



Auskunfts- und entscheidungsunterstützende Systeme (Teil 3)

von
Wolfgang Giere
Zentrum der Medizinischen Informatik
Klinikum der J.W.Goethe-Universität



Ziel der dritten Vorlesung

- Wiederholung EUS-Schema
- Weiterführung des Kapitels „Expertensysteme“
- Erläuterung der Symptom/Diagnose-Matrix

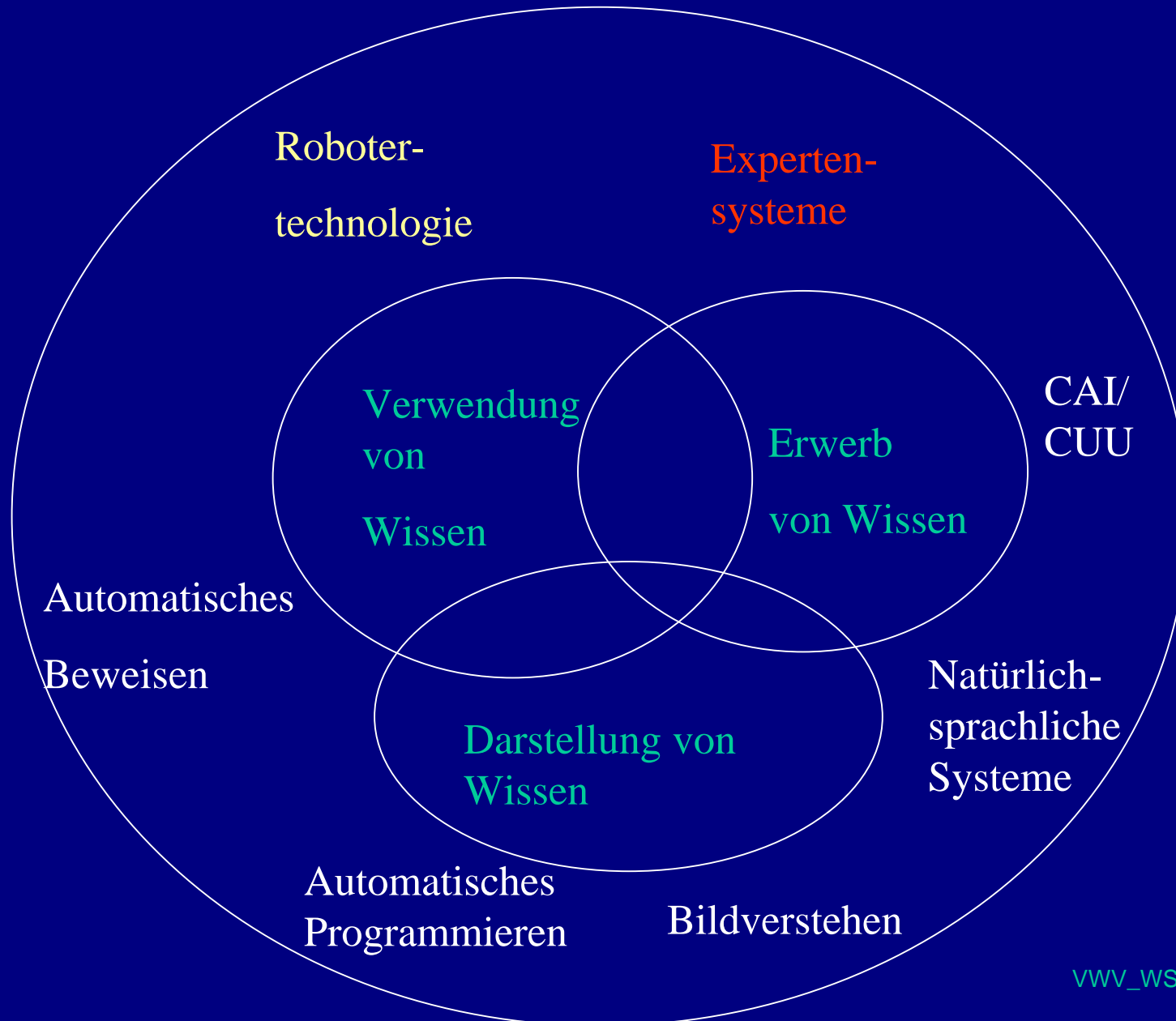
Erinnern Sie sich an das generelle Ziel:
Lieferung von Beurteilungskriterien

Sie sollen sein:

- zeitlos
- grundsätzlich
- erschöpfend

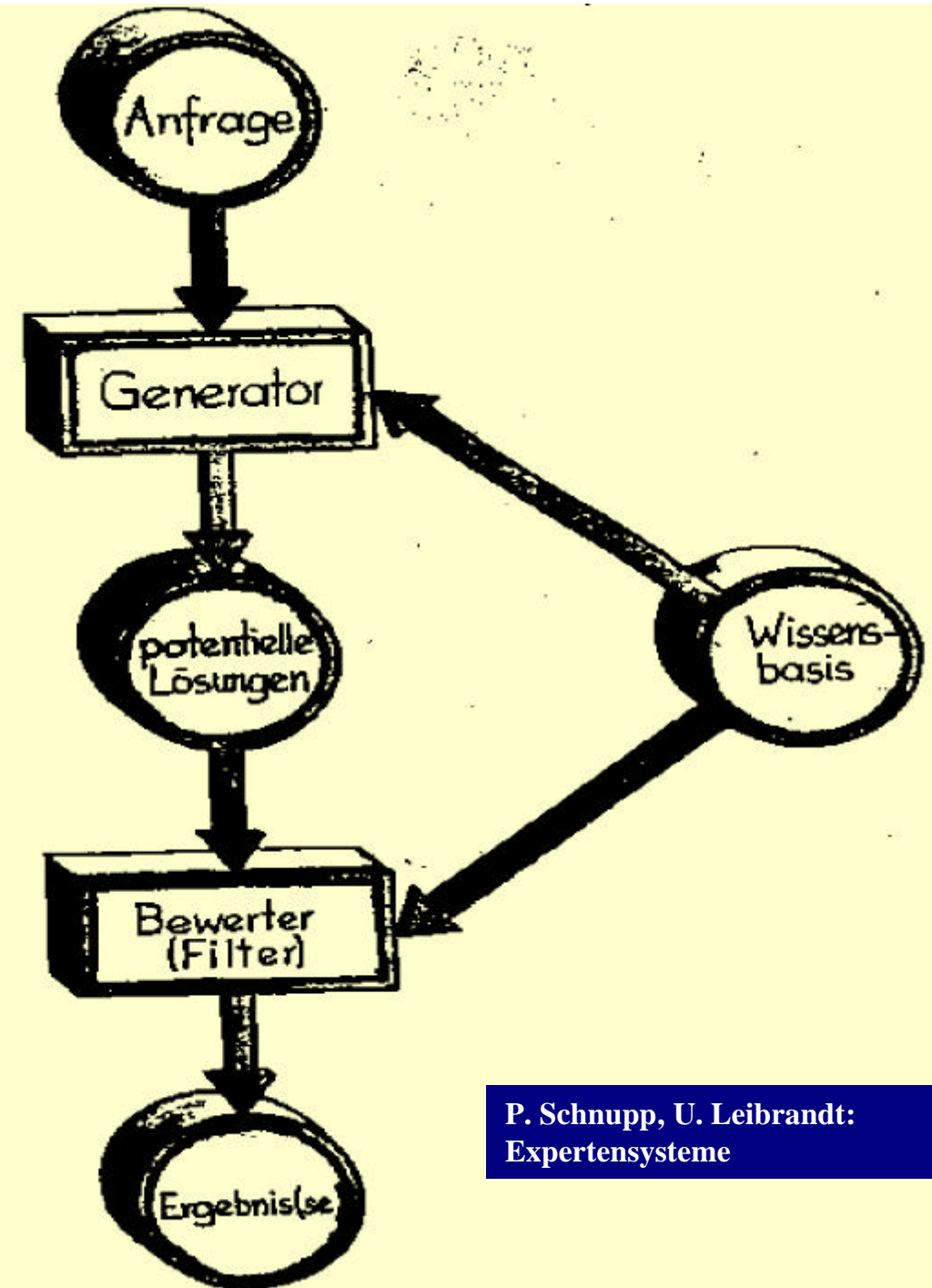


Szenario der künstlichen Intelligenz Roboter



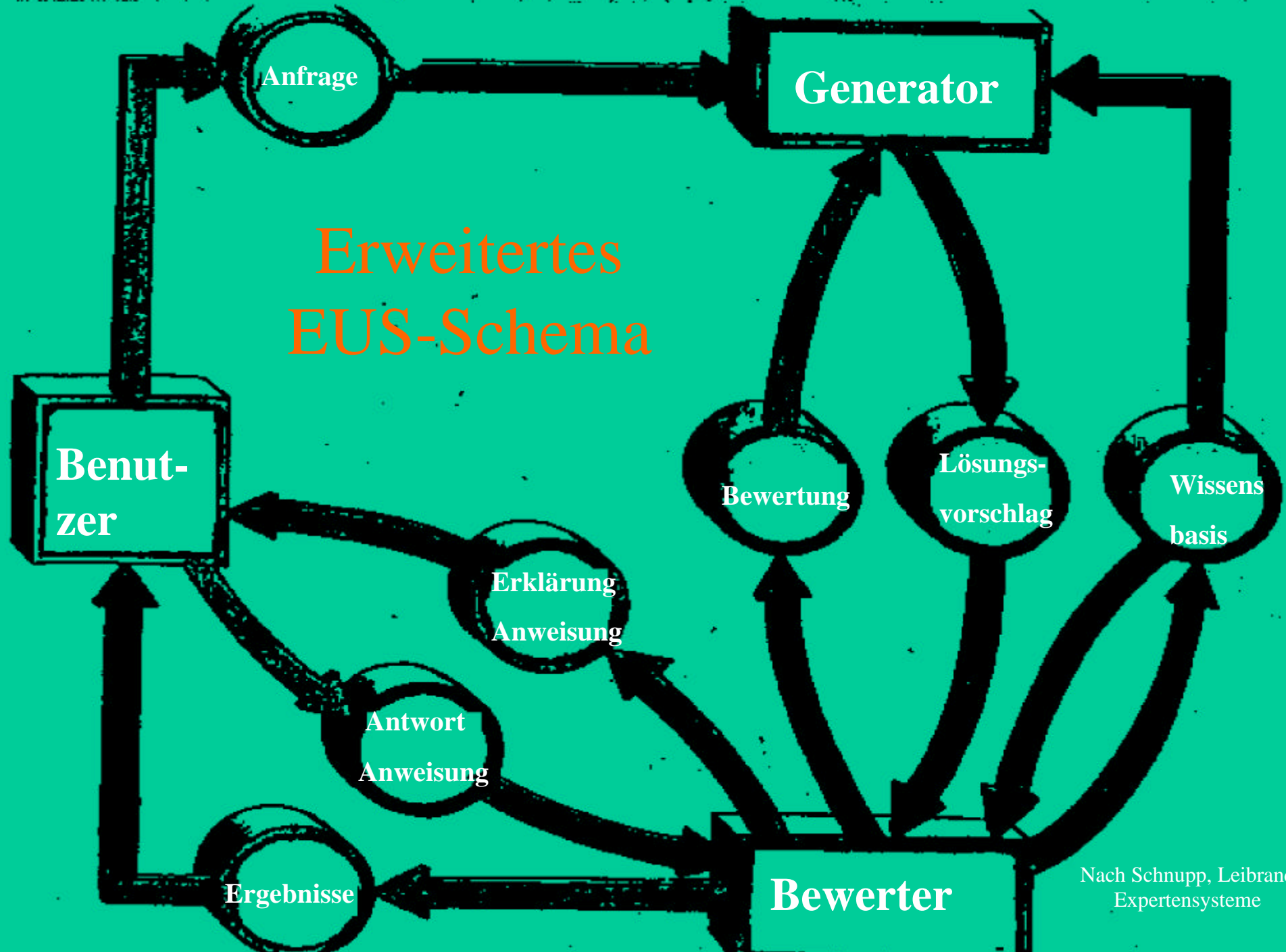


EUS-Schema



P. Schnupp, U. Leibbrandt:
Expertensysteme

Erweitertes EUS-Schema



Nach Schnupp, Leibrand
Expertensysteme



Historie - global

Mitte der sechziger Jahre, d.h. mit Beginn des computings,
zahlreiche Versuche in den USA, z.B.

- Pipberger: EKG
- Ledley, Lusted: Bayes Theorem / Hämatologie

Neben den frühen Versuchen mit neuronalen Netzen
(Perceptron Modell) von Anfang an Benutzung von
Matrizen bzw. Vektoren:



Prinzip der Entscheidungsunterstützung: Symptom/Diagnose-Matrix

Dx Sy	S1	S2	S3	Sn	Summe/Dx
D1						
D2						
...						
Dn						
Summe/Sy						



Symptom-Diagnose-Matrix

enthält

- Zeilenwerte (Merkmale der Diagnosen)
- Spaltenwerte (Merkmale der Symptome)
- Zellenwerte (Merkmale der Kombination)



Prinzip der Entscheidungsunterstützung: Symptom/Diagnose-Matrix (Symptome)

Dx Sy	S1	S2	S3	Sn	Summe/Dx
D1	x		x		x	S(1,3,n)
D2	x	x				S(1,2)
...						
Dn		x	x			S(2,3)
Summe/Sy						



Symptomvektoren

Zu jeder Diagnose wird eine Liste aller Symptome gebildet, die bei der Diagnose vorkommen (Symptomvektor)

Diese Liste enthält nur **positive Symptome**.

Anmerkung: Es gab Experimente auch mit negativen Vektoren, d.h. Listen von Symptomen, welche die Diagnose ausschließen. Sie haben sich nicht bewährt.



Prinzip der Entscheidungsunterstützung: Symptom/Diagnose-Matrix (Diagnosen)

Dx\Sy	S1	S2	S3	Sn	Summe/Dx
D1	x		x		x	
D2	x	x				
...						
Dn		x	x			
Summe/Sy	D(1,2)	D(2,3)	D(1,3)		D(1)	



Diagnosevektoren

Zu jedem Symptom wird eine Liste aller Diagnosen gebildet, bei der dieses Symptom vorkommt (Diagnosevektor)

Diese Liste enthält nur **positive Diagnosen**.

Anmerkung: Es gab Experimente auch mit negativen Vektoren, d.h. Listen von Diagnosen, welche bei Vorliegen dieses Symptoms ausgeschlossen sind. Sie haben sich nicht bewährt.



Prinzip der Entscheidungsunterstützung: Symptom/Diagnose-Matrix (Boole)

Dx\Sy	S1	S2	S3	Sn	Summe/Dx
D1	x		x		x	S(1,3,n)
D2	x	x				S(1,2)
...						
Dn		x	x			S(2,3)
Summe/Sy	D(1,2)	D(2,3)	D(1,3)		D(1)	



Symptom/Diagnose-Matrix

- (1) Zu jeder Diagnose Liste aller Symptome, die bei der Diagnose vorkommen (Symptomvektor)
- (2) Zu jedem Symptom Liste aller Diagnosen, bei denen das Symptom vorkommt (Diagnosevektor)

Anmerkung: Die Listen stellen unterschiedliche Aufbereitungen (Invertierung) derselben Primärinformation dar und können algorithmisch gewonnen und auf Plausibilität geprüft werden

Diese Matrix ist unverzichtbar für E-U-Systeme. In Praxi wird sie meist ergänzt um weitere Angaben:



Entscheidungsunterstützung deterministisch

Grundlagen:

- Logische Sy/Dx Beziehungen (kommt vor/nicht vor)
- Regeln (wenn das und nicht das, dann das oder ...)
- Quelle z.B. Literatur

Mögliche Antworten:

- kommt in Frage
- kommt nicht in Frage

Keine Bewertung!



Deterministische Sy/Dx-Beziehungen

Grabner (Wien) hat die möglichen logischen Beziehungen zwischen Symptomen und Diagnosen untersucht:

- Notwendig und beweisend (Beispiel: Bubo/Pest)
- Notwendig, nicht beweisend (Beispiel: Fieber/Malaria)
- Fakultativ, beweisend (Beispiel: Koplik-Flecken/Masern)
- Fakultativ, nicht beweisend (Beispiel Fieber/Magen-Ca)
- dito auch mit Negationen: Dx-ausschließende Symptome

Problem: Fast alle Beziehungen sind fakultativ, nicht beweisend



Probleme mit der Deterministik

- Ja/Nein-Beziehungen sind schwierig
- Regeln sind mühsam zu erstellen und zu pflegen, oft nicht eindeutig
- Deswegen hat man sich früh auch mit semiquantitativen und/oder probabilistischen Systemen beschäftigt
- Sie basieren auf Berechnungen, Schätzungen, Statistik bzw. Wahrscheinlichkeiten
- Zunächst zu Größen, die sich aus der Literatur berechnen lassen



Entscheidungsunterstützung von Literatur, semi-probabilistisch

Zählung der Nennung von Symptomen bei Diagnosen:

- bei vielen Diagnosen genanntes Symptom trägt wenig zur Diagnose bei
- bei wenigen Diagnosen genanntes Symptom ist spezifischer
- Maß hierfür ist die „Reziproke Häufigkeit“ (MEDIUC) oder Spezifität ...
- Benutzt auch von Sc. BLOIS für das System **Reconsider**, das auf automatisch erfaßter Literatur (CMIT) basiert.



Semi-Probabilistik: Maximum Match

- Ein weiteres Verfahren, aus der Ja/Nein Beziehung eine quantitative abzuleiten, hat sich weitgehend durchgesetzt:
- Zählung und Vergleich der „gefundenen Symptome“.
- Hypothese: Die Diagnose, die mehr Symptome erklärt, als eine andere, ist „richtiger“.
- Verfeinerung durch Ausgabe der „nicht gefundenen Symptome“, also der Symptome, die bei der infrage kommenden Diagnose nicht vorkommen



Prinzip der Entscheidungsunterstützung: Symptom/Diagnose-Matrix (Boole)

Dx\Sy	S1	S2	S3	Sn	Summe/Dx
D1	x		x		x	S(1,3,n)
D2	x	x				S(1,2)
...						
Dn		x	x			S(2,3)
Summe/Sy	D(1,2)	D(2,3)	D(1,3)		D(1)	



Entscheidungsunterstützung deterministisch

Grundlagen:

- Logische Sy/Dx Beziehungen (kommt vor/nicht vor)
- Regeln (wenn das und nicht das, dann das oder ...)
- **Literatur**

Mögliche Antworten:

- kommt in Frage
- kommt nicht in Frage

Keine Bewertung!



Semiquantitative Verfahren (Wiederholung)

- Reziproke Häufigkeit (Spezifität)
- Trefferrate (Maximum Match)



Probleme mit der Deterministik (Wdhlg)

- Ja/Nein-Beziehungen sind schwierig
- Regeln sind mühsam zu erstellen und zu pflegen, oft nicht eindeutig
- Deswegen hat man sich früh auch mit semiquantitativen und/oder probabilistischen Systemen beschäftigt
- Sie basieren auf Berechnungen, Schätzungen, Statistik bzw. Wahrscheinlichkeiten
- Nun zu Größen, die sich **nicht** aus der Literatur berechnen lassen ...



Probabilistik grundsätzlich

- Kann in der Regel nicht aus der Literatur gewonnen werden
- Erfordert die Einbeziehung bzw. Auswertung echter Krankenakten
- automatische Einbeziehung führt zu „selbstlernendem“ System - Problem: Verifikation und Freigabe
- Echte Statistik zur Verfügung zu stellen, ist Nationale Aufgabe, bisher in Deutschland nicht gelöst

Was kann man messen?

Zurück zur Sy/Dx-Matrix:



Symptom/Diagnose-Matrix

Dx-Zähler, „Z1“ (Zeilenwert)

Dx\Sy	S1	S2	S3	Sn	Summe/Dx
D1: 120						
D2: 935						
...						
Dn: 12						
Summe/Sy						



„Z1“ Diagnosenhäufigkeit (a priori)

Welche Quellen gibt es?

- Gesundheitsstatistik
- Krankenhausabrechnung
- KV-Abrechnung
- Tumorzentren (Kinder, DOESAK)
- Spezialregister (z.B. Diabetes, Rheuma in USA)
- Krankengeschichten



Inzidenz und Praevalenz

- Inzidenz
gibt die Zahl der NEU-Erkrankungen an
- Praevalenz
gibt den Kranken-BESTAND an

Unterschiede naturgemäß groß
bei chronischen Erkrankungen



Probleme der Diagnose-Einteilung

- Syndrombenennung
- Erkrankungen ohne Erregernachweis („Grippe“)
- Teildiagnose („Lungenstauung“, „Nierenversagen“)
- usw.

Die Abgrenzung ist oft sehr schwierig.



Symptom/Diagnose-Matrix

Sy-Zähler, „Z2“ (Spaltenwert)

Dx\Sy	S1	S2	S3	Sn	Summe/Dx
	215	20	75		112	
D1						
D2						
...						
Dn						
Summe/Sy						



„Z2“ Symptomhäufigkeit (a priori)

Quellen

- Bücher?
(„häufig“, „eher selten“, „gelegentlich“)
- Statistik?
(mir nicht bekannt, außer für Expertensysteme)
- Krankengeschichtsauswertung

Problem: Was ist ein Symptom?



Zum Begriff „Symptom“ (1)

Wurzel griechisch $\sigma\upsilon\mu\pi\tau\omicron\mu\alpha$:

- Begleiterscheinung
- Beschwerde
- faßbares Krankheitszeichen

Man unterscheidet (unter anderem)

- Subjektive Angaben des Patienten
- Objektive Befunde
- Testergebnisse
- englisch: Symptoms, Signs and Tests



Zum Begriff „Symptom“ (2)

Probleme bei der Symptomklassifikation:

- quantitative Angaben
(z.B. Leukozytose, Hochdruck)
- komplexe Erscheinungen
(z.B. „Kater“, „Schock“)
- „Teildiagnosen“
(z.B. Lungenentzündung, Nierenversagen)

Anmerkung: Die Übergänge sind fließend, die Definitionen abhängig vom Ziel der Einteilung



Symptom/Diagnose-Matrix

Zellen-Zähler, „Z3“ (Zellenwert)

Dx\Sy	S1	S2	S3	Sn	Summe/Dx
D1	99	10	90		95	
D2	90	95	01		10	
...						
Dn	10	95	97		15	
Summe/Sy						



„Z3“ Zellenwert

Häufigkeit der Symptom/Diagnose-Beziehung

- Literatur (lyrisch, unexakt)
- Krankengeschichten (Problem des Bias der Selektion)
- gezielte Studien

Problem:

- Sehr aufwendige Ermittlung,
- nur für wenige Krankheitsbilder verfügbar,
- gerade für die seltenen
(und deswegen interessierenden) nicht



Symptom/Diagnose-Matrix

Z1, Z2, Z3 (komplett)

Dx\Sy	S1	S2	S3	Sn	Summe/Dx
	215	20	75		112	
D1: 120	99	10	90		95	
D2: 935	90	95	01		10	
...						
Dn: 12	10	95	97		15	
Summe/Sy						



Theorem von Bayes (1763)

- Late Sir Reverend Bayes war Pfarrer in England.
- Sein berühmtes Theorem der bedingten Wahrscheinlichkeiten hat er in einem Brief an die Royal Society formuliert, der posthum veröffentlicht wurde
- Es wurde für die computerunterstützte Diagnostik Mitte der sechziger Jahre wiederentdeckt von Ledley und Lusted. Seitdem ist es Basis für alle probabilistischen Ansätze

Einschub: MO/Bayes



Probleme mit der Probabilistik

- Echte Statistiken sind selten
- Gute Statistiken gibt es nur für häufige Erkrankungen
- Für häufige Erkrankungen braucht man kein EUS

Deswegen hat sich der „echte Bayes“ als alleinige
Bewertungsgrundlage nur in einem einzigen System
durchgesetzt: de Dombals Acute Abdominal Pain

Erfolgreichere Systeme simulieren die Statistik durch
Expertenschätzung für die Zellenwerte
(„geschätztes Gewicht“, „expert rating“, „evoking strenght“)



Ende von Teil 3.
Ich bedanke mich
für Ihre Aufmerksamkeit!
(Es folgt Teil 4)