

**FZR-225**

Juni 1998

*Dietrich Hoppe*

**Nutzung der  
statistischen Informationstheorie  
und eines  
tensororientierten Klassifizierungsverfahrens  
zur Signalerkennung an einem  
chemischen Prozeß**

**Archiv-Ex.:**

Herausgeber:  
FORSCHUNGSZENTRUM ROSSENDORF  
Postfach 51 01 19  
D-01314 Dresden  
Telefon (03 51) 26 00  
Telefax (03 51) 2 69 04 61

Als Manuskript gedruckt  
Alle Rechte beim Herausgeber

Fördervorhaben

**Nutzung der  
statistischen Informationstheorie  
und eines  
tensororientierten Klassifizierungsverfahrens  
zur Signalerkennung an einem  
chemischen Prozeß**

Dietrich Hoppe

**Arbeitsbericht**

Das diesem Bericht zugrundeliegende Vorhaben wird mit Mitteln der DFG unter dem Förderkennzeichen WE 2249/1-1 gefördert. Die Verantwortung für den Inhalt der Arbeit liegt allein beim Autor.

## **Kurzfassung**

Die Klassifikation eines chemisch-verfahrenstechnischen Prozesses nach Gefahrenklassen wird unter besonderer Berücksichtigung der Prozeßsicherheit behandelt. Mangels verlässlicher A-priori-Information über den Prozeß (z.B. eines dynamischen Prozeßmodells) werden klassifizierungsrelevante Prozeßeigenschaften (Merkmale) allein aus Prozeßsignalen gewonnen. Fundamentales Hilfsmittel ist dabei die statistische Informationstheorie. Außer durch die Anwendung eines speziellen und - weil zur Kategorie der Assoziativspeicher gehörenden - besonders einfachen Klassifikators wird insbesondere durch die Nutzung der statistischen Informationstheorie erreicht, daß die Klassifizierungsergebnisse im Interesse der Prozeßsicherheit transparent und nachvollziehbar sind; ungeeignete Merkmale können dadurch leicht lokalisiert werden.

Wie exemplarisch gezeigt wird, sind Zusammenhänge, die informationstheoretisch zwischen bestimmten Prozeßsignalen nachgewiesen werden, für die Strukturierung eines dynamischen Prozeßmodells nutzbar. Ein solches Modell kann zur Effektivierung der Prozeßmerkmale und folglich zur genaueren Klassifikation genutzt werden.

## Inhalt

1. Einleitung .....	1
2. Wichtige Begriffe zur Objekterkennung .....	4
3. Informationstheoretische Grundlagen .....	5
4. Präzisierung der Zielstellung .....	10
4.1 Zur Klassifikation .....	10
4.2 Modifikationen .....	12
5. Signalvorverarbeitung und Merkmalsbildung .....	14
5.1 Mitlaufendes Zeitfenster .....	14
5.2 Polynom- und Exponential-Approximation .....	15
5.3 Gleitende Normierung .....	18
5.4 Diskretisierung und Merkmale .....	20
5.5 Bildung von Referenzmustern .....	23
6. Die Transinformation als Bewertungsmaß .....	24
6.1 Bewertung und Modifikation der Klassen .....	24
6.2 Bewertung, Modifikation und Selektion der Merkmale .....	26
6.2.1 Einzelmerkmale .....	27
6.2.2 Merkmalspaare .....	30
7. Der Klassifikator .....	33
8. Klassifikation am chemischen Prozeß .....	36
9. Zusammenfassung und Ausblick .....	38
Anhang 1: Prozeßbeschreibung .....	40
Anhang 2: Theoretische Grundlagen des Klassifikators .....	47
Anhang 3: Strukturelemente eines dynamischen Prozeßmodells .....	52
Literatur .....	56

# 1. Einleitung

Ein Problem der Klassifikation von dynamischen Prozessen wird unter besonderer Berücksichtigung der Prozeßsicherheit sowohl methodisch als auch an einem realen chemisch-verfahrenstechnischen Prozeß<sup>I</sup> behandelt. Gegenstand der Betrachtung sind dabei ausschließlich die aus dem Prozeß zugänglichen Signale; die zur Klassifikation benötigte Information über die Prozeßdynamik wird nur aus diesen Signalen gewonnen.

Ein vergleichsweise einfaches und verlässliches analytisches Modell der vorliegenden chemischen Reaktion ist bekannt [De97]. Es könnte u.a. für eine gute Klassentrennung zur Gewinnung effektiver Signal- oder Objektmerkmale herangezogen werden. Trotzdem wird auf eine Nutzung des Modells verzichtet, denn allgemein ist davon auszugehen, daß industrielle Prozesse und insbesondere solche der chemischen Verfahrenstechnik sehr kompliziert und daher oftmals so wenig untersucht sind, daß es keine verlässlichen Dynamikmodelle gibt.<sup>II</sup>

Die Vorbereitung der Mustererkennung und im besonderen der Klassifikation beinhaltet im wesentlichen

- die Vorgabe genügend problemspezifischer Klassen (falls zur Wahl der Klassen überhaupt Freiheiten bestehen),
- das Auffinden von klassenrelevanten Objekteigenschaften bzw. -merkmalen, deren Bewertung und die Selektion geeigneten Merkmale sowie
- den Nachweis relevanter Zusammenhänge zwischen den Merkmalen und den Klassen.

Um diesen Aspekten quantitativ zu genügen, werden in dieser Arbeit Elemente der statistischen Informationstheorie nach SHANNON [Sh48] genutzt.

An eine Klassifikation besteht allgemein die Forderung, unbekannte (zu klassifizierende) Objekte zu erkennen, also sie anhand der definierten Objektmerkmale und der bekannten Merkmals-Klassen-Beziehungen den vorgegebenen Klassen zuzuordnen.

Unter der besonderen Berücksichtigung der Prozeßsicherheit werden an die Klassifikation folgende Zusatzforderungen gestellt:

- 1) Die Transparenz und die Nachvollziehbarkeit der Erkennungen sollen so weit gewährleistet sein, daß Ursachen fehlerhafter (auch ungenauer oder sonst unbefriedigender) Erkennungen präzise lokalisiert, d.h. die fehlerverursachenden Merkmale gefunden werden.

---

<sup>I</sup> Chemische Reaktion: Säurekatalysierte Veresterung von Essigsäureanhydrid mit Methanol zu Essigsäuremethylester. Klassifiziert wird nach Gefahrenklassen, d.h. Stufen der Prozeßgefährdung, s. Anhang 1.

<sup>II</sup> Der Modellverzicht bezieht sich allerdings nur auf ein vorgegebenes Modell. Nicht ausgeschlossen werden soll, daß Modelle oder Teilmodelle, die erst im Zuge fortschreitender Informationsgewinnung aus dem vorliegenden Prozeß gewonnen (strukturiert und identifiziert) werden, zur Merkmalsgewinnung herangezogen werden.

2) Die Ursachen fehlerhafter Erkennungen sollen in der Weise beseitigt werden, daß die dazu notwendigen Veränderungen

- \* genau (nicht mehr und nicht weniger als) die verursachenden Merkmale betreffen, also daß nicht wie z.B. bei bestimmten statisch nichtlinearen Approximatoren stets auch der gesamte Klassifikator neu optimiert werden muß, und
- \* nicht die schlimmstenfalls unbemerkte Entstehung neuer Fehler mit sich bringen können.

Inwieweit die Forderung 1) erfüllt werden kann, hängt entscheidend davon ab, wie die definierten Merkmale mit den vorgegebenen Klassen zusammenhängen. In dem anstrebenswerten (idealen) Sonderfall, daß die Merkmals-Klassen-Beziehungen eindeutig sind, besteht über die Klassenzugehörigkeit eines unbekanntes Musters keine Ungewißheit und umgekehrt kann genau zurückverfolgt werden, welche Merkmale ein Erkennungsergebnis bewirkt haben.

Den Forderungen 1) und 2) gemeinsam wird um so besser entsprochen, je übersichtlicher der verwendete Klassifikator in Aufbau und Wirkungsweise ist. Dahingehend besonders günstig sind einfache Assoziativspeicher, die ein unbekanntes Objekt bzw. Muster anhand der definierten Merkmale mit vorab klassifizierten (Referenz-)Mustern unmittelbar vergleichen und daraus auf die Klassenzugehörigkeit des unbekanntes Musters schließen. Zu dieser Art von Klassifikatoren gehört das in dieser Arbeit benutzte tensororientierte Klassifizierungsverfahren [H94, H96]. Dieses Verfahren besteht seinem Wesen nach darin, anhand einer nichtlinearen Metrik einen direkten Vergleich zwischen den unbekanntes Mustern und den Referenzmustern vorzunehmen. Das Wesen und die Wirkungsweise dieses nicht allgemein bekannten Verfahrens werden in dieser Arbeit nochmals kurz dargestellt.

Zwischen den Anforderungen, die innerhalb der Mustererkennung an die Merkmale und Klassen einerseits und an die Klassifikation andererseits zu stellen sind, besteht folgende Abhängigkeit: Je besser die Merkmale und die Klassen einander entsprechen, desto geringer sind die Anforderungen an den Klassifikator. Im oben genannten idealen Sonderfall eineindeutiger bzw. klassenidentischer Merkmale ist ein Klassifikator sogar ganz verzichtbar. Demgegenüber ist auch der leistungsfähigste Klassifikator nutzlos, wenn die Merkmale und die Klassen voneinander unabhängig sind.<sup>1</sup> Die Wahl der Merkmale hat offenbar die primäre, wenn nicht sogar weitaus überwiegende Bedeutung gegenüber der Wahl des Klassifikators. Den Merkmalen kommt deshalb die größte Aufmerksamkeit zu. Bei der Merkmalsuche erscheinen folgende Probleme und Besonderheiten:

- Die Suche nach klassenrelevanten Objekteigenschaften ist schwierig und kaum systematisierbar, denn ein reales Objekt besitzt faktisch eine nicht endliche Anzahl von Eigenschaften. Von diesen Eigenschaften hat nur ein verschwindend kleiner und dadurch entsprechend schwierig zu findender Teil überhaupt einen Klassenbezug. Aus diesem Teil, der immer noch als nicht endlich anzusehen ist, zweck-

---

<sup>1</sup> Der Vollständigkeit halber kann zwar auch ein gänzlich unbrauchbarer Klassifikator angenommen werden, doch ist dies wenig realistisch. Die Unbrauchbarkeit von Merkmalen ist dagegen keineswegs realitätsfern.

mäßige Merkmale zu extrahieren, ist zumeist nicht ohne umfassende subjektive Sachkenntnis möglich.

- Wegen der technischen Begrenztheit jedes Klassifikators sind meist bei weitem nicht alle klassenrelevanten Merkmale für die Klassifikation nutzbar. Zur Bewertung der Klassenrelevanz von Merkmalen und für die Selektion der relevantesten Merkmale wird A-priori-Information über die Beziehungen zwischen den Merkmalen und den Klassen benötigt. Zwei Arten solcher Information werden unterschieden:
  - \* Subjektive A-priori-Information (Kenntnisse, Erfahrungen u.ä.). Sie erlaubt eine in gewissem Grade zielgerichtete Merkmalsuche, garantiert aber nicht, daß auch klassenrelevante Merkmale gefunden werden.
  - \* Objektive A-priori-Information. Sie existiert anhand einer Stichprobe aus Referenzmustern, die den Klassenbezug der Merkmale herstellen. Diese Information erlaubt die Bewertung bereits definierter Merkmale, jedoch keine zielgerichtete Merkmalsuche.

Ausgehend davon wird zur Bereitstellung von klassenrelevanten Merkmalen folgende Vorgehensweise gewählt:

- a) Auf weitgehend subjektivem Wege werden sehr viel mehr Merkmale definiert, als vom Klassifikator praktisch nutzbar sind.
- b) Zur Gewinnung möglichst eindeutiger Merkmale-Klassen-Beziehungen werden die Merkmale und - soweit zulässig - auch die Klassen objektiv (informationstheoretisch begründet) modifiziert.
- c) Die Gesamtheit der nach a) und b) gewonnenen Merkmale wird anhand von Referenzmustern informationstheoretisch nach ihrer Klassenrelevanz bewertet und daraufhin selektiv auf eine praktisch nutzbare Anzahl reduziert.

Zur Bewertung des klassenbezogenen Informationsgehaltes von Merkmalen (einzelnen Merkmalen und Merkmalspaaren), des merkmalsbezogenen Informationsgehaltes von Klassen und der informationellen Wechselbeziehung zwischen Merkmalen und Klassen werden die aus der Nachrichtentechnik stammenden informationstheoretischen Kenngrößen Äquivokation, Irrelevanz und Transinformation benutzt [Ma96,SF76]. Selektiert werden nur die mit der höchsten Transinformation bewerteten Merkmale.

Zum Aufbau dieser Arbeit: Ausgehend von der Vereinbarung wichtiger Begriffe (Abschnitt 2) und der Einführung der benötigten informationstheoretischen Grundlagen (Abschnitt 3) wird die Zielstellung dieser Arbeit präzisiert (Abschnitt 4). Die vorwiegend subjektiv motivierten Schritte zur Signalaufbereitung und zur Merkmalsgewinnung beschreibt Abschnitt 5. Die Modifikation von Klassen und Merkmalen ist der Gegenstand von Abschnitt 6. Eine kurze algorithmische Darstellung des tensororientierten Klassifizierungsverfahrens enthält Abschnitt 7. Ergebnisse einer Klassifikation am realen chemisch-verfahrenstechnischen Prozeß werden dann im Abschnitt 8 dargestellt und diskutiert.

Einen kurzen Überblick über die chemischen und technologischen Gegebenheiten



des betrachteten Prozesses gibt Anhang 1. Die Motivation und den theoretischen Hintergrund des tensororientierten Klassifizierungsverfahrens beleuchtet Anhang 2. Wie gefundene informationstheoretische Zusammenhänge für die Strukturierung eines dynamischen Prozeßmodells nutzbar gemacht werden können, verdeutlicht Anhang 3 an einem konkreten Beispiel <sup>1</sup>.

## 2. Wichtige Begriffe zur Objekterkennung

Weil es zur Objekterkennung keine einheitliche Terminologie gibt, werden die wichtigsten der fortan gebrauchten Begriffe charakterisiert.

Ein Objekt ist ein abgegrenzter Gegenstand der Realität. Die Anzahl der Eigenschaften eines Objektes ist im allgemeinen nicht endlich. Im vorliegenden Fall ist ein Objekt eine Gesamtheit von Ein- und -Ausgangssignalen (Stell- und Meßgrößen) aus dem betrachteten chemischen Prozeß. Mehrere Realisierungen solcher Objekte (Signalverläufe aus verschiedenen Experimenten) zeigt Bild A1-2 im Anhang 1.

Für die Objekterkennung und speziell die Klassifikation ausgewählte Objekteigenschaften werden als Merkmale definiert. Aus praktischen Gründen ist nur eine endliche Merkmalsanzahl nutzbar, pro Objekt  $N \in \mathbb{N}$  Merkmale. Der Wertebereich je Merkmal sei diskret mit  $l \in \mathbb{N}$  Werten. Die Objektmerkmale  $x_n$ ,  $n \in \{1, 2, \dots, N\}$  können deshalb die Werte  $x_{n_i}$ ,  $i \in \{1, 2, \dots, l\}$  annehmen. Ein Merkmal hat stets genau einen Wert. Ist dies z.B. der  $j$ -te Werte, so gelte  $x_{n_i} = 1$  für  $i=j$  und  $x_{n_i} = 0$  für  $i \neq j$ .

In dem wichtigen Sonderfall binärer Merkmale ( $l \in \{1, 2\}$ ) wird die Schreibweise  $x_{n_1}$  und  $x_{n_2}$  durch  $x_n = 0$  und  $x_n = 1$  ersetzt, wobei  $x_n = 1/0$  (Merkmal besetzt/nicht besetzt) das Vorhandensein/Nichtvorhandensein einer Objekteigenschaft ausdrückt. Noch kürzer geschrieben steht  $x_n$  für  $x_n = 1$ , während auf eine Darstellung von  $x_n = 0$  verzichtet wird.

Die Gesamtheit aller zu einem Objekt gehörenden Merkmale bilden ein Muster  $x$ . Wenn ein Muster durch einen Code bezeichnet ist, dann ist dieser Code ebenfalls Bestandteil des Musters.

Unter einer Objekterkennung wird eine Zuordnung eines unbekanntes Musters zu einer Anzahl  $K \in \mathbb{N}$  von vorgegebenen Kategorien bzw. Klassen  $y_k$ ,  $k \in \{1, 2, \dots, K\}$  verstanden. Die Klassenbezeichnungen (Elemente des Klassencodes)  $y_k$  sind binärwertig:  $y_k = 1$  drückt die Zugehörigkeit eines Musters zur  $k$ -ten Klasse aus, wohingegen  $y_k = 0$  sowohl die Nichtzugehörigkeit als auch die nicht bekannte Klassenzugehörigkeit ausdrücken kann. Wie bereits angedeutet, sind die  $y_k$  ebenfalls Bestandteil eines Musters.

---

<sup>1</sup> Der im Anhang 3 verwendete methodische Ansatz wird hinsichtlich einer selbstorganisierenden Strukturierung von Modellen dynamischer und speziell regelungstechnischer Prozesse weiterentwickelt; dies ist einer Folgearbeit vorbehalten.

Die Muster sind als  $(N+K)$ -dimensionale Vektoren  $\underline{x}=(x_1, x_2, \dots, x_N, y_1, \dots, y_K)^T$  darstellbar. Muster mit binären Merkmalen und dem ohnehin binären Klassencode können auch als (Merkmals-)Mengen angegeben werden. Die Merkmale mit  $x_n=1$  zusammen mit den Klassenbezeichnungen  $y_k=1$  bilden dann die Elemente der Merkmalsmenge eines Musters. Wegen ihrer Binärwertigkeit ist auf diese Mengen die Boolesche Algebra anwendbar.

Muster mit bekannter Klassenzugehörigkeit heißen Referenzmuster. Als Träger von A-priori-Information über den Zusammenhang zwischen Merkmalen und Klassen sind sie fundamentale Grundlage einer Klassifikation. Gebildet werden  $M \in \mathbb{N}$  Referenzmuster  $x_m$ ,  $m \in \{1, 2, \dots, M\}$  mit den Merkmalen  $x_{mn}$ , deren Werten  $x_{mn_i}$  und den Klassenbezeichnungen  $y_{mk}$ , wobei  $y_{mk}=1/0$  die Zugehörigkeit/Nichtzugehörigkeit des  $m$ -ten Referenzmusters zur Klasse  $y_k$  ausdrückt. Die vektorielle Form eines Referenzmusters ist  $\underline{x}_m=(x_{m1}, x_{m2}, \dots, x_{mN}, y_{m1}, \dots, y_{mK})^T$ . Die Besetzung der Elemente des Klassencodes  $y_{m1}, \dots, y_{mK}$  ist allgemein beliebig, wird aber hier auf disjunkte Klassen eingengt <sup>I</sup>.

Das Muster eines unbekanntes Objektes (unbekannte Klassenzugehörigkeit) heißt unbekanntes Muster  $u$ . Seine Merkmale sind  $u_n$ . Der Klassencode ist im unbekanntes Muster immer null. In vektorieller Form ist ein unbekanntes Muster ein  $(N+K)$ -dimensionaler Vektor  $\underline{u}=(u_1, u_2, \dots, u_N, 0, \dots, 0)^T$ .

**Beispiel 2.1:** Mit  $N=7$  binären Merkmalen und  $K=2$  Klassen ergeben sich  $N+K=9$ -dimensionale Merkmalsvektoren. Beim Referenzvektor  $\underline{x}_m=(x_{m1}, \dots, x_{m7}, y_{m1}, y_{m2})^T$  gehört zur Realisierung  $\underline{x}_m=(010110110)^T$  die Merkmalsmenge  $\{x_m\}=\{x_{m2}, x_{m4}, x_{m5}, x_{m7}, y_{m1}\}$ . Bei unbekanntes Mustern sind die Elemente des Klassencodes nicht besetzt:  $\underline{u}=(u_1, \dots, u_7, 0, 0)^T$ . Die Merkmalsmenge eines unbekanntes Musters enthält deshalb keine Klassenbezeichnungen.

### 3. Informationstheoretische Grundlagen

Die statistische Informationstheorie versteht Information als beseitigte Ungewißheit, z.B. über das Vorliegen einer Objekteigenschaft, den Wert eines Merkmals oder die Zugehörigkeit eines unbekanntes Musters zu einer Klasse. Wegen des statistischen, d.h. auf Wahrscheinlichkeiten begründeten Charakters dieser Theorie unterscheidet sich ihr Informationsbegriff grundsätzlich vom umgangssprachlichen oder semantischen (inhaltlichen, eine Bedeutung betreffenden) Verständnis von Information. Die Terminologie der statistischen Informationstheorie stammt aus der Nachrichtenübertragung.

Der Informationsgehalt  $I$  eines Merkmalswertes  $x_{n_i}$  wird als Anzahl der zur Kodierung dieses Wertes erforderlichen Binärzeichen aufgefaßt und deshalb mit dem

---

<sup>I</sup> Zur Disjunktheit der Klassen: Alle ein Erkennungsproblem betreffenden Muster seien Elemente einer Menge  $\Omega$ . Diese Menge sei erschöpfend in  $K$  Untermengen der Klassen aufgeteilt:  $\Omega_k$ ,  $k \in \{1, 2, \dots, K\}$ ,  $\bigcup_{k=1}^K \Omega_k = \Omega$ . Bei  $\Omega_k \cap \Omega_l = \emptyset$  für  $k \neq l$  sind die Klassen disjunkt (elementefremd), s. [N74] S. 7. Klassendisjunktheit bedeutet, daß zu jedem Referenzmuster nicht mehr als eine Klasse gehört.

binären Logarithmus <sup>I</sup> der Wahrscheinlichkeit  $p(x_{n_i})$  <sup>II</sup> definiert:

$$I(x_{n_i}) = -\log_2(p(x_{n_i})) \quad \text{mit der Pseudoeinheit bit} \quad (3.1-1)$$

Der Informationsgehalt  $I(x_{n_i})$  entspricht der Ungewißheit über den Wert des Merkmals  $x_{n_i}$ , die mit dem Bekanntwerden dieses Wertes beseitigt wird.

Ein mittlerer Informationsgehalt ist eine Entropie, wobei der Entropiebegriff im informationstheoretischen (nicht thermodynamischen) Sinn gebraucht wird. Die Mittelung von  $I(x_{n_i})$  erfolgt anhand der diskreten Wahrscheinlichkeitsverteilung  $\{p(x_{n_1}), \dots, p(x_{n_i})\}$  unter der Nebenbedingung  $\sum_{i=1}^I p(x_{n_i}) = 1$  <sup>III</sup> und ergibt die

### Merkmalsentropie

$$\begin{aligned} H(x_n) &= \sum_{i=1}^I p(x_{n_i}) I(x_{n_i}) \\ &= -\sum_{i=1}^I p(x_{n_i}) \log_2 p(x_{n_i}) \\ &= \sum_{i=1}^I H(x_{n_i}) \end{aligned} \quad (3.2-2)$$

mit den Einzelentropien

$$H(x_{n_i}) = -p(x_{n_i}) \log_2 p(x_{n_i}) \quad (3.2-3)$$

und der Vereinbarung  $0 \cdot \log_2(0) = 0$ . Die Merkmalsentropie  $H(x_{n_i})$  entspricht der mittleren Ungewißheit über den Wert eines Merkmals  $x_{n_i}$  oder äquivalent dem mittleren Informationsgewinn aus dem Bekanntwerden von  $x_{n_i}$ .

Ausdrücke der Form (3.2-2) sind dann maximal, wenn die beteiligten Wahrscheinlichkeiten gleichverteilt sind. Im binären Sonderfall  $H(p) = -p \cdot \log_2 p - (1-p) \cdot \log_2 (1-p)$  ist die Entropiefunktion  $H(p)$  symmetrisch; für  $H' = \frac{dH}{dp} = \log_2 \left( \frac{1-p}{p} \right) - 1 = 0$  ist  $H' < 0$  und deswegen besteht das Maximum  $H(p)_{\max} = 1$  bit für die Wahrscheinlichkeit  $p = 0,5$ , s. Bild 3.1. Bei  $a$  Alternativen besteht das  $H(p)$ -Maximum bei  $p = 1/a$ . Die Bedeutung ist folgende: Die Beseitigung einer Ungewißheit bringt den größtmöglichen mittleren Informationsgewinn dann mit sich, wenn die bestehenden Alternativen (z.B. die Werte eines Signals, s. in Abschnitt 5.4) gleichwahrscheinlich sind.

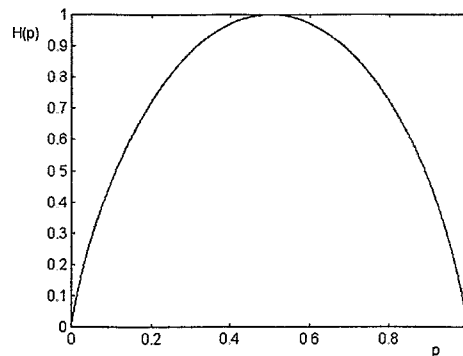
---

<sup>I</sup> Für praktische Rechnungen ist die Substitution  $\log_2(p) = \ln(p)/\ln(2)$  zweckmäßig.

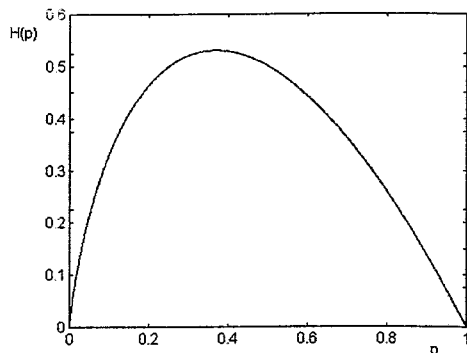
<sup>II</sup> Unter der Annahme einer genügend großen Referenzmusteranzahl  $M$  gelten die relativen Häufigkeiten der Merkmalswerte  $x_{n_i}$  als Wahrscheinlichkeiten:  $1/M \sum_{m=1}^M (x_{mn_i} = 1) \approx p(x_{n_i})$ . Desgleichen gelte auch für alle noch einzuführenden Wahrscheinlichkeiten.

<sup>III</sup> Sie ist stets erfüllt, da ein Signal immer genau einen Wert aufweist.

**Bild 3.1** Die Funktion  $H(p) = -p \cdot \log_2 p - (1-p) \cdot \log_2 (1-p)$  mit ihrem Maximum 1 bit für  $p=0,5$



Die Entropiefunktion  $H(p) = -p \cdot \log_2(p)$  gemäß (3.2-3) ist asymmetrisch; für  $H' = -(\ln(p)+1)/\ln(2) = 0$  ist  $H'' < 0$  und deswegen besteht das Maximum  $H(p)_{\max} = 0,53$  bit für  $p = 1/e \approx 0,368$ , s. Bild 3.2. An den Rändern des Definitionsbereiches von  $p$  gilt für die Ableitung  $H'|_{p \rightarrow 0} \rightarrow -\infty$  und  $H'|_{p \rightarrow 1} \rightarrow -1/\ln(2)$ . Demnach erbringt das Auftreten eines wenig wahrscheinlichen ( $p \approx 0$ ) Merkmalswertes einen im Verhältnis viel größeren Informationsgewinn als das fast sichere ( $1-p \approx 1$ ) Nicht-Auftreten dieses Wertes.



**Bild 3.2** Die Funktion  $H(p) = -p \cdot \log_2(p)$  mit ihrem Maximum für  $p=1/e$  und ihren beträchtlich unterschiedlichen Anstiegen für  $p \approx 0$  und  $p \approx 1$ .

Wird in Gleichungen der Form (3.2-2) nur über eine zweielementige Wahrscheinlichkeitsverteilung gemittelt, d.h. gilt die stets erfüllte Nebenbedingung  $p+(1-p)=1$ , so ist die mittlere Entropie bestimmt durch  $H(p) = -p \cdot \log_2(p) - (1-p) \log_2(1-p)$ . Solche Sonderformen werden später noch öfter benötigt, u.a. für die Bewertung binärer Merkmale.<sup>I</sup>

Wie für die Merkmale wird auch für die Klassen eine Entropie definiert. Die mittlere Entropie aller Klassen ist unter der Nebenbedingung  $\sum_{k=1}^K p(y_k) = 1$ , welche die Disjunktheit der Klassen voraussetzt<sup>II</sup>, die Klassenentropie

<sup>I</sup> Während die Gleichung (3.2-2) auch der statistischen Interpretation der Entropie im Sinne Boltzmanns genügt, existiert die binäre Sonderform nur in der informationstheoretischen Interpretation der Entropie [RF94].

<sup>II</sup> Sollte die Disjunktheit der Klassen nicht von vornherein gegeben sein, kann die Nebenbedingung auf zweierlei Weise erfüllt werden: (1) Klassenkombinationen werden als eigenständige (und damit disjunkte) neue Klassen definiert. Solche komplexeren Klassen sind aber ggf. in zu wenigen Referenzmustern präsent, um statistisch signifikant charakterisiert zu sein. (2) Nach jeder Klasse wird einzeln klassifiziert. Dann brauchen immer nur die disjunkten Fälle Besetztsein/Nichtbesetztsein des jeweiligen Klassencode-Elementes berücksichtigt zu werden. Letzteres wird im weiteren bevorzugt.

$$H(y) = -\sum_{k=1}^K p(y_k) \log_2 p(y_k). \quad (3.2-4)$$

Die Klassenentropie  $H(y)$  entspricht der mittleren Ungewißheit über das Vorliegen der Klassen  $y_k$ .

Merkmals- und Klassenentropie für sich allein besagen noch nichts über Abhängigkeiten zwischen den Merkmalen und den Klassen. Die Beschreibung solcher Abhängigkeiten erfolgt mit Hilfe von Verbundwahrscheinlichkeiten und bedingten Wahrscheinlichkeiten und den daraus abgeleiteten Entropien.

Derjenige Anteil an der Merkmalsentropie  $H(x_n)$ , der für die Klassenzuordnung des Merkmals  $x_n$  bedeutungslos ist, ist unter der Nebenbedingung  $\sum_{i=1}^I \sum_{k=1}^K p(x_{n_i}, y_k) = 1$  eine bedingte Entropie und heißt Äquivokation

$$H(x_n|y) = -\sum_{i=1}^I \sum_{k=1}^K p(x_{n_i}, y_k) \log_2 p(x_{n_i} | y_k) \quad (3.2-5)$$

oder mit  $p(x_{n_i}, y_k) = p(x_{n_i} | y_k) p(y_k)$

$$H(x_n|y) = -\sum_{k=1}^K p(y_k) \sum_{i=1}^I p(x_{n_i} | y_k) \log_2 p(x_{n_i} | y_k). \quad (3.2-5a)$$

Der Wertebereich der Äquivokation ist <sup>1</sup>

$$0 \leq H(x_n|y) \leq H(x_n). \quad (3.2-6)$$

Die Äquivokation  $H(x_n|y)$  entspricht der mittleren Ungewißheit über den Wert eines Merkmals  $x_n$  unter der Bedingung, daß eine Klasse  $y_k$  vorliegt.

Analog zur Äquivokation ist derjenige Anteil an der Klassenentropie  $H(y)$ , der für die Merkmalszuordnung der Klassen bedeutungslos ist, unter der Nebenbedingung  $\sum_{i=1}^I \sum_{k=1}^K p(y_k, x_{n_i}) = 1$  eine bedingte Entropie und heißt Irrelevanz

$$H(y|x_n) = -\sum_{i=1}^I \sum_{k=1}^K p(y_k, x_{n_i}) \log_2 p(y_k | x_{n_i}) \quad (3.2-7)$$

$$= -\sum_{i=1}^I p(x_{n_i}) \sum_{k=1}^K p(y_k | x_{n_i}) \log_2 p(y_k | x_{n_i}). \quad (3.2-7a)$$

Der Wertebereich der Irrelevanz ist analog (3.2-6)

$$0 \leq H(y|x_n) \leq H(y). \quad (3.2-8)$$

---

<sup>1</sup> Die Grenzen des Wertebereiches sind durch zwei Extreme bestimmt: (1) Unabhängigkeit und (2) vollständige Abhängigkeit (Eindeutigkeit) zwischen den Merkmalen  $x_n$  und den Klassen  $y_k$ . Bei (1) ist  $p(x_{n_i} | y_k) = p(x_{n_i}) \forall i$ , weswegen mit  $\sum_{k=1}^K p(y_k) = 1$  folgt  $H(x_n|y) = -\sum_{k=1}^K p(y_k) \sum_{i=1}^I p(x_{n_i}) \log_2 p(x_{n_i}) = \sum_{i=1}^I p(x_{n_i}) \log_2 p(x_{n_i}) = H(x_n)$ . Bei (2) ist  $p(x_{n_i} | y_k) = 1 \forall i$  und deswegen  $H(x_n|y) = -\sum_{k=1}^K p(y_k) \sum_{i=1}^I 1 \log_2(1) = 0$ .

$$0 \leq H(y|x_n) \leq H(y). \quad (3.2-8)$$

Die Irrelevanz  $H(y|x_n)$  entspricht der mittleren Ungewißheit über das Vorliegen einer Klasse  $y_k$  unter der Bedingung, daß der Wert des Merkmals  $x_n$  bekannt ist.

Eine Abhängigkeit zwischen einem Merkmal  $x_n$  und den Klassen  $y_k$  besteht offensichtlich nur dann, wenn die Irrelevanz  $H(y|x_n)$  kleiner als die Klassenentropie  $H(y)$  oder die Äquivokation  $H(x_n|y)$  kleiner als die Merkmalsentropie  $H(x_n)$  ist, also für  $H(y)-H(y|x_n)>0$  oder  $H(x_n)-H(x_n|y)>0$ . Die entsprechende Entropiedifferenz heißt Transinformation (auch gegenseitige Information)

$$H(y;x_n) = H(y) - H(y|x_n) \quad (3.2-9a)$$

$$\begin{aligned} &= -\sum_{k=1}^K p(y_k) \log_2 p(y_k) + \sum_{i=1}^I \sum_{k=1}^K p(y_k, x_{n_i}) \log_2 p(y_k | x_{n_i}), \quad \text{I} \\ &= -\sum_{k=1}^K \sum_{i=1}^I p(y_k, x_{n_i}) (\log_2 p(y_k) - \log_2 p(y_k | x_{n_i})) \\ &= -\sum_{k=1}^K \sum_{i=1}^I p(y_k, x_{n_i}) \log_2 \frac{p(y_k) p(x_{n_i})}{p(y_k, x_{n_i})} \\ &= -\sum_{i=1}^I \sum_{k=1}^K p(x_{n_i}, y_k) \log_2 \frac{p(x_{n_i}) p(y_k)}{p(x_{n_i}, y_k)} \\ &= -\sum_{i=1}^I \sum_{k=1}^K p(x_{n_i}, y_k) (\log_2 p(x_{n_i}) - \log_2 p(x_{n_i} | y_k)) \\ &= -\sum_{i=1}^I p(x_{n_i}) \log_2 p(x_{n_i}) + \sum_{k=1}^K \sum_{i=1}^I p(x_{n_i}, y_k) \log_2 p(x_{n_i} | y_k) \\ &= H(x_n) - H(x_n|y) = H(y;x_n). \end{aligned} \quad (3.2-9b)$$

Aus der Fallunterscheidung

a)  $0 \leq H(y) \leq H(x_n)$  und daraus mit (3.2-9a), (3.2-8)  $0 \leq H(x_n;y) \leq H(y)$ ,

b)  $0 \leq H(x_n) \leq H(y)$  und daraus mit (3.2-9b), (3.2-6)  $0 \leq H(x_n;y) \leq H(x_n)$

folgt der Wertebereich der Transinformation

$$0 \leq H(x_n;y) \leq \min_{H(y), H(x_n)}, \quad (3.2-10)$$

d.h. die Transinformation kann nicht größer sein als die kleinere der beiden Entropien  $H(y), H(x_n)$ . Für  $H(x_n) > H(y) = H(x_n;y)$  ist das bestehende Informations-Überangebot der Merkmale nicht für die Klassentrennung nutzbar. Für  $H(y) > H(x_n) = H(x_n;y)$  wird dagegen die Information, die zur Klassentrennung nötig wäre, von den Merkmalen nicht vollständig bereitgestellt.

---

<sup>I</sup> Hier werden die für Verbundwahrscheinlichkeiten geltenden Beziehungen  $\sum_{i=1}^I p(y_k, x_{n_i}) = p(y_k)$ ,  $p(y_k, x_{n_i}) = p(y_k | x_{n_i}) p(x_{n_i})$  und  $p(x_{n_i}, y_k) = p(y_k, x_{n_i})$  verwendet.

Ergänzend und zur besseren Anschauung wird noch der mittlere Informationsgehalt des gemeinsamen Auftretens von  $x_n$  und  $y_k$  angegeben. Unter der Nebenbedingung  $\sum_{i=1}^I \sum_{k=1}^K p(y_k, x_{n_i}) = 1$  ist die Verbund- oder Gesamtentropie

$$H(x_n, y) = -\sum_{i=1}^I \sum_{k=1}^K p(x_{n_i}, y_k) \log_2 p(x_{n_i}, y_k)$$

bzw. mit  $p(x_{n_i}, y_k) = p(x_{n_i} | y_k) p(y_k)$  und  $\sum_{i=1}^I p(x_{n_i}, y_k) = p(y_k)$  sowie mit (3.2-4) und (3.2-7)

$$H(x_n, y) = -\sum_{i=1}^I \sum_{k=1}^K p(x_{n_i}, y_k) (\log_2 p(x_{n_i} | y_k) + \log_2 p(y_k))$$

$$= -\sum_{i=1}^I \sum_{k=1}^K p(x_{n_i}, y_k) \log_2 p(x_{n_i} | y_k) - \sum_{k=1}^K p(y_k) \log_2 p(y_k)$$

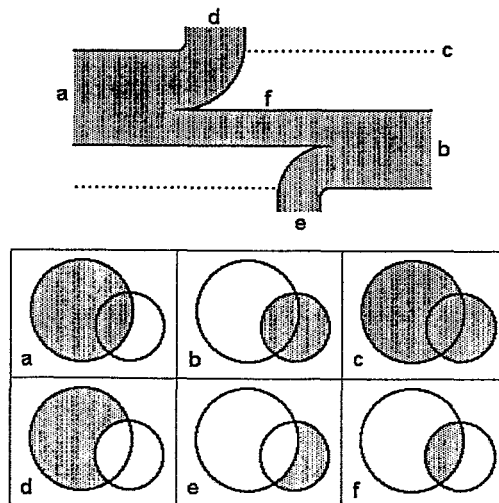
$$= H(x_n | y) + H(y) \tag{3.2-11a}$$

$$= H(y | x_n) + H(x) = H(y, x_n) \tag{3.2-11b}$$

Die Transinformation ist damit auch darstellbar durch

$$H(y; x_n) = H(y) + H(x_n) - H(y, x_n) \tag{3.2-12}$$

Die Beziehungen zwischen den verschiedenen Entropien verdeutlicht Bild 3.3 anhand zweier gebräuchlicher Darstellungsformen: (1) das aus der Nachrichtenübertragung stammende, den Informationsfluß hervorhebende Kanalmodell und (2) die allgemeineren Venn-Diagramme der Entropie.



**Bild 3.3** Die eingeführten Entropien in Flußbild und Venn-Diagrammen

- |                |                  |                 |                  |
|----------------|------------------|-----------------|------------------|
| a: $H(x_n)$    | Merkmalsentropie | d: $H(x_n   y)$ | Äquivokation     |
| b: $H(y)$      | Klassenentropie  | e: $H(y   x_n)$ | Irrelevanz       |
| c: $H(x_n, y)$ | Verbundentropie  | f: $H(x_n; y)$  | Transinformation |

Im Flußbild ist auf den in der Nachrichtenübertragung oft angegebenen Richtungssinn (Informationsfluß von a nach b, d als ab- und e zufließende Information) verzichtet, denn die Entropien sind ihrem Wesen nach ungerichtet. Die Venn-Diagramme vermeiden diesbezügliche Unklarheiten.

## 4. Präzisierung der Zielstellung

### 4.1 Zur Klassifikation

Die allgemeine Forderung an eine Klassifikation, nämlich a) unbekannte Objekte möglichst eindeutig auf die Klassen abzubilden, wird im Hinblick auf die sicherheitsmotiviertete Forderung nach Transparenz und Nachvollziehbarkeit der Erkennung erweitert um die Forderung b) nach möglichst eindeutiger Abbildung der Klassen auf die Objektmerkmale. Zusammengefaßt besteht damit die Forderung c) nach Eineindeutigkeit zwischen den Merkmalen und den Klassen. Diese Forderungen werden nun informationstechnisch formuliert.

- a) Die Ungewißheit über die Zugehörigkeit eines Merkmals  $x_n$  zu einer bestimmten Klasse  $y_k$  soll unter der Bedingung, daß das Merkmal  $x_n$  mit Gewißheit vorliegt, minimal sein. Dies entspricht der Forderung nach minimaler Irrelevanz, ausgedrückt durch das Kriterium (vgl. (3.2-8))

$$H(y_k|x_n) \rightarrow \underset{x_n}{\text{Min!}} = 0. \quad (4.1-1)$$

Im Sonderfall  $H(y_k|x_n)=0$  impliziert das Merkmal  $x_n$  eindeutig die Klasse  $y_k$ . Bei  $H(y_k|x_n)>0$  besteht hingegen keine Gewißheit über die Zugehörigkeit der Klasse  $y_k$  zum Merkmal  $x_n$ ; das Merkmal kann nur unscharf auf die Klasse abgebildet werden. Diese merkmalsbedingt bestmögliche Abbildungsgenauigkeit auch praktisch zu realisieren, wenigstens näherungsweise, ist die Aufgabe eines Klassifikators. Die merkmalsbedingt bestmögliche Abbildungsgenauigkeit ist durch die in den Referenzmustern enthaltene A-priori-Information determiniert. Inwieweit und wie ein Klassifikator diese Information ausnutzt, darin unterscheiden sich die verschiedenen Klassifizierungsverfahren. In dem - praktisch leider nur seltenen - Sonderfall  $H(y_k|x_n)=0$  ist für die Beziehung zwischen  $x_n$  und  $y_k$  eine Klassifikation unnötig.

- b) Die Forderung nach minimaler Ungewißheit über das Vorliegen des Merkmals  $x_n$  unter der Bedingung, daß eine bestimmte Klasse  $y_k$  vorliegt, bedeutet die Minimierung der Äquivokation; das entsprechende Kriterium ist (vgl. (3.2-6))

$$H(x_n|y_k) \rightarrow \underset{y_k}{\text{Min!}} = 0. \quad (4.1-2)$$

- c) Die aus (a) und (b) resultierende allgemeinere Forderung nach eineindeutiger Abbildung zwischen den Merkmalen und den Klassen verlangt

- den maximalen klassenbezogenen Informationsgehalt der Merkmale  $H(x_n)-H(x_n|y_k)=H(x_n;y_k)$  oder äquivalent
- den maximalen merkmalsbezogenen Informationsgehalt der Klassen  $H(y_k)-H(y_k|x_n)=H(y_k;x_n)$ ,



also die maximale Transinformation, ausgedrückt durch das Kriterium <sup>1</sup>

$$H(x_n; y_k) = H(y_k; x_n) \rightarrow \text{Max! } \forall x_n, y_k. \quad (4.1-3)$$

Die Minimierung der Irrelevanz ist bei feststehenden Klassen durch die Modifikation der Merkmale erreichbar. Die Minimierung der Äquivokation wird bei feststehenden Merkmalen durch die Modifikation der Klassen erreicht, sofern dies allerdings überhaupt zulässig ist. Für die Maximierung der Transinformation kommt sowohl die Merkmals- als auch die Klassenmodifikation in Betracht.

## 4.2 Modifikationen

Die Transinformation gestattet die Bewertung von Merkmalen hinsichtlich ihres Klassenbezuges, ihrer Klassenrelevanz. Nicht möglich ist es jedoch, Merkmale mit hoher Klassenrelevanz auf informationstheoretischem Wege systematisch zu gewinnen. Auch allgemein ist die Merkmalsgewinnung problematisch, denn zumeist stehen dafür unübersehbar viele Objekteigenschaften zur Auswahl. Deswegen werden hier die Merkmale wie üblich vorwiegend subjektiv (heuristisch, intuitiv) gewonnen. Erfahrungsgemäß wird so aber keine allzu hohe Ausbeute an klassenrelevanten Merkmalen erreicht. Deshalb werden, um eine Mindestanzahl klassenrelevanter Merkmale sicherzustellen, zunächst viel mehr Merkmale definiert (Abschnitt 5), als letztlich genutzt (bei der Klassifikation praktisch berücksichtigt) werden können. Anschließend (Abschnitt 6) werden die relevantesten dieser Merkmale gemäß (4.1-3) nach der maximalen Transinformation selektiert.

Bestandteil der Merkmalsgewinnung ist es, bereits definierte Merkmale durch

- funktionale und vorzugsweise logische Verknüpfungen (z.B. Konjunktionen, s. Abschnitt 6.2.2) zu verallgemeinern oder durch
- eine Unterteilung, bei der die Teile logisch disjunktiv miteinander verknüpft sind, zu spezialisieren

und auf diesem Wege eigenständige Merkmale zu definieren. Im Unterschied zu den Verallgemeinerungen setzen die Spezialisierungen Zusatzinformation voraus. Diese wird z.B. aus den Referenzmustern und dabei speziell aus Analogien zwischen Merkmalen und Klassen gewonnen, s. Abschnitt 6.1. Weil aber das Vorhandensein der benötigten Zusatzinformation nicht garantiert ist, empfiehlt es sich, die Merkmale von vorn herein so speziell zu definieren, daß nachträgliche Spezialisierungen möglichst unnötig sind.

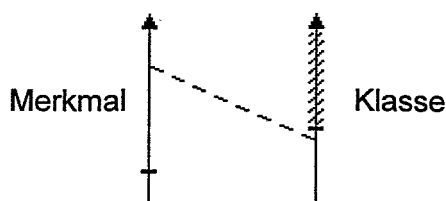
Die Begriffe Verallgemeinerung und Spezialisierung werden unter dem Oberbegriff Modifikation zusammengefaßt.

---

<sup>1</sup> Das Kriterium (4.1.3) ist weitreichender als die Kriterien (4.1-1) und (4.1-2), denn die Transinformation bewertet den wechselseitigen Informationsfluß zwischen den Merkmalen und den Klassen, wohingegen die beiden bedingten Entropien nur einseitige Informationsflüsse bewerten: die Irrelevanz den Informationsfluß von den Merkmalen zu den Klassen und die Äquivokation den Informationsfluß von den Klassen zu den Merkmalen.

Im Unterschied zur Modifikation von Merkmalen, die allenfalls praktischen Grenzen unterliegt, darf eine Modifikation von Klassen der bestehenden Klassifikationsaufgabe nicht zuwiderlaufen. Die Modifikation darf die Aussagekraft bzw. den Informationswert von bereits eingeführten Klassen zwar erhöhen, jedoch nicht verringern. Deswegen muß aus modifizierten Klassen eindeutig zurück auf die ursprünglichen Klassen geschlossen werden können. Folglich sind nur Spezialisierungen von Klassen erlaubt, denn nur aus einer Unterklasse kann mit Gewißheit auf die zugehörige Oberklasse - die ursprüngliche Klasse - geschlossen werden. Dagegen wäre es nicht möglich, aus einer Oberklasse mit Gewißheit auf eine ihrer Unterklassen - die ursprüngliche Klasse - zu schließen <sup>I</sup>.

Zur besseren Anschauung zeigt Bild 4.2-1 schematisch die genannten Möglichkeiten der Verallgemeinerung und Spezialisierung anhand von Abstraktionsskalen <sup>II</sup>. Markiert ist der bestehende Abstraktionsgrad eines Merkmals und einer Klasse. Die Schraffur rechts bedeutet, daß Klassen nicht verallgemeinert werden dürfen. Die Strichlinie zwischen den beiden Skalen bezeichnet die anzustrebende optimale Merkmale-Klassen-Beziehung (maximale Transinformation). Um ausgehend von den beiden Marken zu der optimalen Beziehung zu gelangen, muß das sehr allgemein definierte Merkmal stark verallgemeinert und die Klasse wenig spezialisiert werden.



**Bild 4.2-1** Abstraktionsskalen für ein Merkmal und eine Klasse.

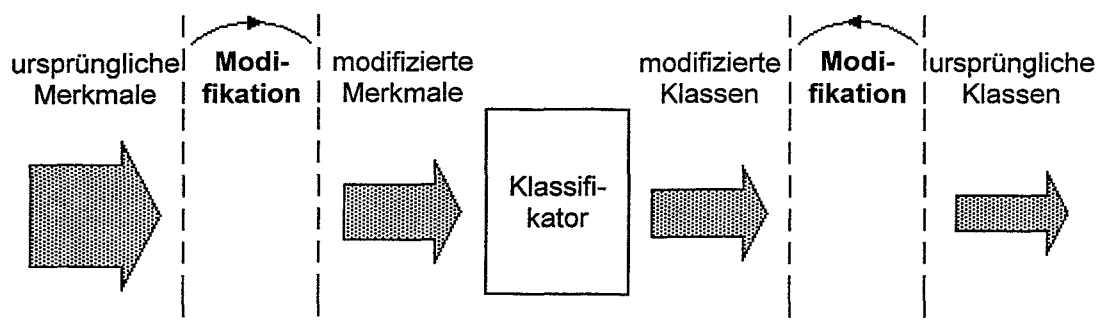
- ▶ : zunehmender Abstraktionsgrad (relativ)
- | : Marken für ein Merkmal und eine Klasse
- //// : nicht erlaubter Bereich
- - - : Merkmal-Klasse-Beziehung mit maximaler Transinformation

In dem bei der Klassifikation wirksamen, von den Merkmalen zu den Klassen gerichteten Informationsfluß befinden sich die modifizierten Merkmale in Vorwärts-, die modifizierten Klassen in Rückwärtsrichtung und der Klassifikator zwischen den

<sup>I</sup> Dies ist informationstheoretisch begründbar: Von zwei gegebenen disjunkten Klassen liege eine mit Gewißheit vor. Die Wahrscheinlichkeitsverteilung für das Vorliegen beider Klassen enthält demnach nur die Werte 1 und 0. Folgerichtig ist gemäß (3.2-4)  $H(p) = -p \cdot \log_2 p - (1-p) \cdot \log_2 (1-p) = -1 \cdot \log_2 1 - (1-1) \cdot \log_2 (1-1) = 0$ ; es besteht keine Ungewißheit. Die beiden Klassen werden zu einer Oberklasse verallgemeinert. Ist nun nur das Vorliegen der Oberklasse bekannt, gelten für die beiden ursprünglichen Klassen gleichverteilte Wahrscheinlichkeiten. Somit besteht für das Vorliegen der ursprünglichen Klassen die mittlere Ungewißheit  $H(p) = -0,5 \cdot \log_2 0,5 - (1-0,5) \cdot \log_2 (1-0,5) = -\log_2 0,5 = 1$  bit, vgl. Bild 3.1; der Informationswert der Oberklasse ist um 1 bit geringer.

<sup>II</sup> Die beiden Abstraktionsskalen haben keinen absoluten und keinen gegenseitigen (Merkmale-Klassen-)Bezug. Die Abstraktheit bezieht sich hier nur auf (relative) Vergleiche zwischen verschiedenen Merkmalen oder zwischen verschiedenen Klassen.

Merkmale und den Klassen, s. Bild 4.2-2.



**Bild 4.2-2** Modifizierte Merkmale und Klassen innerhalb des bei der Klassifikation wirksamen Informationsflusses. Die Dickenabnahme der Pfeile bedeutet möglichen Informationsverlust auf dem Übertragungsweg.

Jede Modifikation kann die optimale Merkmale-Klassen-Beziehung verändern. Die Modifikation von Merkmalen und die Modifikation von Klassen können sich deshalb wechselseitig in Frage stellen und sollten deshalb ebenso wechselseitig (iterativ) vorgenommen werden. Dieses Problem ist Gegenstand weiterführender Untersuchungen.

Im vorliegenden Anwendungsfall sind die Klassenfreiheiten viel stärker als die Merkmalsfreiheiten eingeschränkt. Deswegen werden zuerst die Klassen und danach die Merkmale modifiziert, s. Abschnitt 6. Eine nochmalige Modifikation der Klassen erübrigt sich dann; die Iteration kann nach dem zweiten Schritt abbrechen.

## 5. Signalvorverarbeitung und Merkmalsbildung

### 5.1 Mitlaufendes Zeitfenster

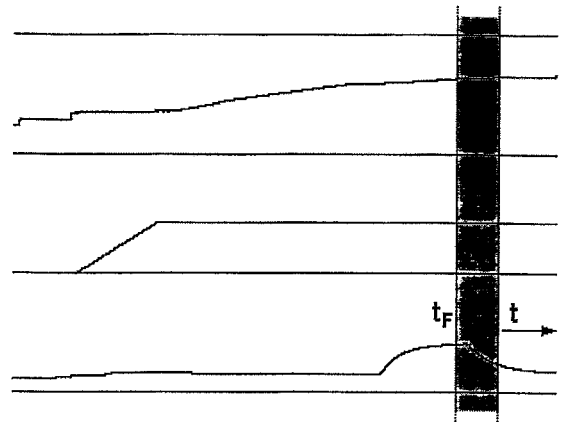
Schon eine einfache Sichtung der gegebenen Signalverläufe  $S_1$  bis  $S_{11}$  (s. Bild A1-2) zeigt, daß die Signale weder als periodisch noch als stochastisch einzuschätzen sind. Deswegen wird auf die Bildung von Frequenz- und Statistik-Kenngrößen verzichtet. Die gleitenden Übergänge innerhalb von verschiedenen Verläufen machen deutlich, daß Signalwerte eines bestimmten Zeitpunktes nicht nur von aktuellen Einflüssen (z.B. Steuereingriffen, Störungen) abhängen, sondern auch von der Vergangenheit. Dies weist auf das Vorhandensein einer Signaldynamik<sup>I</sup> hin. Um diese Eigenschaft zugänglich zu machen, werden die Signale in Anlehnung an [P91] in einem Zeitintervall (Fenster) erfaßt<sup>II</sup>. Im Fenster werden dann verschiedene Signalaufbereitungen vorgenommen und die Merkmale definiert.

Das Zeitfenster erstreckt sich zwischen dem aktuellen Zeitpunkt  $t$  und einem in kon-

<sup>I</sup> Signaldynamik: Der aktuelle Signalwert hängt nicht nur von aktuellen, sondern auch von vergangenen Einflüssen auf das vorliegende System ab.

<sup>II</sup> Der Objektbegriff bezieht sich nunmehr auf alle im Fenster enthaltenen Signalabschnitte.

stantem Abstand zurückliegenden Zeitpunkt  $t_F$  ( $t-t_F$ : Fensterbreite). Das Fenster erfaßt alle Signalverläufe (vgl. Bild 5.1-1).



**Bild 5.1-1** Zeitfenster zur Erfassung der Signaldynamik.  
 $t$ : aktueller Zeitpunkt,  $t-t_F$ : Fensterbreite

Die Wahl der Fensterbreite erfordert einen Kompromiß, denn sie sollte hinsichtlich

- vollständiger Erfassung der Signaldynamik groß sein, aber
- geringen Zeitverzuges bei der späteren Klassifikation <sup>‡</sup> klein sein.

Beim Vorhandensein eines dynamischen Prozeßmodells können bessere bzw. effektivere Merkmale gebildet werden, woraufhin letztlich die Klassifikation mit einer kleineren Merkmalsanzahl auskommen kann. Im besonderen kann die Fensterbreite verringert werden; anhand eines Zustandsraummodells braucht die Prozeßdynamik im Extremfall nur durch die zum aktuellen Zeitpunkt vorliegenden Zustandsgrößen ausgedrückt werden.

## 5.2 Polynom- und Exponential-Approximation

Von den verschiedenen Erscheinungen, die in den vorliegenden Signalen erkennbar sind, wird nur eine, dafür aber besonders deutliche, herausgegriffen: Das Signal  $S_{11}$  (s. Bild A1-2) weist annähernd exponentielle Abschnitte auf. Diese Eigenschaft wird dadurch besser zugänglich gemacht, daß der Fensterinhalt von  $S_{11}$  exponentiell approximiert wird. Der zeitvariante, mit der Fensterposition veränderliche Exponent der Approximationsfunktion wird dann als neues Signal  $S_{12}$  definiert.

Zunächst erfolgt eine Approximation des Fensterinhaltes von  $S_{11}$  (in diesem Abschnitt kurz  $S$ ) durch ein Ausgleichspolynom 2. Grades

$$\hat{S} = a_1 + a_2\tau + a_3\tau^2 \quad (5.2-1)$$

mit  $\tau \in [t_F, t]$ : relative Zeit im Fenster ( $t$ : aktueller Zeitpunkt,  $t-t_F$ : Fensterbreite)  
 $\hat{S} = \hat{S}(\tau)$ : Schätzgröße von  $S=S(\tau)$   
 $a_1, a_2, a_3$ : Parameter

<sup>‡</sup> Die Klassifikation kann frühestens zum Zeitpunkt  $t_F$  beginnen.

unter Berücksichtigung von fünf äquidistant über die Fensterbreite verteilten Werten:  $S_k = S(\tau_k)$  mit  $\tau_k = t_F + (k-1)\Delta T$ ,  $k \in \{1, 2, \dots, 5\}$ ,  $\Delta T = (t - t_F)/4$ . Das Optimierungskriterium

$$\sum_{k=1}^5 (\hat{S}_k - S_k)^2 = \text{Min!} \quad (5.2-2)$$

$a_1, a_2, a_3$

liefert die optimalen Schätzparameter [Sch92]

$$(a_1^* \ a_2^* \ a_3^*) = \underline{m} M^{-1}$$

mit  $\underline{m} = (m_1 \ m_2 \ m_3)$ ,  $m_i = \sum_{k=1}^5 \tau_k^{i-1} S_k$

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}, \quad m_{ij} = \sum_{k=1}^5 \tau_k^{i+j-2}, \quad i, j \in \{1, 2, 3\}.$$

Die drei Schätzwerte

$$\begin{aligned} \hat{S}_I &= \hat{S}^*(\tau_1) = \hat{S}^*(t_F), \\ \hat{S}_{II} &= \hat{S}^*(\tau_3) = \hat{S}^*(t_F + \frac{1}{2}(t - t_F)), \\ \hat{S}_{III} &= \hat{S}^*(\tau_5) = \hat{S}^*(t) \end{aligned}$$

werden interpoliert nach dem exponentiellen Ansatz ( $b_1, b_2, b_3$ : Parameter)

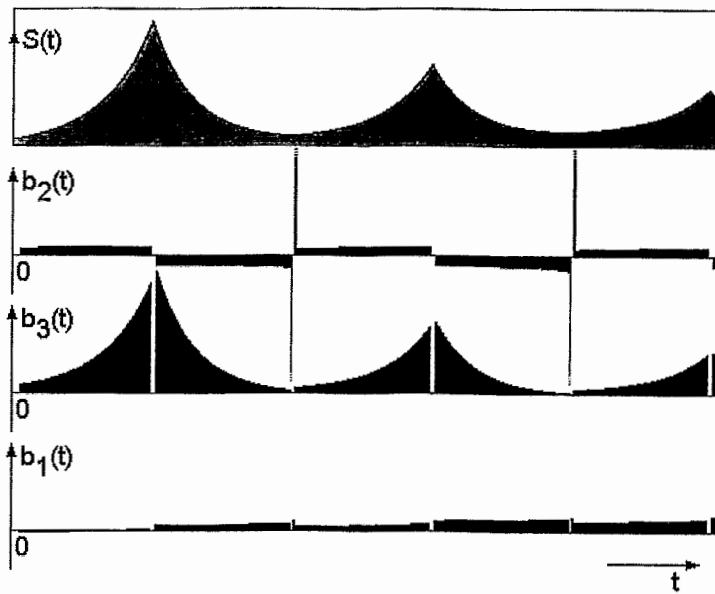
$$b_1 e^{b_2} + b_3. \quad (5.2-3)$$

Die Lösung ist

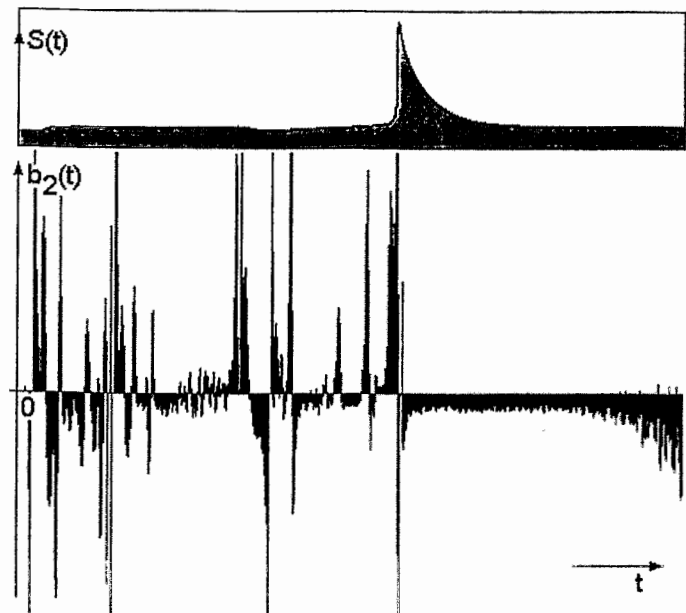
$$b_1 = \frac{(\hat{S}_{II} - \hat{S}_I)^2}{\hat{S}_I + \hat{S}_{III} - 2\hat{S}_{II}}, \quad b_2 = \ln \frac{\hat{S}_{III} - \hat{S}_{II}}{\hat{S}_{II} - \hat{S}_I}, \quad b_3 = \frac{\hat{S}_I \hat{S}_{III} - \hat{S}_{II}^2}{\hat{S}_I + \hat{S}_{III} - 2\hat{S}_{II}}$$

mit den Monotoniebedingungen  $\hat{S}_I < \hat{S}_{II} < \hat{S}_{III}$  und  $\hat{S}_{III} < \hat{S}_{II} < \hat{S}_I$ .

Von den eingeführten Parametern  $a_1^*$ ,  $a_2^*$ ,  $a_3^*$ ,  $b_1$ ,  $b_2$ ,  $b_3$  erweist sich nur  $b_2$  (der Exponent in (5.2-3)) als genügend relevant für die bestehende Klassifizierungsaufgabe. Bild 5.2-1 zeigt einen konstruierten, aus annähernd exponentiellen Teilverläufen zusammengesetzten Signalverlauf  $S(t)$  und die daraus bestimmten Parameterverläufe  $b_1(t)$ ,  $b_2(t)$ ,  $b_3(t)$ . Die Signaleigenschaft "exponentiell" äußert sich darin, daß  $b_2$  zumindest näherungsweise konstant ist; so gesehen ist der absolute Wert von  $b_2$  unerheblich. Bild 5.2-2 zeigt den  $b_2(t)$ -Verlauf für das Signal  $S_{11}$  aus der Realisierung V (s. Bild A1-2.V); die Eigenschaft "exponentiell" ist rechts im Bild deutlich sichtbar.



**Bild 5.2-1** Parameterverläufe für einen konstruierten Verlauf  $S(t)$



**Bild 5.2-2**  $b_2(t)$ -Verlauf für das Signal  $S_{11}$  aus Realisierung V (Bild A1-2.V)

### 5.3 Gleitende Normierung

Die gegebenen Signale (s. Anhang 1) wurden vorab aus technischen Gründen unter Beachtung ihres Langzeitverhaltens auf den Wertebereich  $[0,1]$  normiert. Diese globale Normierung ist für die beabsichtigte situationsbezogene Klassifikation ungünstig. Weil nämlich die Signale

- nur über ein vergleichsweise schmales Zeitfenster erfaßt werden (s. Abschnitt 5.1) und
- nur mit begrenzter Auslösung wertdiskretisiert werden können (Abschnitt 5.4),

kann Information aus sehr flach verlaufenden Signalabschnitten verlorengehen. Deswegen wird eine Transformation vorgenommen, welche die Information aus flachen Signalabschnitten besser zugänglich macht. Diese Transformation ist eine zeitvariante bzw. lokale Normierung, die den Signalabschnitt im Zeitfenster in seinem Wertintervall um so mehr spreizt, je flacher er ist. Diese Transformation wird für jedes Signal unabhängig vorgenommen.

Der Zeitfensterinhalt eines Signals  $S(\tau)$ ,  $\tau \in [t_F, t]$ ,  $S \in [0,1]$  wird transformiert in

$$S^s(\tau) = \frac{S_{\max} S_{\min}^s - S_{\min} S_{\max}^s}{S_{\max} - S_{\min}} + \frac{S_{\max}^s - S_{\min}^s}{S_{\max} - S_{\min}} S(\tau) \quad \text{für } S_{\max} > S_{\min},$$

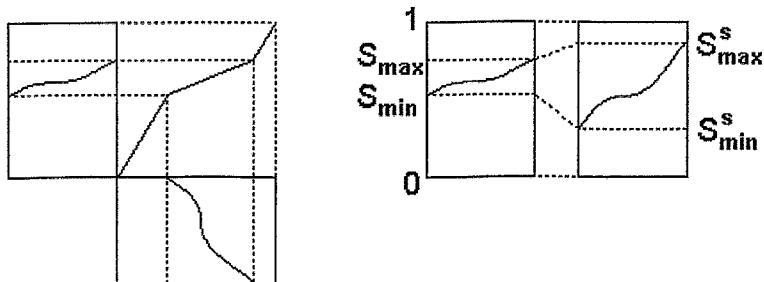
$$= S(\tau) \quad \text{für } S_{\max} = S_{\min}, \quad (5.3-1)$$

mit  $S_{\min}, S_{\max}$ : Extremwerte von  $S(\tau)$

$$\begin{aligned} S_{\min}^s &= S_{\min}(1 - r): \\ S_{\max}^s &= S_{\max} + (1 - S_{\max})r: \end{aligned} \quad \text{Extremwerte von } S^s(\tau)$$

$$r = 0,4 \frac{50(S_{\max} - S_{\min})}{1 - 50(S_{\max} - S_{\min})}: \quad \text{Korrekturfaktor}$$

Die Transformation (5.3-1) vergrößert das  $S(\tau)$ -Wertintervall  $[S_{\min}, S_{\max}]$ , indem sie die Seitenintervalle  $[0, S_{\min}]$ ,  $[S_{\max}, 1]$  verkleinert auf  $[0, S_{\min}^s]$ ,  $[S_{\max}^s, 1]$ , s. Bild 5.3-1. Die Transformation ist nichtlinear, doch wird das Intervall  $[S_{\min}, S_{\max}]$  linear auf das Intervall  $[S_{\min}^s, S_{\max}^s]$  abgebildet. Weiterhin ist  $S^s \in [0,1]$ , d.h. die transformierten Signalabschnitte verlassen den Wertebereich  $[0,1]$  des Signals  $S$  nicht. Bild 5.3-2 verdeutlicht die Zeitvarianz der Spreizung des Fensterinhaltes: Je flacher  $S$ , desto stärker ist die Spreizung in  $S^s$  im Verhältnis zu  $S$ .



**Bild 5.3-1** Zur Verdeutlichung der Transformation (5.3-1)

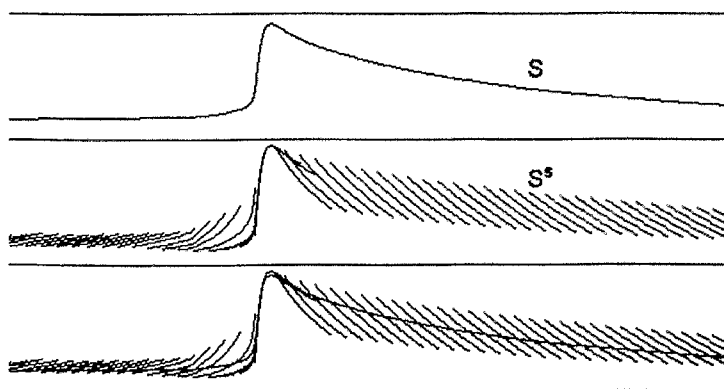
links: Die Transformations-Kennlinie besteht aus drei Geradenstücken. Der Kurvenabschnitt wird unter etwaiger Beibehaltung seiner absoluten Lage linear gespreizt.  
rechts: andere Darstellung der links skizzierten Situation

**Bild 5.3-2** Gleitende Normierung an einem realen Signalverlauf

oben:  $S_{11}$ -Ausschnitt aus Realisierung IV (hier in maximaler zeitlicher Auflösung, vgl. Bild A1-2.IV)

Mitte: Gespreizte Fensterinhalte (tatsächlich werden die  $S^s$ -Segmente lückenlos gebildet).

unten: vergleichende Überlagerung



Zusätzlich zu dem absoluten Signal  $S^s$  wird das relative, d.h. mit seinem rechten Ende bzw. zum aktuellen Zeitpunkt  $t$  auf den Wert 0.5 fixierte Signal

$$S^{sr}(\tau) = S^s(\tau) - S^s(t) + 0,5, \quad \tau \in [t_F, t]$$

gebildet, s. Bild 5.4-1 Mitte rechts. Damit dieses Signal den Wertebereich  $[0, 1]$  nicht verläßt, wird es begrenzt:

$$S^{sr}(\tau) = \begin{cases} 1 \\ 0 \end{cases} \quad \text{für } S^s(\tau) - S^s(t) + 0,5 \begin{cases} > 1 \\ < 0 \end{cases} .$$



## 5.4 Diskretisierung und Merkmale

Die gespreizten Signalabschnitte im Zeitfenster (s. Abschnitt 5.3) gelten als kontinuierlich und analog. Sie werden relativ zum Fenster äquidistant zeitdiskretisiert (abgetastet). Die Signalwerte zu den Abtastzeitpunkten werden dann wertdiskretisiert und dabei so codiert, daß ein Wert durch eine minimale Anzahl binärer Merkmale ausgedrückt wird.

Zur Vermeidung von Informationsverlusten bei der Zeitdiskretisierung fordert das Abtasttheorem eine Abtastfrequenz von mindestens der doppelten oberen Signalgrenzfrequenz. Unter Vernachlässigung von Frequenzanteilen aus Unstetigkeiten und Knicken in den Signalen ist die größte obere Grenzfrequenz, die in den verschiedenen Signalen überhaupt vorkommt, noch vergleichsweise klein. Deswegen wird eine entsprechend grobe Zeitdiskretisierung vorgenommen: Fünf Abtastzeitpunkte im Zeitfenster; sie gelten für alle Signale. In absoluter Zeit sind die Abtastzeitpunkte definiert durch

$$t - (k-1)(t-t_F)/5 \quad \text{für } k=5,4,3,2,1.$$

Sie gehören jeweils zur rechten Seite der fünf gleichgroßen Zeitintervalle, in die das Fenster erschöpfend aufgeteilt ist, s. Bild 5.4-1 Mitte.

Für die Wertdiskretisierung wird der Signalwertebereich  $[0,1]$  in hierarchisch gestaffelte Intervalle aufgeteilt. Die Intervallhierarchie hat

- b Ebenen und
- c (Teil-)Wertintervalle je Ebene

$(b,c \in \mathbb{N})$ , also  $b \cdot c = a$  hierarchische Wertintervalle. Mit den hierarchischen Intervallen sind insgesamt (in höchster Auflösung)  $c^b$  elementare Intervalle unterscheidbar. Bei vorgegebener Anzahl der hierarchischen Intervalle  $a$  soll die Anzahl der elementaren Intervalle  $c^b$  maximiert werden:

$$c^b = \text{Max!}_{b,c} \quad \text{mit der Nebenbedingung } b \cdot c = a \quad (5.4-1)$$

Informationstheoretisch lautet dieselbe Aufgabe so: Aus vorgegebenen  $c^b$  elementaren Intervallen soll ein bestimmtes Intervall mit Hilfe von  $b$  Fragen, wobei jede Frage  $c$ -teilig ist ( $c$  Antwortmöglichkeiten<sup>1</sup> zuläßt), so erfragt werden, daß die einzelne Teilfrage den gemäß (3.2-3) maximalen Informationsgewinn sicherstellt. Die Lösung von (5.4-1) bzw. der Maximalwert der Entropiefunktion  $-1/c \log_2(1/c)$  besteht für  $c=e=2,718$  und ist  $H=0,531$  bit. Eine einzelne Teilfrage erbringt also den maximalen Informationsgewinn dann, wenn die erwartete Antwort die Wahrscheinlichkeit  $1/e$  hat. Unter der Bedingung  $c \in \mathbb{N}$  ist das Optimum allerdings bestimmt durch  $c=3$  und  $H = -1/3 \log_2(1/3) = 0,528$  bit. Somit sind 3-teilige Fragen zu stellen. Aus der Nebenbedingung in (5.4-1) folgt schließlich  $b = \lceil a/c \rceil$ , wobei ungeradzahlige  $b$ -Werte aufgerundet werden müssen.

---

<sup>1</sup> "Der Signalwert liegt im  $c$ -ten Intervall"; nach dem Gegenteil wird nicht gefragt.

Die Anzahl der hierarchischen Intervalle  $a$  bzw. die Anzahl der elementaren Intervalle  $c^b$  ist praktisch nicht frei wählbar. Eine Orientierung für diese Wahl bietet das Nutzsignal-Rausch-Verhältnis [V82]. Dieses Verhältnis bestimmt die sinnvolle Höchstzahl an elementaren Intervallen. Tendenziell gilt: Je größer der Rauschabstand, desto größer ist die sinnvolle Anzahl elementarer Intervalle. Eine feinere Auflösung bringt keinen Informationsgewinn, eine gröbere jedoch Informationsverlust. Gewählt werden  $a=9$  hierarchische Intervalle. Wegen  $c=3$  Intervallen je Hierarchieebene ergeben sich  $b=a/c=3$  Hierarchieebenen. Der Signalwertebereich  $[0,1]$  ist somit in  $c^b=3^3=27$  elementare Intervalle aufgeteilt.

Um durch  $c=3$ -teilige Fragen festzustellen, zu welchem der 27 elementaren Intervalle ein Signalwert gehört, sind bei angenommener Gleichwahrscheinlichkeit der Signalwerte genau  $b=3$  Fragen notwendig ( $b \cdot c=3 \cdot 3=9=a$ ) bei 0,528 bit Informationsgewinn pro Teilfrage. <sup>I II</sup>

Den hierarchischen Intervallen werden nun Merkmale zugewiesen. Weil ein Intervall einen Signalwert nur entweder enthält oder nicht enthält, kommen allein binäre Merkmale in Frage. Im 1. Abschnitt wurde festgelegt, daß der Wert 1 eines binären Merkmals die Zugehörigkeit dieses Merkmals zu einer Merkmalsmenge ausdrückt. Ergänzend dazu ist zu klären, ob das Enthaltensein oder das Nicht-Enthaltensein eines Signalwertes in einem Wertintervall als Merkmalswert 1 ausgedrückt werden soll. Kriterium hierfür sei der Realisierungsaufwand: Um einen Signalwert durch eine minimale Merkmalsanzahl ausdrücken zu können, muß der Informationsgewinn aus dem Bekanntwerden des einzelnen Merkmals maximal sein. Bei angenommener Gleichwahrscheinlichkeit der Signalwerte gelten für die beiden unterscheidbaren Fälle mit  $c=3$  die Wahrscheinlichkeiten

- Signalwert ist im Intervall enthalten:  $p=1/c=1/3$
- Signalwert ist im Intervall nicht enthalten:  $p=(c-1)/c=2/3$ .

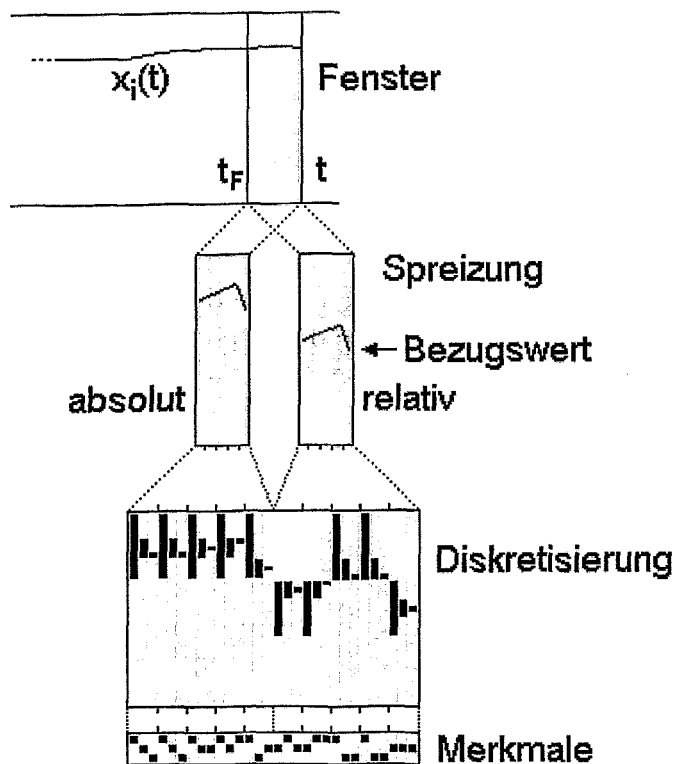
Weil gemäß (3.2-3)  $H(p=1/3)=0,528 \text{ bit} > H(p=2/3)=0,390 \text{ bit}$  ist, erbringt der erstgenannte Fall den größeren Informationsgewinn. Darum wird das Enthaltensein/Nicht-Enthaltensein eines Signalwertes in einem Intervall als Merkmalswert 1/0 ausgedrückt. <sup>III</sup>

<sup>I</sup> Mit  $c=2$ -teiligen Fragen sind mindestens  $b=5$  Fragen erforderlich ( $b \cdot c=5 \cdot 2=10 > 9=a$ ); der mittlere Informationsgewinn pro Teilfrage ist dann nur  $-1/2 \log_2(1/2)=0,5 \text{ Bit}$ .  $c=4$ -teilige Fragen erfordern mindestens  $b=3$  Fragen ( $b \cdot c=3 \cdot 4=12 > 9=a$ ) bei ebenfalls  $-1/4 \log_2(1/4)=0,5 \text{ Bit}$  pro Teilfrage. Bei einer einzigen ( $b=1$ )  $c=27$ -teiligen Frage ( $b \cdot c=1 \cdot 27=27 > 9=a$ ), also ohne die hierarchische Strukturierung ist der Informationsgehalt pro Teilfrage nur noch  $-1/27 \log_2(1/27)=0,176 \text{ Bit}$ .

<sup>II</sup> Die gefundene optimale hierarchische Intervallstrukturierung entspricht der Darstellung einer Zahl im ternären Zahlencode (Basis: 3, Zeichenvorrat: 0,1,2). So werden z.B. die dezimale 16 zur ternären 121 ( $1 \cdot 3^2 + 2 \cdot 3^1 + 1 \cdot 3^0=16$ ) oder die dezimale 26 (als 27. Element des 27-elementigen Zahlenbereiches  $\{0,1,2,\dots,26\}$ ) zur ternären 222.

<sup>III</sup> Auch aus der Erläuterung zum Bild 3.2 geht hervor, daß das Bekanntwerden des Vorliegens einer wenig wahrscheinlichen Objekteigenschaft informativer ist, als das Bekanntwerden des fast sicheren Nicht-Vorliegens dieser Eigenschaft.

Der Erläuterung zu Bild (3.1) folgend ist anzustreben, daß die Signalwerte in den Wertintervallen einer Hierarchieebene gleichwahrscheinlich auftreten. Gleichwahrscheinlichkeit der Signalwerte ist bei den originalen Signalen bei weitem nicht, bei den normierten Signalen (Abschnitt 5.3) aber bereits in guter Näherung gegeben. Um davon ausgehend das gleichwahrscheinliche Enthaltensein der Signalwerte in den Wertintervallen zu erreichen, können die Intervallbreiten an die noch etwas ungleichmäßige Häufigkeitsverteilung der Signalwerte angepaßt werden. Für den vorliegenden Fall wird eine für alle Signale gemeinsame Intervallstrukturierung angestrebt. Bild 5.4-1 unten zeigt dafür die Kompromißlösung. Darüber hinaus verdeutlicht das Bild die hierarchische Aufteilung der Intervalle, die Besetzung der Intervalle in einer konkreten Situation und die dazugehörige Bewertung der Merkmale.



**Bild 5.4-1** Signalvorverarbeitung und Merkmalsbildung an einem Beispiel.

oben: Zeitfenster über einem Signal (t: aktueller Zeitpunkt)

Mitte: gespreizter Signalabschnitt, absolut und relativ (Pfeil markiert festen Bezugspunkt für das rechte Ende des relativen Signalabschnittes). Am unteren Fensterend ist die Zeitdiskretisierung angedeutet.

unten: Wertdiskretisierung durch die Zuordnung der zeitäquidistanten Signalwerte zu hierarchischen (verschiedenen Auflösungen entsprechenden) Wertintervallen. Die Intervallzugehörigkeit der Signalwerte bestimmt direkt die Merkmalsbewertung:  $\blacksquare/\square \rightarrow 1/0$ . Die Merkmalsindizierung ist

$$\begin{array}{l} X_3 \quad X_6 \quad \dots \quad X_{90} \\ X_2 \quad X_5 \quad \dots \quad X_{89} \\ X_1 \quad X_4 \quad \dots \quad X_{87} \end{array}$$

Da jeder abgetastete Signalwert in 3 Auflösungsstufen zu je 3 Merkmalen abgebildet wird, 5 Zeitschritte berücksichtigt werden sowie die absolute und die relative Lage eines Signalabschnittes unterschieden werden, sind pro Signal  $3 \cdot 3 \cdot 5 \cdot 2 = 90$  binäre Merkmale definiert. Bei 12 verwendeten Signalen gibt es mithin  $90 \cdot 12 = 1080$  Merkmale. Wie noch gezeigt wird, sind diese Merkmale bei weitem nicht alle für die Lösung der bestehenden Klassifizierungsaufgabe geeignet.

## 5.5 Bildung von Referenzmustern

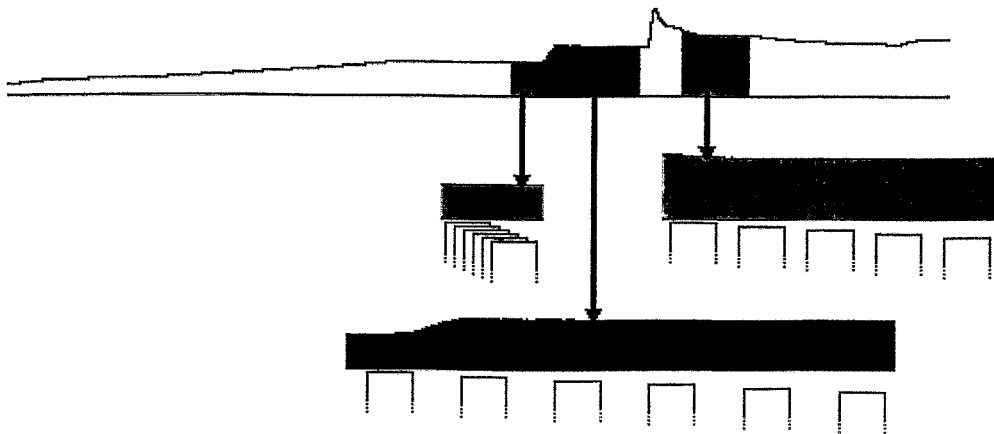
Die Gesamtheit der im Abschnitt 5.4 definierten Merkmale bildet den Merkmals-Teil eines Musters. Entsprechend der Besetzung oder Nichtbesetzung ihres Klassen-Teils werden die Muster in Referenzmuster und unbekannte Muster unterschieden, s. Abschnitt 2.

Unbekannte Muster (Muster mit nicht besetztem Klassen-Teil:  $y_k=0 \forall k$ ) werden fortlaufend (mit der Abtastperiode  $2s$ , siehe Anhang 1) erzeugt.

Referenzmuster können nur mit vorklassifizierten Objekten gebildet werden. Im vorliegenden Fall sind dies die gesamte Signalrealisierung I<sup>z</sup> und die vorklassifizierten Bereiche der Realisierungen II, III, IV (s. Anhang 1, Bild A1-2 unten, obere Zahlenreihe). Die Referenzmuster werden nicht fortlaufend, sondern nur für ausgewählte Zeitpunkte erzeugt.

Die Erzeugung von Referenzmustern veranschaulicht Bild 5.5-1. Die Zeitfenster, aus deren Inhalten die Referenzmuster hervorgehen, werden über die klassifizierten Signalabschnitte so verteilt, daß pro Klasse ähnlich viele Referenzmuster entstehen. Bei Realisierung I werden die Referenz-Zeitfenster über den gesamte Signalverlauf verteilt.

Da zu jedem Referenzmuster nicht mehr als eine Klasse gehört, sind die Klassen disjunkt.



**Bild 5.5-1** Erzeugung von Referenzmustern aus einer vorklassifizierten Signalrealisierung.

oben: Signal  $S_2$  aus Realisierung II mit Klassenzuordnung (unvollständig).

unten: vorklassifizierte Abschnitte aus  $S_2$  in maximaler zeitlicher Auflösung. Jeweils darunter sind Positionen ausgewählter Zeitfenster schematisch angegeben. Diejenigen Merkmale, die zu einem Fenster gehören, bilden gemeinsam mit dem Klassencode ein Referenzmuster.

---

<sup>z</sup> Die Realisierung I beschreibt einen normalen Prozeßverlauf. Diese Normalität entspricht zwar einer eigenständigen Klasse, wird aber nicht als eigenständige Klasse eingeführt, denn diese Klasse würde lediglich das Nichtvorhandensein der Gefahrenklassen ausdrücken und wäre daher redundant.

## 6. Die Transinformation als Bewertungsmaß

### 6.1 Bewertung und Modifikation der Klassen

Den Ausführungen im Abschnitt 4 folgend wird zuerst untersucht, ob die drei vorgegebenen Klassen informationstheoretisch sinnvoll sind oder ob sie besser unterteilt werden sollten. Die Problematik wird an einem einfachen konstruierten Beispiel erläutert.

**Beispiel 6.1:** Gegeben sind Referenzvektoren  $\underline{x}_m, m \in \{1,2,3\}$  mit Merkmalen  $x_{mk}, k \in \{1,2,3\}$  und Klassenbezeichnungen  $y_{mk}, k \in \{1,2\}$ , siehe nachfolgendes Schema (die Zeilen entsprechen den Vektoren).

	$x_{m1}$	$x_{m2}$	$x_{m3}$	$y_{m1}$	$y_{m2}$
$m=1$	1	0	0	1	0
2	0	1	0	1	0
3	0	0	1	0	1

Daraus werden folgende Wahrscheinlichkeiten abgelesen:

$$p(x_1=1)=p(x_2=1)=p(x_3=1)=1/3$$

$$p(x_1=0)=p(x_2=0)=p(x_3=0)=2/3$$

$$p(y_1=1)=2/3$$

$$p(y_2=1)=1/3$$

$$p(y_1=0)=1/3$$

$$p(y_2=0)=2/3$$

$$p(x_1=1|y_1=1)=1/2$$

$$p(x_2=1|y_1=1)=1/2$$

$$p(x_3=1|y_1=1)=0$$

$$p(x_1=0|y_1=1)=1/2$$

$$p(x_2=0|y_1=1)=1/2$$

$$p(x_3=0|y_1=1)=1$$

$$p(x_1=1|y_1=0)=0$$

$$p(x_2=1|y_1=0)=0$$

$$p(x_3=1|y_1=0)=1$$

$$p(x_1=0|y_1=0)=1$$

$$p(x_2=0|y_1=0)=1$$

$$p(x_3=0|y_1=0)=0$$

$$p(x_1=1|y_2=1)=0$$

$$p(x_2=1|y_2=1)=0$$

$$p(x_3=1|y_2=1)=1$$

$$p(x_1=0|y_2=1)=1$$

$$p(x_2=0|y_2=1)=1$$

$$p(x_3=0|y_2=1)=0$$

$$p(x_1=1|y_2=0)=1/2$$

$$p(x_2=1|y_2=0)=1/2$$

$$p(x_3=1|y_2=0)=0$$

$$p(x_1=0|y_2=0)=1/2$$

$$p(x_2=0|y_2=0)=1/2$$

$$p(x_3=0|y_2=0)=1$$

$$p(y_1=1|x_1=1)=1$$

$$p(y_1=1|x_2=1)=1$$

$$p(y_1=1|x_3=1)=0$$

$$p(y_1=0|x_1=1)=0$$

$$p(y_1=0|x_2=1)=0$$

$$p(y_1=0|x_3=1)=1$$

$$p(y_1=1|x_1=0)=1/2$$

$$p(y_1=1|x_2=0)=1/2$$

$$p(y_1=1|x_3=0)=1$$

$$p(y_1=0|x_1=0)=1/2$$

$$p(y_1=0|x_2=0)=1/2$$

$$p(y_1=0|x_3=0)=0$$

$$p(y_2=1|x_1=1)=0$$

$$p(y_2=1|x_2=1)=0$$

$$p(y_2=1|x_3=1)=1$$

$$p(y_2=0|x_1=1)=1$$

$$p(y_2=0|x_2=1)=1$$

$$p(y_2=0|x_3=1)=0$$

$$p(y_2=1|x_1=0)=1/2$$

$$p(y_2=1|x_2=0)=1/2$$

$$p(y_2=1|x_3=0)=0$$

$$p(y_2=0|x_1=0)=1/2$$

$$p(y_2=0|x_2=0)=1/2$$

$$p(y_2=0|x_3=0)=1$$

Daraus folgen nach (2.2-2) die Merkmalsentropien

$$H(x_1) = -p(x_1=1)\log_2 p(x_1=1) - p(x_1=0)\log_2 p(x_1=0) = \underline{0,918 \text{ bit}} = H(x_2) = H(x_3),$$

nach (2.2-4) die Klassenentropien

$$H(y_1) = -p(y_1=1)\log_2 p(y_1=1) - p(y_1=0)\log_2 p(y_1=0) = \underline{0,918 \text{ bit}} = H(y_2),$$

nach (2.2-5a) die Äquivokationen

$$H(x_1|y_1) = -p(y_1=1)[p(x_1=1|y_1=1)\log_2 p(x_1=1|y_1=1) + p(x_1=0|y_1=1)\log_2 p(x_1=0|y_1=1)] \\ -p(y_1=0)[p(x_1=1|y_1=0)\log_2 p(x_1=1|y_1=0) + p(x_1=0|y_1=0)\log_2 p(x_1=0|y_1=0)]$$

$$= \underline{0.667 \text{ bit}} = H(x_2|y_1) = H(x_1|y_2) = H(x_2|y_2)$$

$$H(x_3|y_1) = H(x_3|y_2) = \underline{0}$$

nach (2.2-7a) die Irrelevanzen

$$H(y_1|x_1) = -p(x_1=1)[p(y_1=1|x_1=1)\log_2 p(y_1=1|x_1=1) + p(y_1=0|x_1=1)\log_2 p(y_1=0|x_1=1)] \\ -p(x_1=0)[p(y_1=1|x_1=0)\log_2 p(y_1=1|x_1=0) + p(y_1=0|x_1=0)\log_2 p(y_1=0|x_1=0)]$$

$$= \underline{0.667 \text{ bit}} = H(y_2|x_1) = H(y_1|x_2) = H(y_2|x_2)$$

$$H(y_1|x_3) = H(y_2|x_3) = \underline{0}$$

und nach (2.2-9a) die Transinformationen

$$H(y_1;x_1) = H(y_1) - H(y_1|x_1) = \underline{0.251 \text{ bit}} = H(y_2;x_1) = H(y_1;x_2) = H(y_2;x_2)$$

$$H(y_1;x_3) = H(y_2;x_3) = \underline{0.918 \text{ bit}}$$

Wegen  $H(x_3|y_1)=H(x_3|y_2)=H(y_1|x_3)=H(y_2|x_3)=0$  bzw.  $H(x_3)=H(y_1;x_3)=H(y_2;x_3)=H(y_1)=H(y_2)=0.918 \text{ bit}$  ist der Zusammenhang zwischen dem Merkmal  $x_3$  und den Klassen  $y_1$  und  $y_2$  eineindeutig: Aus  $x_3=1/0$  folgt  $y_1=0/1$ ,  $y_2=1/0$  und umgekehrt.

Dagegen besteht zwischen den Merkmalen  $x_1$ ,  $x_2$  und der Klasse  $y_1$  in beiden Richtungen nicht einmal Eindeutigkeit: z.B. aus  $x_1=1$  folgt  $y_1=1$  und aus  $y_1=0$  folgt  $x_1=0$ , aber aus  $x_1=0$  ist nicht entscheidbar, ob  $y_1=1$  oder  $y_1=0$  vorliegt und aus  $y_1=1$  ist nicht entscheidbar, ob  $x_1=1$  oder  $x_1=0$  vorliegt; folglich ist die Transinformation - da sie ja gegenseitige bzw. ungerichtete Bewertungen vornimmt - mit  $H(y_1;x_1)=0.251 \text{ bit}$  erheblich kleiner als die Merkmals- und die Klassenentropie  $H(x_1)=H(y_1)=0.918 \text{ bit}$ .

Nun wird die Klasse  $y_1$  unterteilt, wobei die dafür notwendige Zusatzinformation aus den Merkmalen hervorgeht. Analog zu den Merkmalen  $x_{m1}$  und  $x_{m2}$  entstehen aus der Klasse  $y_1$  die Unterklassen  $y_{m1.1}$  und  $y_{m1.2}$ :

	$x_{m1}$	$x_{m2}$	$x_{m3}$	$y_{m1.1}$	$y_{m1.2}$	$y_{m2}$
$m=1$	1	0	0	1	0	0
2	0	1	0	0	1	0
3	0	0	1	0	0	1

Wegen  $H(y_{1.1};x_1)=H(y_{1.2};x_2)=H(x_1)=H(x_2)=H(y_{1.1})=H(y_{1.2})=0.918 \text{ bit}$  ist jetzt der Zusammenhang zwischen allen Merkmalen und allen Klassen eineindeutig. Wie für Klassenmodifikationen gefordert (s. Abschnitt 4) kann aus den modifizierten Klassen auf die ursprüngliche Klasse geschlossen werden: Unabhängig davon, ob nur eine der beiden Unterklassen  $y_{1.1}$ ,  $y_{1.2}$  oder beide gemeinsam nachgewiesen wurden, ist damit auch ihre Oberklasse (ursprüngliche Klasse)  $y_1$  nachgewiesen; die Oberklasse ist die logische Disjunktion der beiden Unterklassen:  $y_1=y_{1.1} \vee y_{1.2}$ <sup>†</sup>

Aus dem Beispiel 6.1 folgt allgemein, daß eine Unterteilung einer Klasse (oder eine entsprechende Zusammenfassung von Merkmalen) dann zweckmäßig ist, wenn

- zwar ein eindeutiger Bezug mehrerer Merkmale auf eine Klasse besteht aber
- der Bezug dieser Klasse auf die betreffenden Merkmale mehrdeutig ist.

Solche Umstände bestehen im realen Anwendungsfall für die Klassen  $y_1$  und  $y_2$  (s. Anhang 1). Die Vorklassifikation hat jede dieser Klassen je zwei weitgehend verschiedenen Prozeßsituationen mit entsprechend unterschiedlichen, sich sogar gegenseitig ausschließenden bzw. disjunkten Merkmalen zugeordnet. Mit Bezug zu diesen Merkmalen wird deshalb die Teilung der Klasse  $y_1$  in die Unterklassen  $y_{1.1}$

<sup>†</sup> Die logische Disjunktion der Klassen kann zu den Merkmalen hin verschoben werden, indem  $x_1 \vee x_2$  als neues (verallgemeinertes) Merkmal definiert wird.

und  $y_{1,2}$  sowie von  $y_2$  in  $y_{2,1}$  und  $y_{2,2}$  vorgenommen, s. Bild A1-2.II bis V, jeweils ganz unten.<sup>I</sup>

## 6.2 Bewertung, Modifikation und Selektion der Merkmale

Um festzustellen, welche der 1080 definierten Merkmale (s. Abschnitt 5.4) sich am besten zur Unterscheidung der 5 modifizierten Klassen (s. Abschnitt 6.1) eignen, und ob ggf. neue Merkmale einzuführen sind, werden die Merkmale einzeln und paarweise (logisch konjunktiv) und dabei klassenbezogen (getrennt für jede Klasse) bewertet. Eine bestimmte Anzahl der am höchsten bewerteten Merkmale wird dann selektiert.

Die Stichprobe der Referenzmuster (s. Abschnitt 5.5) repräsentiert auch Mengen aus mehr als zwei Merkmalen, die mehrelementigen Mengen jedoch entsprechend weniger statistisch signifikant; überdies ist ihre Bewertung rechenaufwendiger<sup>II</sup>.

Gründe dafür, Merkmalsmengen überhaupt zu berücksichtigen, sind: Die vorliegenden Signale stammen aus

- einem dynamischen Prozeß. Deshalb ist anzunehmen, daß zu verschiedenen Zeitpunkten gehörende Signalwerte (und damit auch ihre Merkmale) voneinander abhängen (Markov-Eigenschaft).
- demselben Prozeß. Deshalb ist anzunehmen, daß Merkmale verschiedener Signale voneinander abhängen.

Nur bei Unabhängigkeit der Merkmale reicht die Bewertung von einzelnen Merkmalen aus.

Ein Bewertungsziel ist die Maximierung der Transinformation gemäß (4.1-3) mit folgender Besonderheit: Weil die Klassen bereits festgelegt sind (Abschnitt 6.1), ist auch die Klassenentropie festgelegt und folglich kann die Transinformation nicht größer als diese Klassenentropie sein. Das angepaßte Bewertungsziel ist deshalb die Annäherung der Transinformation an die Klassenentropie. Die Berechnung der Transinformation erfolgt nach (3.2-9a); dort wird die merkmalsunabhängige Klassenentropie  $H(y_k)$  verwendet, so daß bei Merkmalsänderungen (die i. allg. öfter als

---

<sup>I</sup> Die alternativ mögliche Verallgemeinerung von Merkmalen beträfe viele Merkmale und wäre deshalb vergleichsweise aufwendig. Im Hinblick auf die praktische Tatsache, daß Merkmale im allgemeinen viel häufiger als Klassen geändert werden, wäre die Merkmalsverallgemeinerung außerdem relativ unflexibel.

<sup>II</sup> Um aus  $N$  Merkmalen  $a$ -elementige Merkmalsmengen zu bilden ( $1 \leq a \leq N$ ), so sind dafür  $\binom{N}{a} = N! / a! (N-a)!$  (Binomialkoeffizient) mögliche Kombinationen zu berücksichtigen. Jede Erhöhung  $a \rightarrow a+1$  vergrößert die Kombinationenanzahl mit dem Faktor  $(N-a+1)/a$ . Für  $a \ll N$  bedeutet das drastische Zunahmen von  $\binom{N}{a}$ . Im vorliegenden Fall, also bei  $N=1080$  Merkmalen (s. Abschnitt 5.4) gilt für zunehmende Mengengrößen  $\binom{N}{1}=1080$ ,  $\binom{N}{2}=\binom{N}{1} \cdot 1079/2=582660$ ,  $\binom{N}{3}=\binom{N}{2} \cdot 1078/3=209369160$  usw. Mit der verfügbaren Rechenkapazität ist höchstens  $a=2$  sinnvoll.

Für Merkmalspaarungen zwischen je zwei Signalen  $S_i, S_j$ ,  $i, j \in \{1, 2, \dots, 12\}$  gibt es wegen 90 Merkmalen je Signal für  $i \neq j$  nur  $90^2=8100$  Merkmalspaare pro Signalpaar. Bei Paarungen innerhalb eines Signals ( $i=j$ ) gibt es wegen der Besonderheit, daß ein Signalwert pro Auflösungsstufe stets genau eines von drei Intervallen besetzt (s. Abschnitt 5.4), nur  $9/2 \cdot 29 \cdot 30 = 3915$  unabhängige Merkmalspaare.

Klassenänderungen vorkommen) nur die Irrelevanz neu gebildet werden muß.

Weiterhin wird gefordert, daß aus der Merkmalsselektion für jede Klasse eine bestimmte, u.a. durch die vorhandene Rechenkapazität eingeschränkte Anzahl von Merkmalen (um 20) hervorgeht. Mit der Forderung einer bestimmten Merkmalsanzahl entfällt eine Forderung nach einem Mindestwert der Transinformation. Aus der Selektion können darum auch Merkmale mit vergleichsweise geringer Transinformation hervorgehen.

### 6.2.1 Einzelmerkmale

Die Transinformation wird nach (3.2-9a) zwischen jedem einzelnen Merkmal  $x_n$  und jeder einzelnen Klasse  $y_k$  bestimmt:

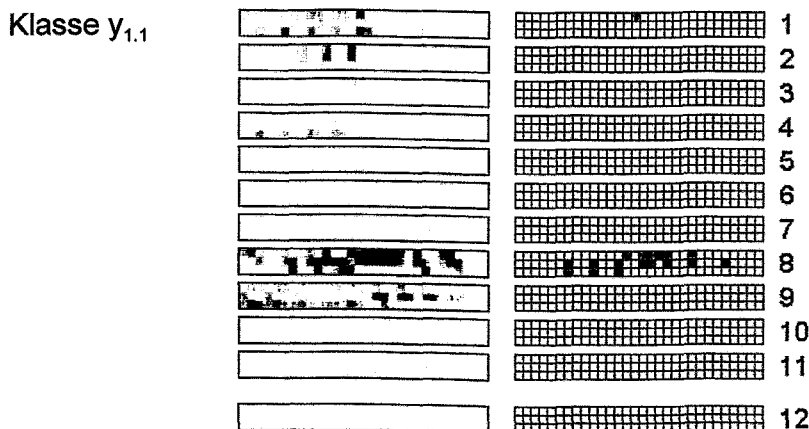
$$H(y_k; x_n) = H(y_k) - H(y_k | x_n)$$

mit  $H(y_k) = - p(y_k=1) \log_2 p(y_k=1) - p(y_k=0) \log_2 p(y_k=0)$

$$H(y_k | x_n) = - p(y_k=1, x_n=1) \log_2 p(y_k=1 | x_n=1) - p(y_k=1, x_n=0) \log_2 p(y_k=1 | x_n=0) \\ - p(y_k=0, x_n=1) \log_2 p(y_k=0 | x_n=1) - p(y_k=0, x_n=0) \log_2 p(y_k=0 | x_n=0)$$

Im eingeführten Merkmalschema (Bild 5.4-1 unten), jeweils für alle 12 Signale und getrennt nach den Klassen zeigt Bild 6.2-1 die Transinformation aller Merkmale als Grauwerte <sup>1</sup>; jeweils rechts sind die klassenbezogen selektierten Merkmale markiert. Die Transinformationswerte der selektierten Merkmale liegen vorwiegend etwa beim 0,3...0,6-fachen der jeweiligen Klassenentropie. Alle selektierten Einzelmerkmale gemeinsam zeigt Bild 6.2-2.

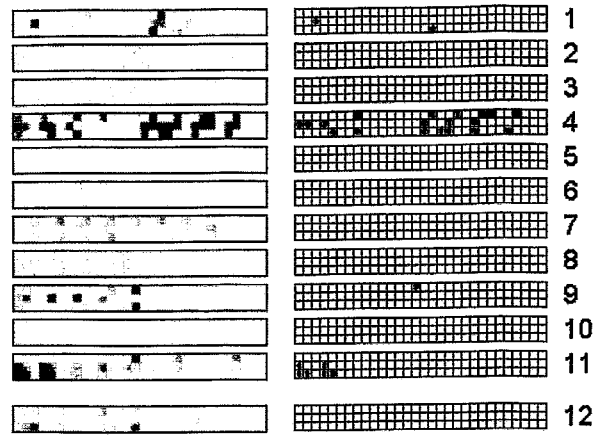
**Bild 6.2-1** Transinformation als Grauwert für alle Merkmale (links) sowie die selektierten Merkmale (rechts), jeweils für alle 12 Signale und getrennt nach den Klassen  $y_{1,1}$ ,  $y_{1,2}$ ,  $y_{2,1}$ ,  $y_{2,2}$ ,  $y_3$ .



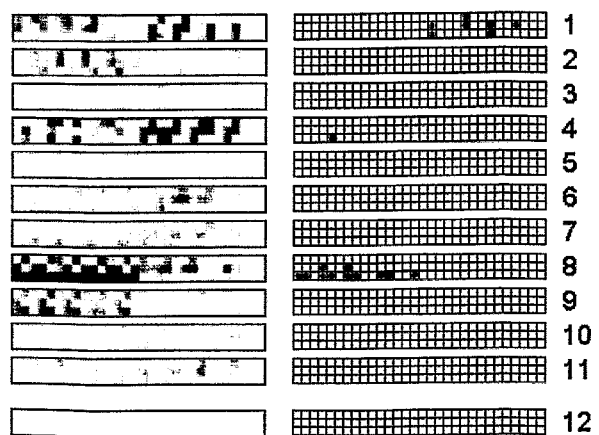
<sup>1</sup>) Weiß bedeutet Transinformation null. Schwarz bedeutet den klassenbezogenen Maximalwert der Transinformation, nicht jedoch den prinzipiell möglichen Maximalwert, die jeweilige Klassenentropie. Deshalb sind die Grauwerte nur klassenintern (in jedem Teilbild) miteinander vergleichbar.



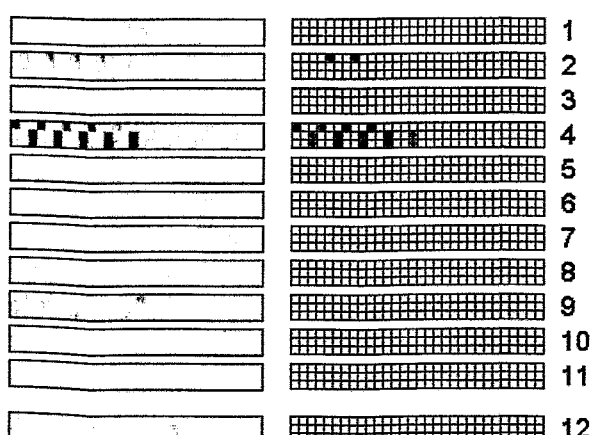
Klasse  $y_{1,2}$



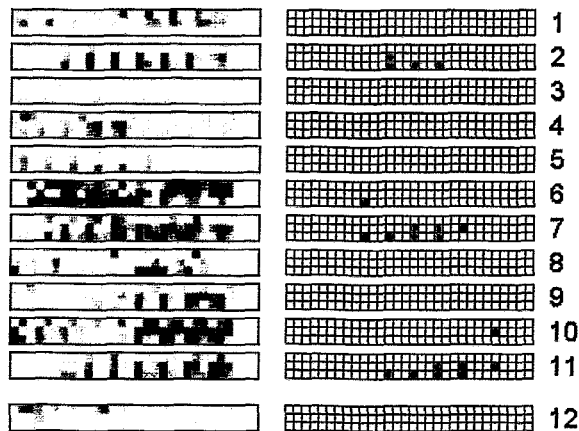
Klasse  $y_{2,1}$



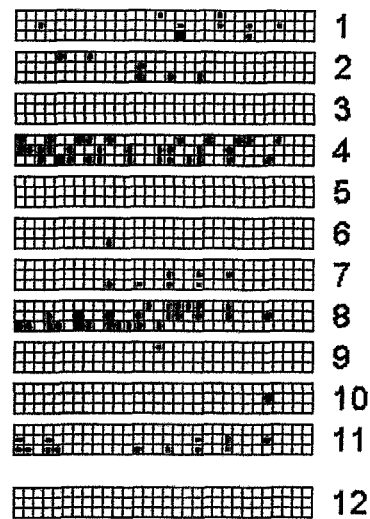
Klasse  $y_{2,2}$



Klasse  $y_3$



**Bild 6.2-2** Überlagerung aller selektierten Einzelmerkmale aus Bild 6.2-1 rechts. Bei Doppelzugehörigkeiten sind die Merkmalsfelder ganz ausgefüllt. Mehrfache Zugehörigkeiten gibt es nicht.



#### Diskussion der Bewertungsergebnisse:

1. Die Mengen der klassenbezogen selektierten Merkmale sind untereinander nahezu disjunkt. Nur vier selektierte Merkmale gehören zu zwei Klassen, s. Bild 6.2-2. Deshalb sind bei der Klassifikation gute Klassentrennungen zu erwarten.<sup>z</sup>
2. Von den Signalen  $S_3$ ,  $S_5$ , und  $S_{12}$  sind keine Merkmale selektiert worden und von  $S_9$  nur ein einziges. Dies weist zwar auf unzureichenden Klassenbezug der betreffenden Einzelmerkmale hin, rechtfertigt aber noch keinen Verzicht auf die vier Signale bzw. ihre Merkmale. Über deren Brauchbarkeit kann erst nach der Überprüfung der Merkmalsmengen entschieden werden.

<sup>z</sup> Hätte die nach einzelnen Klassen getrennte Merkmalsbewertung zu viele Merkmale mit zwei- und mehrdeutigem Klassenbezug ergeben, sollte besser eine über alle Klassen mittelnde Merkmalsbewertung vorgenommen werden, s. Abschnitt 3. Nicht sinnvoll wäre es hingegen, auch bei disjunkten Merkmalen zu mitteln, denn gegenüber der Einzelbewertung bewirkt die Mittelung dann eine unnötig vergrößerte Merkmalsbewertung und folglich eine schlechtere Klassentrennung.

3. Bei den neun Merkmalen  $x_{82}$  bis  $x_{90}$  ist die Transinformation stets gleich null. Der Grund dafür ist, daß das relative Fenstersignal (s. Abschnitt 5.3 oder Bild 5.4-1) zum aktuellen Zeitpunkt auf den Wert 0,5 fixiert ist. Deshalb sind von den genannten neun Merkmalen die drei Merkmale  $x_{83}$ ,  $x_{86}$ ,  $x_{89}$  immer besetzt und die anderen niemals. Also sind für  $x_{82}$  bis  $x_{90}$  die bedingten Wahrscheinlichkeiten  $p(y_k | x_{n_i})$  entweder gleich 1 oder 0, die Verbundwahrscheinlichkeiten  $p(y_k, x_{n_i}) = p(y_k)$ , folglich nach (3.2-7) und (3.2-4) die Irrelevanzen  $H(y_k | x_n) = H(y_k)$  und somit nach (3.2-9a) die Transinformationen  $H(y_k; x_n) = 0$ . Da ihre Werte stets gleich sind, ist auch anschaulich klar, daß diese Merkmale keinen Beitrag zur Klassentrennung liefern können. Auf diese Merkmale wird deshalb ohne zusätzliche Prüfungen verzichtet.

## 6.2.2 Merkmalspaare

Die Transinformation wird nach (3.2-9a) für alle möglichen Merkmalspaare  $\{x_a, x_b\}$  bezüglich jeder einzelnen Klasse  $y_k$  bestimmt:

$$H(y_k; x_a, x_b) = H(y_k) - H(y_k | x_a, x_b)$$

$$\text{mit } H(y_k) = - p(y_k=1) \log_2 p(y_k=1) - p(y_k=0) \log_2 p(y_k=0)$$

$$H(y_k | x_a, x_b) = \begin{aligned} & - p(y_k=1, x_a, x_b=1) \log_2 p(y_k=1 | x_a, x_b=1) \\ & - p(y_k=1, x_a, x_b=0) \log_2 p(y_k=1 | x_a, x_b=0) \\ & - p(y_k=0, x_a, x_b=1) \log_2 p(y_k=0 | x_a, x_b=1) \\ & - p(y_k=0, x_a, x_b=0) \log_2 p(y_k=0 | x_a, x_b=0) \end{aligned}$$

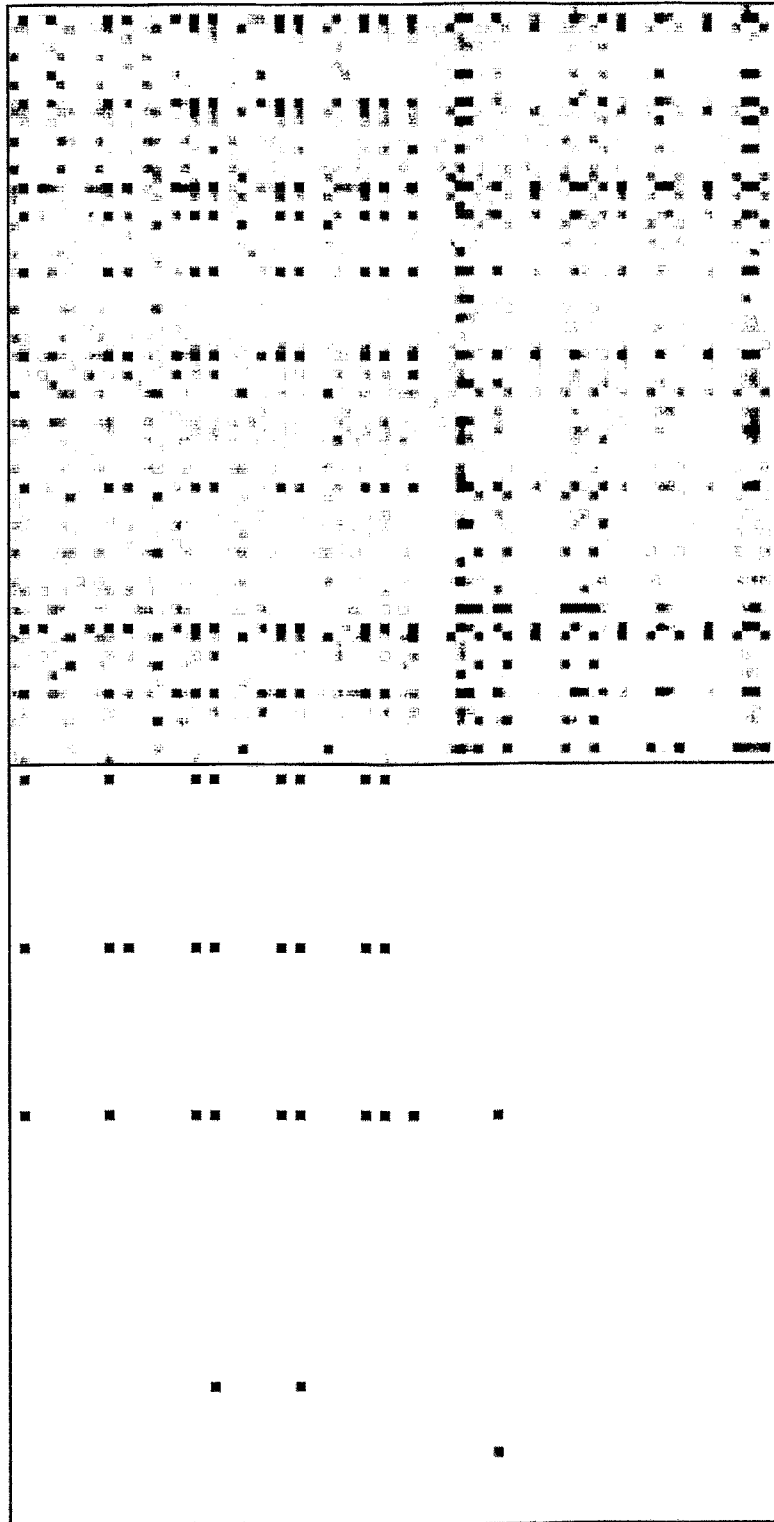
$$x_a, x_b=1 \quad \text{für } x_a=1, x_b=1,$$

$$x_a, x_b=0 \quad \text{für } x_a=0, x_b=1 \text{ oder } x_a=1, x_b=0 \text{ oder } x_a=0, x_b=0$$

Signifikante Transinformationswerte gibt es nur bei Merkmalspaarungen zwischen den Signalen  $S_9$  und  $S_{12}$  und nur bezüglich Klasse  $y_{2,2}$ ; allein diese Signalbeziehung wird weiter betrachtet. Die Merkmalsfolgen  $x_1, x_2, \dots, x_{81}$  aus  $S_9$  und  $S_{12}$  spannen im Bild 6.2-3 eine Grauwert-Matrix auf, oben im Bild für alle Paare und unten nur für die selektierten Paare. Die Transinformation der meisten selektierten Merkmalspaare hat nahezu den größtmöglichen Wert, den der Klassenentropie  $H(y_{2,2})$ .

Für die an den Paaren beteiligten Eigenschaften werden eigenständige (Paar-)Merkmale eingeführt und im Gegenzug wird - wie schon im Abschnitt 6.2.1 angedeutet - auf die an den Paaren beteiligten Einzelmerkmale verzichtet. Trotzdem (zum besseren Vergleich mit den Darstellungen im Abschnitt 6.2.1) zeigt Bild 6.2-4 die an den selektierten Paar-Merkmalen beteiligten Merkmale im Darstellungsschema der Einzelmerkmale (vgl. im Bild 6.2-1 jeweils rechts).

<sup>1</sup> Auf die ursprünglich eingeführten Merkmale  $x_{82}$  bis  $x_{90}$  wurde im Abschnitt 6.2.1 verzichtet.



**Bild 6.2-3** oben: Die Transformation als Grauwert für alle Merkmalspaarungen aus  $\{x_1, x_2, \dots, x_{81}\}$  zwischen den Signalen  $S_9$  ( $\rightarrow$ ) und  $S_{12}$  ( $\downarrow$ ) bezüglich der Klasse  $y_{22}$ . Der Grauwert in einem Matrixelement  $(a,b)$  gibt die Transformation des Merkmalspaars  $\{x_a, x_b\}$  bzw. der logischen Konjunktion  $x_a \wedge x_b$  an. Da unterschiedliche Signale miteinander verglichen werden, ist die Matrix asymmetrisch.  
unten: Selektierte Merkmalspaare (unter Beibehaltung der Grauwerte von oben)

	1
	2
	3
	4
	5
	6
	7
	8
	9
	10
	11
	12

**Bild 6.2-4** Im Paar selektierte Merkmale aus Bild 6.2-3 im Darstellungsschema der Einzelmerkmale.  
Da die selektierten Matrixelemente im Bild 6.2-3 unten über 5 Zeilen und 11 Spalten verteilt sind, gehören zum Signal  $S_{12}$  5 und zum Signal  $S_9$  11 selektierte Merkmale.

### Diskussion der Bewertungsergebnisse:

1. Weil die Transformation der meisten selektierten Merkmalspaare nahezu den größtmöglichen Wert hat, lassen diese Paare fast mit Gewißheit auf die zugehörige Klasse  $y_{2,2}$  schließen. Dies ist insofern bemerkenswert, als die an den Paaren beteiligten Einzelmerkmale ohne nennenswerten Klassenbezug sind - bei keiner Klasse, s. Abschnitt 6.2.1. Dieser Umstand bestätigt sowohl die Sinnfälligkeit der Einführung eigenständiger Merkmale für die Paare als auch den Verzicht auf die an den Paaren beteiligten Einzelmerkmale.
2. Weil das Signal  $S_{12}$  ein exponentielles Verhalten im Signal  $S_{11}$  beschreibt (s. Abschnitt 5.2), deutet der nachgewiesene Zusammenhang zwischen  $S_9$  und  $S_{12}$  auf eine ursächliche Beziehung zwischen  $S_9$  (Temperatur des externen Kühlmittels, s. Anhang 1) und  $S_{11}$  (Temperatur im Reaktor) hin. Dies ist physikalisch plausibel, denn das externe Kühlmittel dient zur Wärmeabführung aus dem Reaktor wie auch zur Steuerung oder Regelung der Reaktortemperatur. Die Existenz einer Beziehung zwischen  $S_9$  und  $S_{11}$  ist Anlaß, zwischen beiden Signalen ein dynamisches Modell zu strukturieren und dies so weit wie möglich zu identifizieren, s. Anhang 3. Wie schon angedeutet (Abschnitt 4.1), kann die Einführung eines dynamischen Prozeßmodells zur Bildung besserer Merkmale ausgenutzt werden. Bessere Merkmale wiederum erlauben eventuell eine Verbesserung des Modells usw.

## 7. Der Klassifikator

Ist die Abbildung der Merkmale auf die Klassen nicht eindeutig, hat ein Klassifikator die Aufgabe, diese Abbildung wenigstens näherungsweise zu realisieren, s. Abschnitt 4. Dabei wird der kleinstmögliche Abbildungsfehler nur dann erreicht, wenn die in den Referenzmustern enthaltene, den Zusammenhang zwischen Merkmalen und Klassen ausdrückende A-priori-Information vollständig ausgenutzt wird. Dies wiederum ist praktisch meist nur näherungsweise erreichbar. An ein Klassifizierungsverfahren besteht deshalb die Forderung,

- die in den Referenzmustern enthaltene, die Merkmale-Klassen-Beziehungen betreffende A-priori-Information bestmöglich auszunutzen.

Hinzu kommen folgende sicherheitsmotivierte Forderungen:

- Hohe Transparenz und Nachvollziehbarkeit einer Erkennung
- Bei fehlerhafter, ungenauer oder falscher Erkennung möglichst genaue Lokalisierung der Fehlerursache
- Die Beseitigung von Fehlerursachen darf weder
  - \* den Verlust klassenrelevanter A-priori-Information noch
  - \* die schlimmstenfalls unbemerkte Entstehung neuer Fehlerursachen mit sich bringen.

Diesen Forderungen wird am ehesten ein unkompliziertes und leicht durchschaubares Klassifizierungsverfahren gerecht, z.B. auf der Grundlage eines Assoziativspeichers, der das unbekannte Muster unmittelbar mit den Referenzmustern vergleicht. Zu dieser Art gehört das tensororientierte Verfahren (eine Arbeitsbezeichnung)<sup>I</sup> [Ho94, Ho96]. Es beruht auf einer nichtlinearen Metrik, die im Sinne eines tensoriell beschriebenen Gütekriteriums optimal ist (s. Anhang 2). Dieses Verfahren arbeitet folgendermaßen:

### Algorithmus für die Klassifikation:

1. Bildung aller Mengendurchschnitte zwischen der Merkmalsmenge des unbekanntes Musters und der Merkmalsmenge jedes einzelnen Referenzmusters:  $\{u\} \cap \{x_m\} \forall m$  (bitweise logische Konjunktion der Vektorelemente). Die Schnittmengen bestehen aus  $g_m \leq N$  Elementen.

**Beispiel 7.1:** Mit  $\{x_m\} = \{x_{m2}, x_{m4}, x_{m5}, x_{m7}, y_{m1}\}$  bzw.  $x_m = (010110110)^T$  aus Beispiel 1.1 und  $\{u\} = \{u_2, u_5, u_6\}$  bzw.  $u = (010011000)^T$  ergibt sich die  $g_m = 2$ -elementige Schnittmenge  $\{u\} \cap \{x_m\} = \{x_{m2}, x_{m5}\}$  bzw.  $(010010000)^T$ .

2. Jeder Klasse  $y_k$ , die im Referenzvektor  $x_m$  mit  $y_{mk} = 1$  codiert ist, wird anhand der Elementezahl  $g_m$  das (Klassen-)Gewicht

---

<sup>I</sup> "tensororientiert", weil Tensoren lediglich zur Begründung der Metrik herangezogen werden. Im Klassifizierungsverfahren selbst treten Tensoren nicht auf.

$$G_m = G_m(y_{mk}=1) = 2^{g_m-1} \quad (7.1)$$

zugeordnet,  $G_m \in \{0, 1, 2, \dots, G_{m, \max} = 2^N - 1\}$ . Dagegen ist  $G_m(y_{mk}=0) = 0$ .

**Beispiel 7.2:** Im Referenzvektor  $\underline{x}_m$  aus Beispiel 7.1 ist der Klassencode  $y_{m1}=1, y_{m2}=0$ . Folglich ergibt sich nach (7.1) mit  $g_m=2$  das Klassengewicht  $G_m(y_{m1})=2^2-1=3$ . Für die andere Klasse gilt dagegen  $G_m(y_{m2})=2^0-1=0$ .

- Die nach (7.1) bestimmten Klassengewichte  $G_m(y_{mk})$  werden für jede Klasse jeweils über alle Referenzmuster summiert:  $\sum_{m=1}^M G_m(y_{mk}) \forall k$ . Die entstehende K-elementige Verteilung der resultierenden Klassengewichte bildet - eventuell geeignet normiert - das Klassifizierungsergebnis.  
Bei disjunkten Klassen ist es möglich und zweckmäßig, die erhaltene Gewichtsverteilung mit der Summe aller erhaltenen Gewichte zu normieren. Die normierten Klassengewichte sind dann als Wahrscheinlichkeiten für die Klassenzugehörigkeit des unbekanntes Musters interpretierbar.

Infolge der exponentiellen Klassengewichtung nach (7.1) dominiert die Klassenzugehörigkeit desjenigen Referenzvektors  $\underline{x}_m$ , der die meisten Merkmalsübereinstimmungen mit dem unbekanntes Vektor aufweist, die Gewichtsverteilung im Klassifizierungsergebnis ggf. bei weitem. Bestimmend dafür sind - und darin besteht die wichtigste Besonderheit des Klassifizierungsverfahrens - die Differenzen der  $g_m$ -Werte, nicht deren Absolutwerte.

**Beispiel 7.3:** Seien  $g_a$  und  $g_b$  die Elementezahlen zweier Schnittmengen gemäß Punkt 1 des obigen Algorithmus. Für  $g_a > g_b$  (Im Klassifizierungsergebnis dominiert die zum a-ten Referenzmuster gehörende Klasse gegenüber der zum b-ten Referenzmuster gehörenden Klasse) ist das Verhältnis der Klassengewichte  $G_a/G_b = (2^{g_a} - 1)/(2^{g_b} - 1) \approx 2^{g_a}/2^{g_b} = 2^{g_a - g_b} > 1$ . Für  $g_a=102$  und  $g_b=100$  ist  $G_a/G_b = 2^{102-100} = 2^2 = 4$ , weswegen die Klassenzugehörigkeit des Referenzvektors  $\underline{x}_a$  etwa 4-fach stärker als die Klassenzugehörigkeit des Referenzvektors  $\underline{x}_b$  in das Klassifizierungsergebnis eingeht, obwohl beide Referenzvektoren in fast gleich vielen Elementen ( $102 \approx 100$ ) mit dem unbekanntes Vektor übereinstimmen.

Das exponentielle Ähnlichkeitsmaß erleichtert es außerdem, die Klassenzugehörigkeit eines unbekanntes Musters auch dann sicher nachzuweisen, wenn die richtige Klasse nur durch vergleichsweise wenige Referenzmuster repräsentiert ist. Dies entschärft das häufige praktische Problem, daß über Gefahren- und Ausnahmesituationen viel weniger Datenmaterial als über die "Normalität" verfügbar ist.

**Beispiel 7.4:** Ein unbekanntes Muster stimme in  $A=10$  Referenzmustern mit  $g_a=10$  Merkmalen überein und in  $B=1$  Referenzmuster mit  $g_b=20$  Merkmalen. Dann dominiert die Klassenzugehörigkeit des letztgenannten Referenzmusters etwa  $B \cdot 2^{g_b} / A \cdot 2^{g_a} = 2^{20} / (10 \cdot 2^{10}) = 1048576 / 10240 \approx 102,4$ -fach im Klassifizierungsergebnis; gegenüber der Anzahl der übereinstimmenden Merkmale ist die Anzahl der Referenzmuster von untergeordneter Bedeutung.

Die starke Nichtlinearität der exponentiellen Klassengewichtung rechtfertigt folgende Vereinfachung: Schnittmengen  $\{u\} \cap \{x_m\}$  (s. Punkt 1 des obigen Algorithmus), die nur Teilmengen anderer Schnittmengen sind, tragen vergleichsweise wenig zum Klassifizierungsergebnis bei und werden deshalb vernachlässigt. In Betracht bleiben dann nur die miteinander unvereinbaren (Ober-)Mengen.

Das folgende Beispiel soll die Wirkungsweise des Klassifizierungsverfahrens insgesamt veranschaulichen.

**Beispiel 7.5:** Gegeben sind (1) Referenzvektoren  $x_m$ ,  $m \in \{1, 2, \dots, M=4\}$  mit Merkmalen  $x_{mn}$ ,  $n \in \{1, 2, \dots, N=6\}$  und disjunkt besetzten Klassencode-Elementen  $y_{mk}$ ,  $k \in \{1, \dots, K=M\}$  sowie (2) zwei unbekannte Vektoren  $u$  mit je  $N=6$  Merkmalen  $u_n$  und unbesetztem (da unbekanntem) Klassencode, siehe nachfolgendes Schema (die Zeilen entsprechen den Vektoren).

(1)	$x_{m1}$	$x_{m2}$	$x_{m3}$	$x_{m4}$	$x_{m5}$	$x_{m6}$	$y_{m1}$	$y_{m2}$	$y_{m3}$	$y_{m4}$
m=1	0	1	1	0	1	1	1	0	0	0
2	0	1	0	0	1	0	0	1	0	0
3	1	0	0	1	1	0	0	0	1	0
4	0	1	0	1	1	1	0	0	0	1

(2)	$u_1$	$u_2$	$u_3$	$u_4$	$u_5$	$u_6$				
a)	1	0	0	0	1	1	0	0	0	0
b)	1	0	1	0	1	1	0	0	0	0

Nach dem obigen Algorithmus ergeben sich für die Fälle a) und b) folgende Schnittmengen  $\{u\} \cap \{x_m\}$ , Elementanzahlen  $g_m$ , Klassengewichte  $G_m$  und normierte Klassifizierungsergebnisse  $y_m$  (fett hervorgehoben), wobei außer den exakten Lösungen auch die (unter alleiniger Verwendung der Obermengen) vereinfachten Lösungen angegeben sind:

a)

	$\{x_m\}$	$\{u\}$	$\{u\} \cap \{x_m\}$	$g_m$	$G_m$	$y_m$ (normiert)
m=1	$\{x_2, x_3, x_5, x_6, y_1\}$	$\{u_1, u_5, u_6\}$	$\{x_5, x_6\}$	2	$2^2-1=3$	<b>0,3</b>
2	$\{x_2, x_5, y_2\}$	"	$\{x_5\}$	1	$2^1-1=1$	<b>0,1</b>
3	$\{x_1, x_4, x_5, y_3\}$	"	$\{x_1, x_5\}$	2	$2^2-1=3$	<b>0,3</b>
4	$\{x_2, x_4, x_5, x_6, y_4\}$	"	$\{x_5, x_6\}$	2	$2^2-1=3$	<b>0,3</b>

Vereinfachung:  $\{x_5\} \subset \{x_5, x_6\}$

m=1	$\{x_2, x_3, x_5, x_6, y_1\}$	$\{u_1, u_5, u_6\}$	$\{x_5, x_6\}$	2	$2^2-1=3$	<b>0,333</b>
3	$\{x_1, x_4, x_5, y_3\}$	"	$\{x_1, x_5\}$	2	$2^2-1=3$	<b>0,333</b>
4	$\{x_2, x_4, x_5, x_6, y_4\}$	"	$\{x_5, x_6\}$	2	$2^2-1=3$	<b>0,333</b>

b)

	$\{x_m\}$	$\{u\}$	$\{u\} \cap \{x_m\}$	$g_m$	$G_m$	$y_m$ (normiert)
m=1	$\{x_2, x_3, x_5, x_6, y_1\}$	$\{u_1, u_3, u_5, u_6\}$	$\{x_3, x_5, x_6\}$	3	$2^3-1=7$	<b>0,5</b>
2	$\{x_2, x_5, y_2\}$	"	$\{x_5\}$	1	$2^1-1=1$	<b>0,071</b>
3	$\{x_1, x_4, x_5, y_3\}$	"	$\{x_1, x_5\}$	2	$2^2-1=3$	<b>0,214</b>
4	$\{x_2, x_4, x_5, x_6, y_4\}$	"	$\{x_5, x_6\}$	2	$2^2-1=3$	<b>0,214</b>

Vereinfachung:  $\{x_5\} \subset \{x_5, x_6\} \subset \{x_3, x_5, x_6\}$

m=1	$\{x_2, x_3, x_5, x_6, y_1\}$	$\{u_1, u_3, u_5, u_6\}$	$\{x_3, x_5, x_6\}$	3	$2^3-1=7$	<b>0,7</b>
3	$\{x_1, x_4, x_5, y_3\}$	"	$\{x_1, x_5\}$	2	$2^2-1=3$	<b>0,3</b>

Die Ergebnisse zeigen: (1) Zusätzliche Übereinstimmungen zwischen  $\{x_m\}$  und  $\{u\}$  um nur ein einziges Merkmal ( $g_m - g_m + 1$ ) bewirken durch die exponentielle Klassengewichtung (vgl. (7.1)) mehr als eine Verdopplung des jeweiligen Klassengewichtes  $G_m$ . (2) Die Vereinfachungen vergrößern die Ergebnisse zwar, ändern sie aber nicht qualitativ; die Dominanz der Klassen  $y_1, y_3, y_4$  bei a) und von Klasse  $y_1$  bei b) bleibt erhalten.



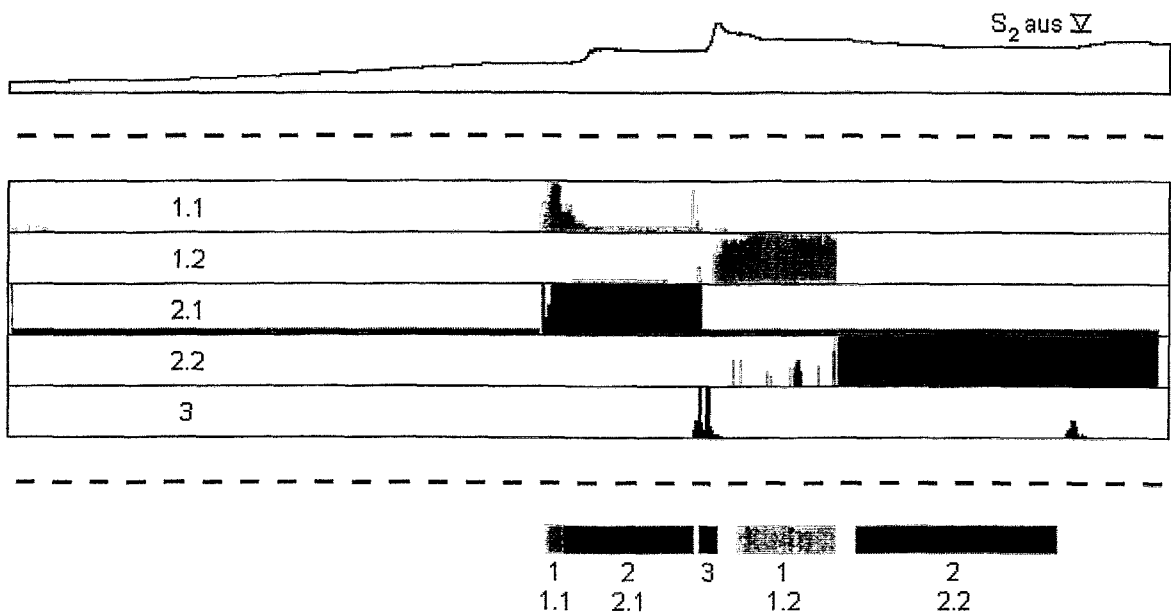
## 8. Klassifikation am chemischen Prozeß

Die Prozeßsignale  $S_1, S_2, \dots, S_{11}$  (s. Anhang 1, Bild A1-2) und das zusätzliche Signal  $S_{12}$  (s. Abschnitt 5.2) liegen als vorklassifizierte Realisierungen I,II,III,IV,V vor. Referenzmuster gibt es nur aus den Realisierungen I bis IV, s. Abschnitt 5.5. Die Realisierung V gilt als unbekannt und soll nach dem im Abschnitt 7 beschriebenen tensororientierten Verfahren klassifiziert werden; verwendet wird allerdings die vereinfachte Variante dieses Verfahrens. Die Vorklassifikation der Realisierung V dient lediglich zur Einschätzung des Klassifizierungsergebnisses.

Von den ursprünglich insgesamt 1080 definierten Merkmalen werden nur die 112 einzeln selektierten Merkmale und von den paarweise selektierten Merkmalen werden die 30 neu definierten Paar-Merkmale (s. Bilder 6.2-2 und 6.2-3) zur Klassifikation herangezogen.

Klassifiziert wird nach den fünf neuen (modifizierten) Klassen (s. Abschnitt 6.1) und nach jeder Klasse einzeln <sup>†</sup>. Weil die vorgenommene Klassenunterteilung eindeutig umkehrbar ist, kann aus dem Klassifizierungsergebnis auch auf das Vorliegen der drei ursprünglichen Klassen geschlossen werden.

Das unbekannte Muster wird fortlaufend klassifiziert. Darum entsteht das Klassifizierungsergebnis in Form von Funktionen über der Zeit, s. Bild 8.2-1.



**Bild 8.2-1** Klassifizierung der als unbekannt geltenden Signalrealisierung V bei getrennter Klassifikation nach den einzelnen Klassen.

Oben: Für die zeitliche Orientierung: Signal  $S_2$  aus Realisierung V

Mitte: Klassifizierungsergebnisse in den einzelnen Klassen. Die Zugehörigkeit zu den drei ursprünglichen Klassen kommt an den gleichen Grautönen zum Ausdruck.

Unten: Vorgegebene Klassenzuordnung

<sup>†</sup> Da die Klassen nahezu disjunkt sind, könnte fast gleichberechtigt auch nach allen Klassen gleichzeitig klassifiziert werden. Die separate Klassifikation wird dennoch bevorzugt, weil sie allgemeiner und deswegen auch bei nichtdisjunkten Klassen anwendbar ist, s. Abschnitt 3.

### Diskussion des Klassifizierungsergebnisses:

1. Alle Klassen sind in denjenigen Signalabschnitten, denen sie vorab zugeordnet wurden, auch per Klassifikation im wesentlichen richtig erkannt worden. Lediglich außerhalb dieser Abschnitte gibt es nennenswerte, nachfolgend genannte Besonderheiten.
2. Die Klasse  $y_3$  wurde mit geringem Gewicht auch am Ende des dargestellten Zeitbereiches erkannt. Dies darf als richtig gelten, denn auch der erfahrene Anlagenbediener weist besagtem Zeitabschnitt die Gefahrenklasse  $y_3$  zu; bei der Vorklassifikation blieb dies unberücksichtigt.
3. Die Klasse  $y_{2,2}$  wurde etwas zu früh aber vor allem erheblich zu lange erkannt. Als Ursache dieser Ungenauigkeit werden zweifelsfrei - und darin bestätigt sich die Transparenz des verwendeten Klassifizierungsverfahrens - bestimmte Merkmale nachgewiesen: 13 der insgesamt 16 für die Klasse  $y_{2,2}$  selektierten Einzelmerkmale (s. Bild 6.2-1) haben nur sehr geringe Transinformation, und auch die Transinformation der drei höchstbewerteten Merkmale erreicht längst nicht den theoretisch möglichen Maximalwert, die Klassenentropie  $H(y_{2,2})$ .

Auf eine quantitative Gesamteinschätzung des Klassifizierungsergebnisses in Form eines Güterwertes wird mangels eines dafür geeigneten Gütekriteriums verzichtet; ein solches Kriterium müßte auch den in dieser Arbeit besonders beachteten aber schwer quantisierbaren Sicherheitsaspekt berücksichtigen.

Zur Behebung der in 3. genannten Ungenauigkeit ist folgendes anzumerken: Wenn zur Erkennung der Klasse  $y_{2,2}$  nur die drei höchstbewerteten Einzelmerkmale herangezogen werden, dann wird  $y_{2,2}$  nicht mehr außerhalb der vorklassifizierten Signalabschnitte erkannt. Statt dessen ist dann aber innerhalb dieser Abschnitte die Erkennung nur noch lückenhaft; hinsichtlich Prozeßsicherheit bedeutet dies eine Verschlechterung. Zwar wurden zur Klasse  $y_{2,2}$  auch Merkmalspaare selektiert, die zum Teil sogar fast eindeutigen Klassenbezug besitzen (s. Abschnitt 6.2-2), doch kommen diese Paare in der unbekanntem Signalrealisierung zu selten vor, um für die Erkennung von  $y_{2,2}$  wesentlich zu sein.

Um diejenigen Merkmale, welche die ungenaue Erkennung verursachen, durch bessere Merkmale zu ersetzen, kann zunächst die A-priori-Information der Referenzmuster tiefgründiger erschlossen werden, z.B. indem (wenigstens einige ausgewählte) mehrelementige Merkmalsmengen oder kompliziertere logische Merkmalsverknüpfungen überprüft werden; untersucht wurde dies nicht. Die Bildung anderer oder einer größeren Anzahl Referenzmuster erbrachte in allen dazu durchgeführten Experimenten keinen nennenswerten Vorteil. Im weiteren können von vorn herein aussichtsreichere Merkmale definiert werden, indem z.B. andere Signaleigenschaften als diejenigen aus Abschnitt 5 berücksichtigt werden; die Wahlfreiheit für solche Eigenschaften ist faktisch nicht eingeschränkt. Auch ein analytisches Prozeßmodell oder Teile eines solchen können zur Merkmalbildung herangezogen werden.

## 9. Zusammenfassung und Ausblick

Um für die gegebene Klassifizierungsaufgabe sowohl relevante als auch transparente Merkmale-Klassen-Beziehungen zu finden, erweist sich folgendes Vorgehen als zweckmäßig:

1. Erfassung der Dynamik der gegebenen Prozeßsignale anhand eines den aktuellen Zeitpunkt begleitenden Zeitfensters.
2. Einführung des Exponenten eines im Zeitfenster exponentiell approximierten Signals als zusätzliches Signal; exponentielles Verhalten zeigte sich subjektiv und objektiv als besonders markante Signaleigenschaft.
3. Zwecks besserer Zugänglichkeit der Information aus sehr flach verlaufenden Signalabschnitten wird der Fensterinhalt gleitend und damit zeitvariant normiert und so eine wertmäßige Spreizung der flachen Signalabschnitte erreicht.
4. Zeit- und Wertdiskretisierung des gespreizten Fensterinhaltes, wobei die Wertdiskretisierung mit informationstheoretischer Begründung hierarchisch gestaffelt erfolgt. Jede Zeit-/Wert-Stützstelle definiert dann ein binäres Merkmal.
5. Bildung von Referenzmustern aus vorklassifizierten Signalabschnitten an ausgewählten, innerhalb der Signalabschnitte etwa gleichmäßig verteilten Positionen. Pro Klasse werden ähnlich viele Referenzmuster gebildet. Von den insgesamt fünf gegebenen, mit Vorklassifizierungen versehenen Signalrealisierungen dienen nur vier zur Gewinnung der Referenzmuster. Die fünfte wird als unbekannt aufgefaßt; ihre Vorklassifikation dient zur Beurteilung des gewonnenen Klassifizierungsergebnisses.
6. Modifikation von Klassen und Merkmalen, indem
  - Klassen merkmalsbezogen in logisch disjunktiv verknüpfte (Unter-)Klassen aufgeteilt und dadurch spezialisiert werden
  - und Merkmale paarweise logisch konjunktiv verknüpft und dadurch verallgemeinert werden.
7. Selektion der definierten Merkmale und Merkmalspaare nach dem Kriterium der maximalen Transinformation sowie danach, daß für jede Klasse eine bestimmte Anzahl von Merkmalen übrigbleibt.

Das Ergebnis der Klassifizierung der unbekannteren Signalrealisierung ist insgesamt und vor allem auch aus sicherheitsorientierter Sicht zufriedenstellend. Im einzelnen ist das Ergebnis wie folgt charakterisiert:

- Die höchste Gefahrenklasse wird sicher erkannt.
- Die Klassen geringerer Gefahr werden
  - \* in den vorklassifizierten Signalabschnitten sicher erkannt und
  - \* teilweise auch außerhalb der vorklassifizierten Signalabschnitte erkannt; diese Ungenauigkeit ist im Sinne der Prozeßsicherheit unbedenklich (besser eine Gefahr irrtümlich vermutet, als eine wirkliche Gefahr nicht erkannt).

Als Ursache der ungenauen Erkennung werden bestimmte Merkmale zweifelsfrei nachgewiesen. Dies bestätigt am praktischen Beispiel die Transparenz des verwendeten Klassifizierungsverfahrens. Zur Beseitigung der Ungenauigkeit brauchen die betreffenden Merkmale einfach nur weggelassen zu werden; die anderen Teile des Klassifizierungsergebnisses werden dadurch nicht beeinträchtigt. Allerdings sollten die weggelassenen durch bessere Merkmale ersetzt werden. Für diesen Zweck können beispielsweise

- kompliziertere (mehr als nur paarweise logisch konjunktive) Merkmalsverknüpfungen, die ihrerseits eigenständige Merkmale definieren, auf höhere Transinformation hin überprüft werden oder
- von vorn herein aussichtsreichere Merkmale definiert werden, vorzugsweise durch die Berücksichtigung anderer Signaleigenschaften, ggf. aber auch anhand eines analytischen Prozeßmodells.

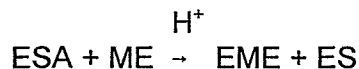
Über die vordergründige Zielstellung der Signalerkennung hinaus wird ausgehend vom Ergebnis einer paarweisen informationstheoretischen Merkmalsbewertung gezeigt, wie nachgewiesene Merkmalsabhängigkeiten im konkreten Fall zur Generierung von Strukturelementen eines dynamischen Prozeßmodells ausgenutzt werden können. Dabei wird deutlich, daß ein fortschreitender Ausbau der Modellstruktur (qualitative Seite der Modellbildung) nicht ohne begleitende Parameteridentifikation (quantitative Seite der Modellbildung) möglich ist. Die Bildung oder Vervollkommnung des Modells schafft Voraussetzungen für die Gewinnung effektiverer Merkmale und Klassen. Auf deren Grundlage wiederum ist u.U. eine abermalige Modellverbesserung erreichbar usw. Die sich damit abzeichnende Wechselbeziehung zwischen

- der Bereitstellung informationstheoretisch optimaler Merkmale und Klassen und
- der Generierung von Modellen dynamischer und vor allem regelungstechnischer Prozesse

ist ein Gegenstand weiterführender Untersuchungen.

## Anhang 1: Prozeßbeschreibung

Betrachtet wird der chemische Prozeß der Veresterung von Essigsäureanhydrid (ESA) mit Methanol (ME) - mit Schwefelsäure ( $H^+$ ) als Katalysator - zu Essigsäuremethylester (EME) und als Nebenprodukt Essigsäure (ES):



Zwar kann für diesen Prozeß und seine verfahrenstechnischen Randbedingungen (s. Bild A1-1) ein vergleichsweise einfaches analytisches Modell angegeben werden, doch wird darauf absichtlich verzichtet.

Für die Erkennung von Gefahrensituationen stehen ausschließlich folgende Prozeßsignale zur Verfügung:

- $S_1$ : Volumen des Reaktorinhaltes
- $S_2$ : Druck im Reaktor
- $S_3$ : Masse des ESA (Essigsäureanhydrid) im Reaktor
- $S_4$ : Masse des ME (Methanol) im Reaktor
- $S_5$ : Masse von  $H_2SO_4$  (Katalysator) im Reaktor
- $S_6$ :  $dS_{11}/dt$
- $S_7$ : bilanzierte Reaktionswärme (Gesamt-Wärmeabgabe des Reaktorinhaltes)
- $S_8$ : Rührerdrehzahl
- $S_9$ : Temperatur des externen Kühlmittels
- $S_{10}$ : Temperatur des Kühlmittels im Reaktormantel
- $S_{11}$ : Temperatur im Reaktor

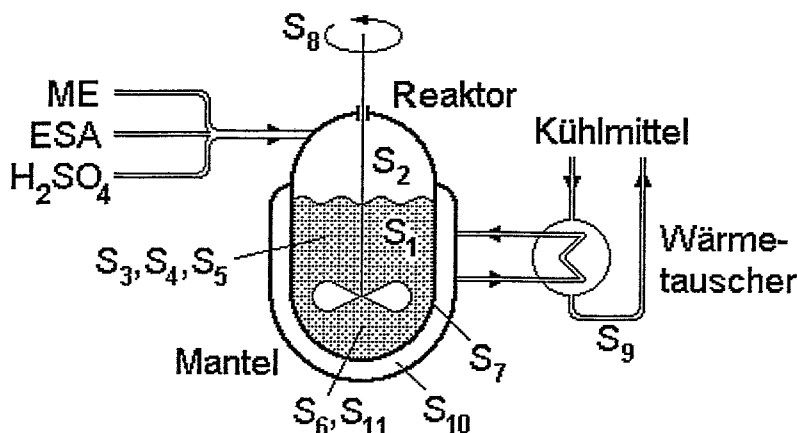


Bild A1-1 Technologisches Prozeßschema

Ausgewählte Verläufe der Prozeßsignale  $S_1, S_2, \dots, S_{11}$  (s. Bild A1-2) und des zusätzlich erzeugten Signals  $S_{12}$  (s. Abschnitt 5.2) sind als vorklassifizierte Realisierungen I, II, III, IV, V gegeben. Alle Signale sind auf den Wertebereich  $[0, 1]$  normiert. Die Signale sind zwar als Zeitreihen und wertdiskretisiert gegeben, allerdings mit vergleichsweise hochfrequenter Abtastrate (Abtastzeit 2s) sehr feinstufiger Wertdis-

cretisierung; sie werden deshalb als (zeit-)kontinuierlich und (wert-)analog angenommen. Der im Bild A1-2 dargestellte Zeitbereich umfaßt etwa 2500 Abtastwerte (5000s).

Die Realisierung I (Bild A1-2.I) zeigt einen normalen Prozeßverlauf. Die Realisierungen II bis V (Bild A1-2, II bis V) enthalten verschiedene außergewöhnliche (gestörte, bezüglich der Prozeßsicherheit bedenkliche) Signalabschnitte. Diesen Signalabschnitten werden deshalb folgende, den Grad der Prozeßgefährdung kennzeichnende Gefahrenklassen zugewiesen (im Bild A1-2 erkennbar an den unterschiedlichen Grauwerten sowie an den Klassen-Indizes jeweils unten im Bild, obere Zahlenreihe):

Klasse  $y_1$ : Vorwarnung

Klasse  $y_2$ : Warnung

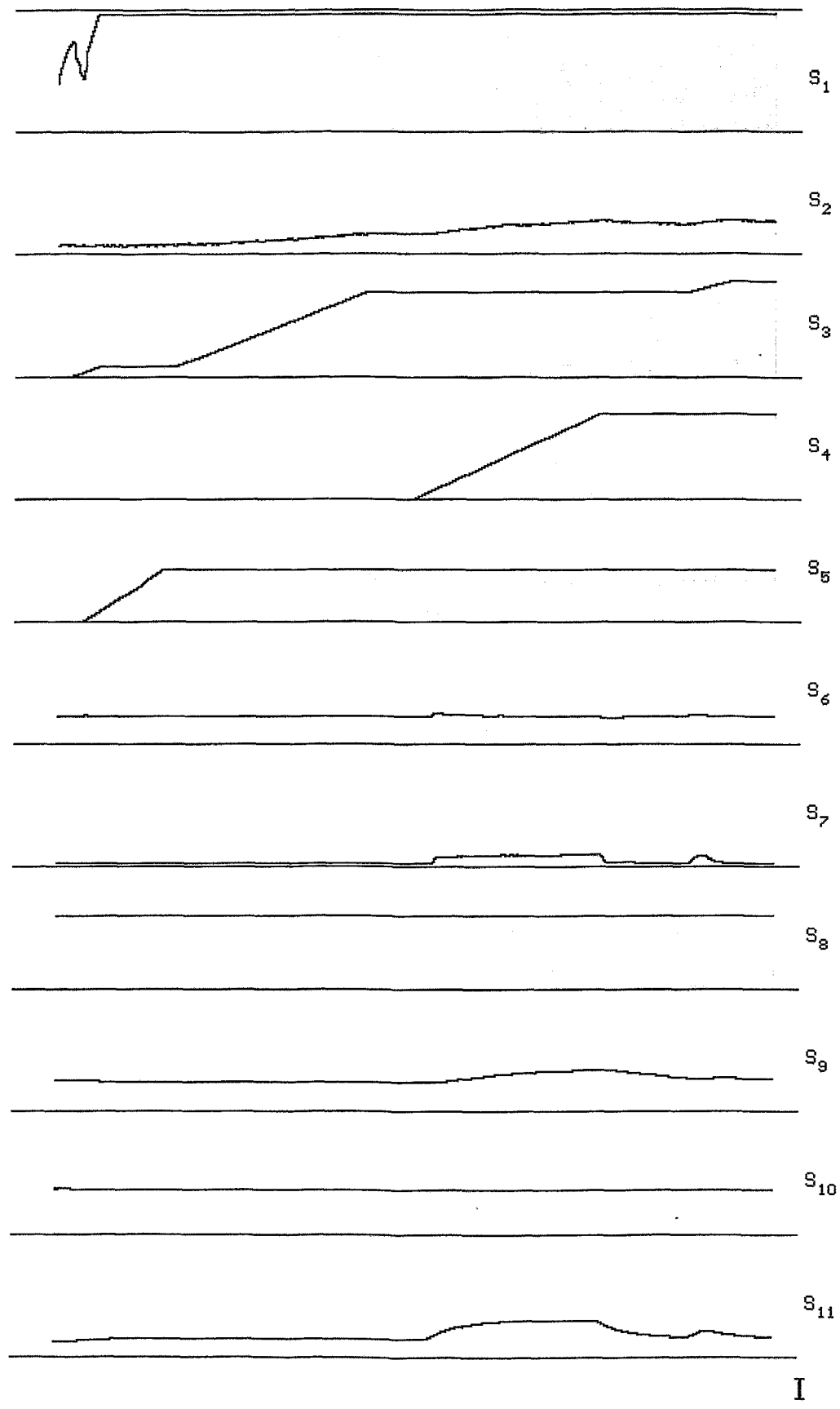
Klasse  $y_3$ : Gefahr

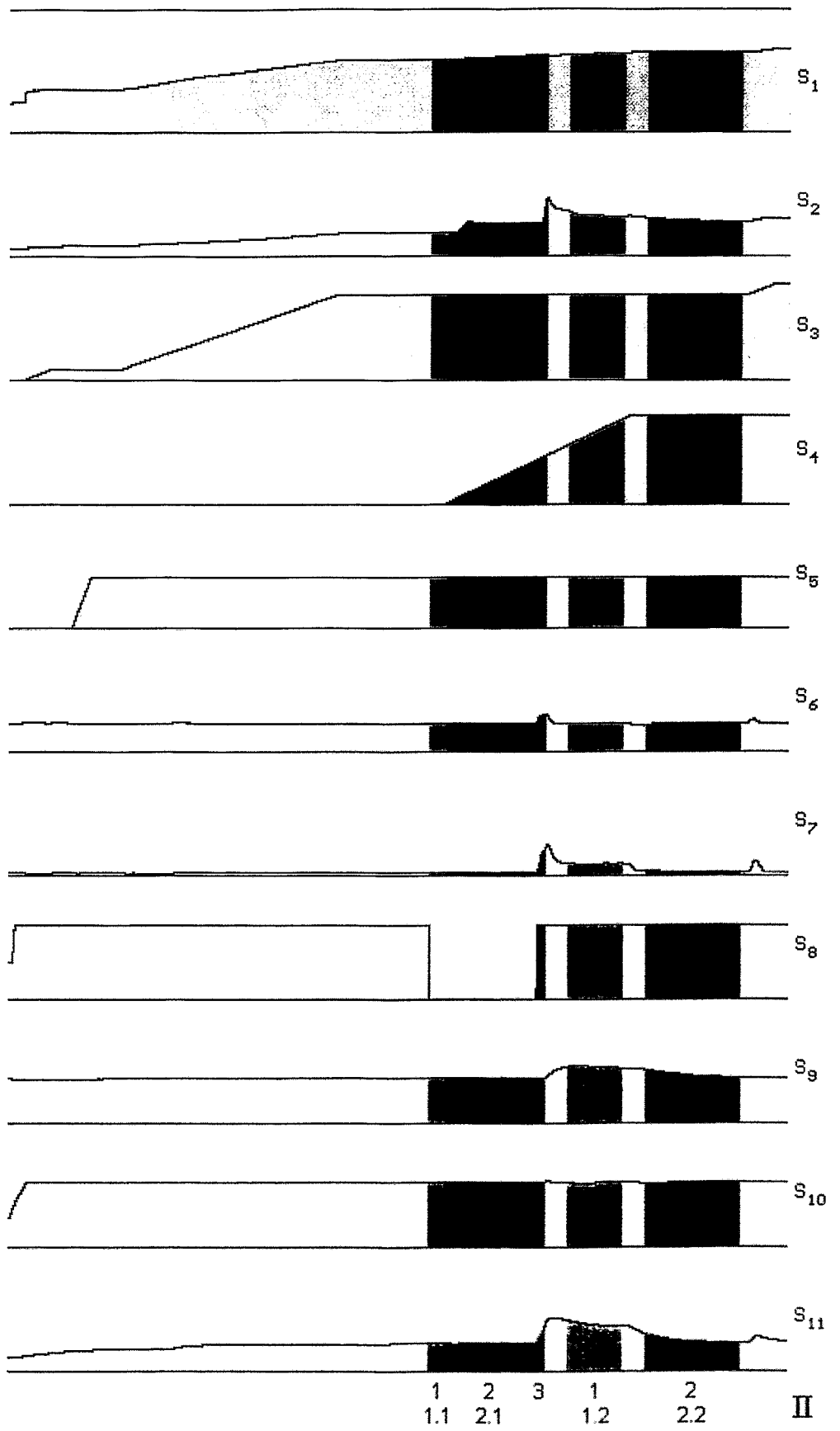
Diese Vorklassifikation wird im Abschnitt 6.1 noch unterteilt. Bei der Signalerkennung wird die Vorklassifikation der Realisierung V als unbekannt aufgefaßt; sie dient nur zur Einschätzung des Klassifizierungsergebnisses.

Die Normalität des Prozeßverlaufes in der Realisierung I wird zwar intern als selbständige Klasse behandelt und auch durch spezielle Referenzmuster repräsentiert, auf eine Klassifikation nach dieser Klasse wird jedoch aus folgenden Gründen verzichtet:

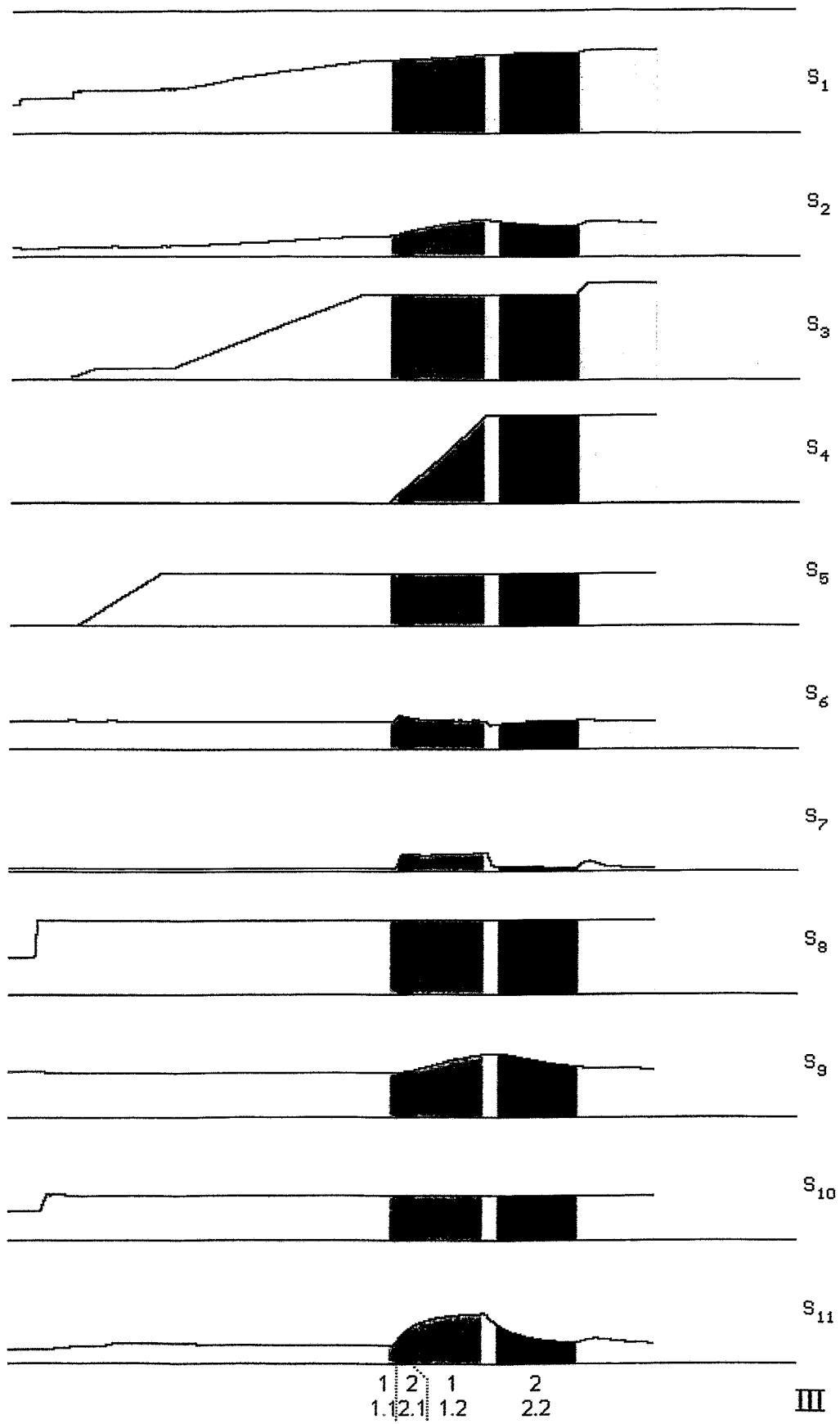
- Die Normalität bedeutet lediglich die Abwesenheit von Gefahr und ist insofern - rein subjektiv gesehen - kaum interessant.
- Die Normalität ist der bei weitem häufigste (wahrscheinlichste) Betriebszustand und hat darum - wie sich informationstheoretisch leicht zeigen läßt - den mit Abstand kleinsten Informations- bzw. Aussagewert. So gesehen ist es auch objektiv kaum zu rechtfertigen, den Anlagenbediener mit einer derart wenig informativen Aussage zu belasten.

**Bild A1-2** (s. nachfolgende Seiten) Die gegebenen Prozeßsignale in den Realisierungen I bis V. Die vorgegebene Klassenzuordnung von Signalabschnitten ist an den unterlegten Grauwerten erkennbar. Nur die Gefahrenklassen - auf sie wird in der Arbeit ausschließlich Bezug genommen - sind außerdem jeweils unten im Bild durch die Indizes der Klassenbezeichnungen angegeben, davon oben: die drei ursprünglichen Klassen, unten: fünf modifizierte (unterteilte) Klassen (s. Abschnitt 6.1).



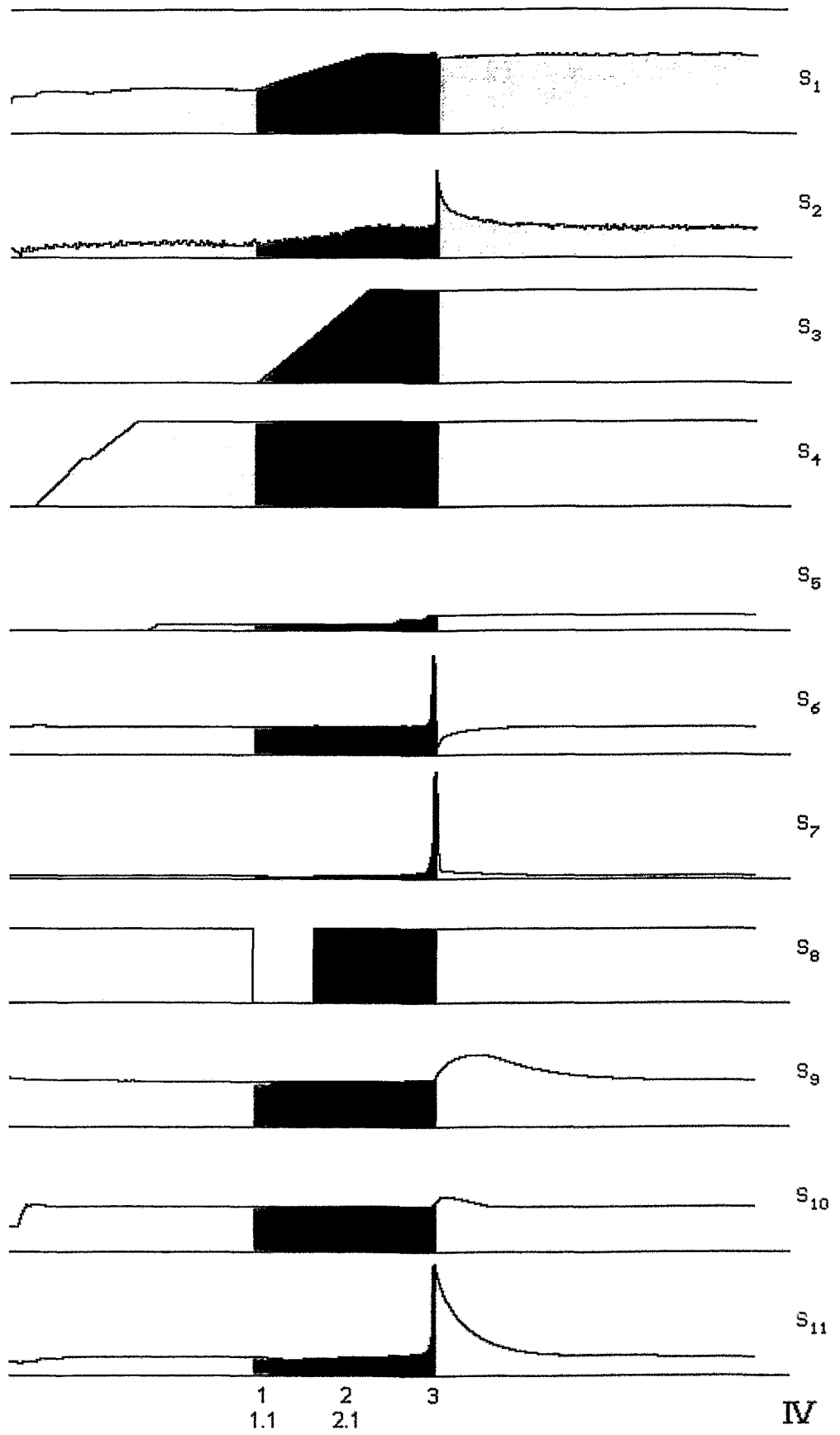


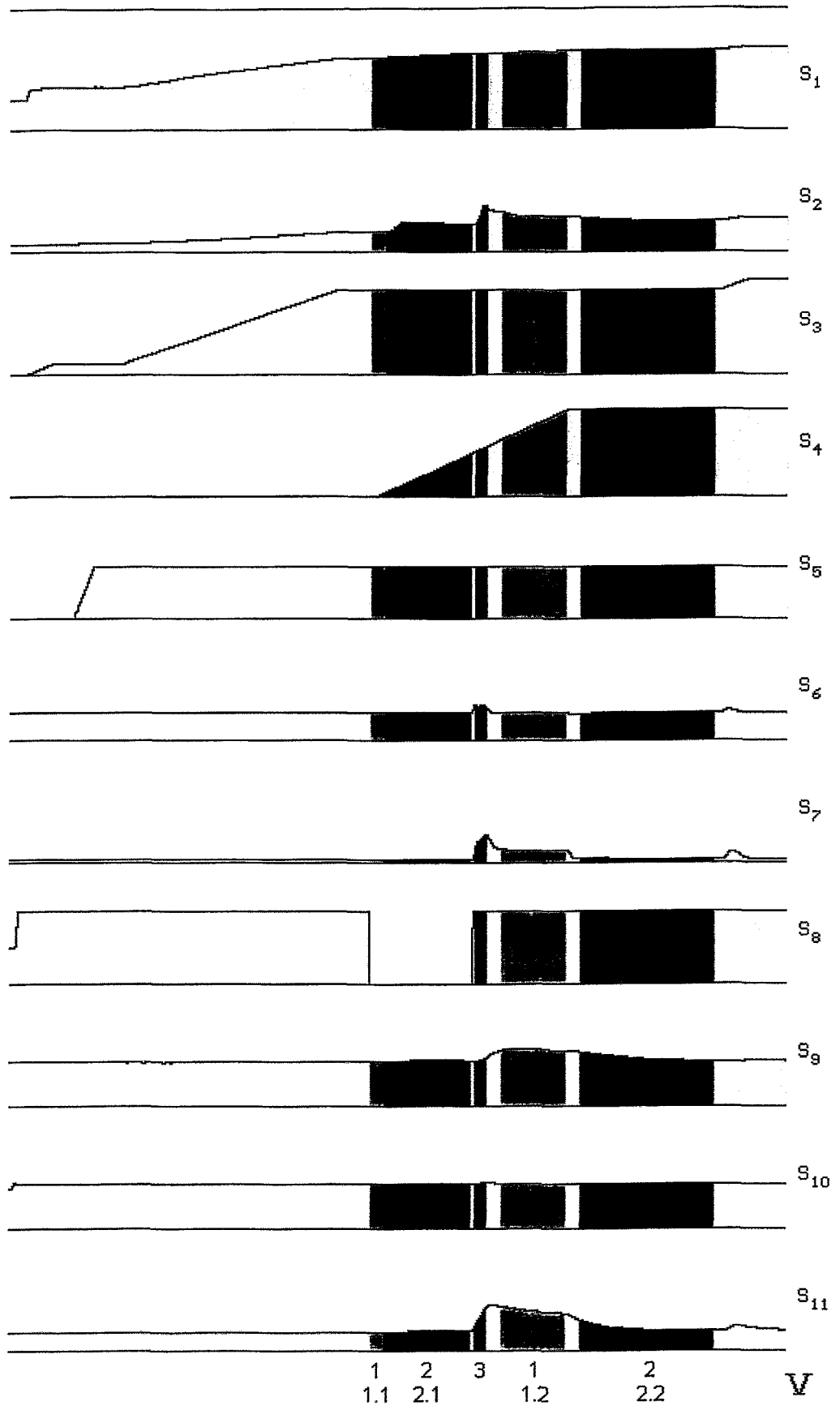




1 2 1 2  
 1.1 2.1 1.2 2.2

III





## Anhang 2: Theoretische Grundlagen des Klassifikators

Die N-dimensionalen Referenzvektoren  $\underline{x}_m$  mit  $m \in \{1, 2, \dots, M\}$ <sup>I</sup> werden zur (N,M)-Referenzmatrix  $\underline{X} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_M)$  zusammengefaßt. Durch Matrizenmultiplikation entsteht die Autokorrelationsmatrix  $\underline{X}\underline{X}^T$ . Wie sich anhand des Orthogonalitätsprinzips [Kr91] zeigen läßt<sup>II</sup>, ist diese Matrix unter bestimmten (allerdings stark einschränkenden) Voraussetzungen als Klassifikator geeignet;  $\underline{X}\underline{X}^T$  heißt diesbezüglich auch assoziative Korrelationsmatrix. Wegen der Schwäche des Matrixkonzeptes, die in den strengen Voraussetzungen begründet ist, werden Tensoren anstelle von Merkmalsvektoren eingeführt.

Die Matrix  $\underline{X}\underline{X}^T$  entspricht einem 2-stufigen Tensor (zur Tensoralgebra siehe z.B. [KI66])

$$\underline{X}^{(2)} = \sum_{i,j=1}^N x^{ij} \underline{e}_i \underline{e}_j \quad (\text{A1a})$$

2 Indizes:  $i, j \in \{(1,1), (1,2), \dots, (1,N), (2,1), \dots, (N,N)\}$

$\underline{e}_i, \underline{e}_j$ : Einheitsvektoren im tensoriellen Produkt

mit den insgesamt  $n^2$  Elementen

$$x^{ij} = \sum_{m=1}^M x_{mi} x_{mj}, \quad x^{ij} \in \{0, 1, \dots, M\}. \quad (\text{A1b})$$

Die Matrix  $\underline{X}\underline{X}^T$  bzw. der Tensor  $\underline{X}^{(2)}$  ist symmetrisch. Aus der Zweistufigkeit des Tensors folgt, daß die Tensorelemente alle 2-elementigen Merkmalsmengen und deren 1-elementigen Teilmengen, die aus der Merkmalsmenge  $\{x_1, x_2, \dots, x_N\}$  gebildet werden können, repräsentieren. Wegen der genannten Symmetrie ist eine 2-elementige Merkmalsmenge durch zwei Matrix- oder Tensorelemente repräsentiert. Zur Erfassung mehrelementiger Merkmalsmengen ist das zweidimensionale Schema der Matrix  $\underline{X}\underline{X}^T$  nicht geeignet. Dies ist ein grundsätzlicher Nachteil der assoziativen Korrelationsmatrix, denn üblicherweise gibt es in den Referenzmustern Merkmalsmengen aus weit mehr als zwei Elementen; die Zusammenhänge zwischen den Elementen solcher Mengen werden von der Matrix nur unvollständig repräsentiert. Zur Überwindung dieses Nachteils wird der Tensor  $\underline{X}^{(2)}$  verallgemeinert zum N-stufigen

---

<sup>I</sup> Damit sind nur die N Merkmale berücksichtigt. Die ebenfalls zu den Referenzvektoren gehörenden K Elemente des Klassencodes (s. Abschnitt 2) werden der einfacheren Schreibweise halber nachfolgend nicht berücksichtigt.

<sup>II</sup> **Orthogonalitätsprinzip:** Der in einem n-dimensionalen Raum definierte unbekannte Merkmalsvektor  $\underline{u}$  soll mit minimalem Fehlerquadrat in den Unterraum projiziert werden, der von den Referenzvektoren  $\underline{x}_m$  aufgespannt wird. Der vektorielle Ansatz ist  $\underline{u} = \underline{X}\underline{g}$  mit  $\underline{g}$  als Vektor von Wichtungsfaktoren;  $\underline{u}$  soll mit  $\underline{g}$  als Linearkombination der  $\underline{x}_m$  ausgedrückt werden. Weil diese Aufgabe im allgemeinen nicht exakt lösbar ist, wird eine Optimalschätzung nach dem Kriterium des minimalen quadratischen Fehlers vorgenommen. Unter der (strengen) Voraussetzung, daß die  $\underline{x}_m$  untereinander orthogonal sind, ergibt sich als optimaler Projektionsoperator die Matrix  $\underline{X}\underline{X}^T$ . Die Elemente des optimal geschätzten Gewichtsvektors  $\underline{\hat{g}}$  gelten als Klassifizierungsergebnis.

gen Referenztensor <sup>1</sup>

$$X^{(N)} = \sum_{i,j,\dots,r=1}^N x^{ij\dots r} \underline{e}_i \underline{e}_j \dots \underline{e}_r \quad (\text{A2a})$$

N Indizes:  $i,j,\dots,r \in \{(1,1,\dots,1), \dots, (N,N,\dots,N)\}$

mit den insgesamt  $N^N$  Elementen

$$x^{ij\dots r} = \sum_{m=1}^M x_{mi} x_{mj} \dots x_{mr}, \quad x^{ij\dots r} \in \{0,1,2,\dots,M\}. \quad (\text{A2b})$$

Die Symmetrieeigenschaft besteht auch hier. In Verallgemeinerung der Symmetrie von  $XX^T$  bzw.  $X^{(2)}$  wird aber beim höherstufigen Tensor von vollständiger Symmetrie gesprochen.

Das unbekannte Muster bildet den ebenfalls N-stufigen unbekannten Tensor

$$U^{(N)} = \sum_{i,j,\dots,r=1}^N u^{ij\dots r} \underline{e}_i \underline{e}_j \dots \underline{e}_r \quad (\text{A3a})$$

mit

$$u^{ij\dots r} = u_i u_j \dots u_{mr}, \quad u^{ij\dots r} \in \{0,1\}. \quad (\text{A3b})$$

Der Wert eines Elementes  $x^{ij\dots r}$  gibt an, wie häufig die Merkmalsmenge  $\{x_i, x_j, \dots, x_r\}$  in den Referenzmustern vorkommt. Jede mögliche Teilmenge, die aus der Merkmals-Gesamtmenge  $\{x_1, x_2, \dots, x_N\}$  gebildet werden kann, ist im Referenztensor  $X^{(N)}$  durch ein oder - wegen der vollständigen Symmetrie des  $X^{(N)}$  - durch mehrere Tensorelemente repräsentiert. Desgleichen gilt für den unbekanntem Tensor.

Wegen der vollständigen Symmetrie der eingeführten Tensoren ist die Anzahl  $a$  von Tensorelementen, die zu einer  $s$ -elementigen Merkmalsmenge gehört,  $s \in \{1, 2, \dots, N\}$ , bestimmt durch

$$\begin{aligned} a(N,s) &= 1 && \text{für } s=1 \\ &= N! && \text{für } s=N \\ &= s[a(N-1,s-1) + a(N-1,s)] && \text{für } 1 < s < N, 2 < N. \end{aligned} \quad (\text{A4})$$

Da jede Merkmalsmenge und -teilmenge, die in den Referenzmustern vorkommt, durch mindestens ein Tensorelement repräsentiert ist, enthält  $X^{(N)}$  im Unterschied zu  $XX^T$  bzw.  $X^{(2)}$  stets die vollständige Information über die Referenzmuster. Folglich sind aus  $X^{(N)}$  sämtliche Referenzmuster rekonstruierbar. Hinsichtlich der Rekonstruktion kann  $X^{(N)}$  formal als Summe über  $M$  Teiltensoren (Referenztensoren jedes einzelnen Referenzmusters) dargestellt werden; aus (A2) folgt unter Vertauschung der Summationen

---

<sup>1</sup> Die Tensorstufe  $N$  ist hinreichend, um alle möglichen Merkmalsmengen und -teilmengen ausdrücken zu können, die sich mit  $N$  Merkmalen bilden lassen. Größere Tensorstufen sind erlaubt aber nicht notwendig. Kleinere Tensorstufen sind dann ausreichend, wenn die gegenseitige Abhängigkeit der Merkmale entsprechend gering ist; andernfalls treten Informationsverluste auf.

$$X^{(N)} = \sum_{m=1}^M X_m^{(N)} \quad (\text{A5a})$$

mit

$$X_m^{(N)} = \sum_{i,j,\dots,r=1}^N x_m^{ij\dots r} e_i e_j \dots e_r \quad (\text{A5b})$$

$$x_m^{ij\dots r} = x_{mi} x_{mj} \dots x_{mr}, \quad x_m^{ij\dots r} \in \{0, 1\}. \quad (\text{A5c})$$

Den Wert eins haben diejenigen Tensorelemente  $x_m^{ij\dots r}$ , welche die Merkmalsmenge des Referenzmusters  $\{x_m\}$  oder deren Teilmengen repräsentieren. Das zur vollständigen Menge  $\{x_m\}$  und damit zum m-ten Referenzmuster gehörende Element  $x_m^{ij\dots r}$  ist daran erkennbar, daß es die meisten unterschiedlichen Indizes hat; aus diesen Indizes kann unmittelbar das m-te Referenzmuster rekonstruiert werden.

Die Herleitung eines Klassifikators, der auf der tensorieller Musterdarstellung beruht, folgt wieder dem Orthogonalitätsprinzip:

Um den unbekanntem Tensor  $U^{(N)}$  durch den gewichteten Referenztensor  $X^{(N)}$  auszudrücken, kommt zur Wichtung nur ein skalares Gewicht  $g$  in Betracht. Aus dem tensoriellen Ansatz

$$U^{(N)} = X^{(N)} g \quad (\text{A6a})$$

folgt das Optimierungsproblem

$$[U^{(N)} - X^{(N)} g] \cdot [U^{(N)} - X^{(N)} g] = \text{Min!}_g, \quad (\text{A6b})$$

wobei der Punkt als Tensorprodukt zu verstehen ist. Das Tensorprodukt in (A6b) erstreckt sich über alle Tensorstufen, so daß im Ergebnis bestimmter tensorieller Operationen (Überschiebung und Verjüngung) ein Tensor nullter Stufe, d.h. ein Skalar entsteht. Daraus folgt als optimales Schätzwert

$$\hat{g} = (X^{(N)} \cdot X^{(N)})^{-1} (X^{(N)} \cdot U^{(N)}) \quad (\text{A7a})$$

mit

$$X^{(N)} \cdot U^{(N)} = \sum_{i,j,\dots,r=1}^N x_m^{ij\dots r} u^{ij\dots r}. \quad (\text{A7b})$$

Der Term  $X^{(N)} \cdot X^{(N)}$  ist skalar und deshalb stets invertierbar <sup>‡</sup>. Weil dieser Term nicht unterscheidungsrelevant und später durch Normierung ohnehin wegfällt, wird er nicht weiter beachtet. Nur der ebenfalls skalare Term  $X^{(N)} \cdot U^{(N)}$  drückt die unterscheidungsrelevanten Beziehungen zwischen den Referenzmustern und dem unbekanntem Muster aus.

Während die Beziehungen des unbekanntem Musters zu den Klassen im vektoriiellen Fall direkt durch die Elemente des Gewichtsvektors  $\hat{g}$  zum Ausdruck kommen, zei-

---

<sup>‡</sup> Abgesehen von dem praktisch bedeutungslosen Fall, daß alle Referenzvektoren Nullvektoren und folglich alle Elemente des Tensors  $X_{(N)}$  gleich null sind.

gen sich diese Beziehungen im tensoriellen Fall nur indirekt an den problembezogenen Anteilen der einzelnen Teiltensoren  $X_m^{(N)}$  am optimalen Schätzwert  $\hat{g}$ . Die Problembezogenheit dieser Anteile wird durch die Beziehungen der Referenzmuster zum unbekanntem Muster bestimmt, also durch den Term  $X^{(N)} \cdot U^{(N)}$  in (A7). Wie aus der Summendarstellung dieses Terms hervorgeht, s. (A7b), selektieren die Tensorelemente  $u^{i \dots r}$  die gleichindizierten Tensorelemente  $x^{i \dots r}$ . Weil nur dies für die Klassifikation von Bedeutung ist, braucht die tensorielle Multiplikation  $X^{(N)} \cdot U^{(N)}$  nicht wirklich ausgeführt zu werden. Statt dessen werden die Produkte  $x^{i \dots r} u^{i \dots r}$  als Tensorelemente  $x_m^{i \dots r}$  eines neuen, reduzierten (an das unbekanntem Muster angepaßten) Referenzensors  $X_m^{(n)}$  angesehen. Dieser Tensor kann analog (A5) in Teiltensoren zerlegt werden:

$$X_m^{(N)} = \sum_{m=1}^M X_m^{(N)} \quad (A8a)$$

mit

$$X_m^{(N)} = \sum_{i,j,\dots,r=1}^N x_m^{i,j,\dots,r} e_i e_j \dots e_r \quad (A8b)$$

$$x_m^{i,j,\dots,r} = (x_{mi} u_i)(x_{mj} u_j) \dots (x_{mr} u_r), \quad x_m^{i,j,\dots,r} \in \{0,1\} \quad (A8c)$$

Der Teiltensor  $X_m^{(N)}$  ist der gesuchte problembezogene Anteil des Teiltensors  $X_m^{(N)}$  am optimalen Schätzwert  $\hat{g}$ . Dieser Gewichtsanteil ist gemäß (A8b) durch die Summe über alle Elemente des Teiltensors bestimmt:  $\sum_{i,j,\dots,r=1}^N x_m^{i,j,\dots,r}$ . Diese Summe ist wegen der Tensorsymmetrie von der Tensorstufe abhängig, s. (A4), woraufhin die Tensorstufe einen (nicht gerechtfertigten) Einfluß auf das Klassifizierungsergebnis erhalten kann. Die Abhängigkeit von der Tensorstufe wird vermieden, indem nur über alle nichtsymmetrischen Tensorelemente summiert wird, d.h. indem für jede Merkmalsmenge und -teilmenge, die im Tensor repräsentiert ist, nur ein einziges Tensorelement berücksichtigt wird. Die praktische Durchführung der Summation kann entfallen, weil ihr Ergebnis gleichermaßen durch die Anzahl aller im Teiltensor  $X_m^{(N)}$  repräsentierten Merkmalsmengen und -teilmengen bestimmt ist: Besteht die Obermenge aller dieser Mengen aus  $g_m$  Merkmalen, ist die Gesamtzahl dieser Mengen und somit der gesuchte Gewichtsanteil  $G_m = 2^{g_m} - 1$ , vgl. (6.1).

Der Wert  $G_m$  wird dem Referenzvektor  $x_m$  zugeordnet und insbesondere - da klassifiziert werden soll - den darin enthaltenen Klassenbezeichnungen  $y_{mk}$ . Durch die Summation der Produkte  $G_m y_{mk}$  über alle  $m$  Referenzvektoren entsteht die Gesamtgewichtung der Klasse  $y_k$ . Dies wird für alle Klassen vorgenommen und liefert eine Gewichtsverteilung - das Klassifizierungsergebnis; eine geeignete Normierung kann sinnvoll sein.

Weil die Klassengewichte  $G_m$  unmittelbar aus Mengenbetrachtungen hergeleitet werden können, ist die praktische Bildung von Tensoren unnötig. Für das Klassifizierungsverfahren ist es daher nur noch von formaler Bedeutung, daß das Schätzwert  $\hat{g}$  im Sinne des tensoriellen Kriteriums (A6) optimal ist.

Ergänzend wird überprüft, ob das tensororientierte Verfahren den für Klassifikatoren wichtigen Metrikbedingungen genügt. Ein wichtiges Kriterium für die Eignung eines

Ähnlichkeits- bzw. Distanzmaßes besteht darin, ob mit Hilfe der Distanz ein metrischer Raum strukturierbar ist. Die Ähnlichkeit zwischen dem unbekanntem Muster und einem Referenzmuster wird im vorliegenden Fall durch das Klassengewicht  $G_m$  ausgedrückt. Um die zugehörige Distanz zu formulieren, wird zunächst die Hamming-Distanz betrachtet.

Die Hamming-Distanz gibt die Anzahl unterschiedlich besetzter Merkmale an, in denen sich das unbekanntem Muster  $u$  von einem Referenzmuster  $x_m$  unterscheidet:

$$d_m = \sum_{n=1}^N |u_n - x_{mn}|, \quad d_m \in \{0, 1, \dots, N\}.$$

Zu dieser Distanz äquivalent ist die Ähnlichkeit

$$g_m = N - d_m, \quad g_m \in \{0, 1, \dots, N\}.$$

Sie gibt die Anzahl der in den Mustern  $u$  und  $x_m$  gemeinsam besetzten (mit eins bewerteten) Elemente an. Die kleinstmögliche Hamming-Distanz ist 0. Die größtmögliche Ähnlichkeit ist folglich die Merkmalsanzahl  $N$ .

In umgekehrter Betrachtung folgt aus dem Maximalwert der oben bestimmten Ähnlichkeit  $G_m$ , nämlich  $G_{m,\max} = 2^N - 1$ , die maximale Distanz

$$\begin{aligned} D_m &= G_{m,\max} - G_m \\ &= 2^N - 1 - (2^{sm} - 1) = 2^N - 2^s, \quad D_m \in \{0, 1, \dots, 2^N - 1\}. \end{aligned}$$

Sind  $a, b, c$  drei beliebige Punkte in einem Merkmalsraum und  $D_{ab}, D_{bc}, D_{ac}$  die Abstände zwischen ihnen, so läßt sich zeigen, daß die drei Metrikbedingungen [Kr91]

- 1)  $D_{ab} \geq 0, D_{ab} = 0 \leftrightarrow a = b$  (Axiom der Identität)
- 2)  $D_{ab} = D_{ba}$  (Axiom der Symmetrie)
- 3)  $D_{ac} \leq D_{ab} + D_{bc}$  (Dreiecksungleichung)

erfüllt sind.  $D_m$  ist damit eine Metrik und folglich ist die Ähnlichkeit  $G_m$  für die Klassifikation geeignet.



### Anhang 3: Strukturelemente eines dynamischen Prozeßmodells

Die im Abschnitt 6.2.2 informationstheoretisch nachgewiesene Abhängigkeit zwischen den Signalen  $S_9$  (Temperatur des externen Kühlmittels) und  $S_{11}$  (Temperatur im Reaktor) wird zum Anlaß genommen, ein Modell desjenigen Teils des vorliegenden dynamischen Prozesses zu generieren, welcher die nachgewiesene Signalabhängigkeit verursacht.

Die informationstheoretisch günstigste zeit- und wertmäßige Signalauflösung, in der das Modell gelten sollte, ist bereits indirekt durch die Definition der Merkmale bestimmt; bei der Merkmalsbildung wurden bereits geeignete Zeit- und Wertaufösungen gefunden, s. Abschnitt 5.4.

Daß sich das Modell auf wirklich interessierende (besonders sicherheitsrelevante) Vorgänge bezieht, ist dadurch gewährleistet, daß die nachgewiesene Signalabhängigkeit bezüglich der (Gefahren-)Klasse  $y_{2.2}$  besteht.

Die Abhängigkeit zwischen den Signalen  $S_9$  und  $S_{11}$  wurde anhand der Transinformation nachgewiesen. Weil die Transinformation aber nichts über eine Ursache-Wirkungs-Richtung innerhalb der Signalabhängigkeit besagt, wird als allgemeinster Fall die wechselseitige Kopplung der beiden Signale angenommen. Davon ausgehend werden

- den (Prozeß-)Ausgangsgrößen  $S_9$  und  $S_{11}$  formale Eingangsgrößen  $u_9$  und  $u_{11}$  gegenübergestellt,
- als Strukturansatz ein Zweigrößen-Übertragungsglied in V-kanonischer Struktur [177] eingeführt (s. Bild A3-1) und
- für die Identifikation der im Bild A3-1 dargestellten Eingrößen-Übertragungsglieder  $G_9(s)$ ,  $G_{11}(s)$ ,  $V_{9,11}(s)$ ,  $V_{11,9}(s)$  sowie für Kombinationen derselben die folgende allgemeine Übertragungsfunktion benutzt:

$$\frac{b_0 + b_1s + \dots + b_{m-1}s^{m-1} + b_ms^m}{a_0 + a_1s + \dots + a_{n-1}s^{n-1} + a_ns^n} \cdot e^{-T_t} \quad (T_t: \text{Totzeit}) \quad (\text{A3-1})$$

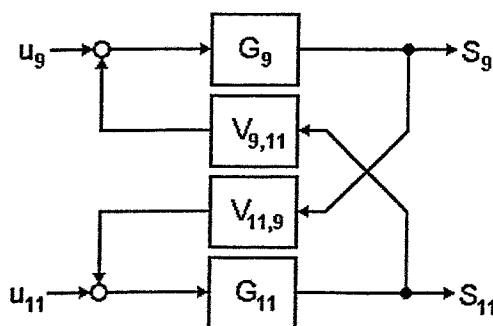


Bild A3-1 V-kanonischer Strukturansatz für ein Teilmodell zur Beschreibung der Prozeßdynamik.

Aus Bild A3-1 wird abgelesen:

$$S_9 = G_9(u_9 + V_{9,11}S_{11}) \quad (\text{A3-2a})$$

$$S_{11} = G_{11}(u_{11} + V_{11,9}S_9) \quad (\text{A3-2b})$$

Weil nur die Ausgangsgrößen  $S_9$  und  $S_{11}$  real vorliegen, die Eingangsgrößen  $u_9$  und  $u_{11}$  aber nicht bekannt sind, wird unter der willkürlichen (aber im nachhinein zu prüfenden) Annahme

$$u_9 \equiv u_{11} \equiv 0 \quad (\text{A3-3})$$

eine Konkretisierung der Teilsysteme  $(V_{9,11}G_9)$ ,  $(V_{11,9}G_{11})$  versucht, indem über ein (a) nichtparametrisches Frequenzgangmodell ein (b) parametrisches dynamisches Modell gemäß (A3-1) gewonnen wird.

Zu der im Schritt (a) notwendigen diskreten Fourier-Transformation oder äquivalent zur Bildung der Koeffizienten der Fourier-Reihe von  $S_9$  und  $S_{11}$  sind die gegebenen Abschnitte dieser Signale formal periodisch fortzusetzen. Wegen der günstigen Form der Signalverläufe<sup>I</sup> erübrigt sich dabei eine Glättung der Übergänge (Hanning-Fenster, Dolph-Tschebyscheff-Fenster o.ä., [Kr91] S.144).

Aus (A3-2) mit (A3-3) folgt

$$\frac{S_9}{S_{11}} = V_{9,11}G_9 \quad , \quad \frac{S_{11}}{S_9} = V_{11,9}G_{11} \quad .$$

Die Division im Frequenzbereich hat nur für  $(V_{11,9}G_{11})$  ein physikalisch sinnvolles Ergebnis, den Frequenzgang

$$\frac{S_9(\omega)}{S_{11}(\omega)} = \frac{\mathcal{F}[S_9(t)]}{\mathcal{F}[S_{11}(t)]} = \mathcal{F}[V_{9,11}G_9] \quad . \quad (\text{A3-4})$$

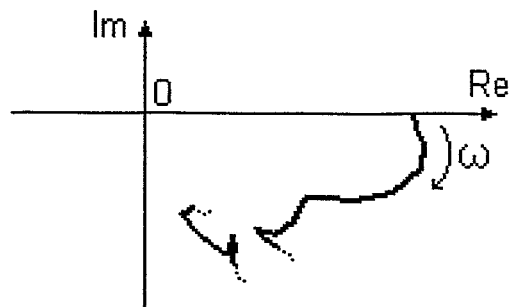
Damit ist auch die Frage nach der kausalen Abhängigkeit der beiden Signale beantwortet: In den durchgeführten Versuchen wirkt  $S_{11}$  rückwirkungsfrei auf  $S_9$ <sup>II</sup>.

Die praktische Gewinnung eines parametrischen dynamischen Modells gemäß Schritt (b) aus dem Frequenzgang (A3-4) ist dadurch erschwert, daß der Prozeß im modellierungsrelevanten Frequenzband mit sehr unterschiedlicher Intensität und zuweilen überhaupt nicht angeregt ist. Dadurch ist der Frequenzgang frequenzvariant fehlerbehaftet oder er existiert für manche Frequenzen gar nicht. Der Frequenzgangfehler ist etwa reziprok dem Betrag des Nenners in (A3-4) und wächst mit zunehmender Frequenz. Für Abschnitte mit vertretbar kleinem Fehler zeigt Bild A3-2 den Real- und Imaginärteil von  $\mathcal{F}[V_{9,11}G_9]$  als Frequenzgang-Ortskurve.

---

<sup>I</sup> Weil die Signalabschnitte in Wert und Anstieg (fast) ebenso enden wie sie beginnen, können Anfang und Ende dieser Abschnitte direkt miteinander verbunden werden, ohne daß durch die Übergänge wesentlich parasitäre (nicht vom Prozeß stammende) Spektralanteile entstehen.

<sup>II</sup> Dies schließt nicht aus, daß durch andere Versuche auch die Gegenrichtung angeregt werden kann.



**Bild A3-2** Ortskurve des Frequenzganges  $\mathcal{F}[V_{9,11}G_9]$  für Abschnitte mit geringem Fehler

Wegen der beträchtlichen Fehler im Frequenzgang ist die große Fülle der bekannten analytischen und grafischen Verfahren der Systemanalyse [St75] kaum nutzbar. Das weitere Vorgehen ist deshalb etwas heuristisch geprägt.

Weil die Ortskurve allenfalls im untersten Frequenzbereich günstig interpretierbar ist, werden aus diesem Bereich subjektiv folgende Hinweise auf die Struktur des parametrischen Modells gewonnen: <sup>1</sup>

- Die Ortskurve beginnt auf der positiven reellen Achse. Dies weist auf proportionales Übertragungsverhalten des Teilsystems  $(V_{9,11}G_9)$  hin. Also sind im Ansatz (A3-1) die Parameter  $a_0, b_0 \neq 0$ .
- Die Ortskurve tritt aus der reellen Achse nach unten hin aus. Das bedeutet verzögerndes Übertragungsverhalten. Folglich muß für  $(V_{9,11}G_9)$  im Ansatz (A3-1) der Nennergrad größer als der Zählergrad ( $n > m$ ) sein.
- Weil sich die Ortskurve anfangs vom Koordinatenursprung entfernt, der Betrag des Frequenzganges also zunimmt, ist schwingfähiges Verhalten von  $(V_{9,11}G_9)$ , d.h. ein konjugiert komplexes Polpaar in (A3-1) zu vermuten.
- Weil sich die Ortskurve im weiteren Verlauf schnell dem Koordinatenursprung nähert, wird trotz des rapide zunehmenden Frequenzgangfehlers ein totzeitbehaftetes oder nichtminimalphasiges Verhalten ausgeschlossen.

Von diesen Hinweisen ausgehend werden für  $(V_{9,11}G_9)$  die strukturellen Ansätze

- Verzögerungsglied 1. Ordnung (ein reeller Pol der Übertragungsfunktion)
- Verzögerungsglied 2. Ordnung (zwei reelle Pole),
- Schwingungsglied mit Verzögerung (ein konjugiert komplexes Polpaar und ein reeller Pol)

untersucht. Die optimale Parameterschätzung nach der bekannten Methode der kleinsten Fehlerquadrate (MKQ) unter Verwendung des verallgemeinerten Fehlers erbringt nur für das  $PT_1$ -Glied ein günstiges Ergebnis:

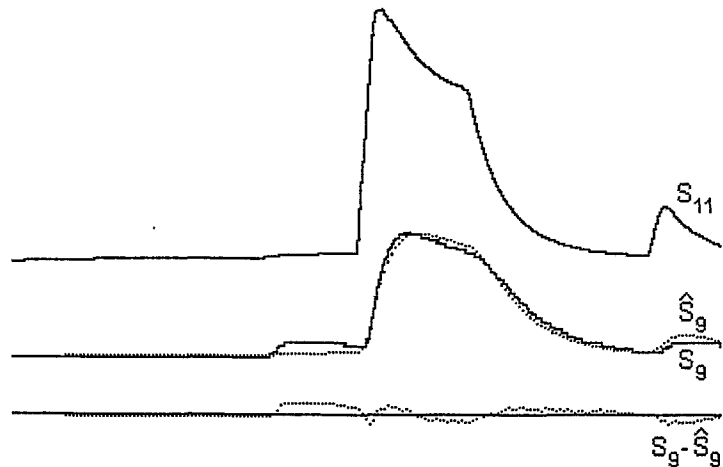
---

<sup>1</sup> In weiterführenden Untersuchungen werden solche Informationen objektiv per Klassifikation gewonnen und für diesen Zweck die Merkmalsdefinition analog zur Zeit-Wert-Rasterung (vgl. Abschnitt 5.4) durch eine ortskurven-spezifische Segmentierung der komplexen Zahlenebene vorgenommen.

$$V_{9,11}(s)G_9(s) = \frac{S_9(s)}{S_{11}(s)} = \frac{b_0}{1 + a_1s} = \frac{K}{1 + T_1s} \quad (\text{A3-5})$$

mit  $K$ : Verstärkungsfaktor ( $K=0,60$  bezüglich des Kleinsignalverhaltens)  
 $T_1$ : Zeitkonstante ( $T_1=63\text{s}$ ).

Probehalber wird aus  $S_{11}$  anhand von (A3-5) das Schätzsignal  $\hat{S}_9(t)$  bestimmt, s. Bild A3-3.



**Bild A3-3** Zur Beurteilung des Schätzfehlers

Wie Bild A3-3 zeigt, ist der Schätzfehler  $S_g - \hat{S}_g$  recht klein. Dies bestätigt, daß

- die Annahme  $u_g \approx 0$  aus (A3-3) sinnvoll ist,
- die Signalabhängigkeit im vorliegenden Fall von  $S_{11}$  nach  $S_g$  gerichtet ist und
- die Annahme des  $PT_1$ -Charakters für das Teilsystem ( $V_{9,11}G_9$ ) sinnvoll ist, zumindest als dominierender Modellanteil.

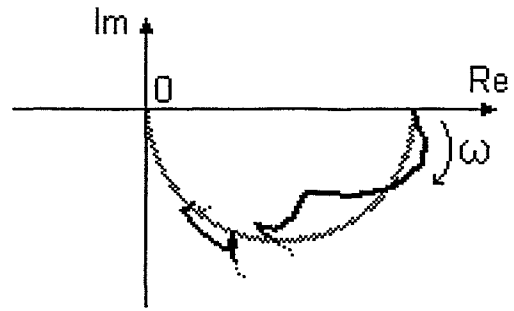
Auch physikalisch ist das gewonnene Ergebnis plausibel, denn

- die Wärmefreisetzung erfolgte ursächlich durch die exotherme chemische Reaktion im Reaktor, also seitens  $S_{11}$  oder  $u_{11}$  und
- eine Anregung seitens  $S_{11}$  kann sich, da ein Wärmetransport aus dem Reaktor in das Kühlmittel erfolgt, nur verzögert und ohne prompten Anteil auf  $S_g$  auswirken.

(A3-5) kann somit als Grundbaustein eines eventuell ausbaufähigen Modells für ( $V_{9,11}G_9$ ) angesehen werden.

Eine Totzeit wäre nach der Anschauung zwar naheliegend. Sie tritt aber in der modellrelevanten Zeitaufösung nicht in Erscheinung; sie ist zu klein.

Eine gemeinsame Darstellung der zum identifizierten  $PT_1$ -Glied gehörenden Ortskurve und der Ortskurve des empirisch gewonnenen Frequenzganges (s. Bild A3-2) zeigt Bild A3-4.



**Bild A3-4** Empirische Ortskurvendarstellung aus Bild A3-2 gemeinsam mit der zum identifizierten  $PT_1$ -Glied gehörenden (gepunkteten) Ortskurve

Der Vergleich der empirischen und der geschätzten Ortskurve ist bedeutsam dafür, daß vor allem die Unterschiede zwischen beiden Kurven eventuell bestehende Möglichkeiten zur weiteren Modellverfeinerung hervortreten lassen: Die im wenig fehlerbehafteten (und somit durchaus ernstzunehmenden) Anfangsabschnitt der empirischen Ortskurve liegende, auf Schwingungsverhalten hinweisende Ausbuchtung tritt im Bild A3-4 wegen ihres Gegensatzes zur  $PT_1$ -Ortskurve besonders deutlich hervor. Dies bekräftigt die Sinnfälligkeit der Einführung eines Schwingungsansatzes für die nächste Modell-Ausbaustufe <sup>I</sup>. Physikalische Ursache für das Schwingungsverhalten, welches sich in der empirischen Ortskurve andeutet, kann eine Periodizität im Kühlmittelumlauf sein.

Die Tatsache, daß aus dem ersten Modell-Strukturansatz ( $PT_1$ ) nicht unmittelbar, sondern erst nach erfolgter Parameteridentifikation ein besser fundierter Hinweis für den weiteren Ausbau der Modellstruktur hervorgeht, hat allgemeingültige Bedeutung:

Ein fortschreitender Ausbau der Modellstruktur (die qualitative Seite der Modellbildung) ist ohne eine begleitende Parameteridentifikation (die quantitative Seite der Modellbildung) nicht möglich.

Zur weiteren Konkretisierung des Strukturansatzes in Bild A3-1 sind die gegebenen Signalverläufe nicht ausreichend <sup>II</sup>. Der Strukturansatz gibt aber eine Orientierung für die Planung spezieller Versuche zur Modellbildung:

- Zur Bestimmung des Teilsystems ( $V_{11,9}G_{11}$ ) ist eine Anregung seitens  $S_9$  oder  $u_9$  bzw. durch eine primäre Temperaturänderung im Kühlmittelumlauf notwendig.
- Zur Bestimmung der einzelnen Eingrößenglieder müssen beide Eingänge definiert und zweckmäßigerweise getrennt voneinander angeregt werden, z.B. durch die Einspeisung elektrischer Heizleistung.

<sup>I</sup> Der Schwingungsansatz konnte nur im ersten Schritt der Modellstrukturierung mit dem gegebenen Datenmaterial nicht bestätigt werden. Das muß im zweiten Schritt nicht ebenfalls so sein, denn bei Richtigkeit der Koeffizienten in (A3-5) erfolgt die Suche nach weiteren Modellanteilen bereits in einem dimensionsreduzierten Suchraum; darauf wird hier nicht näher eingegangen.

<sup>II</sup> Die Experimente galten ja ursprünglich auch nicht vordergründig der Modellbildung, sondern der Signalerkennung.

Bei thermohydraulischen Experimenten an einer vergleichbaren anderen Anlage, bestehend aus zwei Wärmespeichern (Reaktor und Kühlmittelumlauf) und mit Wärmeübergängen zwischen den Speichern sowie zur Umgebung hin konnte das V-kanonisch angesetzte Modell vollständig identifiziert werden. Darüber hinaus verhalf das verbesserte Modell dazu, bislang unerkannte Widersprüche zwischen dem Modell und dem realen Prozeß zu erkennen und daraus Hinweise für Modellverbesserungen abzuleiten.

Die Problematik von Modellverbesserungen in Verbindung mit der Lösung von Widersprüchen wird in einem Folgebeitrag genauer behandelt.

## Literatur

- [Ge97] Deerberg, G.: Zur sicherheitstechnischen Beurteilung von Semibatch-Prozessen mit Gas/Flüssigkeits-Systemen. Fraunhofer IRB Verlag, 1997
- [H94] Hoppe, D.: Ein Vergleich zwischen assoziativen Korrelationsmatrizen und assoziativen Tensoren mit Bezug zur optimalen Signalschätzung. Automatisierungstechnik, (1994)8, 337-344
- [H96] Hoppe, D.: Anwendung eines tensororientierten Mustererkennungsverfahrens zur Leckortung an Druckanlagen. GMA-Kongreß '96, VDI-Berichte 1282, VDI-Verlag
- [I77] Isermann, R.: Digitale Regelsysteme. Springer-Verlag, Berlin u.a., 1977
- [KI66] Klingbeil, E.: Tensorrechnung für Ingenieure. Bibliographisches Institut Mannheim, Hochschulbücher-Verlag 1966.
- [Kr91] Kronmüller, H.: Digitale Signalverarbeitung. Springer-Verlag, Berlin, Heidelberg u.a., 1991
- [N74] Niemann, H.: Methoden der Mustererkennung. Akademische Verlagsgesellschaft, Frankfurt a.M., 1974
- [P91] Peters, H.: Situationserkennung bei kontinuierlichen Prozessen. atp - Automatisierungstechnische Praxis, R. Oldenbourg Verlag, 30(1991)11
- [RF94] Römer, H., Filk, T.: Statistische Mechanik. VCH Verlagsgesellschaft mbH, Weinheim, 1 994
- [SF76] Steinhagen, H.-E., Fuchs, S.: Objekterkennung - Einführung in die mathematischen Methoden der Objekterkennung. Verlag Technik, Berlin, 1976
- [Sh48] Shannon, C.E.: A mathematical theory of communication. The Bell System Technical Journal 27, 379-423, 623-656, 1948
- [St75] Strobel, H.: Experimentelle Systemanalyse. Akademie-Verlag, Berlin, 1975
- [UGB74] Unbehauen, H., Göring, B., Bauer, B.: Parameterschätzverfahren zur Systemidentifikation. Oldenbourg, München, 1974
- [V82] Völz, H: Information I. Akademie-Verlag, Berlin, 1982