformen. Die *Zeichen* oder *Symbol* sollen einem endlichen entnommen sein und bilden die kleinste Nachrichteneinheit. Sie werden mit den Elementarereignissen eines Wahrscheinlichkeitsraums identifiziert.

Definition 1.1 Es sei $(A, \mathcal{P}(A), p)$ ein endlicher *Wahrscheinlichkeitsraum.* Hierbei ist $A = \{a_1, a_2, \cdots, a_n\}$ die Menge der *Elementarereignisse,* $\mathcal{P}(A)$ die *Potenzmenge* von A und p ein *Wahrscheinlichkeitsmaß,* welches jedem Ereignis aus der Potenzmenge (Menge aller Teilmengen) einen Wert zuordnet mit den Eigenschaften:

1. $\forall Q \in \mathcal{P}(A) : \; p(Q) \geq 0;$
2. $p(A) = 1$
3. $\forall Q, R \in \mathcal{P}(A)$ mit $Q \cap R = \emptyset$ gilt: $p(Q \cup R) = p(Q) + p(R);$

Für die Elementarereignisse wird die vereinfachte Schreibweise $p(a_i) \stackrel{\text{def}}{=} p_i$ benutzt. Die genannten Eigenschaften beinhalten auch die Normierungsbedingung $\sum_{i=1}^{n} p_i = 1$. Hat $Q \in \mathcal{P}(A)$ die Wahrscheinlichkeit $p(Q)$, so wird dem Ereignis Q der *Informationsgehalt*

$$I(Q) = -\log p(Q) \quad \text{bzw.} \quad I(a_i) = -\log p_i \quad \text{für Elementarereignisse zugeordnet.} \tag{1.1}$$

Bei dieser Definition lassen wir die Basis b des Logarithmus' unbestimmt. Rechnet man zur Basis $b = 2$, so wird in der *Informationseinheit bit* gemessen, was einer it binären Darstellung der Information mit nur zwei Zeichen z.B. 0 und 1 entspricht.

Die Umrechnung des Informationsgehalts von der Rechnung in einer anderen Basis b in die Einheit bit erfolgt dann durch Multiplikation mit dem Faktor $\log_2 b$.

Die Definition entspricht der anschaulichen Vorstellung, daß der Informationsgehalt, den das Ergebnis eines Zufallsexperiments liefert, um so größer ist, je kleiner die Wahrscheinlichkeit für das eintretende Ereignis ist. Das sichere Ereignis, das mit Wahrscheinlichkeit 1 eintritt, hat den Informationsgehalt 0, da der Ablauf des Experiments schon vorher bekannt ist. Faßt man mehrere Ereignisse, die unabhängig voneinander mit Wahrscheinlichkeiten $q_1, q_2, \cdots, q_n$ eintreten, zu einem gemeinsamen (Und-)Ereignis zusammen, so ist seine Wahrscheinlichkeit das Produkt $q_1 \cdot q_2 \cdots q_n$ und sein Informationsgehalt entspricht gerade der Summe aus den Einzelinformationen $I = \sum_{i=1}^{n} -\log q_i$.

Der mittlere Informationsgehalt bezogen auf alle Elementarereignisse eines Wahrscheinlichkeitsraums wird als Entropie bezeichnet. Sie kann als Maß für den Informationsgewinn interpretiert werden, den der Ausgang eines Zufallsexperiments im Mittel liefert oder umgekehrt für die Unsicherheit, die vor seiner Durchführung hinsichtlich des Ergebnisses besteht.

Definition 1.2 Unter den Voraussetzungen von Definition 1.1 bezeichnet man

$$H(A) = \sum_{i=1}^{n} p(a_i) I(a_i) = -\sum_{i=1}^{n} p_i \log p_i = H(p_1, p_2, \cdots, p_n) \tag{1.2}$$

für $n \geq 2$ als *Entropie*. $H(p_1)$ mit nur einem Argument wird als abkürzende Schreibweise für $H(p_1, 1 - p_1)$ benutzt. ⌋

Die Entropie ist eine nur vom Wahrscheinlichkeitsmaß abhängige n-stellige Funktion. Wegen $\lim_{x \to 0} x \log x = 0$ ordnet man für $p_i = 0$ dem zugehörigen Term $p_i \log p_i$ ebenfalls den Wert Null zu. Wird der Logarithmus zu einer bestimmten Basis b gebildet, so dient ein Index $H_b(A)$ als Kennzeichnung.

Beispiel 1.1 Ein *HDTV-Fernsehbild* kann als Matrix von Bildpunkten mit etwa 1900 Zeilen und 1150 Spalten angesehen werden. Jeder Bildpunkt kann 2^8 verschiedene Helligkeits- oder Farbwerte annehmen. Es sind demnach etwa $2^{8 \cdot 1900 \cdot 1150} \approx 2^{17\,500\,000}$ verschiedene Fernsehbilder möglich. Unter der Annahme, daß die Werte der Bildpunkte voneinander unabhängig und jeweils gleichverteilt sind, ist der Informationsgehalt $I(s_i)$ eines Fernsehbildes

$$I(s_i) \approx -\log 2^{-17\,500\,000} = 1.75 \cdot 10^7 \text{ bit}$$

und die Entropie

$$H(S) \approx \sum_{i=1}^{2^{17\,500\,000}} \frac{1}{2^{17\,500\,000}} \cdot 17\,500\,000 \log 2 = 1.75 \cdot 10^7 \text{ bit}.$$

1.2 Eigenschaften der Entropie

(i) $H(A) = 0 \Leftrightarrow \exists\, i \in \{1, \ldots, n\} : p(a_i) = 1 \quad \wedge \quad \forall j \neq i : \; p(a_j) = 0;$

(ii) Es sei $p_i \geq 0$ und $q_i > 0$ für $i = 1, \ldots, n$, sowie $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i = 1$. Dann gilt:

$$H(p_1, \ldots, p_n) = -\sum_{i=1}^n p_i \log p_i \leq -\sum_{i=1}^n p_i \log q_i. \tag{1.3}$$

Beweis: Für jede positive reelle Zahl x gilt

$$\log x \leq (x-1)\log e. \quad \text{Es folgt:} \tag{1.4}$$

$$\sum_{i=1}^n p_i \log q_i - \sum_{i=1}^n p_i \log p_i = \sum_{\substack{i=1 \\ p_i \neq 0}}^n p_i \log \frac{q_i}{p_i} \leq \sum_{\substack{i=1 \\ p_i \neq 0}}^n p_i \left(\frac{q_i}{p_i} - 1\right) \log e$$

$$= \sum_{\substack{i=1 \\ p_i \neq 0}}^n (q_i - p_i) \log e \leq \log e \sum_{i=1}^n (q_i - p_i) = \log e \left(\underbrace{\sum_{i=1}^n q_i}_{=1} - \underbrace{\sum_{i=1}^n p_i}_{=1} \right) = 0$$

wobei wegen $p_i = 0 \Rightarrow p_i \log q_i = p_i \log p_i = 0$ nur Summanden mit $p_i \neq 0$ berücksichtigt werden. □

(iii) Für eine feste Anzahl $|A| = n$ von Elementarereignissen nimmt die Entropie ihr Maximum im Fall der Gleichverteilung an ($\sum_i p_i = 1$ ist für jede Verteilung vorausgesetzt):

$$H(p_1, \cdots, p_n) \leq H\Big(\underbrace{\frac{1}{n}, \cdots, \frac{1}{n}}_{n-\text{mal}} \Big). \tag{1.5}$$

Beweis: Setzt man in (1.3) $q_i = \frac{1}{n}$, so folgt:

$$H(p_1, \cdots, p_n) = -\sum_{i=1}^n p_i \log p_i \leq -\sum_{i=1}^n p_i \log \frac{1}{n} = -\log \frac{1}{n} \sum_{i=1}^n p_i = \log n.$$

□

Man sieht, daß das bei einer Gleichverteilung erreichbare Maximum der Entropie mit der Zahl n der Elementarereignisse monoton wächst.

(iv) Es gilt:

$$\begin{aligned} H(p_1, \ldots, p_n) = {} & H(p_1, \ldots, p_{i-1}, p_i + p_{i+1}, p_{i+2}, \cdots, p_n) \\ & + (p_i + p_{i+1}) H\left(\frac{p_i}{p_i + p_{i+1}}, \frac{p_{i+1}}{p_i + p_{i+1}} \right). \end{aligned} \tag{1.6}$$

Beweis: Es genügt, den Fall $i = 1$ zu betrachten, da eine Vertauschung der Argumente den Funktionswert von H nicht beeinflußt:

$$H(p_1 + p_2, p_3, \ldots, p_n) = -(p_1 + p_2)\log(p_1 + p_2) - \sum_{i=3}^{n} p_i \log p_i;$$

$$(p_1 + p_2)\, H\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right) = -p_1 \log \frac{p_1}{p_1 + p_2} - p_2 \log \frac{p_2}{p_1 + p_2}$$

$$= (p_1 + p_2)\log(p_1 + p_2) - p_1 \log p_1 - p_2 \log p_2.$$

Eine Addition beider Gleichungen bestätigt die Behauptung. □

(v)

$$H\Big(\underbrace{\frac{1}{n\,l}, \frac{1}{n\,l}, \ldots, \frac{1}{n\,l}}_{n\,l\text{-mal}}\Big) = H\Big(\underbrace{\frac{1}{n}, \ldots, \frac{1}{n}}_{n\text{-mal}}\Big) + H\Big(\underbrace{\frac{1}{l}, \ldots, \frac{1}{l}}_{l\text{-mal}}\Big) \tag{1.7}$$

Dies folgt direkt aus

$$H\Big(\underbrace{\frac{1}{n\,l}, \frac{1}{n\,l}, \ldots, \frac{1}{n\,l}}_{n\,l\text{-mal}}\Big) = \log(n\,l) = \log n + \log l.$$

Satz 1.1 Die Entropie $H(A)$ ist die einzige stetige Funktion, die die Bedingungen (iii)–(v) erfüllt. ⌋

Der Beweis für diesen Eindeutigkeitssatz der Entropie wird für insgesamt äquivalente Bedingungen z.B. in [DuJü77] oder [HeHo70] geführt, siehe auch Übungsaufgabe 1.7. Man beachte, daß die Logarithmus-Funktion in den Eigenschaften (1.4)-(1.6) nicht einmal benannt wird.

1.3 Gedächtnislose Informationsquellen

Bevor die Informationsquelle in Anlehnung an den Begriff eines stochastischen Prozesses eingeführt wird, sind zunächst einige weitere Bezeichnungsweisen aus der Stochastik anzusprechen.
Eine reellwertige Funktion $f(a_i)$ der Elementarereignisse $a_i \in A$ eines Wahrscheinlichkeitsraums $(A, \mathcal{P}(A), p)$ wird durch eine *Zufallsvariable* bezeichnet.[1]
Das Wahrscheinlichkeitsmaß p des zugrundeliegenden Wahrscheinlichkeitsraums ist auf solche Zufallsvariablen anwendbar. Die Schreibweise $p(X = f(a_i)) = p_i$ besagt, daß die Zufallsvariable X mit Wahrscheinlichkeit p_i den Wert $f(a_i)$ annimmt.
Für $f(a_i) = i$ kann man insbesondere die Aufzählung der Elementarereignisse als Zufallsvariable X darstellen, so daß $p_i = p(X = i)$ ist.

[1] Zur Unterscheidung von einfachen Variablen werden für Zufallsvariable nur Großbuchstaben benutzt.

Definition 1.3 Der *Erwartungswert* oder *Mittelwert* $\mathrm{E}(X)$ einer Zufallsvariable X ist durch $E(X) = \sum_i p(a_i)\, f(a_i)$ bestimmt. X_1 und X_2 seien Zufallsvariablen im Wahrscheinlichkeitsraum $Q_1 = (A, \mathcal{P}(A), p_A)$ und $Q_2 = (B, \mathcal{P}(B), p_B)$. Die Zufallsvariablen bzw. die damit beschriebenen Zufallsexperimente heißen *unabhängig*, wenn gilt:

$$\forall\ a, b \in \mathbb{R}: \quad p(X_1 = a \text{ und } X_2 = b) = p(X_1 = a)\, p(X_2 = b) = p(a)\, p(b). \quad (1.8)$$

Eine Folge von (unabhängigen) Zufallsvariablen $\{X_t;\ t \in \mathbb{N}_0\}$ bildet einen *(gedächtnislosen) stochastischen Prozeß.* ⌋

Definition 1.4 Die Bezeichnungsweise $p(a|B)$ steht für die *bedingte Wahrscheinlichkeit* eines Ereignisses a unter einer Bedingung B, die als logische Aussage unter Einbeziehung von Zufallsvariablen formuliert werden kann. In diesem Zusammenhang spricht man auch von einem bedingten Informationsgehalt und einer bedingten Entropie. ⌋

Mit einer Bedingung B kann ein Vorwissen zum Ausdruck gebracht werden, das mit dem Ergebnis eines Zufallsexperiments zusammenhängt, so daß das Wahrscheinlichkeitsmaß davon beeinflußt wird. Für Zufallsvariable X_1 und X_2 gilt:

$$p(X_1 = a|X_2 = b) = p(X_1 = a \text{ und } X_2 = b)/p(X_2 = b), \quad (1.9)$$

sofern $p(X_2 = b) > 0$.
Sind X_1 und X_2 unabhängig, so folgt allerdings $p(X_1 = a|X_2 = b) = p(X_1 = a)$.

Beispiel 1.2 Das Zufallsexperiment "Werfen eines Würfels" liefert mit Wahrscheinlichkeit je $1/6$ eine der Zahlen $1, \cdots, 6$ als Ergebnis. Seien X_1 und X_2 Zufallsvariable für die Ergebnisse von zwei unabhängigen Würfen und $S = X_1 + X_2$ für deren Summe. Dann gilt:

$$p(S = i) = \begin{cases} \sum_{j=1}^{i-1} p(X_1 = j)\, p(X_2 = i - j) & \text{für } 2 \le i \le 7 \\ \sum_{j=i-6}^{6} p(X_1 = j)\, p(X_2 = i - j) & \text{für } 7 \le i \le 12 \end{cases}$$
$$= \begin{cases} (i-1)/36 & \text{für } 2 \le i \le 7 \\ (13-i)/36 & \text{für } 7 \le i \le 12. \end{cases}$$

Man erhält z.B. als bedingte Wahrscheinlichkeiten:

$$p(S = 7|X_1 = 1) = 1/6;\ p(S = 10|X_1 = 5) = 1/6;\ p(S = 10|X_1 = 2) = 0$$
$$\text{oder auch} \qquad p(X_1 = 5|S = 10) = 1/3;\ p(X_1 = 5|S = 6) = 1/5.$$

Definition 1.5 Eine (gedächtnislose) diskrete *Informationsquelle* sendet eine Folge von (unabhängigen) Quellsymbolen, die durch Zufallsvariable $\{X_t;\ t \in \mathbb{N}_0\}$ in einem endlichen Wahrscheinlichkeitsraum $Q = (S, \mathcal{P}(S), p)$ dargestellt werden. Dabei heißt $S = \{s_1 \ldots, s_q\}$ das *Quellalphabet* und $H(p_1, \cdots, p_q)$ die Entropie der Quelle mit den *Signalwahrscheinlichkeiten* $p_1 = p(s_1), \cdots, p_q = p(s_q)$. ⌋

In vielen Fällen werden mehrere von einer Informationsquelle ausgesendete Zeichen zu Wörtern zusammengefaßt. Ein *Wort* σ ist eine Aneinanderreihung von Quellsymbolen $\sigma = s_{j_1} s_{j_2} \dots s_{j_n}$. Die Anzahl der Symbole ergibt die Länge n des Wortes. Mit der n-ten Erweiterung einer Quelle wird eine Anpassung der Bezeichnungsweise bezogen auf Wörter der Länge n vorgenommen.

Definition 1.6 Für die *n-te Erweiterung* einer Informationsquelle, die n-stellige Wörter über dem Alphabet S aussendet, wird der Wahrscheinlichkeitsraum $Q^n = \{S^n, \mathcal{P}(S^n), p^n\}$ in entsprechender Erweiterung von $Q = \{S, \mathcal{P}(S), p\}$ zugrunde gelegt. ⌋

Geht man von unabhängigen Symbolen in einem Wort aus, so erhält man als Wahrscheinlichkeit für das Auftreten eines Wortes

$$\forall\, \sigma = s_{j_1} \cdots s_{j_n} \in S^n : \qquad p(\sigma) = p(s_{j_1}) \cdot \dots \cdot p(s_{j_n}). \tag{1.10}$$

Beispiel 1.3 Q habe das Quellalphabet $S = \{s_1, s_2, s_3\}$ mit den Wahrscheinlichkeiten

$$p_1 = \frac{1}{2}, \quad p_2 = p_3 = \frac{1}{4}.$$

Dann ist die Entropie von Q

$$H(S) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\log 4 = \frac{3}{2}\text{bit}.$$

Das Quellalphabet von Q^2 ist

$$S^2 = \{\tilde{s}_1 = s_1 s_1,\ \tilde{s}_2 = s_1 s_2,\ \tilde{s}_3 = s_1 s_3,\ \tilde{s}_4 = s_2 s_1,\ \tilde{s}_5 = s_2 s_2,$$
$$\tilde{s}_6 = s_2 s_3,\ \tilde{s}_7 = s_3 s_1,\ \tilde{s}_8 = s_3 s_2,\ \tilde{s}_9 = s_3 s_3\}.$$

Die Wahrscheinlichkeiten der Wörter sind bei Unabhängigkeit ihrer Symbole

$$\tilde{p}_1 = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4};\ \tilde{p}_2 = \tilde{p}_3 = \tilde{p}_4 = \tilde{p}_7 = \frac{1}{8};\ \tilde{p}_5 = \tilde{p}_6 = \tilde{p}_8 = \tilde{p}_9 = \frac{1}{16}.$$

Die Entropie von Q^2 ist dann

$$H(S^2) = \frac{1}{4}\log 4 + 4\frac{1}{8}\log 8 + 4\frac{1}{16}\log 16 = 2\frac{3}{2}\text{bit}.$$

Es ist kein Zufall, daß die Entropie der zweiten Erweiterung gerade doppelt so groß ist wie die Entropie der Quelle. Allgemein gilt:

Satz 1.2 Für die Entropie der n-ten Erweiterung Q^n eines Wahrscheinlichkeitsraums Q gilt bei Unabhängigkeit der Symbole in einem Wort:

$$H(S^n) = n\, H(S). \tag{1.11}$$⌋

Beweis: Sei $\sigma = s_{i_1} \dots s_{i_n} \in S^n$. Dann ist

$$H(S^n) = -\sum_{\sigma \in S^n} p(\sigma) \log p(\sigma).$$

Nun ist wegen der Unabhängigkeit der Einzelsymbole $s_{i_1}, \dots, s_{i_n}$ von σ

$$\begin{aligned}
H(S^n) &= -\sum_{i_1=1}^{q} \sum_{i_2=1}^{q} \cdots \sum_{i_n=1}^{q} p(s_{i_1})p(s_{i_2}) \cdot \dots \cdot p(s_{i_n}) \log p(s_{i_1})p(s_{i_2}) \cdot \dots \cdot p(s_{i_n}) \\
&= -\sum_{i_1=1}^{q} \cdots \sum_{i_n=1}^{q} p(s_{i_1}) \cdot \dots \cdot p(s_{i_n}) \left(\log p(s_{i_1}) \cdot \dots \cdot p(s_{i_{n-1}}) + \log p(s_{i_n}) \right) \\
&= -\sum_{i_1=1}^{q} \cdots \sum_{i_{n-1}=1}^{q} p(s_{i_1}) \cdot \dots \cdot p(s_{i_{n-1}}) \log p(s_{i_1}) \cdot \dots \cdot p(s_{i_{n-1}}) \sum_{i_n=1}^{q} p(s_{i_n}) \\
&\quad - \sum_{i_1=1}^{q} \cdots \sum_{i_{n-1}=1}^{q} p(s_{i_1}) \cdot \dots \cdot p(s_{i_{n-1}}) \sum_{i_n=1}^{q} p(s_{i_n}) \log p(s_{i_n}) \\
&= H(S^{n-1}) \sum_{i_n=1}^{q} p(s_{i_n}) + \left(\sum_{i_1=1}^{q} p(s_{i_1}) \right) \cdot \dots \cdot \left(\sum_{i_{n-1}=1}^{q} p(s_{i_{n-1}}) \right) \cdot H(S) \\
&= H(S^{n-1}) + H(S).
\end{aligned}$$

Damit folgt der Satz durchvollständige Induktion. □

1.4 Markov'sche Informationsquellen

Es werden nun Quellen betrachtet, bei denen die Gedächtnislosigkeit und die damit einhergehende Unabhängigkeit der ausgesendeten Symbole nicht erfüllt ist. Stattdessen kann die Wahrscheinlichkeit, daß ein bestimmtes Folgesymbol auftritt je nach seinen Vorgängersymbolen unterschiedlich ausfallen. Beispielsweise folgt im vorliegenden Text nach einem Zeichen 'c' oder nach den beiden Zeichen 'sc' relativ häufig das Zeichen 'h', während es andere Buchstaben und Buchstabenfolgen gibt, nach denen nur selten ein 'h' erscheint.
Einen Ansatz zur Erfassung solcher Abhängigkeiten in diskreten stochastischen Prozessen bieten die Markov-Ketten:

Definition 1.7 Eine Folge $\{X_t,\ t \in \mathbb{N}_0\}$ mit endlichem Wertebereich $X_t \in \{s_1, \cdots, s_q\}$ heißt endliche *Markovkette m-ter Ordnung*, wenn

$$\begin{aligned}
&\forall\, t > m \ \text{ und } \forall\, i, j_{m+1-t}, \cdots, j_m \in \{1, \cdots, q\} : \\
&p(X_t = s_i | X_0 = s_{j_{m+1-t}}, \cdots, X_{t-m} = s_{j_1}, \cdots, X_{t-1} = s_{j_m}) \qquad (1.12) \\
&= p(X_t = s_i | X_{t-m} = s_{j_1}, \cdots, X_{t-1} = s_{j_m}).
\end{aligned}$$

Weiterhin heißt eine Markovkette m-ter Ordnung *homogen*, wenn

$$\begin{aligned} \forall\, t > m: \quad & p(X_t = s_i | X_{t-m} = s_{j_1}, \cdots, X_{t-1} = s_{j_m}) \\ & = p(X_m = s_i | X_0 = s_{j_1}, \cdots, X_{m-1} = s_{j_m}). \end{aligned} \tag{1.13}$$

Eine Markov-Kette heißt *stationär*, wenn

$$\forall\, i, j: \quad \lim_{n\to\infty} p(X_{t+n} = s_i | X_t = s_j) = \lim_{n\to\infty} p(X_{t+n} = s_i) = p(s_i). \tag{1.14}$$

Eine Informationsquelle mit Alphabet $S = \{s_1, \ldots, s_q\}$, deren Ausgabefolge die genannten Eigenschaften einer Markov-Kette hat, wird entsprechend als homogene bzw. stationäre Markov-Quelle m-ter Ordnung bezeichnet. ⌋

Für eine Markov-Quelle m-ter Ordnung ist die Abhängigkeit von der Vergangenheit des Prozesses allein aus den letzten m Zeichen abzulesen, die damit einen aktuellen Zustand der Quelle beschreiben, der sämtliche für den weiteren Verlauf des stochastischen Prozesses relevanten Informationen beinhaltet. Dies bedeutet aber keineswegs, daß Ausgabezeichen mit einem Abstand von mehr als m Zeiteinheiten voneinander unabhängig sind.
Ist die Markov-Quelle homogen, so ist die Abhängigkeit von den vorhergehenden Symbolen zeitlich invariant.

In Verallgemeinerung der angegebenen Definition wird die Markov-Quelle in der Literatur oft auch als ein *Hintergrundprozeß* eingeführt. Statt der hier vorliegenden direkten Identifikation von Ausgabezeichen und Zuständen der Markoff-Kette werden dabei in den Zuständen nach einer jeweiligen Wahrscheinlichkeitsverteilung verschiedene Symbole ausgegeben, so daß man von der ausgegeben Zeichenfolge nicht mehr sicher auf die dabei durchlaufenen Zustände zurückschließen kann.

Stationäre Quellen haben die Eigenschaft, daß Zeichen der ausgegebenen Folge mit wachsendem zeitlichem Abstand mehr und mehr voneinander unabhängig werden. Es existiert dann eine stationäre Grenzverteilung $p_i = \lim_{t\to\infty} p(X_t = s_i)$ der Wahrscheinlichkeiten für die ausgegebenen Symbole, die im Laufe der Zeit unabhängig vom Startzustand angestrebt wird. Für Markov-Quellen kann man folgende Bedingungen für die Stationarität angeben.

Definition 1.8 Jede mögliche Folge $\sigma_j = s_{j_1} \cdots s_{j_m} \in S^m$ von Symbolen kann als *Zustand* einer homogenen Markovkette m-ter Ordnung aufgefaßt werden, der die gesamte verfügbare Information über den weiteren Verlauf des stochastischen Prozesses beinhaltet.
Die *Übergangswahrscheinlichkeiten*

$$p(s_i | \sigma_j) = p(X_1 = s_i | X_{t-m} \cdots X_{t-1} = \sigma_j) \tag{1.15}$$

sind dann entscheidend für das nächste Ausgabesymbol und damit für den Folgezustand $\tilde{\sigma}_j = s_{j_2} \cdots s_{j_m} s_i$.
Zustände, welche mit Wahrscheinlichkeit 1 im weiteren Verlauf des Prozesses wieder auftreten, heißen *rekurrent*. Benötigt die Rückkehr in einen Zustand im Mittel endlich viele Schritte, so heißt der Zustand *positiv rekurrent*. Zustände heißen

periodisch, wenn eine Rückkehr nur in Schrittzahlen möglich ist, die Vielfache einer ganzen Zahl $n \geq 2$ betragen. Sonst heißt der Zustand *aperiodisch*. Kann jeder Zustand einer Markov-Quelle von jedem Zustand aus in einem oder mehreren Übergängen erreicht werden, so heißt die Markov-Quelle *irreduzibel*, andernfalls *reduzibel*. Eine *irreduzible Markov-Quelle*, deren Zustände positiv rekurrent und aperiodisch sind, heißt *ergodisch*. ⌋

Satz 1.3 Für ergodische Markov-Quellen existieren die *stationären Zustandswahrscheinlichkeiten*

$$p(\sigma_j) = p(s_{j_1} \dots s_{j_m}) = \lim_{t \to \infty} p(X_{t-m} = s_{j_1}, \cdots, X_{t-1} = s_{j_m}). \tag{1.16}$$

Sie sind Lösung der Übergangsgleichungen

$$p(s_{j_1} \dots s_{j_m}) = \sum_{i=1}^{q} p(s_{j_m} | s_i s_{j_1} \dots s_{j_{m-1}})\, p(s_i s_{j_1} \dots s_{j_{m-1}}) \tag{1.17}$$

für alle $\sigma_j = s_{j_1} \dots s_{j_m} \in S^m$. Durch dieses lineare, homogene Gleichungssystem und die Normierungsbedingung $\sum_{\sigma_j \in S^m} p(\sigma_j) = 1$ sind die stationären Zustandswahrscheinlichkeiten eindeutig bestimmt.
Für die hier betrachteten endlichen Markov-Ketten sind insbesondere alle rekurrenten Zustände gleichzeitig auch positiv rekurrent. ⌋

Ein Beweis dieses für endliche, ergodische Markovketten grundlegenden Satzes findet sich z.B. in [Kohl77].

Beispiel 1.4

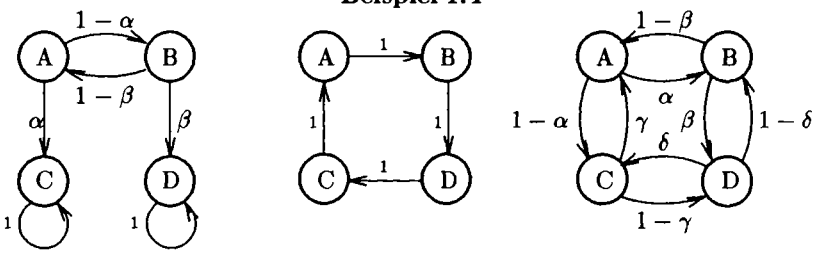

Abbildung 1.1: Übergangsdiagramme von reduziblen und periodischen Markov-Ketten

In der Abbildung 1.1 sind die Übergangsgraphen einer reduziblen Markov-Kette sowie von zwei periodischen Markov-Ketten erster Ordnung mit Alphabet $S = \{A,B,C,D\}$ dargestellt. Die Zustände werden darin als Knoten und die Zustandsübergänge mit einer von Null verschiedenen Wahrscheinlichkeit als Kanten notiert. Eine Kantenmarkierungen gibt die Übergangswahrscheinlichkeit vom

Start- zum Zielzustand an, wobei hier $0 < \alpha, \beta, \gamma, \delta < 1$ gelten soll. Im ersten Fall sind von den Zuständen C und D keine anderen Zustände erreichbar, so daß die Kette reduzibel ist. In den beiden periodischen Ketten ist dagegen eine Rückkehr in denselben Zustand nur möglich, wenn die Anzahl der Übergänge ein Vielfaches von 4 bzw. 2 ist.

Beispiel 1.5 Es sei $S = \{0, 1\}$ und $m = 2$.
Die Übergangswahrscheinlichkeiten seien gegeben durch:

$$p(0|00) = 0.5;\ p(0|01) = 0.3;\ p(0|10) = 0.4;\ p(0|11) = 0.2;$$
$$p(1|00) = 0.5;\ p(1|01) = 0.7;\ p(1|10) = 0.6;\ p(1|11) = 0.8.$$

Die damit beschriebene Markov-Quelle ist ergodisch.
Für die stationären Zustandswahrscheinlichkeiten gilt

$$\begin{aligned} p(00) &= p(0|00)\,p(00) + p(0|10)\,p(10) \\ p(01) &= p(1|00)\,p(00) + p(1|10)\,p(10) \\ p(10) &= p(0|01)\,p(01) + p(0|11)\,p(11) \\ p(11) &= p(1|01)\,p(01) + p(1|11)\,p(11) \end{aligned}$$

Durch Einsetzen der Übergangswahrscheinlichkeiten

$$\begin{aligned} p(00) &= 0.5p(00) + 0.4p(10) \\ p(01) &= 0.5p(00) + 0.6p(10) \\ p(10) &= 0.3p(01) + 0.2p(11) \\ p(11) &= 0.7p(01) + 0.8p(11) \end{aligned}$$

und unter Berücksichtigung der Normierungsbedingung

$$p(00) + p(01) + p(10) + p(11) = 1$$

erhält man daraus die Lösung

$$p(00) = \frac{8}{63};\ p(01) = \frac{10}{63};\ p(10) = \frac{10}{63};\ p(11) = \frac{35}{63}.$$

Beispiel 1.6 Sei $S = \{a, b, c\}$; und $m = 3$. Es gibt $q^m = 3^3 = 27$ Zustände und es sind $q^{m+1} = 81$ Übergänge möglich. Die Übergangsgleichungen sind von der Form:

$$p(x_1x_2x_3) = p(x_3|ax_1x_2)\,p(ax_1x_2) + p(x_3|bx_1x_2)\,p(bx_1x_2) + p(x_3|cx_1x_2)\,p(cx_1x_2).$$

Satz 1.4 Für ergodische Markov-Quellen m-ter Ordnung erhält man die Signalwahrscheinlichkeiten $p(s_i)$ durch Summation der stationären Wahrscheinlichkeiten aller Zustände, die s_i z.B. als letztes Symbol enthalten:

$$p(s_i) = \sum_{\sigma_k \in S^{m-1}} p(\sigma_k s_i) = \sum_{j_1,\ldots,j_{m-1}=1}^{q} p(s_{j_1} s_{j_2} \ldots s_{j_{m-1}} s_i). \qquad (1.18)$$

Beispiel 1.7 Für die Markov-Quelle aus Beispiel 1.5 erhält man:

$$p(0) = p(10) + p(00) = \frac{2}{7},$$
$$p(1) = p(01) + p(11) = \frac{5}{7}.$$

Im Beispiel 1.6 gilt für $p(a)$ und entsprechend für $p(b)$ und $p(c)$:

$$\begin{aligned} p(a) =& p(aaa) + p(aba) + p(aca) + p(baa) + p(bba) \\ &+ p(bca) + p(caa) + p(cba) + p(cca). \end{aligned}$$

1.5 Die Entropie von Markov-Quellen

Auch hinsichtlich des Informationsgehalts und der Entropie einer Quelle ist die Abhängigkeit eines Symbols von den vorangegangenen Symbolen zu berücksichtigen.

Definition 1.9 Der *bedingte Informationsgehalt* $I(s_i|\sigma_j)$ und die *bedingte Entropie* $H(S|\sigma_j)$ einer Markov-Quelle m-ter Ordnung sind gegeben durch

$$H(S|\sigma_j) = \sum_{i=1}^{q} p(s_i|\sigma_j)\, I(s_i|\sigma_j); \quad \text{und} \quad I(s_i|\sigma_j) = -\log p(s_i|\sigma_j). \tag{1.19}$$

Dabei sei $\sigma_j = s_{j_1} \dots s_{j_m} \in S^m$ die Folge der m vorangehenden Symbole. ⌋

Als Entropie einer ergodischen Markov-Quelle m-ter Ordnung erhält man:

$$H(S) = \sum_{\sigma_j \in S^m} p(\sigma_j)\, H(S|\sigma_j). \tag{1.20}$$

Die Berechnung der Entropie mit den eben genannten Beziehungen ist auch für irreduzible periodische Markov-Quellen durchführbar. Die Irreduzibilität einer Markov-Kette gewährleistet die Existenz einer eindeutigen Lösung von Zustandswahrscheinlichkeiten $p(\sigma_i)$, die die Übergangsgleichungen und die Normierungsbedingung erfüllen. Allerdings sind die Voraussetzungen für Stationarität bei periodischen Markov-Ketten nicht gegeben.

Beispiel 1.8 Die Markov-Quelle zweiter Ordnung aus Beispiel 1.5 hat demgemäß die Entropie:

$$\begin{aligned} H(S) =& -p(00)\,(p(0|00)\log p(0|00) + p(1|00)\log p(1|00)) \\ &- p(01)\,(p(0|01)\log p(0|01) + p(1|01)\log p(1|01)) \\ &- p(10)\,(p(0|10)\log p(0|10) + p(1|10)\log p(1|10)) \\ &- p(11)\,(p(0|11)\log p(0|11) + p(1|11)\log p(1|11)) \\ =& \frac{1}{63}\,(8\,H(0.5) + 10\,H(0.3) + 10\,H(0.4) + 35\,H(0.2)) \\ \approx& 0.822\,\text{bit}. \end{aligned}$$

Definition 1.10 Sei Q eine Markov-Quelle m-ter Ordnung mit Quellalphabet $S = \{s_1, \ldots, s_q\}$ und mit Signalwahrscheinlichkeiten $p(s_i)$, $i = 1, \ldots, q$.
Dann heißt die gedächtnislose Quelle $\overline{Q}$ mit übereinstimmendem Alphabet $\overline{S} = S$ und Signalwahrscheinlichkeiten $p(s_i)$ die zur Quelle Q *adjungierte Quelle.* ⌋

Die Entropie der adjungierten Quelle ist also

$$H(\overline{S}) = -\sum_{i=1}^{q} p(s_i) \log p(s_i).$$

Für die Markov-Quelle aus Beispiel 1.5 ist $p(0) = 2/7$, $p(1) = 5/7$ und somit $H(\overline{S}) \approx 0.863$ bit.

Wegen ihrer Gedächtnislosigkeit verfügt die adjungierte Quelle nicht über die bei der zugehörigen Markov-Quelle anfallende Vorinformation über das nächste Symbol. Demzufolge ist die Entropie als Maß für den erwarteten Informationsgehalt für die adjungierte Quelle größer.

Satz 1.5 Ist $H(S)$ die Entropie einer Markov-Quelle m-ter Ordnung und $H(\overline{S})$ die Entropie der adjungierten Quelle, so gilt:

$$H(S) \leq H(\overline{S}). \tag{1.21}$$ ⌋

Auf einen Beweis des Satzes wird hier verzichtet.

Bei der Untersuchung von Sprachstrukturen oder z.B. auch bei der Analyse von Tonfolgen in Musikstücken kann die Theorie der Markov-Informationsquellen eine nützliche Hilfestellung leisten.
Um den mittleren Informationsgehalt eines Buchstabens z.B. in deutsch- oder englisch-sprachigen Texten zu ermitteln, kann man aus einer möglichst großen, repräsentativen Stichprobe solcher Texte Statistiken über die Wahrscheinlichkeiten für einzelne Buchstaben und für Kombinationen aus m ($m \in \mathbb{N}$) aufeinanderfolgenden Buchstaben erstellen. Die damit bestimmten Zustandswahrscheinlichkeiten für eine angenommene Markov-Quelle m-ter Ordnung hängen nicht nur von der Sprache ab, sondern variieren z.B. je nach dem persönlichen Stil eines Autors. So läßt sich beispielsweise Aufschluß darüber gewinnen, ob ein aufgetauchter Text einem bestimmten Schriftsteller zugeordnet werden kann.
Umgekehrt kann man Näherungen an eine Sprachstruktur erreichen, indem man Buchstabenfolgen durch Markov-Quellen m-ter Ordnung mit Hilfe eines Zufallsgenerators erzeugt. Der Aufwand bei der Darstellung einer Quelle m-ter Ordnung wächst allerdings exponentiell mit m, so daß es z.B. bereits mehr als eine Milliarde verschiedene Buchstabenfolgen der Länge $m = 7$ gibt. Syntaktische oder gar semantische Strukturen eines Textes bleiben bei dieser Vorgehensweise unberücksichtigt.
Für die mit $H^{(m)}$ bezeichnete Entropie eines Buchstabens in einem deutschsprachigen Text unter Berücksichtigung der letzten m Buchstaben erhält man ($m = 0$