Forschungszentrum Karlsruhe

Technik und Umwelt

Wissenschaftliche Berichte

FZKA 6390

Informationstheoretische Betrachtungen zur Strukturermittlung

Uwe Getzlaff

Institut für Angewandte Informatik

Arbeitsbericht zum DFG-Fördervorhaben

Modellstrukturierung von sicherheitsrelevanten technischen Systemen mittels tensororientierter Mustererkennung und Genetischen Algorithmen (BR 1303/3-1)

Forschungszentrum Karlsruhe GmbH, Karlsruhe

2000

Forschungszentrum Karlsruhe GmbH
Postfach 3640, 76021 Karlsruhe

Mitglied der Hermann von Helmholtz-Gemeinschaft
Deutscher Forschungszentren (HGF)

ISSN 0947-8620

Zusammenfassung

Dies ist ein Arbeitsbericht zum DFG-Fördervorhaben „Modellstrukturierung von sicherheitsrelevanten technischen Systemen mittels tensororientierter Mustererkennung und Genetischen Algorithmen“ (BR 1303/3-1). In vorausgegangenen Arbeiten wurde die Strukturierung von Modellen aus informationstheoretischer Sicht betrachtet. Dabei wird die Modellbildung als Informationsübertragung von einem Objekt auf ein Modell verstanden.

In diesem Bericht wird untersucht:

- Wie kann die Information optimal kodiert werden? Welche Art der Kodierung von Merkmalen ist für die Modellbildung optimal (binär, ternär, binär kodierte Ternärzahl)?
- Ist aus informationstheoretischer Sicht eine Modellbildung ohne Strukturannahme möglich?
- Welche Auswirkungen hat eine beliebige Struktur auf die Parameter eines Modells?

Für vollständige Merkmale wird nachgewiesen, dass die Information optimal mit dem Dualsystem kodiert wird. Dabei wird, entgegen der in der Literatur üblichen Beweisführung, der duale Logarithmus im Beweis weder angesetzt noch verwendet. Für unvollständige Merkmale wird die Basis e als optimal für die Entscheidungsstufung nachgewiesen. Auch hier wurde vermieden, den natürlichen Logarithmus beim Beweis anzusetzen, was sonst in der Literatur üblich ist. An Beispielen wird gezeigt, dass Entscheidungsstrategien für unvollständige Merkmale und mit nicht ganzzahliger Stufung für die Modellbildung denkbar, aber wenig praktikabel sind.

Für eine endliche Menge an Information, die zur Modellbildung verfügbar ist, wird gezeigt, dass der Wertebereich eines Parameters endlich sein muss, damit der Fehler des Parameters endlich bleiben kann und dass der Wertebereich eines Parameters die Null ausschließen muss, damit der relative Fehler kleiner als Eins bleiben kann.

Für den Fall, dass die Struktur eines Modells nicht durch die Festlegung auf eine Modellklasse eingeschränkt ist, wird gezeigt, dass eine Struktur mit einem oder mehreren Parametern durch eine Struktur ohne diese Parameter ersetzt werden kann. Daraus folgt, dass Gütemaße für solche Modelle nicht nur die Anzahl der Parameter sondern auch die Komplexität der Struktur bewerten müssen.

Abstract

On the Determination of Structures - A Theoretical Approach of Information Science

The report describes two proofs as a basis for optimal coding of information without using any logarithm in the formulation of the problem as is usually found in the literature. We proof also that, if we have finite information, than the range of values has to be finite to get a finite error of parameter and that the range of values must exclude zero to get a relative error less than one. Furthermore, we show an equivalence between structure and parameters. That means, each parameter can be replaced by a structure. Construction of a model is only sensible after reducing the considered class of the model to a finite number of structures and with prior knowledge about the range of parameters.

Inhaltsverzeichnis

1 Einleitung **1**

2 Optimale Darstellung der Information **2**

2.1 Optimale Darstellung der Information bei vollständigen Merkmalen 5

2.2 Optimale Darstellung der Information bei unvollständigen Merkmalen 6

2.3 Konstruktion von nicht ganzzahligen, unvollständigen Tests auf diskreten Elementen 8

3 Parameterfehler bei endlicher Information **15**

4 Äquivalenz von Struktur und Parametern **17**

5 Schlußfolgerung und Zusammenfassung **18**

Abbildungsverzeichnis

1 Entscheidung bei nicht erkanntem Fehlen von Zeichen 8

2 Entscheidung bei erkanntem Fehlen von Zeichen 10

3 Darstellung der doppelten Summe 11

4 Entscheidung für $z(n \to \infty) = \mathrm{e}$ 14

5 Äquivalenz zwischen einer Struktur und dem Parameterwert 1 17

1 Einleitung

Der vorliegende Bericht wurde im Rahmen des DFG-Fördervorhabens „Modellstrukturierung von sicherheitsrelevanten technischen Systemen mittels tensororientierter Mustererkennung und Genetischen Algorithmen“ angefertigt. Ziel des Forschungsprojektes ist die Entwicklung eines Verfahrens zur Strukturermittlung. In vorausgehenden Berichten [GG99, Hop98a, Hop98b] wurde die Strukturermittlung von informationstheoretischer Seite aus betrachtet. Dabei wird die Modellbildung als Informationsübertragung von einem Objekt auf ein Modell verstanden. Speziell die tensororientierter Mustererkennung [Hop94, Hop96] erfordert eine Kodierung der Information. In diesem Zusammenhang sind folgende Fragestellungen aufgetreten:

- Wie kann die Information optimal kodiert werden?
- Wieviel Information muss das Datenmaterial von einem technischen System enthalten, damit die Struktur dieses Systems bestimmt werden kann?
- Wieviel Information muss das Datenmaterial von einem technischen System enthalten, damit ein Modellparameter ohne Einschränkung der Modellstruktur bestimmt werden kann?

Diese Fragestellungen werden im vorliegenden Bericht behandelt.

Für die optimale Kodierung von Information werden in der Literatur [ME69, PS80, You75, Zem59] die Basis 2 und die Basis e angegeben. Die Basis 2 führt auf die duale Kodierung und die e nächste ganzzahlige Basis 3 auf eine ternäre Codierung (siehe z. B. [Hop98a]). Die verschiedenen Basen sind verknüpft mit den Begriffen vollständige und unvollständige Entscheidung, bzw. vollständige und unvollständige Merkmale. Die beiden Begriffe werden weiter unten genauer betrachtet. In den Herleitungen der Basen 2 und e werden in der Literatur (willkürlich) die Logarithmen zur Basis 2 bzw. e angesetzt. Die Herleitungen in diesem Beitrag kommen ganz ohne den Ansatz eines Logarithmus oder irgend einer Basis aus.

Für unvollständige Entscheidungen ist weiterhin interessant:

- Wie kann ein unvollständiger Test aussehen?
- Wie kann eine nicht disjunkte Entscheidung aussehen?
- Wie kann ein nicht ganzzahliger Test aussehen?
- Wie kann eine nicht ganzzahlige Stufung von Entscheidungen aussehen?

Dazu wird in Abschnitt 2.3 gezeigt, dass nicht ganzzahlige unvollständige Tests konstruierbar sind.

Wenn die Information über ein System kodiert wurde, kann deren Betrag ermittelt werden. Da ermittelt werden kann, wie viel Information die Daten über das System enthalten, ist interessant, wie viel Information zur Struktur- und Parameteridentifikation notwendig ist. Im Abschnitt 3 wird gezeigt, dass mit einem endlichen Code weder ein endlicher absoluter Fehler noch ein relativer Fehler < 1 garantiert werden kann, wenn der Wertebereich nicht eingeschränkt wird. Das heißt, dass a-priori Informationen über den Wertebereich der Parameter vorhanden sein müssen.

Des Weiteren wird in Abschnitt 4 gezeigt, dass jeder Parameter beliebig genau durch eine Struktur ersetzt werden kann. Daraus folgt, dass die Komplexität einer Struktur nicht nur durch die Anzahl der in ihr vorkommenden Parameter bewertet werden darf.

2 Optimale Darstellung der Information

Die natürlichen (und technischen) Systeme sind überwiegend analog. Um die Information über ein System zu speichern und zu verarbeiten, werden die Messungen zeit- und wertdiskretisiert. So wird die Zeit z. B. als 1998-07-15,08:30:00 notiert. Die Basis ist dekadisch für die ersten vier Stellen. Dann ist die Basis 12, 31, 24, 60 und 60. Die Information über eine bestimmte Sekunde (im Intervall beginnend mit der Zeitrechnung bis 9999-12-31,23:59:59) entspricht der Information über neun Entscheidungen:

1. eins aus 10
2. eins aus 10
3. eins aus 10
4. eins aus 10
5. eins aus 12
6. eins aus 31
7. eins aus 24
8. eins aus 60
9. eins aus 60 .

Der Bezeichnungsaufwand dafür beträgt $4 * 10 + 12 + 31 + 24 + 60 + 60 = 227$ Symbole. Die gleiche Sekunde kann aber auch als 63.028.571.400 kodiert werden[1]. Diese Bezeichnung, wenn auch nicht sehr anschaulich, kommt mit $4 + 11 * 10 = 114 \mathrel{\widehat{=}} 50\%$ Symbolen aus[2]. In diesem Abschnitt wird untersucht, wie ein Wert (oder Sachverhalt) kodiert werden muss, damit der Bezeichnungsaufwand (die Anzahl der notwendigen Symbole) minimal wird.

Ein Wert oder Sachverhalt kann z. B. ein kleinstes unterscheidbares Intervall eines analogen Messgerätes, eine Sekunde in 10.000 Jahren, ein bestimmter Zustand eines Objektes oder irgend ein anderes Element einer Menge sein. Die Menge heißt Grundgesamtheit Ω, und die Anzahl der Elemente der Menge heißt Elementarvorrat C.

Jedes Element der Menge muss durch seine Merkmale identifiziert werden, z. B. durch die Nummer des Elements. Wenn alle Elemente von 1 bis C durchnummeriert werden und die Nummer des Elementes als **ein** Symbol angegeben wird, erfordert das C verschiedene Symbole. Im obigen (Sekunden-) Beispiel würde das eine Vergrößerung des Bezeichnungsaufwandes um mehr als acht Zehnerpotenzen bedeuten. Offensichtlich ist es sinnvoll, mehrere Merkmale b_i , $i = 1, ..., a$ zu nutzen, um die Elemente der Menge zu identifizieren. Die Anzahl der verschiedenen Werte, die ein Merkmal b_i annehmen kann, ist $B_i = N_{b_i}$. Die Auswahl eines Elementes ist dann eine Variation von a Tests mit B_i Alternativen.

Für die optimale Kodierung von Information werden die Basis 2 und die Basis e angegeben. Die verschiedenen Basen sind verknüpft mit den Begriffen vollständige und unvollständige Entscheidung, bzw. vollständiges und unvollständiges Merkmal.

Eine vollständige Entscheidung kann getroffen werden, wenn ein vollständiges Merkmal zugrunde liegt. Ein Merkmal ist vollständig, wenn es genau einen Wert aus einer endlichen Anzahl diskreter Werte annehmen kann. Ein solches Merkmal ist z. B. die Farbe des Pixels eines LCD-Farbmonitors. Sie kann entweder rot, grün oder blau sein. Wenn bekannt ist, dass ein Pixel defekt ist, stellt sich die Frage nach der Farbe des defekten Pixels. Diese Frage ist mit einer Entscheidung für rot, grün oder blau beantwortet. Der erste Test wird mit einem roten Testbild durchgeführt. Das Ergebnis sei fehlerfrei. Der zweite Test mit einem grünen Testbild sei

[1] Sekunden seit Christi Geburt, Kalendersprünge und Schaltjahre unberücksichtigt

[2] 10.000 Jahre entsprechen 315.360.000.000 Sekunden

ebenfalls fehlerfrei. Der dritte Test mit einem blauen Testbild ist überflüssig, da er nur liefern kann, was schon bekannt ist: Der defekte Pixel ist blau. Die Vollständigkeit eines Merkmals, das genau einen aus n Werten annimmt, führt dazu, dass höchstens $n-1$ Tests durchgeführt werden müssen, um den Wert zu ermitteln. Wenn der Glöckner nach dem Stundenschlag befragt wird, wie oft er geläutet hat — „Einmal?“ „Nein!“; „Zweimal?“ „Nein!“ … „Elfmal?“ „Nein!“ — kann nach dem elften „Nein“ die Befragung abgebrochen werden, da es ein dreizehntes mal aller Erfahrung nach nicht schlägt. (Wer trotzdem weiter fragt, will nur endlich doch noch ein „Ja!“ hören.)

Eine unvollständige Entscheidung folgt z. B. auf die Frage nach der Farbe einer Blume. Der Wertevorrat dieses Merkmals kann unendlich sein, die Werte müssen nicht diskret sein und selbst die Forderung „genau ein Wert“ ist problematisch. Letzteres Problem besteht nicht nur weil die Blume bunt sein kann, sondern auch dann noch, wenn die Farbwerte auf eine vorgegebene endliche Menge beschränkt sind. So führt z. B. ein positiver Test auf die Farbe Rot noch nicht zu einer zuverlässigen Entscheidung, da ein späterer positiver Test auf die Farbe Weinrot die voreilige Entscheidung revidieren würde. Erst wenn alle n möglichen Werte in n Versuchen abgetestet wurden kann die Entscheidung gefällt werden.

Entscheidungen können in Stufen gefällt werden. Herr X betritt in Eile den Bahnhof und muss sich für eine sinnvolle Handlungssequenz entscheiden. Das erste Merkmal der Situation ist die Anwesenheit des Zuges mit den beiden möglichen Werten „Zug da“ und „Zug nicht da“. Das zweite Merkmal der Situation ist die Möglichkeit noch einzusteigen. Dieses zweite Merkmal wird in Abhängigkeit vom Wert des ersten Merkmals gegebenenfalls mit unterschiedlichen Tests bewertet: Für den Fall „Zug da“, direkt mit dem Versuch einzusteigen und für den Fall „Zug nicht da“, mit einem Test ob der Zug noch kommt. In beiden Fällen hat das zweite Merkmal die beiden möglichen Werte „einsteigen noch möglich“ und „einsteigen nicht mehr möglich“. Die Entscheidung wird in zwei Stufen mit jeweils zwei Alternativen gefällt. Wenn die Anzahl der Alternativen (die Anzahl der möglichen Werte des Merkmals) in allen Stufen zwei ist, heißt der Vorgang mehrstufige Entscheidung mit der Basis 2. Die insgesamt vier Alternativen in diesem Beispiel sind

1. „warten“,
2. „einsteigen“,
3. „Schaffner suchen“, (Die Situation „Zug da“ und „einsteigen nicht mehr möglich“ will Herr X zu seinem Gunsten ändern.) und
4. „Taxi oder Hotel nehmen“. (Herr X war wie immer auf den letzten Zug angewiesen.)

Bei einer Entscheidung für eine aus vier verschiedenen Alternativen beträgt der Elementarvorrat 4.

In den Herleitungen der Basen 2 und e wird in der Literatur stillschweigend vorausgesetzt, dass eine gleichmäßige Stufung der Entscheidungen und damit ein und die selbe Basis für alle Stufen optimal ist. Weiterhin werden (willkürlich) die Logarithmen zur Basis 2 bzw. e angesetzt, mit dem Argument, dass die Wahl der Basis nur ein Maßstabsfaktor sei. Schlüssiger währen die Beweise, wenn eine beliebige Basis x angesetzt und nachgewiesen wird, dass das Beweisergebnis unabhängig von dem „Maßstabsfaktor“ ist. Eine andere Möglichkeit ist, gar keinen Logarithmus anzusetzen.

Um die Frage nach der optimalen Basis zu klären, wurden zwei Beweise ausgearbeitet, die ganz ohne den Ansatz irgend eines Logarithmus auskommen. Auch, dass eine gleichmäßige Stufung optimal, ist wurde nachgewiesen.

Jedes Merkmal stellt eine Menge B_i aus B_i Elementen dar. Im Folgenden wird vorausgesetzt, dass die Merkmale unabhängig und die Mengen der Merkmale inhaltlich disjunkt sind[3], so dass die Merkmale einzeln

[3]Zwei Merkmale dürfen durchaus formal die gleichen Elemente enthalten (z. B. Merkmal „Länge“, Element „2m“ und Merkmal „Breite“, Element „2m“), solange die Mengen inhaltlich verschieden sind. Dagegen ist Merkmal „Länge in Meter“, Element

betrachtet werden können. Des Weiteren wird vorausgesetzt, dass ein Merkmal (bzw. Attribut) nur höchstens einen aus B_i Werten annehmen kann[4].

Die Anzahl der Elemente der Vereinigungsmenge aller B_i (Mengen der Merkmale) heißt Merkmalsvorrat und ist

$$B_V = \sum_i B_i \quad . \tag{1}$$

Zur Darstellung der Information müssen, wie oben erläutert, zwei Fälle unterschieden werden:

- vollständige Merkmale und
- unvollständige Merkmale.

Definition 2.1 Ein Merkmal ist vollständig, wenn für jedes beliebige Element des Elementarvorrates das Merkmal genau einen Wert aus der Menge des Merkmals annimmt.

$$b_i(x) \in \mathsf{B}_i = \{b_1, b_2, b_3, ...b_{B_i}\} \quad , \quad N_{b_i(x)} = 1 \quad \forall x \in \Omega \tag{2}$$

Ein Merkmal ist unvollständig, wenn es nicht vollständig ist. •

Die Vollständigkeit eines Merkmals impliziert, dass die Information über $B_i - 1$ Werte ausreicht, denn es gilt

$$b_i(x) \in \{b_1, b_2, b_3, ...b_{B_i-1}\} \quad \longleftrightarrow \quad b_i(x) \neq b_{B_i} \tag{3}$$

$$b_i(x) \notin \{b_1, b_2, b_3, ...b_{B_i-1}\} \quad \longleftrightarrow \quad b_i(x) = b_{B_i} \quad , \tag{4}$$

d. h. es müssen nur $B_i - 1$ Tests durchgeführt werden, wenn B_i Alternativen bestehen.

Die Anzahl von Tests $D_\sqcup$, die maximal notwendig sind, um den Wert eines Merkmals zu bestimmen, beträgt im Fall vollständiger Merkmale

$$D_{[]} = \sum_i B_i - 1 \tag{5}$$

und im Fall unvollständiger Merkmale

$$D_{()} = B_V = \sum_i B_i \quad . \tag{6}$$

Der maximal identifizierbare Elementarvorrat C ist durch die Variation der Merkmale bestimmt

$$C = \prod_i B_i \quad . \tag{7}$$

Die optimale Darstellung der Information ist diejenige, bei der mit minimaler Anzahl von Tests $D_\sqcup$ ein Element aus einer maximalen Anzahl von Elementen C einer Grundgesamtheit identifiziert werden kann. Die Basis für eine solche optimale Darstellung wird im Folgenden für vollständige Merkmale und danach für unvollständige Merkmale hergeleitet. Die hier dargestellten Herleitungen sind etwas umfangreicher als die aus der Literatur bekannten, dafür vermeiden sie die Logarithmen ld und ln die das Ergebnis darstellen, schon im Ansatz vorzugeben.

„2“ und Merkmal „Länge in Millimeter“, Element „2000“ nicht erlaubt.

[4] Das ist keine Einschränkung der Allgemeinheit. Kann „ein Merkmal“ gleichzeitig n Werte annehmen, muss es in n Merkmale aufgeteilt werden.

2.1 Optimale Darstellung der Information bei vollständigen Merkmalen

Vollständige Merkmale heißt, ein Merkmal führt **immer** eine Entscheidung herbei. Die Anzahl der Tests in einer Stufe beträgt

$$D_i = B_i - 1 \quad . \tag{8}$$

Die Information ist dann optimal dargestellt, wenn mit einer minimalen Anzahl von Tests ein Element aus einem maximalen Elementarvorrat identifiziert werden kann.

$$\frac{D_{[]}}{C} = \frac{\sum_i B_i - 1}{\prod_i B_i} = \frac{\sum_i D_i}{\prod_i (D_i + 1)} \quad \Longrightarrow \quad \min \tag{9}$$

Der Testaufwand $D_{[]} = \sum_i D_i$ sei fest. Wann wird der identifizierbare Elementarvorrat $\prod_i D_i + 1$ maximal? Zuerst wird die Aufteilung der Variabilität der Entscheidungen und dann die Aufteilung des Testaufwands auf eine unterschiedliche Anzahl von Entscheidungsstufen a betrachtet.

Wenn die Anzahl der Tests zweier beliebiger Stufen so variiert, dass der Testaufwand gleich bleibt, dann ist die Zielfunktion

$$\frac{\sum_{i=1}^{a-2} D_i + (D_{a-1} + x) + (D_a - x)}{\prod_{i=1}^{a-2} (D_i + 1) \cdot (D_{a-1} + x + 1) \cdot (D_a - x + 1)} \quad \Longrightarrow \quad \min \quad . \tag{10}$$

Bei variablem x sind die Summe im Zähler $\sum_{i=1}^{a-2} D_i + (D_{a-1} + x) + (D_a - x)$ und der erste Faktor des Produktes im Nenner $\prod_{i=1}^{a-2}(D_i + 1)$ konstant.

$$\hookrightarrow \quad f(x) = (D_{a-1} + x + 1) \cdot (D_a - x + 1) \quad \Longrightarrow \quad \max_x \tag{11}$$

Da über die Zuordnung der Entscheidungen zu i nichts vorausgesetzt wurde, die Reihenfolge der Indizes also beliebig ist, reicht es völlig aus zwei beliebige Stufen zu betrachten, und das Ergebnis auf alle Stufen zu verallgemeinern. Aus der Ableitungen $f'(x) = -2x + D_a - D_{a-1} = 0$ ergibt sich

$$x = 1/2(D_a - D_{a-1}) \tag{12}$$

und wegen $f''(x) = -2 < 0$ wird $f(x)$ maximal wenn $(D_{a-1} + x + 1) = (D_a - x + 1)$. Daraus folgt, dass die Variabilität auf allen Entscheidungsstufen gleich sein muss.

$$\frac{D_{[]}}{C} = \min | D_i = \text{ const} \tag{13}$$

Nachdem festgestellt wurde, dass die Aufteilung der Tests auf alle Stufen möglichst gleichmäßig sein muss, wird nun die optimale Anzahl von Stufen betrachten. Bei festem Testaufwand und gleicher Variabilität auf allen Entscheidungsstufen gilt

$$\frac{\sum_i D_i}{\prod_i (D_i + 1)} = \frac{\sum_{i=1}^{a} D_{[]}/a}{\prod_{i=1}^{a} (D_{[]}/a + 1)} \tag{14}$$

für die Aufteilung des Testaufwands auf eine unterschiedliche Anzahl a von Stufen. Wegen des bezüglich a konstanten Nenners folgt als Zielfunktion

$$f(a) = \prod_{i=1}^{a} (D_{[]}/a + 1) = \frac{(D_{[]} + a)^a}{a^a} \quad \Longrightarrow \quad \max_a \tag{15}$$

Da mit positivem Testaufwand $D_{[]}$ der Nenner $(D_{[]} + a)^a$ für steigendes a immer schneller wächst als der Zähler a^a ist $a \to \infty$ das Supremum von $f(a)$.

Daraus folgt der

Satz 2.1 Wenn, bei vollständigen Merkmalen, der Testaufwand auf eine größtmögliche Anzahl von Entscheidungsstufen mit einer kleinstmöglichen Anzahl von Tests je Stufe aufgeteilt wird, kann mit einer minimalen Anzahl von Tests (minimaler Testaufwand) ein Elemente aus einer maximalen Anzahl von Elementen (maximaler Elementarvorrat) ausgewählt werden. •

Für ganzzahlige Tests ist ein Test je Stufe $D = 1$ die kleinstmögliche Anzahl. Das führt mit $B = D + 1 = 2$ auf das bekannte Binärsystem, bei dem mit $a = D_{[]}$ Entscheidungsstufen $C = 2^a$ Elemente bestimmt werden können. Zu bemerken ist, dass die Basis 2 als optimale Basis nicht direkt aus dem Beweis folgt, sondern sich erst aus der Beschränkung auf positive ganzzahlige Tests ergibt. Auch bei dem Beweis in [Zem59], führt erst diese Einschränkung zum gewünschten Ergebnis.

Das gleiche Resultat entsteht mit variablem Testaufwand $D_{[]}$ und festem Elementarvorrat C.

Die Kodierung im Dualsystem mit fester Kodelänge ist nur optimal, wenn alle Merkmale gleichwahrscheinlich sind und die Anzahl der Merkmale eine Potenz von 2 ist. Ist das nicht der Fall, können mit variablen Kodelängen Annäherungen an das Optimum gefunden werden. Dieses Vorgehen entspricht einer nicht ganzzahligen Stufung nicht ganzzahliger Tests, da einige Entscheidungsstufen und damit auch einige Tests nur mit einer bestimmten Wahrscheinlichkeit durchgeführt werden.

2.2 Optimale Darstellung der Information bei unvollständigen Merkmalen

Unvollständige Merkmale heißt, ein Merkmal führt **nicht immer** eine Entscheidung herbei. Aus $b_i(x) \notin \{b_1, b_2, b_3, ...b_{B_i-1}\}$ kann **nicht** gefolgert werden, ob $b_i(x) = b_{B_i}$ ist. Damit ist die Anzahl der Tests gleich der Anzahl der Merkmale.

$$D_i = B_i \tag{16}$$

$$\frac{D_{()}}{C} = \frac{\sum_i B_i}{\prod_i B_i} = \frac{\sum_i D_i}{\prod_i D_i} \Longrightarrow \min \tag{17}$$

Für die Aufteilung des Testaufwands entsteht analog zum vorigen Abschnitt das gleiche Ergebnis. Es gilt allgemein der

Satz 2.2 Wenn der Testaufwand auf eine gleiche Anzahl von Tests je Entscheidungsstufe aufgeteilt wird, kann mit einer minimalen Anzahl von Tests (minimaler Testaufwand) ein Elemente aus einer maximalen Anzahl von Elementen (maximaler Elementarvorrat) ausgewählt werden, d. h. die Variabilität auf allen Entscheidungsstufen muss gleich sein.

$$\frac{D_{[]}}{C} = \min | D_i = \text{ const} \tag{18}$$

•

Für die Aufteilung des Testaufwands auf eine unterschiedliche Anzahl von Entscheidungsstufen a ist das Ergebnis anders. Der Elementarvorrat sei fest. $C = \prod_i B_i$. Die Zielfunktion ist

$$\frac{D_{()}}{C} = \frac{\sum_{i=1}^{a} \sqrt[a]{C}}{\prod_{i=1}^{a} \sqrt[a]{C}} \quad \Longrightarrow \quad \min \quad . \tag{19}$$

Bei variablem a ist mit $B_i = \sqrt[a]{C}$ das Produkt

$$C = \prod_i B_i = \prod_{i=1}^{a} \sqrt[a]{C} = C^{a/a} = C^1 \tag{20}$$

konstant.

$$\hookrightarrow \quad f(a) = \sum_{i=1}^{a} \sqrt[a]{C} = a \cdot \sqrt[a]{C} \quad \Longrightarrow \quad \min_x \tag{21}$$

Mit der Substitution $v = C^{1/a}$, $v' = C^{1/a} \cdot \ln C \cdot \frac{-1}{a^2}$ liefert die Ableitung

$$\begin{aligned} f'(a) &= C^{1/a} + a \cdot C^{1/a} \cdot \ln C \cdot \frac{-1}{a^2} = 0 && (22)\\ 0 &= C^{1/a}(1 - \frac{1}{a}\ln C) \\ 1 &= \frac{1}{a}\ln C \\ a &= \ln C \quad . && (23) \end{aligned}$$

Die zweite Ableitung bestätigt das Maximum.

$$\begin{aligned} f''(a) &= C^{1/a} \cdot \frac{1}{a}\ln C \cdot \left(\frac{1}{a}(1 + \ln C) - 1\right) \\ f''(\ln C = a) &= \mathrm{e}\left(\frac{1+\mathrm{a}}{\mathrm{a}} - 1\right) > 0 \quad \text{für} \quad \mathrm{a} > 0 && (24) \end{aligned}$$

Die Anzahl der Stufen ist abhängig vom Elementarvorrat. Die Anzahl der Alternativen je Stufe ist $B_i = \sqrt[a]{C} = \sqrt[\ln C]{C} = \mathrm{e}$ und der für einen bestimmten Elementarvorrat notwendige Testaufwand $D_{()} = a \cdot \mathrm{e} = \mathrm{e} \cdot \ln \mathrm{C}$.

Daraus folgt der

Satz 2.3 Wenn, bei unvollständigen Merkmalen, der Testaufwand auf $D_{()}/\mathrm{e}$ Entscheidungsstufen mit e Alternativen aufgeteilt wird, kann mit minimalem Testaufwand ein Element aus einem maximalem Elementarvorrat ausgewählt werden. •

Das gleiche Ergebnis entsteht mit variablem Elementarvorrat C und festem Testaufwand $D_{[]}$.

2.3 Konstruktion von nicht ganzzahligen, unvollständigen Tests auf diskreten Elementen

Ein Beispiel für einen nicht ganzzahligen, unvollständigen Test ist in Abbildung 1 dargestellt. Es werden binärwertige Zeichen vorausgesetzt. Bei einer fünfstufigen Auswahl wird zugelassen, dass auf einer oder mehreren Stufen keine Entscheidung getroffen wird, weil der Test nicht eindeutig ist, z. B. weil das entsprechende Zeichen nicht erkannt wird. Dabei sei ungewiss, welche Zeichen nicht erkannt wurden, d. h. es bleibt unbeachtet, auf welcher Stufe keine Entscheidung getroffen wird.

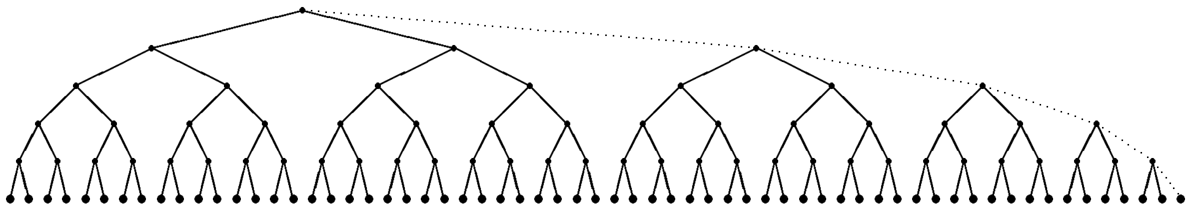

Abbildung 1: Entscheidung bei nicht erkanntem Fehlen von Zeichen

Im Weiteren werden Folgen diskutiert. Darum wird fortan statt a für die Anzahl der Merkmale bzw. Entscheidungsstufen das Formelzeichen n für das n-te bzw. letzte Glied der Folge verwendet. So entspricht die n-te Summe der Alternativen $S_A(n)$ dem Elementarvorrat C_a mit a Merkmalen.

Das Ergebnis der Auswahl hängt von den erkannten Zeichen selbst und von deren Anzahl i (Anzahl der gefällten Entscheidungen) ab. Bei i erkannten binärwertigen Zeichen können jeweils 2^i Alternativen unterschieden werden. Für alle möglichen $i = 0 \ldots n$ sind das insgesamt

$$S_A(n) = \sum_{i=0}^{n} 2^n = 2^{n+1} - 1 = 2 \cdot S_A(n-1) + 1 \tag{25}$$

Alternativen. Damit enthält jede Stufe im mittel $\bar{I}(n) = \mathrm{ld}(2^{n+1} - 1)/n$ Bit. Für $n = 5$ sind das 1.2 Bit. Wenn alle Alternativen in Abbildung 1 gleichwahrscheinlich sind, ist die Wahrscheinlichkeit dass i Entscheidungen getroffen werden

$$p(i) = \frac{2^i}{2^{i+1} - 1} \quad . \tag{26}$$

Damit liefern im Durchschnitt $\bar{n}(n) = 1/n \sum_{i=0}^{n} i \cdot p(i)$ Tests je Stufe ein Ergebnis. Für $n = 5$ sind das 0.8 Tests je Stufe. Nicht ganzzahlige unvollständige Tests sind also konstruierbar.

Die Anzahl $2^{n+1} - 1$ der Alternativen nach Abbildung 1 soll mit der Anzahl 2^n der Alternativen bei ganzzahligen vollständigen binären Entscheidungen verglichen werden. Dazu wird der Quotient der Elementarvorräte beider Varianten gebildet $(2^{n+1} - 1)/2^n$. Der Grenzwert

$$\lim_{n\to\infty} \frac{2^{n+1} - 1}{2^n} = \lim_{n\to\infty} 2 - \frac{1}{2^n} = 2 \tag{27}$$

weist für die Entscheidungsstrategie nach Abbildung 1 die doppelte Anzahl von Alternativen aus. Die Information ist aber auf n Stufen verteilt, so dass der mittlere Informationsgehalt je Stufe $(\mathrm{ld}C)/n$ bei beiden Entscheidungsstrategien für große n gleich ist.

$$\lim_{n\to\infty} \left(\frac{\mathrm{ld}(2^{n+1} - 1)}{\mathrm{ld}(2^n)} \cdot \frac{n}{n} \right) = \lim_{n\to\infty} \left(\frac{\mathrm{ld}(2^{n+1} - 1)}{n} \right) = \lim_{n\to\infty} \left(\frac{1}{n} \mathrm{ld} \left(2 - \frac{1}{2^n} \right) + 1 \right) = 1 \tag{28}$$

Also unterscheidet sich die Strategie der nicht ganzzahligen unvollständigen Entscheidung nach Abbildung 1 von der ganzzahligen vollständigen binären Entscheidung fast nicht.

Als nächstes wird folgender Fall betrachtet:

- unvollständige, mehrstufige Entscheidung mit binärwertigen Zeichen,
- wird ein Zeichen erkannt, steht fest, ob und wie viele Zeichen der vorherigen Stufen nicht erkannt wurden,
- wurden vorhergehende Zeichen nicht erkannt, kann mit dem ersten erkannten Zeichen keine Entscheidung getroffen werden.

Der Entscheidungsgraph für diese Entscheidungsstrategie ist in Abbildung 2 dargestellt.

Die Anzahl der Alternativen in Abhängigkeit von der Anzahl der Stufen ergibt sich aus den Alternativen durch Entscheidung $2 \cdot S_A(n-1)$ plus den Alternativen durch erkannte Fehler $\sum_{i=1}^{n-2} S_A(i)$ (siehe Abbildung 2).

$$S_A(n) = 2 \cdot S_A(n-1) + \sum_{i=0}^{n-2} S_A(i) = S_A(n-1) + \sum_{i=0}^{n-1} S_A(i) \quad . \tag{29}$$

Ohne Entscheidung gibt es nur eine Alternative.

$$S_A(0) = 1 \tag{30}$$

Eine andere Darstellung entsteht, wenn $S_A(n-1)$, $S_A(n-2) \ldots S_A(n-n)$ iterativ als Summen aufgelöst werden.

$$\begin{aligned}
S_A(n) &= \sum_{i=0}^{n-1} S_A(i) + S_A(n-1) \\
&= \sum_{i=0}^{n-1} S_A(i) + \sum_{i=0}^{n-2} S_A(i) + S_A(n-2) \\
&\vdots \\
&= \sum_{i=0}^{n-1} S_A(i) + \sum_{i=0}^{n-2} S_A(i) + \ldots + \sum_{i=0}^{1} S_A(i) + \sum_{i=0}^{0} S_A(i) + S_A(0) \\
S_A(n) &= 1 + \sum_{j=0}^{n-1} \sum_{i=0}^{j} S_A(i) && (31) \\
S_A(n) &= 1 + \sum_{j=0}^{n-1} S_{A0}(j) \qquad S_{A0}(j) = \sum_{i=0}^{j} S_A(i) && (32)
\end{aligned}$$

Das n-te Glied der Folge kann auch als Summe aller Differenzen der benachbarten Glieder und der Differenz

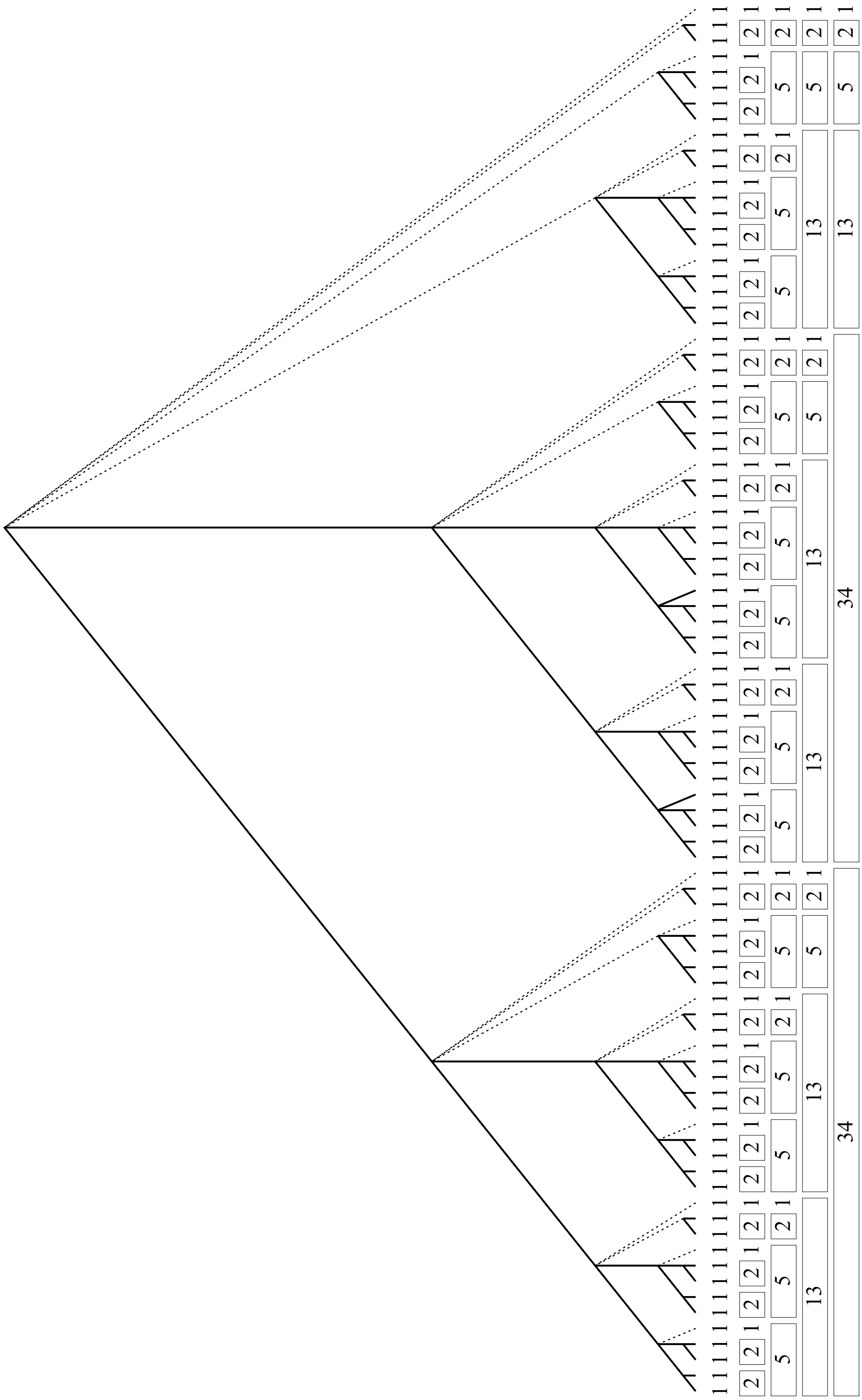

Abbildung 2: Entscheidung bei erkanntem Fehlen von Zeichen

zwischen dem ersten Glied $S_A(0)$ und Null aufgefasst werden. Mit $S_A(-1) = 0$ gilt

$$
\begin{aligned}
S_A(n) &= \sum_{j=1}^{n} \big(S_A(j) - S_A(j-1)\big) + \big(S_A(0) - 0\big) \\
&= \sum_{j=0}^{n} \big(S_A(j) - S_A(j-1)\big) \\
S_A(n) &= \sum_{j=0}^{n} \sum_{i=0}^{j-1} S_A(i) \qquad (33) \\
S_A(n) &= \sum_{j=0}^{n} S_{A1}(j) \qquad S_{A1}(j) = \sum_{i=0}^{j-1} S_A(i) \quad . \qquad (34)
\end{aligned}
$$

Der Vergleich von (32) und(34) liefert

$$S_{A0}(j-1) = S_{A1}(j) \qquad (35)$$

Das Bildungsgesetz der Folge $S_A(n)$ ist in Abbildung 3 dargestellt.

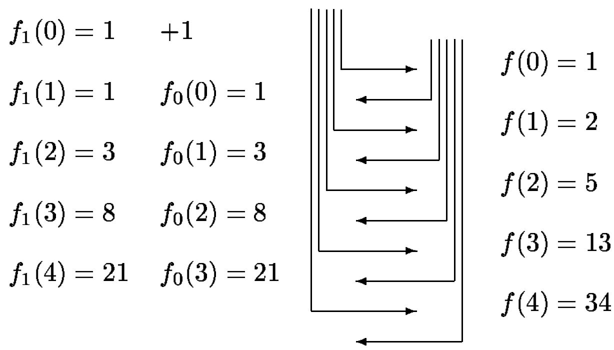

Abbildung 3: Darstellung der doppelten Summe

Aus (29) und mit $\sum_{i=1}^{n-1} S_A(i) = S_A(n) - S_A(n-1) = \sum_{i=1}^{n-2} S_A(i) + S_A(n-1)$ kann die rein rekursive Darstellung hergeleitet werden.

$$
\begin{aligned}
S_A(n-1) &= S_A(n-2) + \sum_{i=0}^{n-2} S_A(i) &&= S_A(n-2) - S_A(n-1) + \sum_{i=0}^{n-1} S_A(i) \\
2 \cdot S_A(n-1) &= S_A(n-2) + \sum_{i=0}^{n-1} S_A(i) &&= S_A(n-2) - S_A(n-1) + S_A(n) \\
S_A(n) &= 3 \cdot S_A(n-1) - S_A(n-2) && \qquad (36)
\end{aligned}
$$

Die Folge verhält sich für $n \to \infty$ wie eine geometrische Folge, d. h. der Quotient der Folge $S_A(n)/S_A(n-1)$ konvergiert gegen einen festen Grenzwert. Um den Grenzwert zu bestimmen, wird der Quotient als Zustand eines zeitdiskreten Systems betrachtet.

$$z(n) = S_A(n)/S_A(n-1) \qquad (37)$$

Der Zustand $z(n)$ beschreibt gerade die Anzahl der Alternativen je Stufe für die Stufe n. Die Folge (36) wird mit dem Zustand nach (37) zur zeitdiskreten Zustandsdifferenzengleichung

$$z[n+1] = 3 - \frac{1}{z[n]} \quad . \qquad (38)$$

Alle stationären Punkte des Systems werden ermittelt und nachgewiesen, dass genau einer davon stabil ist. Gelingt das, ist dieser stabile stationäre Punkt der Grenzwert des Quotienten der Folge. Das System (38) ist ein Spezialfall der Riccati-Differenzengleichung mit konstanten Koeffizienten,

$$x_{n+1} = \frac{ax_n + b}{cx_n + d}, \quad n = 0, 1, \ldots \quad , \quad ad - bc \neq 0 \quad , \quad c \neq 0 \tag{39}$$

mit $a = 3$, $b = -1$, $c = 1$ und $d = 0$.

In [KL93] wird bewiesen, dass die Riccati-Differenzengleichung auf eine lineare Differenzengleichung zweiter Ordnung zurück geführt werden kann und damit aus dem Poincaré-Perron's Theorem folgt, dass das System in der oberen Ruhelage stabil ist. Für das spezielle System (38) kann der Nachweis einfacher wie folgt durchgeführt werden.

Mit der Bedingung für die Ruhelage

$$\tilde{x}[n+1] = \tilde{x}[n] \tag{40}$$

liefert (38) die quadratische Gleichung

$$0 = \tilde{x}^2[n] - 3 \cdot \tilde{x}[n] + 1 \tag{41}$$

mit den zwei Lösungen

$$\tilde{x}_{1/2} = 3/2 \pm \sqrt{5/4} \quad . \tag{42}$$

Das System (38) hat also zwei Ruhelagen. Nun wird die Linearisierung des Systems (38) mittels der Jacobi-Matrix $\mathfrak{J} = \frac{\partial f(x)}{\partial x^T}$ betrachtet. Sind die Beträge der Eigenwerte der Jacobi-Matrix in der Ruhelage verschieden von eins, $|\lambda(\mathfrak{J}(\tilde{x}))| \neq 1$, ist die Ruhelage hyperbolisch. In der Nähe der hyperbolischen Ruhelage ist nach dem Satz von Grobmann und Hartman die Differenzengleichung[5] topologisch konjugiert zu ihrer Linearisierung [BSMM95]. Ein stationärer Punkt $\tilde{x}$ des nichtlinearen, zeitinvarianten, diskreten Systems ist stabil, wenn alle Eigenwerte $\lambda(\mathfrak{J})$ im Punkt $\tilde{x}$ kleiner eins sind. Im hiesigen eindimensionalen Fall gilt

$$z[n+1] = f(x[n]) = 3 - \frac{1}{z[n]}$$

$$\begin{array}{lcll} \lambda(\mathfrak{J}(\tilde{x})) & = & \frac{1}{\tilde{x}^2} & \\ \lambda(\mathfrak{J}(\tilde{x}_1)) & = & \frac{1}{(1{,}5+1{,}1)^2} < 1 & \text{stabil} \\ \lambda(\mathfrak{J}(\tilde{x}_2)) & = & \frac{1}{(1{,}5-1{,}1)^2} > 1 & \text{instabil} \end{array} \quad .$$

Das System ist im größeren stationären Punkt stabil. Somit beträgt der Grenzwert des Quotienten der Folge der Anzahl der Alternativen

$$\lim_{n\to\infty} \left(\frac{S_A(n)}{S_A(n-1)} \right) = \frac{3+\sqrt{5}}{2} = 2{,}618 = a \quad . \tag{43}$$

Der Quotient a liegt deutlich dichter an e als an 2. Die Strategie mit erkannten Fehlern ist somit eine recht gute Näherung an das Optimum für unvollständige, nicht ganzzahlige Entscheidungen.

Bemerkung 2.1 Die Elemente der Hilfsfolgen $S_{A0}(n-1) = S_{A1}(n)$ stellen das ganzzahlig gerundete arithmetische Mittel $\sqrt{S_A(n)S_A(n-1)}$ der Element der Folge S_A dar und treffen mit steigendem n das arithmetische Mittel immer genauer. Es gilt

$$b = \lim_{n\to\infty} \left(\frac{S_A(n)}{S_{A1}(n)} \right) = \lim_{n\to\infty} \left(\frac{S_{A1}(n+1)}{S_A(n)} \right) = \lim_{n\to\infty} \left(\frac{S_A(n+1)}{S_A(n)} \right) - 1 = \lim_{n\to\infty} \left(\frac{S_{A1}(n+1)}{S_{A1}(n)} \right) - 1$$

[5]Die Eigenschaft des Diffeomorphismus ist in der Nähe der zwei Ruhelagen erfüllt.

$$a - 1 = b^2 - 1 = b = \frac{1}{b} + 1$$

$$\left(\frac{3}{2} + \sqrt{\frac{5}{4}}\right) - 1 = \left(\frac{1}{2} + \sqrt{\frac{5}{4}}\right)^2 - 1 = \frac{1}{2} + \sqrt{\frac{5}{4}} = \frac{1}{\left(\frac{1}{2} + \sqrt{\frac{5}{4}}\right)} + 1 = \left(-\frac{1}{2} + \sqrt{\frac{5}{4}}\right) + 1$$

$$\sqrt{2{,}618} = 1{,}618 = \frac{1}{0{,}618} \quad .$$

Die Zahlen 2,618, 1,618, 1, und 0,618 liegen im „goldenen Schnitt“.

Die Vereinigung der Folgen $S_A(n) = S_{A2}(2n)$ und $S_{A1}(n+1) = S_{A2}(2n+1)$ zur Folge S_{A2} führt auf die einfache Bildungsvorschrift

$$S_{A2}(n) = S_{A2}(n-1) + S_{A2}(n-2) \quad \text{mit} \quad S_{A2}(-1) = 0 \quad \text{und} \quad S_{A2}(0) = 1 \quad . \tag{44}$$

•

Oben wurde gezeigt, dass die Strategie mit erkannten Fehlern eine recht gute Näherung an das Optimum für unvollständige, nicht ganzzahlige Entscheidungen darstellt, da nach vielen Entscheidungsstufen die Anzahl der Alternativen je Stufe $z[n]$ gegen $1/2(3+\sqrt{5}) = 2,618 \approx \mathrm{e}$ strebt. Im Folgenden soll untersucht werden, wie eine Entscheidungsstrategie aussehen muss, damit die Anzahl der Alternativen je Stufe genau gegen e strebt, also

$$\lim_{n\to\infty}\left(\frac{S_{A\mathrm{e}}(n)}{S_{A\mathrm{e}}(n-1)}\right) = \mathrm{e} = \lim_{\mathrm{n}\to\infty}\sum_{\mathrm{i}=0}^{\mathrm{n}}\frac{1}{\mathrm{i}!} \tag{45}$$

wird. Wenn die Argumente der Grenzwerte gleich sind, sind auch die beiden Grenzwerte gleich.

$$\frac{S_{A\mathrm{e}}(n)}{S_{A\mathrm{e}}(n-1)} = \sum_{i=0}^{n}\frac{1}{i!} \tag{46}$$

$$\begin{aligned} S_{A\mathrm{e}}(n) &= \sum_{i=0}^{n}\frac{1}{i!}\cdot S_{A\mathrm{e}}(n-1) \\ S_{A\mathrm{e}}(n-1) &= \sum_{i=0}^{n-1}\frac{1}{i!}\cdot S_{A\mathrm{e}}(n-2) \\ S_{A\mathrm{e}}(n-2) &= \sum_{i=0}^{n-2}\frac{1}{i!}\cdot S_{A\mathrm{e}}(n-3) \\ &\ddots \end{aligned} \tag{47}$$

$$S_{A\mathrm{e}}(n) = \prod_{j=0}^{n}\sum_{i=0}^{j}\frac{1}{i!} = \prod_{j=0}^{n}\sum_{i=0}^{j}\frac{1}{\prod_{k=0}^{i} k} \tag{48}$$

Die Entscheidungsstrategie nach (48) ist in Abbildung 4 dargestellt. Die Realisierbarkeit einer solchen Entscheidungsstrategie ist jedoch fraglich. Durch die Teilung der Entscheidung mit $1/i!$, $i = 0,1,2,3,4,5,\ldots$ ist es kaum möglich, einen systematischen Algorithmus anzugeben, so dass die „Teil"-Entscheidungen zu „ganzen“ Alternativen gruppiert werden. So gibt es nach der vierten Stufe zwar 4! Elemente, die mit dem Faktor 4! gewichtet werden, aber 3! Elemente davon werden mit $3!\cdot 4!$ und davon wieder 2! Elemente mit dem Faktor $2!\cdot 3!\cdot 4!$ gewichtet.

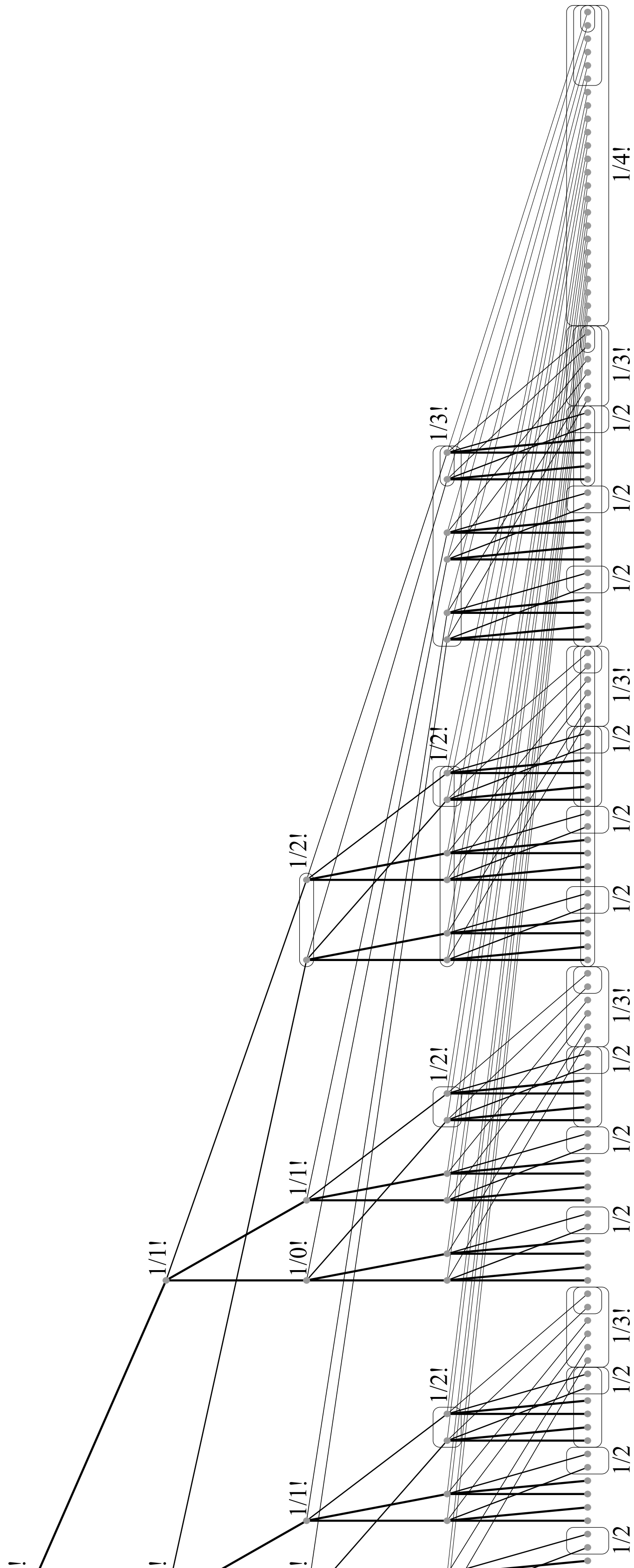

Abbildung 4: Entscheidung für $z(n \to \infty) = \mathrm{e}$

3 Parameterfehler bei endlicher Information

Wünschenswert ist, für einen Parameter in einer Struktur keine Einschränkungen machen zu müssen. Der Wertebereich sollte von $-\infty$ bis $+\infty$ reichen, und der Kodierungsfehler sollte bekannt sein. Es besteht also die Frage, wie kann eine Zahl x mit einem Minimum an Information als metrischer Wert x^* kodiert werden, so dass der Code den Bereich $[-\infty, +\infty]$ abdeckt und der Fehler endlich bleibt; bzw. etwas vorsichtiger: wie groß muss der zulässige Fehler ${}^{e}f(x)$ sein, damit die notwendige Information $H(x^*)$ endlich bleibt. Eine erste Antwort liefert der

Satz 3.1 Es gibt keinen Code endlicher Länge, der für jede beliebige Zahl aus einem unendlichen Intervall einen Wert mit einem endlichen Fehler generiert. •

Beweis 3.1 Ein Code endlicher Länge kann nur eine endliche Anzahl J verschiedener Werte x_j^* annehmen. Um den endlichen Fehler $a = \max_x {}^{e}f(x) = \max_x(\min_{x^*} |x - x^*|)$ zu garantieren, darf der Abstand zwischen zwei benachbarten Werten x_i^* und x_{i+1}^* nicht größer als $2a$ sein. Aus der endlichen Anzahl J endlicher Intervalle a folgt das maximal abgedeckte Intervall von $2aJ$, das nicht unendlich ist. •

Die Forderung nach einem Bereich $[-\infty, +\infty]$ ist praktisch nicht unbedingt notwendig, denn die Identifikation eines Parameters von $\geq 100!^{100!^{100!}}$ ist recht unwahrscheinlich, eine rechentechnische Darstellung ist mit gängigen Systemen nicht möglich und selbst ein Fehler von einer Trilliarde dürfte für den Einsatz als Parameter belanglos sein. Der Sachverhalt lässt sich aber recht gut als $1/0$ darstellen. Eine Antwort auf die Frage, wie gut Zahlen nahe Null dargestellt werden können, liefert der

Satz 3.2 Es gibt keinen Code endlicher Länge, der für jede beliebige Zahl aus einem Intervall, das die Null und eine von Null verschiedene Zahl a einschließt, einen Wert mit einem relativen Fehler $r < 1$ generiert. •

Beweis 3.2 Das kleinste Intervall, welches die Bedingung erfüllt ist $[0, a]$. Der Abstand zwischen zwei benachbarten Werten x_i^* und x_{i+1}^* ($x_i^* > x_{i+1}^*$ für $a > 0$) darf bei einem relativen Fehler r nicht größer sein als $rx_i^* + rx_{i+1}^*$. Die Werte mit maximalem Abstand können über die geometrische Reihe

$$x_i^* = x_{i+1}^*(\frac{1-r}{1+r})$$

ermittelt werden. Das abgedeckte Intervall mit unendlich vielen Werten ist

$$b_\infty = \sum_{i=0}^{\infty} 2rx_0^*(\frac{1-r}{1+r})^i = x_0^*(1+r).$$

Ohne Einschränkung der Allgemeinheit kann $x_0^* = a/(1+r)$ gesetzt werden, d.h. der Toleranzbereich des größten Wertes x_0^* erreicht genau a und der Toleranzbereich des kleinsten Wertes x_∞^* erreicht genau die 0.

$$b_\infty = a - 0$$

Aus einer endlichen Anzahl J verschiedener Werte aus dem endlichen Code entsteht das Intervall

$$\begin{aligned} b_J &= x_0^*(1+r) - x_0^*(1+r)(\frac{1-r}{1+r})^J \\ &= a - R \quad . \end{aligned}$$

Für beliebiges $r < 1$ sind die Basis $(\frac{1-r}{1+r})$, der Faktor $(1+r)$ und mit $x_0^* > 0$ auch das Restglied $x_0^*(1+r)(\frac{1-r}{1+r})^J$ immer positiv, d.h. der Toleranzbereich des kleinsten Wertes x_J^* erreicht die Null nicht. •

Die Größe des abgedeckten Bereiches hat keinen Einfluss auf die Größe des relativen Fehlers.

Eine endliche Information erfordert immer eine Einschränkung des Wertebereiches des zu schätzenden Parameters. Soll der relative Fehler $r < 1$ sein, muss der Wertebereich infinitesimal kleine Werte ausschließen. Soll der absolute Fehler endlich bleiben, muss der Wertebereich in Richtung $\pm\infty$ begrenzt werden. Die Begrenzung des Wertebereiches ist praktisch meist durch das numerische Format der rechentechnischen Zahlendarstellung vorgegeben. Die Auflösung und der Wertebereich der Messwerte geben weitere Hinweise zur Einschränkung des Wertebereichs der Parameter.

4 Äquivalenz von Struktur und Parametern

Bei der datenbasierten Strukturermittlung werden die betrachteten Strukturen mittels Gütekriterien bewertet. In der Literatur sind häufig Gütekriterien zu finden, in die nur die Anzahl der Parameter eingeht und die Komplexität der Struktur nicht anderwertig berücksichtigt wird. Solche Gütekriterien sind ungünstig, da von zwei Modellen mit gleichem Eingangs-/ Ausgangsverhalten, immer das Modell mit der minimalen Anzahl von Parametern als bestes bewertet wird, egal wie komplex seine Struktur auch sein mag. Wenn die Struktur nicht feststeht, ist das nicht akzeptabel, weil jede Vielfalt von Parametern aus einem einzigen Parameter in Verbindung mit Strukturelementen und Verknüpfungen erzeugt werden kann. Wenn das Einselement als Strukturelement akzeptiert wird oder Konstruktionen wie in Bild 5 als gültig angesehen werden, sind Parameter überhaupt nicht mehr notwendig.

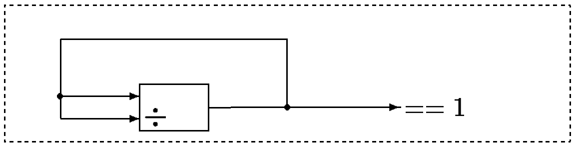

Abbildung 5: Äquivalenz zwischen einer Struktur und dem Parameterwert 1

Entsprechend des Aufbaus der Zahlenbereiche, der natürlichen Zahlen (Null $= a - a$, Eins $= a/a$, Nachfolger von Eins ist Zwei usw.), der ganzen Zahlen (Differenz$\langle a, b\rangle$), der rationalen Zahlen (Verhältnis$\langle a, b\rangle$) kann jeder Parameter beliebig genau durch Strukturelemente dargestellt werden. So ist z. B. die Zahl $3,1416 \approx \pi$ darstellbar durch eine Struktur entsprechend folgender Beschreibung: $\forall a \in \Omega$; $b_1 = a/a$; $b_2 = b_1 + b_1$; $b_8 = b_2 * b_2 * b_2$; $b_{64} = b_8 * b_8$; $b_{1024} = b_{64} * b_8 * b_2$; $b_{1/8} = b_1/b_8$; $b_{1/64} = b_1/b_{64}$; $b_{1/1024} = b_1/b_{1024}$; $b_2 + b_1 + b_{1/8} + b_{1/64} + b_{1/1024} = 3,1416 \approx \pi$.

Es gilt offensichtlich der

Satz 4.1 Parameterwerte können beliebig genau durch Strukturen ersetzt werden. (Oder genauer: Jeder Parameterwert aus einer Metrik, deren System auf den Ordinalzahlen 0 und 1 aufbaut, kann durch eine Struktur entsprechend der Metrik ersetzt werden.) •

5 Schlußfolgerung und Zusammenfassung

In Abschnitt 2 wurde nachgewiesen, dass Informationen über vollständige Merkmale optimal mit dem Dualsystem kodiert werden. Dabei wurde, entgegen der in der Literatur üblichen Beweisführung, der duale Logarithmus im Beweis weder angesetzt noch verwendet.

Auch für Informationen über unvollständige Merkmale wurde, wie hinreichend bekannt, die Basis e als optimal für die Entscheidungsstufung nachgewiesen. Auch hier wurde vermieden den natürlichen Logarithmus beim Beweis anzusetzen, was sonst in der Literatur üblich ist. Im beschriebenen Beweis ergibt sich der natürliche Logarithmus vielmehr durch Differentiation.

An Beispielen wurde gezeigt, dass Entscheidungsstrategien für unvollständige Merkmale und mit nicht ganzzahliger Stufung denkbar, aber wenig praktikabel sind. Besser ist, die Merkmale so zu wählen, dass sie disjunkt sind, gleiche Wahrscheinlichkeit aufweisen und wenn die Komplementärmenge aller Merkmale nicht leer ist, diese als zusätzliches Merkmal mit einzubeziehen. Bei metrisch geordneten Merkmalen sollte die Komplementärmenge in eine untere und eine obere Komplementärmenge aufgeteilt werden, wenn die Elemente der Komplementärmenge bezüglich der Metrik unterscheidbar sind und sämtlich außerhalb der ursprünglichen Merkmale liegen. Mit diesen zwei zusätzlichen Merkmalen bleibt eine sinnvolle Nachbarschaftsbeziehung erhalten.

Wird die Modellbildung als Informationsfluss von einem Objekt auf ein Modell betrachtet, und umfasst der Informationsfluss nur endlich viel Information, so muss das Modell mit endlicher Information aufgebaut werden. Somit müssen auch die Parameter mit endlicher Information bestimmt werden. Im Abschnitt 3 wurde gezeigt, dass der Wertebereich eines Parameters endlich sein muss, damit der Fehler des Parameters endlich bleiben kann und dass der Wertebereich eines Parameters die Null ausschließen muss, damit der relative Fehler kleiner als Eins bleiben kann.

Außerdem wurde in Abschnitt 4 gezeigt, dass sobald die Struktur eines Systems beliebig ist, jeder Parameter (infinitesimal klein oder unendlich groß) beliebig genau (absolut und relativ) angenähert werden kann. Bei endlicher Information müssen also nicht nur die Parameter, sondern auch die Menge der verschiedenen Strukturen eingeschränkt werden. Ohne a-priori Information über diese Einschränkung ist eine Modellbildung nicht durchführbar.

Die Möglichkeit Parameter durch Strukturen zu ersetzen, zeigt außerdem, dass Gütemaße welche die Komplexität der einer Struktur nur durch die Anzahl der Parameter bewerten, für eine Strukturermittlung ungeeignet sind. Zur Bewertung der Komplexität einer Struktur sollten

- die Anzahl der Elemente,
- die Anzahl der Kopplungen,
- die Komplexität der Elemente (z. B. Schaltelement, statisches Element, dynamisches Element) und
- die Komplexität der Kopplung (z. B. binär, diskret, reell, komplex)

bewertet werden.

Die Untersuchungen zeigen, dass zur Modellbildung Strukturannahmen unabdingbar sind. Eine Strukturermittlung kann nur auf diesen Strukturannahmen aufbauen.

Literatur

[BSMM95] I. N. Bronstein, K. A. Semendjajew, G. Musiol und H. Mühlig. *Taschenbuch der Mathematik.* Verlag Harri Deutsch, Thun, Frankfurt am Main, 1995.

[GG99] U. Getzlaff und L. Gröll. *Indikation von Strukturelementen durch Dimensionsbetrachtungen,* FZKA 6327. Karlsruhe: Forschungszentrum Karlsruhe GmbH, Technik und Umwelt, 2000.

[Hop94] D. Hoppe. Ein Vergleich zwischen assoziativen Korrelationsmatrizen und assoziativen Tensoren mit Bezug zur optimalen Signalschätzung. *at - Automatisierungstechnik,* 42(8):337–344, 1994.

[Hop96] D. Hoppe. *Anwendung eines tensororientierten Mustererkennungsverfahrens zur Leckortung an Druckanlagen,* Band 1282 aus *GMA-Kongreß'96, VDI-Berichte.* Düsseldorf: VDI Verlag, 1996.

[Hop98a] D. Hoppe. *Nutzung der statistischen Informationstheorie und eines tensororientierten Klassifikationsverfahrens zur Signalerkennung an einem chemischen Prozess,* Band FZR-225 aus *Forschungsberichte.* Forschungszentrum Rossendorf, PF 51 01 19, D-01314 Dresden, 06/1998.

[Hop98b] D. Hoppe. *Modellbildung durch Auswertung von Fehlerdimensionen,* Band FZR-244 aus *Forschungsberichte.* Forschungszentrum Rossendorf, PF 51 01 19, D-01314 Dresden, 12/1998.

[KL93] V. L. Kocic und G. Ladas. *Global behavior of nonlinear difference equations of higher order with applications.* Kluwer Academic Publisher, 1993.

[ME69] W. M.-Eppler. *Grundlagen und Anwendungen der Informationstheorie.* Kommunikation und Kybernetik in Einzeldarstellungen. Band 1. Berlin/Heidelberg/New York: Springer-Verlag, 1969.

[PS80] D. Poetschke und F. Sobik. *Mathematische Informationstheorie, Probleme u. neuere Ergebnisse,* Band 16 aus *Elektronisches Rechnen und Regeln.* Berlin: Akademie-Verlag Berlin, 1980.

[You75] J. F. Young. *Einfuehrung in die Informationstheorie.* Einfuehrung in die Nachrichtentechnik. München/Wien: R. Oldenbourg Verlag, 1975.

[Zem59] H. Zemanek. *Elementare Informationstheorie.* München/Wien: R. Oldenbourg Verlag, 1959.