

Universität Koblenz
Institut für Wirtschaftsinformatik



Studienarbeit Studiengang Diplom-Informatik
SS 1999

Inventory - Nr. 5602

Stu.

Probabilistische Verfahren zur Diagnoseunterstützung in der Medizin

vorgelegt von:

Bodo Ludwig van Laak
Kur-Kölner-Str.28
57614 Burglahr
Matrikelnr. 9520002
E-Mail: bvl@uni-koblenz.de

Alexander Wlachos
Gartenstraße 2
56220 Kaltenengers
Matrikelnr. 17449
E-Mail: alexsis@uni-koblenz.de

Betreuer:
Dr. rer. pol. P. Rittgen

eingereicht am:
02.03.2000

Stu.

Journal of the
Royal Society of Medicine



Volume 47, Part 1
1954

Probabilistic / Diagnostic Testing in the Medicine

Edited by

Dr. W. G. I. Bell
Dr. J. H. G. Williams
Dr. R. G. L. Jones
Dr. A. G. H. Smith

Dr. J. H. G. Williams
Dr. R. G. L. Jones
Dr. A. G. H. Smith
Dr. W. G. I. Bell

London: Butterworths
1954

Price 12s. 6d.
Pp. 200

Inhaltsverzeichnis

Vorwort	xi
1 Bayes-Statistik und Probleme	1
1.1 Grundlagen der Statistik	1
1.1.1 Zufallsexperimente, Ereignisse und Wahrscheinlichkeiten	1
1.1.2 Zufallsvariable und Hypothese	2
1.1.3 Mittelwert und Standardabweichung	3
1.1.4 Bedingte Wahrscheinlichkeiten	4
1.1.5 Eine erste Anwendung und ihre Probleme	5
1.2 Vektoren und Matrizen	10
1.3 Zentrale Aussage der Bayes-Statistik	12
1.3.1 Unabhängige Ereignisse	12
1.3.2 Satz von Bayes	13
1.3.3 Verwendung des Satzes von Bayes in der Diagnose	13
1.4 Fortschreibung von Wahrscheinlichkeiten	16
1.4.1 Chancen	16
1.4.2 Fortschreibung von Chancen	19
2 Wahrscheinlichkeitsrelevanz	21
2.1 Einführung der Wahrscheinlichkeitsrelevanz	21
2.2 Wahrscheinlichkeitsrelevanzfaktor	23
2.3 Einsatzmöglichkeiten	26
2.4 Probleme	31

3	Bayessche Beliefnetze	33
3.1	Aufbau von Beliefnetzen	34
3.1.1	Qualitativer Aufbau	34
3.1.2	Quantitativer Aufbau	37
3.2	Initialisierung von Beliefnetzen	42
3.2.1	Initialisierung der Wurzelknoten	46
3.3	Propagierung	50
3.3.1	Evidenzen.	50
3.3.2	Evidenz an Blattknoten.	52
3.3.3	λ -Nachrichten	53
3.3.4	Verarbeitung einer λ -Nachricht	55
3.3.5	π -Nachrichten	57
3.3.6	Verarbeitung der π -Nachricht	57
3.4	Probleme und Lösungsvorschläge	59
3.4.1	a-priori-Wahrscheinlichkeiten	59
3.4.2	Knoten für allgemeine Kontexte	61
3.4.3	<i>link</i> -Matrizen	62
3.4.4	Aufbauformen	63
4	Faktorenanalyse	65
4.1	Ausgangsdatenmatrix	66
4.2	Korrelationsrechnung	66
4.3	Fundamentaltheorem	71
4.4	Hauptkomponentenanalyse	74
4.5	Bestimmung der Anzahl der Faktoren	77
4.5.1	Eigenwert-Kriterium	77
4.5.2	Scree-Test	78
4.5.3	kumulierte erklärte Varianz	78
4.5.4	Gewünschte Datenreduktion	79
4.6	reduzierte Faktorladungsmatrix	80
4.7	Faktorinterpretation	82
5	Clusteranalyse	85

INHALTSVERZEICHNIS

v

5.1 Ausgangsdaten	85
5.2 Proximitätsmaße	87
5.3 Arten von Gruppierungsalgorithmen	92
5.4 Agglomerative Verfahren	95
5.5 Kritik	99
6 Synthese	103
6.1 FA zur Modellierung von BBN	103
6.2 CA zur Modellierung von BBN	106
7 Zusammenfassung und Ausblick	109
Literaturverzeichnis	111

1. Einleitung 1

2. Die Bedeutung der ... 2

3. Die ... 3

4. Die ... 4

5. Die ... 5

6. Die ... 6

7. Die ... 7

8. Die ... 8

9. Die ... 9

10. Die ... 10

11. Die ... 11

12. Die ... 12

13. Die ... 13

14. Die ... 14

15. Die ... 15

16. Die ... 16

17. Die ... 17

18. Die ... 18

19. Die ... 19

20. Die ... 20

21. Die ... 21

22. Die ... 22

23. Die ... 23

24. Die ... 24

25. Die ... 25

26. Die ... 26

27. Die ... 27

28. Die ... 28

29. Die ... 29

30. Die ... 30

31. Die ... 31

32. Die ... 32

33. Die ... 33

34. Die ... 34

35. Die ... 35

36. Die ... 36

37. Die ... 37

38. Die ... 38

39. Die ... 39

40. Die ... 40

41. Die ... 41

42. Die ... 42

43. Die ... 43

44. Die ... 44

45. Die ... 45

46. Die ... 46

47. Die ... 47

48. Die ... 48

49. Die ... 49

50. Die ... 50

51. Die ... 51

52. Die ... 52

53. Die ... 53

54. Die ... 54

55. Die ... 55

56. Die ... 56

57. Die ... 57

58. Die ... 58

59. Die ... 59

60. Die ... 60

61. Die ... 61

62. Die ... 62

63. Die ... 63

64. Die ... 64

65. Die ... 65

66. Die ... 66

67. Die ... 67

68. Die ... 68

69. Die ... 69

70. Die ... 70

71. Die ... 71

72. Die ... 72

73. Die ... 73

74. Die ... 74

75. Die ... 75

76. Die ... 76

77. Die ... 77

78. Die ... 78

79. Die ... 79

80. Die ... 80

81. Die ... 81

82. Die ... 82

83. Die ... 83

84. Die ... 84

85. Die ... 85

86. Die ... 86

87. Die ... 87

88. Die ... 88

89. Die ... 89

90. Die ... 90

91. Die ... 91

92. Die ... 92

93. Die ... 93

94. Die ... 94

95. Die ... 95

96. Die ... 96

97. Die ... 97

98. Die ... 98

99. Die ... 99

100. Die ... 100

Abbildungsverzeichnis

1.1	Ein einfacher Ansatz	6
2.1	mögliche Ausgabe des Algorithmus 4	27
2.2	Diagnosewahrscheinlichkeiten nach erster Untersuchung . . .	28
2.3	Diagnosewahrscheinlichkeiten nach zweiter Untersuchung . .	29
2.4	Diagnoserelevanzfaktoren	30
2.5	Diagnosewahrscheinlichkeiten nach dritter Untersuchung . . .	31
3.1	Beispielnetz (qualitativ) für dieses Kapitel	35
3.2	Quantifiziertes Beliefnetz	43
3.3	Äquivalenz der Gleichungen 3.26 und 3.24	46
3.4	Initialisiertes Beliefnetz	51
3.5	regionale Schwankungen der Tuberkulosesterblichkeit zwischen Osnabrück und Oppeln in Oberschlesien nach [BDH ⁺ 54a] (leicht veränderte Darstellung)	60
3.6	regionale Schwankungen der Tuberkulosesterblichkeit in fin- nischen Städten und Landgemeinden nach [BDH ⁺ 54a] (leicht veränderte Darstellung)	60
3.7	mögliche Verfeinerung durch Hinzufügung von Knoten (Zu- fallsvariablen)	61
3.8	Beliefnetz zur Diagnose von Brustkrebs aus [KRW ⁺ 99]	64
4.1	Hinführung zum Fundamentaltheorem	73
4.2	vorläufige Faktorladungsmatrix	77
4.3	Kaiserkriterium	78
4.4	Scree test	79
4.5	kumulierte Varianz	80

4.6	reduzierte Faktorladungsmatrix mit 3 Faktoren	82
5.1	Prinzip der Clusteranalyse	86
5.2	Struktur der Proximitätsmaße nach [BEPW90]	90
5.3	Übersicht verschiedener Clusterverfahren nach [BEPW90]	93
5.4	Baumartige Struktur bei der Clusterung	95
5.5	Ablauf des hierarchisch agglomerativen Verfahrens nach [BEPW90]	100
5.6	Linkage-Strategien	101
5.7	Dendrogramm zum Sportlerbeispiel	102
6.1	Topologietypen	103

Tabellenverzeichnis

1.1	Tabelle mit nur 3 Symptomen	7
1.2	Tabelle mit 5 unabhängigen Symptomen	15
1.3	Anzahl der verschiedenen Diagnosen bisher	15
3.1	Wahrscheinlichkeiten für Zustände von Morbus Addison	40
3.2	Wahrscheinlichkeiten für Knoten B	41
3.3	Wahrscheinlichkeiten für Zustände von L	41
3.4	Wahrscheinlichkeiten für die Zerstörung der NNR	42
4.1	Blutwerte nach $\frac{3}{4}$ Stunden Gymnastik, 13 Studentinnen, Auszug aus [vL35]	67
4.2	Ausgangsdatenmatrix	68
4.3	standardisierte Ausgangsdatenmatrix	69
4.4	Korrelationsmatrix	70
5.1	Blutwerte: 10 Studenten, 800m-Lauf, Auszug aus [vL35]	87
5.2	Blutwerteänderungen: 10 Studenten, 800m-Lauf	88
5.3	standardisierte Blutwerteänderungen	89
5.4	Distanzen zwischen Personen im Beispiel	91
5.5	Parameter für Linkage-Strategien nach [BEPW90]	97
5.6	Distanzen nach erster Fusion	98

Vorwort

Der Wunsch des Menschen ist seit jeher, sich Werkzeuge auszudenken, die dann zum Einsatz kommen, wenn sein Tätigkeitsbereich dies erforderlich macht in der Hinsicht, ihm bei der Ausführung seiner Tätigkeiten zu helfen. Die ständige Entwicklung neuer Werkzeuge ist dabei kein Phänomen moderner Zeiten: Werkzeuge kennt der Mensch schon seit Anbeginn der Steinzeit.

Die Zeiten haben sich geändert. Werkzeuge sind heute nicht mehr aus Stein und Bronze. Vielmehr existieren auch geistige Werkzeuge, wie z.B. die Mathematik, Medizin und Wirtschaftswissenschaften oder allgemein Naturwissenschaften. Diese machen sich Menschen zu eigen, um heute z.B. Software für Computer in allen Bereichen menschlichen Schaffens zu konstruieren.

Die IT-Welt bringt ständig neue Konzepte für die Unterstützung des Anwenders und seiner Aufgaben hervor. Das Konzept der Expertensysteme ist dabei im Rahmen der Forschung nach „künstlicher Intelligenz“ von entscheidender Bedeutung und prägt bis heute die Konstruktion von vielen Softwareprodukten. Immer öfter werden solche Systeme unter Verwendung von Statistik und Wahrscheinlichkeitsrechnung konzipiert.

Expertensysteme, oder besser gesagt wissensbasierte Systeme, beschränken sich nicht auf die Berechnung von quantitativen Ergebnissen, sondern sind stattdessen in der Lage, aus bekanntem Wissen von Fachleuten, dargestellt durch die sogenannte Wissensbasis, und Metawissen (Regeln, manchmal auch vages Wissen oder Heuristiken) neues Fachwissen zu schlußfolgern (Inferenz) und somit Lösungswege für verschiedene Probleme aufzuzeigen[Wla98].

Die klassische Anwendung für Expertensysteme ist die Diagnoseunterstützung. In der Medizin werden schon lange statistische Methoden verwendet, um Wissen zu erlangen. Zwei dieser Verwendung findenden Methoden sind die Faktorenanalyse und die Clusteranalyse. Mittels dieser Methoden wird versucht, Zusammenhänge der Medizin sowohl zu finden als auch zu bestätigen.

Ein Diagnoseunterstützungssystem soll das erlangte Wissen dazu benutzen, bei der Diagnose von Krankheiten hilfreich zur Seite zu stehen, sei es, indem es Diagnose- und Untersuchungsvorschläge gibt, oder indem es prüft, ob vermutete Diagnosen korrekt sein können. Es wird versucht eine Lösung zu

finden, die durch eine Auswahl aus einer Menge vorgegebener Alternativen entsteht [Wla98]. Diese Auswahl ist in der Medizin eben eine Auswahl von Diagnosemöglichkeiten, aus denen die korrekte zu wählen ist.

Der Bau eines Diagnoseunterstützungssystems erfordert eine für den Rechner verständliche Darstellung (sprich Zahlendarstellung) von Wissen. Wir stellen in dieser Arbeit eine dieser Darstellungsmöglichkeiten vor, die sowohl für den Rechner gut umsetzbar als auch für den Modellierer des Wissens durch seine graphische Darstellung leicht verständlich zu sein scheint: die Bayesschen Beliefnetze.

Ziel der Arbeit ist es, verschiedene statistische Instrumente zur Erlangung und Repräsentation von Wissen für diagnostische Zwecke vorzustellen und eventuelle Probleme und Zusammenhänge zu kommentieren.

In Kapitel 1 dieser Arbeit geben wir einen Überblick über Grundlagen der Statistik und zeigen eine Anwendungsmöglichkeit naiver statistischer Methoden für ein Diagnoseunterstützungssystem. Kapitel 2 führt Überlegungen zur Selektion von durch einen Mediziner zu untersuchenden Informationen (z.B. in der Anamnese, Symptome, Indikationen) vor, die die Diagnosesicherheiten stark beeinflussen können. In Kapitel 3 stellen wir eine Möglichkeit zur Repräsentation vagen und unsicheren Wissens sowie zur Inferenz über selbigem dar. Diese Möglichkeit sind Bayessche Beliefnetze. Kapitel 4 und 5 befassen sich mit der Erlangung von Wissen und der Reduktion von Datenmengen mittels der multivariaten statistischen Analysemethoden Faktorenanalyse und Clusteranalyse. In Kapitel 6 folgen Gedanken zu gemeinsamen Einsatzmöglichkeiten und Zusammenhängen zwischen Bayesschen Beliefnetzen, der Faktorenanalyse und der Clusteranalyse. Kapitel 7 stellt das Schlußwort mit einem Ausblick dar.

Kapitel 1

Bayes-Statistik und Probleme

Im ersten Kapitel der Arbeit möchten wir einige benötigte Grundkenntnisse über die Stochastik vermitteln. Dabei werden wir auch erste Ansätze zur Unterstützung der Diagnostik durch Methoden der *Bayes-Statistik* erläutern und kritisch bewerten bzw. Probleme erörtern. Dieser Abschnitt enthält dabei auch einen großen Anteil der Vorkenntnisse, die in den Abschnitten *Bayessche Belief Netze* und *Multivariate Verfahren* benötigt werden.

1.1 Grundlagen der Statistik

1.1.1 Zufallsexperimente, Ereignisse und Wahrscheinlichkeiten

In vielen Versuchen ist der Versuchsausgang, den wir als *Ergebnis* bezeichnen, unabhängig von äußeren Faktoren. Standardbeispiel für einen solchen Versuch ist das Werfen eines Würfels. Das Ergebnis (hier die Augenzahl) ist bei einem idealen Würfel unabhängig von der Wurfhöhe, der Drehgeschwindigkeit, der Temperatur oder ähnlichen Einflüssen. Man kann bei einem solchen Würfelexperiment zunächst nicht vorhersagen, welche Augenzahl fällt. Festgelegt ist jedoch die Menge \mathbb{E} der möglichen Versuchsausgänge (hier ist $\mathbb{E} = \{1, 2, 3, 4, 5, 6\}$). Die Menge \mathbb{E} bezeichnen wir auch als *Ergebnisraum*. Jedes Element des Ergebnisraums heißt Ergebnis (1, 2, 3, 4, 5, 6 sind jeweils Ergebnis des Zufallsexperiments „einmaliges Werfen eines Würfels“). Das Auftreten eines Ergebnisses heißt *Ereignis* (z.B. das Ereignis „bei einmaligem Werfen eines Würfels wurden 4 Augen gewürfelt“). Wir bezeichnen daher die Menge \mathbb{E} im Kontext eines gegebenen Zufallsexperiments auch als *Ereignisraum* (mit den Ereignissen den Elementen dieser Menge).

Wird ein Zufallsexperiment mehrmals durchgeführt, so kann man abzählen, wie oft ein bestimmtes Ereignis e eingetreten ist. Die so gewonnene Zahl heißt *absolute Häufigkeit* $h(e)$ des Eintretens des Ereignisses. Unter der *relativen Häufigkeit* eines Ereignisses e bei n -maliger Durchführung des Zufallsexperiments versteht man den Quotienten

$$r(e) = \frac{h(e)}{n}. \quad (1.1)$$

Die relative Häufigkeit ist immer eine reelle Zahl zwischen 0 und 1. Die relative Häufigkeit eines Ereignisses nähert sich mit steigendem n einer bestimmten Zahl, die *Wahrscheinlichkeit* des Ereignisses $p(e)$ genannt wird¹. p ist eine Funktion mit Parameter e .

Für Wahrscheinlichkeiten gelten die drei *Axiome von Kolmogorov*²:

1. Durch eine Wahrscheinlichkeitsfunktion $p : \mathbb{E} \rightarrow [0, 1]$, $p(e) = x$ wird jedem Ereignis e eine reelle Zahl x zwischen 0 und 1 zugeordnet:

$$0 \leq p(e) \leq 1. \quad (1.2)$$

2. Das Ereignis, daß mit absoluter Sicherheit eintritt (*sicheres Ereignis* Ω) hat die Wahrscheinlichkeit 1:

$$p(\Omega) = 1. \quad (1.3)$$

3. Das Ereignis, daß mit absoluter Sicherheit nicht eintritt (*unmögliches Ereignis*) hat die Wahrscheinlichkeit 0.

4. Wenn sich zwei Ereignisse e_A und e_B gegenseitig ausschließen (man spricht von *disjunkten Ereignissen*), dann gilt

$$p(e_A \vee e_B) = p(e_A) + p(e_B). \quad (1.4)$$

1.1.2 Zufallsvariable und Hypothese

Mehrere Ereignisse können innerhalb einer Variable zusammengefaßt werden. So können die Ereignisse, daß beim Werfen eines Würfels 1, 2, 3, 4, 5,

¹Man beachte, daß $\lim_{n \rightarrow \infty} r(e) = p(e)$ nicht zwingend gelten muß. Die Erfahrung ([LR88], S. 76) zeigt aber, daß die Abweichungen zwischen $r(e)$ und $p(e)$ mit steigendem n abnimmt.

²Kolmogorov, Alexander N., russischer Mathematiker, * Tambow 25.4.1903, Prof. in Moskau, förderte die intuitionistische Logik, begründete die moderne mathematische Wahrscheinlichkeitstheorie. K. erhielt 1963 den Balzan-Preis für Mathematik. (aus [Lin73])

oder 6 Augen erscheinen, zu einer Zufallsvariable „Augenzahl“ zusammengefaßt werden. Die Zufallsvariable kann dann die Belegungen 1 bis 6 haben und stellt damit die Ereignisse

$$\text{Augenzahl} = i \quad (1.5)$$

dar. Die Ereignisse „Augenzahl = 1“, „Augenzahl = 2“ etc. werden dann auch als Hypothesen bezeichnet.

Statt einem Ereignis e eine Wahrscheinlichkeit $p(e)$ zuzuordnen, wird der Zufallsvariable eine Wahrscheinlichkeitsverteilung zugeordnet. Eine Wahrscheinlichkeitsverteilung für eine Zufallsvariable X mit der Wertemenge

$$\mathbb{W}(X) = \{x_1, x_2, \dots, x_m\} \quad (1.6)$$

ist eine Funktion p , die jeder Zahl $x_i \in \mathbb{W}(X)$ die Wahrscheinlichkeit

$$p(x_i) = p(X = x_i) \quad (1.7)$$

zuordnet. Diese Funktion wird z.B. in [LR88] wird auch als Wahrscheinlichkeitsfunktion bezeichnet.

Die Wahrscheinlichkeitsverteilung für das Würfelbeispiel ist

$$p(x_i) = \frac{1}{6} \quad \forall i \in \mathbb{N}, 1 \leq i \leq 6 \quad (1.8)$$

Eine Wahrscheinlichkeitsverteilung, die allen x_i die gleiche Wahrscheinlichkeiten zuordnet, heißt Gleichverteilung.

Wahrscheinlichkeitsverteilungen können diskret oder stetig sein. Wir gehen zunächst davon aus, daß die Zufallsvariablen diskret sind, d.h. endlich viele oder überabzählbar unendlich viele Hypothesen enthalten. Die Zuordnung von Wahrscheinlichkeiten zu den Hypothesen einer Zufallsvariable wird Verteilung oder Wahrscheinlichkeitsverteilung genannt. Bei diskreten Zufallsvariablen sprechen wir von diskreten Verteilungen. So ist die „Augenzahl“ eine diskrete Zufallsvariable, und Gleichung 1.8 stellt eine diskrete Wahrscheinlichkeitsverteilung dar.

Wir benötigen Wahrscheinlichkeitsverteilungen u.a. in den Kapiteln 3, 4 und 5.

1.1.3 Mittelwert und Stanardabweichung

Zwei wichtige Konzepte der Stochastik, die wir in Kapitel 4 und 5 benötigen, sind Mittelwerte und Stanardabweichungen. Sei der Wert einer Zufallsvariable X n -mal gemessen worden, d.h. es sei ein Meßreihe der Länge n mit Untersuchungsmerkmal X durchgeführt worden (z.B. 10-maliges Würfeln

mit 2 Würfeln mit der Zufallsvariable „Augenzahl“ als Untersuchungsmerkmal.

Dann heißt die Zahl

$$\mu(X) = \frac{\sum_{i=1}^n x_i}{n} \quad (1.9)$$

mit x_i dem Wert der Zufallsvariable bei der i -ten Versuchsdurchführung Mittelwert der Versuchsreihe.

Die Zahl

$$\sigma(X) = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu(X))^2}{n}} \quad (1.10)$$

heißt Standardabweichung der Variable X . Die Standardabweichung gibt an, wie stark die Werte einer Versuchsdurchführung im Mittel vom Mittelwert abweichen.

1.1.4 Bedingte Wahrscheinlichkeiten

Dabei sind die Wahrscheinlichkeiten in einem Kontext k zu sehen, wobei k eine Menge allgemeiner Annahmen über den Zustand der Welt während der Versuchsdurchführung darstellt. Daher müßte man statt $p(e)$ eigentlich $P(e | k)$ schreiben (dies ist die Schreibweise für „Wahrscheinlichkeit für das Eintreten des Ereignisses e unter der Voraussetzung, daß k gilt“). Wir schreiben dennoch nur $p(e)$, wenn angenommen werden kann, daß der allgemeine Kontext über die Versuchsdurchführung hinweg konstant bleibt ([Pea88], S. 29 f.). $p(e_A|e_B)$ wird *bedingte Wahrscheinlichkeit* für e_A unter Bedingung e_B genannt. Für bedingte Wahrscheinlichkeiten gilt nach [Vic97], S. 6 u.a.:

- $0 \leq p(e_A|e_B) \leq 1$.
- Unter der Annahme, daß sich die Ereignisse e_A und e_B gegenseitig ausschließen (disjunkte Ereignisse), gilt:

$$p(e_A \vee e_B|e_C) = p(e_A|e_C) + p(e_B|e_C). \quad (1.11)$$

- Falls das Eintreten von e_B bereits bekannt ist und damit der Informationsstand von e_C auf $e_C \wedge e_B$ steigt, gilt ohne weitere Einschränkungen³:

$$p(e_A \wedge e_B|e_C) = p(e_A|e_B \wedge e_C) \cdot P(e_B|e_C). \quad (1.12)$$

³und damit auch ohne Annahmen über die Abhängigkeit oder Unabhängigkeit der Ereignisse e_A , e_B und e_C

Bedingte Wahrscheinlichkeiten sind der Ausgangspunkt für den Bayesschen Formalismus ([Pea88], S. 30). Für die bedingte Wahrscheinlichkeit $p(e_A|e_B)$ gilt nach [KKKK89], S.69, [Bos95], S. 30 und anderen folgende wichtige Gleichung der Bayesschen Statistik:

$$p(e_A|e_B) = \frac{p(e_A \wedge e_B)}{p(e_B)}. \quad (1.13)$$

Die Umformung der Gleichung zu

$$p(e_A \wedge e_B) = p(e_A|e_B)p(e_B) \quad (1.14)$$

ist auch unter dem Namen *Multiplikationssatz* bekannt.

1.1.5 Eine erste Anwendung und ihre Probleme

Wir können bedingten Wahrscheinlichkeiten bereits für kleine diagnoseunterstützende Berechnungen einsetzen (s. Abbildung 1.1). Dazu benötigt man eine Menge \mathcal{S} von Symptomen und eine Menge \mathcal{D} von Diagnosen. Aus der Menge von Symptomen wählt der Anwender eine Teilmenge \mathcal{S} von Symptomen aus, die im vorliegenden Krankheitsfall als bekannt vorhanden (S_i) oder als bekannt nicht vorhanden ($\neg S_i$) ermittelt werden konnten. Das Ergebnis der Symptomermittlung ist das Ereignis $e_{\mathcal{S}}$, das eine Kombination vorliegender bzw. nicht vorliegender Symptome darstellt. Auf diese Eingabe wird der Algorithmus 1 (siehe S. 5) angewendet.

Algorithmus 1 Diagnosesicherheit mit bedingten Wahrscheinlichkeiten

```

erzeuge leere Liste  $L$ ;
 $e_{\mathcal{S}} := \bigwedge_{S_j \in \mathcal{S}} S_j \wedge \bigwedge_{\neg S_k \in \mathcal{S}} \neg S_k$ ;
for all Diagnosen  $D_i \in \mathcal{D}$  do
  berechne  $p(e_{D_i} | e_{\mathcal{S}})$  nach Gleichung 1.13;
  füge das Tupel  $(D_i, p(e_{D_i} | e_{\mathcal{S}}))$  zu  $L$  hinzu;
end for
sortiere  $L$  nach Wahrscheinlichkeitswerten;
zeige  $L$  an.

```

Die Werte $p(e_{D_i} | e_{\mathcal{S}})$ bezeichnen die Wahrscheinlichkeit für die Richtigkeit einer Diagnose D_i . Wir bezeichnen diese daher mit dem Begriff *Diagnosesicherheit* für D_i .

Um die Diagnosesicherheit nach Gleichung 1.13 berechnen zu können, benötigt man die Wahrscheinlichkeiten $p(e_{D_i} \wedge e_{\mathcal{S}})$ und die Wahrscheinlichkeiten $p(e_{\mathcal{S}})$. Um diese zu erhalten braucht man eine Tabelle (s. Tabelle 1.1 als Beispiel), in der in den Zeilen die Diagnosen D_i , und in der in den Spalten *Symptomkombinationen* notiert sind. Die Einträge in der Tabelle entsprechen den

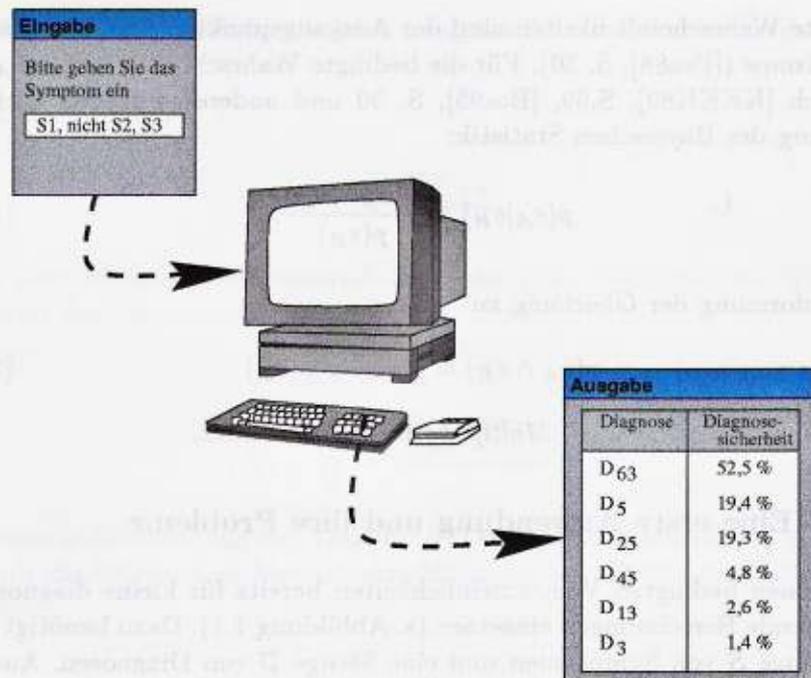


Abbildung 1.1: Ein einfacher Ansatz

Anzahlen der Krankheitsfälle, in denen die Symptomkombination der Spalte aufgetaucht ist und gleichzeitig die Diagnose D_i korrekt gestellt wurde.

Beispiel 1 Der Anwender hat die Symptome S_1 und S_3 (s. Tabelle 1.1) bei einem Krankheitsfall festgestellt. Das Symptom S_2 wurde nicht festgestellt. Er möchte nun wissen, welche der Diagnosen D_1, D_2, D_3, D_4 die wahrscheinlichste ist.

Aus der Tabelle erhält man mit n der Gesamtzahl der in der Tabelle eingetragenen Krankheitsfälle (d.h. Summe aller Einträge in der Tabelle):

$$\begin{aligned}
 p(e_{D_1} | e_{S_1 \wedge \neg S_2 \wedge S_3}) &= \frac{p(e_{D_1} \wedge e_{S_1 \wedge \neg S_2 \wedge S_3})}{p(e_{S_1 \wedge \neg S_2 \wedge S_3})} & (1.15) \\
 &= \frac{h(e_{D_1} \wedge e_{S_1 \wedge \neg S_2 \wedge S_3})}{n} \\
 &= \frac{h(e_{S_1 \wedge S_2 \wedge S_3})}{n} \\
 &= \frac{6}{13713} \\
 &= \frac{6+572+0+355}{13713} \\
 &= \frac{6}{933}
 \end{aligned}$$

Diagnose	Symptomkombinationen			
	$\neg S_1 \wedge \neg S_2 \wedge \neg S_3$	$S_1 \wedge \neg S_2 \wedge \neg S_3$	$\neg S_1 \wedge S_2 \wedge \neg S_3$	$\neg S_1 \wedge \neg S_2 \wedge S_3$
D_1	0	1234	821	338
D_2	0	0	23	823
D_3	0	12	1	36
D_4	0	482	251	29

Diagnose	Symptomkombinationen			
	$S_1 \wedge S_2 \wedge \neg S_3$	$S_1 \wedge \neg S_2 \wedge S_3$	$\neg S_1 \wedge S_2 \wedge S_3$	$S_1 \wedge S_2 \wedge S_3$
D_1	2333	6	3754	112
D_2	433	572	333	311
D_3	3	0	63	3
D_4	985	355	311	89

Tabelle 1.1: Tabelle mit nur 3 Symptomen

$$\approx 0.64\%$$

Analog wird berechnet:

$$\begin{aligned}
 p(e_{D_2} | e_{S_1 \wedge \neg S_2 \wedge S_3}) &= \frac{\frac{572}{13713}}{\frac{933}{13713}} & (1.16) \\
 &= \frac{572}{933} \\
 &\approx 61.31\%
 \end{aligned}$$

$$p(e_{D_3} | e_{S_1 \wedge \neg S_2 \wedge S_3}) = \frac{0}{933} = 0\% \quad (1.17)$$

$$p(e_{D_4} | e_{S_1 \wedge \neg S_2 \wedge S_3}) = \frac{355}{933} \approx 38.05\% \quad (1.18)$$

Damit ist D_2 die wahrscheinlichste Diagnose. Danach folgen D_4 und mit geringer Diagnosesicherheit D_1 sowie D_3 .

Zur Beurteilung eines Patienten werden jedoch nicht nur die akuten Symptome herangezogen, sondern die gesamte Krankengeschichte. Die Krankengeschichte setzt sich nach [MS62], S. 1 ff. aus vier Teilen zusammen:

- **Anamnese** (Vorgeschichte) bestehend aus teilweise unsicheren Angaben:
 - Beruf des Patienten,
 - Krankheiten in der Familie (Familienanamnese),

- besondere Erkrankungen (Tuberkulose, Bluthochdruck etc.),
 - Kinderkrankheiten,
 - Schutzimpfungen,
 - Rauch- und Trinkgewohnheiten,
 - Lebensgewohnheiten (Sport, Berufserkrankungen, Gefangenschaft etc.),
 - Vorboten der Krankheit (z.B. Müdigkeit, Appetitlosigkeit, Gewichtsverlust),
 - gegenwärtige Klagen, wie Schmerzen, Erbrechen, Husten, Schwindel.
- Status praesens (*Befund*) mit teilweise unsicheren oder unbekanntem Informationen. Der Befund wird mittels der körperlichen Untersuchung ermittelt. In einen Befund gehören unter anderem
 - Körpertemperatur,
 - Augenstellung und Augenreflexe,
 - Ohrenspiegelbefund, Hörtest,
 - Lippen- (Blässe, Feuchtigkeit), Gebiß- (Kronen etc.), Mundschleimhaut- und Zungenbeschaffenheit sowie
 - Beschaffenheit von Hals, Wirbelsäule, Brust, Lunge, Herz, Bauch, Gliedmaßen und Nervensystem.
 - Wenn nötig, weitere Untersuchungen (Nachträge), z.B.
 - Harn, Kot,
 - Darmsaft, Mageninhalt,
 - Pulsfrequenz, elektrokardiographische Untersuchungen, Atemfrequenz,
 - Röntgenuntersuchungen, Fieberkurve, und vieles mehr.
 - *Epikrise* (Schlußbetrachtung): Sie enthält die Zusammenfassung von Anamnese, Befund und Nachträgen. Hinzu kommt die endgültige Diagnose, differentialdiagnostische Überlegungen und Angaben über Morbidität (Verhältnis der Zahl der Erkrankten zur Zahl der Gesamtbevölkerung), Mortalität (Verhältnis zwischen Anzahl der Gestorbenen zur Zahl der Gesamtbevölkerung), und Letalität (Verhältnis der Zahl der Gestorbenen zur Zahl der Erkrankten).

Unsichere oder gar nicht bekannte Angaben können mit der im obigen Beispiel dargestellten Methode nicht erfaßt werden. Solche Angaben sind jedoch häufig zu finden. Beispielsweise ist es in der Notfallmedizin oft nicht möglich

herauszufinden, ob der Patient Schmerzen an bestimmten Körperteilen oder Organen verspürt, da er sich im Zustand der Bewußtlosigkeit befindet. Die Anamnese ist dann ebenso nicht oder zumindest nicht sicher festzustellen (findet man am bewußtlosen Patienten Tabak so weiß man nicht sicher, ob der Patient Raucher ist oder nicht, man kann es nur mit einer gewissen Wahrscheinlichkeit vermuten).

Beschränkt man sich bei den Symptomfeststellungen auf die beiden Ausprägungen „Symptom vorhanden“ und „Symptom nicht vorhanden“ verliert man wichtige Informationen, die sich am leichtesten am Beispiel von Schmerzen darstellen lassen: Schmerzen können vielfältige Formen haben, beginnend bei leichtem Unwohlsein über stechenden Schmerzen, chronische und akute Schmerzen, starke und schwache Schmerzen differenziert nach schmerzempfindlichen und schmerzunempfindlichen Patienten etc. Ein weiteres Beispiel für die Unzulänglichkeiten einer binären Ausprägung eines Symptoms ist das Fieber. Nicht nur, daß Fieber für verschiedene Personen bei verschiedenen Temperaturen beginnt, es muß auch unterschieden werden zwischen verschieden starkem Fieber. Zwar könnte man unterschiedliche Ausprägungen eines Symptoms wiederum als verschiedene Symptome definiert werden (logisch möglich, medizinisch jedoch nicht adäquat), dies würde aber in extremen Dimensionen der Tabelle ausufern.

Weiteres Problem ist, daß kausale Abhängigkeiten zwischen Krankheiten und Symptomen⁴ nicht ausgenutzt werden. So kann eine Krankheit völlig unabhängig davon sein, ob es einen positiven oder negativen Befund bei der röntgenologischen Untersuchung gibt. In obigem Beispiel müßte die Information über einen röntgenologischen Befund gespeichert und bei der Eingabe angegeben werden⁵, ohne daß dies die Diagnosesicherheit beeinflussen würde.

Ferner wird dadurch Speicherplatz verschwendet: Zu jeder Symptomkombination mit positivem röntgenologischen Befund wäre eine Symptomkombination mit negativem röntgenologischen Befund nötig, auch wenn diese Information nichts zur Diagnosesicherheit beiträgt.

Die im obigen Beispiel genannte Methode ist hinsichtlich der Speicherkomplexität der nötigen Tabelle untragbar:

Beispiel 2 Nehmen wir an, es seien 30 Symptome zu identifizieren und 10000 mögliche Diagnosen⁶ ständen zur Auswahl. Nehmen wir weiter an,

⁴oder zwischen Symptomen und anderen Symptomen oder zwischen Krankheiten und anderen Krankheiten

⁵...was wiederum bedeutet, daß ein Röntgenbild gemacht werden müßte, was nicht unerhebliche Kosten zur Folge hätte...

⁶In [ICD93], einem internationalen Schlüsselverzeichnis für Diagnosen, sind grob geschätzt über 40000 Diagnosemöglichkeiten aufgezählt

es würde pro Eintrag in der Tabelle ein Eintrag von einem Halbbyte⁷ Länge erforderlich. Dann bräuchte man $2^{30} * 10000 * 0.5$ Bytes⁸ Speicher, dies entspricht ca. 4,9 TByte.

1.2 Vektoren und Matrizen

Dieser Abschnitt paßt eigentlich nicht recht in dieses Kapitel. Wir wollen hierfür kein gesondertes Kapitel beginnen, da wir nur sehr kurz erklären wollen, wie mit Vektoren und Matrizen gerechnet wird. Genauere Erläuterungen finden sich in [NBES82] und vielen anderen. Wir benötigen diese Kenntnisse in den weiteren Kapiteln.

Ein Vektor \vec{x} ist ein geordnetes Tupel von Zahlen. Man kann dieses Tupel von Zahlen in Form einer Zeile

$$\vec{x} = (x_1, x_2, \dots, x_n) \quad (1.19)$$

schreiben. Dann nennen wir den Vektor Zeilenvektor. Einen in Form einer Spalte geschriebenen Vektor

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (1.20)$$

bezeichnet man als Spaltenvektor. Speziell Vektoren sind der Einsvektor $\vec{1}$, der nur Einsen enthält, und der Nullvektor $\vec{0}$, der nur Nullen enthält.

Oft ist es gleichgültig, welche Form der Vektor annimmt. Wichtig ist in diesen Fällen nur, daß der Vektor ein geordnetes Tupel mit n Werte ist (n -Tupel). n ist die Größe des Vektors.

Für beliebige Vektoren \vec{x} der Größe n und beliebige reelle Zahlen a gilt

$$a \cdot \vec{x} = \begin{pmatrix} a \cdot x_1 \\ a \cdot x_2 \\ \vdots \\ a \cdot x_n \end{pmatrix} \quad (1.21)$$

Dieses Produkt heißt skalares Produkt, die zugehörige Operation S-Multiplikation.

⁷ein Halbbyte wird auch als Nibble bezeichnet

⁸es gibt 2^{30} mögliche Symptomkombinationen bei 30 Symptomen binärer Ausprägung (vorhanden, nicht vorhanden)

Wir definieren den Operator \times (Vorsicht! Dieses Produkt kann leicht mit dem Vektorprodukt⁹ verwechselt werden!) zwischen zwei Vektoren der Größe n als elementweise Multiplikation der Werte in den Vektoren:

$$\vec{x} \times \vec{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 \cdot y_1 \\ x_2 \cdot y_2 \\ \vdots \\ x_n \cdot y_n \end{pmatrix} \quad (1.22)$$

Eine Matrix M ist eine zeilenweise Anordnung von Zeilenvektoren \vec{x}

$$M = \begin{pmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_m \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,m} \\ x_{2,1} & x_{2,2} & \dots & x_{2,m} \\ \dots & \dots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,m} \end{pmatrix} \quad (1.23)$$

oder von Spaltenvektoren. Eine spezielle Matrix ist die Einheitsmatrix E . Die Einheitsmatrix ist eine quadratische ($m \times m$) Matrix der Form

$$E = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \quad (1.24)$$

Die transponierte Matrix M^T entsteht, in dem man die Zeilen einer Matrix als Spalten und die Spalten einer Matrix als Zeilen schreibt, z.B. ist

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \quad (1.25)$$

So entsteht aus einer ($m \times n$)-Matrix (mit m Zeilen und n Spalten) eine ($n \times m$)-Matrix (mit n Zeilen und m Spalten).

Das Produkt einer ($m \times n$)-Matrix A mit den Einträgen $a_{i,j}$ und einem Vektor \vec{x} der Größe n ist definiert als

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \dots & \dots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \quad (1.26)$$

⁹auch als Kreuzprodukt bezeichnet

mit

$$y_i = \sum_{j=1}^n a_{i,j} \cdot x_j. \quad (1.27)$$

Das Produkt zwischen zwei Matrizen ist dann definiert, wenn die eine eine $(m \times n)$ - und die andere eine $(n \times m)$ -Matrix ist:

$$\begin{aligned} AB &= \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \dots & \dots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix} \begin{pmatrix} b_{1,1} & b_{1,2} & \dots & b_{1,m} \\ b_{2,1} & b_{2,2} & \dots & b_{2,m} \\ \dots & \dots & \ddots & \vdots \\ b_{n,1} & b_{n,2} & \dots & b_{n,m} \end{pmatrix} \\ &= \begin{pmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,m} \\ c_{2,1} & c_{2,2} & \dots & c_{2,m} \\ \dots & \dots & \ddots & \vdots \\ c_{m,1} & c_{m,2} & \dots & c_{m,m} \end{pmatrix} \end{aligned} \quad (1.28)$$

mit

$$c_{i,j} = \sum_{k=1}^n a_{i,k} \cdot b_{k,j} \quad (1.29)$$

Die Multiplikation von Matrizen ist nicht kommutativ. Wir kommen damit nun wieder zurück zur Bayesschen Statistik.

1.3 Zentrale Aussage der Bayes-Statistik

1.3.1 Unabhängige Ereignisse

Wir werden feststellen, daß eine Datenreduktion durch eine Annahme über die *Unabhängigkeit* von Ereignissen möglich ist. Zwei Ereignisse e_A und e_B heißen stochastisch unabhängig¹⁰ genau dann, wenn gilt (vgl. [LW92], S. 35)

$$p(e_A|e_B) = p(e_A), \quad (1.31)$$

wenn e_B also nichts zur Erklärung von e_A beiträgt.

¹⁰An anderen Stellen, wie z.B. in [FSS86], S. 63 und [Vie97], S. 15, wird die Unabhängigkeit von e_A und e_B anders definiert: e_A und e_B heißen unabhängig genau dann, wenn gilt

$$p(e_A \wedge e_B) = p(e_A) \cdot p(e_B). \quad (1.30)$$

Beide Definitionen sind jedoch gleichwertig.

1.3.2 Satz von Bayes

„Bedingte Wahrscheinlichkeiten bilden die Grundlage für statistische Diagnosehilfen in der Medizin (Anwendung des sog. Bayesschen Satzes)“ erklärt [Wer92], Kapitel 4.4. Auch für Bayessche Belief-Netze ist der Satz von Bayes von zentraler Bedeutung. Sei e_D eine Diagnose und e_S ein Symptom.

Es gilt nach Gleichung 1.13

$$p(e_D|e_S) = \frac{p(e_D \wedge e_S)}{p(e_S)} \quad (1.32)$$

und

$$p(e_S|e_D) = \frac{p(e_D \wedge e_S)}{p(e_D)}. \quad (1.33)$$

Umformung von Gleichung 1.33 ergibt

$$p(e_D \wedge e_S) = p(e_S|e_D)p(e_D). \quad (1.34)$$

Aus der Einsetzung von Gleichung 1.34 in Gleichung 1.32 folgt der Satz von Bayes:

$$p(e_D|e_S) = \frac{p(e_S|e_D)p(e_D)}{p(e_S)} \quad (1.35)$$

Gleichung 1.35 gilt für alle, sprich sowohl unabhängige als auch voneinander abhängige Ereignisse e_D und e_S (Satz von Bayes, vergleiche [Wic90], [Sie86] u.a.). Dieses Faktum ist von sehr großer Bedeutung, da das Vorliegen einer bestimmte Krankheit (Diagnose, Ereignis e_D) das Auftreten entsprechender Symptome (Ereignis e_S) bedingt, e_D und e_S also gerade nicht als unabhängig angenommen werden können.

Die Wahrscheinlichkeit $p(e_D|e_S)$ bezeichnet man als *Posteriori-Wahrscheinlichkeit*. Sie entspricht in unserem Fall der gesuchten Diagnosesicherheit¹¹. $p(e_D)$ und $p(e_S)$ heißen *Priori-Wahrscheinlichkeiten*. $p(e_S|e_D)$ wird im deutschen wie im englischen Sprachgebrauch „als *Likelihood* bezeichnet, da es im deutschen kein zweites Wort für Wahrscheinlichkeit gibt“ ([Kle80], S. 115). Auch wir wollen uns damit begnügen und den Begriff *Likelihood* verwenden.

1.3.3 Verwendung des Satzes von Bayes in der Diagnose

Der Satz von Bayes bringt uns dem Ziel der Diagnoseunterstützung ein Stück näher. Auf der linken Seite von Gleichung 1.35 steht das gesuchte Ergebnis $p(e_D|e_S)$: Das Vorliegen eines Symptoms S bei einem Patienten ist bekannt und man möchte herausfinden, wie hoch die Diagnosesicherheit

¹¹Zur Wiederholung: Es handelt sich um die Wahrscheinlichkeit der Richtigkeit einer Diagnose D , wenn man weiß, daß das Symptom S vorliegt.

für die Diagnose D ist, sprich mit welcher Wahrscheinlichkeit der Patient unter D leidet. Auf der rechten Seite von Gleichung 1.35 findet man die einfachen Wahrscheinlichkeiten $p(e_D)$ und $p(e_S)$, die recht leicht ermittelt werden können (siehe bisherige Methode). Die Likelihood $p(e_S|e_D)$ kann von einem Mediziner abgeschätzt werden, denn dies ist gerade das, was ein Mediziner bereits im Studium lernt: Welche Krankheit bringt welche Symptome (mit welcher Wahrscheinlichkeit) vor.

Den Satz von Bayes kann man (wie auch die bedingten Wahrscheinlichkeiten) erweitern, so daß auch mehrere Symptome in die Berechnung der Wahrscheinlichkeit für eine Diagnose einbezogen werden können. Wir rechnen:

$$p(e_D|e_{S_1} \wedge e_{S_2} \wedge \dots \wedge e_{S_n}) = \frac{p(e_{S_1} \wedge e_{S_2} \wedge \dots \wedge e_{S_n}|e_D)p(e_D)}{p(e_{S_1} \wedge e_{S_2} \wedge \dots \wedge e_{S_n})}. \quad (1.36)$$

Mit Gleichung 1.36 kann die Diagnosesicherheit für die Diagnose D bei Bekanntsein¹² mehrerer Symptome S_i berechnet werden. Die für die einzelnen Symptome stehenden Zufallsvariablen können mit den Werten „liegt vor“ und „liegt nicht vor“ belegt werden. Wir schreiben e_{+S_i} für ein vorliegendes Symptom S_i und e_{-S_i} für ein nicht vorliegendes Symptom S_i sowie e_S als zusammenfassung dafür, daß einer dieser Fälle vorliegt (sprich: ein Symptom ist bekannt, z.B. dadurch, daß der Diagnostiker untersucht hat, ob das Symptom vorliegt oder nicht). Wir nehmen nun an, daß die Symptome untereinander stochastisch unabhängig sind. Unter der Annahme der stochastischen Unabhängigkeit der Symptome gelten nach [Sie86], S. 29 und [Pea88], S. 37, die Gleichungen 1.37 und 1.38.

$$p\left(\bigwedge_{i=1}^n e_{S_i}\right) = \prod_{i=1}^n p(e_{S_i}) \quad (1.37)$$

$$p\left(\bigwedge_{i=1}^n e_{S_i} \middle| e_D\right) = \prod_{i=1}^n p(e_{S_i}|e_D) \quad (1.38)$$

Gleichung 1.38 entspricht dabei der Definition der stochastischen Unabhängigkeit in Bezug auf ein Ereignis e_D aus [Kle80], S. 126.

Dies bedeutet, daß wir den Satz von Bayes für n bekannte, unabhängige Symptome S_i , $1 \leq i \leq n$ schreiben können als

$$p\left(e_D \middle| \bigwedge_{i=1}^n e_{S_i}\right) = \frac{\prod_{i=1}^n p(e_{S_i}|e_D) \cdot p(e_D)}{\prod_{i=1}^n p(e_{S_i})} \quad (1.39)$$

Durch Gleichung 1.39 kann man sich nun von den notwendigen 2^n Symptomkombinationen befreien. Man benötigt nur noch die Häufigkeiten des

¹²zur Erinnerung: Unter dem Bekanntsein ist gemeint, daß man weiß, daß das Symptom vorliegt, oder daß man weiß, daß das Symptom nicht vorliegt

Auftretens der einzelnen Symptome bei den einzelnen Diagnosen. Die Anzahl der Tabellenspalten reduziert sich damit auf n . Zusätzlich benötigt man noch eine Tabelle, in der festgehalten wird, welche Diagnose wie oft gestellt wurde. Bei 10000 Diagnosen, 30 Symptomen¹³ und 1 Nibble pro Tabelleneintrag benötigt man nur noch

$$10000 \cdot 30 \cdot 0.5 \cdot 2 + 10000 \cdot 0.5 \text{Byte} = 150000 \text{Byte} \approx 302,8 \text{kByte} \quad (1.40)$$

Die Größe der Datenbank ist auf ca. $5,8 \cdot 10^{-6} \%$ (!) der ursprünglichen Größe geschrumpft.

Diagnose	unabhängige Symptome									
	e_{S_1}		e_{S_2}		e_{S_3}		e_{S_4}		e_{S_5}	
	e_{+S_1}	e_{-S_1}	e_{+S_2}	e_{-S_2}	e_{+S_3}	e_{-S_3}	e_{+S_4}	e_{-S_4}	e_{+S_5}	e_{-S_5}
D_1	102	270	472	290	24	1003	1113	20	175	175
D_2	579	0	13	110	1	520	60	50	465	117
D_3	552	17	559	560	3	310	15	50	1	120
D_4	1359	1	2530	517	251	700	29	100	905	10
$\sum e_{+/-S_i}$	2592	288	3574	1477	279	2533	1217	220	1546	422
$\sum e_{S_i}$	2880		5051		2812		1437		1968	

Tabelle 1.2: Tabelle mit 5 unabhängigen Symptomen

Diagnose	Anzahl
D_1	1413
D_2	601
D_3	1210
D_4	3300
Krankheitsfälle gesamt	6524

Tabelle 1.3: Anzahl der verschiedenen Diagnosen bisher

Beispiel 3 Wir nehmen in diesem Beispiel an, die Symptome S_1, S_2, S_3, S_4 und S_5 und damit die Ereignisse $e_{S_1}, e_{S_2}, e_{S_3}, e_{S_4}$ und e_{S_5} seien unabhängig. Nach Tabelle 1.2 und Tabelle 1.3 soll die Diagnosesicherheit für die Diagnose D_2 berechnet werden, wenn bekannt ist, daß S_1 und S_5 vorliegen und S_3 nicht vorliegt.

$$\begin{aligned}
 & p(e_{D_1} | e_{+S_1} \wedge e_{-S_2} \wedge e_{+S_3}) \\
 &= \frac{p(e_{+S_1} | e_{D_1}) \cdot p(e_{-S_2} | e_{D_1}) \cdot p(e_{+S_3} | e_{D_1}) \cdot p(e_{D_1})}{p(e_{+S_1}) \cdot p(e_{-S_2}) \cdot p(e_{+S_3})} \quad (1.41)
 \end{aligned}$$

¹³ die doppelt aufzuführen sind: einmal für „vorliegend“, einmal für „nicht vorliegend“

$$\begin{aligned}
 &= \frac{579 \cdot 290 \cdot 465 \cdot 601}{601 \cdot 601 \cdot 601 \cdot 6524} \\
 &= \frac{2592 \cdot 2533 \cdot 1546}{2880 \cdot 2812 \cdot 1968} \\
 &\approx 9,4\%
 \end{aligned}$$

Stellt der Diagnostiker fest, daß auch das Symptom S_2 vorliegt, so erhält man auf analoge Weise einen Wert von ca. 0,3%, was sich dadurch erklären läßt, daß in Tabelle 1.2 e_{+S_2} relativ zu e_{-S_2} selten vorkommt. Erwartungsgemäß erhöht sich die Diagnosesicherheit, wenn man feststellt, daß S_2 nicht vorliegt: Man erhält als Diagnosesicherheit ca. 10,1%.

Im allgemeinen kann nach dem Algorithmus 2 vorgegangen werden.

Algorithmus 2 Diagnosesicherheit mit dem Satz von Bayes

```

prüfe, ob Symptome aus  $S_1$  bis  $S_5$  bekannt sind;
Bekannt vorhandene Symptome  $S_i$  werden als  $e_{+S_i}$  behandelt;
Bekannt nicht vorhandene werden als Ereignis  $e_{-S_i}$  behandelt;
for all Diagnosen do
    Diagnosesicherheit wie in Beispiel 3 gezeigt berechnen;
end for
if Diagnose  $D_x$  erweist sich als korrekt then
    aktualisiere die Spalten für bekannte Symptome in der für die Diagnose
     $D_x$  stehenden Zeile;
end if

```

Durch die letzte *if*-Anweisung in Algorithmus 2 wird gewährleistet, daß die Datenbanktabelle immer wieder auf den neuesten Stand gebracht wird und das System aufgrund dessen mit der Zeit und mit jedem bestätigten Krankheitsfall lernt.

1.4 Fortschreibung von Wahrscheinlichkeiten

1.4.1 Chancen

Im folgenden werden wir die *Verhältnisgleichungen* nach [Pea88] vorstellen und damit einen wesentlichen Grundstein für das Verständnis Bayesscher Beliefnetze legen. Wir benötigen eine Unterscheidung, ob eine Diagnose D vorliegt oder nicht. Wir schreiben analog zur Ereignisschreibweise für Symptome e_{+D} für das Ereignis, daß D vorliegt und e_{-D} für das Ereignis, daß D nicht vorliegt.

Es ist

$$p(e_{+D} | e_S) = \frac{p(e_S | e_{+D}) \cdot p(e_{+D})}{p(e_S)} \quad (1.42)$$

und gleichsam

$$p(e_{-D} | e_S) = \frac{p(e_S | e_{-D}) \cdot p(e_{-D})}{p(e_S)}. \quad (1.43)$$

Aus diesen Gleichungen setzt sich das Verhältnis

$$\begin{aligned} O(e_{+D} | e_S) &= \frac{p(e_{+D} | e_S)}{p(e_{-D} | e_S)} & (1.44) \\ &= \frac{\frac{p(e_S | e_{+D}) \cdot p(e_{+D})}{p(e_S)}}{\frac{p(e_S | e_{-D}) \cdot p(e_{-D})}{p(e_S)}} \\ &= \frac{p(e_S | e_{+D}) \cdot p(e_{+D})}{p(e_S)} \cdot \frac{p(e_S)}{p(e_S | e_{-D}) \cdot p(e_{-D})} \\ &= \frac{p(e_S | e_{+D}) \cdot p(e_{+D})}{p(e_S | e_{-D}) \cdot p(e_{-D})} \end{aligned}$$

zusammen. Wir bezeichnen dieses Verhältnis der Posteriori-Wahrscheinlichkeiten als *Posteriorichance*. In [Pea88] wird das Verhältnis mit dem Begriff „posterior odds“ bezeichnet.

Gleichung 1.45 stellt den Zusammenhang zwischen Chance und Wahrscheinlichkeit dar.

$$p(A) = \frac{O(A)}{1 + O(A)} \quad (1.45)$$

[Pea88] bildet noch zwei weitere Verhältnisse:

- „prior odds“ (*Priorichance*):

$$O(e_{+D}) = \frac{p(e_{+D})}{p(e_{-D})} \quad (1.46)$$

- „likelihood ratio“ (*Likelihoodchance*):

$$L(e_S | e_{+D}) = \frac{p(e_S | e_{+D})}{p(e_S | e_{-D})} \quad (1.47)$$

Die Posteriorichance läßt sich schreiben als Produkt der beiden letztgenannten Chancen (Priorichance und Likelihoodchance):

$$O(e_{+D} | e_S) = O(e_{+D}) \cdot L(e_S | e_{+D}) \quad (1.48)$$

Die Diagnose (Hypothese) D wird durch die beiden Komponenten $O(e_{+D})$ und $L(e_S | e_{+D})$ erklärt. Da $O(e_{+D})$ in einem allgemeinen Kontext K (siehe hierzu Seite 4) zu sehen ist, erklärt diese Komponente die Posteriorichance mittels dieses allgemeinen Kontextes. Die Komponente $L(e_S |$

e_{+D}) enthält den durch das Ereignis e_S bewirkten Erklärungsanteil. $L(e_S | e_{+D})$ unterstützt die Posterioriechance hingegen dadurch, daß das Vorhandensein/Abhandensein eines Symptoms bekannt ist.

Das folgende Beispiel aus [Pea90] zeigt, wie nötig es sein kann, die Verhältnisgleichung zu verwenden.

Beispiel 4 *Eine Alarmanlage heult. Man möchte wissen, wie hoch die Wahrscheinlichkeit dafür ist, daß ein Einbruch stattgefunden hat. Bekannt sind folgende Daten:*

- mit 95-prozentiger Wahrscheinlichkeit wird durch einen Einbruch ein Alarm ausgelöst: $p(\text{Alarm}|\text{Einbruch}) = 0.95$.
- Es gibt eine kleine Wahrscheinlichkeit von 0.01, daß der Alarm durch ein anderes Ereignis ausgelöst wird: ($p(\text{Alarm}|\text{kein Einbruch}) = 0.01$).
- Die Chancen, daß in einem gegebenen Haus in einer beliebigen Nacht eingebrochen wird stehen laut Polizeistatistik bei 1:10000, d.h. $p(\text{Einbruch}) = 0.0001$.

Nach dem Satz von Bayes müßte man rechnen:

$$p(\text{Einbruch}|\text{Alarm}) = \frac{p(\text{Alarm}|\text{Einbruch}) \cdot p(\text{Alarm})}{p(\text{Einbruch})} \quad (1.49)$$

Leider ist jedoch $p(\text{Alarm})$ nicht bekannt. Daher kann man stattdessen die Verhältnisgleichung verwenden:

$$\begin{aligned} O(\text{Einbruch}|\text{Alarm}) &= O(\text{Einbruch}) \cdot L(\text{Alarm}|\text{Einbruch}) & (1.50) \\ &= \frac{p(\text{Einbruch})}{p(\text{kein Einbruch})} \cdot \frac{p(\text{Alarm}|\text{Einbruch})}{p(\text{Alarm}|\text{kein Einbruch})} \\ &= \frac{0.0001}{1 - 0.0001} \cdot \frac{0.95}{0.01} \\ &= 0.0095 \end{aligned}$$

Damit ist

$$p(\text{Einbruch}|\text{Alarm}) = \frac{O(\text{Einbruch}|\text{Alarm})}{1 + O(\text{Einbruch}|\text{Alarm})} = 0.00941 \quad (1.51)$$

Erstaunlich ist im oben genannten Fall die geringe Einbruchswahrscheinlichkeit bei heulendem Alarm. Dies liegt vor allem daran, daß die Wahrscheinlichkeit für einen Einbruch so gering ist. Man beachte jedoch auch das Verhältnis zwischen $p(\text{Einbruch})$ und $p(\text{Einbruch}|\text{Alarm})$ (0.0001:0.00941). Dadurch, daß man den Alarm gehört hat wird die Einbruchswahrscheinlichkeit ver-94-facht. Die Hypothese, daß ein Einbruch stattgefunden hat wird also durch den Likelihood-Faktor stark unterstützt.

1.4.2 Fortschreibung von Chancen

Wir erweitern die Berechnung von Chancen nun auf mehrere Symptome. Dazu setzen wir voraus, daß die Symptome S_i , $1 \leq i \leq n$ stochastisch unabhängig voneinander sind. Wir ersetzen das Ereignis e_S aus Gleichung 1.48 durch das zusammengesetzte Ereignis $\bigwedge_{i=1}^n e_{S_i}$.

$$\begin{aligned}
 O(e_{+D} | \bigwedge_{i=1}^n e_{S_i}) &= O(e_{+D}) \cdot L(\bigwedge_{i=1}^n e_{S_i} | e_{+D}) & (1.52) \\
 &= O(e_{+D}) \cdot \frac{p(\bigwedge_{i=1}^n e_{S_i} | e_{+D})}{p(\bigwedge_{i=1}^n e_{S_i} | e_{-D})} \\
 &= O(e_{+D}) \cdot \frac{\prod_{i=1}^n p(e_{S_i} | e_{+D})}{\prod_{i=1}^n p(e_{S_i} | e_{-D})} \\
 &= O(e_{+D}) \cdot \prod_{i=1}^n \frac{p(e_{S_i} | e_{+D})}{p(e_{S_i} | e_{-D})} \\
 &= O(e_{+D}) \cdot \prod_{i=1}^n L(e_{S_i} | e_{+D})
 \end{aligned}$$

Die Berechnung der Diagnosesicherheiten wird etwas komplizierter. Dies ist der Preis, der gezahlt werden muß, um mit wesentlich kleineren Tabellen arbeiten zu können. Man benötigt nur eine Tabelle mit den absoluten Häufigkeiten des Auftretens bestimmter Symptome und eine Tabelle mit den absoluten Häufigkeiten der Diagnosen.

Der Vorteil von Gleichung 1.52 ist, daß man mit einer Tabelle absoluter Wahrscheinlichkeiten mit N Diagnosen in N Zeilen und n Symptomen in n Spalten auskommt.

Einen interessanten und im weiteren Verlauf wichtigen Aspekt der Bayes-Statistik unter der Annahme der stochastischen Unabhängigkeit von Ereignissen ist die zeitliche Fortschreibung von Wahrscheinlichkeiten bzw. Chancen. So schreibt [Pit75], S. 142:

„The form of Bayes' theorem that applies when information is conditionally independent (...) No matter how long sequence of information, each event may be encoded as a likelihood ratio and then forgotten.“ (Zitiert nach: [Kle80], S. 127).

Nehmen wir an, wir kennen bereits die Chance $O(e_D | \bigwedge_{i=1}^n e_{S_i})$ und es kommt ein Symptom S_{n+1} hinzu. Dann ist

$$O\left(e_D \left| \bigwedge_{i=1}^{n+1} e_{S_i} \right.\right) = O\left(e_D \left| \bigwedge_{i=1}^n e_{S_i} \right.\right) \cdot L(e_{S_{n+1}} | e_D) \quad (1.53)$$

nach [Pea88], S.38.

S_{n+1} ist ein Symptom, das zu den bereits bekannten Symptomen hinzukommt. Dieses Symptom kann aus der Antwort eines Patienten auf eine Frage des Arztes folgen („Leiden Sie oft unter Lustlosigkeit?“), oder auch aus medizinischen Tests (z.B. Messung der Leukozytenzahl¹⁴) folgen.

¹⁴die Leukozytenzahl ist die Anzahl der weißen Blutzellen pro Kubikzentimeter. Sie kann z.B. bei Pneumonie, Appendizitis und anderen Entzündungen erhöht sein. In diesem Fall spricht man als Symptom von Leukozytose ([Psc64])

Kapitel 2

Wahrscheinlichkeitsrelevanz

In diesem Abschnitt stellen wir eine Möglichkeit vor, wie die Fortschreibung von Chancen eingesetzt werden kann, um die Unterstützung der Diagnosesicherheit durch medizinische Untersuchungen und Tests vorherzusagen. Dazu führen wir eine Kennzahl ein, die wir Wahrscheinlichkeitsrelevanz nennen. Uns erscheint diese Zahl interessant aus wirtschaftlichen und aus zeitlichen Gründen.

2.1 Einführung der Wahrscheinlichkeitsrelevanz

Durch Einsetzung von Gleichung 1.53 in Gleichung 1.45 erhält man

$$p \left(e_D \left| \bigwedge_{i=1}^{n+1} e_{S_i} \right. \right) = \frac{O(e_D | \bigwedge_{i=1}^n e_{S_i}) \cdot L(e_{S_{n+1}} | e_D)}{1 + (O(e_D | \bigwedge_{i=1}^n e_{S_i}) \cdot L(e_{S_{n+1}} | e_D))} \quad (2.1)$$

als eine Gleichung zur Diagnosesicherheits-Prognose. Sie kann eingesetzt werden, wenn man medizinische Tests plant.

Mit dieser Gleichung kann im Vorfeld eines medizinischen Test festgestellt werden, ob man, wenn man bei dem Test das neue Symptom S_{n+1} feststellt, mit einer Veränderung der Wahrscheinlichkeit für die Diagnose D rechnen kann oder nicht. Bei

$$p \left(e_D \left| \bigwedge_{i=1}^{n+1} e_{S_i} \right. \right) - p \left(e_D \left| \bigwedge_{i=1}^n e_{S_i} \right. \right) \approx 0 \quad (2.2)$$

verändert sich die Diagnosesicherheit für Diagnose D kaum, so daß der Test auf das Symptom S_{n+1} kaum lohnt¹.

¹Wir werden die vage Phrase „kaum lohnt“ noch weiter spezifizieren.

Definition 1 (Wahrscheinlichkeitsrelevanz) Die reelle Zahl

$$\mathcal{R}_\Delta = \left| p \left(e_D \left| \bigwedge_{i=1}^{n+1} e_{S_i} \right. \right) - p \left(e_D \left| \bigwedge_{i=1}^n e_{S_i} \right. \right) \right| \quad (2.3)$$

ist ein Maß für die Unterstützung einer Diagnose durch ein neues Faktum. Wir bezeichnen diese Zahl mit dem Begriff *Wahrscheinlichkeitsrelevanz* Wahrscheinlichkeitsrelevanz. \diamond

Die Wahrscheinlichkeitsrelevanz zeigt dem Anwender an, um welchen Wert sich die Diagnosesicherheit durch das Bekanntsein eines neuen Symptoms verändert. Symptome mit großer Wahrscheinlichkeitsrelevanz sind die wichtigsten nachzuprüfenden Symptome.

Nehmen wir nun an, das Symptom S_x könnte mehrere Ausprägungen² annehmen. Dann schreiben wir für das Ereignis, daß das Symptom S_x die Ausprägung a annimmt $e_{S_x=a}$. Damit definieren wir:

Definition 2 (Diagnoserelevanz von S_x unter Ausprägung a) Die Zahl

$$\mathcal{R}_\Delta(S_x, a) = \left| p \left(e_D \left| \bigwedge_{i=1}^n e_{S_i} \wedge e_{S_x=a} \right. \right) - p \left(e_D \left| \bigwedge_{i=1}^n e_{S_i} \right. \right) \right| \quad (2.4)$$

bezeichnen wir als *Diagnoserelevanz von S_x unter Ausprägung a* . \diamond

Die Zahl $\mathcal{R}_\Delta(S_x, a)$ ist der Wert, um den sich die Diagnosesicherheit verändert, wenn man ein Symptom S_x nachprüft und feststellt, daß die Ausprägung a vorliegt. Wir wollen nun herausfinden, welchen Einfluß das Symptom auf die Diagnosesicherheit höchstens haben kann. Dieser Einfluß ist bei der Ausprägung a von S_x erreicht, bei dem die Diagnoserelevanz von S_x unter Ausprägung a den größten Wert annimmt, so daß wir definieren:

Definition 3 (Diagnoserelevanz von S_x) Die Zahl

$$\mathcal{R}_\Delta(S_x) = \max_a \left\{ \left| p \left(e_D \left| \bigwedge_{i=1}^n e_{S_i} \wedge e_{S_x=a} \right. \right) - p \left(e_D \left| \bigwedge_{i=1}^n e_{S_i} \right. \right) \right| \right\} \quad (2.5)$$

heißt *Diagnoserelevanz von S_x* . \diamond

Das Zeichen \mathcal{R} steht für den Begriff der Relevanz. Δ soll wie in der Mathematik üblich daran erinnern, daß die Relevanz aus einer Differenz entstanden ist.

²die schließt den binären Fall ein

Mit der Definition 3 haben wir nun ein Hilfsmittel zur Verfügung gestellt, mit dem wir leicht feststellen können, welches Symptom den größten Einfluß auf die Diagnosesicherheit haben könnte und damit, welches Symptom als nächstes zu prüfen ist, um eine gute Diagnose im Sinne einer mit hoher Wahrscheinlichkeit korrekten Aussage gefunden werden soll. Wir verwenden die Definitionen 2 und 3 in Algorithmus 3 für eben diese Aufgabe.

Algorithmus 3 Aufdeckung der wichtigsten zu prüfenden Symptome

```

erzeuge eine leere Liste L;
wähle eine Menge  $\mathbb{S} \subset \mathcal{S}$ ;
 $\mathbb{S}_u = \mathcal{S} - \mathbb{S}$ ;
wähle die vermutete Diagnose  $D \in \mathcal{D}$ ;
for all  $S_x \in \mathbb{S}_u$  do
  for all Ausprägungen  $a$  von  $S_x$  do
    max. Wert = 0;
    berechne  $\mathcal{R}_\Delta(S_x, a)$ ;
    if  $\mathcal{R}_\Delta(S_x, a) >$  max. Wert then
      max. Wert =  $\mathcal{R}_\Delta(S_x, a)$ ;
    end if
  ; der max. Wert entspricht nun  $\mathcal{R}_\Delta(S_x)$ ...
  end for
  füge Tupel  $(S_x, \mathcal{R}_\Delta(S_x))$  zu L hinzu;
end for
sortiere L absteigend nach  $\mathcal{R}_\Delta(S_x)$ ;
zeige L an.

```

Nun weiß man, welches Symptom welchen maximalen Einfluß auf die Sicherheit einer bereits vermuteten Diagnose haben kann.

In welche Richtung die Diagnosesicherheit verändert wird kann festgestellt werden, in dem man die Beträge in den Berechnungsformeln für die Kennzahlen \mathcal{R}_Δ , $\mathcal{R}_\Delta(S_x)$ und $\mathcal{R}_\Delta(S_x, a)$ entfernt und so die gerichteten Relevanzen $\mathcal{R}_{\rightarrow}$, $\mathcal{R}_{\rightarrow}(S_x)$ und $\mathcal{R}_{\rightarrow}(S_x, a)$ erhält. Negative Werte gerichteter Relevanzen bedeuten dann eine Verringerung der Diagnosesicherheit, d.h. eine Verkleinerung der Wahrscheinlichkeit, daß diese Diagnose korrekt ist. Positive Werte gerichteter Relevanzen bedeuten, daß sich die Diagnosesicherheit erhöht.

2.2 Wahrscheinlichkeitsrelevanzfaktor

Ein zweiter möglicher Ansatz, die Relevanz eines Symptoms für die Diagnosesicherheit, ist die Berechnung von Relevanzfaktoren. Sie gestaltet sich prinzipiell etwas einfacher, haben aber dafür den Nachteil, daß man nicht

auf den ersten Blick sehen kann, wie stark³ sich das Bekanntsein eines Symptoms auf die Diagnosesicherheit auswirkt.

Wir bilden hier keine Differenz sondern einen Quotienten unter Verwendung des Satz von Bayes und der Annahme der stochastischen Unabhängigkeit der Symptome wie folgt:

$$\begin{aligned}
 \frac{p(e_D | \bigwedge_{i=1}^{n+1} e_{S_i})}{p(e_D | \bigwedge_{i=1}^n e_{S_i})} &= \frac{\frac{p(\bigwedge_{i=1}^{n+1} e_{S_i} | e_D) p(e_D)}{p(\bigwedge_{i=1}^{n+1} e_{S_i})}}{\frac{p(\bigwedge_{i=1}^n e_{S_i} | e_D) p(e_D)}{p(\bigwedge_{i=1}^n e_{S_i})}} & (2.6) \\
 &= \frac{p(\bigwedge_{i=1}^{n+1} e_{S_i} | e_D) p(e_D)}{p(\bigwedge_{i=1}^{n+1} e_{S_i})} \cdot \frac{p(\bigwedge_{i=1}^n e_{S_i})}{p(\bigwedge_{i=1}^n e_{S_i} | e_D) p(e_D)} \\
 &= \frac{p(\bigwedge_{i=1}^{n+1} e_{S_i} | e_D) p(e_D) p(\bigwedge_{i=1}^n e_{S_i})}{p(\bigwedge_{i=1}^{n+1} e_{S_i}) p(\bigwedge_{i=1}^n e_{S_i} | e_D) p(e_D)} \\
 &= \frac{p(\bigwedge_{i=1}^n e_{S_i} | e_D) p(e_{S_{n+1}} | e_D) p(e_D) p(\bigwedge_{i=1}^n e_{S_i})}{p(\bigwedge_{i=1}^n e_{S_i}) p(e_{S_{n+1}}) p(\bigwedge_{i=1}^n e_{S_i} | e_D) p(e_D)} \\
 &= \frac{p(e_{S_{n+1}} | e_D)}{p(e_{S_{n+1}})}
 \end{aligned}$$

Offensichtlich ist der letzte Quotient eine Zahl, mit der man $p(e_D | \bigwedge_{i=1}^n e_{S_i})$ multiplizieren kann, um als Ergebnis $p(e_D | \bigwedge_{i=1}^{n+1} e_{S_i})$ zu erhalten. Dieses Ergebnis bedeutet, daß nicht nur eine Fortschreibung von Chancen nach den Ausführungen in Abschnitt 1.4.2 möglich ist, sondern auch eine direkte Fortschreibung von Wahrscheinlichkeiten. Hat man eine Diagnosewahrscheinlichkeit berechnet und kommt nun ein neues Symptom hinzu, so wird die bisherige Diagnosesicherheit nur noch mit $\frac{p(e_{S_{n+1}} | e_D)}{p(e_{S_{n+1}})}$ multipliziert, und man erhält die neue Diagnosesicherheit. Wir wollen diesen Quotienten deshalb definieren als:

Definition 4 (Wahrscheinlichkeitsrelevanzfaktor) Die Zahl

$$\mathcal{R}_{\odot} = \frac{p(e_{S_{n+1}} | e_D)}{p(e_{S_{n+1}})} \quad (2.7)$$

bezeichnen wir als *Wahrscheinlichkeitsrelevanzfaktor*. \diamond

Das Zeichen \odot soll daran erinnern, daß wir Wahrscheinlichkeiten durch Multiplikation mit dem Wahrscheinlichkeitsrelevanzfaktor „updaten“ können.

Analog zu den Definitionen 2 und 3 definieren wir daher weiter:

³ „um wie viele Prozent“

Definition 5 (Diagnoserelevanzfaktor von S_x unter Ausprägung a)

Die Zahl

$$\mathcal{R}_{\odot}(S_x, a) = \frac{p(e_{S_{n+1}=a} | e_D)}{p(e_{S_{n+1}=a})} \quad (2.8)$$

heißt *Diagnoserelevanzfaktor von S_x unter Ausprägung a* . \diamond **Definition 6 (Diagnoserelevanzfaktor von S_x)** Die Zahl

$$\mathcal{R}_{\odot}(S_x) = \max_a \left\{ \frac{p(e_{S_{n+1}=a} | e_D)}{p(e_{S_{n+1}=a})} \right\} \quad (2.9)$$

heißt *Diagnoserelevanzfaktor von S_x* . \diamond

Das Verfahren, den Diagnoserelevanzfaktor von S_x als eine Zahl aufzufassen ist nicht unproblematisch, wie folgendes Beispiel zeigt:

Beispiel 5 Das Symptom S habe drei Ausprägungen a , b und c . Es sei

$$\mathcal{R}_{\odot}(S_x, a) = 1.1 \quad (2.10)$$

$$\mathcal{R}_{\odot}(S_x, b) = 0.8 \quad (2.11)$$

$$\mathcal{R}_{\odot}(S_x, c) = 0.99 \quad (2.12)$$

Dann kann das Wissen über S_x einen diagnosebestätigenden (bei Ausprägung a) oder einen diagnoseentkräftigenden (bei Ausprägung b und c) Effekt haben.

Dies bedeutet, daß ein Symptom die Diagnosesicherheit in zwei Richtungen beeinflussen kann. Demnach müssen wir, wenn wir den Diagnostiker bei der Erhöhung der Diagnosesicherheit oder bei der Entkräftung einer Diagnosevermutung unterstützen wollen, dieses Symptom zweimal angeben: Einmal als bestätigendes, einmal als entkräftendes Mittel.

Die beiden Werte fassen wir zum sogenannten *Diagnoserelevanztupel* zusammen:

Definition 7 (Diagnoserelevanztupel) Das Tupel

$$(\mathcal{R}_{\odot+}(S_x), \mathcal{R}_{\odot-}(S_x)) \quad (2.13)$$

mit

$$\mathcal{R}_{\odot+}(S_x) = \max_a \left\{ \frac{p(e_{S_{n+1}=a} | e_D)}{p(e_{S_{n+1}=a})} \right\} \quad (2.14)$$

und

$$\mathcal{R}_{\odot-}(S_x) = \min_a \left\{ \frac{p(e_{S_{n+1}=a} | e_D)}{p(e_{S_{n+1}=a})} \right\} \quad (2.15)$$

bezeichnen wir als *Diagnoserelevanztupel*. \diamond

Mittels Algorithmus 4 erhalten wir zwei Listen: Eine Liste mit den möglicherweise diagnosebestätigenden Symptomen sowie eine mit möglicherweise diagnosezurückweisenden Symptomen.

Algorithmus 4 Aufdeckung der wichtigsten zu prüfenden Symptome II

```

erzeuge leere  $\mathcal{R}_{\odot+}$ -Liste;
erzeuge leere  $\mathcal{R}_{\odot-}$ -Liste;
wähle eine Menge bereits geprüfter Symptome  $\mathbb{S} \subset \mathcal{S}$ ;
 $\mathbb{S}_u = \mathcal{S} - \mathbb{S}$ ;
wähle die vermutete Diagnose  $D \in \mathcal{D}$ ;
for all  $S_x \in \mathbb{S}_u$  do
  Berechne  $(\mathcal{R}_{\odot+}(S_x), \mathcal{R}_{\odot-}(S_x))$ ;
  if  $\mathcal{R}_{\odot+}(S_x) > 1$  then
    füge das Tupel  $(S_x, \mathcal{R}_{\odot+}(S_x))$  zur  $\mathcal{R}_{\odot+}$ -Liste hinzu;
  end if
  if  $\mathcal{R}_{\odot-}(S_x) < 1$  then
    füge das Tupel  $(S_x, \mathcal{R}_{\odot-}(S_x))$  zur  $\mathcal{R}_{\odot-}$ -Liste hinzu;
  end if
end for
sortiere die  $\mathcal{R}_{\odot+}$ -Liste absteigend;
sortiere die  $\mathcal{R}_{\odot-}$ -Liste aufsteigend;
zeige beide Listen an.

```

Die Ausgabe des Algorithmus 4 könnte wie in Abbildung 2.1 dargestellt aussehen.

2.3 Einsatzmöglichkeiten

Die Diagnoserelevanzen und das Diagnoserelevanztupel haben einige interessante Anwendungsmöglichkeiten. Sie können dazu eingesetzt werden, Diagnosevermutungen systematisch zu bestätigen oder ebenso systematisch zu widerlegen.

Möchte man die Diagnosevermutung bestätigen, so findet man in der $\mathcal{R}_{\odot+}$ -Liste eine Auswahl von Symptomen, mit denen die Diagnosesicherheit maximiert werden könnten. Analog findet man in der $\mathcal{R}_{\odot-}$ -Liste Symptome, mit denen die Diagnosesicherheit minimiert und damit ausgeschlossen werden könnte.

Die verschiedenen Einsatzmöglichkeiten und Gründe für den Einsatz der Diagnoserelevanzen, Diagnoserelevanzfaktoren und vor allem der Diagnoserelevanztupel sollen nun in einem hypothetischen Beispiel erläutert werden.

Beispiel 6 Der Arzt A verfügt in seiner Praxis über ein Diagnoseunterstützungs-

Diagnose: XYZ-Syndrom	
<u>Bestätigende Symptome</u>	<u>Wiederlegende Symptome</u>
Fieber: 1.3	Blutdruck: 0.5
Mandeln: 1.1	Haarausfall: 0.6
Leukozytenzahl: 1.04	Fieber: 0.65
Blutzucker: 1.03	Fibrinogenwert: 0.93
Hämatokritwert: 1.03	Blutzucker: 0.99
	Oestrogene: 0.992

Abbildung 2.1: mögliche Ausgabe des Algorithmus 4

system, das mit *Diagnoserelevanzen*, *Diagnoserelevanzfaktoren* und *Diagnoserelevanztupeln* arbeiten kann. Das *Diagnoseunterstützungssystem* ist mit einem zentralen Datenbankserver verbunden, der seine Daten über *Diagnosen* und *Symptome* über das Internet zur Verfügung stellt.

A empfängt einen Patienten P. Schnell stellt er fest, daß P ein schwieriger Fall ist. A vermutet zunächst, daß P an der Tropenkrankheit K leidet. Erste Untersuchungen von Symptomen gibt A in sein DUS (*Diagnoseunterstützungssystem*) ein in Form von Wertepaaren (*Symptom*, *Symptomausprägung*). Das DUS meldet sich zurück mit einer Auflistung von *Diagnosewahrscheinlichkeiten* für einige Krankheiten (s. Abbildung 2.2)

Mit diesem Ergebnis ist A noch nicht zufrieden. Zum einen ist ihm die *Diagnosesicherheit* von 70.2% zu wenig, zum anderen gibt es noch die *Diagnosemöglichkeiten* D31 und T mit ähnlich hohen *Diagnosesicherheiten*. A ist kein Tropenexperte und ist daher froh, daß ihn das DUS auf die Idee gebracht hat, daß auch D31 und T als *Diagnose* durchaus in Frage kommen.

Er vermutet trotzdem, daß P unter K leidet. Dies möchte er durch weitere Untersuchungen bestätigen. Er glaubt, daß eine Untersuchung von Symptom S11 Aufklärung bringt darüber, woran P tatsächlich erkrankt ist. Er läßt sich nun von seinem DUS den *Diagnoserelevanzfaktor* $\mathcal{R}_{\odot}(S11)$ berechnen. Das Ergebnis von 1.2 bestärkt ihn in seinem Vorhaben.

A führt nun die Untersuchung nach S11 durch und stellt die Ausprägung h fest. Mit dieser neuen Information füttert er das *Diagnoseunterstützungssystem* und erhält die Ausgabe von Abbildung 2.3.

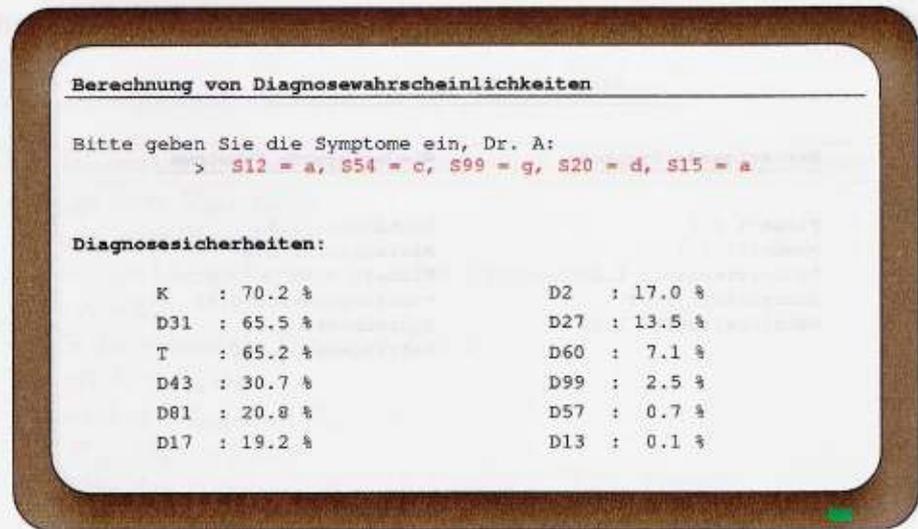


Abbildung 2.2: Diagnosewahrscheinlichkeiten nach erster Untersuchung

Die Diagnosesicherheit hat sich nicht erhöht sondern mehr als halbiert. Er weiß aber nun, daß K nicht mehr in Frage kommt. A prüft nun welche Symptome die Diagnosesicherheit von T vergrößern könnten. Er möchte die Symptome $S88$, $S77$ und $S66$ untersuchen. Für das erstere ist eine CT^4 notwendig, für das zweite eine Blutprobe, für das dritte eine Biopsie⁵.

Aufgrund des Diagnoserelevanzfaktors von 1.2 für $S77$ entscheidet sich der Arzt A , die Blutprobe zuerst durchzuführen. Er stellt weiter fest, daß er P die physiologischen und psychischen Belastungen der Biopsie ersparen kann, da die in der entfernten Datenbank inhärent gespeicherten Erfahrungen von Medizinern weltweit zeigen, daß eine Biopsie zwar zur Erklärung der Krankheit P beitragen kann, aber nur in geringem Maße.

A stellt bei der Blutprobe für $S77$ die Ausprägung f fest und gibt auch diese in das DUS ein. Abbildung 2.5 zeigt das Ergebnis.

Die Diagnosesicherheit für T ist A hoch genug, und es gibt auch keine weiteren Diagnosemöglichkeiten, die eine hohe Diagnosewahrscheinlichkeit haben. A informiert P über die Diagnose T und leitet die Behandlung ein.

Die Diagnoserelevanzen $\mathcal{R}_\Delta(S_x)$ und $\mathcal{R}_\rightarrow(S_x)$ und der Diagnoserelevanzfaktor $\mathcal{R}_\odot(S_x)$ haben einige interessante Konsequenzen für den effizienten Einsatz medizinischer Informationssysteme.

Sie können, wie wir in diesem Kapitel bereits erwähnten, eingesetzt wer-

⁴Computertomographie

⁵Gewebeprobe

Berechnung von Diagnosewahrscheinlichkeiten			
Bitte geben Sie die Symptome ein, Dr. A:			
> S12 = a, S54 = c, S99 = g, S20 = d, S15 = a			
S11 = d			
Diagnosesicherheiten:			
T	: 76.3 %	D43	: 10.3 %
D31	: 70.8 %	D60	: 9.1 %
K	: 30.2 %	D57	: 4.3 %
D81	: 21.6 %	D27	: 3.5 %
D2	: 17.1 %	D99	: 2.7 %
D17	: 13.9 %	D13	: 0.1 %

Abbildung 2.3: Diagnosewahrscheinlichkeiten nach zweiter Untersuchung

den, um die Diagnosesicherheit zu erhöhen oder Diagnosevermutungen zu entkräften.

Dadurch können

1. von vorne herein medizinische Tests wegen zu geringer Diagnoserelevanz ausgeschlossen und die Kosten für diese Tests eingespart werden, wobei das Verhältnis von Diagnoserelevanz zu den entstehenden Kosten von Bedeutung ist. So findet man zum Thema Lebensversicherungsmedizin: „Für die versicherungsmedizinische Risikoeinschätzung stehen praktisch nur die vom Antragssteller beantworteten Gesundheitsfragen oder ein ärztlicher Untersuchungsbericht und Hausarztbericht (...) zur Verfügung. Diese Unterlagen reichen aber in einem Teil der Fälle zu einer Risikoeinschätzung nicht aus. (...) In diesen und vielen anderen Fällen ist es noch möglich, durch entsprechende Zusatzuntersuchungen ergänzungswürdige Befunde weiter zu klären. [...] Auch die Kostenfrage ist bei der Anordnung von Zusatzuntersuchungen zu berücksichtigen, da die Untersuchungskosten zur Beitragsleistung des Versicherungsnehmers in einer gewissen Relation stehen müssen.“ [Rae80]
2. gezielt Testreihen geplant werden, die die Diagnose weiter bestätigen oder widerlegen, indem man für alle noch nicht in die Betrachtung eingebrachten Symptome (d.h. alle möglichen S_{n+1}) die daraus folgenden neuen Diagnosesicherheiten berechnet und die, die mit der größten Wahrscheinlichkeitsrelevanz aufwartenden Symptome mittels geeigne-

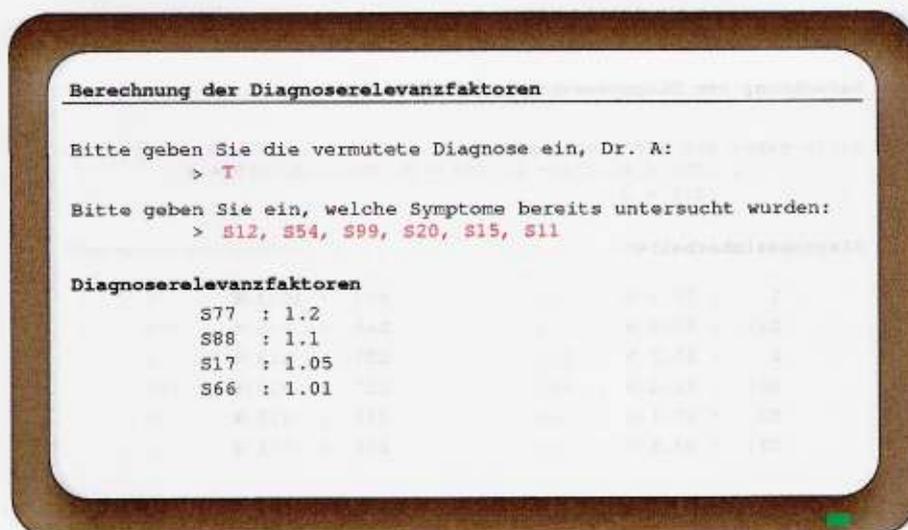


Abbildung 2.4: Diagnoserelevanzfaktoren

ter medizinischer Tests einbringt unter Beachtung der in Punkt 1 genannten Kostenaspekte. Alternativ können selbstverständlich statt der Diagnoserelevanzen \mathcal{R}_Δ die Diagnoserelevanzfaktoren \mathcal{R}_\odot eingesetzt werden. Eine entsprechende Software könnte z.B. alle möglichen weiteren zu prüfenden Symptome auswerten und in nach Diagnoserelevanzfaktoren geordnete Tabelle mit den Inhalten Diagnoserelevanzfaktor, Kosten des Tests bzw. der Untersuchung, Belastung des Patienten etc. anbieten.

3. Patienten sowohl physiologisch als auch psychisch entlastet werden. Durch die systematische Untersuchungsplanung können Untersuchungen so ausgewählt sind, daß sie minimal invasiv sind⁶. Durch eine solche Untersuchungsplanung könnten vielleicht einige Untersuchungen gespart werden, so daß der Patient eben weniger Untersuchungen über sich ergehen lassen müßte und die damit verbundenen psychischen Belastungen des Patienten in manchen Fällen reduziert werden könnten
4. Ärzte eine formale Begründung vorlegen, warum sie welche Untersuchungen durchgeführt haben (nämlich aufgrund großer Diagnoserelevanzen) und warum sie die endgültige Diagnose gestellt haben (hohe Diagnosesicherheit konnte aus der Statistik gewonnen werden und es

⁶Beispielsweise durch Vermeidung von Biopsien, Darmspiegelungen und sonstiger unangenehmer Untersuchungsmethoden, oder durch Ersatz durch weniger invasive Untersuchungen mit ähnlich hohen Diagnoserelevanzen

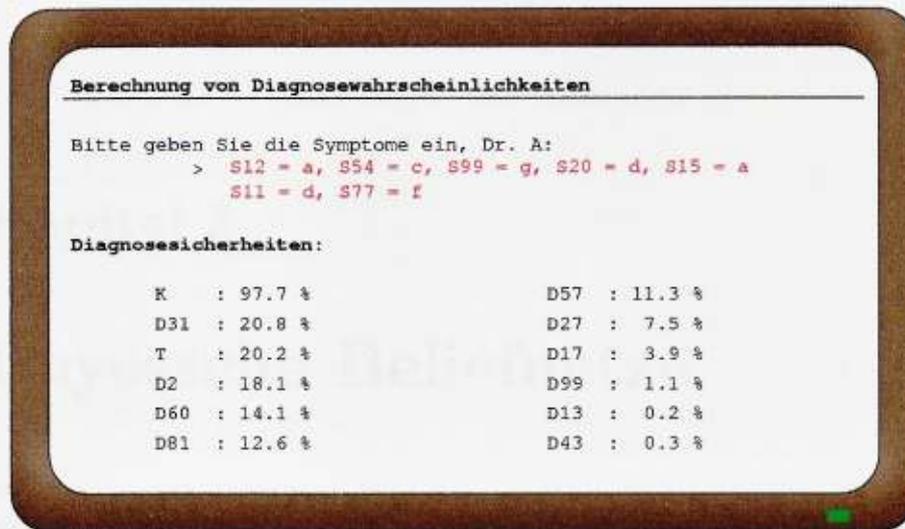


Abbildung 2.5: Diagnosewahrscheinlichkeiten nach dritter Untersuchung

gab keine weiteren Symptome, die eine hohe Diagnoserelevanz haben).

2.4 Probleme

Problematisch bleibt immer noch die fehlende Beachtung von Unsicherheiten bei der Feststellung der Symptome. Ferner wurde das Wissen über kausale Zusammenhänge zwischen Symptomen/Krankheiten noch nicht in die Berechnungen einbezogen. Die Darstellung der Berechnungen sowie der verwendeten Daten bleiben weiterhin schlecht visualisierbar. Eine gute visuelle Präsentation ist wichtig für das Verständnis der Vorgänge im Diagnoseunterstützungssystem, damit für das Vertrauen in die Ergebnisse, die das DUS hervorbringt, und aufgrund dessen letztlich von entscheidender Bedeutung für die Benutzungshäufigkeit von DUS.

Im folgenden Kapitel stellen wir Bayessche Beliefnetze vor. Diese erlauben eine Visualisierung von Wissen (von Kausalzusammenhängen) und die Einbeziehung unsicherer oder unbekannter Information.

The image shows a document page with a table. The table has two columns. The left column contains numerical values, and the right column contains text labels. The text is mirrored and difficult to read due to the image quality.

Diagnosewert	Klassifikationswert
0,00	0,00
0,01	0,01
0,02	0,02
0,03	0,03
0,04	0,04
0,05	0,05
0,06	0,06
0,07	0,07
0,08	0,08
0,09	0,09
0,10	0,10

Abbildung 2.3: Diagnosewertebenen nach dritter Untersuchung

gab keine weiteren Symptome, die eine hohe Diagnosewert haben

2.4 Probleme

Problematisch bleibt immer noch die fehlende Bestimmung von Ursache- und Wirkungszusammenhängen. Die Darstellung der Beziehungen zwischen den verschiedenen Symptomen/Klassifikationswerten ist nicht in die Form eines Entscheidungsbaums überführt worden. Die Darstellung der Beziehungen zwischen den verschiedenen Symptomen/Klassifikationswerten ist nicht in die Form eines Entscheidungsbaums überführt worden. Die Darstellung der Beziehungen zwischen den verschiedenen Symptomen/Klassifikationswerten ist nicht in die Form eines Entscheidungsbaums überführt worden.

Im folgenden Kapitel sollen wir versuchen, die Bedeutung von Diagnosewertebenen (von Klassifikationswerten) und die Bedeutung von Diagnosewertebenen (von Klassifikationswerten) zu verdeutlichen.

Kapitel 3

Bayessche Beliefnetze

In diesem Kapitel beschäftigen wir uns mit einer Form der Repräsentation von und Inferenz über Wissen: Bayessche Beliefnetze. Bayessche Beliefnetze dienen einerseits der graphischen Darstellung von Informationen, insbesondere unsicheren oder vagen Informationen, und können als Wissensbasis in Expertensystemen eingesetzt werden. Andererseits sind mit Bayesschen Beliefnetzen gewisse Methoden und Formalismen entwickelt worden, mittels derer auch auf diesem Wissen inferiert, resonniert bzw. geschlußfolgert werden kann.

Wir wollen in diesem Kapitel – und das wurde uns mit zunehmendem Fortschritt der Arbeit ein immer größeres Bedürfnis – eine Einführung in Bayessche Beliefnetze geben, die klar und verständlich ist. Wir versuchen uns damit der in der Belief-Netze-Literatur scheinbar bereits zum guten Stil gewordenen didaktischen Mangelhaftigkeit zu widersetzen. Uns ist aufgefallen, daß die meisten der verwendeten Quellen mittlere bis schwere Fehler und Auslassungen aufweist und sie sich (nicht allein) deshalb völlig unbrauchbar für das Erlangen von Erkenntnis über die Anwendung Bayesscher Beliefnetze erweisen. Besonders zu empfehlen ist zur Einführung in das Thema der Bericht [KBPS95]. Eine ausführliche formale Behandlung von Beliefnetzen findet man in [Pea88]¹.

Wir gehen weiterhin anhand eines durchgehenden Beispiels (in der Literatur leider ebenfalls Mangelware!) auf die spezifischen Probleme für insbesondere medizinische Diagnosesysteme bei der Modellierung Bayesscher Beliefnetze ein und geben Lösungsvorschläge und/oder -strategien an.

Abschnitt 1 dieses Kapitels beschreibt zunächst den qualitativen und quantitativen Aufbau von Beliefnetzen. Die Initialisierung wird in Abschnitt 2 besprochen. In Abschnitt 3 stellen wir die Propagierung von Wahrscheinlichkeiten in Bayesschen Netzen vor. Zu jedem dieser Abschnitte werden

¹nur dem hartgesottenen Leser zu empfehlen

wir auf die Probleme, die sich in der Anwendung zu insbesondere medizinischen Diagnosezwecken ergeben eingehen und schlagen Lösungen vor. In Abschnitt 4 dieses Kapitels gehen wir darauf ein, welche Merkmale Netze zur medizinischen Diagnose aufweisen können.

Zur Darstellung von Vagheit und Ungewißheit werden in Bayesschen Beliefnetzen Wahrscheinlichkeitswerte verwendet, dabei vor allem bedingte Wahrscheinlichkeiten verwendet. Dieses Vorgehen ist nicht allein in Bayesschen Beliefnetzen zu finden. So schreibt [TS89] allgemein über die Merkmale wissensbasierter Diagnosesysteme:

„Vages, unsicheres Wissen und die Behandlung von Evidenzen² wird als [...] Bewertung der Gewißheit des Ergebnisses in Abhängigkeit von den Gewißheitsgraden der Bedingungen ausgedrückt.“

Alternativ zu dem Begriff „Bayessches Beliefnetz“ werden auch die Begriffe „Beliefnetz“, „Bayessches Netz“ und „kausal probabilistisches Netz“ verwendet.

3.1 Aufbau von Beliefnetzen

Ein Bayessches Beliefnetz ist ein gerichteter azyklischer Graph³ $G(V, E)$. Dabei ist V eine Menge von Knoten („engl. vertexes“) und E eine Menge von geordneten Tupeln (V^x, V^y) , $V^x \in V$, $V^y \in V$. In einem Beliefnetz repräsentiert jedes Element von V eine Zufallsvariable und jedes Tupel $(V^x, V^y) \in E$ eine direkte Abhängigkeit [Pea88] der Zufallsvariable V^y von V^x . Abbildung 3.1 zeigt den qualitativen Aufbau eines Bayesschen Beliefnetzes.

Jede Zufallsvariable V^x kann mehrere Werte f_i annehmen. Wir bezeichnen das Vorliegen des Ereignisses $V^x = f_i$ als Hypothese.

3.1.1 Qualitativer Aufbau

Wir möchten nun den qualitativen Aufbau eines Beliefnetzes anhand eines Beispiels aus der Medizin darstellen. Das dazu benötigte Informationsmaterial stammt aus [MS62], [ALS], [BDH⁺54a] und [BDH⁺54b]. Dargestellt wird durch das Netz in Abbildung 3.1 der folgende Sachverhalt.

Das Vorhandensein der Krankheit **Tuberkulose** kann eine Zerstörung oder

²Der Begriff stammt ab vom lateinischen „evidencia“, was soviel bedeutet wie „Augenfälligkeit“. Eine Evidenz wird beschrieben als „innere Gewißheit der Gültigkeit einer Erkenntnis“ ([Lin73]). Auch [Pea88] benutzt den englischen Begriff „evidence“, der sich als „Offenkundigkeit“ oder „Klarheit“ übersetzen läßt.

³Ein Graph ist eine „Modell zur Beschreibung von Objekten die untereinander in Beziehung stehen“ [Dud93]. Gerichtete Graphen werden auch als „Digraph“ ([Aig96]), azyklische Digraphen als „DAG“ bezeichnet.

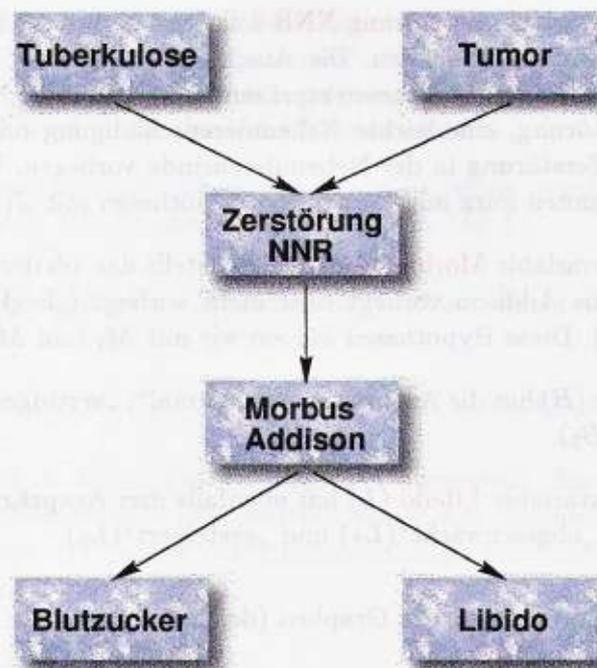


Abbildung 3.1: Beispielnetz (qualitativ) für dieses Kapitel

zumindest Schädigung der **Nebennierenrinde (NNR)** nach sich ziehen. Auch ein **Tumor** kann die Nebennierenrinde **angreifen**. Eine zu großen Teilen zerstörte Nebennierenrinde **verursacht** das Krankheitsbild **Morbus Addison**. Dieses wiederum **äußert sich in einer Erhöhung** des **Blutzuckerspiegels**. In seltenen Fällen **ist dies auch verantwortlich** für eine Abschwächung der **Libido** des Patienten.

Rot markiert sind all jene Dinge, bei denen Ereignisse auftreten können. Für jede rote Markierung kann eine Zufallsvariable und damit ein Knoten im Netz erzeugt werden. Hier benötigen wir 6 Zufallsvariablen und dazu korrespondierend 6 Knoten:

- Die Zufallsvariable **Tuberkulose** (zukünftig mit S abgekürzt, da Tuberkulose auch als Schwindsucht bezeichnet wird) kann die Ausprägungen „liegt vor“ und „liegt nicht vor“ annehmen. Dadurch werden zwei Hypothesen dargestellt: 1. der Patient leidet unter Tuberkulose, und 2. der Patient leidet nicht unter Tuberkulose. Wir bezeichnen diese Hypothesen auch mit S_1 und S_2 .
- Die Zufallsvariable **Tumor** (zukünftig mit T abgekürzt) kann ebenfalls die Ausprägungen „liegt vor“ und „liegt nicht vor“ annehmen. Analog zu S bezeichnen wir die dargestellten Hypothesen mit T_1 und T_2 .

- Die Zufallsvariable **Zerstörung NNR** soll den Grad der Schädigung der Nebennierenrinde darstellen. Die Ausprägungen „nein“, „leicht“ und „schwer“ sollen die Hypothesen repräsentieren, daß keine Nebennierenrindenzerstörung, eine leichte Nebennierenschädigung oder ein hoher Grad der Zerstörung in der Nebennierenrinde vorliegen. Wir bezeichnen den Knoten kurz mit Z bzw. die Hypothesen mit Z_1 , Z_2 und Z_3 .
- Die Zufallsvariable **Morbus Addison** (M) stellt dar, ob der Krankheitsbild Morbus Addison vorliegt oder nicht vorliegt („liegt vor“, „liegt nicht vor“). Diese Hypothesen kürzen wir mit M_1 und M_2 ab.
- **Blutzucker** (B) hat die Ausprägungen „normal“, „verringert“ und „erhöht“ (B_1 , B_2 , B_3).
- Die Zufallsvariable **Libido** (L) hat ebenfalls drei Ausprägungen: „normal“ (L_1), „abgeschwächt“ (L_2) und „gesteigert“ (L_3).

Damit ist die Knotenmenge des Graphen (des Beliefnetzes)

$$V = \{S, T, Z, M, B, L\} \quad (3.1)$$

ermittelt.

Die Kanten zwischen den Knoten sind kausale Beziehungen (Ursache-Wirkungs-Kanten). Es ist von entscheidender Bedeutung, sich bei der Modellierung von kausal probabilistischen Netzen darüber klar zu werden, was die Ursache und was die Wirkung ist, um die Richtung der Kanten korrekt in den Graph einzutragen. Die gerichteten Kanten führen immer von der Ursache weg hin zur Wirkung⁴. Die Zusammenhänge, für die Kanten modelliert werden müssen, sind im darzustellenden Sachverhalt (s.o.) grün markiert.

- **„nach sich ziehen“**: Eine Tuberkulose kann einer Zerstörung der NNR nach sich ziehen und kann damit Ursache einer NNR-Schädigung sein. Dieser Zusammenhang wird durch eine gerichtete Kante von S nach Z modelliert. Diese Kante ist ein Tupel (S, Z) .
- **„angreifen“**: Ein Tumor kann die NNR angreifen. Ein Tumor ist also auch eine Ursache, deren Wirkung die Schädigung der NNR ist. Wir fügen daher die Kante (T, Z) ein.
- **„verursacht“**: Eine Zerstörung der NNR verursacht Morbus Addison. Wir fügen eine Kante (Z, M) ein.

⁴Es sei an dieser Stelle darauf hingewiesen, daß die Forderung, das Netz als DAG darstellen zu müssen, jede Modellierungen von Rückkopplungen (Wirkung verstärkt Ursache) ausschließt, was sich als ein Mangel erweisen wird.

- „**äußert sich**“: Morbus Addison hat einen Einfluß auf die Ausprägung „erhöht“ der Variable „Blutzucker“. Die Variable „Blutzucker“ enthält also die Wirkung von Morbus Addison. Zur Modellierung dieses Sachverhalts wird die Kante (M, B) verwendet.
- „**ist verantwortlich**“: Morbus Addison hat einen negativen Einfluß auf die Libido. Wir benötigen also die Kante (M, L) .

Wir haben nun qualitativ das medizinische Wissen qualitativ modelliert als Graph:

$$G = (\{S, T, Z, M, B\}, \{(S, Z), (T, Z), (Z, M), (M, B), (M, L)\}) \quad (3.2)$$

Der Graph ist in Abbildung 3.1 graphisch dargestellt. Ein Bayessches Beliefnetz unterscheidet sich aber deutlich von einem allgemeinen DAG durch die Möglichkeit, die Stärke der Zusammenhänge zwischen den Knoten zu quantifizieren.

3.1.2 Quantitativer Aufbau

Bisher ist das Beliefnetz nur eine graphische Notation dafür, daß gewisse Zusammenhänge und Abhängigkeiten existieren. Zur Quantifizierung der Sicherheit bzw. Unsicherheit der durch die Knoten repräsentierten Hypothesen werden diesen nun Wahrscheinlichkeiten zugeordnet (vgl. Zitat von [TS89] auf Seite 34).

Bei der Modellierung des quantitativen Aufbaus des Beliefnetzes ordnen wir den Knoten unter anderem Wahrscheinlichkeitsverteilungen zu. Wir unterscheiden zwischen Wurzelknoten⁵ und inneren Knoten⁶.

Wurzelknoten.

In unserem Beispielnetz existieren zwei Wurzelknoten S und T . Wir ordnen jeder Zufallsvariable in einem Wurzelknoten eine Wahrscheinlichkeitsverteilung zu [Pea90]. Dazu wird zu jeder Hypothese der Zufallsvariable eine a-priori-Wahrscheinlichkeit für das Eintreffen der selben angegeben.

Die Krankheit Tuberkulose kommt nur selten vor. Daher wählen wir für den Knoten S :

$$p(S_1) = 0.01 \quad (3.3)$$

$$p(S_2) = 0.99 \quad (3.4)$$

⁵Knoten ohne Vorgängerknoten

⁶Knoten mit mindestens einem Vorgänger

Dies ist nur eine Abschätzung. Reales Zahlenmaterial liegt uns zur Zeit nicht vor. Die Abschätzung stellt die Gewißheit dar, mit der die Hypothesen S_1 und S_2 zutreffen. Die zugeordneten Wahrscheinlichkeiten für die Hypothesen gelten in einem Beliefnetz jedoch nur so lange, wie keine weiteren Informationen in das Netz einfließen, in diesem Fall also so lange, wie ein Patient noch nicht vorstellig und untersucht wurde.

Die Hypothesen S_1 und S_2 sind disjunkt⁷ und zur Zufallsvariable S kann außer S_1 und S_2 keine weitere disjunkte Hypothese gefunden werden. Damit entspricht $S_1 \vee S_2$ dem sicheren Ereignis, woraus folgt, daß

$$p(S_1 \vee S_2) = \sum_{i=1}^2 p(S_i) = 1 \quad (3.5)$$

gelten muß.

Auch wenn dies nirgends explizit gemacht wird, geht man für jeden Knoten K in einem Beliefnetz mit k Hypothesen K_i , $i \in \mathbb{N}$, $1 \leq i \leq k$, zunächst davon aus, daß

$$\sum_{i=1}^k p(K_i) = 1 \quad (3.6)$$

ist, die Hypothesen also eine vollständige Ereignisdisjunktion⁸ bilden.

Der zweite Wurzelknoten T wird ebenfalls mit a-priori-Wahrscheinlichkeiten belegt. Wir belegen T_1 mit einer höheren Wahrscheinlichkeit als S_1 , da ein Tumor (ebenfalls hypothetisch) häufiger vorkommt als die heute seltene Schwindsucht.

$$p(T_1) = 0.02 \quad (3.7)$$

$$p(T_2) = 0.98 \quad (3.8)$$

Auch hier sind die Werte von uns nur abgeschätzt. Konkrete Zahlen aus der Wirklichkeit liegen uns nicht vor.

Die Wahrscheinlichkeitsverteilung von a-priori-Wahrscheinlichkeiten einer Zufallsvariable K mit k Hypothesen K_1, \dots, K_k wird auch oft in Form eines Vektors

$$\vec{p}(K) = \begin{pmatrix} p(K_1) \\ p(K_2) \\ \vdots \\ p(K_k) \end{pmatrix} \quad (3.9)$$

⁷d.h. sie schließen sich gegenseitig aus

⁸der gesamte Ergebnisraum der Zufallsvariable wird abgedeckt, ohne daß sich die Hypothesen überschneiden

dargestellt. Wir wollen diese Notation verwenden und schreiben

$$\vec{p}(S) = \begin{pmatrix} 0.01 \\ 0.99 \end{pmatrix} \quad (3.10)$$

$$\vec{p}(T) = \begin{pmatrix} 0.02 \\ 0.98 \end{pmatrix} \quad (3.11)$$

Innere Knoten.

Alle Knoten, die mindestens einen Elternknoten⁹ besitzen, werden als innere Knoten bezeichnet. In einem Beliefnetz bedeutet das Vorhandensein eines Elternknotens, daß der innere Knoten bzw. die durch ihn dargestellten Hypothesen abhängig sind bzw. beeinflusst werden von anderen Hypothesen, nämlich denen des oder der Elternknoten ([Pea90]).

Der Zusammenhang zwischen einem Knoten K und einem Vorgängerknoten V stellt in einem qualitativen DAG eine Regel der Form

„Wenn K , dann V “

dar. Im Unterschied dazu sieht eine solche Regel in einem Bayesschen Beliefnetz wie folgt aus (mit K_i einer Hypothese des Knotens K und V_u einer Hypothese des Knotens V):

„Wenn K_i , dann V_j mit einer Gewißheit/Wahrscheinlichkeit.“

Um die Verteilung der durch den inneren Knoten repräsentierten Zufallsvariable in Abhängigkeit von anderen Zufallsvariablen darzustellen, werden aus diesem Grund bedingte Wahrscheinlichkeiten benutzt.

Hat ein Knoten K Hypothesen K_i und n Vorgängerknoten V^j mit den Hypothesen $V_{u_j}^j$, so muß jede mögliche bedingte Wahrscheinlichkeit

$$p \left(K_i \mid \bigwedge_{j=1}^n V_{u_j}^j \right) \quad \forall i, \forall u_j \quad (3.12)$$

In Abbildung 3.1 liegen 4 innere Knoten vor (Z , M , B und L). Betrachten wir zunächst M genauer. Morbus Addison ist (laut unseres einfachen Beliefnetzes) nur abhängig davon, ob die Nebennierenrinde geschädigt ist oder nicht. Je weiter die Schädigung der Nebennierenrinde (Z) fortgeschritten ist, desto höher ist die Wahrscheinlichkeit, daß das Krankheitsbild Morbus

⁹„Elternknoten“ ist ein dem Wort „Vorgängerknoten“ äquivalenter Begriff

Addison erscheint. Dagegen kann Morbus Addison nicht vorliegen, wenn die Nebennierenrinde noch in gesundem Zustand ist. Die bedingten Wahrscheinlichkeiten für Morbus Addison belegen wir daher wie folgt:

$$p(M_1|Z_1) = 0.0, \quad p(M_2|Z_1) = 1.0 \quad (3.13)$$

$$p(M_1|Z_2) = 0.3, \quad p(M_2|Z_2) = 0.7 \quad (3.14)$$

$$p(M_1|Z_3) = 0.9, \quad p(M_2|Z_3) = 0.1 \quad (3.15)$$

Diese Wahrscheinlichkeiten können auch als Tabelle (s. Tabelle 3.1) dargestellt werden.

Morbus Addison	Zerstörung NNR		
	nein	leicht	schwer
liegt vor	0.0	0.3	0.9
liegt nicht vor	1.0	0.7	0.1

Tabelle 3.1: Wahrscheinlichkeiten für Zustände von Morbus Addison

Eine weitere Schreibweise ist die Matrixnotation:

$$\text{link}(M|Z) = \begin{pmatrix} 0.0 & 0.3 & 0.9 \\ 1.0 & 0.7 & 0.1 \end{pmatrix} \quad (3.16)$$

Die Matrix *link* wird als Verbindungsmatrix („link matrix“) bezeichnet. Die Werte der Verbindungsmatrix $\text{link}(M|Z)_{i,u}$ entsprechen den bedingten Wahrscheinlichkeiten $p(M_i|Z_u)$.

Die bedingten Wahrscheinlichkeiten können aus Statistiken stammen oder abgeschätzt werden. Wir schätzen die Inhalte der Verbindungsmatrizen in dieser Arbeit ab, indem wir Worten wie „meist“, „gelegentlich“, „selten“ o.ä. aus medizinischen Lehrbüchern in Form höherer oder niedrigerer Wahrscheinlichkeitswerte Ausdruck und mathematische Greifbarkeit verleihen. Die Abschätzungen sind selbst äußerst grob. In der praktischen Anwendung wird man dazu Spezialisten zu Rate ziehen müssen, wenn keine Statistiken vorliegen.

Eine solche Verbindungsmatrix wird jedem inneren Knoten zugeordnet. Die Anzahl der Dimensionen der Verbindungsmatrix ist festgelegt durch die Anzahl der Elternknoten (d.h. der bedingenden Zufallsvariablen). Hat ein Knoten K n Vorgängerknoten V_i , $1 \leq i \leq n$, so hat die Verbindungsmatrix $n+1$ Dimensionen. Es ist jedoch darauf zu achten, daß unter jeder Bedingung die Wahrscheinlichkeiten für die Hypothesen in K 1 ergeben (s. dazu auch S. 38), denn man verlangt auch hier, daß die Hypothesen in K eine vollständige

Ereignisdisjunktion bilden. Habe K k Hypothesen K_i , $1 \leq i \leq k$, und haben die n Vorgängerknoten V^j , $1 \leq j \leq n$ die Hypothesen $V_{j_u}^j$, dann muß gelten

$$\sum_{i=1}^k \text{link}(K|V^1, V^2, \dots, V^n)_{i, j_1, j_2, \dots, j_n} \quad (3.17)$$

Für den Knoten M gilt dies:

$$\begin{aligned} \sum_{i=1}^2 \text{link}(M|Z)_{i,1} &= \sum_{i=1}^2 \text{link}(M_i|Z_1) \\ &= \text{link}(M_1|Z_1) + \text{link}(M_2|Z_1) \\ &= 0.0 + 1.0 \\ &= 1.0 \end{aligned} \quad (3.18)$$

und analog für die Hypothesen Z_2 und Z_3 .

Die Knoten B und L haben als einzigen Elternknoten M . Wir ordnen den Knoten daher wieder Wahrscheinlichkeitsverteilungen bzw. Wahrscheinlichkeitsmatrizen mit je zwei Dimensionen zu (Tabellen 3.2 und 3.3). Durch Morbus Addison wird der Blutzuckerspiegel in den meisten Fällen erhöht ($p(B_3|M_1) = 0.94$). Selten findet man dabei auch eine Abschwächung der Libido ($p(L_2|M_1) = 0.2$ gegenüber $p(L_2|M_2) = 0.1$, die geringe Wahrscheinlichkeitserhöhung spiegelt die Seltenheit des Ereignisses wieder).

Blutzucker	Morbus Addison	
	liegt vor	liegt nicht vor
B_1 , normal	0.05	0.88
B_2 , verringert	0.01	0.07
B_3 , erhöht	0.94	0.05

Tabelle 3.2: Wahrscheinlichkeiten für Knoten B

Libido	Morbus Addison	
	liegt vor	liegt nicht vor
L_1 , normal	0.7	0.8
L_2 , abgeschwächt	0.2	0.1
L_3 , gesteigert	0.1	0.1

Tabelle 3.3: Wahrscheinlichkeiten für Zustände von L

Der letzte zu quantifizierende innere Knoten ist Z . Z ist jedoch von zwei Knoten abhängig: S und T . Wir benötigen nun eine dreidimensionale Ma-

trix $link(Z|S, T)$. In dieser Tabelle stehen die Wahrscheinlichkeiten dafür, daß eine bestimmte Hypothesen von Z eintritt unter der Bedingung, daß eine bestimmte Hypothese aus S und zugleich eine bestimmte Hypothese aus T wahr sind (logische Konjunktion). Tabelle 3.4 zeigt diese dreidimensionale Matrix und Abbildung 3.2 gibt eine Zusammenfassung der bisherigen Arbeiten an.

Zerstörung NNR	S_1		S_2	
	T_1	T_2	T_1	T_2
Z_1 , nein	0.01	0.05	0.1	0.98
Z_2 , leicht	0.09	0.25	0.3	0.01
Z_3 , schwer	0.9	0.7	0.6	0.01

Tabelle 3.4: Wahrscheinlichkeiten für die Zerstörung der NNR

Das Beliefnetz ist nun fertig modelliert. Wie wir bereits erwähnten handelt es sich hier um ein stark vereinfachendes Modell der Wirklichkeit. Es dient in dieser Arbeit als Erklärungsmodell und in der Anwendung¹⁰ als Entscheidungsmodell.

Die Verbindungsmatrizen zu den Tabellen lauten:

$$Z = \begin{pmatrix} 0.01 & 0.05 & 0.1 & 0.98 \\ 0.09 & 0.25 & 0.3 & 0.01 \\ 0.9 & 0.7 & 0.6 & 0.01 \end{pmatrix} \quad (3.19)$$

$$M = \begin{pmatrix} 0.0 & 0.3 & 0.9 \\ 1.0 & 0.7 & 0.1 \end{pmatrix} \quad (3.20)$$

$$B = \begin{pmatrix} 0.5 & 0.88 \\ 0.01 & 0.07 \\ 0.94 & 0.05 \end{pmatrix} \quad (3.21)$$

$$L = \begin{pmatrix} 0.7 & 0.8 \\ 0.2 & 0.1 \\ 0.1 & 0.1 \end{pmatrix} \quad (3.22)$$

3.2 Initialisierung von Beliefnetzen

In diesem Abschnitt gehen wir darauf ein, wie den in Beliefnetzen dargestellten Zufallsvariablen, sprich den Knoten, Wahrscheinlichkeitsverteilungen zu-

¹⁰in realiter wäre ein derart stark vereinfachstes Modell wohl kaum adäquat

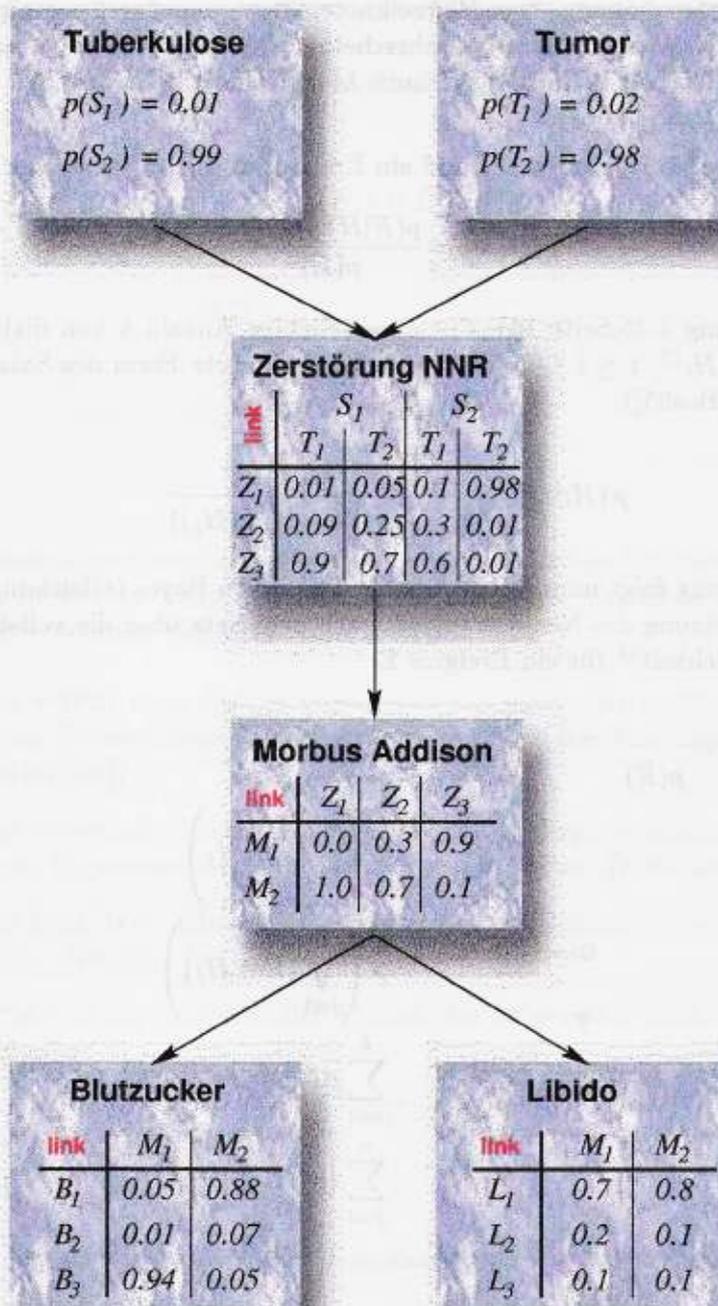


Abbildung 3.2: Quantifiziertes Beliefnetz

geordnet werden können. Den Wurzelknoten bzw. den durch sie repräsentierten Variablen sind ja bereits Wahrscheinlichkeitsverteilungen¹¹ zugeteilt worden. In diesem Abschnitt wird auch klar, woher die Bayesschen Beliefnetze ihren Namen haben.

Für eine einzelne Hypothese H und ein Ereignis E gilt der Satz von Bayes

$$p(H|E) = \frac{p(E|H) \cdot p(H)}{p(E)} \quad (3.23)$$

(vgl. Gleichung 1.35, Seite 13). Für eine beliebige Anzahl k von disjunkten Hypothesen H_i ¹², $1 \leq i \leq k$, gilt der verallgemeinerte Form des Satzes von Bayes (vgl. [Bos95]):

$$p(H_i|E) = \frac{p(H_i) \cdot p(E|H_i)}{\sum_{j=1}^k (p(H_j) \cdot p(E|H_j))} \quad (3.24)$$

Diese Gleichung folgt unmittelbar aus dem Satz von Bayes (Gleichung 3.23) und der Ersetzung des Nenners ($p(E)$) nach dem Satz über die vollständige Wahrscheinlichkeit¹³ für ein Ereignis E :

$$\begin{aligned}
 p(E) & \stackrel{\text{Gl. 1.3}}{=} p(E \wedge \Omega) \\
 & \stackrel{\text{s.o.}}{=} p\left(E \wedge \bigvee_{j=1}^k H_j\right) \\
 & \stackrel{\text{Gesetze der Logik}}{=} p\left(\bigvee_{j=1}^k (E \wedge H_j)\right) \\
 & \stackrel{\text{Gl. 1.4}}{=} \sum_{j=1}^k p(E \wedge H_j) \\
 & \stackrel{\text{Gl. 1.14}}{=} \sum_{j=1}^k (p(H_j) \cdot p(E|H_j)) \quad (3.25)
 \end{aligned}$$

(vgl. hierzu auch [Bos95]).

Gleichung 3.24 kann umgeformt werden zu

$$BEL(K_i) = \alpha \cdot \lambda(K_i) \cdot \pi(K_i) \quad (3.26)$$

¹¹nämlich a-priori-Wahrscheinlichkeiten

¹²damit Gleichung 3.24 gilt, verlangt man sogar, daß die Hypothesen H_i eine vollständige Ereignisdisjunktion bilden

¹³auch Satz von der totalen Wahrscheinlichkeit genannt

mit K_i einer Hypothese (wir verwenden den Bezeichner K , um darauf hinzuweisen, daß es sich um eine Hypothese aus einem Knoten eines Beliefnetzes handelt). Die Umwandlung ergibt sich aus den folgenden Ersetzungsmaßnahmen:

$$BEL(K_i) = P(H_i|E) \quad (3.27)$$

$$\alpha = \frac{1}{\sum_{j=1}^k (P(H_j) \cdot P(E|H_j))} \quad (3.28)$$

$$\lambda(K_i) = P(E|H_i) \quad (3.29)$$

$$\pi(K_i) = P(H_i) \quad (3.30)$$

Diese Ersetzungen werden nicht allein aus syntaktischen Gründen vollzogen, sondern sie isoliert Werte mit besonderer Semantik, um damit die Sinnhaftigkeit des weiteren Vorgehens zu unterstreichen:

- der π -Wert einer Zufallsvariable gibt die gegenwärtige Stärke der kausalen Unterstützung einer Hypothese durch die Vorgängerknoten an ([KBPS95]),
- der λ -Wert gibt die gegenwärtige Stärke der diagnostischen Unterstützung einer Hypothese durch die Nachfolgerknoten an ([KBPS95],
- der BEL -Wert gibt das Gesamtvertrauen („BELief“) in eine Hypothese an ([Pea90]).
- α ist nur ein normierender Faktor, der so gewählt wird, daß die Bedingung

$$\sum_{i=1}^k BEL(K_i) = 1 \quad (3.31)$$

gilt (vgl. [Hec95]).

Abbildung 3.3 verdeutlicht den Zusammenhang der Gleichungen 3.26 und 3.24.

Nun stellt in einem Beliefnetz jeder Knoten mehrere Hypothesen dar. Da jede Hypothese einen eigenen BEL -, π - und λ -Wert erhält, werden wir diese, wie bereits die Wahrscheinlichkeiten auf den vorhergehenden Seiten, zu Vektoren $\vec{BEL}(K)$, $\vec{\pi}(K)$ und $\vec{\lambda}(K)$ komprimieren. Wir verwenden dazu im Gegensatz zu allen anderen im Literaturverzeichnis angegebenen Quellen die Notation mit Vektorpfeilen, da das Auslassen der selben erfahrungsgemäß gerade am Anfang nicht unerhebliche Verwirrung stiftet.

$$BEL(K_j) = \alpha \pi(K_j) \lambda(K_j)$$

$$P(H_j | E) = \frac{P(H_j) P(E | H_j)}{\sum_{k=1}^n P(H_k) P(E | H_k)}$$

Abbildung 3.3: Äquivalenz der Gleichungen 3.26 und 3.24

3.2.1 Initialisierung der Wurzelknoten

Ein Wurzelknoten K läßt sich initialisieren, indem man zunächst in den Vektor $\vec{\pi}(K)$ die a-priori-Wahrscheinlichkeiten einsetzt ([Pea88]). $\pi(K_i)$ beschreibt im allgemeinen die kausale Unterstützung der Hypothese K_i durch den Vorgängerknoten. Wurzelknoten haben zwar keine Vorgängerknoten, nach Abschnitt 1.1.4 stehen a-priori-Wahrscheinlichkeiten aber in einem allgemeinen Kontext, der (theoretisch) als Elternknoten des Wurzelknoten eingefügt werden könnte. Daher ist die Erklärung von π als „kausale Unterstützung durch den Vorgängerknoten“ auch im Fall der Wurzelknoten durchaus sinnvoll. Der Kontext entspricht der Ursache für die kausale Unterstützung π .

In unserem Beispiel gilt für die Wurzelknoten

$$\vec{\pi}(S) = \begin{pmatrix} \pi(S_1) \\ \pi(S_2) \end{pmatrix} = \begin{pmatrix} p(S_1) \\ p(S_2) \end{pmatrix} = \begin{pmatrix} 0.01 \\ 0.99 \end{pmatrix} \quad (3.32)$$

$$\vec{\pi}(T) = \begin{pmatrix} 0.02 \\ 0.98 \end{pmatrix} \quad (3.33)$$

λ entspricht laut Gleichung 3.29 der Wahrscheinlichkeit für das Zutreffen einer Hypothese im Wurzelknoten unter der Bedingung, daß eine Evidenz des Nachfolgeknotens vorliegt. Über den Nachfolgeknoten ist diese Information jedoch noch nicht bekannt. Daher setzen wir für $\vec{\lambda}$ zunächst eine Gleichverteilung ein ([KBPS95]). Die Größe der Werte in $\vec{\lambda}$ ist nicht von Interesse. Wichtig ist nur, daß alle Komponenten des Vektors den gleichen Wert haben. Wir setzen hierfür den 1-Vektor (alle Komponenten 1) ein.

$$\vec{\lambda}(S) = \begin{pmatrix} \lambda(S_1) \\ \lambda(S_2) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (3.34)$$

$$\vec{\lambda}(T) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (3.35)$$

$\vec{\lambda}(K)$ erhält eine Gleichverteilung (alle Hypothesen gleiche Werte), da es noch keine Evidenzen in unserem Netz gibt. Dies bedeutet nicht, daß die kausale Unterstützung durch die Nachfolgerknoten 0 wäre, sondern, daß keiner Hypothese in K eine höhere diagnostische Unterstützung zukommt als einer anderen: darin kann der Einsatz einer Gleichverteilung begründet werden.

Nun, da die π - und λ -Werte vorliegen, kann das Gesamtvertrauen BEL berechnet werden. Dazu verwenden wir den \times -Operator, der die Vektoren Zeilenweise multipliziert.

$$\begin{aligned} \overrightarrow{BEL}(S) &= \begin{pmatrix} BEL(S_1) \\ BEL(S_2) \end{pmatrix} \\ &= \begin{pmatrix} \alpha \cdot \pi(S_1) \cdot \lambda(S_1) \\ \alpha \cdot \pi(S_2) \cdot \lambda(S_2) \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot 1 \cdot 0.01 \\ 1 \cdot 1 \cdot 0.99 \end{pmatrix} \\ &= \begin{pmatrix} 0.01 \\ 0.99 \end{pmatrix} \end{aligned} \quad (3.36)$$

Alle Komponenten der Vektoren werden mit dem gleichen α multipliziert. Daher führen wir die Multiplikation von α als skalare Multiplikation mit den Vektoren durch. Dies führt zum gleichen Ergebnis.

$$\overrightarrow{BEL}(T) = \alpha \cdot \vec{\pi}(T) \times \vec{\lambda}(T) \quad (3.37)$$

$$\begin{aligned} &= \alpha \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \times \begin{pmatrix} 0.02 \\ 0.98 \end{pmatrix} \\ &\stackrel{\text{mit } \alpha=1}{=} \begin{pmatrix} 0.02 \\ 0.98 \end{pmatrix} \end{aligned} \quad (3.38)$$

Damit sind alle Wurzelknoten initialisiert.

Initialisierung der inneren Knoten

Auch den inneren Knoten werden BEL -, π - und λ -Werte zugeordnet. Der Unterschied zu den Wurzelknoten besteht darin, daß Wahrscheinlichkeiten

aus dem oder den Vorgängerknoten dazu benötigt werden ([Pea88]). Um $\pi(K_i)$ zu berechnen (s. Gl. 3.30) benötigen wir für einen Knoten K mit den Hypothesen K_i die Wahrscheinlichkeiten $p(K_i)$. Habe der Knoten K einen einzelnen Vorgängerknoten V^1 mit den v Hypothesen V_u^1 , $1 \leq u \leq v$. Dann gilt für eine Hypothese K_i der Satz von der totalen Wahrscheinlichkeit:

$$\pi(K_i) = p(K_i) = \sum_{u=1}^v (p(K_i|V_u^1) \cdot p(V_u^1)) \quad (3.39)$$

Der Satz von der totalen Wahrscheinlichkeit wird z.B. in [DE92], [Bos95] und anderen genauer beschrieben. Teilweise findet man stattdessen auch die Bezeichnung „Satz von der vollständigen Wahrscheinlichkeit“ (wie in [Bos95]). Wir möchten den Leser darauf hinweisen, daß diese Gleichung (3.39) auch für Wurzelknoten bzw. Hypothesen darin gilt. Daher ist die Initialisierung der Wurzelknoten nicht anderes als die Initialisierung eines inneren Knoten, wie man fälschlicherweise glauben könnte (und wie es auch in der Literatur zu erwähnen unterlassen wird). Für einen Wurzelknoten K entspricht der Vorgängerknoten dem allgemeinen Kontext (s. Abschnitt 1.1.4):

$$\begin{aligned} \pi(K_i) &= p(K_i) \\ &= \sum_1^1 p(K_i|\text{Kontext}) \cdot p(\text{Kontext}) \\ &= p(K_i|\Omega) \cdot p(\Omega) \\ &= p(K_i) \cdot 1 \\ &= p(K_i) \end{aligned} \quad (3.40)$$

Ein Knoten K kann natürlich (wie unser Beispiel zeigt) auch von mehreren Zufallsvariablen bzw. Knoten abhängig sein. Für einen Knoten K mit k Hypothesen K_i , $1 \leq i \leq k$ und n Vorgängerknoten V_j , $1 \leq j \leq n$, die jeweils eine Anzahl von Hypothesen v_j enthalten, gilt die Erweiterung von Gleichung 3.39:

$$\pi(K_i) = p(K_i) = \sum_{u_1} \sum_{u_2} \dots \sum_{u_n} (p(K_i|V_{u_1}^1 \wedge V_{u_2}^2 \wedge \dots \wedge V_{u_n}^n) \cdot p(V_{u_1}^1) \cdot p(V_{u_2}^2) \cdot \dots \cdot p(V_{u_n}^n)) \quad (3.41)$$

Die Werte $p(K_i|V_{u_1}^1 \wedge \dots \wedge V_{u_n}^n)$ sind per definitionem in der Verbindungsmatrix des Knoten K zu finden. Die Werte $p(V_{u_j}^j)$, $1 \leq j \leq n$ entsprechen den π -Werten aus den Knoten V_j .

In unserem Diagnosenetz gestaltet sich dies so:

$$\begin{aligned}
 \pi(Z_1) &= \sum_{s=1}^2 \sum_{t=1}^2 m(Z|S,T)_{1,s,t} \cdot \pi(S_s) \cdot \pi(T_t) \\
 &= \sum_{s=1}^2 \sum_{t=1}^2 m(Z_1|S_s, T_t) \cdot \pi(S_s) \cdot \pi(T_t) \\
 &= p(Z_1|S_1 \wedge T_1) \cdot p(S_1) \cdot p(T_1) \\
 &\quad + p(Z_1|S_1 \wedge T_2) \cdot p(S_1) \cdot p(T_2) \\
 &\quad + p(Z_1|S_2 \wedge T_1) \cdot p(S_2) \cdot p(T_1) \\
 &\quad + p(Z_1|S_2 \wedge T_2) \cdot p(S_2) \cdot p(T_2) \\
 &= 0.01 \cdot 0.01 \cdot 0.02 + 0.05 \cdot 0.01 \cdot 0.98 \\
 &\quad + 0.1 \cdot 0.99 \cdot 0.02 + 0.98 \cdot 0.99 \cdot 0.98 \\
 &= 0.000002 + 0.00049 + 0.00198 + 0.950796 \\
 &= 0.93268 \tag{3.42}
 \end{aligned}$$

$$\pi(Z_2) = 0.01811 \tag{3.43}$$

$$\pi(Z_3) = 0.028622 \tag{3.44}$$

woraus folgt (wir runden auf 3 Nachkommastellen)

$$\vec{\pi}(Z) = \begin{pmatrix} 0.953 \\ 0.018 \\ 0.029 \end{pmatrix} \tag{3.45}$$

$\vec{\lambda}(Z)$ setzen wir aus dem bekannten Grund wieder auf den 1-Vektor. Dann folgt daraus

$$\overrightarrow{BEL}(Z) = \alpha \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \times \begin{pmatrix} 0.953 \\ 0.018 \\ 0.029 \end{pmatrix} \stackrel{\alpha=1}{=} \begin{pmatrix} 0.953 \\ 0.018 \\ 0.029 \end{pmatrix} \tag{3.46}$$

Analog laßt sich der Knoten M berechnen. Er besitzt jedoch nur einen einzigen Vorgängerknoten (Z):

$$\begin{aligned}
 \pi(M_1) &= 0.0 \cdot 0.953 + 0.3 \cdot 0.018 + 0.9 \cdot 0.029 \\
 &= 0.0 + 0.0054 + 0.0261 \\
 &= 0.0315 \tag{3.47}
 \end{aligned}$$

$$\begin{aligned}
 \pi(M_2) &= 1.0 \cdot 0.953 + 0.7 \cdot 0.018 + 0.1 \cdot 0.029 \\
 &= 0.953 + 0.0126 + 0.0029 \\
 &= 0.9685
 \end{aligned} \tag{3.48}$$

Wir runden hier $\pi(M_1)$ korrekt auf 3 Stellen auf, runden aber, damit die Summe der π -Werte 1 ergibt, d.h. aus didaktischen Gründen, $\pi(M_2)$ ab. Damit ist

$$\overrightarrow{BEL}(M) = \alpha \cdot \vec{\lambda}(M) \times \vec{\pi}(M) \stackrel{\alpha=1}{=} \begin{pmatrix} 0.032 \\ 0.968 \end{pmatrix} \tag{3.49}$$

Die Werte für Blutzucker B und Libido L werden ebenfalls analog berechnet. Das Ergebnis, sprich das fertig initialisierte Belief-Netz zeigt Abbildung 3.4.

3.3 Propagierung

3.3.1 Evidenzen.

An einem initialisierten Beliefnetz können Wahrscheinlichkeiten für das Zutreffen von Hypothesen abgelesen werden. Entscheidend ist dafür der \overrightarrow{BEL} -Vektor eines Knotens, der das Gesamtvertrauen in die einzelnen Hypothesen des Knoten angibt. Damit kann das Netz prinzipiell bereits als Informationsträger – oft wird ein solcher Informationsträger übermütig als „Wissensbasis“¹⁴ bezeichnet – benutzt werden.

Die im initialisierten Beliefnetz ablesbaren Wahrscheinlichkeiten sind jedoch für sich allein stehend nicht sonderlich interessant. Man könnte zwar einem Patienten sagen, daß der Patient mit einer Wahrscheinlichkeit von 3.2% unter Morbus Addison leidet, dies wird ihm aber verständlicherweise nicht weiterhelfen. Der Patient wird stattdessen untersucht bzw. es wird ein persönliches Gespräch mit ihm geführt. Dies dient dazu, sich Klarheit über Anamnese und Befund zu verschaffen (wir haben bereits den den Seiten 7f. festgestellt, was alles dazu gehören kann).

Die gewonnene Information (Klarheit über einen Sachverhalt) wird unter den Statistikern als Evidenz¹⁵ bezeichnet.

Für ein Beliefnetz bedeutet eine Evidenz, daß Informationen in Form von Wahrscheinlichkeiten hinzugefügt wird oder vorhandene Informationen verändert werden. Diese Informationen sind Werte eines Vektors

¹⁴Der wesentliche Unterschied zwischen Daten im herkömmlichen Sinn und Wissen ist, daß eine Wissensbasis auch prozedurales Wissen enthalten kann (vgl. [TS89]).

¹⁵aus dem lateinischen „evidencia“, was soviel wie „Klarheit“, „Augenscheinlichkeit“ oder „Offenkundigkeit“ bedeutet

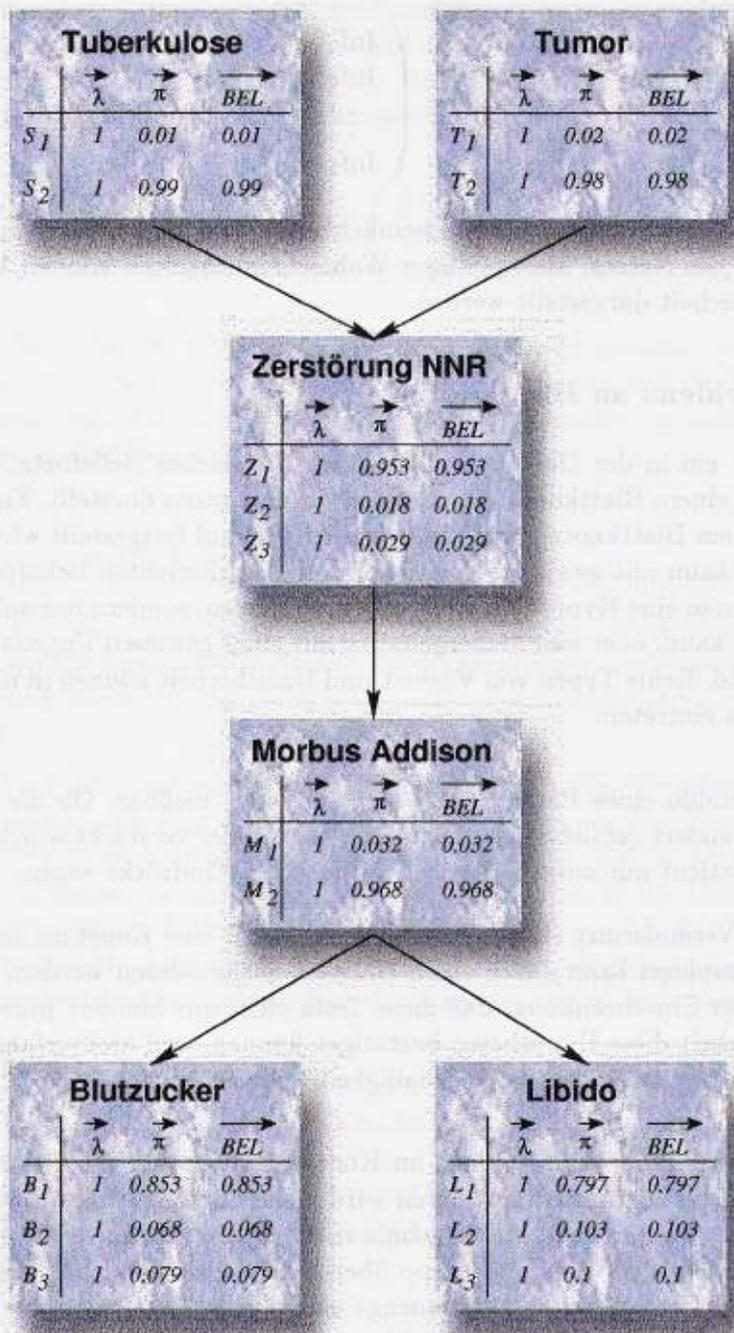


Abbildung 3.4: Initialisiertes Beliefnetz

$$Inf = \begin{pmatrix} Inf_1 \\ Inf_2 \\ \vdots \\ Inf_k \end{pmatrix} \quad (3.50)$$

Inf_i , $1 \leq i \leq k$ sind dabei Wahrscheinlichkeiten von Hypothesen in einem der Knoten des Netzes. Mittels dieser Wahrscheinlichkeiten können Vagheit und Unsicherheit dargestellt werden.

3.3.2 Evidenz an Blattknoten.

Typisch für ein in der Diagnose eingesetztes Bayessches Beliefnetz ist eine Evidenz an einem Blattknoten¹⁶, da dieser ein Symptom darstellt. Eine Evidenz an einem Blattknoten tritt ein, wenn ein Befund festgestellt wird. Dieser Befund kann mit gewissen Vagheiten und Unsicherheiten behaftet sein, sei es, weil man eine Hypothese nicht objektiv messen, sondern nur subjektiv einschätzen kann, oder weil Meßergebnisse mit einer gewissen Ungenauigkeit behaftet sind. Beide Typen von Vagheit und Unsicherheit können in unserem Beispielnetz eintreten:

- Die Libido eines Patienten ist nicht objektiv meßbar. Ob die Libido unverändert geblieben ist, nachgelassen hat oder verstärkt wurde, kann der Patient nur aufgrund seiner subjektiven Eindrücke sagen.
- Eine Veränderung (Erhöhung, Senkung) oder eine Konstanz im Blutzuckerspiegel kann durch einen Bluttest nachgewiesen werden, jedoch mit der Einschränkung, daß diese Tests nicht mit hundert prozentiger Sicherheit diese Hypothesen bestätigen können, weil Meßverfahren immer mit einer gewissen Ungenauigkeit behaftet sind.

Greifen wir zunächst eine Evidenz im Knoten Blutzucker für die weiteren Erklärungen heraus: Einem Patienten wird eine Blutprobe entnommen und diese im Labor untersucht. Als Ergebnis tritt ein Blutzucker von $150.5 \text{ mg}\%$ ein. Läge dieser Wert liegt nur knapp über dem Normalbereich¹⁷, wäre hier aufgrund der kontinuierlichen Wertemenge eine stetige Zufallsvariable „Blutzucker“ wünschenswert. Wir haben sie aber aus didaktischen Gründen in drei Hypothesen diskretisiert und handeln uns eventuell (nämlich z.B. bei einem Blutzuckerspiegel von $120.5 \text{ mg}\%$) damit das Problem ein, welcher Hypothese (B_1 , B_2 , B_3) der Blutzuckerspiegel zugeordnet werden soll, denn

¹⁶Knoten ohne Nachfolgerknoten

¹⁷Nach der Methode von Hagedorn-Jensen, Frank-Kirberger, Folin-Wu liegt der Normalbereich bei $80 \text{ mg}\%$ bis $120 \text{ mg}\%$

die Ränder des Normalbereiches sind „unscharf“ (nicht exakt definierbar)¹⁸. Wir ordnen aufgrund der Tatsache, daß der Wert des Blutzuckers recht weit von dem Normalbereich entfernt ist, dem Knoten Blutzucker den Vektor (und damit die Wahrscheinlichkeitsverteilung)

$$\vec{\lambda} = \begin{pmatrix} 0.01 \\ 0.01 \\ 0.98 \end{pmatrix} \quad (3.51)$$

zu. Der Wert für die Hypothesen B_1 und B_2 wurde gewählt, weil z.B. ein Test dem Patienten falsch zugeteilt worden sein könnte, weil die Blutproben verwechselt sein könnten, oder weil der Blutzuckertest an sich mit einer gewissen Vagheit behaftet sein könnte.

3.3.3 λ -Nachrichten

Man ordnet diese Wahrscheinlichkeitsverteilung zum λ -Vektor zu, weil gerade dieser die gegenwärtige Stärke der diagnostischen (retrospektiven) Unterstützung darstellt, die ein Knoten von seinen Nachfolgern erhält (vgl. [KSH91]).

$\vec{\lambda}$ ist wohl der problematischste Vektor bezüglich des Verständnisses der Berechnungen, die durch die Beliefnetze ausgeführt werden. Die Erklärung nach [KSH91] (s.o.) findet sich auch in [KBPS95]. Sie ist aber dahingehend irreführend, da danach $\vec{\lambda}(K)$ in einem Knoten K die Unterstützung der Hypothesen in K durch die Nachfolger von K angibt. Zur entgeltigen Klärung, was denn der $\vec{\lambda}$ -Vektor genau bedeutet, sind noch einige wenige Zwischenüberlegungen notwendig. Wir kommen also gleich wieder darauf zurück.

Durch das Eintreten der Evidenz wird der Knoten, an dem die Evidenz eingetreten ist, aktiviert. Aktivierte Knoten rechnen, inaktive Knoten befinden sich in einer Ruhephase. Ein aktivierter Knoten muß die BEL -Werte nach Gleichung 3.26 neu berechnen ([BWW97]), da sich der BEL -Vektor (im allgemeinen $\vec{\lambda}$ oder auch $\vec{\pi}$) geändert hat.

$$\overline{BEL}(L) = \alpha \cdot \begin{pmatrix} 0.01 \\ 0.01 \\ 0.98 \end{pmatrix} \times \begin{pmatrix} 0.853 \\ 0.068 \\ 0.079 \end{pmatrix}$$

¹⁸Die Wertemenge für den Blutzuckerspiegel kann man also als 3 unscharfe Mengen betrachten, wovon eine den Normalbereich darstellt. Dies führt letztendlich auf die Fuzzy-Mengen-Theorie, die sich gerade mit diesem Problem beschäftigt. Dem interessierten Leser empfehlen wir hierzu das einführende Buch [Gra95].

$$\alpha = 11.5433 \cdot \begin{pmatrix} 0.00853 \\ 0.00068 \\ 0.07742 \end{pmatrix} = \begin{pmatrix} 0.098 \\ 0.008 \\ 0.894 \end{pmatrix} \quad (3.52)$$

Alle Endergebnisse werden von uns auf drei Nachkommastellen gerundet. Daher können durchaus Rundungsfehler entstehen, von denen wir hier jedoch abstrahieren wollen.

Da nun bekannt ist, daß der Blutzucker mit großer Wahrscheinlichkeit erhöht ist, liegt kein Grund mehr vor anzunehmen, daß die Wahrscheinlichkeit des Vorliegens von Morbus Addison noch so klein ist wie vorher, da ein erhöhter Blutzucker als Symptom von Morbus Addison auch Indikator für selbigen ist. Der Knoten M muß also von den Änderungen in B informiert werden.

Ein Knoten K sendet bei einer Änderung seiner BEL -Werte Nachrichten an seine Nachbarknoten ([Pea88]). Die Vorgängerknoten erhalten λ -, die Nachfolgerknoten π -Nachrichten ([Pea88]). Da B keine Nachfolgerknoten besitzt, sendet er nur für jeder Hypothese in M eine λ -Nachricht an M . Die λ -Nachrichten, die ein Knoten K mit k Hypothesen K_i an einen Vorgängerknoten V sendet, betreffen stets eine Hypothese V_j des Vorgängerknotens. Nach [KSH91] lauten diese Nachrichten

$$\begin{aligned} \lambda_K(V_j) &= \sum_{i=1}^k m(K|V)_{i,j} \cdot \lambda(K_i) \\ &= \sum_{i=1}^k p(K_i|V_j) \cdot \lambda(K_i) \end{aligned} \quad (3.53)$$

Die Nachrichten lassen sich wieder zu einem Vektor

$$\vec{\lambda}_K(V) = \begin{pmatrix} \lambda_K(V_1) \\ \lambda_K(V_2) \\ \vdots \\ \lambda_K(V_v) \end{pmatrix} \quad (3.54)$$

mit v der Anzahl der Hypothesen in V zusammenfügen. Nun wird der Knoten K wieder deaktiviert.

Der Knoten „Blutzucker“ sendet demnach den Vektor

$$\vec{\lambda}_B(M) = \begin{pmatrix} \sum_{i=1}^3 m(B|M)_{i,1} \lambda(B_i) \\ \sum_{i=1}^3 m(B|M)_{i,2} \end{pmatrix}$$

$$\begin{aligned}
 &= \begin{pmatrix} 0.05 \cdot 0.098 + 0.01 \cdot 0.008 + 0.94 \cdot 0.894 \\ 0.88 \cdot 0.098 + 0.07 \cdot 0.008 + 0.05 \cdot 0.894 \end{pmatrix} \\
 &= \begin{pmatrix} 0.845 \\ 0.1315 \end{pmatrix} \quad (3.55)
 \end{aligned}$$

3.3.4 Verarbeitung einer λ -Nachricht

Durch das Senden der Nachrichten an den Vorgängerknoten entsteht bei diesem wiederum eine Evidenz [Pea88] und er wird aktiviert. Da die Evidenz durch eine Nachricht ausgelöst wurde, wird diese zunächst verarbeitet.

Wenn an einem Knoten K für eine Hypothese K_i von m Nachfolgerknoten W^l λ -Nachrichten eingehen, so verarbeitet K diese nach [KBPS95] nach der Gleichung

$$\lambda(K_i) = \prod_{l=1}^m \lambda_{W^l}(K_i) \quad (3.56)$$

K berechnet also seine neuen λ -Werte aus dem Produkt der eingehenden λ -Nachrichten. Danach verarbeitet K eventuelle π -Nachrichten, auf die wir aus gegebenen Anlaß (in diesem Beispiel kommen am aktivierten Knoten M keine π -Nachrichten an) noch verzichten. Wie holen die Besprechung später nach.

Im Knoten M gehen nur Nachrichten von einem einzigen Nachfolger ein. Dies könnte im Beispielnetz auch anders sein: Zu der Evidenz in B hätte auch eine Evidenz in L hinzukommen können. L hätte dann ebenfalls λ -Nachrichten an M gesendet. So aber vereinfacht sich die Verarbeitung der λ -Nachrichten im Knoten M zu (wir fassen die K_i wieder zu einem Vektor zusammen)

$$\vec{\lambda}(M) = \vec{\lambda}_B(M) = \begin{pmatrix} 0.845 \\ 0.1315 \end{pmatrix} \quad (3.57)$$

Ein Knoten muß nach Verarbeitung der Nachrichten wieder seine BEL -Werte anpassen:

$$\begin{aligned}
 \overline{BEL}(M) &= \alpha \cdot \begin{pmatrix} 0.845 \\ 0.1315 \end{pmatrix} \times \begin{pmatrix} 0.032 \\ 0.968 \end{pmatrix} \\
 &= \alpha \cdot \begin{pmatrix} 0.02704 \\ 0.127292 \end{pmatrix} \\
 &\stackrel{\alpha=6.47954}{=} \begin{pmatrix} 0.175 \\ 0.825 \end{pmatrix} \quad (3.58)
 \end{aligned}$$

Wie wir sehen, ist das Vertrauen in die Hypothese M_1 (Morbus Addison liegt vor) zwar immer noch gering¹⁹, jedoch schon fast vier mal so groß wie bisher.

Nun, da bekannt ist, wie λ -Nachrichten entstehen und verarbeitet werden, können wir auch genauer erklären, was die λ -Nachrichten bedeuten. Ein Knoten K habe m Nachfolgerknoten W^l , die jeweils l_w Hypothesen enthalten. K enthält wieder k Hypothesen K_i , $1 \leq i \leq k$.

K berechnet seine $BEL(K_i)$ -Werte nach Gleichung 3.26. Dazu wird der Wert $\lambda(K_i)$ benötigt, welcher die Wahrscheinlichkeit $p(E|K_i)$ darstellt (vgl. Gleichung 3.29) mit E_W einer Evidenz im Nachfolgerknoten W ([Pea88]). Bei mehreren Evidenzen E_W in den Nachfolgerknoten stellt dementsprechend $\lambda(K_i)$ die Wahrscheinlichkeit

$$p\left(\bigwedge_W E_W | K_i\right) = \prod_W p(E_W | K_i) \quad (3.59)$$

dar.

Daraus erklärt sich Gleichung 3.56. Die Wahrscheinlichkeiten $p(E_W | K_i)$ entsprechen prinzipiell den Wahrscheinlichkeitsverteilungen in den Verbindungsmatrizen der Nachfolgerknoten W ([Pea88]). Da diese im Knoten K nicht zugreifbar sind²⁰, müssen sie von W an K gesendet werden.

Wie in [Pea88] nachzulesen ist, kann $\lambda(W)$ trotzdem nicht direkt aus der Verbindungsmatrix von W kopiert werden, weil W indirekt von einem weiteren Knoten X (d.h. K, W, X bilden eine Kette) beeinflusst wird. Die Evidenz E , die dann bei K eintrifft, ist dann eine Kombination der Evidenzen E_W in W und in E_X in X , nicht eine Evidenz in W allein. Darum gilt

$$\begin{aligned} \lambda(K_i) = p(E|K_i) &= \sum_{E_W} p(E|E_W \wedge K_i) \cdot p(E_W | K_i) \\ &= \sum_{E_W} p(E|K_i) \cdot p(E_W | K_i) \\ &= m(E_W | K_i) \cdot \lambda(E_W) \end{aligned} \quad (3.60)$$

wodurch die Gleichung 3.53 entsteht.

¹⁹ was nicht verwundert, da Morbus Addison aufgrund selten vorkommender Ursachen auch eher selten vorkommt und ein erhöhter Blutzucker bei sehr vielen Krankheiten auftreten kann

²⁰ das Berechnungsparadigma von [Pea88] gibt vor, daß jeder Knoten für sich autonom ist; vergleichbar mit Objekten aus dem Paradigma der Objektorientierung verfügt ein Knoten nur über seine eigenen Daten. Daten aus anderen Knoten (Objekten) müssen durch Nachrichten (Methodenaufrufe) ausgetauscht werden

3.3.5 π -Nachrichten

Kommen wir zurück zur Propagierung der Evidenz im Knoten „Blutzucker“. Die *BEL*-Werte von *M* haben sich geändert.

Der nun aktivierte Knoten *M* sendet wiederum, genau wie eben noch *B*, λ -Nachrichten an seinen Vorgänger. *M* erzeugt aber auch Nachrichten an seine Nachfolger.

Ein Knoten *K* mit Nachfolgerknoten W^l sendet nach seiner Aktivierung und der Berechnung neuer *BEL*-Werte π -Nachrichten an alle Nachfolgerknoten W^l außer an die, durch die die Aktivierung ausgelöst wurde ([KBPS95]). Für jede Hypothese K_i wird eine solche Nachricht an die Nachfolgerknoten weitergeleitet. Die Nachricht π von K_i an die m Nachfolgerknoten W^l wird erzeugt nach [KBPS95] mittels der Gleichung

$$\pi_{W^l}(K_i) = \pi(K_i) \cdot \prod_{r=1, r \neq l}^m \lambda_{W^r}(K) \quad (3.61)$$

Die an die Nachbarknoten weitergeleiteten Nachrichten bestehen also aus dem Produkt aller eingehenden λ -Nachrichten und dem π -Wert des Knotens. Dadurch ist gewährleistet, daß alle Knoten im Beliefnetz, auch die, für die es keinen Pfad durch das Netz zum Knoten, an dem die Evidenz aufgetreten ist, gibt.

Auch die π -Nachrichten schreiben wir als Vektor. *M* erzeugt nur einen solchen Vektor, den er an *L* sendet. *B* erhält keine π -Nachricht, da *B* die Änderung in *M* hervorgerufen hat. So berechnet *M* die Nachricht

$$\begin{aligned} \vec{\pi}_L(M) &= \begin{pmatrix} \pi(M_1) \cdot \lambda_B(M_1) \\ \pi(M_2) \cdot \lambda_B(M_2) \end{pmatrix} \\ &= \begin{pmatrix} 0.032 \cdot 0.845 \\ 0.968 \cdot 0.1315 \end{pmatrix} \end{aligned} \quad (3.62)$$

M sendet diesen Vektor an den Nachfolgerknoten *L*. Zusätzlich sendet er einen λ -Vektor. Dieser wird analog zum λ -Vektor von *B* an *M* berechnet. Dann wird *M* deaktiviert.

3.3.6 Verarbeitung der π -Nachricht

L empfängt nun die π -Nachricht. Daraus werden die *L*-eigenen π -Nachrichten berechnet. Empfängt ein Knoten *K* mit k Hypothesen K_i von seinen n Vorgängerknoten V^n , die jeweils über Hypothesen v_n $V_{v_n}^n$ verfügen, so berechnet sich $\pi(K_i)$ nach [KSH91] aus der Gleichung

$$\pi(K_i) = \sum_{u_1=1}^{v_1} \dots \sum_{u_n=1}^{v_n} M(K_i|V_{u_1}^1, \dots, V_{u_n}^n) \cdot \pi_K(V_{u_1}^1) \cdot \dots \cdot \pi_K(V_{u_n}^n) \quad (3.63)$$

was einer Verallgemeinerung des Satzes von der totalen Wahrscheinlichkeit entspricht. In unserem Beispiel rechnen wir daher für den Knoten L :

$$\begin{aligned} \pi(L_1) &= p(L_1|M_1) \cdot \pi_L(M_1) + p(L_1|M_2) \cdot \pi_L(M_2) \\ &= 0.7 \cdot 0.027 + 0.8 \cdot 0.127 \\ &= 0.121 \end{aligned} \quad (3.64)$$

$$\begin{aligned} \pi(L_2) &= p(L_2|M_1) \cdot \pi_L(M_1) + p(L_2|M_2) \cdot \pi_L(M_2) \\ &= 0.2 \cdot 0.027 + 0.1 \cdot 0.127 \\ &= 0.018 \end{aligned} \quad (3.65)$$

$$\begin{aligned} \pi(L_3) &= p(L_3|M_1) \cdot \pi_L(M_1) + p(L_3|M_2) \cdot \pi_L(M_2) \\ &= 0.1 \cdot 0.027 + 0.1 \cdot 0.127 \\ &= 0.015 \end{aligned} \quad (3.66)$$

Daraus folgt der neue Belief-Vektor des Knotens L :

$$\begin{aligned} \overrightarrow{BEL} &= \alpha \cdot \begin{pmatrix} 0.121 \\ 0.018 \\ 0.015 \end{pmatrix} \times \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 0.785 \\ 0.117 \\ 0.097 \end{pmatrix} \end{aligned} \quad (3.67)$$

L hat keine Nachfolgerknoten und keine Vorgängerknoten außer M , so daß der Knoten L zunächst stabil bleibt.

Fassen wir nun den Verlauf von Nachrichten zusammen. Wir nehmen an, es würde eine Evidenz im Knoten Blutzucker stattfinden. B wird aktiviert und sendet λ -Nachrichten an M . B wird deaktiviert. M wird aktiviert und sendet π -Nachrichten an L sowie λ -Nachrichten an Z . M wird deaktiviert. L und Z werden aktiviert. L sendet keine Nachrichten und wird deaktiviert. Z sendet λ -Nachrichten an S und T . Z wird deaktiviert. Dafür werden S und T aktiviert. Beide senden jedoch keine Nachrichten mehr und werden wieder deaktiviert und alle Wahrscheinlichkeiten des Netzes sind wieder konsistent.

3.4 Probleme und Lösungsvorschläge

3.4.1 a-priori-Wahrscheinlichkeiten

Im Idealfall stammen die a-priori-Wahrscheinlichkeiten, die zur Initialisierung der Wurzelknoten benötigt werden, aus Statistiken, aus denen hervorgeht, wie wahrscheinlich das Vorliegen einer Krankheit ist, bevor man den Patienten untersucht hat. Es kann sich dabei um den Anteil der Kranken an der Weltbevölkerung, an der Bevölkerung eines Landes, an der Anzahl der bisher behandelten Patienten o.ä. sein.

Die Wahl der Grundmenge ist hier von besonderer Bedeutung, da „unbedingte“ Wahrscheinlichkeiten vorliegen, die jedoch wie bereits auf Seite 1 erwähnt, nur in einem allgemeinen Kontext gelten.

Es werden jedoch keine Angaben über den Grad der Allgemeinheit oder Art des allgemeinen Kontextes gemacht. Gerade der räumliche Kontext enthält für die Medizin viele Informationen über die a-priori-Wahrscheinlichkeiten, da bestimmte Krankheiten oder Symptome bzw. die Häufigkeit der selben großen regionalen Schwankungen ausgesetzt sein können.

Ein Beispiel für die regionalen Schwankungen ist die Tuberkulosesterblichkeit in den Regierungsbezirken Osnabrück und Oppeln aus [BDH⁺54a] (s. Abbildung 3.5). Ein weiteres Beispiel aus Finnland (ebenfalls nach [BDH⁺54a]) zeigt Abbildung 3.6. Die Tuberkulosesterblichkeit war dort in ländlichen Gebieten zeitweilig nur halb so groß wie in den Städten.

Solch drastische Differenzen haben einen entsprechenden Einfluß auf die Wurzelknoten und auf das gesamte Beliefnetz. Es fragt sich nun, welchen a-priori-Wahrscheinlichkeiten im Falle des Vorliegens mehrerer unterschiedlicher Statistiken zu bevorzugen sind.

Wenn eine Auswahl verschiedener Statistiken und einhergehend Kontexten existiert (vielfach wird man nur eine einzige oder gar keine Statistik finden, die man den Wahrscheinlichkeitswerten zugrundelegen kann²¹), so empfehlen wir den Kontext zu wählen, der die größte Semantik²² beinhaltet, die alle Fälle (Krankheitsfälle, Patienten) der konkreten Anwendungsdomäne (Praxis, Klinik, etc.) gerade noch einschließt.

Je mehr Semantik in die Wahrscheinlichkeiten des Wurzelknotens eingebracht wird, desto spezialisierter ist der Wurzelknoten. Aus diesem Grund wird die Genauigkeit der Wahrscheinlichkeiten im Beliefnetz (leider auf Kosten der Flexibilität) größer.

²¹für den Fall, daß keine Statistiken existieren, müssen die a-priori-Wahrscheinlichkeiten abgeschätzt werden. Dazu wird jedoch viel Erfahrung von Experten auf dem Gebiet benötigt.

²²im Sinne einer starken Spezialisierung der in der Statistik verwendeten Grundmenge

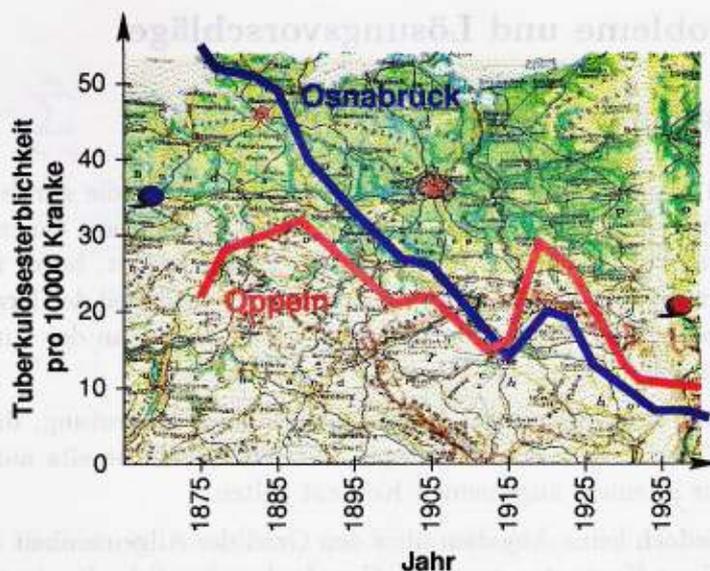


Abbildung 3.5: regionale Schwankungen der Tuberkulosesterblichkeit zwischen Osnabrück und Oppeln in Oberschlesien nach [BDH⁺54a] (leicht veränderte Darstellung)



Abbildung 3.6: regionale Schwankungen der Tuberkulosesterblichkeit in finnischen Städten und Landgemeinden nach [BDH⁺54a] (leicht veränderte Darstellung)

Soll das Beliefnetz als Wissensbasis flexibel in verschiedenen Domänen eingesetzt werden, so empfiehlt sich die Wahl eines allgemeineren Kontextes, der weniger Semantik enthält. Ein Beispiel für die Anwendung eines Netzes, deren Wahrscheinlichkeiten aus Statistiken mit niedrigerem semantischen Niveau stammen, könnte ein globales via Internet zugreifbares Diagnoseunterstützungssystem sein.

Beispiel 7 Sei in diesem rein hypothetischen Beispiel (in Anlehnung an Abbildung 3.6) die Wahrscheinlichkeit für das Auftreten von Tuberkulose in einem Staat 0.03. Seien weiterhin die Wahrscheinlichkeiten für Tuberkulose unter der Stadtbevölkerung 0.07 und unter der Landbevölkerung 0.01. Ist der Diagnostiker, der ein Beliefnetz zur Unterstützung verwendet, auf dem Lande tätig und bekommt nur selten städtischen Besuch in seiner Praxis, so sollte er die Landbevölkerungsstatistik mit $p(\text{Tuberkulose}) = 0.01$ zum quantitativen Aufbau verwenden. Diese Statistik bietet ihm genau auf seine Patienten passende Informationen. Die Staatsstatistik bietet weniger Information/Semantik an (die Information, daß die betrachtete Bevölkerung vom Lande stammt ist nicht enthalten). Die Stadtbevölkerungsstatistik kann nicht verwendet werden, da deren Grundmenge (Stadtbevölkerung) seine Kundenschaft (Landbevölkerung) nicht umfaßt.

3.4.2 Knoten für allgemeine Kontexte

Eine weitere Möglichkeit solche Informationen in das Netz einzubringen ist die Modellierung dieser Informationen als zusätzliche Knoten (qualitative und quantitative Modellierung). Für das oben genannte Beispiel könnte eine Verfeinerung des Netzes wie in Abbildung 3.7 dargestellt aussehen.

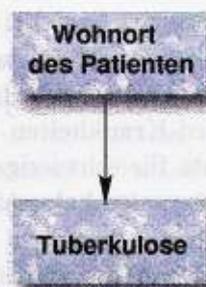


Abbildung 3.7: mögliche Verfeinerung durch Hinzufügung von Knoten (Zufallsvariablen)

Diese Lösung verursacht zum einen eine Verlagerung des Problems in andere Knoten (nämlich in den oder die neuen Wurzelknoten), zum anderen läuft

man bei der Modellierung des Beliefnetzes Gefahr, letztendlich „die Welt“ zu modellieren. Die Netze können schnell ungeahnte Größen annehmen. Dies ist der Preis dafür, entweder mit genaueren Zahlen arbeiten zu können, oder das Problem der Abschätzung der a-priori-Wahrscheinlichkeiten zu vereinfachen. Dies darf aber nicht darüber hinwegtäuschen, daß man sich nun ein weiteres neues Problem einhandelt: Die Wahrscheinlichkeit des bisherigen Wurzelknotens (der ja jetzt ein innerer Knoten ist, da ein neuer Wurzelknoten eingefügt wurde) sind jetzt keine a-priori-Wahrscheinlichkeiten mehr, sondern bedingte Wahrscheinlichkeiten (abhängig von dem neuen Wurzelknoten), die manchmal nicht weniger schwierig abzuschätzen sind. Wir kommen später auf dieses Problem zurück.

3.4.3 *link*-Matrizen

Hier ist eine gute Stelle um die Diagnostik reinzubringen, und wo man sonst Statistiken herkriegern kann

Grundsätzlich begegnet man bei der Verwendung von Statistiken den selben Problemen, wie sie schon auf Seite 59f. erläutert wurden. Zusätzlich taucht bei bedingten Wahrscheinlichkeiten die Schwierigkeit auf, daß die bedingten Wahrscheinlichkeiten weniger leicht zu erhalten sind. Benötigt werden in diesem Fall z.B. Tabellen wie auf Seite 7 dargestellt.

Tabellen dieser Art sind unseres Wissens nicht allgemein zugreifbar. Die Idee, derartige Daten über ein globales Netz (z.B. Internet) zugreifbar zu machen, eventuell mit einer Verbindung aller ärztlicher und klinischer Patientendatenbanken, in denen für die Patienten ihre Krankheiten, Krankengeschichte, Beschwerden, Symptome, Indikationen etc. gespeichert sind, ist sicherlich zur Zeit noch Zukunftsmusik.

Man mag hier einwenden, daß es nicht nötig sei, Daten aus aller Welt zu sammeln. Dies mag wohl richtig sein, wenn es sich um Standard-Erkrankungen mit entsprechender Häufigkeit handelt. Auch bei kleineren bekannten Krankheitsfällen wird sich bereits eine relative Häufigkeit stabilisiert haben, die als Wahrscheinlichkeit übernommen werden kann. Interessant sind Diagnoseunterstützungssysteme für Standard-Krankheiten aber sowieso nicht. Interessant werden sie erst beim Einsatz für schwierige Differentialdiagnosen. Dabei aber existieren eventuell sehr wenige bekannte Fälle, wie z.B. bei der vor wenigen Wochen erst an einer besonderen Form des Lassa-Fieber verstorbenen Studentin. Dieser Fall sorgte für große Schlagzeilen und bereitete den behandelnden Ärzten einiges Kopfzerbrechen. Ein auf solche Fälle spezialisiertes Diagnoseunterstützungssystem hätte vielleicht den Diagnoseprozeß beschleunigen und somit vielleicht ein Leben retten können.

Die Verbindungsmatrizen müssen meist wohl aus Ermangelung an entsprechenden Statistiken geschätzt werden. Hier kommt der Vorteil der Beliefnetze zum Tragen, die gerade derart gestaltete bedingte Wahrscheinlichkeiten verwenden, die auch geschätzt werden können, nämlich Wahrscheinlichkeiten

ten für eine Wirkung unter der Bedingung, daß eine bestimmte Ursache vorhanden ist.

Dies ist genau das, was Mediziner im Studium zu lernen pflegen und sollte damit, zumindest wenn man nur grobe Schätzungen verlangt, kein echtes Problem werden.

3.4.4 Aufbauformen

Das Wissen des Mediziners kann im einfachsten Fall in einem Netz modelliert werden, daß von einer 3-Ebenen-Struktur diktiert wird:

- Ursachen für Krankheiten
- Krankheiten
- Symptome und Indikationen

Diese 3-Ebenenstruktur kann jedem Diagnosesystem zugrunde gelegt werden, auch z.B. einem Motordiagnosesystem oder einem anderen technischen Diagnosesystem. Dabei ist nur das Wort „Krankheit“ durch „Fehler“ zu ersetzen. Innerhalb der Ebenen gibt es wieder Unterstrukturen, die ebenfalls als Subebenen die Ebene in mehrere Ebenen aufteilen. Abbildung 3.8 zeigt ein Beliefnetz zur Diagnose von Brustkrebs.

Deutlich erkennt man auch die 3-Ebenen-Struktur mit einer zusätzlichen Subebene in Ebene 3. Für den in der Abbildung dargestellten 1-Krankheits-Fall reichen auch die einfach verbundenen Bayesschen Beliefnetze aus. Schwierigkeiten mit den einfach verbundenen Bayesschen Beliefnetzen kommen aber schnell zu Tage, wenn das System mehr als nur eine Krankheit diagnostizieren können soll bzw. dafür eine Unterstützung anbieten muß. Dann führt die Situation, daß ein und die selbe Ursache nicht nur gleichzeitig einen wie auch immer gearteten Einfluß auf beide Krankheiten hat, und zugleich beide Krankheiten mindestens eine gleiche ein Symptom oder eine Indikation repräsentierende Zufallsvariable beeinflussen. Dann kommt man nicht mehr an den mehrfach verbundenen Netzen vorbei.

Mehrfach verbundene Netze werden außer in [Pea88] auch in [BFGK97] und [BWW97] beschrieben.

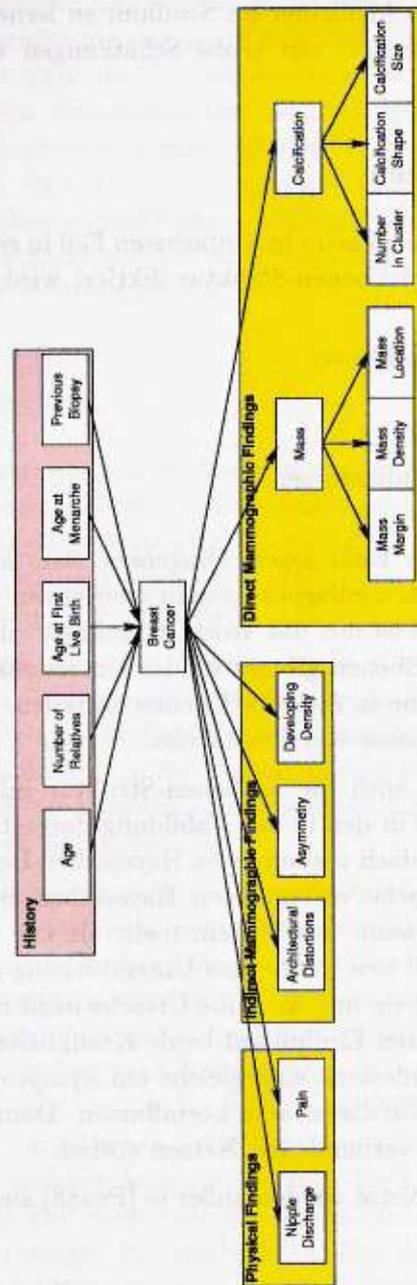


Abbildung 3.8: Beliefnetz zur Diagnose von Brustkrebs aus [KRW+99]

Kapitel 4

Faktorenanalyse

Ursprünglich wurde die Faktorenanalyse in eine enge Verbindung mit der psychologischen Intelligenz- und Persönlichkeitsforschung gebracht und in diesem Anwendungsbereich weiterentwickelt. Erst später setzte man sie auch bei anderen Fragestellungen ein. Inzwischen findet sie in verschiedenen Natur- bzw. verhaltenswissenschaftlichen Fachgebieten Verwendung, unter anderem in der Medizin ([CTC89], [FH84], [HE84]).

Die Faktorenanalyse findet insbesondere dann Anwendung, wenn im Rahmen einer statistischen Erhebung eine Vielzahl von Zufallsvariablen zu einer bestimmten Fragestellung erhoben wurden und der Anwender der Erhebung an einer Reduktion der Datenmenge interessiert ist ([BEPW90]).

Ziel der Faktorenanalyse ist es, mannigfache (mehrere) Merkmale, die zu einem zu klärenden Sachverhalt erhoben wurden, auf wenige zentrale Faktoren zu dezimieren. Es wird angestrebt, aus einer Vielzahl von Variablen voneinander unabhängige Einflußfaktoren herauszukristallisieren und damit einen Beitrag zur Entdeckung von unabhängigen Beschreibungs- und Erklärungsvariablen zu ermitteln ([BEPW90]).

Die Faktoren, von denen hier die Rede ist, sind den Zufallsvariablen sehr ähnlich. Jedoch besteht laut ([EMT95]) zwischen ihnen ein wichtiger Unterschied:

- Variablen sind direkt beobachtbare oder meßbare Werte, wie z.B. Gewicht, Temperatur, Anzahl geschenkter Blumen, Blutzuckerspiegel etc.
- Faktoren hingegen sind latente, nicht direkt beobachtbare oder meßbare Werte, wie z.B. Kreativität, Gesundheit, Liebe, Wirtschaftlichkeit etc.

4.1 Ausgangsdatenmatrix

Wir werden ebenso wie bereits in Kapitel 3 die Faktorenanalyse anhand eines durchgängigen Beispiels aus der Medizin erläutern. Eine Erhebung muß nicht unbedingt eine Befragung sein. Es kann sich um jedwede andere Art der Messung handeln. Das Zahlenmaterial für dieses Beispiel stammt aus [vL35]. Dort werden Messungen des weißen Blutbildes durchgeführt. Hinter den gemessenen Variablen (eosinophile Granulocyten, stabkernige Granulocyten, segmentkernige Granulocyten, basophile Granulocyten, Monocyten, Übergangsformen, Lymphocyten) suchen wir „nicht meßbare“ Faktoren, die das Zustandekommen der Variablenwerte erklären könnten.

Wir benötigen zunächst eine Matrix der Werte der gemessenen Variablen. In [vL35] werden die Variablen Eosinophile, Stabkernige etc. an 13 Sportstudentinnen unter anderem vor und nach 0.75 Stunden gymnastischer Übungen gemessen. Das Ergebnis dieser Messungen ist Tabelle 4.1 zusammengefaßt.

Dabei ist Entnahme I jeweils das Blutbild vor Beginn der Gymnastik, Entnahme II sind die Werte des Blutes nach der Gymnastik. Wir wollen die Blutbildveränderung untersuchen und erzeugen, in dem wir die Werte von Entnahme I von den Werten der Entnahme II subtrahieren und so die Veränderung des Blutbild während der sportlichen Betätigung erhalten, die Ausgangsdatenmatrix (Tabelle 4.2).

4.2 Korrelationsrechnung

Faktoren, die hinter den Variablen stehen, findet man durch Bündelung von Variablen. Mehrere Variablen entsprechen einem Faktor.

Beispiel 8 (Bündelung von Variablen) *Zwei Variablen „Haltbarkeit“ und „Preis“ eines Produkts können zu einem latenten Faktor „Wirtschaftlichkeit“ zusammengefaßt werden ([BEPW90]).*

Um Variablen zu finden, die man evtl. zusammenfassen könnte, sucht man nach Korrelationen zwischen den Variablen. Zwei Variablen X und Y heißen korreliert dann und genau dann, wenn die Kovarianz

$$\mu(X \cdot Y) - \mu(X) \cdot \mu(Y) \quad (4.1)$$

mit $\mu(X)$ dem Erwartungswert der Variable X , einen Wert ungleich 0 hat (vgl. [LR88]). Andernfalls heißen X und Y unkorreliert¹. Die Korrelation

¹Stochastisch unabhängige Variablen sind stets unkorreliert. Dagegen sind aber nach [LR88] nicht alle unkorrelierten Variablen stochastisch unabhängig

Name	Entnahme	Eosinophile	Stabkernige	Segmentkernige	Basophile	Monocyten	Übergangsformen	Lymphocyten
Bo.	I	2.5	2	46.5	0	5	4.5	39.5
	II	1.5	8	49	3	3	2	33.5
Mü.	I	7	4	39	1.5	2	5	41.5
	II	3	3.5	56.5	1.5	2	7.5	26
Hu.	I	2	6	47.5	1.5	1.5	3.5	38
	II	2	9.5	46	1	1.5	3.5	36.5
Bk.	I	1.5	4	48	0	1.5	1.5	43.5
	II	2	7	48	1.5	4.5	7	30
Jo.	I	1	4	42	1	2	2.5	47.5
	II	1.5	4	42.5	0.5	5	4	42.5
Br.	I	1.5	1.5	46.5	0	2.5	7	41
	II	1	9	55	2	0.5	1	31.5
Sm.	I	3	9	44	0	1.5	1	41.5
	II	3	13.5	56.5	2	4.5	2.5	18
Ge.	I	4.5	8	41.5	1.5	2	4	38.5
	II	1	7.5	45	0.5	4.5	5	36.5
Fr.	I	3.5	9	34.5	0	2	4.5	46.5
	II	4	11	44	0.5	3	2.5	35
Ja.	I	6	3	40	1.5	2.5	2	45
	II	6.5	8	48.5	1	3.5	2.5	30
Pl.	I	2	9.5	41.5	0.5	1.5	4	41
	II	1.5	19	41	0.5	2.5	7	28.5
Al.	I	8	10	36.5	1.5	3	5	36
	II	10.5	7.5	39	2.5	2.5	0.5	37.5
Sch.	I	4	4.5	37.5	1	3.5	4	47.5
	II	2.5	9.5	48	2	2.5	0	35.5

Tabelle 4.1: Blutwerte nach $\frac{3}{4}$ Stunden Gymnastik, 13 Studentinnen, Auszug aus [vL35]

zweier Variablen kann man sich vorstellen als Zusammenhang

$$X \text{ ist groß} \leftrightarrow Y \text{ ist groß}$$

oder

$$X \text{ ist klein} \leftrightarrow Y \text{ ist klein.}$$

Um aber gerade diese Korrelationen berechnen zu können, benötigt man ein mathematisches Hilfsmittel: die Korrelationsrechnung. Unter Zuhilfenahme der Ausgangsdatenmatrix in Tabelle 4.2 versucht man, die Korrelationen aller Variablen untereinander aufzuklären ([BEPW90]).

Name	Eosinophile	Stabkernige	Segmentkernige	Basophile	Monocyten	Übergangsformen	Lymphocyten
Bo.	-1.0	6	2.5	3	-2	-2.5	-7
Mü.	-4	-0.5	17.5	0	0	2.5	-15.5
Hu.	0	3.5	-1.5	-0.5	0	0	-1.5
Bk.	0.5	3	0	1.5	3	5.5	-13.5
Jo.	0.5	0	0.5	-0.5	3	1.5	-5
Br.	-0.5	7.5	8.5	2	-2	-6	-9.5
Sm.	0	4.5	12.5	2	3	1.5	-23.5
Ge.	-3.5	-0.5	3.5	-1	2.5	1	-2
Fr.	0.5	2	9.5	0.5	1	-2	-11.5
Ja.	0.5	5	8.5	-0.5	1	0.5	-15
Pl.	-0.5	9.5	-0.5	0	1	3	-12.5
Al.	2.5	-2.5	2.5	1	-0.5	-4.5	1.5
Sch.	-1.5	5	10.5	1	-1	4	-12.5

Tabelle 4.2: Ausgangsdatenmatrix

Um zwei Variablen X und Y vergleichen zu können, muß man sie standardisieren. Schließlich können die beiden Variablen Werte in vollkommen unterschiedlichen Größenordnungen aufweisen, z.B. dadurch, daß die Maßeinheiten ungeeignet gewählt wurden, oder weil es den Variablen inhärente Unterschiede in Mittelwerten und Varianzen gibt. Ebenfalls problematisch ist, daß die Variablen in verschiedenen nicht metrischen Skalenniveaus erhoben worden sein können, z.B. in Nominal-, Ordinal-, Intervall- und Verhältnisskalen (vgl. [KSV96]), auf die wir im Rahmen der Arbeit jedoch nicht einhegen wollen. Die Standardisierung zielt darauf ab, alle Variablen derart zu verändern, daß deren Mittelwerte 0 ergeben, und alle die Standardabweichung 1 vorweisen (vgl. [BEPW90]).

Um dies für eine Variable X zu erreichen werden die Werte jeder einzelnen Erhebung x_i von X modifiziert:

$$z_i = \frac{x_i - \bar{X}}{s_X} \quad (4.2)$$

mit z_i dem standardisierten Wert von x_i , \bar{X} dem Mittelwert aller Beobachtungen von X und s_X der Standardabweichung der selben. Diese Gleichung wenden wir auch Tabelle 4.2 an. Zunächst benötigen wir die Mittelwerte und Standardabweichungen aller Variablen:

$$\overline{\text{Eosinophile}} = \frac{-1 - 4 + 0.5 + 0.5 - 0.5 - 3.5 + 0.5 + 0.5 - 0.5 + 2.5 - 1.5}{13}$$

$$\begin{aligned}
 &= \frac{-6.5}{13} \\
 &= -0.5
 \end{aligned} \tag{4.3}$$

$$\begin{aligned}
 s_X &= \sqrt{\frac{1}{13} \cdot \sum_{i=1}^{13} (\text{Eosinophile}_i - \overline{\text{Eosinophile}})^2} \\
 &= \sqrt{\frac{1}{13} \cdot ((-0.5)^2 + (-3.5)^2 + 1^2 + 0^2 + (-3)^2 + 1^2 + 1^2 + 0^2 + 3^2 + (-1)^2)} \\
 &= 1.66
 \end{aligned} \tag{4.4}$$

Analog dazu berechnet man Mittelwert und Standardabweichung aller Variablen und erhält nach Gleichung 4.2 Tabelle 4.3.

Name	Eosinophile	Stabkernige	Segmentkernige	Basophile	Monocyten	Übergangsformen	Lymphocyten
Bo.	-0.083	0.226	-0.155	0.558	-0.43	-0.248	0.115
Mü.	-0.583	-0.312	0.575	-0.155	-0.11	0.187	-0.239
Hu.	0.083	0.019	0.35	-0.274	-0.11	-0.03	0.366
Bk.	0.167	-0.022	-0.277	0.201	0.368	0.448	-0.156
Jo.	0.167	-0.271	-0.253	-0.274	0.368	0.1	0.199
Br.	0	0.35	0.137	0.32	-0.43	-0.552	0.011
Sm.	0.083	0.102	0.331	0.32	0.368	0.1	-0.572
Ge.	-0.5	-0.312	-0.107	-0.393	0.288	0.057	0.324
Fr.	0.167	-0.105	0.185	-0.037	0.049	-0.204	-0.072
Ja.	0.167	0.143	0.137	-0.274	0.049	0.013	-0.218
Pl.	0	0.516	-0.301	-0.155	0.049	0.231	0.114
Al.	0.5	-0.478	0.155	0.082	-0.19	-0.421	0.47
Sch.	-0.167	0.143	0.234	0.082	-0.27	0.318	-0.114

Tabelle 4.3: standardisierte Ausgangsdatenmatrix

In dieser Tabelle, die als „standardisierte Ausgangsdatenmatrix“ bezeichnet wird, kann man manchmal, aber nicht immer, schon gewisse Beziehungen zwischen Variablen sehen (z. B. eine Variable ist immer dann klein, wenn eine bestimmte andere Variable auch klein ist und umgekehrt). Betrachtet man sich aber Tabelle 4.3, so ist offensichtlich, daß dies wohl nicht reichen kann, die Korrelation zwischen den Variablen herauszufinden. Man verwendet stattdessen nach [BEPW90] die Gleichung

$$R = \frac{1}{k-1} Z^T Z \tag{4.5}$$

mit Z der standardisierten Ausgangsdatenmatrix, Z^T der transponierten standardisierten Ausgangsdatenmatrix, und R der gesuchten Korrelationsmatrix. Die Korrelationsmatrix ist eine quadratische Matrix, in der für alle Variablenpaare (X_1, X_2) verzeichnet ist, wie stark sie korrelieren.

Sind in einer Ausgangsmatrix k Erhebungen (hier Studentinnen) $x_{1,i}$, $1 \leq i \leq k$ der Variable X_1 und ebenso k Erhebungen $x_{2,i}$, $1 \leq i \leq k$ der Variable X_2 notiert, so kann man die Werte $r(X_1, X_2)$ der Korrelationsmatrix, ebenfalls nach [BEPW90], mittels der Gleichung

$$r(X_1, X_2) = \frac{\sum_{i=1}^k (x_{1,i} - \bar{X}_1)(x_{2,i} - \bar{X}_2)}{\sqrt{\sum_{i=1}^k (x_{1,i} - \bar{X}_1)^2 \cdot \sum_{i=1}^k (x_{2,i} - \bar{X}_2)^2}} \quad (4.6)$$

berechnen. Wir wollen dies nun nicht am Beispiel demonstrieren und geben statt dessen nur die fertige Korrelationsmatrix R in Tabelle 4.4 an.

	Eosinophile	Stabkernige	Segmentkernige	Basophile	Monocyten	Übergangsformen	Lymphocyten
Eosinophile	1	0.024	-0.426	0.208	0.066	-0.315	0.163
Stabkernige		1	-0.080	0.363	-0.296	0.036	-0.421
Segmentkernige			1	0.136	-0.155	-0.007	-0.628
Basophile				1	-0.413	-0.304	-0.282
Monocyten					1	0.578	-0.248
Übergangsformen						1	-0.428
Lymphocyten							1

Tabelle 4.4: Korrelationsmatrix

Die Hauptdiagonale dieser Matrix enthält zwingend ausnahmslos Einsen. Dies ist zwingend, weil jede Variable ja mit sich selbst vollständig korreliert². Die Korrelationsmatrix ist nur oberhalb der Hauptdiagonalen gefüllt. Zwar ist sie mathematisch auch unterhalb der Hauptdiagonalen nicht Werte-frei, aber die Werte entsprechen denen über der Hauptdiagonale (X_1 korreliert mit X_2 in gleichem Maße wie X_2 mit X_1).

Meist wird es empfohlen, die Ausgangsdatenmatrix zu standardisieren, weil somit erstens eine Vergleichbarkeit von Variablen ermöglicht wird und zweitens die Korrelationsrechnung erleichtert wird ([BEPW90]). Wir empfinden dies als unnötigen Umweg, da „interessante“ Korrelationsrechnungen aus der Realwelt (mit entsprechend vielen Variablen) sowieso nicht mehr per

²eine Korrelation von 1 bedeutet zwar nicht immer, daß die Variablen identisch sind, aber, daß sich ihrer Werte proportional verhalten

Hand sondern mittels Rechnerunterstützung berechnet werden und somit schnell ein Ergebnis vorliegt. Die standardisierte Ausgangsmatrix ist in unseren Augen nur dann interessant, wenn man schnell einen Überblick über das Verhalten von Variablen erhalten möchte und die Rechnung „per Hand“ durchgeführt wird.

Anhand der Korrelationsmatrix können die Zusammenhänge zwischen Variablen begutachtet werden. Die Korrelation zwischen zwei Variablen ist immer eine Zahl x , $-1 \leq x \leq 1$. Ist die Korrelation nahe an 1, so liegt eine hohe positive Korrelation vor. Hohe positive Korrelationen zwischen zwei Variablen X_1 und X_2 sind ein Ausdruck dafür, daß, wenn X_1 einen hohen Wert hat, auch X_2 einen hohen Wert hat, zumindest im Schnitt. Eine hohe negative Korrelation hingegen bedeutet für hohe Werte von X_1 besonders niedrige Werte für X_2 . Ein Korrelationskoeffizient von 0 hingegen sagt aus, daß die Werte von X_1 und X_2 sich nicht zu beeinflussen scheinen oder anders gesagt kein Zusammenhang zwischen den Variablen existiert ([vL98]).

Besteht zwischen zwei Variablen eine betragsmäßig hohe Korrelation, so liegt es nahe, diese zu einem latenten Faktor, der zunächst noch nicht interpretiert sei, zusammenzufassen.

Durch die Zusammenfassung zweier oder mehrerer Variablen gehen Informationen (z.B. über die genauen Werte, den Mittelwert, die Varianz etc. der Variablen) verloren. Mit dem Informationsverlust einher geht aber auch die Größe der Datenmenge, die es zu verarbeiten gilt. Ziel der Faktorenanalyse ist es nun, die Größe der Datenmenge zu reduzieren, und gleichzeitig den einhergehenden Informationsverlust gering zu halten (vgl. hierzu u.a. [BEPW90] und [FH84]).

Wir werden noch sehen, daß es mit Angaben zur Reduktionsstärke und zum zu erlaubenden Informationsverlust gerade keine Aussagen zu machen gilt. Wie stark die Information verdichtet werden soll und wie viel davon dabei verloren gehen darf liegt allein in der Hand dessen, der die Faktorenanalyse durchführt.

Dies ist auch ein wichtiger Kritikpunkt an der Faktorenanalyse: Die hochgradige Indeterminiertheit der Ergebnisse durch Eingriffe des Anwenders.

4.3 Fundamentaltheorem

Die Idee der Faktorenanalyse haben wir weiter oben bereits kennengelernt. Es stellt sich aber die Frage, wie man die Faktoren rein rechnerisch aus den Variablen (konkret aus der Korrelationsmatrix) ermitteln kann. Wir gehen davon aus, daß jede Variable bzw. jeder Wert derselben aus einer Kombina-

tion von Faktoren³ dargestellt werden kann. Man ist in der Faktorenanalyse sogar noch strenger und verlangt, daß es sich bei angesprochener Kombination um eine Linearkombination handelt, die einen Wert der Variable in der Z -Matrix (d.h. den standardisierten Wert) darstellt (vgl. [Moo78], [BEPW90]):

$$\begin{aligned} z_{i,k} &= a_{1,k}f_{i,1} + a_{2,k}f_{i,2} + \dots + a_{j,k}f_{i,j} + \dots + a_{r,k}f_{i,r} \\ &= \sum_{j=1}^r a_{j,k}f_{i,j} \end{aligned} \quad (4.7)$$

mit $z_{i,j}$ dem Wert der i -ten Beobachtung der k -ten Variable, $f_{i,j}$ dem Wert der i -ten Beobachtung des j -ten Faktors und $a_{j,k}$ der Gewichtung der k -ten Variable durch den j -ten Faktor. Dies können wir auch in Matrix-Schreibweise darstellen:

$$Z = FA^T \quad (4.8)$$

mit Z der standardisierten Ausgangsmatrix, A^T der transponierten Matrix der Gewichtungen (Faktorladungsmatrix) und F der Matrix der Faktorwerte (Faktorwertematrix) (s. Abbildung 4.1).

Das verwenden eine Linearkombination ist eine Vorweg-Vereinfachung. Es handelt sich hier um einen kritischen Punkt der Faktorenanalyse, denn man kann anbringen, daß die Unterstellung eines linearen Zusammenhangs nicht immer haltbar ist und evtl. auf andere Zusammenhänge (z.B. quadratische etc.) zurückgegriffen werden müßte. Die Art des Zusammenhangs ist jedoch, da zu diesem Zeitpunkt weder die Faktoren selbst noch deren Interpretation bekannt sind. Andere Arten von Zusammenhängen werden in der Literatur jedoch nicht angewendet.

Es folgt aus Gleichung 4.8 und Gleichung 4.5 durch symbolische Manipulation:

$$\begin{aligned} R &\stackrel{\text{Gl. 4.5}}{=} \frac{1}{K-1} Z^T Z \\ &\stackrel{\text{Ersetzung nach Gl. 4.8}}{=} \frac{1}{K-1} (FA^T)^T (FA^T) \\ &\stackrel{(XY)^T = Y^T X^T \text{ und } (X^T)^T = X}{=} \frac{1}{K-1} (AF^T)(FA^T) \\ &= \frac{1}{K-1} A(F^T F)A^T \end{aligned}$$

³ die noch nicht vorliegen

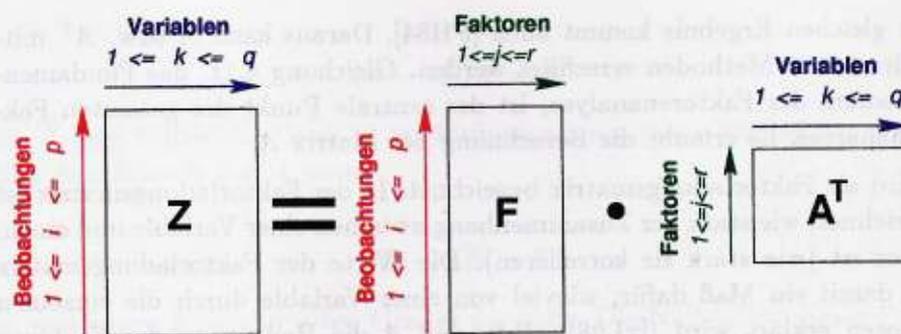


Abbildung 4.1: Hinführung zum Fundamentaltheorem

$$\begin{aligned}
 &= \\
 \text{s. Gl. 4.10} \quad &= \frac{1}{K-1} (F^T F) A^T \\
 &= A C A^T \quad (4.9)
 \end{aligned}$$

mit

$$C = \frac{1}{K-1} F^T F \quad (4.10)$$

der Korrelationsmatrix der Faktoren.

Man wünscht sich, daß die Faktoren untereinander nicht korrelieren ([CTC89]). Eine Korrelation zwischen zwei Variablen oder zwei Faktoren bedeutet, daß eine Variable (Faktor) einen Teil der anderen Variable (des anderen Faktors) erklärt. Man versucht aber in der Faktorenanalyse gerade die Variablen, die korrelieren und sich gegenseitig erklären bzw. voneinander abhängen, zusammenzufassen und damit „mehrfach vorhandene“ Information herauszufiltern.

Unter der Annahme und dem Wunschenken, daß die Faktoren untereinander unabhängig sind, d.h. nicht korrelieren, enthält die Korrelationsmatrix nur Nullen⁴. Daraus folgt wiederum, daß die Korrelationsmatrix C der Faktoren der Einheitsmatrix⁵ E entspricht.

$$\begin{aligned}
 R &\stackrel{\text{Gl. 4.9}}{=} A C A^T \\
 &\stackrel{C \stackrel{!}{=} E}{=} A E A^T \\
 \text{Distributivität der Matrizenmultiplikation} &\stackrel{=} {=} (A E) A^T \\
 A E &\stackrel{=} {=} A \quad (4.11) \\
 &\stackrel{=} {=} A A^T
 \end{aligned}$$

⁴außer natürlich auf der Hauptdiagonalen, da die Korrelation mit sich selbst immer eins ist

⁵Einheitsmatrix: eine quadratische Matrix, die nur 1 auf der Hauptdiagonale und nur 0 über und unter der Hauptdiagonale aufweist

Zum gleichen Ergebnis kommt auch [FH84]. Daraus kann A bzw. A^T mittels iterativer Methoden errechnet werden. Gleichung 4.11, das Fundamentaltheorem der Faktorenanalyse, ist der zentrale Punkt der gesamten Faktorenanalyse. Es erlaubt die Berechnung der Matrix A .

A wird als Faktorladungsmatrix bezeichnet. In der Faktorladungsmatrix ist verzeichnet, wie stark der Zusammenhang zwischen einer Variable und einem Faktor ist (wie stark sie korrelieren). Die Werte der Faktorladungsmatrix sind damit ein Maß dafür, wieviel von einer Variable durch die einzelnen Faktoren erklärt wird ([vL98]). D.h. daß A die Bedeutung der Faktoren beschreibt.

Die einzelnen Werte $A_{i,j}$ mit einer Variablen X_i und einem Faktor F_j werden Faktorladung oder manchmal auch Gewichtszahl genannt.

4.4 Hauptkomponentenanalyse

Ziel der Faktorenanalyse ist nun die Auffindung von möglichst unabhängigen Faktoren aus der Korrelationsmatrix der Variablen sowie die Erklärung der Variablen durch die Faktoren. Nach dem Fundamentaltheorem läßt sich die Korrelationsmatrix aus dem Produkt der Faktorladungsmatrizen darstellen. Dazu benötigt man das Konzept der Eigenvektoren und Eigenwerte (empfohlen hierzu sei [FH84] und [LB82]).

Aus dem Fundamentaltheorem der Faktorenanalyse läßt sich⁶ nach der Gleichung

$$A = T\Lambda^{\frac{1}{2}} \quad (4.12)$$

die Faktorladungsmatrix berechnen.

Dabei ist

$$T = (\vec{t}_1, \vec{t}_2, \dots, \vec{t}_k, \dots, \vec{t}_q) \quad (4.13)$$

die Matrix der Eigenvektoren der Korrelationsmatrix R . Λ entspricht der Matrix der Eigenwerte der Korrelationsmatrix R .

Nun ist R eine quadratische $q \times q$ -Matrix. Gibt es einen Vektor \vec{x} und eine Zahl λ , für die die Gleichung

$$R\vec{x} = \lambda\vec{x} \quad (4.14)$$

gilt, so heißt λ Eigenwert und \vec{x} zugehöriger Eigenvektor von R . Durch Umformung der Gleichung erhält man

⁶für einen Beweis befasse sich der interessierte Leser z.B. mit [FH84]

$$(R - \lambda E)\vec{x} = \vec{0} \quad (4.15)$$

mit $\vec{0}$ dem Nullvektor⁷ und E der Einheitsmatrix. Diese Formel hat genau dann eine Lösung $x \neq 0$, wenn

$$|R - \lambda E| = 0 \quad (4.16)$$

ist, was einem Polynom q -ten Grades darstellt. Die Nullstellen dieses Polynoms heißen Eigenwerte von R .

Die Eigenwerte unseres Beispiels lauten

$$\lambda_1 = 1.295 \quad (4.17)$$

$$\lambda_2 = 0.876 \quad (4.18)$$

$$\lambda_3 = 0.012 \quad (4.19)$$

$$\lambda_4 = 2.117 \quad (4.20)$$

$$\lambda_5 = 0.241 \quad (4.21)$$

$$\lambda_6 = 0.446 \quad (4.22)$$

$$\lambda_7 = 2.013 \quad (4.23)$$

Daraus ergibt sich die Matrix

$$\Lambda^{1/2} = \begin{pmatrix} \sqrt{1.295} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{0.876} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{0.012} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{2.117} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{0.241} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{0.446} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{2.013} \end{pmatrix} \quad (4.24)$$

Den Eigenvektor \vec{x}_i eines Eigenwertes λ_i berechnet man aus der Gleichung

$$(R - \lambda_i E)\vec{x}_i = \vec{0} \quad (4.25)$$

Ein Vektor, für den diese Gleichung gilt, heißt korrespondierenden Eigenvektor \vec{x}_i zu λ_i .

Wir schreiben die Eigenvektoren als Spaltenvektoren und erhalten nach Gleichung 4.13 die Matrix

⁷Vektor, der nur aus Nullen besteht

$$T = \begin{pmatrix} -0.528 & -0.531 & -0.119 & 0.368 & 0.426 & 0.317 & -0.091 \\ -0.466 & 0.543 & -0.352 & 0.006 & -0.186 & 0.369 & 0.44 \\ 0.474 & -0.377 & -0.529 & -0.329 & 0.149 & 0.255 & 0.397 \\ -0.211 & -0.244 & -0.143 & 0.209 & -0.127 & -0.736 & 0.524 \\ -0.329 & -0.328 & -0.298 & -0.407 & -0.606 & -0.036 & -0.402 \\ -0.27 & 0.238 & -0.186 & -0.556 & 0.615 & -0.348 & -0.168 \\ 0.233 & 0.242 & -0.662 & 0.487 & 0.031 & -0.184 & -0.42 \end{pmatrix} \quad (4.26)$$

Nach der Gleichung 4.12 erhalten wir eine erste Faktorladungsmatrix:

$$A = \begin{pmatrix} -0.601 & -0.497 & -0.013 & 0.536 & 0.209 & 0.212 & -0.129 \\ -0.53 & 0.508 & -0.038 & 0.008 & -0.091 & 0.247 & 0.625 \\ 0.54 & -0.353 & -0.057 & -0.479 & 0.073 & 0.171 & 0.563 \\ -0.24 & -0.228 & -0.015 & 0.304 & -0.062 & -0.492 & 0.743 \\ -0.37 & -0.307 & -0.032 & -0.592 & -0.297 & -0.024 & -0.57 \\ -0.307 & 0.222 & -0.02 & -0.809 & 0.302 & -0.233 & -0.238 \\ 0.265 & 0.226 & -0.071 & 0.709 & 0.015 & -0.123 & -0.596 \end{pmatrix} \quad (4.27)$$

A ist eine Faktorladungsmatrix für sieben Variablen und sieben Faktoren. Die Faktoren können jedoch nicht einfach den Variablen zugeordnet werden (1 Variable - 1 Faktor). Wäre dies der Fall, so würde A der Korrelationsmatrix R entsprechen. An diesem Punkt angelangt hat man die Faktorenanalyse noch nicht beendet. Schließlich ist es nicht wünschenswert, aus q Variablen auch q Faktoren zu machen. Der aktuelle Stand ist nur ein Sprungbrett.

Es ist an der Zeit, die Faktorladungsmatrix neu zu berechnen, um eine kleinere Faktorladungsmatrix mit weniger Faktoren zu erhalten. Die Idee ist, daß man mit einer geringeren Zahl von Faktoren ebenfalls einen vielleicht sogar recht hohen Anteil an Varianz der Ausgangsvariablen erklären kann. Man sagt, daß ein Faktor oder eine Menge von Faktoren eine Variable oder eine Menge von Variablen erklärt, wenn er oder sie die Varianz der Ausgangsvariable(-n) erklärt. Gesucht ist nun eine Anzahl r von Faktoren aus den q vorhandenen, die möglichst viel Varianz erklären. Um diese Faktoren herauszufinden kann man mehrere Methoden verwenden. Drei von diesen möchten wir hier vorstellen und verwenden dazu [BEPW90], [LB82] und [FH84]. Im Anschluß wird ein weiteres einfaches aber fundamentales Kriterium angesprochen.

4.5 Bestimmung der Anzahl der Faktoren

4.5.1 Eigenwert-Kriterium

Jeder Faktor besitzt einen Eigenwert. Dieser Eigenwert kann der Faktorladungsmatrix entnommen werden. Er berechnet sich aus der Summe der Quadrate einer Spalte (s. Abbildung 4.2):

$$\text{eigen}(f_j) = \sum_{k=1}^q A_{jk}^2 \quad (4.28)$$

	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	Faktor 7	Kommunität
Eosinophile	-0.601	-0.497	-0.013	0.536	0.209	0.212	-0.129	1.0
Stabkernige	-0.53	0.508	-0.038	0.008	-0.091	0.247	0.625	1.0
Segmentkernige	0.54	-0.353	-0.057	-0.479	0.073	0.171	0.563	1.0
Basophile	-0.24	-0.228	-0.015	0.304	-0.062	-0.492	0.743	1.0
Monocyten	-0.37	-0.307	-0.032	-0.592	-0.297	-0.024	-0.57	1.0
Übergangsformen	-0.307	0.222	-0.02	-0.809	0.302	-0.233	-0.238	1.0
Lymphocyten	0.265	0.226	-0.071	0.709	0.015	-0.123	-0.596	1.0
Eigenwert	1.293	0.876	0.012	2.117	0.241	0.447	2.013	7.0

Abbildung 4.2: vorläufige Faktorladungsmatrix

Der Eigenwert $\text{eigen}(f_j)$ eines Faktors f_j stellt die durch den Faktor f_j erklärte Varianz an allen Variablen zusammen dar. Im Durchschnitt wird von allen Variablen die Gesamtvarianz der Daten erklärt. Da die Daten auf eine Varianz von 1 pro Variable standardisiert sind, ergibt sich in unserem Fall eine Gesamtvarianz der Ausgangsdaten von 7. Da 7 Variablen zur Erklärung der Gesamtvarianz zur Verfügung stehen, erklärt jede Variable durchschnittlich eine Varianz von 1.

Nach dem Eigenwertkriterium, daß nach seinem Verfasser (s. [KD59]) auch Kaiserkriterium genannt wird, werden nur solche Faktoren als wesentlich erachtet, die einen Eigenwert größer als 1 haben. Ein Faktor mit einem Eigenwert kleiner als eins erklärt weniger von der Gesamtvarianz als eine Variable selbst und steht damit dem Sinn der Faktorenanalyse, Variablen durch eine geringe Anzahl von Faktoren zu erklären, entgegen.

Nach Abbildung 4.3 erhält man aufgrund des Eigenwertkriterium (die Faktoren sind nach ihren Eigenwerten sortiert) eine Anzahl von 3 wesentlichen Faktoren: Faktor 4, Faktor 7 und Faktor 1.

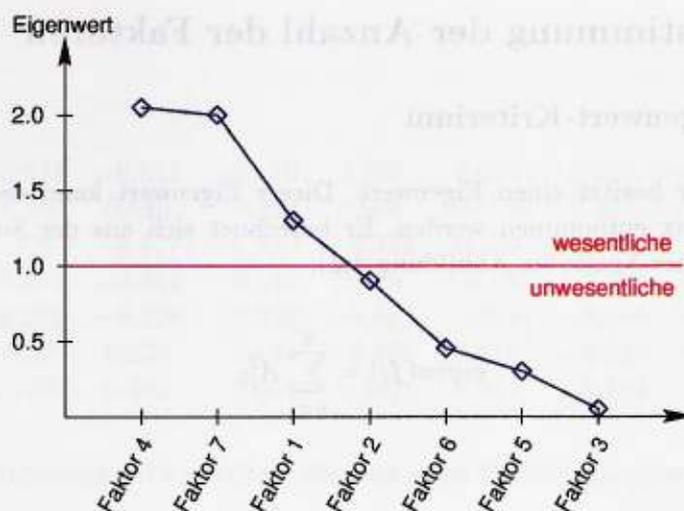


Abbildung 4.3: Kaiserkriterium

4.5.2 Scree-Test

Der Scree-Test ist ein „optisches“ Verfahren. Empirische Korrelationsmatrizen (die Faktorladungsmatrix ist eine Korrelationsmatrix), d.h. solche die aus reinen unabhängige Zufallszahlen entstanden sind, zeigen typischerweise immer ähnliche Eigenwerte, oder bildlich gesprochen einen flachen Verlauf der Kurve der Eigenwerte ([FH84]). Daher empfindet man die Faktoren als wesentlich, für die die Eigenwertkurve steil verläuft, andere erachtet man als unwesentlich. Der Scree-Test ist, wie unser Beispiel zeigt in Abbildung 4.4 zeigt, recht schwierig durchzuführen und von der persönlichen Semantik von „Steilheit einer Kurve“ abhängig. Er ist unserer Ansicht daher nur bedingt geeignet.

Nach dem Scree-Test und unserer Interpretation der Kurve erhält man als wesentliche Faktoren f_4, f_7, f_1, f_2 und f_6 .

4.5.3 kumulierte erklärte Varianz

Nach diesem Kriterium aus [FH84] erklärt man so viele Faktoren zu wesentlichen Faktoren, die notwendig sind, um einen gewissen Anteil der Gesamtvarianz zu erklären. Auch hier benötigt man die Faktoren nach Eigenwerten sortiert. Der Erklärungsanteil eines Faktors f_j an der Gesamtvarianz ergibt sich aus dem Quotienten

$$\frac{\text{eigen}(f_j)}{p} \quad (4.29)$$

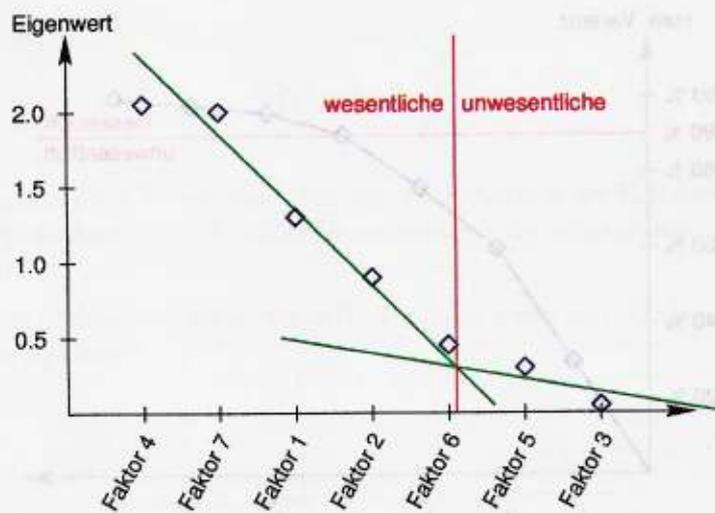


Abbildung 4.4: Scree test

mit p der Anzahl der Variablen, hier 7, und $eigen(f_j)$ dem Eigenwert von f_j . Faktor 4 erklärt somit 30.2 % der Gesamtvarianz, Faktor 7 erklärt 27.8 % etc. Diese Werte werden kumuliert solange, bis ein bestimmter Varianzanteil erreicht ist, den man von den Faktoren insgesamt erklärt sehen möchte.

In unserem Beispiel sollen 90% der Gesamtvarianz erklärt werden. Nach Abbildung 4.5 und Gleichung ergibt sich für uns eine Anzahl von 4 wesentlichen Faktoren (Faktor 7, Faktor 4, Faktor 1 und Faktor 2).

4.5.4 Gewünschte Datenreduktion

Da das Ziel der Faktorenanalyse neben der Aufdeckung von Zusammenhängen zwischen Variablen eine Datenreduktion ist, kann man als Kriterium auch direkt eine Anzahl von Faktoren auswählen. Wünscht man eine Datenreduktion um mindestens 50%, dann würde man in unserem Beispiel 3 Faktoren wählen. Die oben genannten anderen Verfahren sind dann ein Maß für die Sinnhaftigkeit dieses Wunsches. Nach dem Kaiserkriterium wäre der Wunsch auf 3 Faktoren zu reduzieren noch sinnvoll. Die erklärte kumulierte Varianz von ca. 78% durch 3 Faktoren ist zwar nicht sonderlich hoch, soll aber unseren Ansprüchen hier genügen (der Wunsch nach Datenreduktion ist höher als der Wunsch nach hoher Varianzerklärung). Die Wahl der Anzahl der Faktoren ist trotz aller Auswahlverfahren immer noch eine Aufgabe für die persönliche Einschätzung.

Dies ist ein Streitpunkt der Faktorenanalyse. Von einem mathematischen Verfahren erwartet man mehr Determiniertheit und weniger persönliche

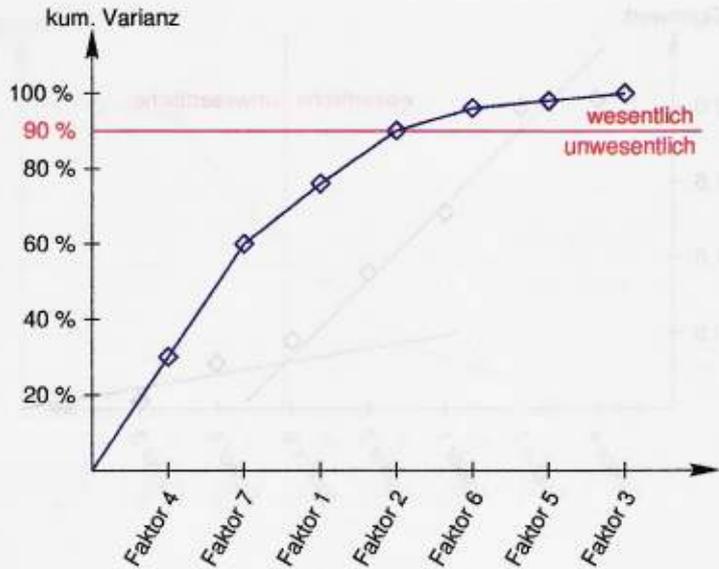


Abbildung 4.5: kumulierte Varianz

Eingriffe von außen.

4.6 reduzierte Faktorladungsmatrix

Sei also die Anzahl der Faktoren $r = 3$. Die einzelnen Faktoren nummerieren wir neu in Reihenfolge ihrer Eigenwerte und damit Wichtigkeit:

1. Faktor 4 wird nun zu Faktor I (römisch 1)
2. Faktor 7 wird nun zu Faktor II (römisch 2)
3. Faktor 1 wird nun zu Faktor III (römisch 3)

Die reduzierte Faktorladungsmatrix ist eine Faktorladungsmatrix wie die bisherige auch. Sie ist zwar nicht quadratisch, dennoch entspricht sie wieder einer Matrix von Korrelationen zwischen Faktoren und Variablen. Die neue Matrix entsteht durch eine einfache Streichung der unwesentlichen Faktoren. Dadurch wird die bisherige Matrix von 7×7 auf ein Format von 7×3 reduziert.

Die Einträge $a_{j,k}^{\text{neu}}$ der neuen Faktorladungsmatrix A^{neu} mit j der Nummer der Variable k der Nummer des Faktors ergibt sich in Anlehnung an [LB82] aus der Gleichung

$$a_{j,k}^{\text{neu}} = t_{j,k} \cdot \sqrt{\frac{\lambda_k}{s_j^2}} \quad (4.30)$$

wobei $t_{j,k}$ den j -ten Werte des k -ten Eigenvektors aus der Eigenvektormatrix T (s.o.) bezeichnet. Die Standardabweichungen s_j entstammen den Zeilen der Matrix A .

In unserem Beispiel rechnen wir z.B. wie folgt dann mit λ_4 dem Eigenwert des neuen Faktors I

$$\begin{aligned} a_{\text{Eosinophile, Faktor I}} &= t_{1,4} \cdot \sqrt{\frac{\lambda_4}{s_{\text{Eosinophile}}^2}} \\ &= 0.368 \cdot \sqrt{\frac{2.117}{1^2}} \\ &= 0.368 \cdot \sqrt{\frac{2.117}{1}} \\ &= 0.368 \cdot \sqrt{2.117} \\ &= 0.368 \cdot 1.455 \\ &= 0.535 \end{aligned} \quad (4.31)$$

für die Korrelation zwischen dem erwähnten (aufgrund des hohen Eigenwertes) Faktor I und der Variable „Eosinophile“. Man beachte $t_{1,4}$. In T entspricht die 4. Spalte dem Eigenvektor des Faktor 4, der (s.o.) nun unser Faktor I sein soll. Weiterhin können wir für die Standardabweichung s_1 sowie für alle anderen Standardabweichungen weiterer Variablen 1 einsetzen, da wir mit einer bereits standardisierten Ausgangsmatrix gearbeitet haben. Für alle anderen Korrelationen geht man analog vor.

Man erhält als Ergebnis die Korrelationsmatrix (Faktorladungsmatrix)

$$A^{\text{neu}} = \begin{pmatrix} 0.535 & -0.129 & -0.601 \\ 0.009 & 0.624 & -0.53 \\ -0.479 & 0.563 & 0.539 \\ 0.304 & 0.743 & -0.24 \\ -0.592 & -0.57 & -0.374 \\ -0.809 & -0.238 & -0.307 \\ 0.709 & -0.596 & 0.265 \end{pmatrix} \quad (4.32)$$

Abbildung 4.6 zeigt diese Matrix noch einmal in Tabellenform mit annotierten Eigenwerten und neuen Kommunalitäten.

	Faktor I	Faktor II	Faktor III	Kommunalität
Eosinophile	0.535	-0.129	-0.601	0.664
Stabkernige	0.009	0.624	-0.530	0.670
Segmentkernige	-0.479	0.563	0.539	0.837
Basophile	0.304	0.743	-0.240	0.702
Monocyten	-0.592	-0.57	-0.374	0.815
Übergangsformen	-0.809	-0.238	-0.307	0.805
Lymphocyten	0.709	-0.596	0.265	0.928
Eigenwert	2.117	2.011	1.295	5.422 / 7.0

Abbildung 4.6: reduzierte Faktorladungsmatrix mit 3 Faktoren

4.7 Faktorinterpretation

Abbildung 4.6 zeigt die Zusammenfassung von Variablen zu Faktoren. Eine Variable wird stets dem Faktor zugeordnet, der durch die Variable betragsmäßig am höchsten geladen wird. Es ergeben sich daraus die Zusammenfassungen

$$\text{Faktor I} = \{\text{Monocyten}, \text{Lymphocyten}, \text{Übergangsformen}\} \quad (4.33)$$

$$\text{Faktor II} = \{\text{Stabkernige}, \text{Segmentkernige}, \text{Basophile}\} \quad (4.34)$$

$$\text{Faktor III} = \{\text{Eosinophile}\} \quad (4.35)$$

Die Faktorladungen besagen, daß Faktor I „viel zu tun hat“ mit Monocyten, Lymphocyten und Übergangsformen, Faktor II mit Stabkernigen, Segmentkernigen und Basophilen, und Faktor III mit Eosinophilen. Die Kommunalitäten, die die durch aller Faktoren zusammen dargebotenen erklärte Varianz einer Variable angibt⁸, sind zumindest größer als 0.5 und damit „gar nicht einmal so schlecht“. Die Eigenwerte der Faktoren⁹ bleiben, und das ist mathematisch nicht anders möglich, gleich den Eigenwerten der Faktoren bei der Extraktion von 10 Faktoren. Die Summe der Kommunalitäten ist gleich der Summe der Eigenwerte und gibt die erklärte Varianz aller Variablen durch alle Faktoren an. Sie ist in unserem Beispiel 5.422 (s. Abbildung 4.6, grün coloriert). Der erklärte Gesamtvarianzanteil ergibt sich zu

$$\frac{5.422}{7.0} \approx 77.5\% \quad (4.36)$$

⁸berechnet sich aus der Summe der Quadrate der jeweiligen Zeile

⁹Summe der Quadrate der entsprechenden Spalte

mit 7.0 der Varianz aller Variablen (7 standardisierte Variablen = 7.0 Gesamtvarianz) und dies korrespondiert mit den im Kriterium „kumulierte erklärte Varianz“ gewünschten und erwarteten Ergebnissen.

Die letzte Interpretation der Faktoren (z.B. die Beschreibung der selben durch einen passenden Namen) steht außerhalb der statistisch-mathematischen Möglichkeiten und bleibt daher den Fachexperten (in diesem Fall für das weiße Blutbild) überlassen.

Wir können aber immerhin soviel sagen, daß durch die Faktorenanalyse mit Ausnahme der Eosinophilen alle anderen Granulocytentypen (Stabkernige, Basophile, Segmentkernige) als zusammengehörig in Faktor II erkannt wurden.

mit 7.0 der Varianz aller Variablen (7 standardisierte Variablen = 7.0 Ge-
 samtheitsvarianz) und dies korrespondiert mit den im Kriterium „kumulierte
 erklärte Varianz“ gewinnte hier auch erwarteter Ergebnisse.
 Die letzte Information der Faktoren (z.B. die Beschreibung der beiden
 durch einen kumulierten Faktor) stellt sich als das Ergebnis der statistischen
 Möglichkeiten und nicht über die Faktoren im ersten Fall für die
 weitere Herleitung ablesen.
 Wir können aber feststellen, dass durch die Faktorenanalyse mit
 Ausnahme der Beobachtungen alle anderen (ökonomischen) Informationen
 (Bausätze, Regenerationsrate als zusammenfassend in Faktor II erfasst wird
 den.



Ergebnis der Faktorenanalyse (rot markiert) und die Faktoren

4.1 Faktorenanalyse

Die Faktorenanalyse ist eine statistische Methode zur Identifizierung von
 gemeinsamen Faktoren, die die beobachteten Variablen erklären können.
 Sie wird häufig verwendet, um die Dimensionalität von Datensätzen zu
 reduzieren und die zugrunde liegenden Strukturen zu verstehen.

Die Faktorenanalyse ist eine statistische Methode zur Identifizierung von

gemeinsamen Faktoren, die die beobachteten Variablen erklären können.

Sie wird häufig verwendet, um die Dimensionalität von Datensätzen zu

reduzieren und die zugrunde liegenden Strukturen zu verstehen. Die
 Faktorenanalyse ist eine statistische Methode zur Identifizierung von
 gemeinsamen Faktoren, die die beobachteten Variablen erklären können.
 Sie wird häufig verwendet, um die Dimensionalität von Datensätzen zu
 reduzieren und die zugrunde liegenden Strukturen zu verstehen.

Ergebnis der Faktorenanalyse

(rot markiert)

Die Faktorenanalyse ist eine statistische Methode zur Identifizierung von
 gemeinsamen Faktoren, die die beobachteten Variablen erklären können.

Kapitel 5

Clusteranalyse

Die Clusteranalyse gehört wie die Faktorenanalyse zu den struktur-entdeckenden Verfahren. Während die Faktorenanalyse, welche im vorigen Kapitel vorgestellt wurde, eine Zusammenfassung einzelner Variablen zur Aufgabe hat, wird in der Clusteranalyse eine Zusammenfassung und Einteilung von beobachteten Objekten aufgrund ihrer Merkmale (Ausprägungen von Zufallsvariablen) vorgenommen. Dadurch erzielt man eine Komprimierung/Verdichtung von Objekten, die bezüglich ihrer Beschreibungsmerkmale bzw. Zufallsvariablenwerte ähnlich sind ([BEPW90], [FH84], [CTC89]).

Bei diesem Analyseverfahren werden Gruppen, sogenannte Cluster gebildet, wobei die Objekte einer Gruppe möglichst homogen¹, die Gruppen untereinander jedoch möglichst heterogen² sollen ([CTC89]).

5.1 Ausgangsdaten

Die Ausgangsmenge der Untersuchungsobjekte wird zunächst als heterogene Grundgesamtheit angenommen. Ziel ist es, homogene Gruppen innerhalb der Grundgesamtheit zu ermitteln (s. Abbildung 5.1, soll nicht an einen Teller mit Spinat, Lasagne, Ei und Frikadelle erinnern!)

Wenn versucht wird, Variablen (Merkmale) oder Objekte (Merkmalsträger) aufgrund ihrer Ähnlichkeit oder Unterschiedlichkeit (Distanz) in einen Cluster zusammenzuführen, dann wendet man der Verfahren der Clusteranalyse an.

Die Clusteranalyse findet man in vielen Anwendungsbereichen. Um nur einige Beispiele zu nennen:

¹ gleichartig

² unterschiedlich, verschiedenartig

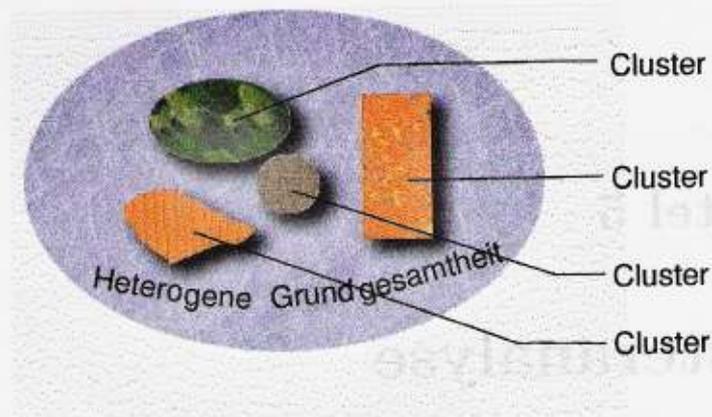


Abbildung 5.1: Prinzip der Clusteranalyse

- Archäologie
- Medizin
- Biologie
- Sozialwissenschaften
- Betriebswirtschaft

[FH84] nennt weitere Anwendungsgebiete.

Wir möchten die Clusteranalyse anhand eines medizinischen Beispiels (die Daten dazu entstammen wiederum [vL35]) demonstrieren.

Bei der Clusteranalyse wird immer von einer Matrix mit m Eigenschaften und n Objekten ausgegangen. In unserem Beispiel sieht diese wie in Tabelle 5.1 dargestellt aus.

Wie bereits bei der Faktorenanalyse subtrahieren wir jeweils die zweite von der ersten Zeile einer jeden Versuchsdurchführung (also jeder Person) und erhalten die Blutwertänderungen in Tabelle 5.2 sowie nach Standardisierung Tabelle 5.3

Wir erhalten so unsere Ausgangsmatrix, mit der wir im weiteren Verlauf arbeiten werden. Die Durchführung der Clusteranalyse besteht aus zwei wesentlichen Schritten:

- Wahl und Berechnung des Proximitätsmaßes
- Wahl und Ausführung eines Fusionierungsalgorithmus

Name	Entnahme	Eosinophile	Stabkernige	Segmentkernige	Basophile	Monocyten	Übergangsformen	Lymphocyten
Bu.	I	4	4	43.5	0	5	2.5	41
	II	1	4	30.5	1	5	2	56.5
Lo.	I	4.5	3	44	2.5	2	4.5	39.5
	II	2	5.5	44	1.5	4	2	41
Sch.	I	5.5	6	50.5	0	1.5	3	33.5
	II	2	6.5	44.5	1	2	0	44
We.	I	5.5	9	40	0.5	3	1.5	40.5
	II	1	2.5	24.5	2	6.5	4	59.5
Be.	I	6	5.5	45	0	2	2	39.5
	II	3	2	40.5	2	4.5	1	47
Ha.	I	2.5	7.5	45	1	5	3.5	35.5
	II	1	6	34	2	5	1	51
Ga.	I	2	3.5	41.5	1.5	5	6.5	40
	II	1	0	40.5	1	5	3	49.5
Gü.	I	3	3	42.5	0	2	7	42.5
	II	2	9	26	2	3.5	5	52.5
Go.	I	3.5	8	22.5	0.5	4.5	2	59
	II	1	4	33	1	7	0	54
La.	I	4	5.5	41.5	1.5	1	3	43.5
	II	0.5	7.5	38.5	2	2	1	48.5

Tabelle 5.1: Blutwerte: 10 Studenten, 800m-Lauf, Auszug aus [vL35]

5.2 Proximitätsmaße

Dieser erste Schritt der Clusteranalyse dient dazu, ein Maß für die Ähnlichkeit oder Proximität bzw. Unterschiedlichkeit oder Distanz der zu untersuchenden Objekte zu finden ([BEPW90],[FH84]).

Dazu stehen mehrere Maße für verschiedene Skalenniveaus, in denen die Daten erhoben worden sein können, zu Auswahl. Die objektbezogenen (in unserem Beispiel auf Personen bezogenen) Variablenwerte in der Matrix (s. Tabelle 5.3) sind metrisch (d.h. sie bestehen aus Zahlenwerten) oder nicht metrisch (z.B. wenn die Zufallsvariable Werte „ja“ und „nein“ annehmen kann (vgl. [KSV96], [FH84], [BEPW90]).

Alle metrischen Distanzmaße basieren auf der Gleichung

$$d_{i_1, i_2} = \left(\sum_{j=1}^m (x_{i_1, j} - x_{i_2, j})^r \right)^{\frac{1}{r}} \quad (5.1)$$

Name	Eosinophile	Stabkernige	Segmentkernige	Basophile	Monocyten	Übergangsformen	Lymphocyten
Bu.	-3	0	-13	+1	0	-0.5	+15.5
Lo.	-2.5	+2.5	0	-1	+2	-2.5	+2.5
Sch.	-3.5	+0.5	-6	+1	+0.5	-3	+10.5
We.	-4.5	-6.5	-15.5	+1.5	+3.5	+2.5	+19.5
Be.	-3	-3.5	-4.5	+2	+2.5	-1	+7.5
Ha.	-1.5	-1.5	-11	+1	0	-2.5	+15.5
Ga.	-1	-3.5	-1	-0.5	0	-3.5	+9.5
Gü.	-1	+6	-16.5	+2	-1.5	-2	+10
Go.	-2.5	-4	+10.5	+1.5	+2.5	-2	-5
La.	-3.5	+2	-3	+0.5	+1	-2	+5

Tabelle 5.2: Blutwerteänderungen: 10 Studenten, 800m-Lauf

([CTC89]) oder in Wurzelform

$$d_{i_1, i_2} = \sqrt[r]{\sum_{j=1}^m (x_{i_1, j} - x_{i_2, j})^r} \quad (5.2)$$

Dabei ist d_{i_1, i_2} die Distanz zwischen zwei Objekten X_{i_1}, X_{i_2} , m die Anzahl der Variablen oder Merkmale der Objekte, und $x_{i, j}$ die Ausprägung des j -ten Merkmals des i -ten Objekts. Die identischen Gleichungen 5.1 und 5.2 werden als Minkowski-Metrik oder L_r -Norm mit der Minkowski-Konstante $r \geq 1$ bezeichnet ([BEPW90]).

Das Maß d_{i_1, i_2} gibt an, wie sehr sich zwei Objekte ähneln oder eben nicht ähneln, oder anders gesagt, wie groß die Distanz zwischen zwei Objekten ist. Dabei spielt die reelle Zahl r eine wichtige Rolle ([CTC89]).

Für verschiedene Werte, die die Konstante r annimmt, ergeben sich auch verschiedene Distanzmaße bzw. verschiedene Metriken. Je größer der Wert des Distanzmaßes, desto verschiedener sind die Objekte. Neben den Distanzmaßen gibt es auch echte Ähnlichkeitsmaße, für die gilt, daß ein hoher Wert des Maßes eine hohe Ähnlichkeit zwischen den Objekten darstellt. Einige Metriken aus [BEPW90] und [CTC89] sind in Abbildung 5.2 (Proximitätsmaße) dargestellt.

Bei Nominalskalen haben die Merkmale bzw. Variablen nominale Ausprägungen. Ein Sonderfall der Nominalskalen sind die binären Nominalskalen, die sich dadurch auszeichnen, daß die Variablen nur zwei Werte („nein, ja“, „vorhanden, nicht vorhanden“ etc.) annehmen können. Diese nominalen Werte

Name	Eosinophile	Stabkernige	Segmentkernige	Basophile	Monocyten	Übergangsformen	Lymphocyten
Bu.	-0.367	0.226	-0.890	0.106	-0.721	0.713	0.957
Lo.	0.092	0.933	0.763	-2.014	0.652	-0.527	-0.971
Sch.	-0.825	0.368	0.000	0.106	-0.378	-0.837	0.215
We.	-1.742	-1.612	-1.207	0.636	1.682	2.572	1.550
Be.	-0.367	-0.763	0.191	1.166	0.995	0.403	-0.230
Ha.	1.008	-0.198	-0.636	0.106	-0.721	-0.527	0.957
Ga.	1.467	-0.763	0.636	-1.484	-0.721	-1.147	0.067
Gü.	1.467	1.923	-1.335	1.166	-1.750	-0.217	0.141
Go.	0.092	-0.905	2.097	0.636	0.995	-0.217	-2.084
La.	-0.825	0.792	0.381	-0.424	-0.034	-0.217	-0.601

Tabelle 5.3: standardisierte Blutwerteänderungen

können besonders leicht in eine metrische Skala übersetzt werden, in dem man der einen Ausprägung den Wert 0 („nicht vorhanden“), der anderen den Wert 1 („vorhanden“) zuordnet.

In einer solchen binären Struktur können ebenfalls Ähnlichkeiten zwischen Objekten gesucht werden. [BEPW90] verwendet dazu die Ähnlichkeitsfunktion

$$s_{i_1, i_2} = \frac{a + \delta \cdot e}{a + \delta \cdot e + \lambda \cdot (b + c)} \quad (5.3)$$

mit s_{i_1, i_2} der Ähnlichkeit zwischen den Objekten X_{i_1} und X_{i_2} , δ und λ möglichen konstanten Gewichtungsfaktoren, a der Anzahl der Eigenschaften, die bei beiden Objekten vorhanden sind, b den Eigenschaften, die nur bei X_{i_2} vorhanden sind, c den Eigenschaften, die nur bei X_{i_1} vorhanden sind und e den Eigenschaften, die bei beiden Objekten nicht vorhanden sind.

Je nach Wahl der beiden Gewichtungsfaktoren erhält man die auch in Abbildung 5.2 dargestellten nominalen Metriken, die ebenfalls in [BEPW90] angesprochen werden:

- Tamimoto-Koeffizient,
- M-Koeffizient,
- RR-Koeffizient,
- Kulczynski-Koeffizient und
- Dice-Koeffizient.



Abbildung 5.2: Struktur der Proximitätsmaße nach [BEPW90]

Da sich nominale Skalen in metrische Skalen übersetzen lassen, die besondere Vertiefung nominaler Skalen und Metriken den Rahmen der Arbeit sprengen würde und zudem nicht zum Verständnis der clusteranalytischen Vorgehensweise beiträgt, beschränken wir uns auf metrische Skalen.

Ausgehend von Gleichung 5.1 und Gleichung 5.2 ergeben sich eine Vielzahl von Metriken, wie z.B. die L_1 -Norm³, L_2 -Norm⁴, Mahalanobis-Distanz und Q-Korrelationskoeffizient.

Es sollte jedoch beachtet werden, daß bei der Anwendung der Minkowski-Metriken die Vergleichbarkeit der Maßeinheiten im Vordergrund steht. Wenn diese Voraussetzung nicht vorliegt, dann ist es nötig, die Ausgangsdaten zu standardisieren⁵. Wir haben dies in unserem Beispiel bereits getan (Tabelle 5.3).

Das gebräuchlichste Maß ist die euklidische Distanz. Je nach Wahl von r unterscheidet man die verschiedenen Maße. Je größer die Minkowski-Konstante r gewählt wird, desto stärker wird die Distanz von zwei Objekten überbewertet. Ist r hingegen klein, so werden die Distanzen eher unterbewertet. Die euklidische Distanz folgt aus dem Einsetzen von $r = 2$. Aus diesem Grund heißt die euklidische Distanz auch L_2 -Norm:

³dies wird auch City-Block- oder Manhattan- oder Taxifahrer-Metrik genannt

⁴entspricht der euklidischen Distanz zwischen zwei Objekten, z.B. zwei Körpern im Raum

⁵siehe dazu auch Abschnitt 4.2

$$d_{i_1, i_2} = \sqrt{\sum_{j=1}^m |x_{i_1, j} - x_{i_2, j}|^2} \quad (5.4)$$

Die Distanz zwischen den Objekten ergeben eine quadratische Matrix ähnlich der der Korrelationsmatrix in Abschnitt 4.2, jedoch mit anderer Semantik. Für unser Beispiel ergeben sich die Distanzen wie in Tabelle 5.4 dargestellt.

	Bu.	Lo.	Sch.	We.	Be.	Ha.	Ga.	Gü.	Go.	La.
Bu.	0.000	3.883	2.023	3.903	2.778	1.916	3.667	3.187	4.873	2.488
Lo.	3.883	0.000	2.967	6.145	3.881	3.765	2.900	4.948	3.692	2.059
Sch.	2.023	2.967	0.000	4.914	2.504	2.202	3.115	3.585	3.833	1.332
We.	3.903	6.145	4.914	0.000	3.632	5.089	6.381	6.681	6.020	5.061
Be.	2.778	3.881	2.504	3.632	0.000	3.039	4.002	4.579	2.822	2.602
Ha.	1.916	3.765	2.202	5.089	3.039	0.000	2.419	2.853	4.624	2.941
Ga.	3.667	2.900	3.115	6.381	4.002	2.419	0.000	4.478	4.120	3.262
Gü.	3.187	4.948	3.585	6.681	4.579	2.853	4.478	0.000	5.868	3.937
Go.	4.873	3.692	3.833	6.020	2.822	4.624	4.120	5.868	0.000	3.324
La.	2.488	2.059	1.332	5.061	2.602	2.941	3.262	3.937	3.324	0.000

Tabelle 5.4: Distanzen zwischen Personen im Beispiel

Man benötigt diese Tabelle in Form einer Matrix D ⁶:

$$D = \begin{pmatrix} 0.000 & 3.883 & 2.023 & 3.903 & 2.778 & 1.916 & 3.667 & 3.187 & 4.873 & 2.488 \\ 3.883 & 0.000 & 2.967 & 6.145 & 3.881 & 3.765 & 2.900 & 4.948 & 3.692 & 2.059 \\ 2.023 & 2.967 & 0.000 & 4.914 & 2.504 & 2.202 & 3.115 & 3.585 & 3.833 & 1.332 \\ 3.903 & 6.145 & 4.914 & 0.000 & 3.632 & 5.089 & 6.381 & 6.681 & 6.020 & 5.061 \\ 2.778 & 3.881 & 2.504 & 3.632 & 0.000 & 3.039 & 4.002 & 4.579 & 2.822 & 2.602 \\ 1.916 & 3.765 & 2.202 & 5.089 & 3.039 & 0.000 & 2.419 & 2.853 & 4.624 & 2.941 \\ 3.667 & 2.900 & 3.115 & 6.381 & 4.002 & 2.419 & 0.000 & 4.478 & 4.120 & 3.262 \\ 3.187 & 4.948 & 3.585 & 6.681 & 4.579 & 2.853 & 4.478 & 0.000 & 5.868 & 3.937 \\ 4.873 & 3.692 & 3.833 & 6.020 & 2.822 & 4.624 & 4.120 & 5.868 & 0.000 & 3.324 \\ 2.488 & 2.059 & 1.332 & 5.061 & 2.602 & 2.941 & 3.262 & 3.937 & 3.324 & 0.000 \end{pmatrix} \quad (5.5)$$

Die Matrix enthält (Anzahl Objekte)² Einträge (hier $7^2 = 49$). Davon sind (Anzahl Objekte) Stück immer 0, nämlich die Hauptdiagonaleinträge. Von den restlichen Distanzen sind immer zwei gleich ($d_{i_1, i_2} = d_{i_2, i_1}$), so daß

$$\binom{\text{Anzahl Objekte}}{2} \quad (5.6)$$

⁶D wie Distanz

verschiedene Distanzen übrig bleiben.

In der Praxis finden sich auch andere Proximitätsmaße. Beliebte ist die City-Block-Metrik oder L_1 -Norm. Sie ergibt sich aus der Minkowski-Metrik mit $r = 1$:

$$d_{i_1, i_2} = \sum_{j=1}^m |x_{i_1, j} - x_{i_2, j}| \quad (5.7)$$

Auch die Metrik „Q-Korrelationskoeffizient“ findet man häufig. Im Unterschied zu den beiden obigen Metriken ist diese Metrik kein Distanz-, sondern ein Ähnlichkeitsmaß. Je größer der Wert dieses Maßes ist, desto stärker ist der Zusammenhang und die Ähnlichkeit zwischen den Objekten. Der Q-Korrelationskoeffizient berechnet sich aus der Gleichung

$$r_{i_1, i_2} = \frac{\sum_{j=1}^m (x_{i_1, j} - \bar{x}_{i_1}) \cdot (x_{i_2, j} - \bar{x}_{i_2})}{\left(\sum_j (x_{i_1, j} - \bar{x}_{i_1})^2 \sum_j (x_{i_2, j} - \bar{x}_{i_2})^2 \right)^{\frac{1}{2}}} \quad (5.8)$$

([BEPW90], [FH84]). Diese Berechnung der Ähnlichkeit zwischen zwei Objekten durch den Q-Korrelationskoeffizienten findet unter Berücksichtigung aller Variablen und der Bildung des Durchschnittwertes aller Variablenwerte eines Objektes (\bar{x}_i) statt.

Wie bereits bei den Distanzen geschehen würde man bei Verwendung des Q-Korrelationskoeffizienten eine Matrix bilden, die dann jedoch nicht mehr Distanzmatrix sondern Ähnlichkeitsmatrix hieße. Vergleicht man die Werte einer Ähnlichkeitsmatrix mit denen einer Distanzmatrix so wird schnell deutlich, daß die beiden mehr sind als umgekehrt proportionale Maße. Ein Variablenpaar, dem durch eine euklidische Distanzmatrix ein hoher Abstand und damit eine große Verschiedenartigkeit bescheinigt wird, kann in der Q-Korrelationskoeffizientenmatrix plötzlich ebenso einen sehr hohen Wert, sprich eine große Ähnlichkeit, zugeordnet werden. Dies führt zu dem Ergebnis, daß die Wahl der Proximitätsmaßes, die letztlich das Ergebnis der Clusteranalyse beeinflusst, aufgrund inhaltlicher oder intuitiver Gesichtspunkte durchgeführt werden muß. Manchmal ringt man sich auch dazu durch, die Clusteranalyse mit verschiedenen Proximitätsmaßen zu durchlaufen, bis man zu einem leichter interpretierbaren Ergebnis kommt.

5.3 Arten von Gruppierungsalgorithmen

Die gewonnenen Daten aus den Proximitätsmaßen dienen als Ausgangspunkt für die Bildung der Cluster. Die zu verwendenden Clusteralgorithmen

sind mit dem Ziel verbunden, die verschiedenen Objekte zu eben diesen genannten Clustern oder Gruppe zusammenzufassen. In der einschlägigen Literatur wird eine Vielzahl solcher Gruppierungsalgorithmen, von denen wir nur wenige herausgreifen können, beschrieben. Die Vielzahl der Algorithmen ist ein Schwachpunkt der Clusteranalyse: Auch hier wird das Ergebnis durch die manuelle Wahl des Clusterverfahrens zuweilen stark verändert.

Aufgrund der Anzahl der Merkmale (Variablen), die ein Merkmalsträger (Objekt) trägt, unterscheidet man zwischen monothetischen und polythetischen Verfahren. Die monothetischen Verfahren treten dann zu Tage, wenn als Gruppierungskriterium nur eine einzige Variable Verwendung finden kann oder soll. Da aber die Clusteranalyse den Vorteil bietet, bei der Gruppierung simultan alle Beschreibungsmerkmale (Variablen) heranziehen zu können (polythetische Verfahren) lassen wir die weniger interessanten monothetischen Verfahren außer Acht.

In Abbildung 5.3 findet sich ein Überblick über verschiedene Clusterungsverfahren.

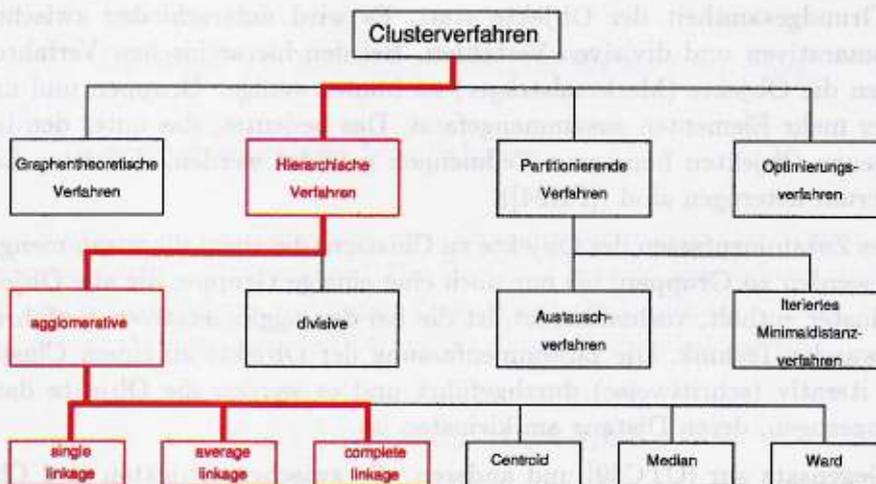


Abbildung 5.3: Übersicht verschiedener Clusterverfahren nach [BEPW90]

Bei den polythetischen Verfahren unterscheidet man zwischen den

- partitionierenden Verfahren und
- hierarchischen Verfahren

Da die anderen Verfahren eine untergeordnete Rolle bei der Clusteranalyse spielen werden diese auch nicht zum Gegenstand dieser Arbeit gemacht.

Bei den partitionierenden Verfahren geht man von einer vorgegebener Anzahl von Clustern aus. Diese vorgegebene Gruppierung von Objekten wird

auch als Startpartition bezeichnet. Diese Clusterung wird durch vorher definierten Kriterien festgelegt. Vorteil dieses Verfahrens ist, daß einmal durchgeführte Partitionen immer wieder revidiert werden können ([BEPW90]). Der grosse Nachteil ist, daß dieses Verfahren sehr zeit- und speicheraufwendig ist ([CTC89]). So sind für eine Einteilung von 12 Objekten zu 5 Gruppen

$$5^{12} = 244.140.625 \quad (5.9)$$

Möglichkeiten eine Startpartition aufzustellen aufgezeigt. Aufgrund dieser schlechten Performanz wird dieses Verfahren nicht weiter in unserer Arbeit erläutert.

Bei den hierarchischen Verfahren findet eine fortschreitende Untergliederung der Grundgesamtheit der Objekte statt. Es wird unterschieden zwischen agglomerativen und divisiven Verfahren. Bei den hierarchischen Verfahren werden die Objekte (Merkmalsträger) zu immer weniger Gruppen und mit immer mehr Elementen zusammengefasst. Das bedeutet, dass unter den heterogenen Objekten homogene Teilmengen gebildet werden, die unter sich wiederum heterogen sind ([FH84]).

Dieses Zusammenfassen der Objekte zu Clustern, die ebenfalls zusammengefasst werden zu Gruppen, bis nur noch eine einzige Gruppe, die alle Objekte/Cluster enthält, vorhanden ist, ist die bei den agglomerativen Verfahren angewandte Technik. Die Zusammenfassung der Objekte zu einem Cluster wird iterativ (schrittweise) durchgeführt und es werden die Objekte dazu herangezogen, deren Distanz am kleinsten ist.

Im Gegensatz zur [CTC89] und anderen, die zwischen Objekten und Clustern unterscheiden, wollen wir in diesem Kapitel diese Unterscheidung nicht konsequent durchführen und stattdessen einzelne Objekte als einelementige Cluster auffassen. Aus der ständigen Zusammenfassung von Objekten und Clustern ergibt sich eine baumartige Struktur. Abbildung 5.4 zeigt dies.

Zusätzlich zu den agglomerativen Verfahren existieren unter den hierarchischen Clusteranalyseverfahren die divisiven Verfahren, die sich von obigen durch eine Umkehrung des Clusterungsvorgangs von „Bottom-Up“ (aus Objekten Cluster machen bis nur noch ein großes Supercluster übrig bleibt) zu „Top-down“ (beginnend mit einem großen Supercluster werden schrittweise kleinere heterogene Subcluster identifiziert) unterscheiden. „Da den divisiven Verfahren bei der praktischen Anwendung jedoch kaum eine Bedeutung zukommt“ ([BEPW90]) betrachten wir dieses Verfahren nicht näher und beschränken uns auf das hierarchisch agglomerative Verfahren.

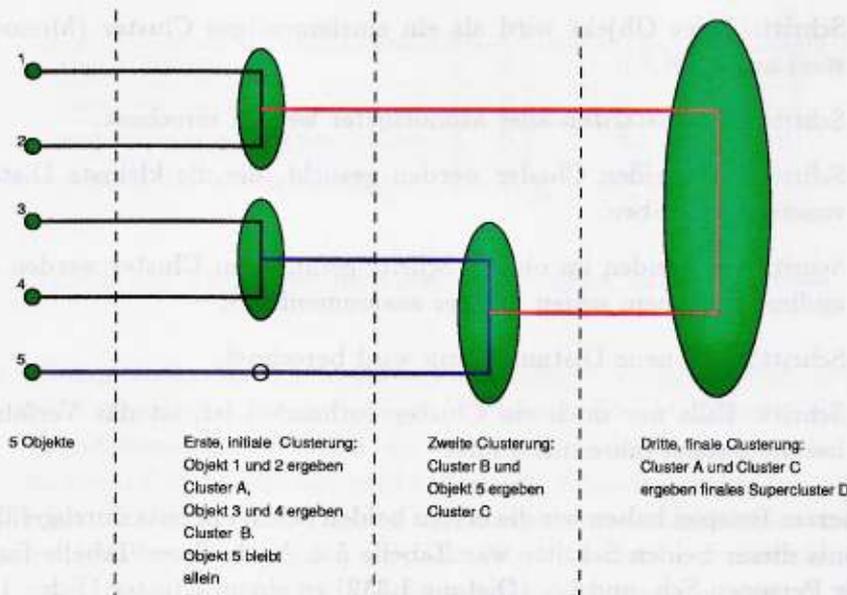


Abbildung 5.4: Baumartige Struktur bei der Clustering

5.4 Agglomerative Verfahren

Eine Partition ist die Zerlegung einer Grundgesamtheit. Der Vorgang des Zerlegens heißt „partitionieren“. Die Partition besteht aus disjunkten Mengen von Objekten der gegebenen Grundgesamtheit, die als Cluster („Ansammlungen“) fungieren. [FH84] drückt diesen Sachverhalt so aus:

$$C = \bigcup_{k=1}^g C_k = I, C_k \cap C_j = \emptyset, k \neq j \tag{5.10}$$

mit C der Partition und C_1 bis C_g den Clustern und I der Grundgesamtheit der Objekte.

In der Clusteranalyse sucht man meist nach einer Partition, auch wenn es, konträrer Weise, häufig durchaus sinnvoll sein kann, ein Objekt mehreren Clustern zuzuordnen. Dann ist ein Cluster jedoch keine Partition mehr. Wird die Disjunktheits-Bedingung aufgegeben handelt es sich um nichtdisjunkte Klassifikationsverfahren („clumping methods“, [FH84]), für die wir auf [Boc74] verweisen möchten.

Die Berechnung der Distanz bei den agglomerativen Verfahren basiert meistens auf der euklidischen Distanz. Der Ablauf des hierarchisch agglomerativen Verfahrens kann anhand des Diagramms in Abbildung 5.5, welche aus [BEPW90] stammt, veranschaulicht werden.

1. Schritt: Jedes Objekt wird als ein einelementiges Cluster (Monocluster) aufgefaßt.
2. Schritt: Die Distanzen aller Monocluster werden errechnet.
3. Schritt: Die beiden Cluster werden gesucht, die die kleinste Distanz voneinander haben.
4. Schritt: Die beiden im obigen Schritt gefundenen Cluster werden kumuliert (zu einem neuen Cluster zusammenfaßt).
5. Schritt: Eine neue Distanzmatrix wird berechnet.
6. Schritt: Falls nur noch ein Cluster vorhanden ist, ist das Verfahren beendet, sonst fahre mit 2 fort.

In unserem Beispiel haben wir die ersten beiden Schritte bereits durchgeführt. Ergebnis dieser beiden Schritte war Tabelle 5.4. Nach dieser Tabelle fassen wir die Personen Sch. und La. (Distanz 1.332) zu einem Cluster {Sch., La.} zusammen.

Wir haben aus zwei Personen (Objekten/Monoclustern) ein neues Cluster gemacht. Das bedeutet, daß die Distanzmatrix um eine Zeile und eine Spalte verkleinert werden muß (zwei Cluster sind entfallen, eines neu hinzugefügt worden).

Bei der Berechnung der neuen Distanzmatrix muß nun auch die Distanz zu einem (Nichtmono)-Cluster berechnet werden. Dafür gibt es wiederum verschiedene Strategien unter denen es sich für eine zu entscheiden gilt. Möchte man die Distanz $D(R, P+Q)$ eines Clusters R von einem Cluster $P+Q$, welches aus Clustern P und Q entstanden ist, berechnen, so kann dies mittels der folgenden parametrisierten Gleichung aus [BEPW90] geschehen:

$$D(R, P+Q) = A \cdot D(R, P) + B \cdot D(R, Q) + E \cdot D(P, Q) + G \cdot |D(R, P) - D(R, Q)| \quad (5.11)$$

Dabei ist $D(R, P)$ die Distanz zwischen den Clustern R und P , etc.

Die Parameter A , B , E und G sind wählbare Parameter. Anhand der Parameter werden die verschiedenen Strategien unterschieden (s. Tabelle 5.5).

In dieser Tabelle steht $|P|$ für die Anzahl der Objekte/Monocluster im Cluster P und $|Q|$ für ebendiese in Q .

- „Single linkage“. Bei diesen Verfahren wird die Distanz zwischen zwei Clustern bestimmt durch die Distanz der beiden am nächsten zusammen liegenden Objekte (minimal-Objekt-Distanz). Aus diesem Grund wird „single linkage“ häufig auch als „nearest neighbour“ bezeichnet ([EMT95]).

	A	B	E	G
single linkage	0.5	0.5	0	-0.5
complete linkage	0.5	0.5	0	0.5
ungewichtetes average linkage	0.5	0.5	0	0
gewichtetes average linkage	$\frac{ P }{ P + Q }$	$\frac{ P }{ P + Q }$	0	0

Tabelle 5.5: Parameter für Linkage-Strategien nach [BEPW90]

- „Complete-linkage“. Das „Complete-linkage“ Verfahren (vgl. [FH84]) stellt eine dem „Single-linkage“ Verfahren Objekt-Distanz-reziproke Vorgehensweise dar: Nicht die minimal-Objekt-Distanz sondern die maximal-Objekt-Distanz wird verwendet. Zunächst sucht man also für jedes Clusterpaar die Distanz der am weitesten voneinander entfernten Objekte, in dem man z.B. alle Objektdistanzen D kalkuliert und $\max(D)$ als Maximal-Objektdistanz des Clusterpaares notiert. Unter den Clusterpaaren wird dasjenige als zu verbindend ausgewählt, welches die geringste Maximal-Objektdistanz aufweist.
- ungewichtetes „Average-linkage“. Vom Prinzip her ist dieses Verfahren eine Mischung aus „Single-linkage“- und „Complete-Linkage“-Strategie. Beim „Average-linkage“ wird die Distanz zwischen den Clustern bestimmt durch die mittlere Distanz der beiden am weitesten voneinander entfernten und den beiden am nächsten aneinander liegenden Objekten.
- das gewichtete „Average-linkage“ zeigt, daß neben Konstanten auch Variablen, wie eben die Anzahl der Objekte in den Clustern, in eine Distanzbewertung einbezogen werden kann.

Abbildung 5.6 verdeutlicht die Unterschiede zwischen den drei Verfahren.

Die Formel zur Berechnung der Distanz zwischen zwei Objekten bei dem von uns in diesem Beispiel verwendeten „Single Linkage“-Verfahren läßt sich (s. [BEPW90]) vereinfachen zu

$$D(R, P + Q) = \min\{D(R, P), D(R, Q)\} \quad (5.12)$$

Bu. hat z.B. von Sch. die Distanz 2.023 und von La. die Distanz 2.488. Nach der obigen Gleichung ist dann die Distanz zwischen Bu. und dem Cluster (Schu.,La.) 2.023. Diese Berechnung wird nun für alle Cluster (auch Monocluster) durchgeführt und wir erhalten die neue Distanzmatrix 5.6.

Wir wiederholen:

	(Sch.,La.)	Bu.	Lo.	We.	Be.	Ha.	Ga.	Gü.	Go.
(Sch.,La.)	0	2.023	2.059	4.914	2.504	2.202	3.115	3.585	3.324
Bu.		0	3.883	3.903	2.778	1.917	3.667	3.187	4.873
Lo.			0	6.145	3.981	3.765	2.900	4.948	3.692
We.				0	3.632	5.089	6.381	6.681	6.020
Be.					0	3.039	4.002	4.579	2.822
Ha.						0	2.419	2.853	4.624
Ga.							0	4.478	4.120
Gü.								0	5.868
Go.									0

Tabelle 5.6: Distanzen nach erster Fusion

- Beim ersten Schritt werden aus der Distanzmatrix die zwei Objekte zu einem Cluster zusammengeführt die die kleinste Distanz aufweisen. Mit diesen neugewonnenen Cluster 1 wird eine neue Distanzmatrix mit den übriggebliebenen Objekten anhand der Formel errechnet.
- Danach wird Cluster 1 mit dem Objekt(x) zu einen neuen Cluster 2, das wir auch als Obercluster bezeichnen können, vereinigt, welches die kleinste Distanz aufweist. Es kann sein dass sich auch zwei einzelne Objekte zu einen neuen Cluster 2a zusammenfassen lassen, wenn diese die kleinste Distanz aufweisen so das wir entweder Cluster 2 und die restlichen Objekte erhalten oder Cluster 1, Cluster 2a und restliche Objekte.
- Dann wird wieder eine neue Distanzmatrix mit den neu gewonnenen Clustern und den übrig gebliebenen Objekten (oder Clustern) gebildet und wir führen diese Berechnung weiter bis wir nur noch einen einzigen Cluster erhalten der auch Supercluster genannt werden kann. Beim jedem Schritt wird die Distanzmatrix um genau ein Objekt oder Cluster reduziert.

Der Fusionierungsalgorithmus bei diesen sowohl auch bei den anderen Verfahren läßt sich durch ein Dendrogramm darstellen. In einem Dendrogramm kann veranschaulicht werden, welche Cluster oder Monocluster zusammengefaßt werden. Häufig findet man in Dendrogrammen auch Distanzangaben. Abbildung 5.7 zeigt die schrittweise Erstellung eines Dendrogramms für unser Sportlerbeispiel.

Rot dargestellt ist immer das, was nach einem Fusionierungsschritt in das Dendrogramm hinzugefügt wird.

5.5 Kritik

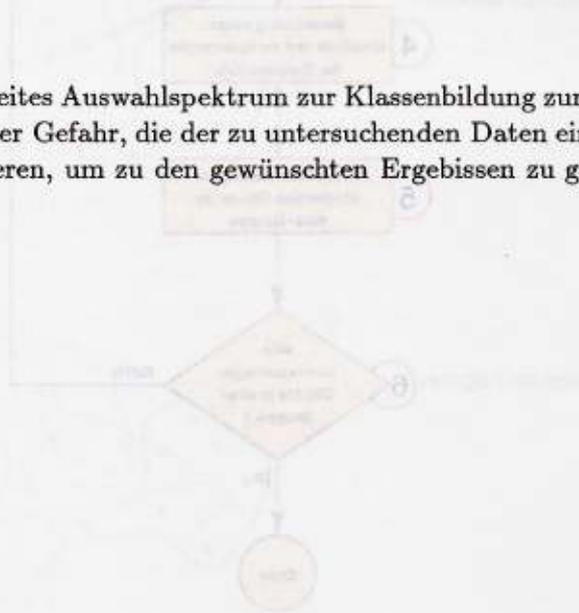
Die Clusteranalyse bietet sehr viele Möglichkeiten zur Klassifizierung (Gruppierung) von Merkmalsträgern in homogene und voneinander deutlich unterschiedliche Klassen (Gruppen). Dadurch steht man des öfteren vor dem Problem, bei der Anwendung der Clusteranalyse für ein geeignetes Verfahren zu votieren.

Ein für eine Anwendung geeignetes Klassifikationsverfahren läßt sich aber schon aufgrund der individuellen Interpretation und Präferenzierung von Fragestellungen Wie z.B.:

- Wie viele Objekte sollen berücksichtigt werden?
- Welche Ähnlichkeits- oder Distanzmaße sollen herangezogen werden?
- Welche Verfahren sollen angewendet werden?
- Wie viele Gruppen sollen gebildet werden?

bestimmen.

Da man aber eines breites Auswahlpektrum zur Klassenbildung zur Verfügung hat läuft der Anwender Gefahr, die der zu untersuchenden Daten einer Erhebung so zu manipulieren, um zu den gewünschten Ergebnissen zu gelangen.



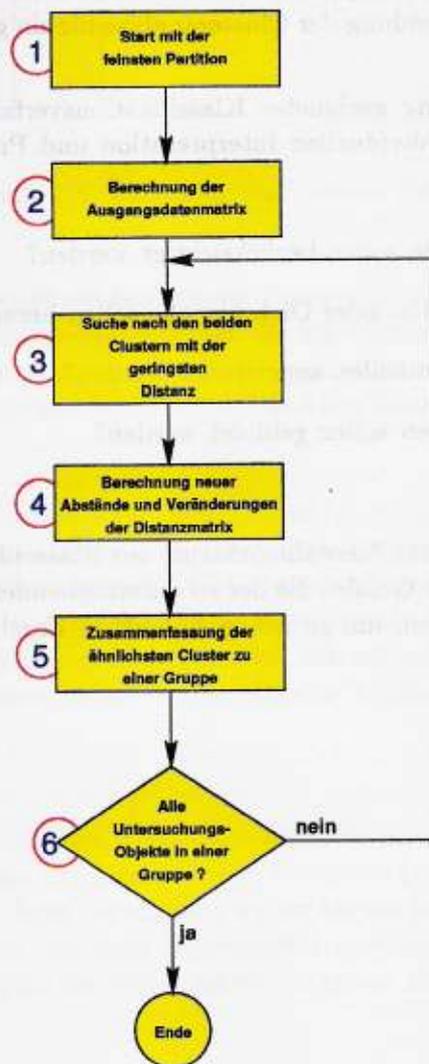


Abbildung 5.5: Ablauf des hierarchisch agglomerativen Verfahrens nach [BEPW90]

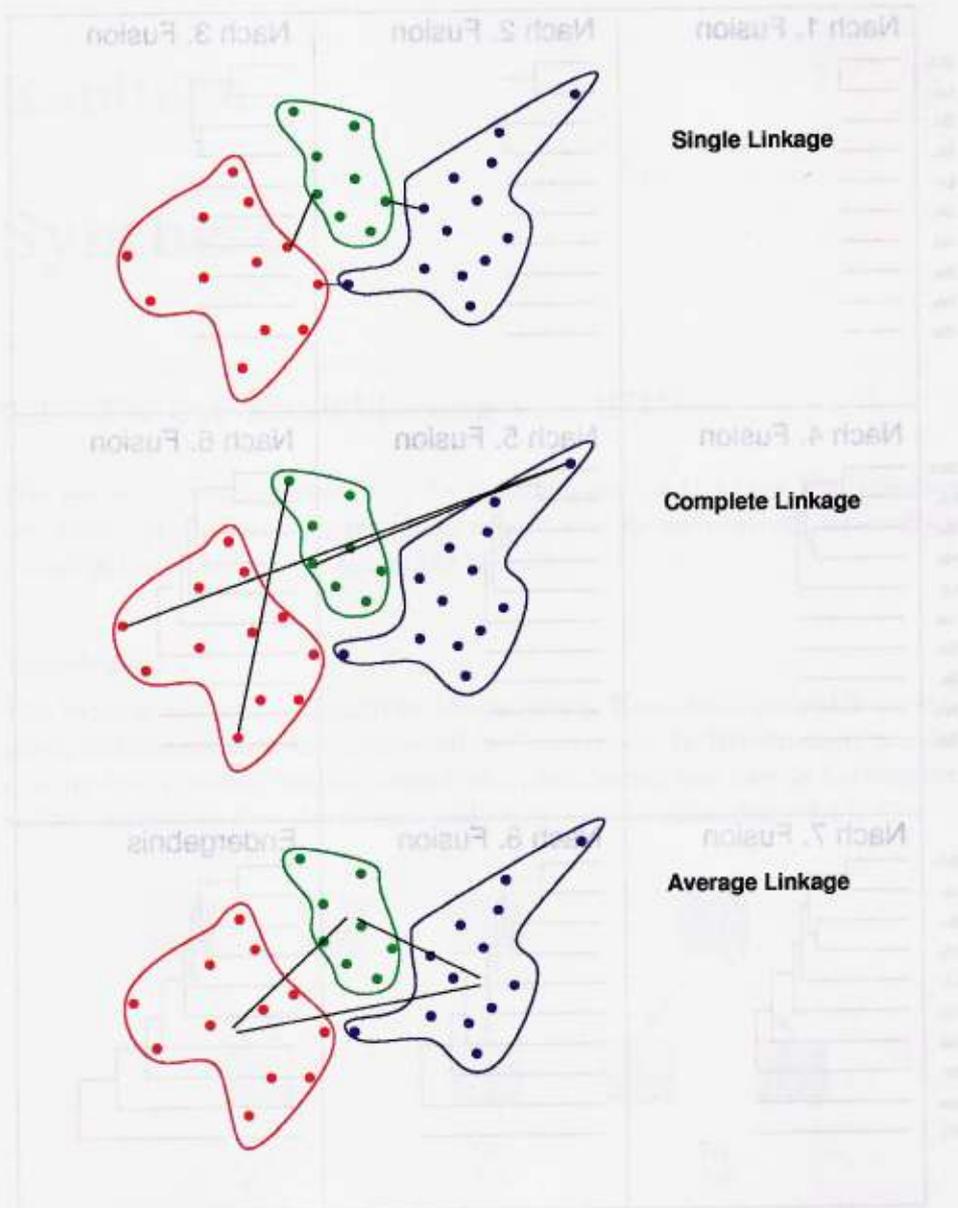


Abbildung 5.6: Linkage-Strategien

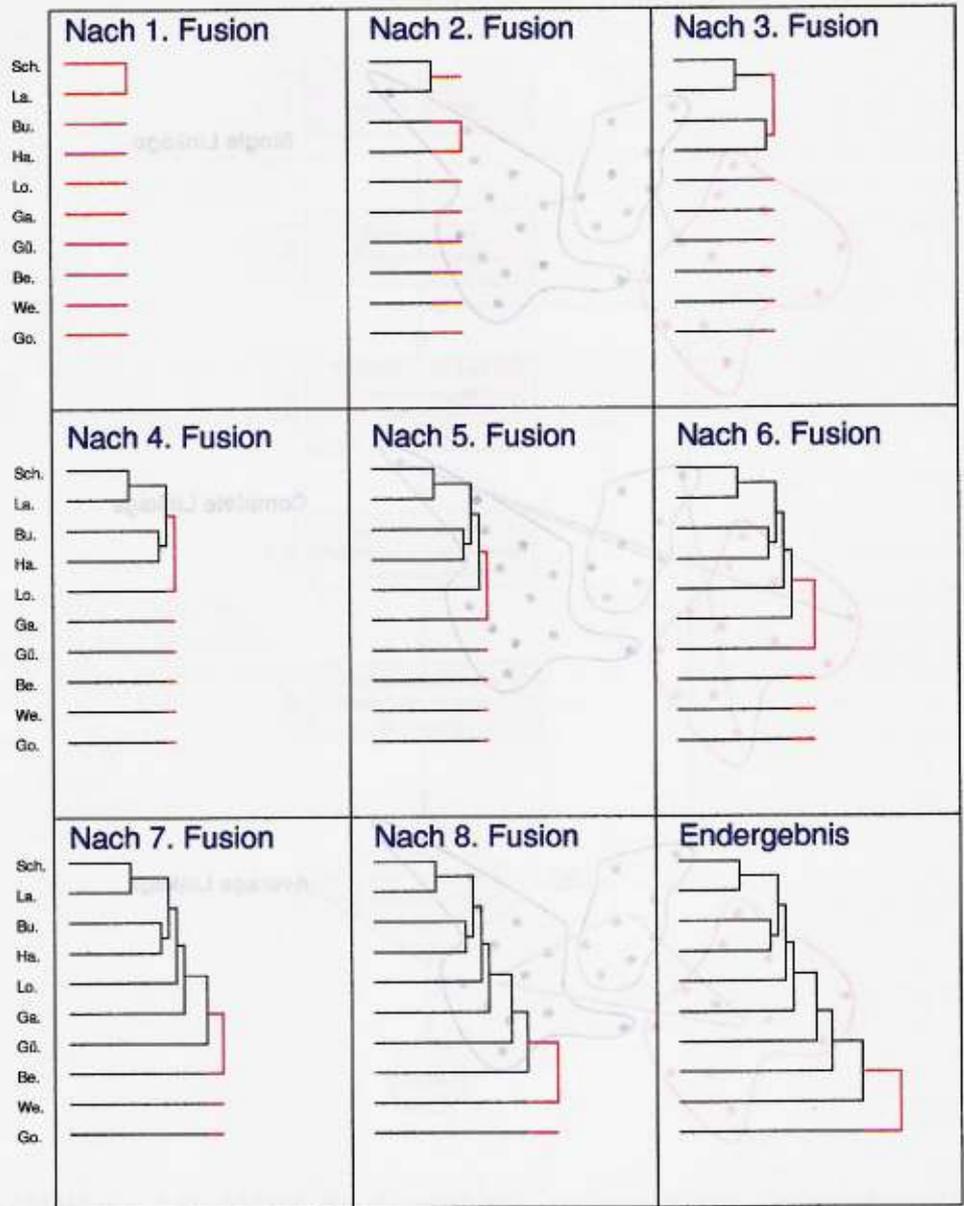


Abbildung 5.7: Dendrogramm zum Sportlerbeispiel

Kapitel 6

Synthese

6.1 FA zur Modellierung von BBN

Wir wollen prüfen, inwiefern sich die Faktorenanalyse (FA) zur Modellierung von Beliefnetzen (BBN) eignen kann, sowohl was die qualitativen als auch die quantitativen Modellierungsaspekte betrifft.

Topologien

Wie benötigen zur quantitativen Modellierung Kausalzusammenhänge zwischen Zufallsvariablen und letztendlich Hypothesen. In Beliefnetzen, zumindest in denen einfach verbundener Natur, auf die wir uns hier ja beschränken wollen, existieren drei Arten von Teiltopologien (s. Abbildung 6.1).

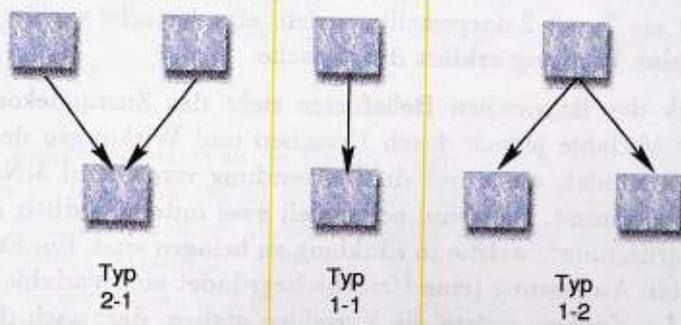


Abbildung 6.1: Topologietypen

Die erste Teiltopologie ist die durch zwei (oder mehr) Ursachen beeinflusste Zufallsvariable (ein Knoten mit mehreren eingehenden Kanten). Entsprechend der Semantik der Faktorenanalyse könnte dies aufgefasst werden als

eine Menge von Zufallsvariablen, die zu einem Faktor (Wirkung) zusammengefasst werden. Die Variablen in Beliefnetz sind dadurch, daß es sich um einfach verbundene Beliefnetze handelt, zumindest nicht direkt voneinander abhängig. In der Faktorenanalyse gilt für die Variablen daß sie zwar abhängig sein können aber nicht müssen. Für mehrere Faktoren, die aus den Variablen extrahiert werden gilt, daß sie stochastisch unabhängig sein sollen. Dies ist für die Wirkungsvariable (den Knoten mit den eingehenden Kanten) keine Einschränkung da es sich sowieso nur um einen einzigen „Faktor“ handelt und daher nicht stochastisch abhängig sein kann.

Kausalzusammenhänge und Faktoren Die Idee ist daher nun folgende:

Wir entdecken einen Zusammenhang zwischen Variablen und einem einzigen Faktor mit den Mitteln der Faktorenanalyse. Wir können für diesen Zusammenhang sagen daß zwischen Variablen und Faktor „irgend ein“ Zusammenhang existiert. Leider, führt diese schwache Aussage zu dem Problem, daß wir keine mathematisch begründete Formulierung für die Art des gefundenen Zusammenhangs erhalten. Wir wissen also nicht ob der Faktor eine Ursache für das Zustandekommen der Variable darstellt oder eine Wirkung, oder für einen nicht kausalen Zusammenhang.

Laut [FH84] und anderen erklärt ein Faktor eine Menge von Variablen bzw. deren Zustände kommen. Betrachten wir die intensionale Semantik der Faktorenanalyse und bewegen wir uns aus einer eher hermeneutischen Position auf dieses Thema zu, so könnten wir die Aussage Fahrmeiers interpretieren als eine vom Faktor ausgehende Kausalbeziehung.

Ein so gefundener Zusammenhang müsste daher nicht vom Typ 2-1 sein, sondern eher als Typ 1-2 dargestellt werden: eine Ursache erklärt eine Wirkung, nicht eine Wirkung erklärt die Ursache.

Die Semantik der Bayesschen Beliefnetze sieht das Zustandekommen des Wertes einer Variable jedoch durch Ursachen und Wirkungen der Variable beiderseits begründet, was durch die Verwendung von π - und λ -Nachrichten zum Ausdruck kommt. Dies sind prinzipiell zwei unterschiedlich Auffassungen von „Begründung“, welche in Einklang zu bringen sind. Ein Faktor stellt nach der ersten Auffassung (eine Ursache begründet eine Variable) einen Elternknoten der Knoten, indem die Variablen stehen, dar; nach der zweiten Auffassung ist ein Faktor sowohl ein Kindknoten einer Menge von Variablen als auch ein Elternknoten mehrerer Variablen.

Faktorladungen und bedingte Wahrscheinlichkeiten Betrachten wir nun zur weiteren Klärung ein Ergebnis der Faktorenanalyse: die Faktorladungsmatrix. Sie erklärt für jede Variable den Zusammenhang zu den Fak-

toren anhand einer Linearkombination.

Diese Linearkombination begegnet uns bei der Berechnung von Belief-Werten wieder. Bei der Berechnung von π -Werten einer Hypothese K_i wird die Summe aus dem Produkt von Verbindungsmatrixeinträge und π -Werte der Hypothesen im Vorgängerknoten gebildet. Ersetzt man alle Matrixeinträge durch die entsprechende Faktorladung und die π -Werte durch die passenden Faktoren, so erhält man eine Äquivalenz zwischen den beiden. Diese Idee könnte zwar mathematisch (symbolisch) begründet werden, semantisch jedoch würde jeder Hypothese des Vorgängerknotens ein Faktor und jeder Hypothese des Nachfolgerknotens eine Variable gleichgestellt. Dies würde dazu führen, daß, da in Nachfolgerknoten mehr Hypothesen als im Knoten selbst vorhanden sein können, mehr Faktoren als Variablen vorhanden sein müssten, was jedoch gerade nicht der Sinn der Faktorenanalyse ist. Umgekehrt, das heißt den Hypothesen des Knotens die Faktoren zuzuordnen und den Hypothesen des Nachfolgerknotens die Variablen, entstehen aus analogen Grund die gleichen Probleme.

Interpretation von Faktorladungen als Werte der Verbindungsmatrix. Die Interpretation von Faktorladungen als Einträge in der *link*-Matrix ist nicht möglich, da nicht bekannt ist (s.o.) welche Art von Zusammenhang die Faktorladungsmatrix darstellt. Es kann sich dabei um kausale Zusammenhänge handeln. Andererseits können die Korrelationen, und Faktorladungen sind nichts anders als Korrelationen, zwischen Faktoren und Variablen auch wechselseitige, Gemeinsamkeits- und formale Korrelationen darstellen (Ausführungen zu den verschiedenen Arten von Korrelationen lassen unter anderem in [KSV96] finden).

Es scheint eine Frage der persönlichen Präferenzen und der fachspezifischen Kenntnisse zu sein, als was die Faktoren und die Korrelationen zu den Variablen interpretiert werden.

Von der Faktorenanalyse indirekt zum Bayesschen Netz. Ohne die Gründe für den praktischen Einsatz von faktoranalytische Methoden ad absurdum führen zu wollen, obwohl häufige negative Kritik zu hören ist, müssen wir leider feststellen, daß allein die Problematik der Interpretation eine geistig derart hohen, und maschinell nicht leistbaren Aufwand darstellt, der das Aufführen eines algorithmisch und nicht manuell beeinflussten Verfahrens unmöglich oder zumindest nicht sinnvoll macht.

Es ist ratsamer, sich darauf zu beschränken die Faktorenanalyse ein Glied der medizinischen, oder allgemein anwendungsfachlichen, Kette der Erkenntnisgewinnung zu betrachten und erst mittels dieser Erkenntnisse (dieses Wissens) Beliefnetze qualitativ zu modellieren. Der Schritt von der Faktorenanalyse zur Modellierung Bayesscher Netze kann daher nur indirekt unter Zuhilfe-

nahme anderer erkenntnisgewinnender Methoden gegangen werden.

Die Übertragung der faktoranalytischen Ladungsmatrizen zu Werten, die der Beschreibung des quantitativen Teils von Beliefnetzen dienlich sein könnten, ist ebenso aufgrund der nicht zu vereinbarenden semantischen Differenzen zwischen Korrelationen in Faktorladungsmatrizen und zum Teil bedingten Wahrscheinlichkeiten in Beliefnetzen unangebracht.

Semantik und Anschaulichkeit. Bayesche Beliefnetze leben gerade zu von ihrer hohen Anschaulichkeit und damit Verwendbarkeit für den nicht auf Probabilistik spezialisierten Anwender, der in unserem Fall einem Mediziner entspricht. Es kann sich jedoch auch um einen Anwender aus anderen Wissenschaftlichen Fachbereichen handeln. Bayessche Netze bieten dem Anwender durch die graphische Repräsentation schnellen Zugang zu einer probabilistischen Methode eines Expertensystems¹ und nimmt somit die Angst vor scheinbar unerklärlichen Phänomenen der Software.

Die Mischung von Faktorenanalyse und Bayesschen Netzen auf rein mathematischer Basis ist, wenn überhaupt möglich, nur unter Hinnahme intersemantischer Störungen zwischen Faktorenanalyse und Bayesschen Netzen möglich. Wir können zwar gewisse Zusammenhänge zwischen diesen Gebieten erkennen, finden jedoch keine sinnvollen und semantisch in beide Berechnungsmodelle passenden Sinnzusammenhänge, die nicht die Klarheit von Beliefnetzen und das Verständnis und die Selbstverständlichkeit derselben invasiv beeinflussen würde.

Der Vorgang der Erkenntnisgewinnung durch die Faktorenanalyse bis zur Darstellung von Kausalwissen in Beliefnetzen hat genau dort seine Lücke, wo es an die Modellierung der Ursache-Wirkungszusammenhänge geht. Aus diesem Grund kann die Faktorenanalyse nicht als Grundlage für die Modellierung von Beliefnetzen verwendet werden.

6.2 CA zur Modellierung von BBN

In der Clusteranalyse (CA) findet man, wenn es die Daten erlauben², Cluster, die homogene Teilmengen einer Grundmenge heterogener Natur darstellen. Diese Gruppen unterscheiden sich mehr oder weniger stark. Die von uns vorgestellten Metriken sowie Verfahren geben uns ein Maß für die Unterschiedlichkeit der Gruppen. Diese Erkenntnisse, daß Objekte aufgrund

¹Wir wollen dabei „Expertensystem“ nicht als Architektur (wie z.B. in [Fra97] angegeben, sondern allgemein als System, daß Aufgaben eines Experten für einen Fachfremden übernimmt, verstanden wissen.

²die von uns untersuchten Daten im Kapitel Clusteranalyse konnten uns diese Freude nicht bereiten

ihrer Eigenschaften unterschiedlichen Gruppen zugeordnet werden können, bringt uns auf die Idee, dies als eine bestimmte Struktur in einem Beliefnetz darzustellen.

Liefert die Clusteranalyse aus einer Grundgesamtheit zwei Cluster, dann unterscheiden sie sich voneinander. Die Unterscheidung kommt zustande aufgrund der Werten (Merkmale) der Objekte innerhalb dieser Cluster. Weiterhin haben wir erkannt, daß es in der Grundgesamtheit Objekte einer Art und Objekte einer zweiten Art gibt.

Wir können dann, zum Beispiel mittels der multivariaten Diskriminanzanalyse, die in der Literatur zu finden ist, weitere Objekte in diese Cluster einfügen oder den Clustern zuordnen. Dies wird auch tatsächlich getan:

Hat man ein Cluster „Kranke“ und ein Cluster „nicht Kranke“ gefunden, so kann man einen neuen Patienten aufgrund seiner Merkmale in eine dieser Gruppen einteilen. Jedes Cluster kann als neue Variable betrachtet und damit in einem Beliefnetz als Knoten dargestellt werden. Ursachen dieser Knoten sind dann die in der Clusteranalyse analysierten Merkmale.

Hätten wir in unseren Sportlerbeispiel zwei Cluster gefunden, so hätten wir sie zunächst interpretieren müssen. Nehmen wir weiter an wir hätten als Interpretation feststellen können, daß die Personen (Objekte) im ersten Cluster als Trainierte und die Personen in zweiten Cluster als Untrainierte identifizieren können. Dann hätten wir daraus zwei Cluster „Trainierte“ und „Untrainierte“ extrahieren können. Wir hätten nun eine Interpretation der Zusammenhänge zwischen den Clustern durchführen müssen. Eine Feststellung eines Ursache-Wirkungs-Zusammenhangs zwischen den Merkmalen und den Clustern hätte wohl dazu geführt, daß die Tatsache, das eine Person Mitglied eines bestimmten Clusters ist, Rückschlüsse auf die Veränderung des Blutbildes und damit der gemessenen Merkmale ziehen lässt.

Wir modellieren daher das Ergebnis der Clusteranalyse als zwei Knoten entsprechend je einem Cluster, die als Elternknoten sieben weiterer Knoten, nämlich der Merkmale (Eosinophile, Stabkernige, Segmentkernige) etc., fungieren. Jeder Elternknoten ist mit jedem Kindknoten verbunden.

In einer anderen Clusteranalyse könnte wiederum der umgekehrte Fall eintreten, daß die Cluster nach der entsprechenden Interpretation als Folgen der Merkmale erkennbar sind. Dann würde die Modellierung als Netz anders herum verlaufen. Die Elternknoten entsprächen den Merkmalen, die Kindknoten den Clustern.

In vielen Fällen ist es noch günstiger verschiedene Gruppen nicht als unterschiedliche Knoten (Zuffalsvariablen) sondern als eine Zufallsvariable mit unterschiedlichen Hypothesen er selben zu modellieren. Die Bezeichnung des die Cluster darstellenden Knotens ergäbe sich dann aus dem allgemeinen Kontext, aus dem die heterogene Grundgesamtheit stammt.

Die Anwendung und Einbindung so entstandener Teile von Beliefnetzen hängt allein vom Anwendungskontext ab.

Die Anwendung und Einbindung so entstandener Teile von Beliefnetzen hängt allein vom Anwendungskontext ab. In der Praxis ist dies oft ein Problem, da die meisten Beliefnetze in einem isolierten Kontext entwickelt werden. Die Integration dieser Teile in ein größeres System erfordert eine sorgfältige Analyse des Anwendungskontextes. Dies umfasst die Identifizierung der Schnittstellen, die Kompatibilität der Datenformate und die Anpassung der Logik an die spezifischen Anforderungen des Systems. Ein weiterer Aspekt ist die Sicherstellung der Konsistenz und der Korrektheit der Daten während der Integration. Dies kann durch die Verwendung von Validierungsroutinen und die Implementierung von Fehlerbehandlungsmechanismen erreicht werden. Die erfolgreiche Einbindung von Beliefnetzen ist entscheidend für die Skalierbarkeit und Flexibilität eines Systems. Sie ermöglicht die Erweiterung des Systems um neue Funktionen und die Integration von Daten aus verschiedenen Quellen. Dies ist besonders wichtig in der heutigen Welt der Daten, in der die Integration von Informationen aus verschiedenen Kanälen eine zentrale Rolle spielt. Die Herausforderung besteht darin, diese Integration so zu gestalten, dass sie nahtlos und effizient erfolgt. Dies erfordert eine enge Zusammenarbeit zwischen den Entwicklern der Beliefnetze und den Architekten des Gesamtsystems. Nur durch eine sorgfältige Planung und Implementierung kann die volle Leistungsfähigkeit der Beliefnetze ausgeschöpft werden. Die Integration von Beliefnetzen ist ein komplexer Prozess, der viel Zeit und Ressourcen erfordert. Es ist jedoch ein Prozess, der sich lohnt, da er die Leistungsfähigkeit und die Flexibilität eines Systems erheblich steigern kann. Die Integration von Beliefnetzen ist ein Schlüsselschritt bei der Entwicklung moderner, datenzentrierter Systeme. Sie ermöglicht die Schaffung von Systemen, die in der Lage sind, mit der raschen Veränderung der Datenlandschaft umzugehen. Die Integration von Beliefnetzen ist ein Prozess, der die Grenzen zwischen verschiedenen Systemen verwischt und die Schaffung eines einheitlichen Ökosystems ermöglicht. Dies ist ein wichtiger Schritt zur Realisierung der Vision einer nahtlos integrierten Datenlandschaft. Die Integration von Beliefnetzen ist ein Prozess, der die Vorteile der künstlichen Intelligenz und der Datenanalyse in vollem Maße ausschöpft. Sie ermöglicht die Schaffung von Systemen, die in der Lage sind, komplexe Aufgaben zu lösen und die Entscheidungsfindung zu unterstützen. Die Integration von Beliefnetzen ist ein Schlüsselschritt bei der Entwicklung von Systemen, die die Anforderungen der modernen Welt erfüllen. Sie ermöglicht die Schaffung von Systemen, die in der Lage sind, mit der raschen Veränderung der Datenlandschaft umzugehen und die volle Leistungsfähigkeit der künstlichen Intelligenz auszuschöpfen. Die Integration von Beliefnetzen ist ein Prozess, der die Vorteile der künstlichen Intelligenz und der Datenanalyse in vollem Maße ausschöpft. Sie ermöglicht die Schaffung von Systemen, die in der Lage sind, komplexe Aufgaben zu lösen und die Entscheidungsfindung zu unterstützen. Die Integration von Beliefnetzen ist ein Schlüsselschritt bei der Entwicklung von Systemen, die die Anforderungen der modernen Welt erfüllen. Sie ermöglicht die Schaffung von Systemen, die in der Lage sind, mit der raschen Veränderung der Datenlandschaft umzugehen und die volle Leistungsfähigkeit der künstlichen Intelligenz auszuschöpfen.

Kapitel 7

Zusammenfassung und Ausblick

Wir haben in unserer Arbeit schnell festgestellt, daß die in ersten Kapitel verwendeten naiven statistischen Methoden nicht ausreichend sind, um zum einen angenehm zum zweiten zeitsparend und zum dritten speicheökonomisch statistisch begründetes Wissen darzustellen und darüber zu resonieren.

Die Bayeschen Beliefnetze brachten eine wesentliche Erleichterung bezüglich der Speicherintensivität als auch der praktischen Anwendbarkeit. Eine graphische Repräsentation von Daten ist von immenser Wichtigkeit dafür, daß der Fachexperte sich traut mit einem wissensbasierten System umzugehen und die Vorgänge darin nicht unbedingt zu verstehen, aber zumindest in Ansätzen nachzuvollziehen. Die bayeschen Netze bieten daher auch dem der Informatik fremden Fachexperten bequemen Zugang zu modernen IT-Konzepten.

Auf dem Markt existieren zur Zeit verschiedene auf bayesschen Beliefnetzen basierende Systeme. Das System MUNIN (S. dazu [AJJ⁺89]) ist ein System, daß in er Elektromyographie zur Diagnose von Muskel- und Nervenkrankheiten eingesetzt wird. HUGIN ist eine allgemeine Shell zum Aufbau von bayesschen Beliefnetzen und zum Rechnen darin.

Bayesche Beliefnetze sind zwar zur medizinischen Diagnose geeignet können, aber auch in anderen Diagnosebereichen eingesetzt werden, da Diagnose ein allgemeiner und nicht nur in der Medizin verwendbarer Begriff ist. Direkt bietet sich hier zum Beispiel die Diagnose von Maschinenschäden, Fehlern in Netzwerken, in Arbeitsabläufen in Fabriken und ähnliches an. Da eine Diagnose auch eine Art von Prognose darstellt, kann man Beliefnetze sogar zu Aktienkursprognosen und Prognosen der Effizienz von Arbeitsabläufen einsetzen. Gerade in der Wirtschaft, wenn es um Märkte, Kunden, Lieferanten und die Beziehungen zwischen ihnen geht, muß man immer mit vagen und

unsicheren Informationen arbeiten. Daher bieten sich Beliefnetze gerade zu Unterstützung betrieblicher und betriebswirtschaftlicher Entscheidungsfindungsprozesse an.

Die Faktorenanalyse war ein weiterer Schwerpunkt unserer Arbeit. Die Idee in Variablen, die sich überschneiden, mehrfach vorhandene Informationen zu entfernen, in den man Faktoren einführt, mit denen man auch mehr als nur die mehrfach vorhandene Information reduzieren kann, scheint zu nächst recht plausibel. So manches mal kann damit in der praktischen Anwendung ein erhebliche Reduktion von Datenmengen erreicht werden. In Betriebswirtschaftlichen Anwendungsbereichen fallen viele Zufallsvariablen an. Die Faktorenanalyse kann hier dazu dienen, die wesentlichen Faktoren daraus zu extrahieren und gezielt diese zu optimieren. Betrachtet man Wirtschaftlichkeit als eine zielgerichtete, planmäßige Handlungsweise, ist die Faktorenanalyse dafür von Bedeutung, da man begrenzte Mittel gezielt einsetzen kann und damit die Ergiebigkeit einer Investition im Sinne einer Ausgabe zur Steigerung des wirtschaftlichen Potentials eines Betriebs steigert.

Die Methoden der Faktorenanalyse bedürfen jedoch einer sorgfältigen Auswahl. Es existiert eine Vielzahl faktoranalytischer Verfahren, die alle zu unterschiedlichen Ergebnissen führen können und dies meistens auch tun. Durch das oft willkürliche Eingreifen des Anwenders in den Prozeß der Faktorenanalyse wird diese hochgradig indeterminiert.

Aus diesen Grund ist die Faktorenanalyse mit äußerster Vorsicht zu genießen, auch dort wo es ein Ergebnis zu interpretieren ist, was gar, von mehreren Personen durchgeführt, zu widersprüchlichen Ergebnissen führen könnte.

Auch die Clusteranalyse kann in wirtschaftlichen Bereich eingesetzt werden. Ein einfaches Beispiel könnte die Kreditwürdigkeitsprüfung einer Person sein, die von einer Bank durchgeführt wird. Anhand von Merkmalen von wie Gehalt, Kontostand, Familienstand u.ä. kann der nach einem Kredit Fragende in die Gruppe würdiger oder in die Gruppe kreditunwürdiger Kunden eingeteilt werden. Ein weiteres Beispiel ist die Gruppierung von Aktien zu bestimmten Klassen und vieles mehr.

Die Clusteranalyse wird nicht ganz so stark wie die Faktorenanalyse urch den Anwender beeinflusst. Trotzdem führt die Auswahl verschiedener Metriken und Verfahren zu zum Teil höchst unterschiedlichen Ergebnissen. Die Wahl der Verfahren kann häufig nicht rational begründet werden und erfolgt intuitiv. Daher ist auch die Clusteranalyse bei nicht ausreichender Ergebnisinterpretation vorsichtig zu behandeln.

Literaturverzeichnis

- [Aig96] Martin Aigner. *Diskrete Mathematik*. Vieweg, Braunschweig, Wiesbaden, 2., durchgesehene Auflage edition, 1996.
- [AJJ⁺89] S. Andreassen, F.V. Jensen, S.K. Jensen, et al. MUNIN-An Expert EMG Assistant, 1989.
- [ALS] W. Albath, H. Lommel, and M. F. Schloss. *Vademecum der medizinischen Laboratoriumsdiagnostik*. Bonn.
- [BDH⁺54a] M. Broglie, H. Dennig, K. Hansen, et al. *Lehrbuch der inneren Medizin*, volume Band 1. Georg Thieme Verlag, Stuttgart, 3. verbesserte Auflage edition, 1954.
- [BDH⁺54b] M. Broglie, H. Dennig, K. Hansen, et al. *Lehrbuch der inneren Medizin*, volume Band 2. Georg Thieme Verlag, Stuttgart, 3. verbesserte Auflage edition, 1954.
- [BEPW90] K. Backhaus, B. Erichson, W. Plinke, and R. Weiber. *Multivariate Analysemethoden*. Springer, Berlin, Heidelberg, 6. Auflage edition, 1990.
- [BFGK97] Craig Boutilier, Nir Friedman, Moises Goldszmidt, and Daphne Koller. Context-Specific Independence in Bayesian networks. In *Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence (UAI-96)*, Portland (Oregon, USA), August 1997.
- [Boc74] H. H. Bock. *Automatische Klassifikation*. Vandenhoeck und Ruprecht, Göttingen, 1974.
- [Bos95] K. Bosch. *Elementare Einführung in die Wahrscheinlichkeitsrechnung*. Vieweg Verlag, Braunschweig/Wiesbaden, 6. auflage edition, 1995.
- [BWW97] B. Brandherm, D. Wagner, and F. Wittig. Propagierungsalgorithmen in mehrfach verbundenen Bayesschen Netzen. Techni-

- cal report, Fachbereich Informatik, Universität des Saarlandes, Saarbrücken, 1997. Lehrstuhl Prof. Dr. W. Wahlster.
- [CTC89] H. Cheaib Tayssir and M. Cord. *Multivariate Statistik in den Natur- und Verhaltenswissenschaften*. Vieweg, Braunschweig, 1989.
- [DE92] H. Dallmann and K. H. Elster. *Einführung in die höhere Mathematik*, volume Band III. Gustav Fischer Verlag, Jena, 1992.
- [Dud93] Duden Informatik, ein Sachlexikon für Studium und Praxis, 1993.
- [EMT95] Engel, Möhring, and Troitzsch. *Multivariate statistische Verfahren*. BI Wissenschaftsverlag, 1995.
- [FH84] L. Fahrmeier and A. Hamerle. *Multivariate statistische Verfahren*. de Gruyter, Berlin, New York, 1984.
- [Fra97] U. Frank. Modellierung betrieblicher informationssysteme, 1997. Foliensatz zu gleichnamigen Vorlesung.
- [FSS86] Wolf-Günter Felmy, August Schmid, and Jörg Start. *Stochastik Grundkurs*. LS Mathematik. Ernst Klett Schulbuchverlag, Stuttgart, 1. Auflage edition, 1986.
- [Gra95] Adolf Grauel. *Fuzzy-Logik: Einführung in die Grundlagen mit Anwendungen*. BI Wissenschaftsverlag, Mannheim, Leipzig, Wien, Zürich, 1995.
- [HE84] J. Hartung and B. Elpelt. *Multivariate Statistik*. Oldenbourg, München, 1984.
- [Hec95] David Heckerman. A Tutorial on Learning with Bayesian Networks. Technical Report, 1995.
- [ICD93] Internationale Klassifikation der Krankheiten (ICD) 1979, März 1993.
- [KBPS95] Bernhard Kipper, Thorsten Brants, Marcus Plach, and Ralph Schäfer. Bayessche Netze: Ein einführendes Beispiel. Bericht des Graduiertenkolleg Kognitionswissenschaft 4, Universität des Saarlandes, Saarbrücken, Germany, 1995.
- [KD59] H. F. Kaiser and K. W. Dickman. Analytic determination of common factors. *American Psychologist*, 14, 1959.
- [KKKK89] I. Knoche, N. Knoche, M. Krallmann, and B. Krallmann. *Duden Abiturhilfen Stochastik I*. Dudenverlag, Mannheim, Wien, Zürich, 1989.

- [Kle80] Gernot D. Kleiter. *Bayes-Statistik: Grundlagen und Anwendungen*. de Gruyter, Berlin, New York, 1980.
- [KRW⁺99] Charles E. Jr. Kahn, Linda M. Roberts, Kung Wang, Deb Jenks, and Peter Haddaway. Preliminary Investigations of a Bayesian Network for Mammographic Diagnosis of Breast Cancer, 1999.
- [KSH91] R. Kruse, E. Schwecke, and J. Heinsohn. *Uncertainty and Vagueness in Knowledge Based Systems*. Springer, Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest, 1991.
- [KSV96] W. Köhler, G. Schachtel, and P. Voleske. *Biostatistik*. Springer, Berlin, Heidelberg, 2. Auflage edition, 1996.
- [LB82] A. Linder and W. Berthold. *Statistische Methoden III*. Birkhäuser, Basel, Boston, Stuttgart, 1982.
- [Lin73] Lingen Lexikon in 20 Bänden. Druck und buchbinderische Verarbeitung durch R. Oldenbourg München, August 1973. Erarbeitet nach den Unterlagen der Lexikon-Redaktion des Verlags F.A. Brockhaus Wiesbaden.
- [LR88] Josef Lauter and Karlheinz Rüdiger. *Mathematik Sekundarstufe II, Wahrscheinlichkeitsrechnung und Statistik*. Cornelsen Verlag Schwann-Girardet, Düsseldorf, 6. Auflage edition, 1988.
- [LW92] J. Lehn and H. Wegmann. *Einführung in die Statistik*. Teubner Studienbücher Mathematik. Teubner, 2. überarbeitete Auflage edition, 1992.
- [Moo78] Helfried Moosbrugger. *Multivariate statistische Analyseverfahren*. W. Kohlhammer, Berlin, Köln, Mainz, Stuttgart, 1978.
- [MS62] Friedrich Müller and Otto Seifert. *Taschenbuch der medizinisch-klinischen Diagnostik*. Verlag von J. F. Bergmann, München, 1962.
- [NBES82] H. Neunzert, A. Blickensdoerfer, W.G. Eschmann, and K. Schelkes. *Analysis II: Eine Einführung in die Vektor- und Matrizenrechnung*. Springer, Berlin, Heidelberg, New York, 2. Auflage edition, 1982. aus der Serie Mathematik für Physiker und Ingenieure.
- [Pea88] Judea Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. The Morgan Kaufmann Series in Representation and Reasoning. Morgan Kaufmann Publishers Inc., San Mateo/California, 1. Auflage edition, 1988.

- [Pea90] Judea Pearl. Bayesian Decision Methods. In Michael B. Morgan, editor, *Readings in Uncertain Reasoning*, pages 345–352. Morgan Kaufmann, San Mateo/California, 1990.
- [Pit75] Pitz. Bayes theorem: can a theory of judgment and inference do without it. *Cognitive Theory*, Vol. 1:131–148, 1975.
- [Psc64] Willibald Pschyrembel. *Klinisches Wörterbuch mit klinischen Syndromen*. Verlag Walter de Gruyter & Co., Berlin, 154. bis 184. Auflage edition, 1964.
- [Rae80] O. Raestrup. *Leitfaden der Lebensversicherungsmedizin*. Verlag Versicherungswirtschaft e.V., Karlsruhe, 1980.
- [Sie86] H. Sieber. *Mathematische Tafeln mit Formelsammlung E*. Ernst Klett Verlag, Stuttgart, 1986. in Formelsammlung E.
- [TS89] N. H. C. Thuy and P. Schupp. *Wissensverarbeitung und Expertensysteme*. Oldenbourg, München, Wien, 1. Auflage edition, 1989. herausgegeben in der Reihe Handbuch der Informatik.
- [Vie97] Reinhard Viertl. *Einführung in die Stochastik mit Elementen der Bayes-Statistik und Ansätzen für die Analyse unscharfer Daten*. Springer, Wien, New York, 2., überarbeitete Auflage edition, 1997.
- [vL35] Carl van Laak. *Veränderungen des weißen Blutbildes nach sportlichen Leistungen*. Inaugural-dissertation zur Erlangung der Doktorwürde der hohen medizinischen Fakultät der westfälischen Wilhelms-Universität zu Münster (Westf.), Hygienisches Institut der Westfälischen Wilhelms-Universität zu Münster in Westfalen, Münster, 1935.
- [vL98] B. van Laak. *Multivariate Datenanalyse, 1998*. Seminararbeit Ausgewählte Methoden der empirischen Wirtschafts- und Sozialforschung.
- [Wer92] Jürgen Werner. *Biomathematik und medizinische Statistik*. Verlag Urban und Schwarzenberg, München, Wien, Baltimore, 2. überarbeitete auflage edition, 1992.
- [Wic90] Dieter Wickmann. *Bayes-Statistik. Einsicht gewinnen und entscheiden bei Unsicherheit*, volume Band 4 of *Mathematische Texte*. BI Wissenschaftsverlag, Mannheim, Wien, Zürich, 1. auflage edition, 1990.
- [Wla98] A. Wlachos. *Wissensbasierte Systeme im Produktionsbereich (Einsatzgebiete und Erfolgspotentiale, 1998)*. Seminararbeit zum Thema Computer Inregrated Manufacturing.