Einführung in die KI

Prof. Dr. sc. Hans-Dieter Burkhard
Vorlesung Winter-Semester 2003/04

Ungenaues/Unsicheres/Unscharfes Wissen
Probabilistisches Schließen
Fuzzy-Theorie

Ungenaues/Unsicheres/Unscharfes Wissen

- Ungenau (impräzise):
  - genau (präzise): *Blau.* *39,635 km.*
  - ungenau (impräzise): *Blau oder grün.* *Zwischen 39 und 40 km.*
- Unsicher
  - sicher: *Bestimmt blau.* *Ganz sicher 20 km.*
  - unsicher: *Möglicherweise blau.* *Vielleicht 50 km.*
- Unscharf (vage)
  - scharf: *RGB=[20,5,230].* *130 277 km.*
  - unscharf (vage): *Blau.* *Ziemlich weit entfernt.*

Ungenaues/Unsicheres/Unscharfes Wissen

3 mögliche Dimensionen bei Aussagen:

Genauigkeit - Sicherheit – Schärfe

Auch in Kombinationen:

- *„vielleicht blau oder grün“*
- *„wahrscheinlich nicht teuer“*

Weitere Begriffe:

- Zuverlässigkeit
- Vertrauenswürdigkeit
- Vermutung

Maßzahlen oft in Prozent:

„70%ig vertrauenswürdig“

„99%ig sicher“

Ungenaues/Unsicheres/Unscharfes Wissen

Formalismen:

- Klassische Logik
  - wahr – falsch , {0,1}
- Mehrwertige Logiken
  - Werte zwischen 0 und 1 , z.B. {0, 1/2, 1} oder [0,1]
- Modale Logik:
  - unmöglich - möglich - sicher (notwendig)
- Wahrscheinlichkeitsrechnung
  - *in 30% aller Fälle.*
  - *mit 90%er Sicherheit.*
- Fuzzy-Logik
  - *ziemlich richtig.*
  - *halbwahr.*
  - *stimmt zu 60%.*

Ungenaues Wissen

- genau (präzise): *„Blau.“* *„39,635 km.“*
- ungenau (impräzise):

  „Blau oder grün.“ *„Zwischen 39 und 40 km.“*

Ungenaues Wissen beschreibbar durch Angabe aller Möglichkeiten, z.B.:

– Alternative: Farbe = blau $\vee$ Farbe = grün

– Intervall: Entfernung $\in$ [39,40]

– Vergleichsprädikate: Entfernung $\geq$ 39 $\wedge$ Entfernung $\leq$ 40

Behandlung mit Mitteln einer genauen Beschreibung.

Unsicherheit, Möglichkeit, Vertrauenswürdigkeit, ...

Unsicheres Wissen

- sicher: *„Bestimmt blau.“* *„Ganz sicher 20 km.“*
- unsicher: *„Möglicherweise blau.“* *„Vielleicht 50 km.“*

Unterschied zu „ungenau“ :
Nicht alle Möglichkeiten angegeben.

Formalisierungen:

- Modale Logik

  Unmöglich - möglich - sicher (notwendig)

- Quantifizierbare Möglichkeits-/Vertrauensgrade, Wahrscheinlichkeiten

  – Wahrscheinlichkeit eines Gewinns

  – *„Regen mit 75%er Wahrscheinlichkeit“*

Unscharfes Wissen

Unschärfe, Vagheit, Unexaktheit

- scharf: *„[0,0,255].“* *„130 277 km.“*
- unscharf (vage): *„Blau.“* *„Ziemlich weit entfernt.“*

Formalismen mit scharfen Begriffen:
Klassische Logik: „wahr“/„falsch“.
Mengezugehörigkeit: „ja“/„nein“
Wahrscheinlichkeit: wie wahrscheinlich ist „Aussage H ist wahr“

Problem scharfer Grenzen:
- nah = 0...99 ?
- weit = 100...9999 ?
- sehr weit = 10000... ?

Wieviele Streichhölzer bilden einen Haufen?

Behandlung von Unsicherheit

Form des Umgangs mit unvollständigem Wissen
- Beschränkung auf bekannte Einflüsse
- Beschränkung auf (meistens) relevante Einflüsse
  - Effizienz
  - Rationalität
- Vergleich von Varianten

Beispiel:
Medizinische Diagnose

Genauer betrachten:
- Konsequenzen aus unvollständigem Wissen
- Inkonsistente Verwendung im Alltag
- Probleme mit Statistiken

Problematik subjektiver Angaben

Linda ist 31 Jahre alt, sie lebt allein, ist offen, sehr begabt und Magister der Philosophie. Während ihres Studiums hat sie sich intensiv mit Fragen sozialer Gerechtigkeit und Diskriminierung befasst und auch an Anti-Atom-Demonstrationen teilgenommen.

Wie glaubhaft (wie wahrscheinlich) sind die folgenden Aussagen:

1. Linda ist Kassiererin bei einer Bank.
2. Linda hat zwei Kinder.
3. Linda ist Kassiererin bei einer Bank und aktiv in der Frauenbewegung.
4. Linda ist geschieden.

Problematik subjektiver Angaben

Test von 2 Medikamenten A und B für Männer bzw. Frauen.

A M-besser als B bedeutet:
Für Männer gilt: Die prozentuale Erfolgsquote von A ist grösser als die prozentuale Erfolgsquote von B.

A F-besser als B bedeutet:
Für Frauen gilt: Die prozentuale Erfolgsquote von A ist grösser als die prozentuale Erfolgsquote von B.

A besser als B bedeutet:
Insgesamt gilt: Die prozentuale Erfolgsquote von A ist grösser als die prozentuale Erfolgsquote von B.

Gilt die folgende Regel?

Wenn A M-besser als B und A F-besser als B , so A besser als B .

Problematik subjektiver Angaben

„Simpsons Paradox“

(Beispiel aus Richter: „Prinzipien der KI“, S. 198f)

	Medikament	Gesamt	Erfolg	Erfolg %
A M-besser als B :	A	200	20	**10**
	B	100	4	4
A F-besser als B :	A	40	20	**50**
	B	100	37	37
Nicht A besser als B:	A	240	40	16,6
	B	200	41	**20,5**

Schließen in unsicheren Bereichen

- nicht alle Regeln sind absolut gültig
- nicht alle Fakten sind absolut verlässlich

Resultat: Unsicherheit von Schlußfolgerungen

(Un-)Sicherheits-, Vertrauens-, Evidenz-Werte für

- Fakten
- Regeln

Der Patient zeigt mit 90% Sicherheit das Symptom S .

Bei Symptom S hat der Patient mit 85% Wahrscheinlichkeit die Krankheit K.

Die Erfolgs-Chancen vonTherapie T bei Krankheit K betragen 80% .

Was ist die Erfolgs-Wahrscheinlichkeit für Therapie T?

Beispiel: Experten-System MYCIN

- Kombination von „measure of belief“ und „measure of disbelief“ (bzgl. eines Symptoms, einer Krankheit, einer Therapie)
- Veränderungen als Folge entsprechender Regelanwendungen

Regeln haben z.B. Form:

Bei Symptomen A und B
ist mit 70% Wahrscheinlichkeit auf Krankheit K zu schließen.

Bedeutet nicht: ... 30% Wahrscheinlichkeit für „nicht K“.

Die „70%“ beschreiben keinen absoluten Vertrauensgrad, sondern den „Zuwachs an Vertrauen“ in die Hypothese.

Probleme bei Interpretation und Abarbeitung (siehe Literatur)

Schließen in unsicheren Bereichen

Der Patient zeigt mit 90% Sicherheit das Symptom S .

Bei Symptomen A und B
ist mit 70% Wahrscheinlichkeit auf Krankheit K zu schließen.

Die Erfolgs-Chancen von Therapie T bei Krankheit K
betragen 80% , sofern keine Allergie auftritt.

Problem:

Propagierung/Verknüpfung unsicherer Werte.

Addition, Multiplikation, Maximum, Minimum, . . . ?

Insgesamt 3 Quellen für unsichere Resultate:
Fakten, Regeln, Inferenz.

Wahrscheinlichkeiten

- Wahrscheinlichkeit, eine 6 zu würfeln: 16,7% bzw. 0,167
- Wahrscheinlichkeit für Niederschlag: 60 % bzw. 0,6

Zuverlässigkeit, Nützlichkeit, ...

Interpretationen:

- Statistische Deutung als Quotient:
  - Anzahl der günstigen Ereignisse / Anzahl aller Ereignisse
- Subjektive Unsicherheit
  - Maß für Überzeugtheit (Wettbereitschaft!)
- Physikalisch begründete Unsicherheit

Wahrscheinlichkeiten als Zahlen aus [0,1]

- Wahrscheinlichkeit $P(H)$ als Maßzahl für Sicherheit des Zutreffens einer Annahme H

Normierung: $P(H) + P(\neg H) = 1$

H ist wahr oder falsch (klass.Logik)

- Wahrscheinlichkeiten $P(X=x)$ als Maßzahl, dass eine Zufallsvariable X einen Wert $x \in W = \{ x_1,\ldots,x_n\}$ annimmt

Normierung: $P(x_1) + \ldots + P(x_n)=1$

- Wahrscheinlichkeiten $P(\mathbf{X}=\mathbf{x})$ als Maßzahl, dass ein Zufallsvektor $\mathbf{X} = [X^{(1)},\ldots,X^{(N)}]$ einen Wert $\mathbf{x} = [x^{(1)},\ldots,x^{(N)}] \in W^{(1)}\times\ldots\times W^{(N)}$ annimmt

Normierung: $\sum P(\mathbf{x}) = 1$

(hier: Diskrete Wahrscheinlichkeiten)

Wahrscheinlichkeiten

- Wahrscheinlichkeiten ermitteln anhand von
  - Häufigkeitsabschätzungen
  - statistischen Auswertungen
  - anderen Wahrscheinlichkeiten
    (Bayessche Formel, Bayessche Netze)
- Wahrscheinlichkeitsberechnung veränderlich bei
  - zusätzlicher Information
  - weiteren Ereignissen

3-Türen-Beispiel

Wahrscheinlichkeiten: Ereignis-Algebra

- Ω : Grundmenge der Elementareignisse (Ereignis-Raum)
- S : System von Ereignissen (Mengen) über Ω .

S heißt Ereignis-Algebra, falls

(1) S enthält Ω (das sichere Ereignis) und $\varnothing$ (das unmögliche Ereignis)

(2) Falls $A \in S$, so auch $\Omega - A \in S$.

(3) Falls $A, B \in S$, so auch $A \cap B \in S$, $A \cup B \in S$.

S heißt σ-Algebra, falls außerdem

(3´) Falls $A_1, A_2, A_3, \ldots \in S$, so auch $\bigcap A_i \in S$, $\bigcup A_i \in S$.

Kolmogorow-Axiome für Wahrscheinlichkeit P

S sei Ereignis-Algebra über endlichem Ereignis-Raum Ω

Axiom 1: P ist eine Funktion $P : S \rightarrow [0,1]$

Axiom 2: $P(\Omega) = 1$

Axiom 3: Für $A, B \in S$ mit $A \cap B = \emptyset$ gilt $P(A \cup B) = P(A)+P(B)$ (Additionsaxiom)

bzw. für σ-Algebra über unendlichen Ereignis-Raum Ω :

Axiom 3´:

Für $A_1, A_2, A_3, \ldots \in S$ mit $A_i \cap A_j = \emptyset$ für alle $i,j \in N$ mit $i \neq j$ gilt

$$P(\bigcup A_i) = \sum P(A_i)$$

Folgerungen aus Kolmogorow-Axiomen

Für beliebige A , $B \in S$ gilt:

- $P(\Omega - A) = 1 - P(A)$
- $P(\emptyset) = 0$
- $A \subseteq B \rightarrow P(A) \leq P(P)$
- $P(A - B) = P(A) - P(A \cap B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cup B) \leq P(A) + P(B)$

Für $A_1, \ldots, A_n \in S$ mit $A_i \cap A_j = \emptyset$ für alle $i \neq j$ gilt:

$$P(\cup A_i) = \Sigma P(A_i)$$

Zufallsgrößen

- Zufallsvariable X kann Werte $x \in W$ annehmen
- $P(x) = P(X = x)$: Wahrscheinlichkeit für „Ereignis $X = x$ "

  $P(x) \in [0,1]$ (0% ... 100 %)

- Verteilungsfunktion: $F(x) = P(X \leq x)$
  - für diskrete Zufallsgrößen: $F(x) = \sum_{x' \leq x} P(x')$
  - für stetige Zufallsgrößen: $F(x) = \int_{-\infty}^{x} f(u)du$

    $f(u)$ heißt Dichte der Verteilung

- Normierung:

  $\sum P(\mathbf{x}) = 1$ bzw. $\int_{-\infty}^{\infty} f(u)du = 1$

Bedingte Wahrscheinlichkeit

Wahrscheinlichkeit von A unter Bedingung, dass B eingetreten ist:

$P(A|B) = P(A \cap B) / P(B)$

- erfüllt Kolmogorow-Axiome
- Division durch $P(B)$ zur Normierung, (Voraussetzung $P(B) \neq 0$)

Produkt-Satz: $P(A \cap B) = P(B) \cdot P(A \mid B)$

$= P(A) \cdot P(B \mid A)$

A heißt (*stochastisch*) *unabhängig von* B (mit $0<P(B)<1$), falls

$P(A \mid B) = P(A)$

gleichwertig: $P(A \cap B) = P(A) \cdot P(B)$

wegen Symmetrie: „ A, B voneinander (*stochastisch*) *unabhängig* "

Bayesscher Satz/Bayessche Formel

Voraussetzung: $B_1 \cup \ldots \cup B_K = \Omega$ und $B_i \cap B_j = \emptyset$ für alle $i \neq j$

Satz über Totale Wahrscheinlichkeit:

$$P(A) = \Sigma_{k=1,\ldots,K}\, P(B_k) \cdot P(A \mid B_k)$$

Bayesscher Satz/Bayessche Formel:

$$P(B_k \mid A) = \frac{P(B_k) \cdot P(A \mid B_k)}{P(A)} = \frac{P(B_k) \cdot P(A \mid B_k)}{\Sigma_{k=1..K}\, P(B_k) \cdot P(A \mid B_k)}$$

Schließen auf „a-posteriori-Wahrsch.“

Satz von Bayes:

$$P(\text{Ursache} \mid \text{Symptom}) = \frac{P(\text{Ursache}) \cdot P(\text{Symptom} \mid \text{Ursache})}{P(\text{Symptom})}$$

Beispiel:

P (Bronchitis) = 0,05

P (Husten) = 0,2

P (Husten|Bronchitis) = 0,8

damit:

P (Bronchitis|Husten) = 0,2

Mehrere Symptome $S_1,\ldots,S_n$

$$P(U|\ S_1\&\ldots\&\ S_n) = \frac{P(U)\cdot P(S_1\&\ldots\&\ S_n\ |U\ )}{P(S_1\&\ldots\&\ S_n)}$$

Idealisierung:

Unabhängigkeit der Symptome *beim Auftreten von U*

$$P(S_1\&\ldots\&\ S_n\ |\ U) = P(S_1\ |\ U)\cdot\ \ldots\ \cdot P(\ S_n\ |\ U)$$

(nicht verlangt: $P(S_1\&\ldots\&\ S_n) = P(S_1)\cdot\ldots\cdot P(\ S_n)$, d.h. Symptome können untereinander abhängig sein)

Vereinfachter Bayesscher Satz:

$$P(U|\ S_1\&\ldots\&\ S_n) = \frac{P(U)\cdot P(S_1\ |\ U)\cdot\ \ldots\ \cdot P(\ S_n\ |\ U)}{P(S_1\&\ldots\&\ S_n)}$$

Schließen auf „a-posteriori-Wahrsch.“

Vereinfachter Bayesscher Satz:

$$P(U|\ S_1\&\ldots\&\ S_n) = \frac{P(U)\cdot P(S_1\ |\ U)\cdot\ \ldots\ \cdot P(\ S_n\ |\ U)}{P(S_1\&\ldots\&\ S_n)}$$

Die **a-posteriori-Wahrscheinlichkeit** $P(U|\ S_1\&\ldots\&\ S_n)$ für U nach Beobachten der Symptome $S_1\ldots S_n$ ist berechenbar aus

- der a-priori-Wahrscheinlichkeit $P(U)$,
- den Wahrscheinlichkeiten $P(S_i\ |\ U)$ für das Auftreten des Symptoms S_i bei der Ursache U
- der Wahrscheinlichkeit $P(S_1\&\ldots\&\ S_n)$.

Wahrscheinlichkeiten sind statistischen Messungen zu entnehmen.

Mehrere Ursachen (Diagnosen) U_i

Relative Wahrscheinlichkeiten:

$$P(U_i \mid S_1\&...\& S_n) = \frac{P(U_i) \cdot P(S_1 \mid U_i) \cdot ... \cdot P(S_n \mid U_i)}{\Sigma_{k=1..K} P(U_k) \cdot P(S_1 \mid U_k) \cdot ... \cdot P(S_n \mid U_k)}$$

Relative a-posteriori-Wahrscheinlichkeit $P(U_i \mid S_1\&...\& S_n)$ für U_i nach Beobachten der Symptome $S_1...S_n$ ist berechenbar aus

- a-priori-Wahrscheinlichkeiten $P(U_k)$ der möglichen Ursachen U_k,
- Wahrscheinlichkeiten $P(S_i \mid U_k)$ für das Auftreten der Symptome S_i bei der Ursache U_k .

Wahrscheinlichkeiten sind statistischen Messungen zu entnehmen.

Normalisierung

$$P(U \mid S_1\&...\& S_n) = \frac{P(U) \cdot P(S_1 \mid U) \cdot ... \cdot P(S_n \mid U)}{P(S_1\&...\& S_n)}$$

Die Nenner in den Formeln dienen jeweils der

$$P(U_i \mid S_1\&...\& S_n) = \frac{P(U_i) \cdot P(S_1 \mid U_i) \cdot ... \cdot P(S_n \mid U_i)}{\Sigma_{k=1..K} P(U_k) \cdot P(S_1 \mid U_k) \cdot ... \cdot P(S_n \mid U_k)}$$

Normierung auf den Wert 1 für Gesamtwahrscheinlichkeit.

Alternative Schreibweisen:

$$P(U \mid S_1\&...\& S_n) = \alpha \cdot P(U) \cdot P(S_1 \mid U) \cdot ... \cdot P(S_n \mid U)$$

$$P(U_i \mid S_1\&...\& S_n) = \beta \cdot P(U_i) \cdot P(S_1 \mid U_i) \cdot ... \cdot P(S_n \mid U_i)$$

mit Normalisierungsfaktoren α, β

α, β können nach Bestimmung der anderen Werte ermittelt werden (praktisches Vorgehen).

Anwendungen

$$P(U_i \mid S_1 \& \ldots \& S_n) = \frac{P(U_i) \cdot P(S_1 \mid U_i) \cdot \ldots \cdot P(S_n \mid U_i)}{\Sigma_{k=1..K} P(U_k) \cdot P(S_1 \mid U_k) \cdot \ldots \cdot P(S_n \mid U_k)}$$

– Entscheidung für wahrscheinlichste Ursache/Diagnose U_i

(oder: Alternativen anbieten)

– **Entscheidungstheorie**:

– Kosten/Nutzen einbeziehen

– Nützlichkeit („Utility“) optimieren

Utility(Aktion) = Nutzen(Aktion) · P(Erfolg der Aktion)

Schließen auf „a-posteriori-Wahrsch.“

Generelle Problematik:

- Voraussetzungen der Wahrscheinlichkeitsrechnung sind schwer zu erfüllen (z.B. unabhängige Symptome).
- a-priori-Wahrscheinlichkeiten werden benötigt.
  Ersatz durch (subjektive) Schätzung problematisch, insbesondere keine Vergleichbarkeit der Maßstäbe.
- Unübersichtlichkeit von Regelanwendungen bei regelbasierten Systemen.
  (Genauere Analyse mit Bayesschen Netzen).

Diskrete Zufallsvektoren

Systemzustand $\mathbf{x}$ beschrieben durch Vektor $\mathbf{x} = [x^{(1)},...,x^{(N)}]$

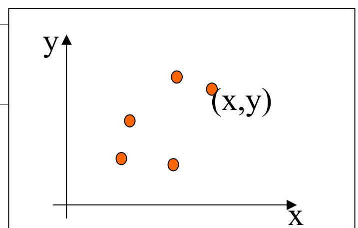
Wahrscheinlichkeit für Zustand $\mathbf{x}$: $P(x^{(1)},...,x^{(N)})$

Ursache	U	mit Werten	$u_1, ..., u_k$
Symptom	S^1	mit Werten	$s^1{}_1,..., s^1{}_{m1}$
Symptom	S^2	mit Werten	$s^2{}_1,..., s^2{}_{m2}$
	. . .		
Symptom	S^n	mit Werten	$s^n{}_1,..., s^n{}_{mn}$

Vollständige Beschreibung der Wahrscheinlichkeiten benötigt

$Anz(W^{(1)}) \cdot ... \cdot Anz(W^{(N)})$ Werte (N-dimensionale Matrix)

Diskrete Zufallsvektoren



- Zufallsvektor $\mathbf{X} = [X^{(1)},...,X^{(N)}]$

  kann Werte $\mathbf{x} = [x^{(1)},...,x^{(N)}] \in W^{(1)} \times ... \times W^{(N)}$ annehmen

- $P(\mathbf{x}) = P(x^{(1)},...,x^{(N)}) := P(\mathbf{X} = \mathbf{x})$:

  Wahrscheinlichkeit für „Ereignis $\mathbf{X} = \mathbf{x}$ “

- Verteilungsfunktion:

  $F(x^{(1)},...,x^{(N)}) = P(X^{(1)} < x^{(1)},..., X^{(N)} < x^{(N)})$

Diskrete Zufallsvektoren

Vollständige Beschreibung der Wahrscheinlichkeiten eines Zufallsvektor $\mathbf{X} = [X^{(1)},...,X^{(N)}]$ durch N-dimensionale Matrix.

Beispiel: $X^{(1)}$= Größe $X^{(2)}$= Gewicht (Angaben in %)

	40kg	50kg	60kg	70kg	80kg	90kg	100kg	Σ
150cm	3	3	2	1	0	0	0	9
160cm	2	3	6	5	3	0	0	19
170cm	3	2	7	7	6	1	0	26
180cm	0	0	3	8	9	7	2	29
190cm	0	0	0	6	6	2	3	17
Σ	8	8	18	27	24	10	5	100

Diskrete Zufallsvektoren

Schlußfolgerungen, Kombinationsmöglichkeiten anhand gegebener Wahrscheinlichkeiten (N-dimensionale Matrix für Zufallsvektor $\mathbf{X} = [X^{(1)},...,X^{(N)}]$):

- für Alternativen
- für bedingte Wahrscheinlichkeiten
- für Wahrscheinlichkeit bzgl einer/mehrerer Dimensionen

P(180cm)

P(Größe > 180cm)

P(„Größe - 100 < Gewicht“)

Randverteilung für eine Komponente $X^{(i)}$

Randverteilung für Komponente $X^{(i)}$:

Wahrscheinlichkeit, dass $X^{(i)}$ den Wert $x^{(i)}$ annimmt

$$P(x^{(i)}) = P(X^{(i)} = x^{(i)}) = \sum \{ P(\mathbf{x}) \mid \mathbf{x}^{(i)} = x^{(i)} \}$$

(als Projektion:

Summierung über (N-1)-dim. „Reihe" der N-dim. Matrix)

- Die Komponenten $X^{(1)},\ldots, X^{(N)}$ heißen *unabhängig*, falls

$$P(x^{(1)},\ldots, x^{(N)}) = P(x^{(1)}) \cdot \ldots \cdot P(x^{(N)})$$

Vollständige Beschreibung der Wahrscheinlichkeiten benötigt bei Unabhängigkeit nur noch $Anz(W^{(1)}) + \ldots + Anz(W^{(N)})$ Werte

Randverteilung für mehrere Komponenten

Randverteilung für $\mathbf{Y}=[Y^{(1)},\ldots,Y^{(M)}] \subseteq \mathbf{X}$:
Wahrscheinlichkeit, dass $\mathbf{Y}$ den Wert $\mathbf{y}=[y^{(1)},\ldots,y^{(M)}]$ annimmt:

$$P(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y}) = \sum \{ P(\mathbf{x}) \mid Proj_Y(\mathbf{x}) = \mathbf{y} \}$$

(als Projektion:

Summierung über (N-M)-dim. „Reihe" der N-dim. Matrix)

- Bedingte Wahrscheinlichkeiten:

$$P(\mathbf{u} \mid \mathbf{v}) = P(\mathbf{u}, \mathbf{v}) \ / \ P(\mathbf{v}) \qquad (\mathbf{u},\mathbf{v} \subseteq \mathbf{x})$$

(Normierte Summierung
über Projektion in „**v**-wertiger Teil-Matrix")

Urnenbeispiel 1

In einer Urne sind 2 weiße Kugeln (w1,w2) und 3 rote Kugeln (r1,r2,r3). Es wird eine Kugel gezogen. Danach wird diese zurückgelegt, und eine erneut Kugel wird gezogen.

P(k1,k2) = Wahrscheinlichkeit(1.Kugel=k1, 2.Kugel=k2)

	w1	w2	r1	r2	r3
w1					
w2					
r1					
r2					
r3					

Urnenbeispiel 1

In einer Urne sind 2 weiße Kugeln (w1,w2) und 3 rote Kugeln (r1,r2,r3). Es wird eine Kugel gezogen. Danach wird diese zurückgelegt, und eine erneut Kugel wird gezogen.

P(k1,k2) = (Wahrscheinlichkeit 1.Kugel=k1, 2.Kugel=k2)

	w1	w2	r1	r2	r3
w1	1/25	1/25	1/25	1/25	1/25
w2	1/25	1/25	1/25	1/25	1/25
r1	1/25	1/25	1/25	1/25	1/25
r2	1/25	1/25	1/25	1/25	1/25
r3	1/25	1/25	1/25	1/25	1/25

Randverteilungen P(k1), P(k2)

	w1	w2	r1	r2	r3	
w1	1/25	1/25	1/25	1/25	1/25	1/5
w2	1/25	1/25	1/25	1/25	1/25	1/5
r1	1/25	1/25	1/25	1/25	1/25	1/5
r2	1/25	1/25	1/25	1/25	1/25	1/5
r3	1/25	1/25	1/25	1/25	1/25	1/5
	1/5	1/5	1/5	1/5	1/5	

K1 und K2 sind unabhängig: P(k1,k2) = P(k1) · P(k2)

Urnenbeispiel 2

In einer Urne sind 2 weiße Kugeln (w1,w2) und 3 rote Kugeln (r1,r2,r3). Es wird eine Kugel gezogen. Danach wird diese **nicht** zurückgelegt, und eine **weitere** Kugel wird gezogen.

P(k1,k2) = Wahrscheinlichkeit (1.Kugel=k1, 2.Kugel=k2)

	w1	w2	r1	r2	r3
w1					
w2					
r1					
r2					
r3					

Urnenbeispiel 1

In einer Urne sind 2 weiße Kugeln (w1,w2) und 3 rote Kugeln (r1,r2,r3). Es wird eine Kugel gezogen. Danach wird diese **nicht** zurückgelegt, und eine **weitere** Kugel wird gezogen.

P(k1,k2) = Wahrscheinlichkeit (1.Kugel=k1, 2.Kugel=k2)

	w1	w2	r1	r2	r3
w1	0	1/20	1/20	1/20	1/20
w2	1/20	0	1/20	1/20	1/20
r1	1/20	1/20	0	1/20	1/20
r2	1/20	1/20	1/20	0	1/20
r3	1/20	1/20	1/20	1/20	0

Randverteilungen P(k1), P(k2)

	w1	w2	r1	r2	r3	
w1	0	1/20	1/20	1/20	1/20	1/5
w2	1/20	0	1/20	1/20	1/20	1/5
r1	1/20	1/20	0	1/20	1/20	1/5
r2	1/20	1/20	1/20	0	1/20	1/5
r3	1/20	1/20	1/20	1/20	0	1/5
	1/5	1/5	1/5	1/5	1/5	

K1 und K2 sind abhängig: $P(k1,k2) \neq P(k1) \cdot P(k2)$

Verbundene Wahrscheinlichkeiten

Es gilt allgemein

$P(x^{(1)},\ldots, x^{(N)})$

$= P(x^{(N)} \mid x^{(N-1)},\ldots, x^{(1)}) \cdot P(x^{(N-1)},\ldots, x^{(1)})$

$= P(x^{(N)} \mid x^{(N-1)},\ldots, x^{(1)}) \cdot P(x^{(N-1)} \mid x^{(N-2)},\ldots, x^{(1)}) \cdot P(x^{(N-2)},\ldots, x^{(1)})$

$= P(x^{(N)} \mid x^{(N-1)},\ldots, x^{(1)}) \cdot P(x^{(N-1)} \mid x^{(N-2)},\ldots, x^{(1)}) \cdot\ldots\cdot P(x^{(2)}|x^{(1)})\cdot P(x^{(1)})$

$= \prod_{i=1,\ldots,n} P(x^{(i)} \mid x^{(i-1)},\ldots, x^{(1)})$

- Berechnung von $P(x^{(1)},\ldots, x^{(N)})$ für beliebige $x^{(1)},\ldots, x^{(N)}$
- Für jede Reihenfolge der Variablen darstellbar.

Verbundene Wahrscheinlichkeiten

In einer Urne sind 2 weiße Kugeln (w1,w2) und 3 rote Kugeln (r1,r2,r3). Es werden nacheinander 4 Kugeln A,B,C,D gezogen.

P(A=r3,B=w1,C=r1,D=w2)

= P(D=w2 | A=r3,B=w1,C=r1)·P(A=r3,B=w1,C=r1)

= P(D=w2 | A=r3,B=w1,C=r1)·P(C=r1 | A=r3,B=w1)·P(A=r3,B=w1)

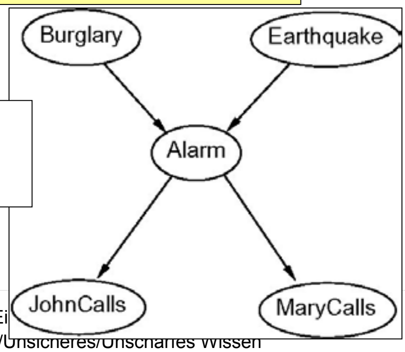
= P(D=w2 | A=r3,B=w1,C=r1)·P(C=r1 | A=r3,B=w1)·P(B=w1 | A=r3) · P(A=r3)

Bayessche Netze

probabilististic networks, belief networks, causal networks, knowledge maps, ...

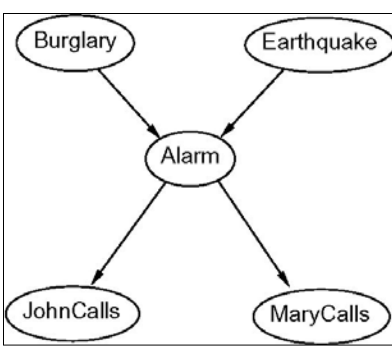
- Darstellung der direkten Abhängigkeiten von *Bedingungen* durch gerichteten zyklenfreien Graph.

Beispiele aus Russell/Norvig, Kapitel 14 bzw.15



Bayessche Netze

- Bedingungen X als Knoten,
- gerichtete Kanten für „direkten Einfluß“

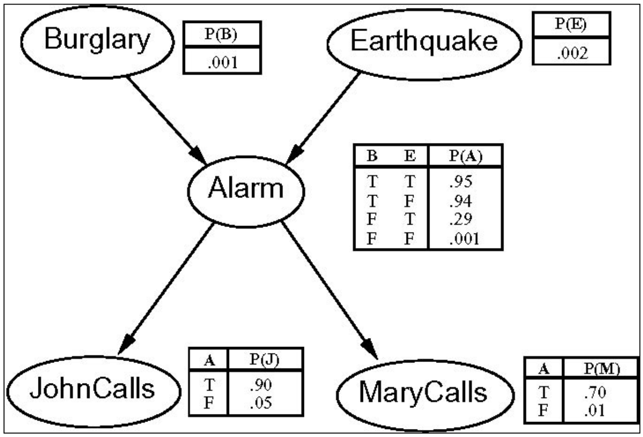


Wahrscheinlichkeiten durch Tabellen angeben, z.B. für Alarm:

		$P(A \mid B,E)$	
Bulgl.	Earthq.	true	false
true	true	0.950	0.050
true	false	0.960	0.040
false	true	0.290	0.710
false	false	0.001	0.999

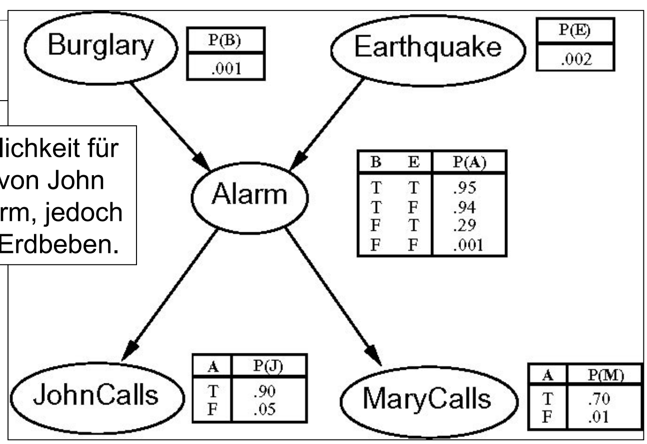
Bayessche Netze

• Einfluss der unmittelbaren Eingangsknoten („Eltern“) für jeden Knoten durch Tabellen bedingter Wahrscheinlichkeiten



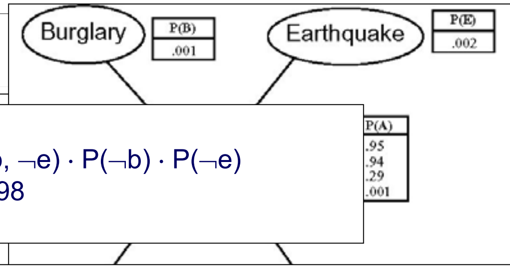
Inferenz

Beispiel:Wahrscheinlichkeit für gleichzeitigen Anruf von John und Mary wegen Alarm, jedoch ohne Einbruch oder Erdbeben.



$$
\begin{aligned}
& P(jc, mc, a, \neg b, \neg e) \\
&= P(jc \mid a) \cdot P(mc \mid a) \cdot P(a \mid \neg b, \neg e) \cdot P(\neg b) \cdot P(\neg e) \\
&= 0.9 \cdot 0.7 \cdot 0.001 \cdot 0.999 \cdot 0.998 \\
&= 0.00062
\end{aligned}
$$

Inferenz



$P(jc,mc,a,\neg b,\neg e)$
$= P(jc \mid a) \cdot P(mc \mid a) \cdot P(a \mid \neg b,\neg e) \cdot P(\neg b) \cdot P(\neg e)$
$= 0.9 \cdot 0.7 \cdot 0.001 \cdot 0.999 \cdot 0.998$
$= 0.00062$

Eigentlich gilt für Verbundwahrscheinlichkeit:
$P(jc,mc,a,\neg b,\neg e)$
$= P(jc \mid mc,a,\neg b,\neg e) \cdot P(mc \mid a,\neg b,\neg e) \cdot P(a \mid \neg b,\neg e) \cdot P(\neg b \mid \neg e) \cdot P(\neg e)$

Annahmen:
JC unabhängig von MC,B,E gegeben A :
$P(jc \mid mc,a,\neg b,\neg e) = P(jc \mid a)$
MC unabhängig von B,E gegeben A :
$P(mc \mid a,\neg b,\neg e) = P(mc \mid a)$
B unabhängig von E :
$P(\neg b \mid \neg e) = P(\neg b)$

Bedingte Unabhängigkeit

Betrachten disjunkte Teilmengen:

- $\mathbf{U}=[U^{(1)},\ldots,U^{(M)}] \subseteq \mathbf{X}$
- $\mathbf{V}=[Y^{(1)},\ldots,Y^{(M)}] \subseteq \mathbf{X}$
- $\mathbf{Z}=[Y^{(1)},\ldots,Y^{(M)}] \subseteq \mathbf{X}$

U , **V** heißen ***unabhängig gegeben*** **Z**, falls

$$P(\mathbf{U} \mid \mathbf{V}, \mathbf{Z}) = P(\mathbf{U} \mid \mathbf{Z})$$

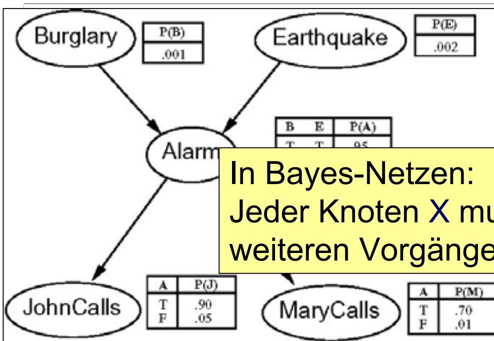
dabei gilt auch (Symmetrie):

$$P(\mathbf{V} \mid \mathbf{U}, \mathbf{Z}) = P(\mathbf{V} \mid \mathbf{Z})$$

Betrachtung im „Kontext“ Z

Jeweils für alle konkreten Werte von **U**,**V**, **Z** .

Bedingte Unabhängigkeit



In Bayes-Netzen:
Jeder Knoten X muss unabhängig sein von den weiteren Vorgängern V bei gegebenen Eltern Z .

X , **V** ***unabhängig gegeben*** **Z**, falls P(**X** | **V, Z**) = P(**U** | **Z**)

$P(jc,mc,a,\neg b,\neg e)$
$= P(jc \mid a) \cdot P(mc \mid a) \cdot P(a \mid \neg b,\neg e) \cdot P(\neg b) \cdot P(\neg e)$

Inferenz

Wahrscheinlichkeit für Einbruch bei Anruf von John und Mary:

$P(b \mid jc,mc) = P(b,jc,mc) \;/\; P(jc,mc) = \alpha \cdot P(b,jc,mc)$

$P(b,jc,mc)$
$= P(b,jc,mc,a,e) + P(b,jc,mc,\neg a,e) + P(b,jc,mc,a,\neg e) + P(b,jc,mc,\neg a,\neg e)$
$= P(jc \mid a) \cdot P(mc \mid a) \cdot P(a \mid b,e) \cdot P(b) \cdot P(e)$
$+ P(jc \mid \neg a) \cdot P(mc \mid \neg a) \cdot P(\neg a \mid b,e) \cdot P(b) \cdot P(e)$
$+ P(jc \mid a) \cdot P(mc \mid a) \cdot P(a \mid b,\neg e) \cdot P(b) \cdot P(\neg e)$
$+ P(jc \mid \neg a) \cdot P(mc \mid \neg a) \cdot P(\neg a \mid b,\neg e) \cdot P(b) \cdot P(\neg e)$
$= 0.00059224$

$P(\neg b \mid jc,mc) = P(\neg b,jc,mc) \;/\; P(jc,mc) = \alpha \cdot P(\neg b,jc,mc)$

$P(\neg b,jc,mc) = \ldots = 0.0014919$

Normierung $P(b \mid jc,mc) + P(\neg b \mid jc,mc) = 1$ mittels α :

$P(b \mid jc,mc) = 0.284$

Wahrscheinlichkeit für Einbruch bei Anruf von John und Mary ≈28%

Inferenz

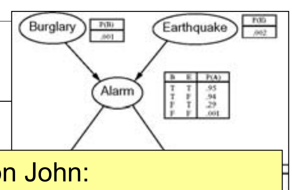
A-posteriori Wahrscheinlichkeiten für Anfrage-Variable ermitteln aus gegebenen Werten für Evidenz-Variable:

P(Anfrage | Evidenz)

für unterschiedliche Anfragen/Evidenzen gemäß (reduzierten) Formeln.

Propagierungsalgorithmen in Bayes Netzen
- über mehrere Schritte
- i.a. NP-hart
- Effizientere Algorithmen für eingeschränkte Netz-Klassen (z.B. einfach zusammenhängende Netze)

Unterschiedliche Inferenzen



- Diagnose

Wahrscheinlichkeit für Einbruch bei Anruf von John:
P(b | jc) = 0.016
Wahrscheinlichkeit für Einbruch bei Anruf von John und Mary:
P(b | jc,mc) = 0.29

- Kausal (Effekte)

Wahrscheinlichkeit für Anruf von John bei Einbruch:
P(jc | b) = 0.86
Wahrscheinlichkeit für Anruf von Mary bei Einbruch:
P(mc| b) = 0.67

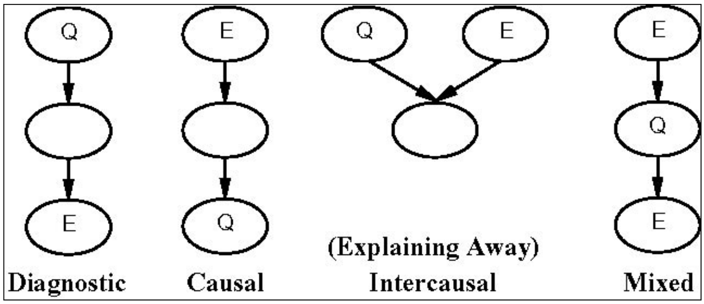
- Interkausal

Wahrscheinlichkeit für Einbruch bei Alarm und Erdbeben:
P(b| a,e) = 0.003

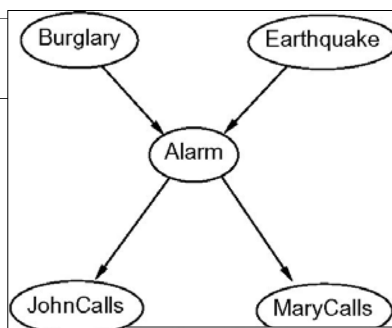
„Explaining away" bzgl. Einbruch, denn P(b|a) = 0.376

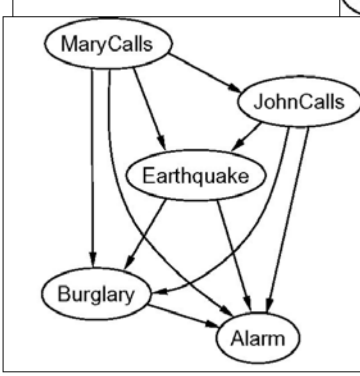
Inferenz in unterschiedlicher Richtung/Bedeutung

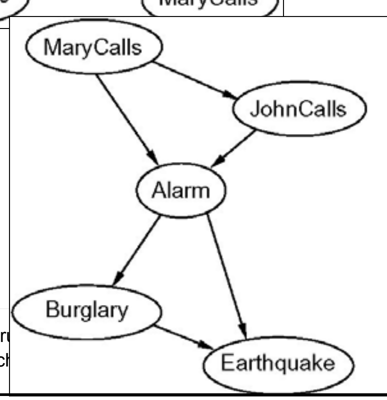
- Diagnose: Ursache aus Effekten
- Kausale Abhängigkeit: Effekte aus Ursachen
- Interkausal: Zwischen Ursachen gemeinsamer Effekte
- Mischungen der vorherigen



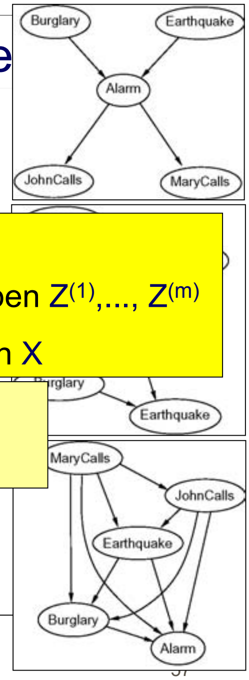
Konstruktion







Konstruktion Bayesscher Netze



Viele Varianten unter Beachtung von

- gerichtete Kanten für „direkten Einfluß“:

  Kanten von Eltern $Z^{(1)},\dots, Z^{(m)}$ nach X bei

  Unabhängigkeit von X und $Y^{(1)},\dots, Y^{(m)}$ gegeben $Z^{(1)},\dots, Z^{(m)}$

  für die mittelbaren Vorgänger $Y^{(1)},\dots, Y^{(m)}$ von X

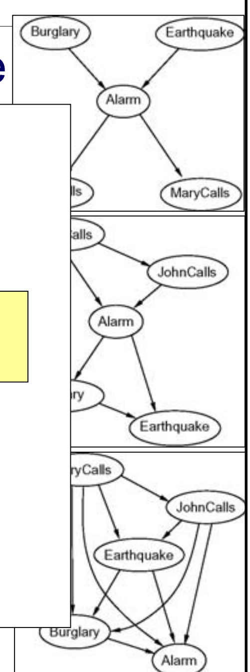
- Tabellen für bedingte Wahrscheinlichkeiten

  $P(X \mid Z^{(1)},\dots,Z^{(m)})$

Wenig Kanten (kleine Tabellen)

bei geschickter Anordnung.

Konstruktion Bayesscher Netze



(0) Reihenfolge $X_1,X_2,\dots, X_n$ wählen.

(i) für jedes $i=1,\dots,n$:

Bestimme $Parents(X_i)$ als eine minimale Teilmenge

von $\{X_1,\dots, X_{i-1}\}$, für die gilt

X_i bedingt unabhängig

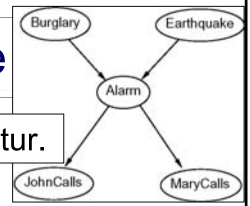
von $\{X_1,\dots, X_{i-1}\}$ - $Parents(X_i)$ gegeben $Parents(X_i)$

Ergänze das Netz durch

- Knoten für X_i
- Kanten von $Parents(X_i)$ nach X_i
- Tabelle für bedingte Wahrscheinlichkeit

  $P(X_i \mid Parents(X_i))$

Konstruktion Bayesscher Netze

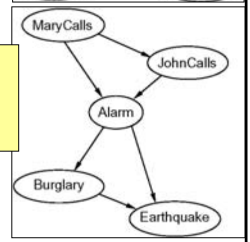


Reihenfolge $X_1, X_2, \ldots, X_n$ bestimmt Modellstruktur.

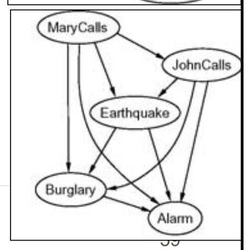
Gesichtspunkt 1:

Einsichtige Abhängigkeiten,

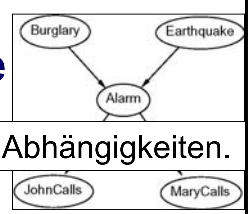
z.B. von Ursachen zu direkten Effekten.



Erleichtert statistische Analyse
(Aufstellung der Tabellen)



Konstruktion Bayesscher Netze

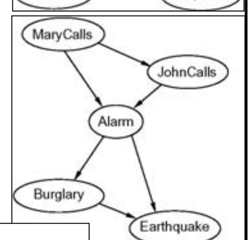


Reihenfolge $X_1, X_2, \ldots, X_n$ bestimmt modellierte Abhängigkeiten.

Gesichtspunkt 2:

Wenig Kanten (kleine Tabellen)

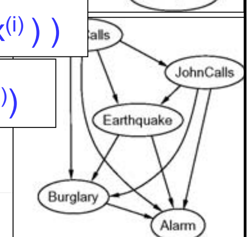
bei geschickter Wahl der Reihenfolge.



Bedingte Unabhängigkeit ausnutzen:

$$P(x^{(1)},\ldots, x^{(N)}) = \prod_{i=1,\ldots,n} P(x^{(i)} \mid \mathrm{Parents}(x^{(i)}))$$

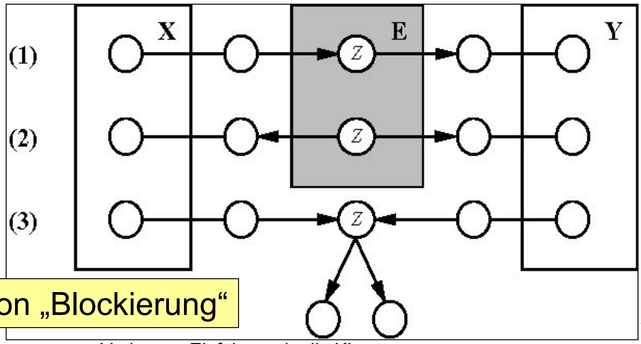
statt $P(x^{(1)},\ldots, x^{(N)}) = \prod_{i=1,\ldots,n} P(x^{(i)} \mid x^{(i-1)},\ldots, x^{(1)})$



Bedingte Unabhängigkeit

Graphische Bestimmung:

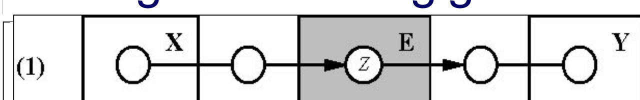
X und Y sind bedingt unabhängig gegeben E, falls jeder ungerichtete Weg von X nach Y durch einen Knoten Z blockiert wird



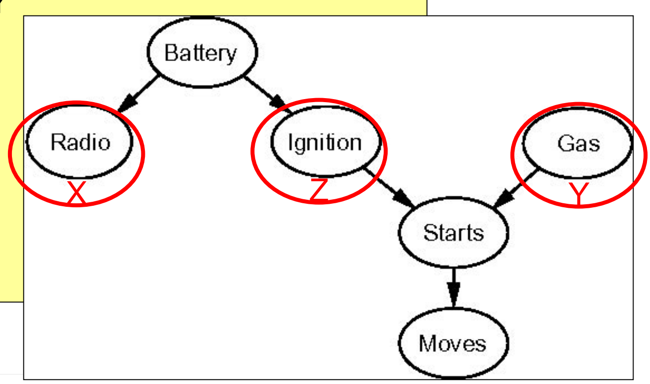
3 Formen von „Blockierung“

3 Formen von „Blockierung“

Bedingte Unabhängigkeit

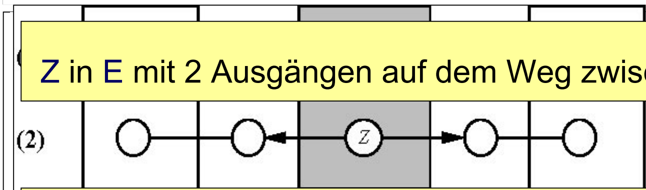


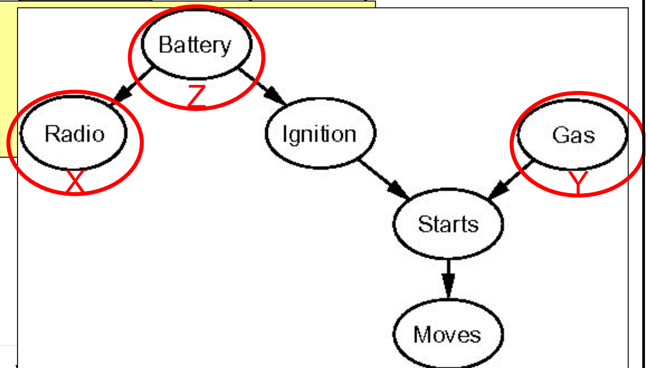
Z in E mit Eingang und Ausgang auf dem Weg zwischen X und Y



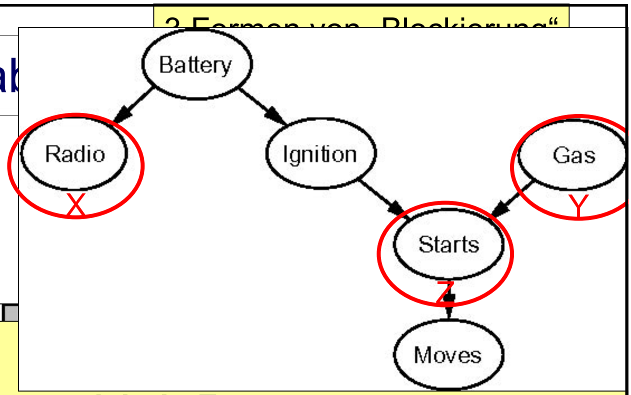
Bedingte Unabhängigkeit

Z in E mit 2 Ausgängen auf dem Weg zwischen X und Y



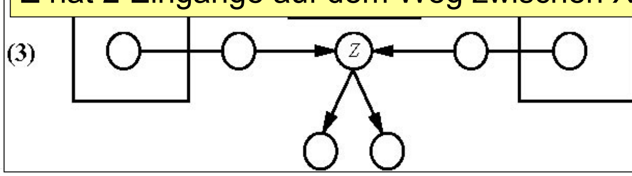


Bedingte Unabhängigkeit



Z und seine Nachfolger **nicht in** E,

Z hat 2 Eingänge auf dem Weg zwischen X und Y



Unscharfe Begriffe

- Klassische Logik, klassische Mengenlehre:
  - Wahrheitswerte: entweder H wahr oder H falsch
  - Mengen: entweder $x \in M$ oder $x \notin M$
- Alltagsbegriffe oft nicht klar abgrenzbar:

billiges Hemd, großer Mann, ...

blaue Farbe	100%ig:	RGB=	0, 0,255
blaue Farbe	100%ig:	RGB=	0, 0,100
blaue Farbe	80%ig:	RGB=	100,100,255
blaue Farbe	10%ig:	RGB=	50,150,255
blaue Farbe	0%ig:	RGB=	255,255, 0

Unschärfe: Mehrwertige Logik

- „Linguistische Terme“ (groß, blau,...) als *unscharfe Begriffe*

Eine Aussage „ Eigenschaft E trifft auf x zu “
wird nicht durch wahr/falsch bewertet,
sondern durch Grad von Wahrheit:
„ Eigenschaft E trifft auf x mit 25% zu “

Wahrheitswerte aus [0,1] (0% - 100%)

Analog: „ x gehört zur Menge M “
wird nicht durch wahr/falsch bewertet,
sondern durch Grad von Zugehörigkeit:
„ x gehört zur Menge M mit 60% “

Unschärfe: Ähnlichkeit

•„Linguistische Terme“ (groß, blau,...) als *Prototypen*

An die Stelle der Aussage „ x hat Eigenschaft E “

tritt die Aussage

„ die Ähnlichkeit von x zur Eigenschaft E beträgt 25% “

Ähnlichkeitsmaß: Wert aus [0,1] (0% - 100%)

Analog:

An die Stelle der Aussage „ x gehört zur Menge M “

tritt die Aussage

„ die Ähnlichkeit von x zur Menge M beträgt 60% “

Unscharfes Schließen

(Alltags-)Regeln auf der Basis linguistischer Terme:

Bei Erhitzung nicht ins kalte Wasser springen.

Bei hellem Objekt in geringer Entfernung kurze Belichtungszeit.

Unempfindliche Textilien bei hoher Verschmutzung heiß waschen

Linguistische Terme beschreiben Prototypen.

Format der Regeln:

IF $X_1 = L_1$ AND $X_2 = L_2$ AND ... AND $X_n = L_n$ THEN $S = L$.

Benötigen logische Verknüpfung linguistischer Terme, z.B.:

(80% $X_1 = L_1$) AND (30% $X_2 = L_2$)

ist äquivalent zu 30% ($X_1 = L_1$ AND $X_2 = L_2$)

Mehrwertige Logik/Fuzzy-Mengen

- Menge von Wahrheitswerten
- Logische Operatoren (AK, PK)

- klassische Operatoren „einbetten“:
  speziell $\neg$, $\wedge$, $\vee$, $\forall$, $\exists$
  unter Beibehaltung z.B. von $p \wedge q = \neg (\neg p \vee \neg q)$

- Operationen für Mengen:
  Analog zu klassischem Zusammenhang mit AK:
  $M \cap N = \{ x \mid x \in M \wedge x \in N \}$
  $M \cup N = \{ x \mid x \in M \vee x \in N \}$
  $U - M = \{ x \mid x \in U \vee x \notin M \}$

Mehrwertige Logik: Syntax (analog zu PK1)

- Menge *K* von aussagenlogischen Operatoren κ
  (klassische Operatoren: $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$)
- Menge Q von Quantoren
  (klassische Quantoren: $\exists$, $\forall$)

Ausdrücke:

- Wenn $H_1, \ldots, H_{n\kappa}$ Ausdrücke sind und κ ein n-stelliger aussagenlogischer Operator ist,
  so ist auch $\kappa(H_1, \ldots ,H_{n\kappa})$ ein Ausdruck.

- Wenn F ein Ausdruck, x ein Individuenzeichen und Quantor q ist,
  so ist auch q x F ein Ausdruck.

Mehrwertige Logik: Semantik (analog PK1)

- Für die Interpretation der Formeln benötigen wir an Stelle der ,,klassischen" Wahrheitsfunktionen des AK

  not : $\{0,1\} \to \{0,1\}$

  and, or, seq, aeq : $\{0,1\} \times \{0,1\} \to \{0,1\}$

  Wahrheitswertfunktionen für jeden Operator $\kappa \in K$:

  $$t_\kappa : \nu^{n_\kappa} \to \nu$$

  und Wahrheitswertfunktionen für jeden Quantor $q \in Q$:

  $$t_q : 2^\nu \setminus \{\varnothing\} \to \nu$$

Definition: ν-wertige Logik

ν sei eine **Menge von Wahrheitswerten**.

Eine ν**-wertige Logik** ist gegeben durch

- eine Menge K aussagenlogischer Operatoren κ

  jeweils mit Stelligkeit n_κ und

  zugeordneter Wahrheitswertfunktion $t_\kappa : \nu^{n_\kappa} \to \nu$

- eine Menge Q von Quantoren q,

  jeweils mit einer Wahrheitswertfunktion $t_q : 2^\nu \setminus\{\varnothing\} \to \nu$

Mehrwertige Logik: Semantik (analog PK1)

Eine Interpretation S=[U,I] über einen Universum (Menge) U ordnet jedem m-stelligen Prädikatenzeichen P eine m-stellige Funktion $I(P) : U^m \to \nu$ zu.

(Klassisch: m-stelliges Prädikat (Relation) über U, d.h. eine m-stellige Funktion $I(P) : U^m \to \{0,1\}$).

Wert eines Ausdrucks
bei Belegung β in der Interpretation S=[U,I]
für Operatoren $\kappa \in K$ bzw. Quantoren $q \in Q$ gemäß:

$$Wert_S(P(t_1,\dots,t_m), \beta) := I(P)(\beta(t_1),\dots,\beta(t_n))$$

$$Wert_S(\kappa(H_1,\dots,H_{n\kappa}),\beta) := t_\kappa(Wert_S(H_1,\beta),\dots,Wert_S(H_{n\kappa},\beta))$$

$$Wert_S(q\ x\ H(x), \beta) := t_q(\{ Wert_S(H(a), \beta) \mid a \in U \})$$

Beispiele

$\nu = [0,1]$

$t_\neg(x) = 1 - x$

$t_\wedge(x,y) = \min(x,y)$

$t_\vee(x,y) = \max(x,y)$

$t_\to(x,y) = \min(1,1-x+y)$

$t_\forall(M) = \inf(M)$

$t_\exists(M) = \sup(M)$

$\nu = [0,1]$

$t_\neg(x) = 1 - x$

$t_\wedge(x,y) = x \bullet y$

$t_\vee(x,y) = x + y - x \bullet y$

$t_\to(x,y) = 1 - x + x \bullet y$

$t_\forall(M) = \inf(M)$

$t_\exists(M) = \sup(M)$

Die klassische Logik ist jeweils „Spezialfall“ bei $\nu = \{0,1\}$.

$x \to y = \neg x \vee y$ muss dabei nicht gelten

T-Norm, T-Conorm

Ziele für Definition mehrwertiger Logiken

mit Wahrheitswerten $\nu = [0,1]$:

- Erweiterung der klassischen Logik
- Operatoren sind intuitiv definiert.
- Beibehaltung gewisser Korrespondenzen, z.B.:

$$t(p \wedge q) = \mathrm{Min}(p,q)$$
$$= 1 - (\mathrm{Max}(1-p,1-q)) = \neg(\neg p \vee \neg q)$$

Klassische Logik als Einschränkung

mit wahr=1 und falsch=0

T-Norm, T-Conorm

Axiome für $t_\wedge$ (**T-Norm**) und $t_\vee$ (**T-Conorm**):

$$t : [0,1] \times [0,1] \to [0,1]$$

heißt *T-Norm* (bzw. *T-Conorm*), falls:

- t ist assoziativ und kommutativ,
- t ist monoton wachsend in beiden Argumenten:

  $t(u_1,v) \leq t(u_2,v)$ und $t(v,u_1) \leq t(v,u_2)$ für $u_1 \leq u_2$
- 1 (bzw. 0) ist neutrales Element:

  $t(1, v) = v$ (bzw. $t(0, v) = v$).

T-Norm, T-Conorm

$n : [0,1] \rightarrow [0,1]$

heißt *strikte Negationsfunktion*, falls

- $n(0) = 1$, $n(1) = 0$
- n ist streng monoton fallend und stetig

(n^{-1} existiert, da n bijektiv ist.)

Bemerkung :

Ist t eine T-Norm und n strikte Negationsfunktion,

so ist $n^{-1}(t(n(u),n(v)))$ eine T-Conorm.

Aussagenlogische Entsprechnung zu $a \vee b = \neg(\neg a \wedge \neg b)$

Paare oft benutzter T-Normen
und T-Conormen,

jeweils zusammen mit der Negationsfunktion
$n(u) = 1 - u$

min

$t(u,v) = \min(u,v)$	$\overline{t}(u,v) = \max(u,v)$

BeschränkteDifferenz

$t(u,v) = \max(0, u+v-1)$	$\overline{t}(u,v) = \min(1, u+v)$

Drastisches Produkt

$t(u,v) = \begin{cases} v & : u = 1 \\ u & : v = 1 \\ 0 & : sonst \end{cases}$	$\overline{t}(u,v) = \begin{cases} v & : u = 0 \\ u & : v = 0 \\ 1 & : sonst \end{cases}$

Algebraisches Produkt

$t(u,v) = u * v$	$\overline{t}(u,v) = u + v - u * v$

Hamacher für $\gamma \geq 0$

$t_\gamma(u,v) = \frac{u*v}{\gamma+(1-\gamma)*(u+v-1)}$	$\overline{t_\gamma}(u,v) = \frac{(u+v-u*v)-(1-\gamma)uv}{1-(1-\gamma)*uv}$

Dombi für $\gamma > 0$

$t_\gamma(u,v) = \frac{u*v}{u*v+\sqrt[\gamma]{(1-u)^\gamma+(1-v)^\gamma}}$	$\overline{t_\gamma}(u,v) = \frac{\sqrt[\gamma]{u^\gamma+v^\gamma}}{1-u-v+uv+\sqrt[\gamma]{u^\gamma+v^\gamma}}$

Kombination lokaler Eigenschaften

- Ermittlung der (globalen) Bewertung eines Objekts aus den (lokalen) Bewertungen seiner Eigenschaften

Wert eines Hauses aus
- Anzahl der Zimmer
- Lage
- Größe
- Alter

Funktionen zur Ermittlung der globalen Werte aus lokalen Werten

Wahrheitswert eines logischen Ausdrucks aus
- Wahrheitswerten der Variablen (Extensionalität)

Kombination lokaler Eigenschaften

- (Globaler) Vergleich von Objekten

durch Vergleich der (lokalen) Werte ihrer Eigenschaften.

Ähnlichkeit von Objekten gemäß Ähnlichkeit der Eigenschaften

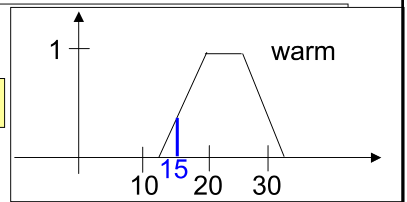
Ähnlichkeit zweier Häuser gemäß
- Anzahl der Zimmer
- Lage
- Größe
- Alter

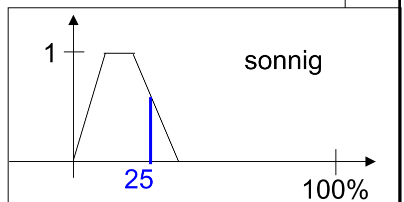
Analog:
Nähe im Raum gemäß Nähe bei den einzelnen Koordinaten.

Funktionen zur Ermittlung der globalen Werte aus lokalen Werten

Kombination lokaler Eigenschaften

Lokal: Ähnlichkeiten zu Prototyp



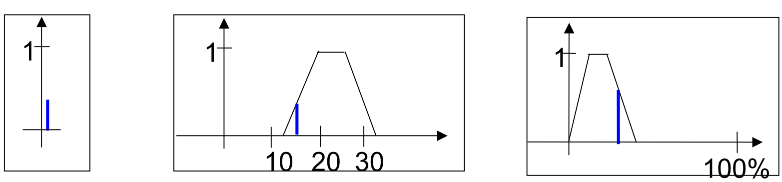


konkreter Wert	linguistischer Term	Ähnlichkeit
Temperatur 15^0	warm	30%
Wolken 25%	sonnig	60%

Kombination lokaler Eigenschaften

Globale Bewertung: Regeln anwenden

Freundlich = warm und sonnig



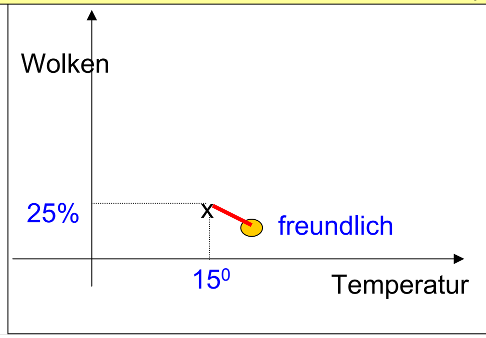
Kombination mittels T-Norm (bzw. T-Conorm)

Kombination lokaler Eigenschaften

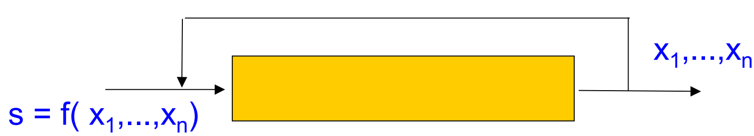
- Andere Möglichkeiten für Kombinationen.

  z.B. Nähe/Distanz als Maß für Vergleich

  Kombination der Distanzen entlang der Dimensionen
  (z.B. Euklidische Distanz, Manhattan-Distanz etc.)



Regelung (analog Steuerung)



(Scharfe) Stell-Werte s

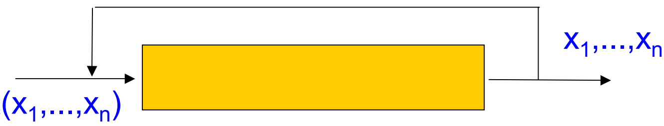
aus (scharfen) Meß-Werten $x_1,\dots,x_n$ berechnen:

Benötigt Funktion f mit $s = f(x_1,\dots,x_n)$ (exaktes Modell)

Alternative Fuzzy-Ansatz:

Unscharfe Regeln mit linguistischen Termen.

Fuzzy-Regelung



Ersatz der Funktion $s = f(x_1,\dots,x_n)$ durch Fuzzy-Regler:

scharfe Meßwerte $x_1,\dots,x_n$

umformen in: linguistische Terme $L_1,\dots,L_n$ — Fuzzifizierung

berechnen: linguistische Stell-Werte $S_1,\dots,S_k$ — Fuzzy-Regeln

umformen: scharfer Stell-Wert s — Defuzzifizierung

Fuzzy-Mengen

Definition :

Sei U ein Universum (eine klassische ,,scharfe" Menge).

Eine *Fuzzy-Teilmenge* von U (unscharfe Menge)

ist eine Funktion $\mu: U \to [0,1]$

- Eine *Fuzzy-Interpretation* eines Prädikats P über U ordnet P eine Fuzzy-Teilmenge von U zu.
- Redeweisen:

,,Belegung der Variablen x mit $a \in U$ erfüllt P mit Grad $\mu(a)$".

,,P(a) wird mit $\mu(a)$ akzeptiert."

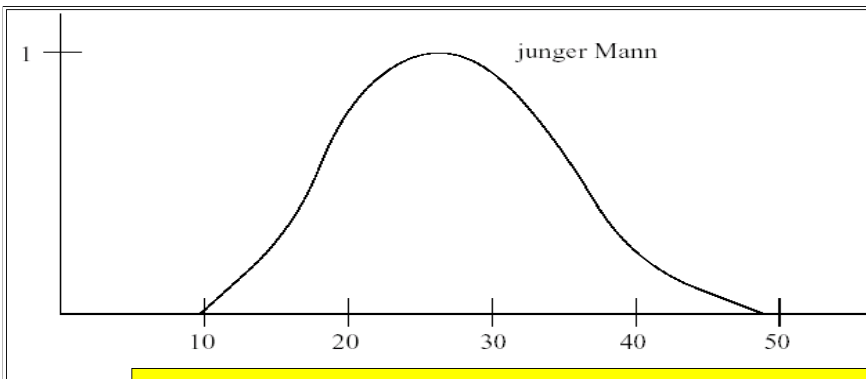
Linguistische Terme/Variable

Definition :

Ein Linguistischer Term ist eine Fuzzy-Teilmenge über einer Grundmenge U.

Eine Linguistische Variable ist eine Variable, die als Ausprägung bestimmte Linguistische Terme annehmen kann, wobei die Grundmenge U der Linguistischen Terme immer gleich bleibt.

Linguistische Terme/Variable



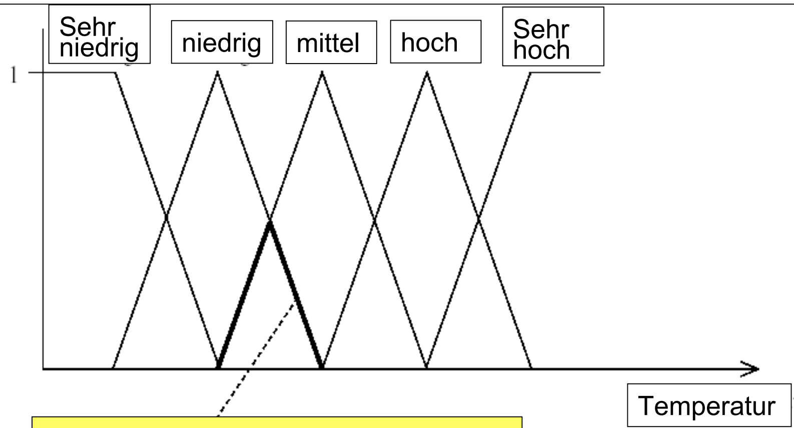
U = scharfe Altersangaben

Linguistischer Term: Junger Mann

Unscharfe Altersangaben für linguistische Variable:

Kleinkind, Junge, junger Mann, ...

Linguistische Terme/Variable



Bei Min als T-Norm Durchschnitt von niedrig und mittel

Mengenoperationen

Mit T-Norm $t_\wedge$ und T-Conorm $t_\vee$:

Durchschnitt: $(\mu \cap \nu)(x) = t_\wedge(\mu(x), \nu(x))$

Vereinigung: $(\mu \cup \nu)(x) = t_\vee(\mu(x), \nu(x))$

z.B.

$(\mu \cap \nu)(x) = \min(\mu(x), \nu(x))$

$(\mu \cup \nu)(x) = \max(\mu(x), \nu(x))$

$\overline{\mu}(x) = 1 - \mu(x)$

Klassisch: $\mu: U \rightarrow \{0,1\}$

$a \in M \subseteq U$ gdw. $\mu_M(a) = 1$

$a \in M \cap N$ gdw. $\mu_M = 1 \wedge \mu_N = 1$

$a \in M \cup N$ gdw. $\mu_M = 1 \vee \mu_N = 1$

$a \in \overline{M}$ gdw. $\neg\mu_M = 1$

Extensionsprinzip (Zadeh 1975)

Sei $\mu: X \to [0,1]$, $f: X \to Y$.

Dann wird definiert:

$$\nu(y) := \sup \{ \mu(x) \mid f(x) = y \}$$

(analog zu $y \in f(M)$ gdw. $\exists x \in M: f(x) = y$)

Repräsentations-Satz:

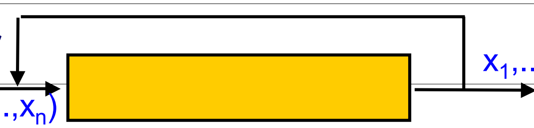
μ ist bestimmt durch die Niveaumengen

$\mu_c := \{ x \mid \mu(x) \geq c \}$ für alle $c \in [0,1]$

Darstellung im Rechner: stückweise lineare Menge.

z.B. sehr gering - gering - mittel - stark - sehr stark.

Fuzzyregler



$x_1,\ldots,x_n$

$s = FR(x_1,\ldots,x_n)$

Aufgabe:

Stellwerte für s aus Meßwerten $x_1,\ldots,x_n$ bestimmen.

Verwendung von Fuzzy-Regeln:

Linguistische Variable $X_1,\ldots,X_n ,S$ anstelle von $x_1,\ldots,x_n,s$.

Wertebereiche:

Linguistische Terme $L_1,\ldots,L_n ,T$ für die Variablen $X_1,\ldots,X_n ,S$.

Wenn Wäsche stark verschmutzt *und* Ölanteil hoch

dann Waschzeit lang.

Regeln

Wenn Hauptobjekt sehr nah *und* Belichtung wenig

dann Schärfentiefe sehr niedrig.

Wenn Hauptobjekt am Bildrand *dann* Schärfentiefe hoch.

Regeln haben Form

„*Wenn* linke Seite, *dann* rechte Seite."

Grad des Zutreffens der Stellgröße S (rechte Seite) berechnen aus Grad des Zutreffens der Meßgrößen (linke Seite)

Regeln

IF $X_{11} = L_{11} \wedge \ldots \wedge X_{1m1} = L_{1m1}$ **THEN** $S = T_1$

… ….

IF $X_{n1} = L_{n1} \wedge \ldots \wedge X_{nmn} = L_{nmn}$ **THEN** $S = T_n$

Ermittlung von Stellgrößen T_i an der Stellvariablen S

in Abhängigkeit von Eingangsgrößen X_{ij}.

X_{ij}, S: Linguistische Variable,

L_{ij}, T_i: Linguistische Terme.

Wenn Hauptobjekt sehr nah *und* Belichtung wenig

dann Schärfentiefe sehr niedrig.

Wenn Hauptobjekt am Bildrand *dann* Schärfentiefe hoch.

Regeln

IF $X_{i1} = L_{i1} \wedge \ldots \wedge X_{imi} = L_{im1}$ **THEN** $S = T_i$

Wenn Hauptobjekt sehr nah *und* Belichtung wenig

dann Schärfentiefe sehr niedrig.

Wenn Hauptobjekt am Bildrand *dann* Schärfentiefe hoch.

Beispiel:
X_{11} : Entfernung_Hauptobjekt
L_{11} : ,,sehr nah"
X_{12} : Ausleuchtung
L_{12} : ,,wenig"
LV_{21} : Position_Hauptobjekt
L_{21} : ,,am_Bildrand"
S : Schärfentiefe
T_1 : ,,sehr niedrig "
T_2 : ,,hoch"

Welches Resultat für:
Entfernung_Hauptobjekt = 2 m
Position_Hauptobjekt
= erstes Viertel des Bildes
Lichtmenge = 22 Lux

IF $X_{i1} = L_{i1} \wedge \ldots \wedge X_{imi} = L_{im1}$ **THEN** $S = T_i$

Fuzzifizierung

Aktivierung der linguistischen Terme L_{ij}

Linguistischer Term ist eine Fuzzy-Teilmenge $U: \to [0,1]$.

- Übereinstimmung von scharfen (Mess-)Werten x_i mit den linguistischen Termen L_i (Fuzzy-Mengen):

  $Act(L_i , x_i) := L_i(x_i)$

- Erweiterbar für Verwendung linguistischer Terme L'_i als Eingabe: $Act(L_i) := T\text{-}Norm(L_i,L'_i)$

Regelauswertung: Aktivierung

Regel anwendbar, falls Vorbedingungen erfüllt.

Hier: Grad der Erfülltheit = Aktivierung

IF $X_{i1} = L_{i1} \wedge \ldots \wedge X_{imi} = L_{im1}$ **THEN** $S = T_i$

- Aktivierung a einer Regel

ergibt sich durch Fuzzy-Kombination mittels T-Norm

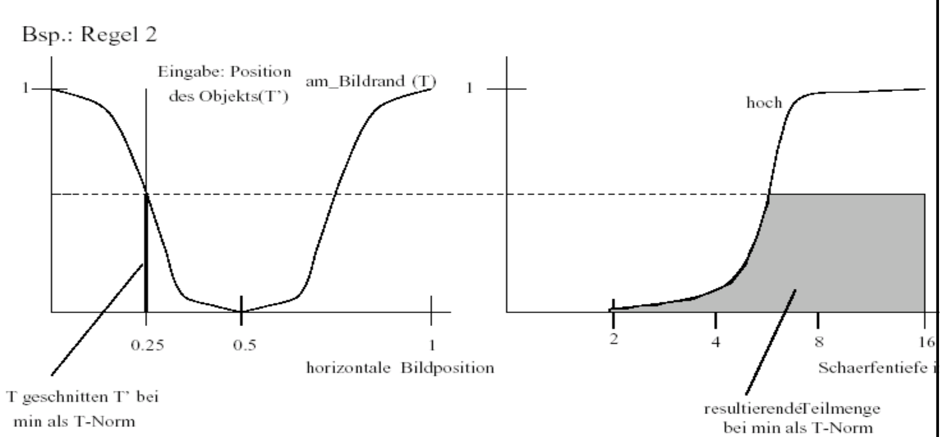
(ggf. auch T-Conorm für „oder“, und Negation für „nicht“)

der Aktivierungen der Linguistischen Terme L_{ij}

Regelauswertung: Aktivierung

Wenn Hauptobjekt am Bildrand *dann* Schärfentiefe hoch.

Position_Hauptobjekt = erstes Viertel des Bildes
Lichtmenge = 22 Lux



Aktivierung der Stellwerte

IF $X_{i1} = L_{i1} \wedge \ldots \wedge X_{imi} = L_{im1}$ **THEN** $S = T_i$

• Aktivierung $act(T_i)$ für linguistischen Term der rechten Seite ergibt sich aus Aktivierung a der Regel (der linken Seite)

Falls mehrere Regeln aktiviert für Stellgröße S: Verknüpfung der entsprechenden linguistischen Terme mittels T-Conorm:

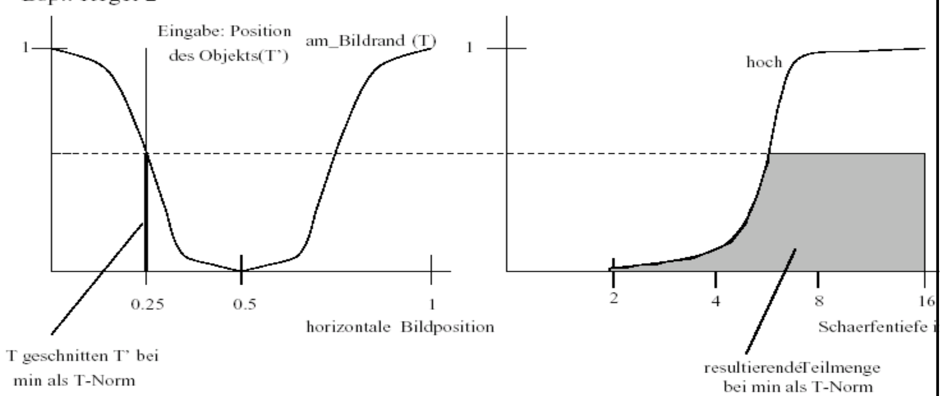
$$T = T_1 \vee T_2 \vee \ldots \vee T_n$$

Alternativen

Defuzzifizierung: Konkrete Stellgröße

Bestimmung des konkreten Stellwertes s aus T .

Bsp.: Regel 2



Defuzzifizierung: Konkrete Stellgröße

Unterschiedliche Varianten:

- Stellwert bei Maximum (nicht eindeutig):

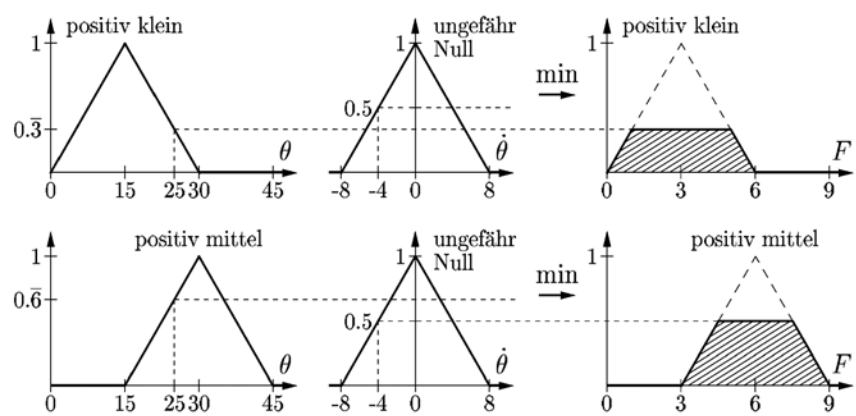
  s mit T(s) maximal

- Stellwert bei „Schwerpunkt“.
- Mittelwert bzgl. aller s mit Maximalwerten

Abhängig von Aufgabe auswählen.

Mögliche Probleme in einigen Fällen:

- Unstetiger Verlauf von $s = FR(x_1,...,x_n)$
- $T(s)=0$.



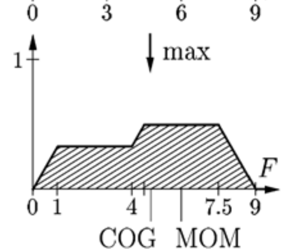
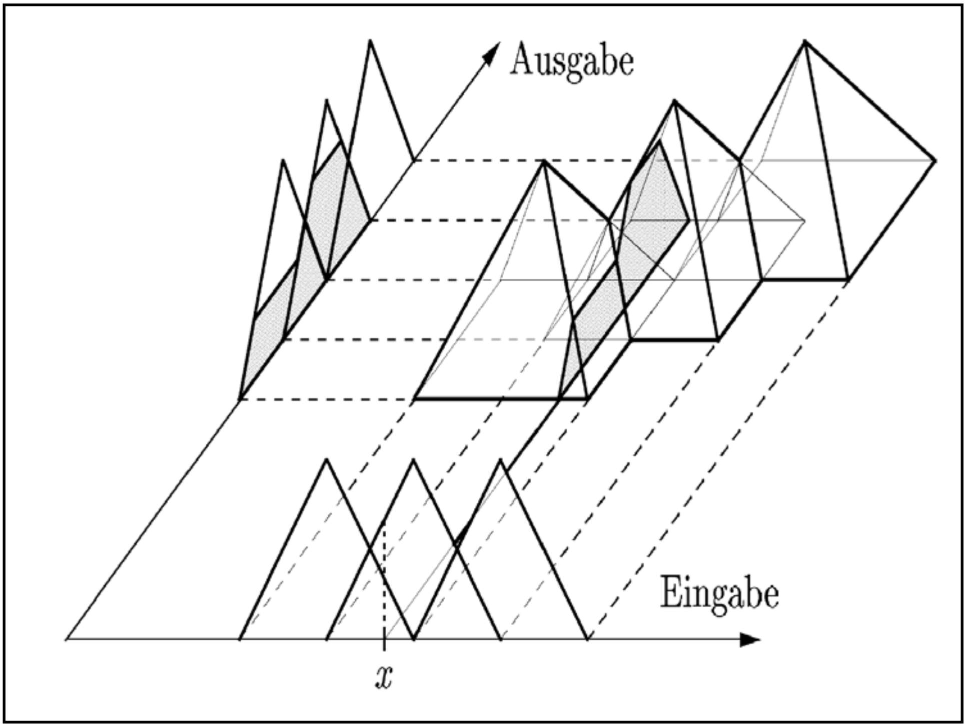


Abbildung 9.18: Regelauswertung bei einem Mamdani-Regler. Das Eingabetupel $(25, -4)$ führt zu der rechts gezeigten unscharfen Ausgabe. Aus dieser Fuzzy-Menge wird der entsprechende Ausgabewert durch Defuzzifikation, z.B. durch die Mean-of-Maxima-Methode (MOM) oder die Schwerpunktmethode (COG), bestimmt.

Aus Görz et al.: Handbuch der KI


Ausgabe
Eingabe
x