

**Automatische Auswahl geeigneter
Visualisierungstechniken für allgemeine
wissenschaftliche Datensätze**

Diplomarbeit

Universität Rostock
Fachbereich Informatik

vorgelegt von:

Holger Theisel
geboren am 16.08.1969 in Jena

Betreuer: Prof. Dr. ing. habil. Heidrun Schumann

Abgabedatum: 01.07.1994

Zusammenfassung:

Zur Visualisierung wissenschaftlicher Daten existiert eine Vielzahl von Visualisierungstechniken, welche mehr oder weniger gut für ein konkretes Problem geeignet sind. Die Auswahl einer geeigneten Technik ist eine Voraussetzung für den Erfolg der visuellen Analyse, eine automatische Hilfe bei dieser Auswahl ist notwendig.

Bisher existierende Ansätze für solche Auswahlhilfen sind nur für spezielle Datenklassen sinnvoll anwendbar.

In dieser Arbeit soll ein Algorithmus eingeführt und diskutiert werden, der für beliebige wissenschaftliche Datensätze geeignete Visualisierungstechniken ermittelt. Dazu wird zunächst eine einheitliche Form der Datenbeschreibung angegeben und ein einheitliches System abstrakter Interpretationsziele eingeführt. Der darauf basierende Algorithmus wird mit bereits vorhandenen Algorithmen verglichen und durch Beispiele unterlegt.

Abstract:

Given a scientific data set there are many different techniques to visualize it. The usage of a suitable visualization technique is a requirement for the success of the visual analysis.

In this paper a new approach is introduced to choosing suitable techniques for any scientific data automatically. For that purpose a general data description is given and a general system of abstract interpretation aims is introduced. The new algorithm is discussed and compared with known algorithms.

CR-Spezifikation:

I.3 Computer Graphics

I.3.3 Display Algorithms

I.3.6 Methodology and Techniques

Key Words:

scientific visualization, visualization techniques, interpretation aims, scientific data record, information theory

Inhaltsverzeichnis:

1. Einführung und Problemstellung	4
2. Datenbeschreibung	7
2.1 Datenbeschreibung als Matrix	7
3. Interpretationsziele	12
3.1 Ein System abstrakter Interpretationsziele	13
3.2 Die Abbildung konkreter Interpretationsziele auf abstrakte	33
4. Der Algorithmus zur Auswahl geeigneter Visualisierungstechniken	36
4.1 Die Wahl der Abstandsfunktion dist	37
5. Beispiele, Tests, Ergebnisse	40
5.1 Charakteristika konkreter Visualisierungstechniken	40
5.2 Beispiele für die Anwendung des Auswahlalgorithmus	45
6. Ungenauigkeiten, Unkorrektheiten	
48	
7. Ausblick	50

1. Einführung und Problemstellung

Die Visualisierung wissenschaftlicher Daten entwickelte sich in den letzten Jahren zu einem Schwerpunkt der modernen Computergraphik. Riesige Datenmengen, die aus verschiedenen Gebieten gewonnen werden, müssen analysiert werden.

Es ist naheliegend, Mittel der Computergraphik zur Analyse zu nutzen, da der Mensch über das Auge eine Vielzahl verschiedener Informationen gleichzeitig aufnehmen kann. Das ergibt jedoch auch völlig neue Aufgaben für die Computergraphik.

Das Ziel der herkömmlichen Computergraphik war es hauptsächlich, realistische Darstellungen natürlicher oder künstlicher Objekte in einer möglichst geringen Zeit zu erzeugen. Für die Visualisierung wissenschaftlicher Daten stehen andere Aufgaben im Vordergrund. Hier kommt es weniger auf Realismus der Bilder an, sondern darauf, daß der Nutzer alle wichtigen Informationen, die im Datensatz enthalten sind, erkennen kann.

Es wurde eine Vielzahl von Techniken entwickelt, um wissenschaftliche Daten zu visualisieren. Jede dieser Techniken ist für bestimmte Fälle gut geeignet, für andere weniger. Wird ein Datensatz mit einer ungeeigneten Technik visualisiert, so können ungenügende oder gar falsche Erkenntnisse über den Datensatz aus der Visualisierung gewonnen werden. Die Wahl einer geeigneten Visualisierungstechnik ist also eine entscheidende Voraussetzung für den Erfolg der visuellen Analyse.

Bei der Auswahl passender Visualisierungstechniken ist der Nutzer oft auf sich allein gestellt, gängige Visualisierungssysteme bieten hier keine Unterstützung an. Eine richtige Auswahlentscheidung erfordert jedoch weiterreichendes Wissen über alle in Frage kommenden Visualisierungstechniken sowie darüber, unter welcher konkreten Zielstellung die Visualisierung erfolgt. Dieses Wissen eines Visualisierungsexperten kann beim Nutzer nicht vorausgesetzt werden.

Notwendig sind also Systeme, die automatisch entscheiden können, welche Visualisierungstechniken für ein konkretes Problem geeignet sind und welche nicht. Mit Hilfe einer solchen automatischen Entscheidungshilfe kann die große Menge der möglichen Visualisierungstechniken reduziert werden auf eine kleinere Menge geeigneter Techniken. Die Entscheidung, welche dieser Techniken letztlich Verwendung finden, bleibt jedoch in der Hand des Nutzers.

Für ein System zur automatischen Auswahl geeigneter Visualisierungstechniken ist es naheliegend, folgenden Grundaufbau anzunehmen:

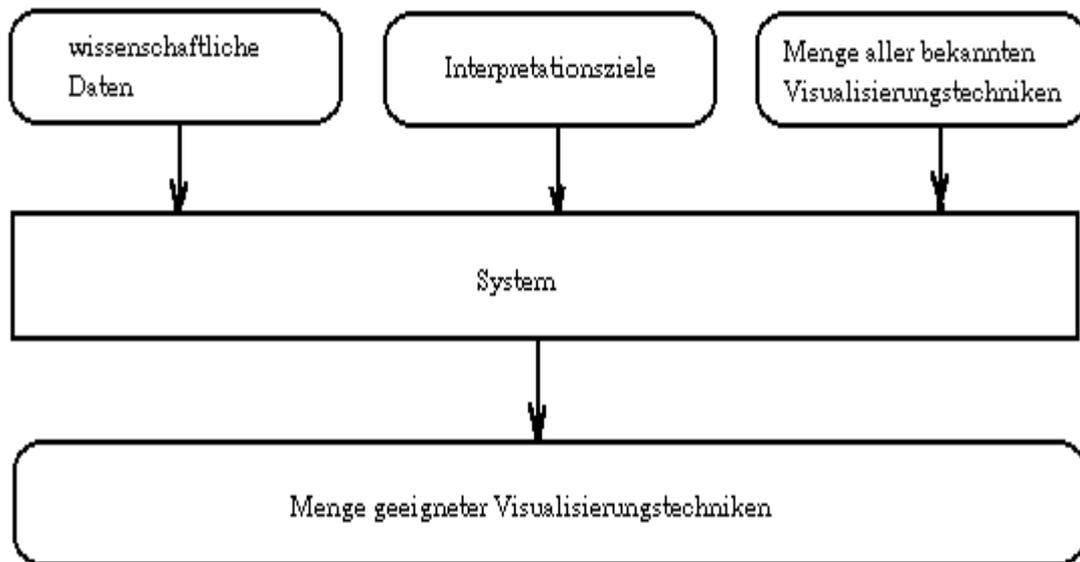


Bild 1: Grundaufbau eines Systems zur automatischen Auswahl geeigneter Visualisierungstechniken

Aus Bild 1 wird ersichtlich, daß ein System zur Auswahl geeigneter Visualisierungstechniken im wesentlichen drei wichtige Eingaben benötigt:

- den wissenschaftlichen Datensatz bzw. die Spezifikation der wichtigsten Eigenschaften von diesem
- nähere Angaben dazu, was mit der Visualisierung erreicht werden soll, welches Ziel der Nutzer mit der Visualisierung verfolgt. Dies wird festgelegt durch eine Menge von Interpretationszielen.
- die Menge aller in Frage kommenden Visualisierungstechniken, unter denen geeignete ausgewählt werden.

Weitere Einflußfaktoren sind denkbar, sollen in dieser Arbeit aber nicht weiter betrachtet werden.

Ein umfassender Ansatz zur Lösung des Problems ist bisher nicht bekannt. In der Literatur finden sich lediglich Teillösungen.

In /Wehr90/ wird die Auswahl passender Visualisierungstechniken mit Hilfe einer Matrix vorgenommen. Zeilen und Spalten der Matrix werden durch Datenklassen und Interpretationsziele gebildet. In diesen Ansatz gehen jedoch nicht alle notwendigen Parameter ein, so findet z.B. die Anzahl der Beobachtungsfälle für eine bestimmte Meßgröße keine Beachtung. Außerdem lassen sich die Einteilungen der Datenklassen und Interpretationsziele nur bedingt verfeinern, da eine zu große Matrix die Möglichkeit der Kombinierbarkeit von Matrixelementen einschränkt. Für eine allgemeine Lösung des Problems ist der Wehrend-Ansatz also wenig geeignet. Für spezielle Datenklassen (z.B. Volumendaten) kann er jedoch annehmbare Ergebnisse liefern (siehe /Schw93/).

In /Robe90/ wird eine Methodologie zur Auswahl geeigneter Visualisierungstechniken basierend auf einem natürlichen Szenenparadigma vorgestellt. Auch bei diesem Ansatz ist zweifelhaft, ob er sich auf beliebige Anwendungen übertragen läßt. Aber auch mit diesem Ansatz ergeben sich annehmbare Lösungen für Spezialfälle (siehe /Will93/).

Insgesamt läßt sich feststellen, daß für bestimmte Datenklassen durchaus akzeptable Ansätze zur Auswahl geeigneter Visualisierungstechniken existieren. Es stellt sich die Frage, ob es überhaupt notwendig ist, nach einer allgemeinen Lösung für einen beliebigen Datensatz zu

suchen. Es zeigt sich jedoch, daß in der Praxis auftretende Datensätze sich nicht immer klar bestimmten Datenklassen (Volumendaten, Strömungsdaten, gestreute Daten, multivariate Daten) zuordnen lassen. Vielmehr liegen viele Datensätze zwischen diesen eng definierten Datenklassen oder lassen sich als gemischter Datensatz aus mehreren Datenklassen auffassen. Spezielle Lösungen für solche Datensätze zu suchen, ist wenig sinnvoll. Ein allgemeiner Ansatz ist notwendig.

Von einem solchen allgemeinen Ansatz muß verlangt werden, daß

- er für spezielle Datenklassen ähnliche Ergebnisse liefert wie die bisher existierenden Speziallösungen und
- er Ergebnisse für beliebige Datensätze liefert.

Folgende Schritte sind notwendig:

- Suche nach einer einheitlichen Datenbeschreibung für beliebige wissenschaftliche Daten. Näheres dazu im Abschnitt 2.
- Schaffung eines einheitlichen Systems von Interpretationszielen für beliebige Datensätze. Dies geschieht im Abschnitt 3.
- Entwicklung eines Algorithmus zur Auswahl geeigneter Visualisierungstechniken. Mehr dazu im Abschnitt 4.

In Abschnitt 5 finden sich Tests, Beispiele und Vergleiche für den neu eingeführten Algorithmus.

2. Datenbeschreibung

Für wissenschaftliche Daten existieren verschiedene Formen der Datenbeschreibung. Für verschiedene Datenklassen werden auch verschiedene, der jeweiligen Datenklasse optimal angepaßte, Datenbeschreibungen verwendet. So werden multivariate Daten oft als Matrix, Volumendaten jedoch als dreidimensionales Array beschrieben.

Zur Schaffung eines einheitlichen Systems zur Auswahl geeigneter Visualisierungstechniken ist es jedoch notwendig, eine einheitliche Form der Datenbeschreibung zu finden. Hierfür sind zwei grundlegende Ansätze möglich:

- a) Die wichtigsten Eigenschaften des Datensatzes werden spezifiziert, nur diese gehen in den Algorithmus ein. Die Form, in der der Datensatz beschrieben ist, spielt für die Auswahl geeigneter Visualisierungstechniken keine Rolle.
- b) Der Datensatz wird in eine einheitliche Beschreibungsform gebracht und geht als solcher in das System ein.

Beide Ansätze haben Vor- und Nachteile. Ansatz b) ist i.a. speicherplatzaufwendiger als a), da die Datenbeschreibung in einer allgemeinen und nicht in einer für eine konkrete Datenklasse optimalen Form erfolgt. Bei Ansatz b) lassen sich viele Eigenschaften des Datensatzes automatisch ermitteln, wo bei a) interaktive Abfragen nötig sind. Für Ansatz b) spricht außerdem, daß spätestens zur informationstheoretischen Analyse (siehe /Thei93/) der Datensatz auf eine einheitliche Form gebracht werden muß. Beide Ansätze lassen aber in etwa gleiche Ergebnisse erwarten.

In /Berg89/ und /Brod92/ finden sich Vorschläge zu Ansatz a). /Arnd92/ vereinigt die Vorteile von /Berg89/ und /Brod92/ in sich.

Im folgenden Abschnitt 2.1 findet sich ein Vorschlag für Ansatz b). Die nachfolgenden Abschnitte beziehen sich auf diese Datenbeschreibung. Wird eine Datenspezifikation gemäß /Arnd92/ verwendet, lassen sich die nachfolgenden Abschnitte sinngemäß ebenfalls anwenden.

2.1. Datenbeschreibung als Matrix

In diesem Abschnitt soll die Beschreibung wissenschaftlicher Daten in Form einer Matrix eingeführt werden. Ursprünglich nur zur Beschreibung multivariater Daten üblich, soll diese Beschreibung auf beliebige wissenschaftliche Daten erweitert werden.

Ein wissenschaftlicher Datensatz soll aufgefaßt werden als eine Matrix M mit N Spalten und Z Zeilen. Jede Spalte entspricht einer gemessenen oder auf andere Weise gewonnenen Größe, jede Zeile entspricht einem Beobachtungsfall. Eine Spalte enthält also Z Realisierungen einer Beobachtungsgröße, eine Zeile enthält je eine Realisierung für jede Beobachtungsgröße. Das Element in der i -ten Spalte und j -ten Zeile von M gibt den Wert der i -ten Beobachtungsgröße im j -ten Beobachtungsfall an.

Die Realisierung einer Beobachtungsgröße liefert Werte eines bestimmten Datentyps. Jeder Beobachtungsgröße, also jeder Spalte von M , muß ein elementarer Datentyp zugeordnet werden. Ein solcher Datentyp kann sein:

- "integer"
- "[$i_1..i_2$]", ein Integer-Intervall, i_1 und i_2 sind Integer-Zahlen, $i_1 \leq i_2$
- "char"

- "[c₁..c₂]", ein Char-Intervall, c₁ und c₂ sind vom Typ Char, c₁ ≤ c₂
- "boolean"
- "real"
- "[r₁,r₂]", ein real-Intervall, r₁ und r₂ sind vom Typ real, r₁ ≤ r₂
- "[a₁,a₂,...,a_k]", Aufzählungstyp
- "string"

Sei typ_i der Datentyp der i-ten Spalte von M. Alle Elemente der i-ten Spalte von M sind also entweder vom Typ typ_i oder nehmen den Sonderwert "unbekannt" an. Damit sind komplexe Strukturen (Vektoren, Felder) als Elemente von M nicht zugelassen.

Einigen der N Beobachtungsgrößen kommt eine gewisse Sonderrolle zu, da in ihnen der räumliche bzw. zeitliche Bezug der Beobachtung steht. Diese Größen sind weniger als Beobachtungsgrößen direkt anzusehen, sondern geben Zeit und Ort der Beobachtung an. Zur Modellierung dieser Sonderrolle wird der Begriff des Gitters eingeführt.

Sei G eine Teilmenge von {1,...,N}. G enthalte die Nummern aller Spalten von M, in denen der räumliche bzw. zeitliche Bezug der Beobachtung angegeben ist, die also das Gitter definieren, auf dem sich die Daten befinden. Da sich die Spalten von M vertauschen lassen, gelte für |G|=n o.B.d.A. G={1,...,n}. Der Wert |G| sei auch mit dem Begriff Dimensionalität des Gitters bezeichnet.

Die räumliche bzw. zeitliche Lage der Beobachtung muß nicht zwangsläufig die Identifizierbarkeit eines Beobachtungsfalles sichern. Es kann durchaus vorkommen, daß mehrere Beobachtungsfälle vom gleichen Ort und/oder der gleichen Zeit vorliegen. Sucht man aber nach identifizierenden Größen, so ist es sicher sinnvoll, räumliche und zeitliche Lage (und, falls nötig, noch weitere Größen) hierfür zu nutzen. Mit anderen Worten: Die Menge der das Gitter definierenden Größen ist eine Teilmenge der Menge aller "unabhängigen" (siehe /Arnd93/) Größen.

In vielen Fällen stimmen die Dimensionalität des Raumes, in dem die Daten beobachtet werden, und die Dimensionalität des Gitters überein. Es ist jedoch auch möglich, daß die Dimensionalität des Gitters kleiner ist als die des Raumes.

Beispiel: Beobachtung von Größen im 3D-Raum entlang einer Linie

Für diesen Fall gibt es zwei Möglichkeiten der Behandlung.

- 1) Zusätzlich zum Gitter existiert eine Abbildung der Menge aller Gitterpunkte in den (höherdimensionalen) Raum.
- 2) Die Dimensionalität des Gitters wird künstlich erhöht, indem die Koordinaten des Beobachtungspunktes im Raum als Gitterkoordinaten betrachtet werden.

Die zweite Möglichkeit hat den Vorteil der einheitlichen Behandlung aller wissenschaftlichen Datensätze: Nach einer künstlichen Erhöhung der Dimensionalität des Gitters stimmen die Dimensionalität von Gitter und Raum stets überein. Außerdem kann durch die informationstheoretische Analyse (siehe /Thei93/) jederzeit festgestellt werden, ob eine künstliche Erhöhung der Dimensionalität des Gitters stattgefunden hat. Nachteilig ist, daß nach der Erhöhung der Dimensionalität unter Umständen die Gitterlinien sehr eng liegen und die Anzahl der Gitterpunkte stark zunimmt. Trotzdem soll im folgenden auf Möglichkeit 2 zurückgegriffen werden. Die Dimensionalität von Gitter und Raum stimmen überein, die Werte zur Beschreibung des Gitterpunktes entsprechen den Koordinaten des Raumpunktes, an dem die Beobachtung stattfindet.

Die in den Gitterpunkten beobachteten Merkmale können verschiedener Struktur (Skalar, Vektor, Tensor) sein. Zur Beobachtung eines 3D-Vektors ist die Realisierung von drei

Beobachtungsgrößen notwendig. Zur Modellierung dieser Tatsache wird eine Klasseneinteilung V der Menge $\{1, \dots, N\} \setminus G$ eingeführt.

Werden in den Gitterpunkten nur Skalare gemessen, so sind alle Elemente von V einelementig. Ein k -dimensionaler Vektor wird durch eine k -elementige Menge in V repräsentiert. Die Elemente dieser Menge sind die Spaltennummern, in denen die Komponenten des Vektors stehen. Der Wert $|V|$ entspricht der Anzahl der in einem Gitterpunkt beobachteten Merkmale.

2D- und 3D-Vektoren können in verschiedenen Koordinatensystemen gegeben sein, die alle ineinander überführt werden können. Zur weiteren Behandlung eines Datensatzes ist es notwendig, sich auf ein einheitliches und festes Koordinatensystem zu einigen. Da es sich bei den Vektoren i.a. um Richtungsvektoren handelt (die Position ist bereits durch den Gitterpunkt gegeben), erscheint die einheitliche Benutzung von Polarkoordinaten sinnvoll: Die Richtungen werden sehr gut durch die Winkel ϕ (und λ) repräsentiert, außerdem steht als weitere gut analysierbare (und koordinatensystem-unabhängige) Größe der Betrag des Vektors zur Verfügung.

Sei $\{n_1, n_2, n_3\} \in V$ mit $n_1 < n_2 < n_3$. In den Gitterpunkten wird also ein 3D-Vektor gemessen. Dann stehe o.B.d.A. in der n_1 -ten Spalte von M der Betrag dieses Vektors, in der n_2 -ten Spalte der ϕ -Winkel und in der n_3 -ten Spalte der λ -Winkel des Vektors in Polarkoordinaten (beide Winkelmaße im Bogenmaß).

Insgesamt muß zur Beschreibung eines wissenschaftlichen Datensatzes also gegeben sein:

- Z (Anzahl der Beobachtungsfälle)
- N (Anzahl der Beobachtungsgrößen)
- $\text{typ}_1, \dots, \text{typ}_N$ (Datentyp der einzelnen Beobachtungsgrößen)
- M (Matrix mit N Spalten und Z Zeilen, dabei ist jedes Element aus der i -ten Spalte vom Typ typ_i oder "unbekannt")
- G (Spalten von M, die das Gitter festlegen)
- V (Menge der in den Gitterpunkten beobachteten Merkmale)

Alle anderen Charakteristika des Datensatzes lassen sich aus diesen Werten automatisch ermitteln.

Beispiele:

1. Ein Volumendatensatz bestehe aus 41 Schnitten mit je 256×256 Punkten. Für jeden dieser 2686976 Gitterpunkte ist ein skalarer Wert zwischen 0 und 255 gegeben.

Dann gilt:

-Z=2686976

-N=4

- $\text{typ}_1 = \text{"char"}$, $\text{typ}_2 = \text{"char"}$, $\text{typ}_3 = \text{"[0..40]"}$, $\text{typ}_4 = \text{"char"}$

$$- M = \begin{pmatrix} 0 & 0 & 0 & c_1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 40 & c_{41} \\ \vdots & \vdots & \vdots & \vdots \\ 255 & 255 & 40 & c_{2686976} \end{pmatrix}$$

mit $c_1, \dots, c_{2686976}$ vom Typ "char"

-G={1,2,3}

-V={{4}}

2. Der in /Arnd92/ vorgestellte mikrobiologische Datensatz umfaßt 343 Beobachtungsfälle von 41 Merkmalen.

Hierfür gilt:

-Z=343

-N=41

-typ₁="[282..50000]", typ₂="[19881129..19890929]", typ₃="[0..274]", ...
 -G=∅

-V={{1},{2},...,{41}}

Dieser Datensatz kann auch aufgefaßt werden als G={1} und V={{2},...,{41}}. Es liegt im Ermessen des Nutzers, der ersten Beobachtungsgröße, der Katalognummer, eine besondere Bedeutung zu geben oder nicht.

3. Auf einem 2D-scattered-data-record wurden an 100 Meßpunkten je ein real-Wert zwischen 0 und 1 aufgenommen. Hier gilt:

-Z=100

-N=3

-typ₁="real", typ₂="real", typ₃="[0,1]"

-M= $\begin{pmatrix} x_1 & y_1 & r_1 \\ \vdots & \vdots & \vdots \\ x_{100} & y_{100} & r_{100} \end{pmatrix}$

mit x₁,...,x₁₀₀,y₁,...,y₁₀₀ vom Typ "real" und r₁,...,r₁₀₀ vom Typ "[0,1]"

-G={1,2}

-V={{3}}

Ein scattered-data-record wird also als ein Datensatz auf regulärem Gitter betrachtet, wobei die Gitterpunkte quasi-unendlich dicht liegen.

4. Auf einem 3D-scattered-data-record wurden an 200 Meßpunkten je ein 3D-Vektor und ein binärer Wert gemessen:

-Z=200

-N=7

-typ₁="real", typ₂="real", typ₃="real", typ₄="real", typ₅="[-π/2,π/2]", typ₆="[-π,π]", typ₇="boolean"

-G={1,2,3}

-V={{4,5,6},{7}}

Aus den bisher eingeführten Datencharakteristika können weitere Charakteristika automatisch berechnet werden. Diese sind:

-Die Umfänge der Wertebereiche der einzelnen Beobachtungsgrößen:

Diese Werte ergeben sich direkt aus den Typen der Wertebereiche. wb_i sei der Umfang des Wertebereichs der i-ten Beobachtungsgröße. Es gilt:

wb_i:= $\begin{cases} 65536 & \text{für typ}_i=\text{"integer"} \\ i_2-i_1 & \text{für typ}_i=\text{"[i}_1\text{..i}_2\text{]"} \\ 256 & \text{für typ}_i=\text{"char"} \\ c_2-c_1 & \text{für typ}_i=\text{"[c}_1\text{..c}_2\text{]"} \\ 2 & \text{für typ}_i=\text{"boolean"} \\ \text{"cont"} & \text{für typ}_i=\text{"[r}_1\text{,r}_2\text{]"} \text{ und } r_1 < r_2 \\ k & \text{für typ}_i=\text{"[a}_1\text{,...,a}_k\text{]"} \end{cases}$

Falls typ_i="string", so ergibt sich wb_i aus der Anzahl der verschiedenen Zeichenketten, die in der i-ten Spalte von M auftreten. Das Sonderzeichen "cont" steht für einen quasi-unendlichen Wertebereich.

-Die Qualität der Datenmenge:

Es ist möglich, daß Datenwerte an einem Gitterpunkt teilweise oder vollständig unbekannt sind. Sind die Werte an einem Gitterpunkt vollständig unbekannt, so taucht ein zu diesem Gitterpunkt gehörender Beobachtungsfall nicht in der Matrix auf. Ein Datensatz ist vollständig, wenn für jeden Gitterpunkt ein Beobachtungsfall existiert und kein Element der Matrix den Wert "unbekannt" hat.

Weiterhin ist es möglich, daß für einen Gitterpunkt mehrere Beobachtungsfälle existieren. Ein Datensatz ist überdefiniert, wenn es einen Gitterpunkt gibt, für den mehr als ein Beobachtungsfall mit nicht vollständig unbekanntem Datenwerten existiert.

Ist ein Datensatz vollständig und nicht überdefiniert, so stimmen die Anzahl der Gitterpunkte mit der Anzahl der Zeilen der Matrix M überein.

Beispiele:

- N=4, Z=4,

typ₁=[0..1], typ₂=[0..1], typ₃="char", typ₄="integer"

G={1,2}, V={{3},{4}}

$$M = \begin{pmatrix} 0 & 0 & 10 & 500 \\ 0 & 1 & \text{unbekannt} & 2 \\ 1 & 0 & 5 & 5 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

unvollständig und
nicht überdefiniert

- N=4, Z=3,

typ₁=[0..1], typ₂=[0..1], typ₃="char", typ₄="integer"

G={1,2}, V={{3},{4}}

$$M = \begin{pmatrix} 0 & 0 & 10 & 500 \\ 0 & 1 & \text{unbekannt} & 2 \\ 0 & 1 & 2 & 5 \end{pmatrix}$$

unvollständig und
überdefiniert

- N=4, Z=5,

typ₁=[0..1], typ₂=[0..1], typ₃="char", typ₄="integer"

G={1,2}, V={{3},{4}}

$$M = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 10 & 100 \\ 1 & 0 & 15 & 20 \\ 1 & 1 & 13 & 13 \\ 1 & 1 & 5 & 5000 \end{pmatrix}$$

vollständig und
überdefiniert

- N=4, Z=4,

typ₁=[0..1], typ₂=[0..1], typ₃="char", typ₄="integer"

G={1,2}, V={{3},{4}}

$$M = \begin{pmatrix} 0 & 0 & 1 & 100 \\ 0 & 1 & 15 & 3000 \\ 1 & 0 & 1 & -15 \\ 1 & 1 & 5 & 0 \end{pmatrix}$$

vollständig und
nicht überdefiniert

-Datensatz 1. (s.oben): vollständig und nicht überdefiniert

-Datensatz 2. (s.oben): vollständig und überdefiniert für die Interpretation G=∅ und V={{1},...,{41}}.
unvollständig und nicht überdefiniert für G={1} und V={{2},...,{41}}.

-Datensatz 3. und 4. (s.oben): unvollständig (wie alle scattered-data-records)

Es sei $\mathfrak{M}_{n,p}$ die Menge aller denkbaren wissenschaftlichen Datensätze, für die $|G|=n$ und $|V|=p$ gilt. Ein (n,p)-Datensatz $M \in \mathfrak{M}_{n,p}$ sei also ein Datensatz, bei dem in einer n-dimensionalen Gitterpunktmenge p verschiedene Merkmale beobachtet werden können.

Bemerkung:

Die einheitliche Form der Datenbeschreibung als Matrix hat für einige Datensätze den Nachteil erhöhter Speicherplatzanforderungen. Für solche Fälle kann es sinnvoll sein, die physische Abspeicherung des Datensatzes nicht in Matrixform vorzunehmen. Voraussetzung hierfür ist jedoch, daß eine schnelle automatische Umrechnung der physischen Speicherung auf das Matrixformat möglich ist. So ist es z.B. sinnvoll, einen Volumendatensatz als dreidimensionales Array zu verwalten. Jedes Element der gedachten Matrix kann dabei schnell aus dem Array gewonnen werden, so daß eine physische Abspeicherung des Datensatzes als Matrix nicht nötig ist.

3. Interpretationsziele

Mit der Visualisierung eines wissenschaftlichen Datensatzes kann der Nutzer verschiedenste Ziele verfolgen. Diese verschiedenen Ziele und Wünsche des Nutzers sind mitentscheidend dafür, welche Techniken für einen gegebenen Datensatz günstig sind. Zur Formulierung der Wünsche des Nutzers wird der Begriff des Interpretationsziels eingeführt. Ein Interpretationsziel beschreibt den genauen Nutzerwunsch für einen konkreten Aspekt der Visualisierung. Der gesamte Nutzerwunsch für die Visualisierung eines Datensatzes ergibt sich dann als ein System mehrerer Interpretationsziele.

Bisherige Untersuchungen zu Interpretationszielen ergaben zwei wichtige Probleme:

- 1) Die Beschreibung von Interpretationszielen kann in verschiedenen Abstraktionsstufen erfolgen. Der Nutzer sollte seine Wünsche in einer möglichst niedrigen Abstraktionsstufe, also möglichst mit Hilfe der natürlichen Sprache, formulieren können. Für ein System zur automatischen Auswahl geeigneter Visualisierungstechniken ist es jedoch notwendig, Interpretationsziele auf möglichst abstraktem Niveau (beispielsweise mit den Mitteln von Mengenlehre und Logik) zu formulieren. Es ergibt sich ein nur unvollständig zu lösender Grundwiderspruch.
- 2) Bei bisherigen Untersuchungen zu Interpretationszielen ergaben sich für verschiedene Datenklassen auch völlig verschiedene (verbal formulierte) Interpretationsziele. Beispielsweise haben die in /Schw93/ für Volumendaten aufgestellten Interpretationsziele fast keinen Zusammenhang zu den in /Arnd91/ eingeführten Interpretationszielen für einen mikrobiologischen Datensatz.

Zur Behandlung von Interpretationszielen soll deshalb folgender in Bild 2 dargestellter Ansatz gewählt werden:

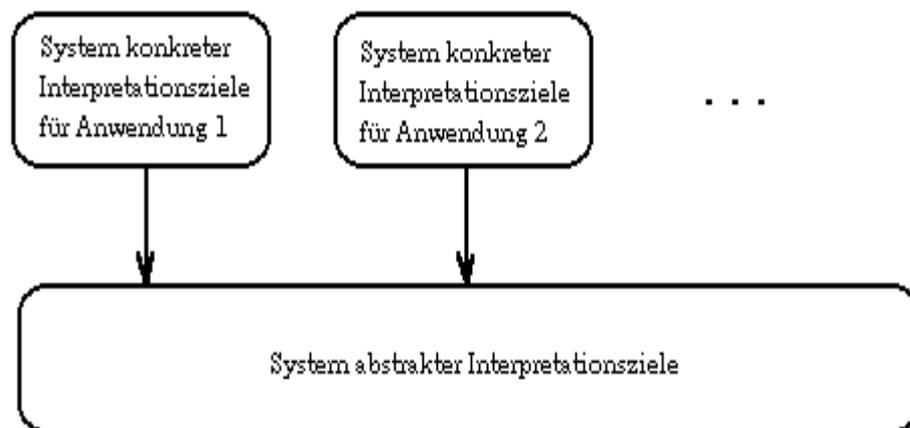


Bild 2: Ansatz zur Behandlung von Interpretationszielen

Bild 2 zeigt, daß ein allgemeines System abstrakter Interpretationsziele zu schaffen ist. Die konkreten Interpretationsziele für die verschiedenen Anwendungen müssen auf das System

abstrakter Interpretationsziele abgebildet werden. Diese Abbildung muß für jedes Anwendungsgebiet neu bestimmt werden.

In Abschnitt 3.1 wird ein System abstrakter Interpretationsziele eingeführt. In Abschnitt 3.2 befinden sich Überlegungen und Beispiele für die Abbildung konkreter Interpretationsziele in dieses System abstrakter Interpretationsziele.

3.1 Ein System abstrakter Interpretationsziele

Zur Einführung eines einheitlichen Systems abstrakter Interpretationsziele soll zunächst die Beschreibung des zu visualisierenden Datensatzes in einer neuen Form erfolgen: mit den Mitteln von Mengenlehre und Prädikatenlogik. Diese Form der Datenbeschreibung hat keine praktische Bedeutung, sie liefert lediglich das theoretische Modell, aus dem das System abstrakter Interpretationsziele hergeleitet werden soll.

Es sei G die Menge aller Gitterpunkte. Mer sei die Menge aller Merkmale, die in einem Gitterpunkt beobachtet werden können. Val sei die Menge aller Datenwerte, die ein beliebiges Merkmal an einem beliebigen Gitterpunkt annehmen kann. G , Val und Mer seien nichtleer. Ein wissenschaftlicher Datensatz D wird dann aufgefaßt als eine Teilmenge von $G \times Val \times Mer$. Dabei gilt $(g,v,m) \in D$ genau dann, wenn im Gitterpunkt g für das Merkmal m der Wert v beobachtet wurde. n sei die Dimensionalität der Raumes, in dem das Gitter liegt. Dieser Raum sei nachfolgend mit dem Begriff "Gitterraum" bezeichnet. $p := |Mer|$ sei die Anzahl der in einem Gitterpunkt zu beobachtenden Merkmale.

Beispiel:

Der Datensatz

- $N=4, Z=5$
- $typ_1=[0..3], typ_2=[0..3] typ_3="char", typ_4="integer"$
- $G=\{1,2\}, V=\{\{3\},\{4\}\}$
- $M = \begin{pmatrix} 0 & 0 & 1 & 15 \\ 0 & 1 & 1 & 15 \\ 0 & 1 & 5 & \text{unbekannt} \\ 2 & 2 & 1 & 1 \\ 3 & 3 & 0 & 0 \end{pmatrix}$

wird mengentheoretisch beschrieben als:

- $G = \{(0,0), \dots, (0,3), (1,0), \dots, (3,3)\}$ mit $|G|=16$
- $Val = "integer" \cup \{\text{unbekannt}\}$
- $Mer = \{\{3\}, \{4\}\}$
- $D = \{((0,0), 1, \{3\}), ((0,0), 15, \{4\}), ((0,1), 1, \{3\}), ((0,1), 15, \{4\}), ((0,1), 5, \{3\}), ((0,1), \text{"unbekannt"}, \{4\}), ((2,2), 1, \{3\}), ((2,2), 1, \{4\}), ((3,3), 0, \{3\}), ((3,3), 0, \{4\})\}$
- $n=2$
- $p=2$

Bemerkung:

Streng genommen ist diese mengentheoretische Beschreibung eines Datensatzes weder exakt noch vollständig. Wird ein Datenwert in einem Gitterpunkt und einem Merkmal mehrfach gemessen, so müßten in D zwei identische Elemente stehen. Außerdem geht aus D nicht hervor, welche gemessenen Datenwerte für einen Gitterpunkt und verschiedene Merkmale zu einem Beobachtungsfall gehören. Diese Probleme sollen hier jedoch nicht weiter beachtet werden, da diese Form der Datenbeschreibung lediglich theoretisch und zum Zwecke der Beschreibung abstrakter Interpretationsziele ist.

Aus den bisher eingeführten Größen zur Beschreibung eines Datensatzes ergeben sich folgende weitere Größen:

- \mathfrak{X}_i : die Menge aller i -dimensionalen Schnitte von G ($i=0, \dots, n$). Ein Element von \mathfrak{X}_i ist also eine Teilmenge von G und enthält alle Gitterpunkte, die sich in einem i -dimensionalen Unterraum des Gitterraums (oder in einer gewissen Umgebung dieses

- Raumes) befinden. Ein Element von \mathfrak{R}_2 ist also eine Menge von Gitterpunkten, die alle in einer Ebene liegen. Weiterhin gilt $\mathfrak{R}_n = \{\text{Git}\}$ und $\mathfrak{R}_0 = \{\{g_1\}, \dots, \{g_{|\text{Git}|}\}\}$ für $\text{Git} = \{g_1, \dots, g_{|\text{Git}|}\}$.
- $\text{Mer}_i := \{X : X \subseteq \text{Mer} \wedge |X| = i\}$ ($i=1, \dots, p$)
sei die Menge aller i -elementigen Merkmalsklassen.
 - $m : \wp(\text{Git}) \times \text{Mer} \rightarrow \text{Val}$
beschreibt eine Mittelwertbildung. Für eine Menge von Gitterpunkten und ein Merkmal wird der Mittelwert aller in diesen Gitterpunkten in diesem Merkmal stehenden Datenwerte bestimmt.
 - $ex : \wp(\text{Git}) \times \text{Mer} \rightarrow \text{Val}$
beschreibt eine Extremwertbildung. Für eine Menge von Gitterpunkten und ein Merkmal wird das Maximum (Minimum) aller in diesen Gitterpunkten in diesem Merkmal stehenden Werte bestimmt.
 - $h_i : \wp(\text{Git}) \times \text{Mer}^i \times \text{Val}^i \rightarrow [0,1]$ ($i=1, \dots, |\text{Mer}|$)
beschreibt die relative Häufigkeit des Auftretens von Wertekombinationen. Mer^i und Val^i seien das i -fache Kreuzprodukt von Mer bzw. Val .
 $h_i(\{g_1, \dots, g_k\}, m_1, \dots, m_i, v_1, \dots, v_i)$ gibt die relative Häufigkeit an, mit der in den Gitterpunkten g_1, \dots, g_k die Wertekombination v_1, \dots, v_i in den Merkmalen m_1, \dots, m_i auftritt. Dieser Wert ist nur sinnvoll, wenn m_1, \dots, m_i paarweise verschieden sind.
 - $unbek : \text{Git} \times \text{Mer} \rightarrow \text{"boolean"}$
gibt an, ob in einem bestimmten Gitterpunkt für ein bestimmtes Merkmal kein Wert beobachtet wurde, ein Wert an dieser Stelle also unbekannt ist.
 - $int : \text{Git} \times \text{Mer} \rightarrow \text{Val}$
gibt einen aus den Nachbarpunkten interpolierten Wert für einen Gitterpunkt und ein Merkmal an. Hier müßte genauer unterschieden werden, welche Interpolationstechnik verwendet wird und welche Gitterpunkte unter den Begriff "Nachbarpunkte" zählen. Mit Hilfe dieser Unterscheidung ließen sich dann z.B. die verschiedenen scattered-data-Techniken charakterisieren. Dies soll hier jedoch ausgespart werden, da für den Spezialfall scattered data noch nicht genügend Kenntnisse über die Auswahl geeigneter Visualisierungstechniken vorliegen.

Allen diesen Größen ist gemeinsam, daß sie sich automatisch aus dem in Matrixform beschriebenen Datensatz ermitteln lassen.

Bei der Anwendung einer Visualisierungstechnik werden alle Merkmale eines Gitterpunktes in einen bestimmten Bildschirmbereich abgebildet. Bildpunkte außerhalb dieses Bildschirmbereiches werden -egal, was im Gitterpunkt beobachtet wird- nicht von diesem Gitterpunkt beeinflusst. Der Bildbereich für einen Gitterpunkt kann im Extremfall ein Pixel oder der gesamte Bildschirm sein. Die Bildbereiche für zwei Gitterpunkte können identisch, disjunkt oder teilweise überlappend sein. Eine Menge von Gitterpunkten heißt bildbereichsdisjunkt, wenn für zwei beliebige Gitterpunkte aus dieser Menge die Bildbereiche disjunkt sind. Diese Eigenschaft soll durch das Prädikat $bdis(G)$ mit $G \subseteq \text{Git}$ ausgedrückt werden.

Eine Eigenschaft des Datensatzes sei *am* Punkt g sichtbar, wenn in der Visualisierung diese Eigenschaft nicht außerhalb des Bildbereichs von g dargestellt wird.

Die Eigenschaften einer Visualisierung lassen sich mit folgenden Prädikaten beschreiben:

- $vis(g, v, m)$ ($g \in \text{Git}, v \in \text{Val}, m \in \text{Mer}$)
In der Visualisierung ist sichtbar, daß im Gitterpunkt g im Merkmal m der Wert v steht.
- $vis(g, h_i(S_j, m_1, \dots, m_i, v_1, \dots, v_i))$ ($g \in \text{Git}, S_j \in \mathfrak{R}_j, m_1, \dots, m_i \in \text{Mer}, v_1, \dots, v_i \in \text{Val}, m_1, \dots, m_i$ paarweise verschieden)

In der Visualisierung ist am Punkt g die relative Häufigkeit des Auftretens der Wertekombination v_1, \dots, v_i in den Merkmalen m_1, \dots, m_i , ermittelt über alle Gitterpunkte des Schnittes S_j , sichtbar.

$$- \text{kvis}(g, S_i) \quad (g \in \text{Git}, S_i \in \mathfrak{R}_i)$$

Alle vorhandenen Korrelationen der Merkmale untereinander bzgl. S_i sind in der Visualisierung am Punkt g sichtbar. "Bzgl. S_i " bedeutet dabei, daß zur Bestimmung und Visualisierung nur diejenigen Beobachtungsfälle genutzt werden, die in einem Gitterpunkt aus S_i liegen.

$$- \text{kvis}(g, S_i, m) \quad (g \in \text{Git}, S_i \in \mathfrak{R}_i, m \in \text{Mer})$$

Am Punkt g sind die Korrelationen des Merkmals m zur Lage im Gitter erkennbar. Dabei werden nur diejenigen Beobachtungsfälle betrachtet, die auf einem Gitterpunkt aus S_i liegen.

$$- \text{zw}(S_i, v, m) \quad (S_i \in \mathfrak{R}_i, v \in \text{Val}, m \in \text{Mer})$$

In der Visualisierung sind alle Punkte zwischen den Gitterpunkten aus S_i sichtbar, in denen sich durch Interpolation aus den Werten im Merkmal m der Wert v ergibt. "Alle Punkte" sind hierbei alle Punkte des quasi-kontinuierlichen Raumes, die in dem durch S_i definierten Unterraum liegen.

Mit diesen eingeführten Größen und Prädikaten sollen nun Interpretationsziele beschrieben werden.

Es ist naheliegend, den aussagenlogischen Ausdruck $(g, v, m) \in D \Rightarrow \text{vis}(g, v, m)$ erfüllen zu wollen: Wird im Gitterpunkt g im Merkmal m der Wert v gemessen, so sollte dies in der Visualisierung sichtbar sein. Werden vor diesen Ausdruck logische Quantifikatoren gesetzt, so ergeben sich die ersten abstrakten Interpretationsziele:

$$(1) \quad \forall g \in \text{Git} \forall v \in \text{Val} \forall m \in \text{Mer} ((g, v, m) \in D \Rightarrow \text{vis}(g, v, m))$$

Das universelle Identifikations- und Lokalisationsproblem. Jeder Wert an jeder Stelle des Datensatzes ist im Bild erkennbar. Jeder Datenwert ist identifizierbar und lokalisierbar.

$$(2) \quad \exists v \in \text{Val} \forall g \in \text{Git} \forall m \in \text{Mer} ((g, v, m) \in D \Rightarrow \text{vis}(g, v, m))$$

Das universelle Lokalisationsproblem. Ein bestimmter Wert wird ausgewählt, in der für diesen Wert ausgewählten Visualisierung sind alle Gitterpunkte und alle Merkmale, in denen dieser Wert auftritt, sichtbar.

$$(3) \quad \exists g \in \text{Git} \forall v \in \text{Val} \forall m \in \text{Mer} ((g, v, m) \in D \Rightarrow \text{vis}(g, v, m))$$

Das universelle Identifikationsproblem. In einem bestimmten Gitterpunkt sind alle Werte in allen Merkmalen erkennbar.

$$(4) \quad \exists m \in \text{Mer} \forall g \in \text{Git} \forall v \in \text{Val} ((g, v, m) \in D \Rightarrow \text{vis}(g, v, m))$$

Das Identifikations- und Lokalisationsproblem, beschränkt auf ein Merkmal. Für ein bestimmtes Merkmal sind die Werte an allen Gitterpunkten identifizierbar, ebenso sind für einen beliebigen Wert alle Gitterpunkte lokalisierbar, in denen dieser Wert im Merkmal m auftritt.

$$(5) \quad \exists g \in \text{Git} \exists v \in \text{Val} \forall m \in \text{Mer} ((g, v, m) \in D \Rightarrow \text{vis}(g, v, m))$$

Für einen bestimmten Gitterpunkt ist ein bestimmter Wert in allen Merkmalen, in denen er auftritt, sichtbar.

$$(6) \quad \exists v \in \text{Val} \exists m \in \text{Mer} \forall g \in \text{Git} ((g,v,m) \in D \Rightarrow \text{vis}(g,v,m))$$

Das spezielle Lokalisationsproblem. Für einen bestimmten Wert und ein bestimmtes Merkmal sind alle Gitterpunkte lokalisierbar, in denen dieser Wert in diesem Merkmal steht.

$$(7) \quad \exists g \in \text{Git} \exists m \in \text{Mer} \forall v \in \text{Val} ((g,v,m) \in D \Rightarrow \text{vis}(g,v,m))$$

Das spezielle Identifikationsproblem. Für einen bestimmten Gitterpunkt und ein bestimmtes Merkmal sind alle dort stehenden Werte sichtbar.

$$(8) \quad \exists g \in \text{Git} \exists v \in \text{Val} \exists m \in \text{Mer} ((g,v,m) \in D \Rightarrow \text{vis}(g,v,m))$$

Für einen speziellen Gitterpunkt, ein spezielles Merkmal und einen speziellen Wert ist in der Visualisierung sichtbar, ob dieser Wert in diesem Merkmal in diesem Gitterpunkt vorkommt.

Zwischen den bisher eingeführten 8 Interpretationszielen existieren Abhängigkeiten. Erfüllt eine Technik etwa das universelle Identifikationsproblem (3), so erfüllt sie auch das spezielle Identifikationsproblem (7). Erfüllt eine Technik das Interpretationsziel (1), so erfüllt sie auch automatisch alle anderen bisher eingeführten Interpretationsziele. Die Abhängigkeiten zwischen den Interpretationszielen (1)-(8) werden im Bild 3 dargestellt:

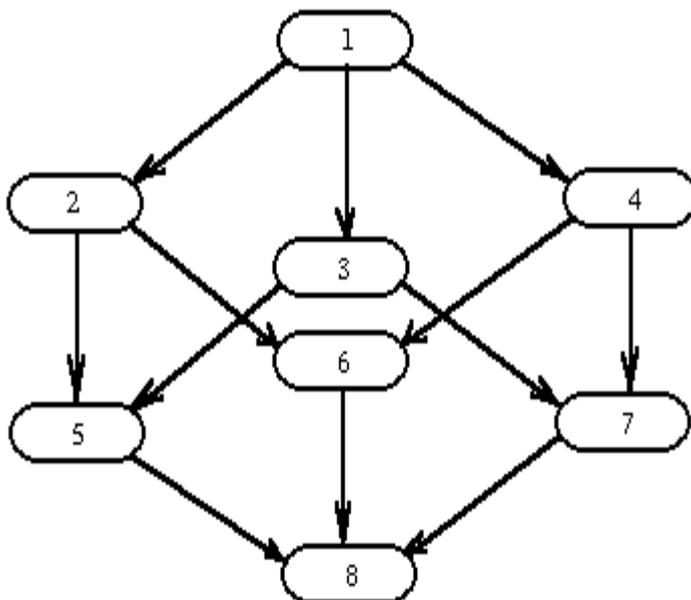


Bild 3: Abhängigkeiten zwischen den Interpretationszielen (1)-(8)

In Bild 3 bedeutet ein Pfeil von (i) nach (j), daß das Interpretationsziel (i) das Interpretationsziel (j) mit einschließt. (j) ist eine Abschwächung von (i). Kann eine Technik das Ziel (i) erfüllen, so erfüllt sie automatisch auch das Ziel (j). Das Interpretationsziel (j) ist aus (i) direkt ableitbar.

Die bisher eingeführten Interpretationsziele stellen nur eine erste grobe Näherung dar, da bei Gitterpunkten, Merkmalen und Werten entweder alle oder nur ein bestimmtes betrachtet wurden. Eine feinere Unterteilung bei den Gitterpunkten und den Merkmalen ist notwendig. Ebenso erscheint eine feinere Unterteilung der Datenwerte auf den ersten Blick sinnvoll, um beispielsweise Interpretationsziele wie "Lokalisieren aller Punkte, in denen ein Datenwert zwischen 0.5 und 0.6 beobachtet wurde" zu beschreiben. Solche Interpretationsziele sind jedoch mit den Zielen (1)-(8) und einer Klasseneinteilung auf Val beschreibbar. Deshalb soll bei den Datenwerten die Betrachtung entweder "von einem" oder "von allen" genügen.

Eine feinere Einteilung bei den Gitterpunkten ist nötig, um Untersuchungen auf Schnitten zu beschreiben. Dazu werden folgende Interpretationsziele eingeführt:

$$(9.i) \quad \exists S_i \in \mathfrak{R}_i \quad \forall g \in S_i \quad \forall v \in \text{Val} \quad \forall m \in \text{Mer} \quad ((g,v,m) \in D \Rightarrow \text{vis}(g,v,m))$$

$i=1, \dots, n-1$

Das universelle Identifikations- und Lokalisationsproblem, beschränkt auf einen i-dimensionalen Schnitt durch den Gitterraum. Es ergeben sich (n-1) neue Interpretationsziele, abhängig von der Dimensionalität n des Gitters. Das fiktive Interpretationsziel (9.0) ist dabei identisch zu (3), das fiktive Interpretationsziel (9.n) ist identisch zu (1).

$$(10.i) \quad \exists v \in \text{Val} \quad \exists S_i \in \mathfrak{R}_i \quad \forall g \in S_i \quad \forall m \in \text{Mer} \quad ((g,v,m) \in D \Rightarrow \text{vis}(g,v,m))$$

$i=1, \dots, n-1$

Das Lokalisationsproblem, beschränkt auf einen i-dimensionalen Schnitt durch den Gitterraum. In einem i-dimensionalen Unterraum des Gitterraums ist für einen bestimmten Wert jeder Gitterpunkt und jedes Merkmal, in welchem dieser Wert auftritt, erkennbar. Es ergeben sich wiederum die (n-1) neuen Interpretationsziele (10.1), ..., (10.n-1). Das fiktive Interpretationsziel (10.0) entspricht dem Interpretationsziel (5), das fiktive Interpretationsziel (10.n) entspricht (2).

$$(11.i) \quad \exists m \in \text{Mer} \quad \exists S_i \in \mathfrak{R}_i \quad \forall g \in S_i \quad \forall v \in \text{Val} \quad ((g,v,m) \in D \Rightarrow \text{vis}(g,v,m))$$

$i=1, \dots, n-1$

In einem i-dimensionalen Schnitt durch den Gitterraum sind alle in einem bestimmten Merkmal stehenden Datenwerte erkennbar. (11.0) entspricht (7), (11.n) entspricht (4).

$$(12.i) \quad \exists v \in \text{Val} \quad \exists m \in \text{Mer} \quad \exists S_i \in \mathfrak{R}_i \quad \forall g \in S_i \quad ((g,v,m) \in D \Rightarrow \text{vis}(g,v,m))$$

$i=1, \dots, n-1$

Ein bestimmter Datenwert in einem bestimmten Merkmal ist in einem bestimmten i-dimensionalen Schnitt in allen Gitterpunkten lokalisierbar. (12.0) entspricht (8), (12.n) entspricht (6).

Zwischen den neu eingeführten Interpretationszielen existieren verschiedene Abhängigkeiten, welche durch folgendes Bild veranschaulicht werden:

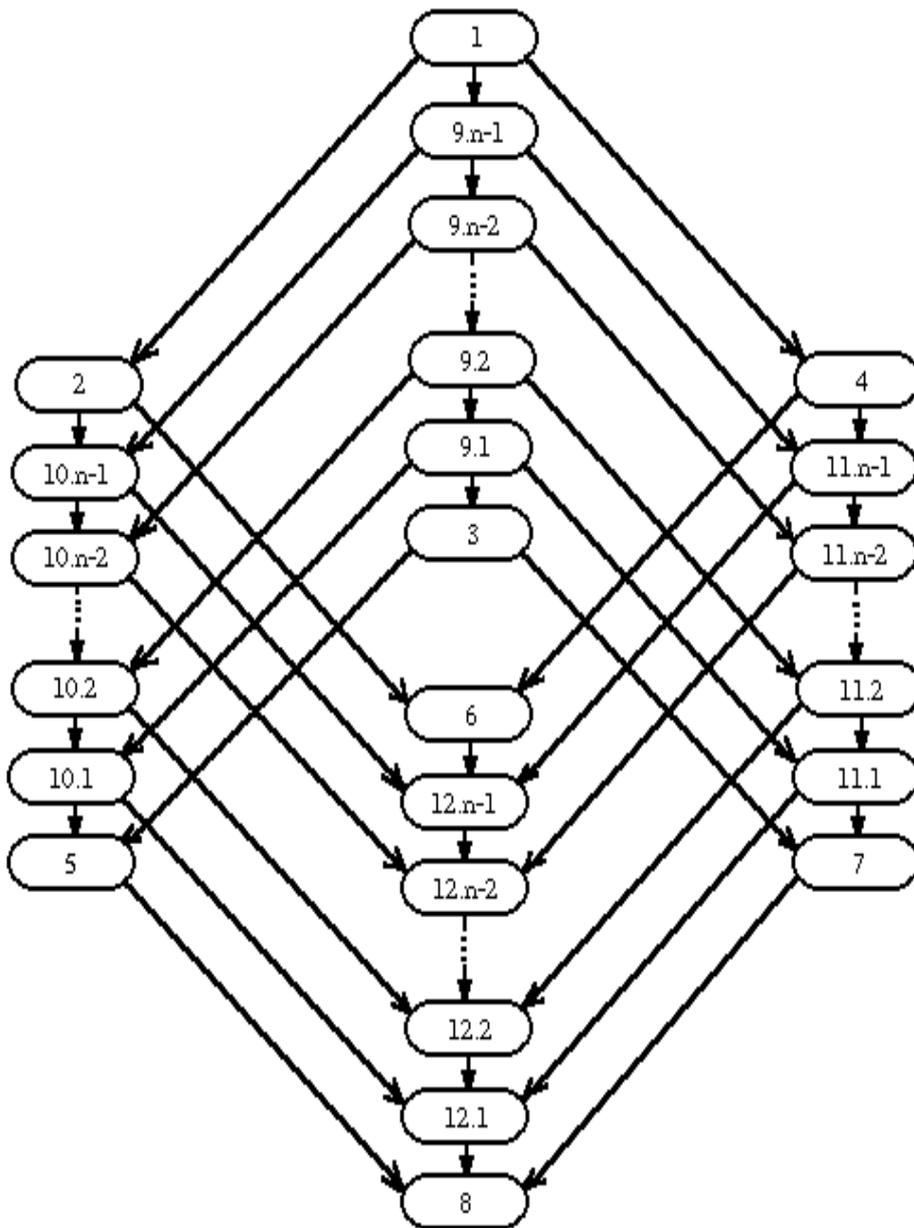


Bild 4: Abhängigkeiten der Interpretationsziele (1)-(8),(9.i)-(12.i)

Den Pfeilen in Bild 4 kommt eine analoge Bedeutung zu wie in Bild 3.

Bei der Formulierung von Interpretationszielen werden oft bestimmte Merkmalsklassen betrachtet. Das bisher eingeführte System von Interpretationszielen, das entweder alle oder nur ein Merkmal unterscheidet, soll mit den folgenden neuen Interpretationszielen verfeinert werden.

$$(13.i.j) \quad \exists S_i \in \mathfrak{R}_i, \exists mk \in Mer_j, \forall g \in S_i, \forall m \in mk, \forall v \in Val, ((g,v,m) \in D \Rightarrow vis(g,v,m))$$

$i=0, \dots, n$
 $j=2, \dots, p-1$

Für einen i -dimensionalen Schnitt und eine j -dimensionale Merkmalskombination ist jeder Datenwert, der in diesem Schnitt in einem Merkmal aus dieser Merkmalskombination beobachtet wurde, in der Visualisierung sichtbar. Das fiktive Interpretationsziel (13.i.p) entspricht (9.i) für $i=1, \dots, n-1$. Das fiktive Interpretationsziel (13.i.1) entspricht (11.i) für

$i=1,\dots,n-1$. (13.n.p) entspricht (1), (13.0.p) entspricht (3), (13.n.1) entspricht (4), (13.0.1) entspricht (7).

$$(14.i.j) \quad \exists v \in \text{Val} \exists S_i \in \mathfrak{R}_i \exists m_k \in \text{Mer}_j \forall g \in S_i \forall m \in m_k ((g,v,m) \in D \Rightarrow \text{vis}(g,v,m))$$

$i=0,\dots,n$
 $j=2,\dots,p-1$

Ein bestimmter Datenwert ist in einem i -dimensionalen Schnitt und einer j -elementigen Merkmalskombination an jeder Stelle seines Auftretens erkennbar. Das fiktive Interpretationsziel (14.i.p) entspricht (10.i) für $i=1,\dots,n-1$. Das fiktive Interpretationsziel (14.i.1) entspricht (12.i) für $i=1,\dots,n-1$. (14.n.p) entspricht (2), (14.0.p) entspricht (5), (14.n.1) entspricht (6), (14.0.1) entspricht (8).

Die Abhängigkeiten für die neu eingeführten Interpretationsziele werden durch die Bilder 5-7 beschrieben.

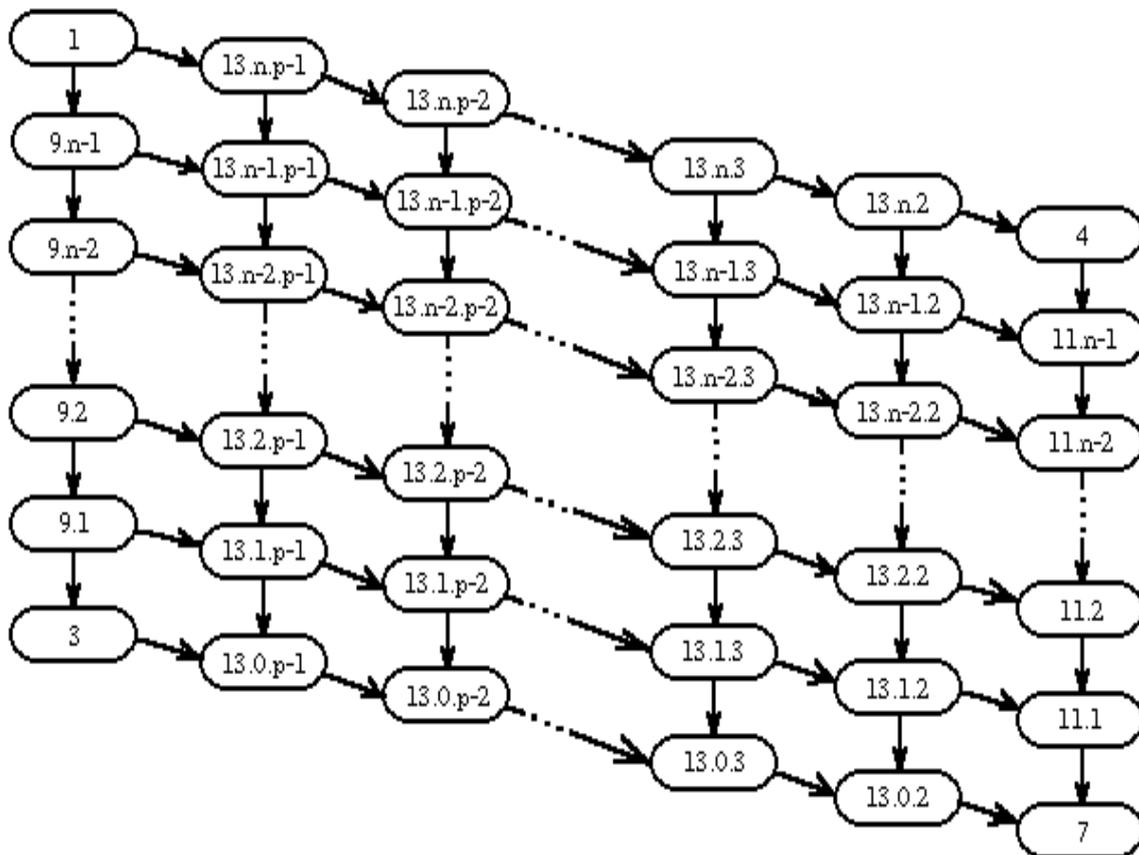


Bild 5: Abhängigkeiten der Interpretationsziele (1), (3), (4), (7), (9.i), (11.i), (13.i.j)

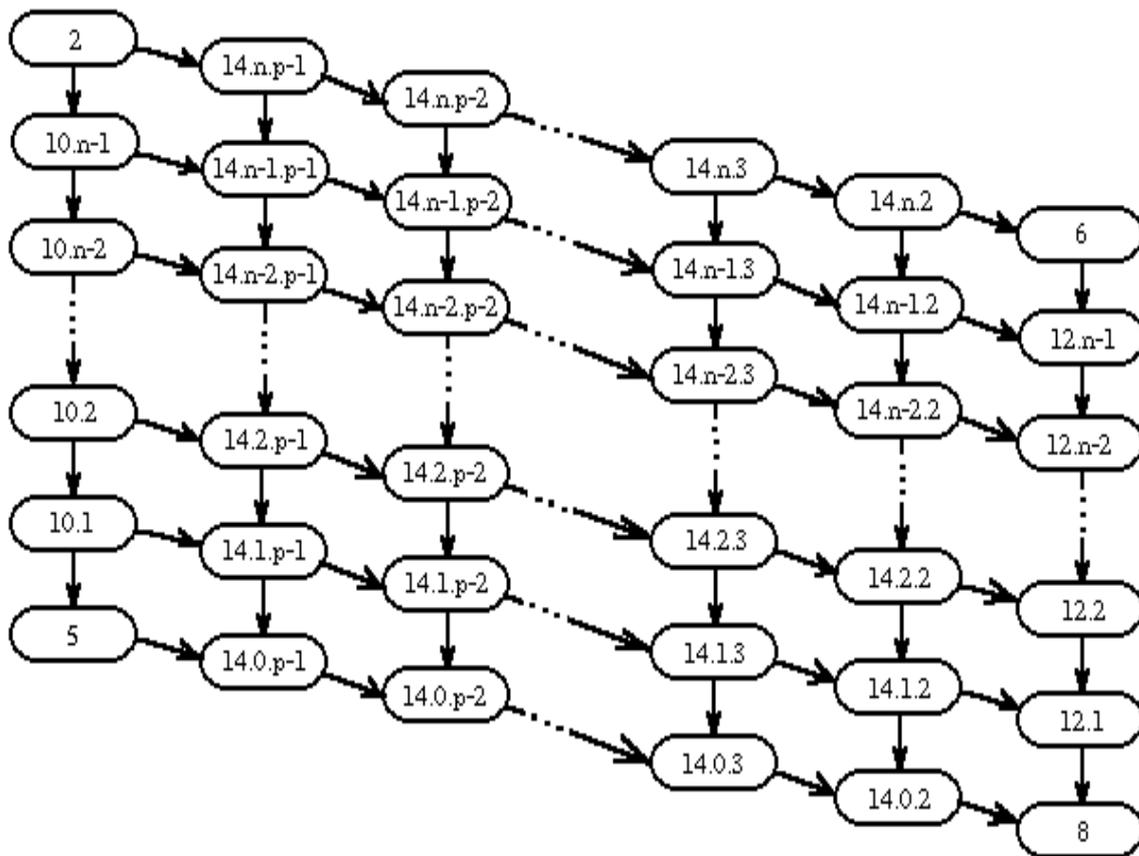


Bild 6: Abhängigkeiten der Interpretationsziele (2), (5), (6), (8), (10.i), (12.i), (14.i.j)

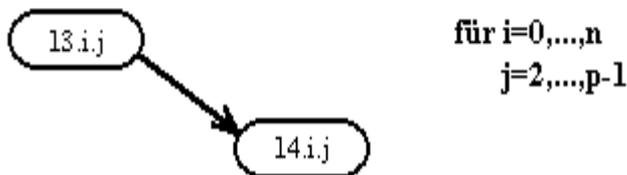


Bild 7: Abhängigkeiten der Interpretationsziele (13.i.j), (14.i.j)

Das Gesamtsystem der Abhängigkeiten zwischen den bisher eingeführten Interpretationszielen ergibt sich als Join-Verbindung der Graphen aus den Bildern 3-7.

Bemerkung:

Werden die Interpretationsziele (1)-(8) und (9.i)-(12.i) als Ausartungsfälle der Interpretationsziele (13.i.j) und (14.i.j) betrachtet, so lassen sich sämtliche bisher eingeführte Abhängigkeiten beschreiben durch:

- (13.i.j) → (13.i-1.j) für $i=1, \dots, n, j=1, \dots, p$
- (14.i.j) → (14.i-1.j) für $i=1, \dots, n, j=1, \dots, p$
- (13.i.j) → (13.i.j-1) für $i=0, \dots, n, j=2, \dots, p$
- (14.i.j) → (14.i.j-1) für $i=0, \dots, n, j=2, \dots, p$
- (13.i.j) → (14.i.j) für $i=0, \dots, n, j=1, \dots, p$

Eine weitere wichtige Klasse von Interpretationszielen ist das Erkennen von Mittelwerten. Datenwerte eines Merkmals werden für mehrere Gitterpunkte zusammengefaßt und als ein Mittelwert in der Visualisierung dargestellt. Dies ermöglicht bei Datensätzen mit höherdimensionalem Gitterraum die Visualisierung eines Überblicks über den gesamten Datensatz.

$$(15.i.j.k) \quad \exists S_i \in \mathfrak{R}_i \exists mk \in \text{Mer}_k \forall g \in S_i \exists S_j \in \mathfrak{R}_j \forall m \in mk \\ (\text{bdis}(S_i) \wedge S_i \cap S_j = \{g\} \wedge \text{vis}(g, m(S_j, m), m))$$

$i=0, \dots, n$
 $j=0, \dots, n-i$
 $k=1, \dots, p$

An jedem Gitterpunkt eines i-dimensionalen Schnittes ist für jedes Merkmal einer k-elementigen Merkmalsklasse der Mittelwert aller in einem j-dimensionalen Schnitt in diesem Merkmal auftretenden Datenwerte erkennbar.

Beispiel:

Für einen Volumendatensatz soll ein Überblick mittels Raytracing gewonnen werden. Für jeden Strahl, der durch das Volumen "geschossen" wird, wird der Mittelwert der Datenwerte aller derjenigen Gitterpunkte gebildet, die nahe dem Strahl liegen. An jedem Punkt einer zweidimensionalen Gitterpunktmenge wird also der Mittelwert einer eindimensionalen Gitterpunktmenge (aller Gitterpunkte in der Nähe eines Strahls) gebildet. Dies entspricht dem Interpretationsziel (15.2.1.1).

Spezialfälle:

(15.n.0.p): An jedem Gitterpunkt ist für jedes Merkmal ein Mittelwert aller dort auftretenden Datenwerte sichtbar.

(15.0.n.p): Für jedes Merkmal ist ein Mittelwert aller im Datensatz in diesem Merkmal stehenden Datenwerte erkennbar.

Bei der Kennzeichnung der Zusammenhänge zwischen den eingeführten Interpretationszielen ergeben sich zwei Arten von Abhängigkeiten:

- Die direkte Ableitbarkeit eines Interpretationsziels aus einem anderen wurde bereits eingeführt und in den bisherigen Abbildungen mit einem Pfeil gekennzeichnet.
- Die intuitive Ableitbarkeit eines Interpretationsziels aus einem anderen soll am Beispiel der Interpretationsziele (3) und (15.0.0.p) erläutert werden. (3) besagt, daß an einem bestimmten Gitterpunkt für jedes Merkmal alle in diesem Punkt stehenden Datenwerte erkennbar sind. (15.0.0.p) besagt, daß an einem bestimmten Gitterpunkt für jedes Merkmal ein Mittelwert aller in diesem Punkt stehenden Datenwerte erkennbar ist. Sind in einem Punkt und einem beliebigen Merkmal alle Datenwerte erkennbar, so ist dem Nutzer intuitiv auch das Erkennen eines Mittelwertes dieser Datenwerte möglich. Wie gut dieser Mittelwert erkannt werden kann, hängt von der konkreten Visualisierungstechnik ab. Es ist möglich, daß der Mittelwert schlechter zu erkennen ist als die Einzelwerte, beispielsweise wenn die Einzelwerte über den gesamten Bildschirm verteilt liegen. Andererseits kann der Mittelwert auch sehr gut erkennbar sein. z.B. wenn er neben den Einzelwerten noch extra codiert ist. Prinzipiell ist aber bei Erkennen der Einzelwerte ein intuitives Schließen auf einen Mittelwert möglich. Das Interpretationsziel (15.0.0.p) ist intuitiv aus (3) ableitbar. Die intuitive Ableitbarkeit soll in der graphischen Darstellung durch einen unterbrochenen Pfeil gekennzeichnet werden.

Die Abhängigkeiten der neu eingeführten Interpretationsziele (15.i.j.k) werden durch die Bilder 8-11 verdeutlicht:

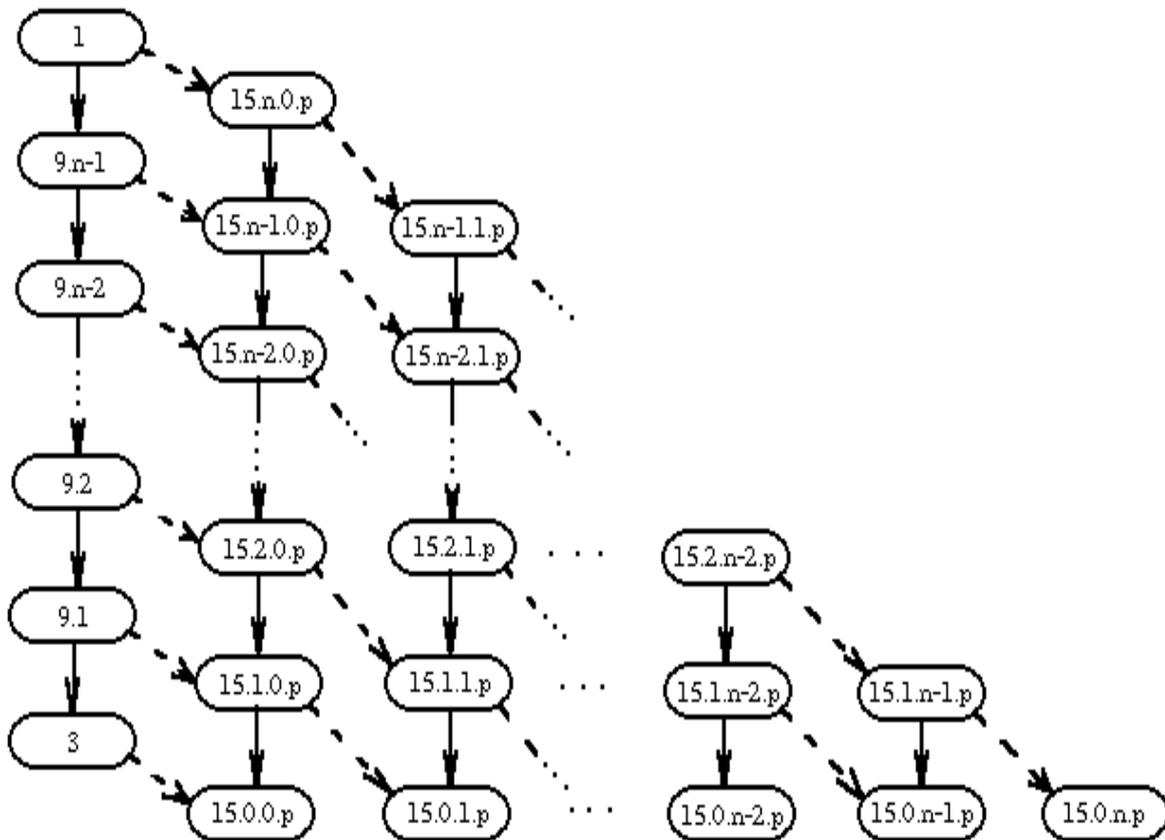


Bild 8: Abhängigkeiten der Interpretationsziele (1), (3), (9.i), (15.i,j,p)

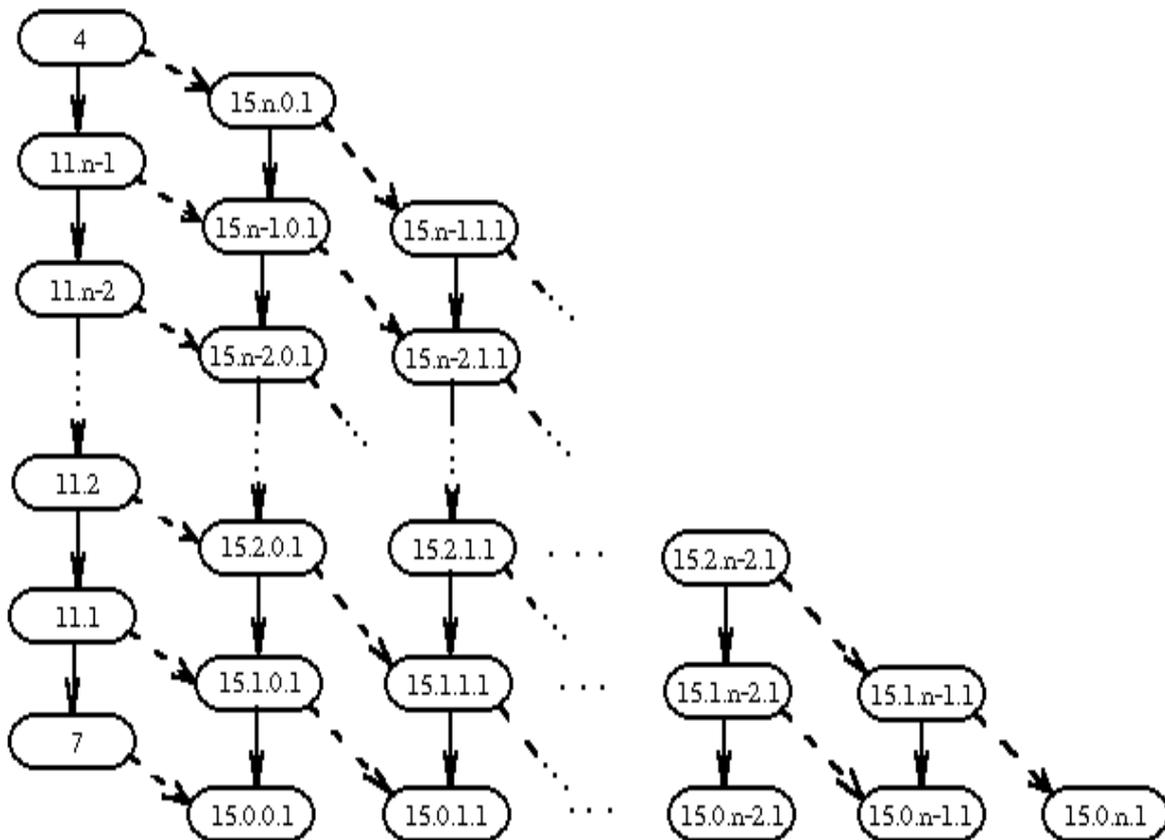


Bild 9: Abhängigkeiten der Interpretationsziele (4), (7), (11.i), (15.i,j,1)

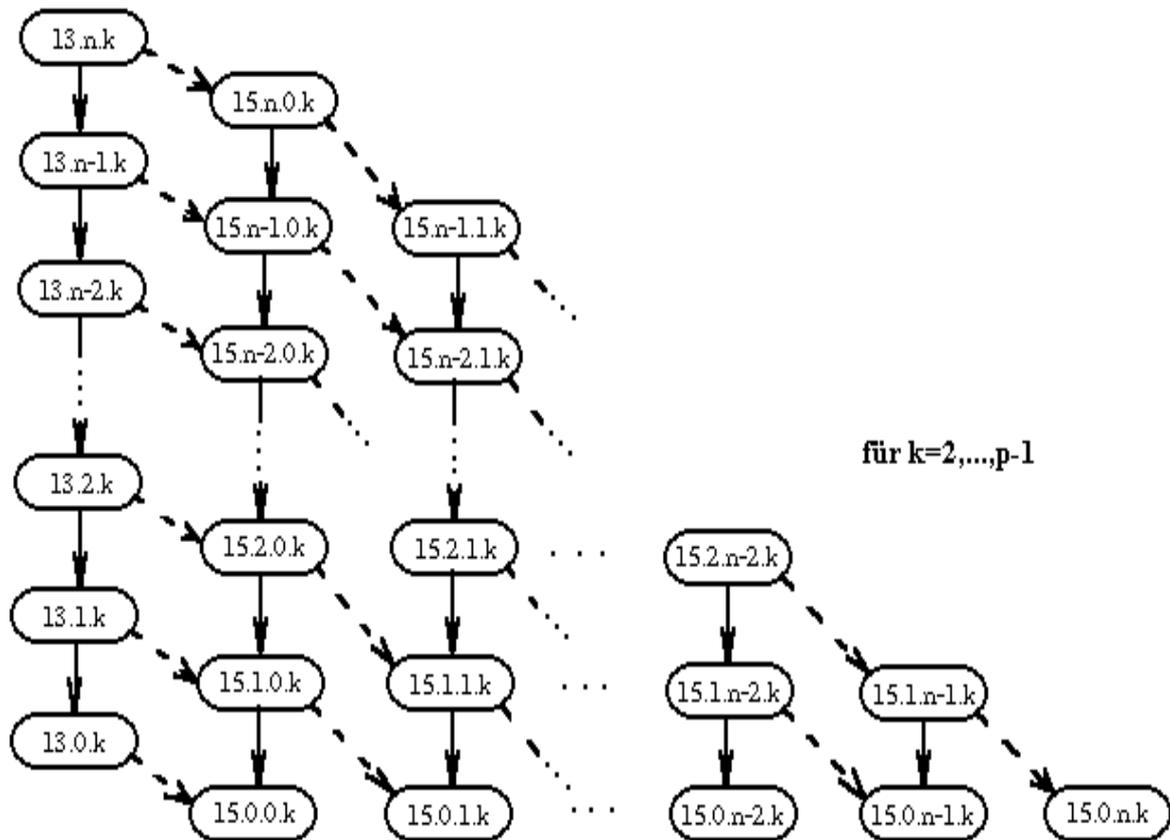


Bild 10: Abhängigkeiten der Interpretationsziele (13.i.k), (15.i.j.k)

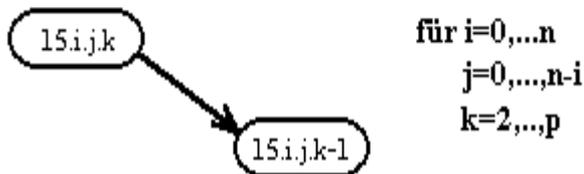


Bild 11: Abhängigkeiten der Interpretationsziele (15.i.j.k)

Eine weitere Klasse von Interpretationszielen läßt sich unter dem Oberbegriff "Erkennen von Häufigkeiten" beschreiben.

$$(16.i.j.k) \quad \exists S_i \in \mathfrak{R}_i \exists m_1, \dots, m_k \in \text{Mer} \forall g \in S_i \exists S_j \in \mathfrak{R}_j \forall v_1, \dots, v_k \in \text{Val} \\ (m_1, \dots, m_k \text{ paarweise verschieden} \wedge \text{bdis}(S_j) \\ \wedge S_i \cap S_j = \{g\} \wedge \text{vis}(g, h_k(S_j, m_1, \dots, m_k, v_1, \dots, v_k)))$$

$$i=0, \dots, n \\ j=0, \dots, n-i \\ k=1, \dots, p$$

An jedem Beobachtungspunkt eines i-dimensionalen Schnittes ist erkennbar, mit welcher Häufigkeit eine beliebige k-elementige Datenkombination in einem j-dimensionalen Schnitt in den Merkmalen m_1, \dots, m_k vorkommt.

$$(17.i.j.k) \quad \exists S_i \in \mathfrak{R}_i \exists m_1, \dots, m_k \in \text{Mer} \exists v_1, \dots, v_k \in \text{Val} \forall g \in S_i \exists S_j \in \mathfrak{R}_j$$

$$(m_1, \dots, m_k \text{ paarweise verschieden} \wedge \text{bdis}(S_i) \\ \wedge S_i \cap S_j = \{g\} \wedge \text{vis}(g, h_k(S_j, m_1, \dots, m_k, v_1, \dots, v_k)))$$

$i=0, \dots, n$
 $j=0, \dots, n-i$
 $k=1, \dots, p$

An jedem Punkt eines i-dimensionalen Schnittes ist erkennbar, mit welcher Häufigkeit eine bestimmte k-elementige Datenwertkombination in einem j-dimensionalen Schnitt in den Merkmalen m_1, \dots, m_k vorkommt.

Die Abhängigkeiten für die neu eingeführten Interpretationsziele ergeben sich aus den Bildern 12-17:

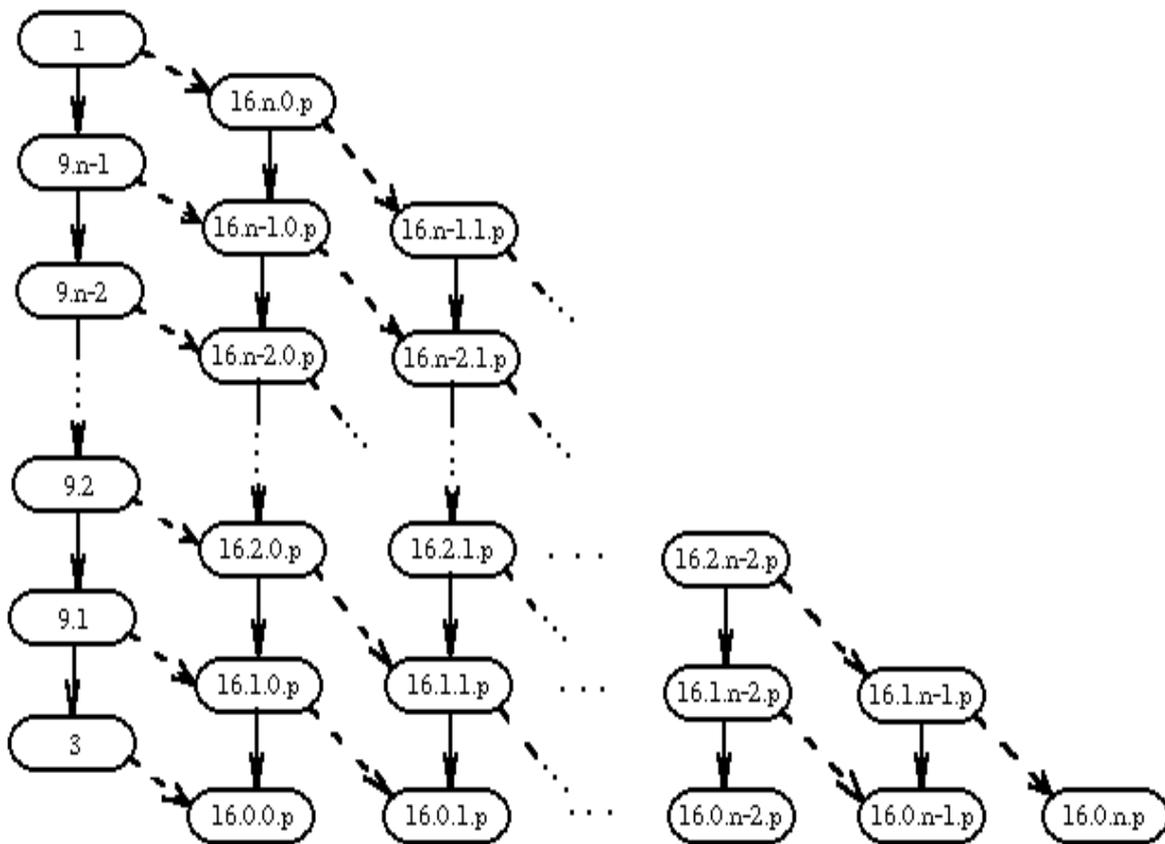


Bild 12: Abhängigkeiten der Interpretationsziele (1), (3), (9.i), (16.i.j.p)

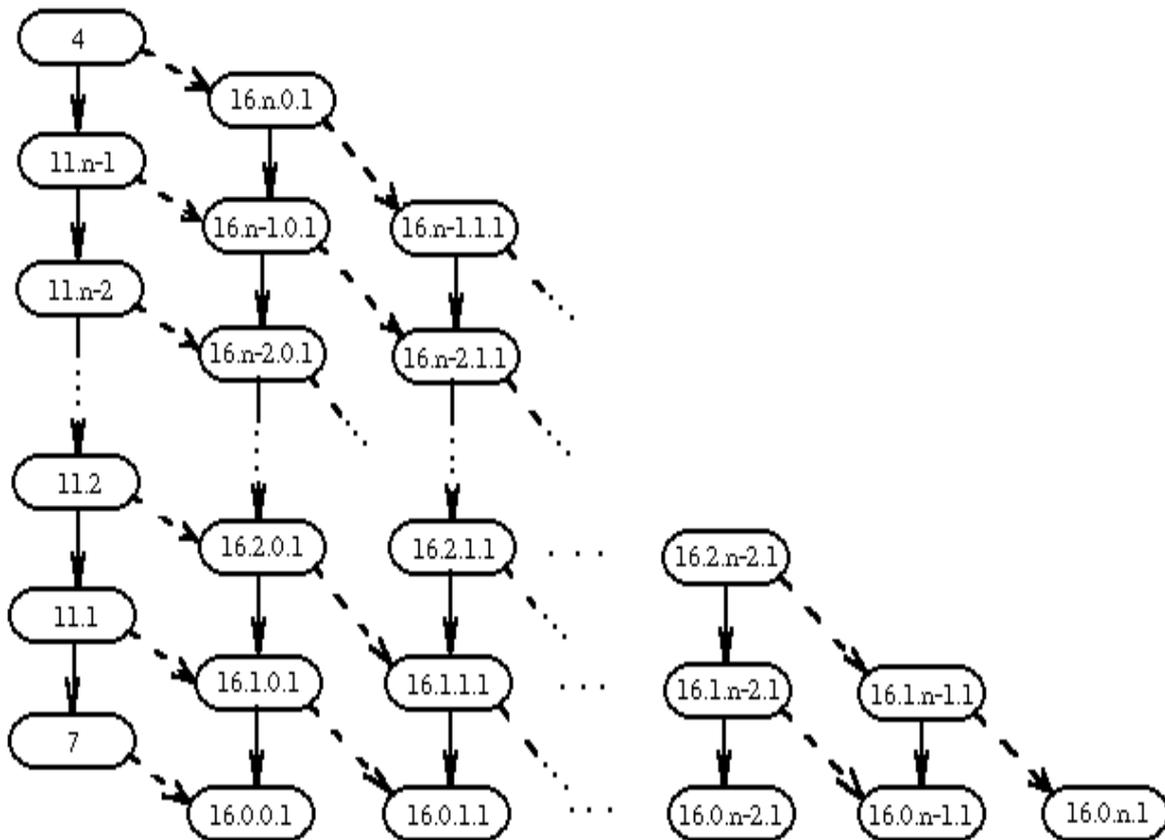


Bild 13: Abhängigkeiten der Interpretationsziele (4), (7), (11.i), (16.i.j.1)

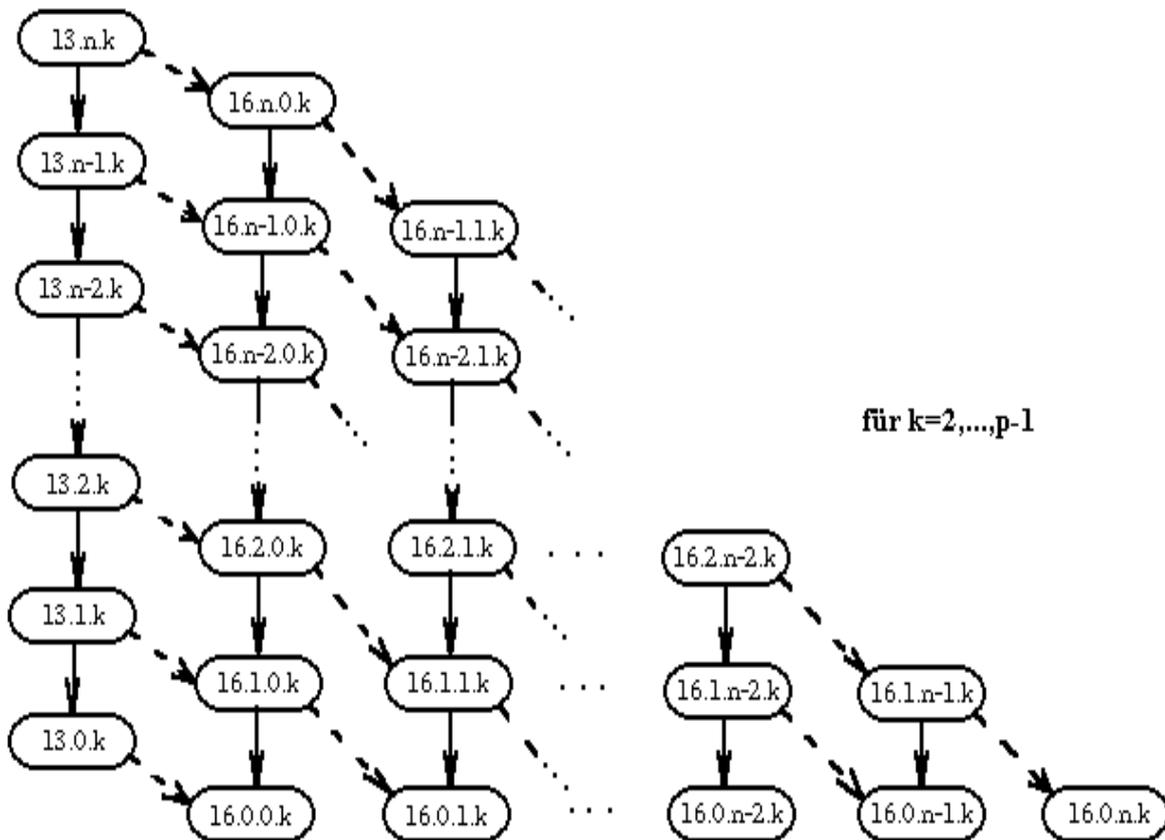


Bild 14.: Abhängigkeiten der Interpretationsziele (13.i.k), (16.i.j.k)

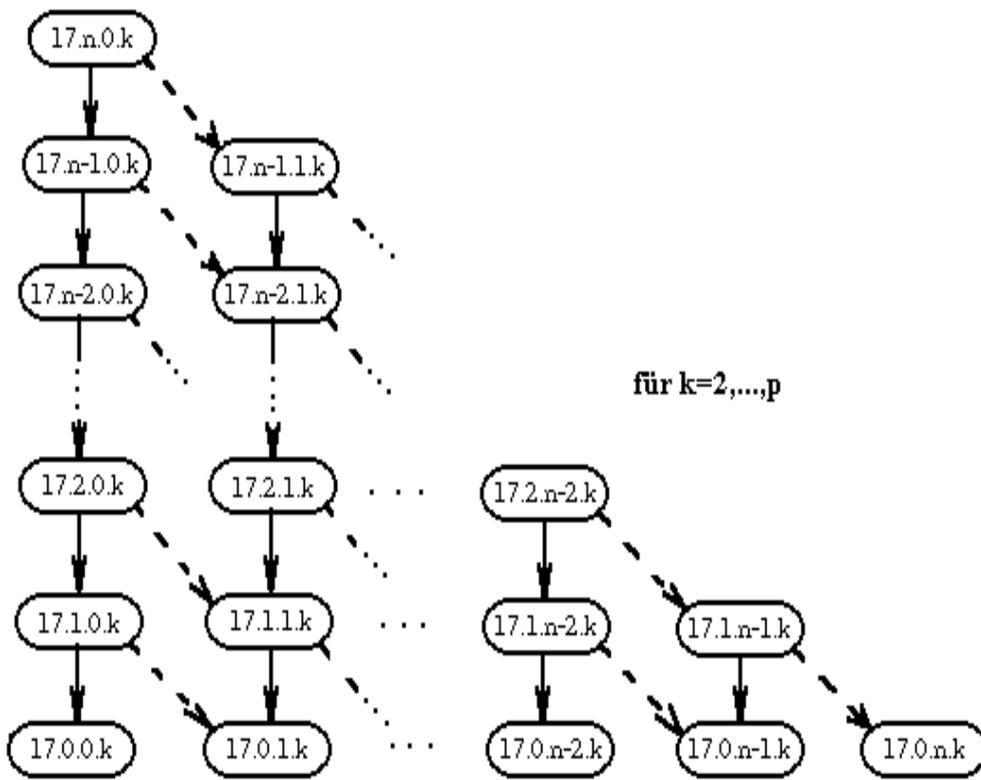


Bild 15: Abhängigkeiten der Interpretationsziele $(17.i.j.k)$

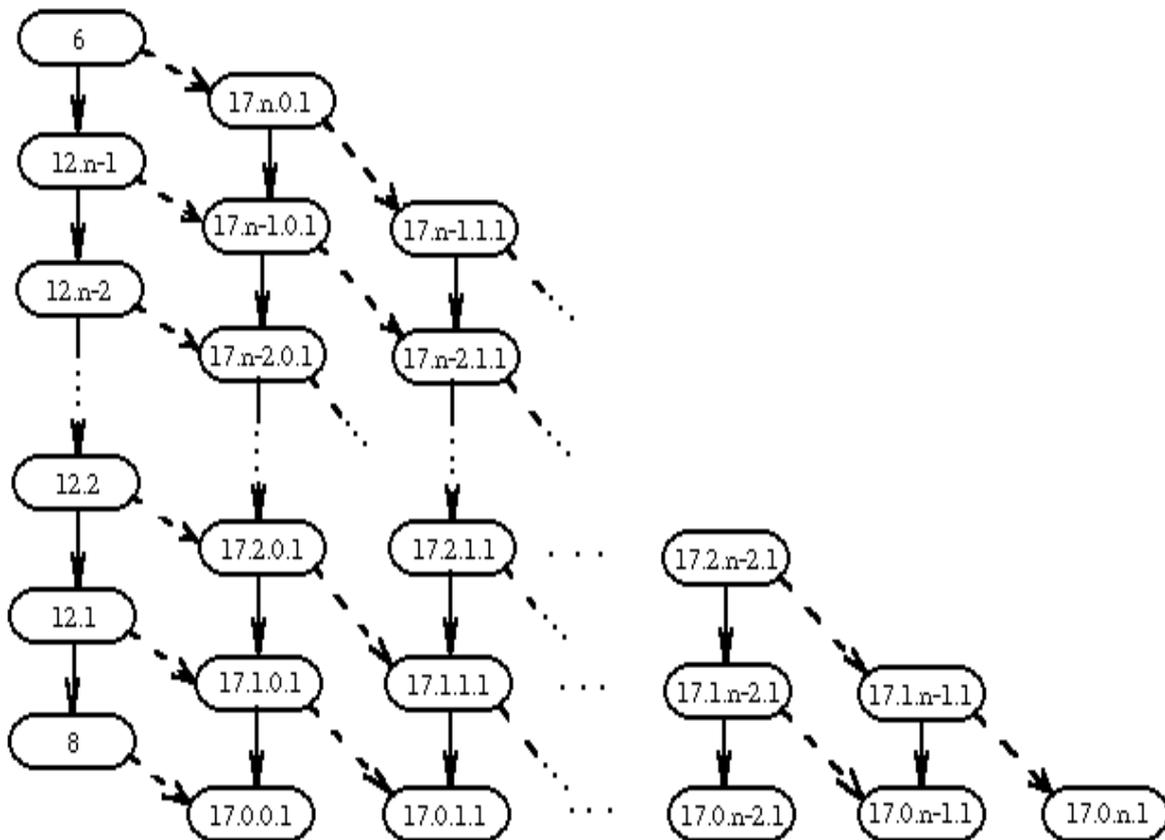


Bild 16: Abhängigkeiten der Interpretationsziele (6) , (8) , $(12.i)$, $(17.i.j.1)$



Bild 17: Abhängigkeiten der Interpretationsziele (16.i.j.k), (17.i.j.k)

Ein weiterer wichtiger Komplex von Interpretationszielen ist das Erkennen von Korrelationen. Für dieses Ziel gibt es folgende Spezialisierungen:

- Erkennen von Korrelationen zwischen den Merkmalen
- Erkennen von Korrelationen zwischen Merkmalen und der Lage im Gitter (Erkennen von Verteilungen)

Die Korrelationen zwischen den Merkmalen können global (über den gesamten Datensatz) oder lokal (für jeden Gitterpunkt extra) betrachtet werden. Zwischenstufen (Betrachtungen auf Schnitten) sind ebenfalls möglich.

$$(18.i.j) \quad \exists S_i \in \mathfrak{R}_i \quad \forall g \in S_i \quad \exists S_j \in \mathfrak{R}_j \quad (bdis(S_i) \wedge S_i \cap S_j = \{g\} \wedge kvis(g, S_j))$$

$i=0,\dots,n$
 $j=0,\dots,n-i$

An jedem Punkt eines i-dimensionalen Schnittes sind alle vorhandenen Korrelationen der Merkmale bzgl. des Schnittes S_j zu erkennen. Es ergeben sich folgende Abhängigkeiten:

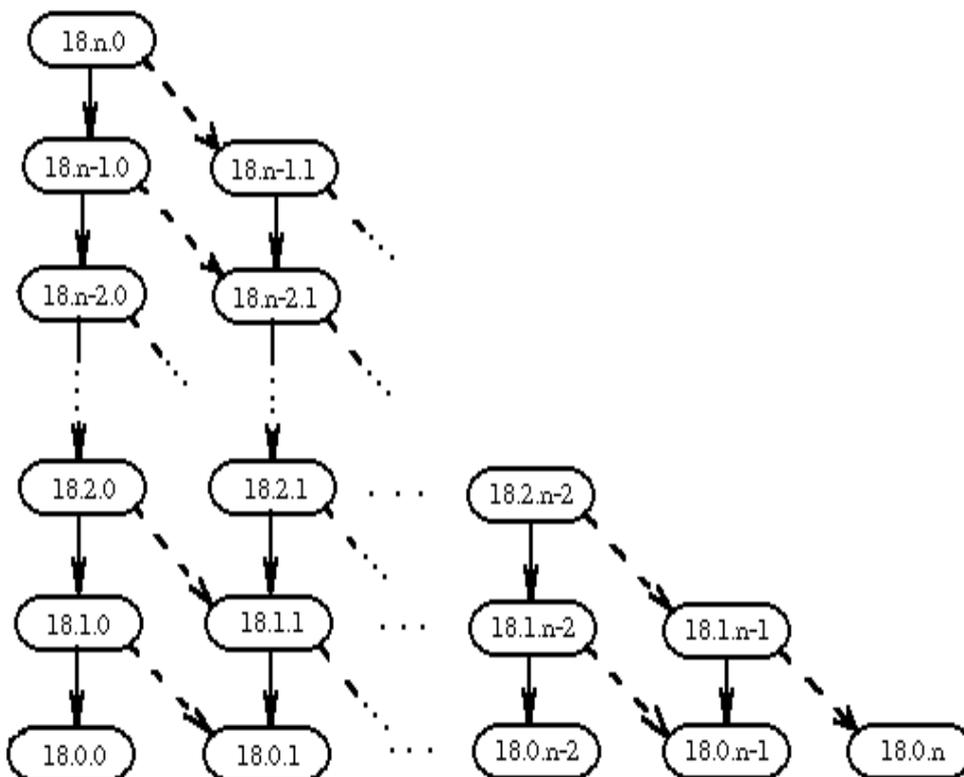


Bild 18: Abhängigkeiten der Interpretationsziele (18.i.j)

Das Erkennen von Korrelationen zwischen Merkmalen und der Lage im Gitter soll durch das folgende Interpretationsziel beschrieben werden:

$$(19.i.j.k) \quad \exists S_i \in \mathfrak{R}_i, \exists m_k \in \text{Mer}_k, \forall g \in S_i, \exists S_j \in \mathfrak{R}_j, \forall m \in m_k \\ (\text{bdis}(S_i) \wedge S_i \cap S_j = \{g\} \wedge \text{kvis}(g, S_j, m))$$

$$i=0, \dots, n-1$$

$$j=1, \dots, n-i$$

$$k=1, \dots, p$$

An jedem Punkt eines i-dimensionalen Schnittes sind für jedes Merkmal einer k-elementigen Merkmalsklasse die Korrelationen zur Lage im Gitter, bezogen auf einen j-dimensionalen Schnitt, erkennbar. Die Bilder 19 und 20 zeigen die Abhängigkeiten der neu eingeführten Interpretationsziele:

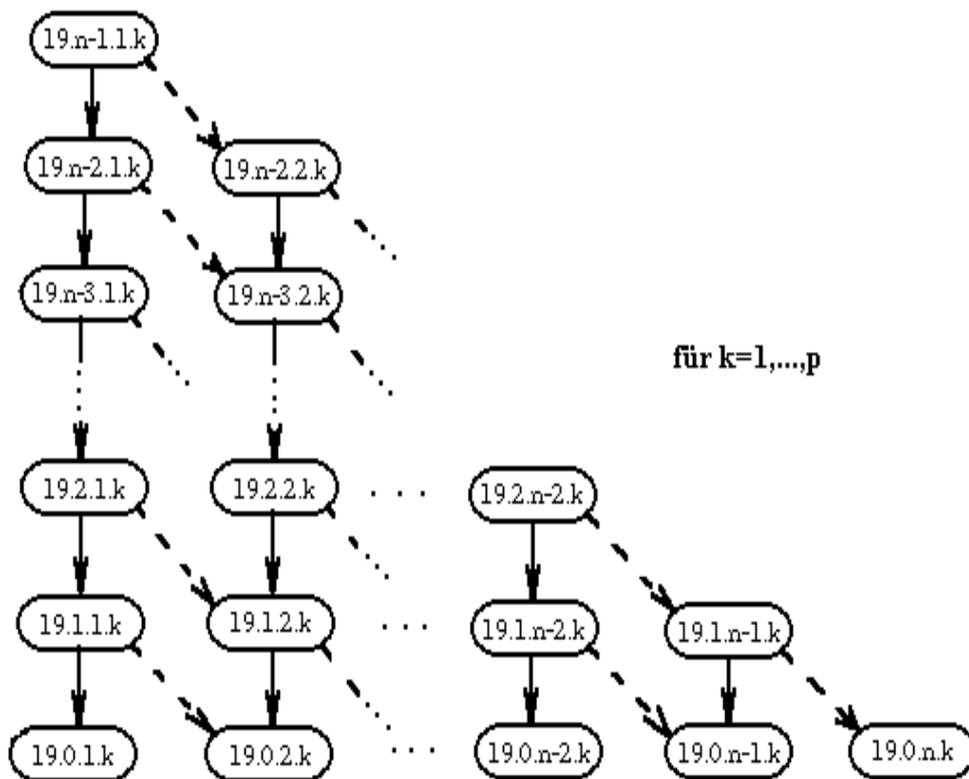


Bild19: Abhängigkeiten zwischen den Interpretationszielen (19.i.j.k)

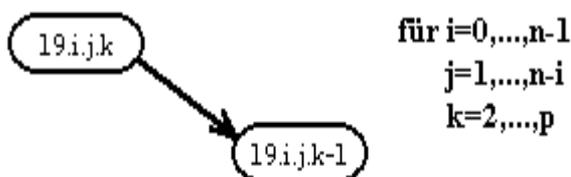


Bild20: Abhängigkeiten der Interpretationsziele (19.i.j.k)

Das nächste betrachtete Interpretationsziel ist die Interpolation unbekannter Datenwerte an Gitterpunkten. Diese kann -je nach Anforderung- vom Nutzer gewünscht oder ausgeschlossen werden. Es wird unterschieden, ob die Darstellung des interpolierten Wertes stets erfolgen soll oder nur dann, wenn die Interpolation einen bestimmten Wert ergibt.

$$(20) \quad \forall g \in \text{Git} \quad \forall m \in \text{Mer} \quad (\text{unbek}(g,m) \Rightarrow \text{vis}(g,\text{int}(g,m),m))$$

In jedem Gitterpunkt und in jedem Merkmal wird ein interpolierter Wert dargestellt, falls dort kein beobachteter Wert bekannt ist.

$$(21) \quad \exists v \in \text{Val} \quad \forall g \in \text{Git} \quad \forall m \in \text{Mer} \quad ((\text{unbek}(g,m) \wedge \text{int}(g,m)=v) \Rightarrow \text{vis}(g,v,m))$$

In einem beliebigen Gitterpunkt und einem beliebigen Merkmal wird ein interpolierter Wert dargestellt, falls die Interpolation einen bestimmten Wert ergibt.

Ein weiteres mögliches Interpretationsziel ist das Erkennen von interpolierten Werten zwischen den Gitterpunkten (nur sinnvoll, wenn die Gitterlinien nicht quasi-unendlich dicht liegen). Dieses Interpretationsziel in seiner allgemeinen Form zu formulieren, ist wenig sinnvoll, da das Ziel in dieser Form nicht erfüllbar ist. Es soll deshalb nur der einzig in der Praxis auftretende Fall (ein Merkmal, ein Datenwert) betrachtet werden.

$$(22.i) \quad \exists S_i \in \mathfrak{R}_i \quad \exists v \in \text{Val} \quad \exists m \in \text{Mer} \quad (\text{zw}(S_i,v,m))$$

$i=1,\dots,n$

In einem i-dimensionalen Schnitt durch den Gitterraum ist jeder Punkt zwischen den Gitterpunkten erkennbar, in dem eine Interpolation im Merkmal m den Wert v ergibt.

Sollen Werte zwischen den Gitterpunkten interpoliert werden, so ist es sicher sinnvoll, bei Bedarf auch unbekannte Werte an den Gitterpunkten zu interpolieren. Dies ergibt folgende Abhängigkeiten:

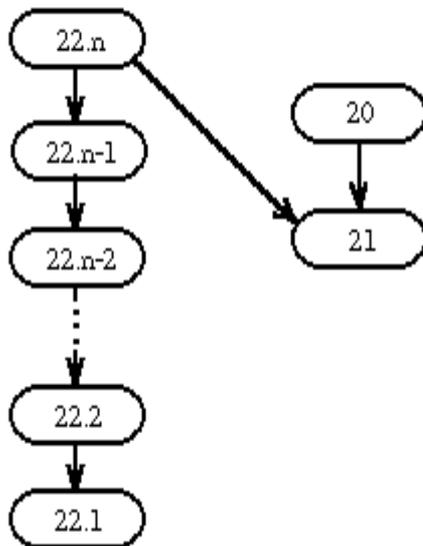


Bild 21: Abhängigkeiten der Interpretationsziele (20), (21) und (22.i)

Eine weitere wichtige Klasse von Interpretationszielen ist das Erkennen von Extremwerten. Dafür wird das folgende Interpretationsziel eingeführt:

$$(23.i.j.k) \quad \exists S_i \in \mathfrak{R}_i \exists m_k \in \text{Mer}_k \forall g \in S_i \exists S_j \in \mathfrak{R}_j \forall m \in m_k \\ (\text{bdis}(S_i) \wedge S_i \cap S_j = \{g\} \wedge \text{vis}(g, \text{ex}(S_j, m), m))$$

$i=0, \dots, n$
 $j=0, \dots, n-i$
 $k=1, \dots, p$

An jedem Gitterpunkt eines i-dimensionalen Schnittes ist für jedes Merkmal einer k-elementigen Merkmalsklasse das Maximum (oder Minimum) aller in einem j-dimensionalen Schnitt in diesem Merkmal auftretenden Werte erkennbar.

Die Abhängigkeiten der neu eingeführten Interpretationsziele (23.i.j.k) ergeben sich aus den Bildern 22-25:

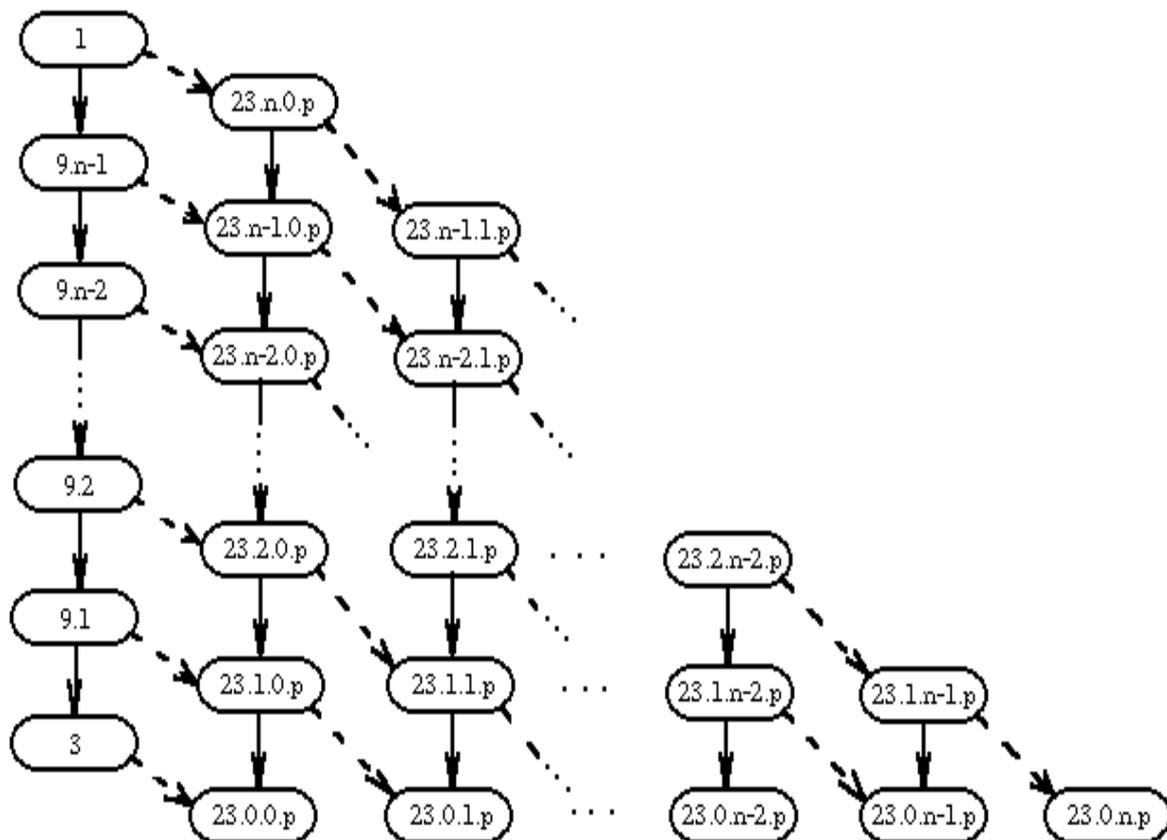


Bild 22: Abhängigkeiten der Interpretationsziele (1), (3), (9.i), (23.i.j.p)

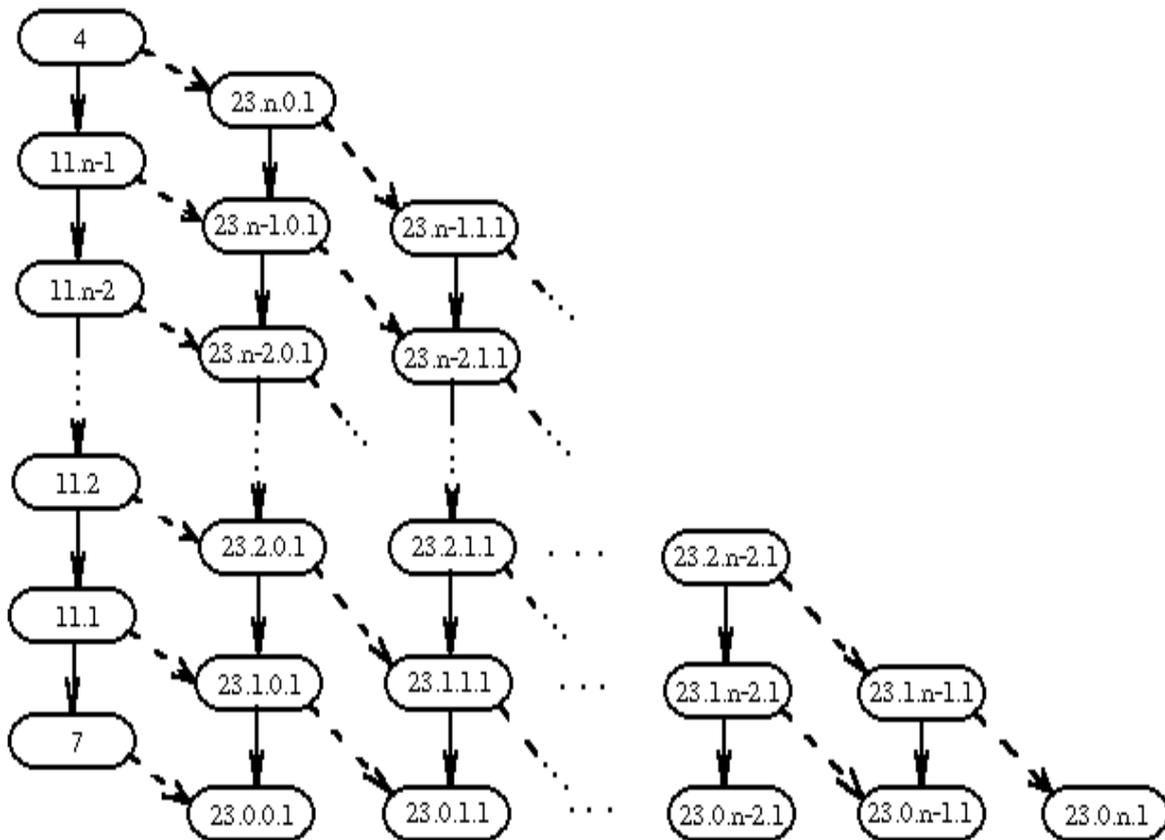


Bild 23: Abhängigkeiten der Interpretationsziele (4), (7), (11.i), (23.i.j.1)

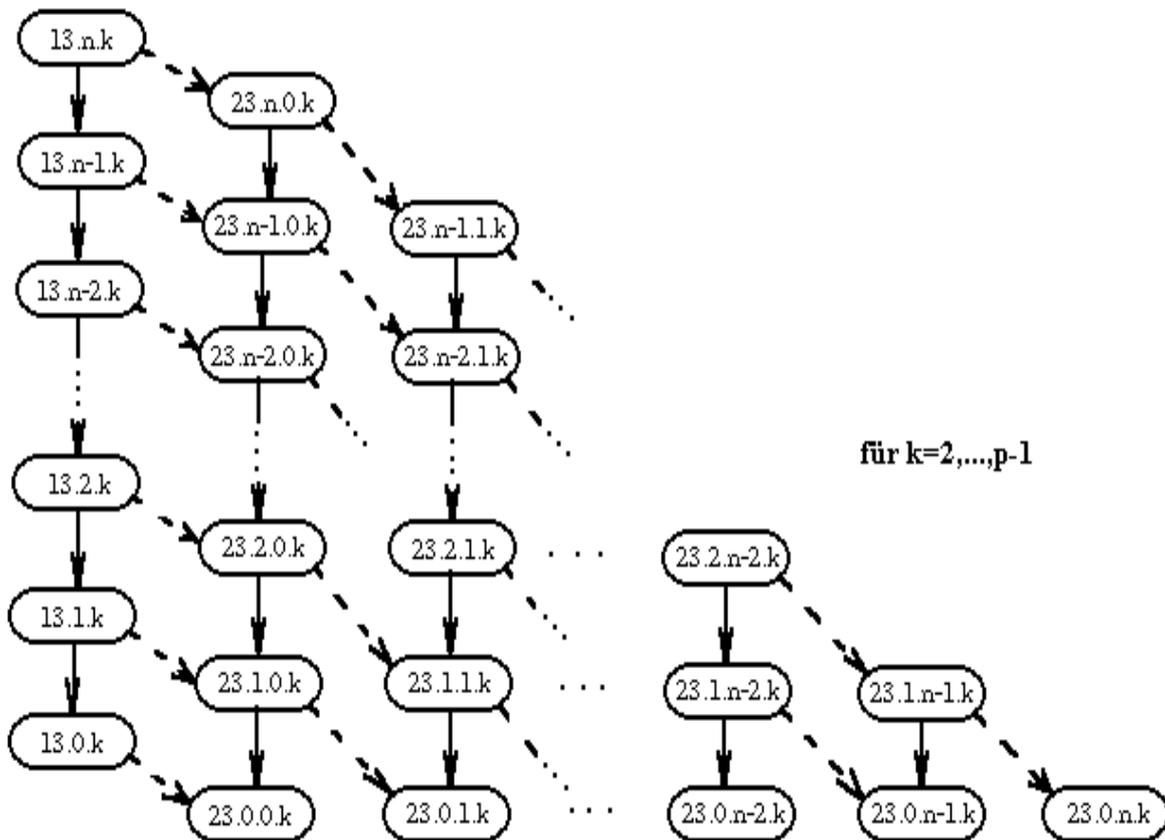


Bild 24: Abhängigkeiten der Interpretationsziele (13.i.k) und (23.i.j.k)

(18.i.j) steht für

$(18.n.0) \text{---} \dots \rightarrow (18.0.0) \text{---} \longrightarrow (18.n-1.1) \text{---} \dots \rightarrow (18.0.1) \text{---} \dots \rightarrow (18.0.n).$

(22.i) steht für

$(22.n) \text{---} \dots \rightarrow (22.1).$

3.2. Die Abbildung konkreter Interpretationsziele auf abstrakte

Das im vorigen Abschnitt eingeführte System abstrakter Interpretationsziele hat den Vorteil der Universalität, d.h., es ist für beliebige wissenschaftliche Datensätze anwendbar. Nachteilig ist jedoch, daß das System so abstrakt und kompliziert ist, daß vom Nutzer nicht verlangt werden kann, damit zu arbeiten. Seine konkreten Interpretationsziele müssen abgebildet werden auf das System abstrakter Ziele. Zur Findung einer solchen Abbildung ist ein Experte nötig als Bindeglied zwischen Nutzer und System.

Die Aufgaben eines solchen Experten sind automatisierbar für bestimmte, oft genutzte und klar abgegrenzte Anwendungsgebiete. Für neue Anwendungsgebiete müssen diese Abbildungen jedoch immer wieder neu gefunden werden. Auf die Möglichkeit der Automatisierbarkeit der Abbildungsfindung soll hier nicht näher eingegangen werden. Es sollen lediglich Beispiele für solche Abbildungen gegeben werden.

Die Abbildung von konkreten in abstrakte Interpretationsziele soll so beschrieben werden, daß sich für einen (n,p)-Datensatz als Bild eine Abbildung $w_{n,p}$ der Menge aller abstrakten Interpretationsziele in das Intervall $[0,1]$ ergibt. $w_{n,p}(i)$ gibt dabei die Wichtigkeit des Interpretationsziels (i) für die Visualisierung an. $w_{n,p}(i)=1$ bedeutet, daß auf die Erfüllung des Interpretationsziels (i) in der Visualisierung ein größtmöglicher Wert gelegt wird. $w_{n,p}(i)=0$ bedeutet, daß das Interpretationsziel (i) für die Visualisierung keine Rolle spielt. Die Abbildung $w_{n,p}$ sei auch als Visualisierungswunsch für einen (n,p)-Datensatz bezeichnet.

Ein Visualisierungswunsch $w_{n,p}$ ist an bestimmte Regeln gebunden, welche sich aus dem System abstrakter Interpretationsziele ergeben. Ist ein Interpretationsziel (j) aus dem Ziel (i) direkt ableitbar, so schließt der Wunsch nach einer Erfüllung des Ziels (i) auch den Wunsch nach Erfüllung des Ziels (j) ein. Es muß $w_{n,p}(i) \leq w_{n,p}(j)$ gelten.

Ist ein Interpretationsziel (j) aus einem Interpretationsziel (i) intuitiv ableitbar, so kann bei vorhandenem Visualisierungswunsch für (i) der Wunsch für (j) nicht 0 sein. Dies soll ausgedrückt werden durch die Bedingung $w_{n,p}(j) \geq w_{n,p}(i)/10$. Der Wert 10 ist dabei willkürlich. Er wurde möglichst groß gewählt, um die Bedingung schwach zu formulieren.

Von einem Visualisierungswunsch sei im folgenden vorausgesetzt, daß er für beliebige Interpretationsziele die Bedingung für die direkte und intuitive Ableitbarkeit erfüllt.

Im folgenden sollen Beispiele für Visualisierungswünsche für konkrete Interpretationsziele gegeben werden. Dabei sollen nur diejenigen Interpretationsziele (i) aufgeschrieben werden, für die $w_{n,p}(i) > 0$ gilt und die sich nicht durch direkte oder intuitive Ableitung aus anderen Interpretationszielen ergeben. Falls also "(i) \longrightarrow (j)" gilt und $w_{n,p}(j)$ nicht explizit angegeben ist, so gilt $w_{n,p}(j) = w_{n,p}(i)$.

Beispiele:

In /Schw93/ werden konkrete Interpretationsziele für spezielle Volumendaten aus der Medizin (Diagnose von Leberkrankheiten) formuliert. Für diese konkreten Interpretationsziele sollen hier die jeweiligen Visualisierungswünsche angegeben werden. Für den Volumendatensatz gilt $n=3$ und $p=1$.

- 1) Erkennen der Glattheit der Oberfläche der Leber:

$$w_{3,1}(2) = 1$$

$$w_{3,1}(22.3) = 1$$

Zur Erfüllung dieses Interpretationsziels ist es wichtig, bestimmte Datenwerte an Gitterpunkten und zwischen diesen lokalisieren zu können.

- 2) Gewinnen eines Überblicks über die Homogenität im Inneren der Leber

$$w_{3,1}(15.2.1.1) = 0.8$$

$$w_{3,1}(19.0.3.1) = 0.8$$

$$w_{3,1}(23.2.1.1) = 1$$

Es ist wichtig, an jedem Punkt eines 2D-Schnittes Mittelwerte und Extremwerte einer eindimensionalen Gitterpunktmenge zu erkennen. Dabei ist das Erkennen von Extremwerten wichtiger als das Erkennen von Mittelwerten, da sich Inhomogenitäten besonders durch Änderungen bei den Extremwerten zeigen. Wichtig ist außerdem das Erkennen globaler Korrelationen zwischen Datenwerten und Lage im Raum.

- 3) Erkennen von Größe und Form von Tumoren

$$w_{3,1}(2) = 1$$

$$w_{3,1}(22.3) = 1$$

Dieses Interpretationsziel ist ähnlich zu 1), es wird lediglich ein anderer Datenwert zur Lokalisation verwendet.

- 4) Erkennen der exakten Lage von Tumoren in der Leber

$$w_{3,1}(15.2.1.1) = 1$$

$$w_{3,1}(19.0.3.1) = 1$$

$$w_{3,1}(23.2.1.1) = 0.8$$

Hier ist das Erkennen von Mittelwerten wichtiger als das Erkennen von Extremwerten, da ein Tumor Mittelwerte ändert, Extremwerte aber unangetastet lassen kann.

- 5) Analyse sonstiger makroskopischen Inhomogenitäten im Inneren der Leber

$$w_{3,1}(15.2.1.1) = 0.8$$

$$w_{3,1}(19.0.3.1) = 0.8$$

$$w_{3,1}(23.2.1.1) = 0.8$$

Dieses Interpretationsziel läßt sich auffassen als Abschwächung von 2) und 4).

- 6) Analyse von Größenrelationen auf Schnitten durch das Lebergewebe

$$w_{3,1}(9.2) = 0.8$$

$$w_{3,1}(19.0.2.1) = 0.8$$

Auf einem 2D-Schnitt soll identifiziert und lokalisiert werden können, außerdem sollen auf dem Schnitt Korrelationen zwischen Datenwerten und Lage im Raum erkannt werden.

- 8) Erkennen mikroskopischer Veränderungen auf Schnitten durch die Leber

$$w_{3,1}(9.2) = 1$$

$$w_{3,1}(19.0.2.1) = 0.8$$

Ein ähnliches Interpretationsziel wie 6), zum Erkennen aller mikroskopischen Veränderungen soll aber besonderen Wert darauf gelegt werden, daß wirklich an allen Gitterpunkten des 2D-Schnittes identifiziert und lokalisiert werden kann.

- 9) Erkennen mikroskopischer Veränderungen im Inneren der Leber.

$$w_{3,1}(15.2.1.1) = 1$$

$$w_{3,1}(19.0.3.1) = 0.8$$

$$w_{3,1}(23.2.1.1) = 1$$

Ähnlich wie 4) und 7).

- 10) Erkennen von Einlagerungen im Lebergewebe

$$w_{3,1}(15.2.1.1) = 0.8$$

$$w_{3,1}(19.0.3.1) = 1$$

$$w_{3,1}(23.2.1.1) = 0.8$$

In /Arnd92/ werden konkrete Interpretationsziele für einen mikrobiologischen Datensatz formuliert. Für diese sollen die Visualisierungswünsche angegeben werden. Für den Datensatz gelte $n=0$ und $p=41$.

- 1) Zur Bewertung einer Gruppe von Merkmalen wird nach dem Auftreten von Mustern dieser Merkmale gesucht. Dabei interessiert, wieviele und welche verschiedenen Muster aufzufinden sind und wie sich die Isolate darauf verteilen.

$$w_{0,41}(13.0.k) = 0.8$$

$$w_{0,41}(16.0.0.k) = 1$$

Über Muster lassen sich am besten Aussagen treffen, wenn für beliebige k -elementige Wertekombinationen Aussagen über die Häufigkeit ihres Auftretens gemacht werden können. Außerdem ist eine Identifizierbarkeit und Lokalisierbarkeit für k -elementige Merkmalskombinationen wichtig. k ist dabei die maximale Größe einer Merkmalsgruppe, für die nach Mustern gefragt wird.

- 2) Beurteilen des Trennvermögens eines Merkmals

$$w_{0,41}(18.0.0) = 1$$

Sind alle vorhandene Korrelationen zwischen den Merkmalen zu erkennen, erkennt man auch das Trennvermögen eines Merkmals sowie die Tatsache, ob zwei Merkmale dieselbe Trennung vornehmen.

- 3) Aufdecken von Zusammenhängen zwischen den Merkmalen.

$$w_{0,41}(18.0.0) = 1$$

4. Der Algorithmus zur Auswahl geeigneter Visualisierungstechniken

Gegeben ist ein (n,p) -Datensatz und ein Visualisierungswunsch $w_{n,p}$. Für jede zur Verfügung stehende Visualisierungstechnik soll überprüft werden, ob sie für den gegebenen Datensatz dem Visualisierungswunsch entspricht oder zumindest nahekommt. Dazu dient folgendes prinzipielles Vorgehen:

- 1) Es wird überprüft, ob die Technik überhaupt sinnvoll auf den Datensatz angewendet werden kann. Der Test hierfür erfolgt in zwei Schritten:
 - a) Es wird geprüft, ob der Datensatz prinzipiell mit der Technik zu visualisieren ist.
Beispiel: Die Marching-Cube-Technik kann nur angewendet werden, wenn $n \geq 3$ gilt. Die Entscheidung hierüber ist für jede Visualisierungstechnik binär und unabhängig vom Visualisierungswunsch. Ist eine Technik prinzipiell ungeeignet, so brauchen die nächsten Schritte gar nicht erst durchgeführt zu werden.
 - b) Es wird der Platzbedarf überprüft, der zur Visualisierung eines Datensatzes mit einer Visualisierungstechnik nötig ist. Übersteigt der Platzbedarf die Bildschirmgröße, so ist die Technik ebenfalls ungeeignet und wird nicht weiter betrachtet. Zur Modellierung des Platzbedarfs werden für jede Technik zwei Funktionen

$$\begin{aligned} pl_x &: \mathbb{M}_{n,p} \rightarrow \mathbb{N} \\ pl_y &: \mathbb{M}_{n,p} \rightarrow \mathbb{N} \end{aligned}$$

eingeführt. pl_x gibt dabei die minimale Pixelanzahl in x-Richtung an, die ein Bildschirmfenster haben muß, um den Datensatz mit der Technik visualisieren zu können. pl_y ist der analoge Wert für die y-Richtung. Ist $pl_x(M_{n,p})$ größer als die maximale Ausdehnung des Fensters in x-Richtung, oder ist $pl_y(M_{n,p})$ größer als die maximale Ausdehnung des Fensters in y-Richtung, so ist die Visualisierungstechnik für den Datensatz prinzipiell ungeeignet.

Die Schritte a) und b) werden zusammengefaßt, indem für jede Technik eine Möglichkeitsfunktion

$$pos_{n,p} : \mathbb{M}_{n,p} \times \mathbb{N} \times \mathbb{N} \rightarrow \{0,1\}$$

eingeführt wird. Gilt $pos_{n,p}(M_{n,p}, aufl_x, aufl_y) = 0$, so ist der (n,p) -Datensatz $M_{n,p}$ bei einer Fenstergröße von $aufl_x \times aufl_y$ nicht mit der Technik darstellbar. Weitere Techniken müssen untersucht werden.

Die Funktion $pos_{n,p}(M_{n,p}, aufl_x, aufl_y)$ muß für jede Technik nur einmal (bei Aufnahme in die Menge der bekannten Visualisierungstechniken) bestimmt werden. Sie ist unabhängig von subjektiven Wahrnehmungseigenschaften des Nutzers.

- 2) Ist eine Visualisierungstechnik für einen Datensatz prinzipiell geeignet, so wird geprüft, ob das erreichbare Ergebnis der Visualisierung annähernd dem Visualisierungswunsch des Nutzers entsprechen würde. Dazu wird für jede Visualisierungstechnik eine Fähigkeitsfunktion

$$ab_{n,p} : \mathbb{M}_{n,p} \times IZ_{n,p} \rightarrow [0,1]$$

eingeführt. Diese Funktion gibt an, wie gut die Technik bei Anwendung auf einen konkreten Datensatz die einzelnen Interpretationsziele erfüllt. $ab_{n,p}(M_{n,p},(i))=0$ bedeutet, daß die Technik bei Anwendung auf $M_{n,p}$ das Interpretationsziel (i) nicht erfüllen kann. Eine bestmögliche Erfüllung liegt bei $ab_{n,p}(M_{n,p},(i))=1$ vor.

Die Funktion $ab_{n,p}$ muß zwei Regeln genügen, die für jede Visualisierungstechnik eingehalten werden müssen. Ist ein Interpretationsziel (j) aus einem Interpretationsziel (i) direkt ableitbar, so muß

$$ab_{n,p}(M_{n,p},(j)) \geq ab_{n,p}(M_{n,p},(i))$$

für jeden Datensatz $M_{n,p}$ gelten.

Ist (j) intuitiv aus (i) ableitbar, so muß

$$ab_{n,p}(M_{n,p},(j)) \geq ab_{n,p}(M_{n,p},(i)) / 10$$

für jeden Datensatz $M_{n,p}$ gelten. Der Faktor 10 ist dabei wieder willkürlich, wurde aber groß gewählt, um die Bedingung schwach (und damit allgemeingültig) zu formulieren.

Um festzustellen, ob eine Technik geeignet ist, wird der Visualisierungswunsch des Nutzers mit der Visualisierungsfähigkeit der Technik verglichen. Hierfür wird eine Abstandsfunktion $dist(w_{n,p}, ab_{n,p})$ eingeführt, die einem Visualisierungswunsch und einer Visualisierungsfähigkeit für jeden Datensatz $M_{n,p}$ eine nichtnegative real-Zahl zuordnet. Ist $dist(w_{n,p}, ab_{n,p})=0$, so stimmen Wunsch und Fähigkeit optimal überein, die Technik ist für den Datensatz und den Visualisierungswunsch geeignet. Je größer $dist(w_{n,p}, ab_{n,p})$, desto ungeeigneter ist die Technik. Die Funktion $dist$ wird im Abschnitt 4.1 eingeführt und diskutiert.

Als Ergebnis des hier beschriebenen Algorithmus erhält man für jede Visualisierungstechnik eine Zahl $dist(w_{n,p}, ab_{n,p})$. Je kleiner diese Zahl ist, desto besser ist die Technik geeignet. Damit existiert also der Eignungsbegriff für eine Technik relativ zu den anderen Techniken. (Es sind also Aussagen möglich wie "Eine Technik ist für das Problem besser geeignet als eine andere".) Notwendig sind aber absolute Aussagen zur Eignung einer Technik. ("Die Technik ist geeignet!") Hierzu muß ein Schwellwert für $dist$ eingeführt werden. Ist $dist$ größer als dieser Schwellwert, so ist die Visualisierungstechnik ungeeignet.

Die Wahl eines solchen Schwellwertes ist ein praktisches Problem. Näheres dazu in Abschnitt 6.

Zusammenfassend läßt sich sagen: Soll eine Visualisierungstechnik neu in das System zur Auswahl geeigneter Techniken aufgenommen werden, so müssen folgende Funktionen für beliebige natürliche Zahlen n und p mit $p \neq 0$ festgelegt werden:

- $pos_{n,p}$ (und hierfür pl_x und pl_y)
- $ab_{n,p}$.

4.1 Die Wahl der Abstandsfunktion dist

Die Abstandsfunktion dist soll angeben, wie weit der Visualisierungswunsch des Nutzers von der Visualisierungsfähigkeit einer Technik abweicht. Gilt für einen Datensatz $M \in \mathbb{M}_{n,p}$ für jedes Interpretationsziel $i \in IZ_{n,p}$

$$\text{ab}_{n,p}(M, (i)) \geq w_{n,p}(i),$$

so ist es naheliegend, daß $\text{dist}(w_{n,p}, \text{ab}_{n,p}) = 0$ gilt.

Die Wahl von dist soll am Beispiel eines fiktiven Systems abstrakter Interpretationsziele diskutiert werden. Für dieses System gelte $IZ_{n,p} = \{(1), (2), (3)\}$ und " $(1) \longrightarrow (2) \longrightarrow (3)$ ". Folglich gilt auch $w_{n,p}(M, (1)) \leq w_{n,p}(M, (2)) \leq w_{n,p}(M, (3))$.

Die erste Möglichkeit für eine Abstandsfunktion dist ist :

$$\text{dist}(w_{n,p}, \text{ab}_{n,p}) := \sum_{i \in IZ_{n,p}} [w_{n,p}(i) - \text{ab}_{n,p}(M, (i))]$$

Dabei sei

$$[x] := \begin{cases} 0 & \text{für } x < 0 \\ x & \text{sonst} \end{cases}$$

Diese Form der Abstandsfunktion hat den Nachteil, daß die innere Struktur des Systems abstrakter Interpretationsziele nicht mit in die Abstandsfunktion eingeht.

Beispiel:

Sei $\text{ab}_{n,p}(M, (1)) = \text{ab}_{n,p}(M, (2)) = \text{ab}_{n,p}(M, (3)) = 0.2$ und $w_{n,p}(1) = w_{n,p}(2) = w_{n,p}(3) = 0.6$. Dann liefert das Interpretationsziel (1) einen Beitrag von 0.4 zur Abstandsfunktion. Die Beiträge der Interpretationsziele (2) und (3) haben keinen eigenständigen Charakter, da sie aus den Werten für (1) abgeleitet sind: Es gilt $\text{ab}_{n,p}(M, (1)) \leq \text{ab}_{n,p}(M, (2)) \leq \text{ab}_{n,p}(M, (3))$ und $w_{n,p}(1) \leq w_{n,p}(2) \leq w_{n,p}(3)$. Dieser Tatsache wird in der Abbildungsfunktion keine Rechnung getragen, für $\text{dist}(w_{n,p}, \text{ab}_{n,p})$ würde sich der übermäßig hohe Wert 1.2 ergeben.

Eine weitere Möglichkeit für die Abstandsfunktion besteht in der Umwandlung von $w_{n,p}$ in $w_{n,p}'$ und von $\text{ab}_{n,p}$ in $\text{ab}_{n,p}'$ nach folgendem Prinzip:

Sei $(i) \in IZ_{n,p}$. Dann sei

$$\text{ubd}(i) := \{(j) \in IZ_{n,p} : "(j) \longrightarrow (i)"\}$$

$$\text{ubi}(i) := \{(j) \in IZ_{n,p} : "(j) \longrightarrow \longrightarrow (i)"\}$$

$$w_{n,p}'(i) := \begin{cases} w_{n,p}(i) & \text{für } \text{ubd}(i) = \emptyset \text{ und } \text{ubi}(i) = \emptyset \\ \min(\{w_{n,p}(i) - w_{n,p}(j) : (j) \in \text{ubd}(i)\} \cup \{w_{n,p}(i) - w_{n,p}(j)/10 : (j) \in \text{ubi}(i)\}) & \text{sonst} \end{cases}$$

$$\text{ab}_{n,p}'(M, (i)) := \begin{cases} \text{ab}_{n,p}(M, (i)) & \text{für } \text{ubd}(i) = \emptyset \text{ und } \text{ubi}(i) = \emptyset \\ \min(\{\text{ab}_{n,p}(M, (i)) - \text{ab}_{n,p}(M, (j)) : (j) \in \text{ubd}(i)\} \cup \{\text{ab}_{n,p}(M, (i)) - \text{ab}_{n,p}(M, (j))/10 : (j) \in \text{ubi}(i)\}) & \text{sonst} \end{cases}$$

Beispiel:

Sei $IZ_{n,p} = \{(1),(2),(3)\}$ und " $(1) \longrightarrow (2) \longrightarrow (3)$ ".

Für $w_{n,p}(1)=0.1$, $w_{n,p}(2)=0.4$, $w_{n,p}(3)=0.5$, $ab_{n,p}(M,(1))=0.3$, $ab_{n,p}(M,(2))=0.5$, $ab_{n,p}(M,(3))=0.9$ gilt:
 $w_{n,p}'(1)=0.1$, $w_{n,p}'(2)=0.3$, $w_{n,p}'(3)=0.1$, $ab_{n,p}'(M,(1))=0.3$, $ab_{n,p}'(M,(2))=0.2$, $ab_{n,p}'(M,(3))=0.4$

Mit $ab_{n,p}'$ und $w_{n,p}'$ soll der Abstand in der bisher üblichen Form bestimmt werden:

$$\text{dist}(w_{n,p}, ab_{n,p}) := \sum_{i \in IZ_{n,p}} [w_{n,p}'(i) - ab_{n,p}'(M,(i))]$$

Diese Form der Abstandsberechnung beachtet zwar die innere Struktur des Systems abstrakter Interpretationsziele, hat dafür aber andere Nachteile.

Beispiel:

1) $IZ_{n,p} = \{(1),(2),(3)\}$ und " $(1) \longrightarrow (2) \longrightarrow (3)$ ",

$w_{n,p}(1) = w_{n,p}(2) = w_{n,p}(3) = 0.6$, $ab_{n,p}(M,(1)) = ab_{n,p}(M,(2)) = ab_{n,p}(M,(3)) = 0.2$.

Dies ergibt $\text{dist}(w_{n,p}, ab_{n,p}) = 0.4$, was den Erwartungen entspricht.

2) $w_{n,p}(1)=0.2$, $w_{n,p}(2)=0.4$, $w_{n,p}(3)=0.7$, $ab_{n,p}(M,(1))=0.6$, $ab_{n,p}(M,(2))=0.6$, $ab_{n,p}(M,(3))=0.6$ ergibt

$w_{n,p}'(1)=0.2$, $w_{n,p}'(2)=0.2$, $w_{n,p}'(3)=0.3$, $ab_{n,p}'(M,(1))=0.6$, $ab_{n,p}'(M,(2))=0$, $ab_{n,p}'(M,(3))=0$

und $\text{dist}(w_{n,p}, ab_{n,p}) = 0.5$, was den Erwartungen ($\text{dist}(w_{n,p}, ab_{n,p}) = 0$) völlig widerspricht. Für diesen Fall versagt diese Form der Abstandsberechnung.

Brauchbare Ergebnisse liefert eine Synthese der beiden bisher eingeführten Abstandsfunktionen:

$$\text{dist}(w_{n,p}, ab_{n,p}) := \sum_{i \in IZ_{n,p}} \min \{ [w_{n,p}(i) - ab_{n,p}(M,(i))] , [w_{n,p}'(i) - ab_{n,p}'(M,(i))] \}$$

Diese Abstandsfunktion, die die Vorteile der beiden vorher eingeführten in sich vereinigt, soll im folgenden verwendet werden.

Bemerkung:

Wird ein unstrukturiertes System abstrakter Interpretationsziele genutzt (siehe z.B. /Luko93/), so stimmt die hier eingeführte Abstandsfunktion mit der in /Luko93/ verwendeten Form überein.

5. Beispiele, Tests, Ergebnisse

In diesem Abschnitt sollen praktische Aspekte des neu eingeführten Algorithmus zur automatischen Auswahl geeigneter Visualisierungstechniken betrachtet werden. Außerdem sollen Beispiele für konkrete Anwendungsfälle und konkrete Visualisierungstechniken gegeben werden.

5.1 Charakteristika konkreter Visualisierungstechniken

In diesem Abschnitt sollen für konkrete Visualisierungstechniken der Platzbedarf, die Möglichkeitsfunktionen und die Fähigkeitsfunktionen bestimmt werden.

Platzbedarf und Möglichkeitsfunktionen sind unabhängig von subjektiven Wahrnehmungseigenschaften des Nutzers. Anders die Fähigkeitsfunktionen. Diese können stark vom konkreten Nutzer abhängen. Zur Bestimmung der Fähigkeitsfunktionen sind zwei grundlegende Herangehensweisen möglich:

1) Die "objektive" Herangehensweise.

Einige Fähigkeiten können von einer Technik unabhängig von den Wahrnehmungseigenschaften des Nutzers gut erfüllt werden. (Werden beispielsweise in der Visualisierung Häufigkeiten direkt codiert, so können diese auch unabhängig vom konkreten Nutzer gut erkannt werden.) Für alle Eigenschaften, die nicht von allen Nutzern erkannt werden können, wird die Technik als unfähig gesetzt. Es ergibt sich eine minimale Fähigkeitsfunktion, die zwar für beliebige Nutzer gilt, jedoch recht schwache Ergebnisse liefert.

2) Die "subjektive" Herangehensweise.

Die Fähigkeitsfunktionen werden auf die Wahrnehmungseigenschaften eines konkreten Nutzers bezogen. Es entstehen optimal angepaßte Fähigkeitsfunktionen. Nachteilig ist, daß für jeden Nutzer seine eigenen Funktionen bestimmt werden müßten. Der dafür notwendige Aufwand macht die reine Anwendung dieser Herangehensweise praktisch unmöglich.

Es soll hier ein Kompromiß beider Herangehensweisen verwendet werden. Für alle Nutzer werden dieselben Fähigkeitsfunktionen verwendet. Diese enthalten neben den "objektiven" Eigenschaften des Nutzers auch solche "subjektiven", von denen erwartet werden kann, daß sie für die meisten Nutzer gelten.

Zur Bestimmung der Fähigkeitsfunktionen sind also praktische Erfahrungen über allgemeine subjektive Wahrnehmungseigenschaften notwendig. Solche Erfahrungen liegen teilweise vor in /Luko93/, /Arnd92/ und /Schw93/. Die hier eingeführten Fähigkeitsfunktionen stützen sich auf die dort beschriebenen Erfahrungen.

Für die Beschreibung der Möglichkeits- und Fähigkeitsfunktionen ist es sinnvoll, Variablen für bestimmte Charakteristika des Datensatzes einzuführen. Die Werte für diese Variablen

sind -falls sie nicht schon bekannt sind- problemlos automatisch aus dem Datensatz in Matrixform zu ermitteln. Solche Variablen sind:

- n: die Dimensionalität des Gitters
- p: die Anzahl der in einem Gitterpunkt beobachteten Merkmale
- anzska := $|\{x \in V: |x|=1\}|$: die Anzahl aller skalaren Merkmale
- maxwb := $\max \{wb_i: \{i\} \in V\}$: der maximale Umfang der Wertebereiche aller skalaren Merkmale
- minwb := $\min \{wb_i: \{i\} \in V\}$: der minimale Umfang der Wertebereiche aller skalaren Merkmale
- Z: die Anzahl der Beobachtungsfälle
- anztup(i): die Anzahl von i-Tupeln von Merkmalen mit relevanter gemeinsamer Information
- maxgitwb: der maximale Wertebereich einer das Gitter definierenden Beobachtungsgröße

Im folgenden sollen Beispiele für Möglichkeits- und Fähigkeitsfunktionen gegeben werden. Dabei sollen nur diejenigen Interpretationsziele (i) aufgeschrieben werden, für die $ab_{n,p}(M,(i)) > 0$ für ein bestimmtes $M \in \mathbb{M}_{n,p}$ gilt. Gilt "(i) \longrightarrow (j)" und ist $ab_{n,p}(M,(j))$ nicht explizit gegeben, so gilt $ab_{n,p}(M,(j)) = ab_{n,p}(M,(i))$ für beliebige $M \in \mathbb{M}_{n,p}$.

Beispiele:

1) parallele Koordinaten (s. /Fins91/)

Bei dieser Technik werden Koordinatenachsen parallel angeordnet, jede Koordinatenachse entspricht einem beobachteten Merkmal. Für jeden Beobachtungsfall werden die Punkte auf benachbart liegenden Koordinatenachsen durch Strecken verbunden. Ein Beobachtungsfall entspricht also einem Streckenzug, dessen Eckpunkte auf den Koordinatenachsen liegen.

$$pl_x(M) = 20 * p$$

Ein Streckenzug soll als solcher erkennbar sein, wenn er mindestens über 19 Pixel in x-Richtung geht. Der minimal mögliche Abstand der Koordinatenachsen sei also mit 19 Pixeln festgelegt.

$$pl_y(M) = \max wb$$

Der Platzbedarf in y-Richtung entspricht dem maximalen Umfang der Wertebereiche der einzelnen Beobachtungsgrößen.

$$pos_{n,p}(M, \text{auflx}, \text{aufly}) = \begin{cases} 1 & \text{für auflx} > pl_x(M) \\ & \text{und aufly} > pl_y(M) \\ & \text{und anzska} \geq 2 \\ 0 & \text{sonst} \end{cases}$$

Parallele Koordinaten sind nur sinnvoll für skalare Merkmale. Merkmale in Form von Vektoren oder Tensoren können mit parallelen Koordinaten nicht visualisiert werden. Es müssen also mindestens 2 skalare Merkmale gegeben sein, um die Technik überhaupt anwenden zu können.

$$ab_{n,p}(M, (13.0.j)) = \begin{cases} 1 - Z / (2 * \min wb) & \text{für } j \leq \text{anzska} \\ & \text{und } Z \leq 2 * \min wb \\ 0 & \text{sonst} \end{cases}$$

Die Technik ist überhaupt nur sinnvoll anwendbar, wenn nur ein Gitterpunkt betrachtet wird, da räumliche Beziehungen bei parallelen Koordinaten nicht beachtet werden. Außerdem müssen alle zu visualisierenden Merkmale skalar sein, Vektoren oder Tensoren sind nicht visualisierbar. Je mehr Beobachtungsfälle existieren, desto unübersichtlicher wird das Bild, die Möglichkeit der Identifizierbarkeit wird geringer. Gegenläufig wirkt der Umfang der Wertebereiche der skalaren Merkmale. Je geringer dieser Wert ist, desto mehr Streckenzüge gehen von einem Punkt aus, desto geringer ist die Möglichkeit der Identifizierbarkeit.

Es wird hier davon ausgegangen, daß eine Identifizierbarkeit nicht mehr gegeben ist, wenn die Anzahl der Beobachtungsfälle (also die Anzahl der Streckenzüge) den Wert $2 \cdot \min_{wb}$ übersteigt.

$$ab_{n,p}(M,(18.0.0)) = \begin{cases} 1 & \text{für } anz\text{tup}(2) \leq anz\text{ska}/2 \\ & \text{und } anz\text{tup}(3) = 0 \\ 2 - (2 \cdot anz\text{tup}(2)) / anz\text{ska} & \text{für } anz\text{ska}/2 < anz\text{tup}(2) \leq anz\text{ska} \\ & \text{und } anz\text{tup}(3) = 0 \\ 0 & \text{sonst} \end{cases}$$

Korrelationen können nur dann erkannt werden, wenn es höchstens Paare von Merkmalen mit relevanter gemeinsamer Information gibt. Die entsprechenden Koordinatenachsen müssen dann in der Visualisierung benachbart angeordnet werden. Maximal sind also $anz\text{ska} - 1$ Korrelationspaare zu erkennen.

Bemerkung:

In dieser knappen Form ist bereits die gesamte Fähigkeitsfunktion für parallele Koordinaten beschrieben. Durch die direkte Ableitbarkeit ergibt sich

$$ab_{n,p}(M,(14.i.j)) = ab_{n,p}(M,(13.i.j)) ,$$

durch intuitive Ableitbarkeit ergibt sich

$$ab_{n,p}(M,(15.0.0.j)) = ab_{n,p}(M,(13.0.j)) / 10$$

$$ab_{n,p}(M,(16.0.0.j)) = ab_{n,p}(M,(13.0.j)) / 10$$

$$ab_{n,p}(M,(23.0.0.j)) = ab_{n,p}(M,(13.0.j)) / 10$$

2) shape coding I (s. /Bedd90/)

Bei dieser Technik werden für jeden Beobachtungsfall die Werte der einzelnen Merkmale in eine Ikone abgebildet und dort farbcodiert. Bei der hier betrachteten Variante des Verfahrens werden räumliche Zusammenhänge nicht beachtet, der Position der Ikone im Bild wird keine Bedeutung beigemessen.

$$pl_x(M) = pl_y(M) = 2 \cdot \sqrt{Z \cdot anz\text{ska}}$$

Zur Farbcodierung eines Wertes soll mindestens ein Bereich von 4 Pixeln zur Verfügung stehen. In jeder Ikone werden maximal $anz\text{ska}$ Werte codiert. Z Ikonen müssen visualisiert werden.

$$pos_{n,p}(M,auf\text{x},auf\text{y}) = \begin{cases} 1 & \text{für } auf\text{x} > pl_x(M) \\ & \text{und } auf\text{y} > pl_y(M) \\ & \text{und } anz\text{ska} \geq 1 \\ 0 & \text{sonst} \end{cases}$$

Die Technik ist nur sinnvoll für skalare Merkmale.

$$ab_{n,p}(M,(13.0.j)) = \begin{cases} 1 - (\max\text{wb}/100) & \text{für } j \leq anz\text{ska} \text{ und } \max\text{wb} \leq 10 \\ 0 & \text{sonst} \end{cases}$$

Die Technik ist nur sinnvoll auf einen Gitterpunkt anwendbar, räumliche Beziehungen werden nicht beachtet. Alle zu visualisierenden Merkmale müssen skalar sein. Es wird davon ausgegangen, daß bis zu 10 verschiedenen Farben/Graustufen unterscheidbar sind.

$$ab_{n,p}(M,(18.0.0)) = 0.8$$

Das Erkennen von Korrelationen ist intuitiv über Mustererkennung möglich.

3) shape coding II (s. /Bedd90/)

Das gleiche Prinzip wie shape coding I, nur daß über die Lage der Ikone auf dem Bildschirm die räumliche Lage im 2D-Gitter codiert wird. Sind an einem Gitterpunkt mehrere Beobachtungsfälle, so wird eine Ikone mit dem Mittelwert aller dort auftretenden Beobachtungsfälle visualisiert. Existiert für einen Gitterpunkt kein Beobachtungsfall, so wird in der Visualisierung der entsprechende Platz mit der Hintergrundfarbe belegt.

$$pl_x(M) = pl_y(M) = 2 * \max_{\text{gitwb}} * \sqrt{\text{anzska}}$$

Für jeden Gitterpunkt soll Platz zur Visualisierung mindestens einer Ikone sein. In einer Ikone werden anzska verschiedene Werte farbcodiert.

$$pos_{n,p}(M, \text{auflx}, \text{aufly}) = \begin{cases} 1 & \text{für auflx} > pl_x(M) \\ & \text{und aufly} > pl_y(M) \\ & \text{und } n \geq 2 \\ & \text{und anzska} \geq 1 \\ 0 & \text{sonst} \end{cases}$$

Die Technik ist nur dann sinnvoll anwendbar, wenn ein mindestens 2-dimensionales Gitter vorliegt und mindestens eine skalare Größe zu visualisieren ist.

$$ab_{n,p}(M, (13.2.j)) = \begin{cases} 1 - (\max_{\text{wb}}/100) & \text{für } j \leq \text{anzska} \\ & \text{und } \max_{\text{wb}} \leq 10 \\ & \text{und M nicht überdefiniert} \\ 0 & \text{sonst} \end{cases}$$

Eine Identifizierung einzelner Werte ist nur möglich, wenn der Datensatz nicht überdefiniert ist.

$$ab_{n,p}(M, (15.2.0.k)) = \begin{cases} 1 & \text{für } k \leq \text{anzska} \\ 0 & \text{sonst} \end{cases}$$

An jedem Punkt einer 2-dimensionalen Gitterpunktmenge können die Mittelwerte der in diesen Punkten beobachteten Größen gut erkannt werden.

$$ab_{n,p}(M, (19.0.2.k)) = \begin{cases} 1 & \text{für } k \leq \text{anzska} \\ 0 & \text{sonst} \end{cases}$$

Die Korrelationen von Datenwerten zur Lage im Raum kann gut erkannt werden.

4) dimensional stacking (s. /LeBl90/)

Bei dieser Technik werden die Merkmale zu Paaren geordnet und "ineinandergeschachtelt". Für jede mögliche Wertekombination wird die Häufigkeit ihres Auftretens farbcodiert.

$$pl_x(M) = pl_y(M) = 2 * \sqrt{\max_{\text{wb}}^{\text{anzska}}}$$

Für jede mögliche Wertekombination soll ein Platz von mindestens 4 Pixeln zur Verfügung stehen.

$$pos_{n,p}(M, \text{auflx}, \text{aufly}) = \begin{cases} 1 & \text{für auflx} > pl_x(M) \\ & \text{und aufly} > pl_y(M) \\ & \text{und anzska} \geq 2 \\ 0 & \text{sonst} \end{cases}$$

Die Technik ist nur sinnvoll anwendbar für skalare Merkmale. Da diese zu Paaren geordnet werden, müssen mindestens 2 existieren.

$$ab_{n,p}(M, (15.0.0.k)) = \begin{cases} 0.1 & \text{für } k \leq \text{anzska} \\ 0 & \text{sonst} \end{cases}$$

Ein Erkennen von Mittelwerten ist intuitiv möglich.

$$ab_{n,p}(M, (16.0.0.k)) = \begin{cases} 1 & \text{für } k \leq \text{anzska} \\ 0 & \text{sonst} \end{cases}$$

Häufigkeiten sind sehr gut erkennbar, da sie direkt farbcodiert sind.

$$ab_{n,p}(M,(18.0.0)) = 0.8$$

Korrelationen zwischen den Merkmalen sind auch bei einer höheren Anzahl von Merkmalen recht gut erkennbar.

5) marching cube (s. /Lore87/)

Bei dieser Technik werden Isoflächen im Datenvolumen gewonnen. Ein "wandernder Würfel" durchläuft die Zellen und ermittelt diejenigen, die von der Isofläche geschnitten werden. Bei diesen werden die geschnittenen Kanten bestimmt und die Isoflächen durch Dreiecke approximiert.

$$pl_x(M) = pl_y(M) = 5 * \max_{\text{gitwb}}$$

In der Visualisierung sollen zwischen zwei Gitterpunkten mindestens 5 Pixel Platz sein, um ein approximiertes Dreieck erkennen zu können

$$pos_{n,p}(M, \text{auflx}, \text{aufly}) = \begin{cases} 1 & \text{für} & \text{auflx} > pl_x(M) \\ & & \text{und aufly} > pl_y(M) \\ & & \text{und } n \geq 3 \\ 0 & \text{sonst} \end{cases}$$

Die Technik ist nur sinnvoll für eine mindestens 3-dimensionale Gitterpunktmenge.

$$ab_{n,p}(M,(14.3.1)) = 0.5$$

$$ab_{n,p}(M,(14.2.1)) = 1$$

Einen bestimmten Wert an allen Gitterpunkten eines 3D-Schnittes zu sehen, ist teilweise möglich, da die Rückseite einer Isofläche verdeckt ist. Für einen 2D-Schnitt ist dies jedoch problemlos möglich.

$$ab_{n,p}(M,(19.0.3.1)) = 0.5$$

Korrelationen von Datenwerten zur Lage im Raum sind teilweise erkennbar.

$$ab_{n,p}(M,(20)) = 1$$

Eventuell auftretende unbekannte Werte an Gitterpunkten müssen stets interpoliert werden.

$$ab_{n,p}(M,(22.3)) = 0.5$$

$$ab_{n,p}(M,(22.2)) = 1$$

Ein unbekannter Datenwert zwischen den Gitterpunkten ist für einen 3D-Schnitt teilweise, für einen 2D-Schnitt vollständig sichtbar.

6) Raycasting

Bei dieser Technik werden Strahlen auf ihrem Weg durch das Volumen verfolgt. Entlang eines solchen Strahls werden die Werte der benachbart liegenden Gitterpunkte ausgewertet. Dargestellt wird der Mittelwert aller in der Nähe des Strahls gefundenen Datenwerte.

$$pl_x(M) = pl_y(M) = \max_{\text{gitwb}}$$

Die Technik ist sinnvoll anwendbar, wenn für jeden Gitterpunkt eines 2D-Schnittes mindestens ein Pixel zur Verfügung steht.

$$pos_{n,p}(M, \text{auflx}, \text{aufly}) = \begin{cases} 1 & \text{für} & \text{auflx} > pl_x(M) \\ & & \text{und aufly} > pl_y(M) \\ & & \text{und } n \geq 3 \\ 0 & \text{sonst} \end{cases}$$

Die Technik ist nur sinnvoll für eine mindestens 3-dimensionale Gitterpunktmenge.

$$ab_{n,p}(M,(15.2.1.1)) = 1$$

Die Werte jeweils eines 1D-Schnittes werden zu einem Mittelwert zusammengefaßt.

$$ab_{n,p}(M,(19.0.3.1)) = 0.5$$

Korrelationen zwischen Datenwerten und Lage im 3D-Raum sind teilweise erkennbar.

7) Slicing-Methode

Der Gitterraum wird mit 2D-Ebenen geschnitten, die entstehenden Polygonflächen werden visualisiert. Existiert für einen Gitterpunkt mehr als ein Beobachtungsfall, wird der Mittelwert aller dort auftretenden Datenwerte visualisiert.

$$pl_x(M) = pl_y(M) = \max_{\text{gitwb}}$$

Für jeden Gitterpunkt eines Schnittes soll mindestens ein Pixel zur Verfügung stehen.

$$pos_{n,p}(M, \text{auflx}, \text{aufly}) = \begin{cases} 1 & \text{für} \quad \text{auflx} > pl_x(M) \\ & \text{und aufly} > pl_y(M) \\ & \text{und } n \geq 2 \\ 0 & \text{sonst} \end{cases}$$

Es muß eine mindestens 2-dimensionale Gitterpunktmenge vorliegen, um die Technik sinnvoll anwenden zu können.

$$ab_{n,p}(M, (13.2.1)) = \begin{cases} 1 & \text{für } M \text{ nicht überdefiniert} \\ 0 & \text{sonst} \end{cases}$$

Auf 2-dimensionalen Schnitten ist die Identifizierbarkeit eines Merkmals sehr gut möglich.

$$ab_{n,p}(M, (15.2.0.1)) = 1$$

In jedem Punkt des 2D-Schnittes sind die Mittelwerte aller dort auftretenden Datenwerte eines Merkmals erkennbar.

$$ab_{n,p}(M, (19.0.2.1)) = 1$$

Im 2D-Schnitt können Korrelationen zwischen den Datenwerten und der Lage in der Ebene erkannt werden.

$$ab_{n,p}(M, (20)) = 1$$

$$ab_{n,p}(M, (22.2)) = 1$$

Unbekannte Datenwerte an Gitterpunkten werden interpoliert. Unbekannte Datenwerte zwischen den Gitterpunkten im 2D-Schnitt werden ebenfalls interpoliert.

An den hier gebrachten Beispielen sind zwei Dinge besonders auffällig. Zum einen ist die Beschreibung der Fähigkeitsfunktionen in recht kurzer Form möglich. Viele Fähigkeiten ergeben sich durch direkte oder intuitive Ableitbarkeit. Zum anderen fällt auf, daß geringfügige Modifikationen der Visualisierungstechniken zu völlig neuen Fähigkeits- und Möglichkeitsfunktionen führen können (s. z.B. shape coding I und II).

5.2 Beispiele für die Anwendung des Auswahlalgorithmus

- 1) Für den in /Arnd92/ behandelten mikrobiologischen Datensatz ($n=0$, $p=41$) möchte der Nutzer das Auftreten von Merkmalsmustern (welche und wieviele verschiedene Muster, wie verteilen sich die Isolate darauf) erkennen. Dabei sollen maximal 10 Merkmale betrachtet werden. Nach Abschnitt 3.2 ergibt das den Visualisierungswunsch

$$w_{0,41}(13.0.10) = 0.8$$

$$w_{0,41}(16.0.0.10) = 1$$

Für die in Abschnitt 5.1 eingeführten Visualisierungstechniken ergeben sich für die Fähigkeits- und Möglichkeitsfunktionen folgende Werte:

- parallele Koordinaten :

$$ab_{n,p}(M,(13.0.10)) = 0, \text{ da } Z = 343 > 2 * \min_{wb} = 2 * 2$$

$$ab_{n,p}(M,(16.0.0.10)) = 0$$

Dies ergibt

$$\text{dist}(w_{n,p}, ab_{n,p}) = 1.8$$

- shape coding I :

$$ab_{n,p}(M,(13.0.10)) = 1 - 10/100 = 0.9$$

$$ab_{n,p}(M,(16.0.0.10)) = 0.09$$

Dies ergibt

$$\text{dist}(w_{n,p}, ab_{n,p}) = 0.91$$

- shape coding II :

$$\text{pos}_{n,p}(M, \text{auflx}, \text{aufly}) = 0, \text{ da } n < 2 .$$

- dimensional stacking :

$$pl_x(M) = 2 * \sqrt{10^{10}}$$

Die Technik ist für 10 Merkmale nicht anwendbar, der Platzbedarf ist zu hoch.

- marching cube :

$$\text{pos}_{n,p}(M, \text{auflx}, \text{aufly}) = 0, \text{ da } n < 3 .$$

- Raycasting :

$$\text{pos}_{n,p}(M, \text{auflx}, \text{aufly}) = 0, \text{ da } n < 3 .$$

- Slicing-Methode :

$$\text{pos}_{n,p}(M, \text{auflx}, \text{aufly}) = 0, \text{ da } n < 2 .$$

Es ergibt sich, daß von den hier betrachteten Techniken shape coding I am geeignetsten ist.

2) Für einen in /Schw93/ behandelten Volumendatensatz aus der Medizin (Diagnose von Leberkrankheiten) soll ein Überblick über die Homogenität im Inneren der Leber gewonnen werden. Es gilt $n=3$ und $p=1$. Nach Abschnitt 3.2 ergibt sich der Visualisierungswunsch

$$w_{3,1}(15.2.1.1) = 0.8$$

$$w_{3,1}(19.0.3.1) = 0.8$$

$$w_{3,1}(23.2.1.1) = 1$$

Für die einzelnen Visualisierungstechniken gilt:

- parallele Koordinaten :

$$\text{pos}_{n,p}(M, \text{auflx}, \text{aufly}) = 0, \text{ da } \text{anzska} < 2$$

- shape coding I :

$$ab_{n,p}(M,(15.2.1.1)) = 0$$

$$ab_{n,p}(M,(19.0.3.1)) = 0$$

$$ab_{n,p}(M,(23.2.1.1)) = 0$$

Dies ergibt

$$\text{dist}(w_{n,p}, ab_{n,p}) = 2.6$$

- shape coding II :

$$ab_{n,p}(M,(15.2.1.1)) = 0$$

$$ab_{n,p}(M,(19.0.3.1)) = 0$$

$$ab_{n,p}(M,(23.2.1.1)) = 0$$

Dies ergibt

$$\text{dist}(w_{n,p}, ab_{n,p}) = 2.6$$

- dimensional stacking :

$$\text{pos}_{n,p}(M, \text{auflx}, \text{aufly}) = 0, \text{ da } \text{anzska} < 2$$

- marching cube :
 - $ab_{n,p}(M,(15.2.1.1)) = 0$
 - $ab_{n,p}(M,(19.0.3.1)) = 0.5$
 - $ab_{n,p}(M,(23.2.1.1)) = 0$
 Dies ergibt
 - $dist(w_{n,p},ab_{n,p}) = 2.1$

- Raycasting
 - $ab_{n,p}(M,(15.2.1.1)) = 1$
 - $ab_{n,p}(M,(19.0.3.1)) = 0.5$
 - $ab_{n,p}(M,(23.2.1.1)) = 0$
 Dies ergibt
 - $dist(w_{n,p},ab_{n,p}) = 1.3$

- Slicing- Methode
 - $ab_{n,p}(M,(15.2.1.1)) = 0$
 - $ab_{n,p}(M,(19.0.3.1)) = 0$
 - $ab_{n,p}(M,(23.2.1.1)) = 0$
 Dies ergibt
 - $dist(w_{n,p},ab_{n,p}) = 2.6$

Es ergibt sich, daß die Raycasting- Methode am geeignetsten ist.

- 3) Für denselben Datensatz wie in 2) sollen Größenrelationen auf Schnitten durch das Lebergewebe analysiert werden. Dies ergibt nach Abschnitt 3.2 den Visualisierungswunsch

$$w_{3,1}(9.2) = 0.8$$

$$w_{3,1}(19.0.2.1) = 0.8$$

Für die einzelnen Techniken gilt:

- parallele Koordinaten :
 - $pos_{n,p}(M,auflx,aufly) = 0$, da $anzska < 2$
- shape coding I :
 - $ab_{3,1}(M,(9.2)) = 0$
 - $ab_{3,1}(M,(19.0.2.1)) = 0$
 Dies ergibt
 - $dist(w_{n,p},ab_{n,p}) = 1.6$
- shape coding II :
 - $ab_{3,1}(M,(9.2)) = 0$
 - $ab_{3,1}(M,(19.0.2.1)) = 1$
 Dies ergibt
 - $dist(w_{n,p},ab_{n,p}) = 0.8$
- dimensional stacking
 - $pos_{n,p}(M,auflx,aufly) = 0$, da $anzska < 2$
- marching cube :
 - $ab_{3,1}(M,(9.2)) = 0$
 - $ab_{3,1}(M,(19.0.2.1)) = 0$
 Dies ergibt
 - $dist(w_{n,p},ab_{n,p}) = 1.6$
- Raycasting :
 - $ab_{3,1}(M,(9.2)) = 0$
 - $ab_{3,1}(M,(19.0.2.1)) = 0$
 Dies ergibt
 - $dist(w_{n,p},ab_{n,p}) = 1.6$
- Slicing- Methode :

$$ab_{3,1}(M,(9.2)) = 1$$

$$ab_{3,1}(M,(19.0.2.1)) = 1$$

Dies ergibt

$$\text{dist}(w_{n,p}, ab_{n,p}) = 0$$

Es ergibt sich, daß die Slicing- Methode optimal geeignet ist.

6. Ungenauigkeiten, Unkorrektheiten

Gegeben ist ein praktisches Problem, welches in der natürlichen Sprache formuliert ist. Damit ist die Verwendung von mehrdeutigen und nicht exakt definierten Begriffen möglich, das Problem kann eine für Maschinen nicht faßbare Unexaktheit, Mehrdeutigkeit und Subjektivität bekommen.

Soll ein solches reales Problem trotzdem maschinell bearbeitet werden, so ist es nötig, es in ein mathematisches Modell zu pressen. Dieses Modell kann noch so kompliziert und filigran gearbeitet sein. Immer stellt es eine Vereinfachung des realen Problems dar.

Es ist also wenig sinnvoll zu fragen, ob ein geschaffenes Modell "richtig" oder "falsch" ist. Der einzig praktikable Gradmesser für die Güte eines Modells ist die Nachvollziehung der bisher mit dem praktischen Problem gemachten Erfahrungen.

Zu einem praktischen Problem gibt es meist praktische Erfahrungen und Faustregeln zur Lösung. Außerdem können Teilprobleme des allgemeinen Problems bereits formalisiert und die Ergebnisse der maschinellen Bearbeitung als praxisrelevant anerkannt sein. Ein mathematisches Modell soll als gültig betrachtet werden, wenn es alle bisher vorliegenden Ergebnisse und Erfahrungen widerspiegelt. Liefert das Modell zusätzlich noch neue Ergebnisse, die nach Tests und Prüfungen zu neuen anerkannten Erfahrungen führen, so war die Schaffung des neuen mathematischen Modells eine lohnenswerte Arbeit.

Da die Gültigkeit eines Modells nur bezüglich bisher vorliegender praktischer Erfahrungen festgelegt werden kann, kann auf Grund neu gewonnener Erfahrung ein Modell jederzeit wieder ungültig werden. Allerdings wird mit der Menge der bereits gemachten Erfahrungen die Wahrscheinlichkeit geringer, neue Erfahrungen zu finden, die ein Modell vollständig ungültig machen.

Das in dieser Arbeit beschriebene neue Modell zur Auswahl geeigneter Visualisierungstechniken basiert auf den Erfahrungen, die in /Schw93/, /Luko93/, /Will93/, /Arnd92/ und /Thei93/ beschrieben sind. Hinzu kommen allgemeine Erfahrungen aus /Wehr90/, /Fins91/, /LeBl90/, /Bedd90/ und /Robe90/, die im Modell beachtet werden sollen.

Es sollen nun die einzelnen Komponenten des Modells näher untersucht werden.

1) Die Datenbeschreibung

Die Beschreibung des Datensatzes stellt bezüglich der Gültigkeit des Modells keine Schwierigkeit dar, da es auch im praktischen Problem um die Visualisierung maschinell ermittelter Daten geht. Jede vollständige Beschreibung des Datensatzes kann für ein gültiges mathematisches Modell verwendet werden, so auch die in dieser Arbeit verwendete Beschreibung in Matrixform.

2) Das System abstrakter Interpretationsziele

In /Schu93/ wurden verschiedene Klassen von Interpretationszielen verbal formuliert. Auf Grundlage dieser Klassen wurde das hier vorliegende System abstrakter Interpretationsziele geschaffen. Erweiterungen dieses Systems wären in zwei Richtungen möglich:

- In der Praxis können neue verbal formulierte Klassen von Interpretationszielen auftreten. Formalisierte Formen dieser neuen Ziele können neu in das System aufgenommen werden.

- Für die vorhandenen Klassen von Interpretationszielen kann die Formalisierung verfeinert werden, wenn dies nötig erscheint. Beispielsweise kann das Interpretationsziel (22.i) weiter verfeinert werden, indem unterschieden wird, wie die Interpolation zwischen den Gitterpunkten erfolgen soll.

Da beide Formen der Erweiterung jederzeit möglich sind, ist es wenig sinnvoll, die Vollständigkeit eines solchen Systems abstrakter Interpretationsziele zu untersuchen. Das in dieser Arbeit eingeführte System ermöglicht aber die Formalisierung aller bisher aus den verschiedenen Anwendungsgebieten betrachteten verbalen Interpretationsziele.

3) Der Algorithmus zur Auswahl geeigneter Techniken

Der Algorithmus schließt alle Vorteile des Ansatzes aus /Wehr90/ in sich ein. Einige offenkundige Mängel dieses Ansatzes sind abgestellt. Ebenfalls eingeschlossen ist der Ansatz aus /Luko93/. Dieser Ansatz kann als Vorläufer oder Spezialfall des hier beschriebenen betrachtet werden.

Die Wahl des Schwellwertes, ab welchem eine Technik als ungeeignet gilt, orientiert sich an praktischen Erfahrungen. Mit einem Wert von 0.5 ergeben sich ähnliche praktische Ergebnisse wie in /Schw93/, /Arnd92/ und /Luko93/. Neue praktische Erfahrungen können diesen Wert jedoch auch wieder verändern.

4) Die Wahl der Abstandsfunktion dist

Für diese Komponente existieren am wenigsten Erfahrungen. Entsprechend groß ist auch die Möglichkeit, besser geeignete Abstandsfunktionen zu finden. Die hier eingeführte Form liefert jedoch brauchbare Ergebnisse für die bisher bekannten Fälle.

Insgesamt läßt sich feststellen: Der in dieser Arbeit vorgestellte Algorithmus erfüllt alle Eigenschaften der Spezialfälle aus /Arnd92/, /Schw93/ und /Luko93/. Ob sich der Algorithmus auch für neue "gemischte" Datensätze (also Datensätze, die sich nicht klar bestimmten Datenklassen zuordnen lassen) bewährt, kann nur abgewartet werden.

7. Ausblick

Um den hier eingeführten Algorithmus zu Implementation und Anwendung zu bringen, müssen für viele verschiedene Visualisierungstechniken die Möglichkeits- und Fähigkeitsfunktionen bestimmt werden. Der Algorithmus selbst ist in folgende Richtungen erweiterbar:

- Verwendung dynamischer Platzanforderungen

In der hier vorliegenden Form des Algorithmus werden pl_x und pl_y für einen Datensatz M bestimmt, ohne zu beachten, ob eventuell nur bestimmte Teile von M (Schnitte, Merkmalsklassen) visualisiert werden können. Dies kann zu geringeren Platzanforderungen führen als in pl_x und pl_y für den gesamten Datensatz bestimmt.

- Verwendung der Komponente Zeit

Ist für eine Technik ein Bild in Echtzeit erstellbar, so kann eine Bildsequenz zur Visualisierung genutzt werden. Dies führt zu einer verbesserten Möglichkeitsfunktion für die Technik.

Beispiel:

Erfüllt eine Technik das Interpretationsziel (9.i), so ist durch eine Sequenz durch die (i+1)-te Dimension auch das Interpretationsziel (9.i+1) erfüllbar. (i=1,...,n-2)

- Schaffung von Verfahren zur automatischen Abbildung konkreter Interpretationsziele auf abstrakte

Dies muß für jedes Anwendungsgebiet extra erfolgen, da jedes Anwendungsgebiet eine eigene Sprache und eigene Erfahrungen zur Beschreibung konkreter Interpretationsziele hat.

Mit dem hier angegebenen Algorithmus können für eine konkrete Anwendung geeignete Visualisierungstechniken ausgewählt werden. Ein weiteres Problem ist noch, wie die einzelnen Merkmale eines Datensatzes auf die Parameter einer ausgewählten Visualisierungstechnik abzubilden sind. Erste Ansätze dazu finden sich in /Luko93/.

Literatur:

- /Arnd92/ S. Arndt: Visualisierung mikrobiologischer Daten, Diplomarbeit, Universität Rostock, Fachbereich Informatik, 1992
- /Arnd93/ S. Arndt: Spezifizierung multivariater Daten zum Zwecke einer systematischen Auswahl ihrer visuellen Repräsentation, Preprint CS-06-93, Universität Rostock, Fachbereich Informatik, 1993
- /Bedd90/ J. Beddow: Shape Coding of Multidimensional Data on a Microcomputer Display, Tagungsband "Visualization '90", San Francisco, 1990
- /Berg89/ R. D. Bergeron, G.G Grinstein: A Reference Model for the Visualization of Multi-dimensional Data, Proceedings Eurographics '89, Hamburg, 1989
- /Brod92/ K. W. Brodlie u.a.: Scientific Visualization, Springer-Verlag, Berlin, 1992
- /Fins91/ R. Finsterwalder: Entscheidungsunterstützende Visualisierung bei der mehrzieligen Entwurfsoptimierung, Workshop Visualisierungstechniken, Stuttgart, 1991
- /LeBl90/ J. LeBlanc, M. O. Ward, N. Wittels: Exploring N-Dimensional Databases, Tagungsband "Visualization '90", San Francisco, 1990
- /Lore87/ W. E. Lorensen, H. E. Cline: Marching Cubes: A High Resolution 3D Surface Construction Algorithm, Computer Graphics, Vol. 21, No 4, 1987
- /Luko93/ K. Lukoschek: Die Expressivität als ein Kriterium für die Wirksamkeit visueller Repräsentation, Preprint CS-07-93, Universität Rostock, Fachbereich Informatik, 1993
- /Robe90/ P. K. Robertson: A Methodology for Scientific Visualization: Choosing Representations Based on a Natural Scene Paradigm, Tagungsband "Visualization '90", San Francisco, 1990
- /Schu93/ H. Schumann: Entscheidungshilfen zur Visualisierung wissenschaftlicher Daten, Rostocker Informatik-Berichte, 13(1992), S. 62-69
- /Schw93/ M. Schwenck: Visualisierung von Volumendaten, Diplomarbeit, Universität Rostock, Fachbereich Informatik, 1993

- /Thei93/ H. Theisel: Analyse mehrdimensionaler Daten mit Hilfe der Informationstheorie, Rostocker Informatik-Berichte, 14(1993), S. 86-96
- /Wehr90/ S. Wehrend, C. Lewis: A Problem-oriented Classification of Visualization Techniques, Tagungsband "Visualization '90", San Francisco, 1990
- /Will93/ A. Will: Visualisierung ornithologischer Daten, Studienarbeit, Universität Rostock, Fachbereich Informatik, 1993

Thesen

- 1) Die Visualisierung ist ein wichtiges Hilfsmittel zur Analyse und Auswertung wissenschaftlicher Daten.
- 2) Die Wahl einer geeigneten Visualisierungstechnik ist eine entscheidende Voraussetzung für den Erfolg der visuellen Analyse. Bisher vorhandene Ansätze zur automatischen Unterstützung einer solchen Wahl genügen nicht allen Anforderungen an Universalität bzw. Vollständigkeit der eingehenden Parameter.
- 3) Es ist möglich, für wissenschaftliche Daten eine einheitliche Beschreibung in Form einer Matrix einzuführen. Auf Grund dieser Datenbeschreibung, die auch für eine informationstheoretische Analyse des Datensatzes verwendbar ist, kann ein einheitlicher Auswahlalgorithmus beschrieben werden.
- 4) Es ist möglich, ein für beliebige wissenschaftliche Daten verwendbares System abstrakter Interpretationsziele zu schaffen. Dies ist ebenfalls notwendig zur Beschreibung eines einheitlichen Auswahlalgorithmus.
- 5) Die Abbildung konkreter Interpretationsziele für eine Anwendung auf das System abstrakter Interpretationsziele ist schwer automatisierbar. Eine Automatisierbarkeit ist nur möglich für bestimmte und klar abgegrenzte Anwendungsgebiete.
- 6) Ein neu eingeführter Auswahlalgorithmus führt bei bekannten Datensätzen zu ähnlichen Ergebnissen wie bisher bekannte Speziallösungen. Gleichzeitig können Ergebnisse für beliebige wissenschaftliche Datensätze geliefert werden.

Danksagung

Meine größte Dankbarkeit gilt Frau Prof. Schumann, die mir gleichzeitig eine optimale fachliche Betreuung und eine maximale wissenschaftliche Freiheit ermöglichte. Sowohl während der Erstellung dieser Arbeit als auch während meines gesamten Studiums konnte ich mich jederzeit mit meinen Fragen an sie wenden.

Diese Arbeit entstand teilweise parallel zum DFG-Projekt "Visualisierungssystem und -hilfe" (Schu 887/1-1). Für den ständigen geistigen Austausch bedanke ich mich bei den Projektmitarbeitern Susanne Arndt und Kai Lukoschek.

Rostock, 01.07.1994

Holger Theisel

Erklärung

Ich erkläre, daß ich die vorliegende Arbeit selbstständig und nur unter Zuhilfenahme der angegebenen Literatur und Hilfsmittel angefertigt habe.

Rostock, 01.07.1994

Holger Theisel