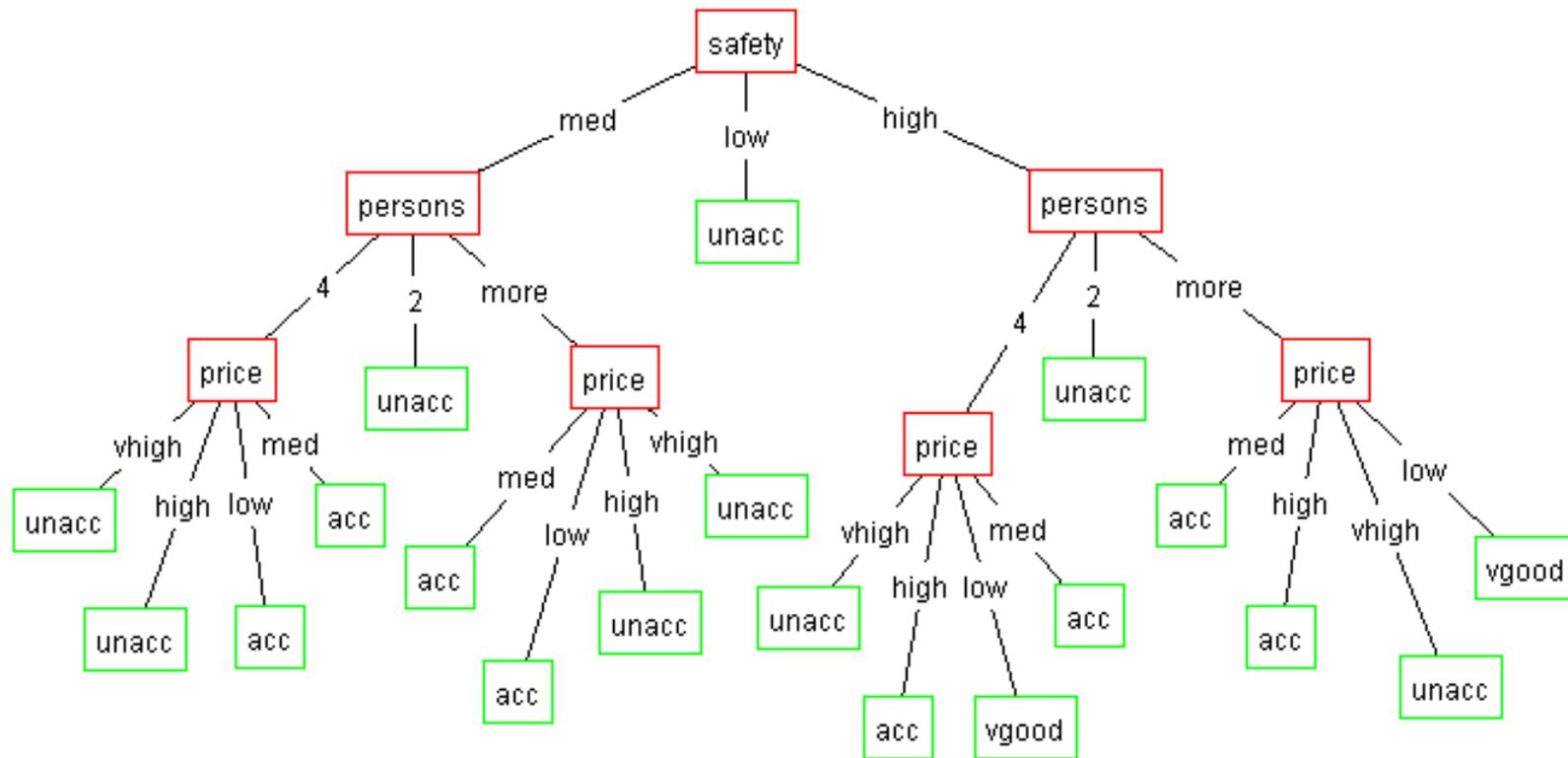


fuzzy-Entscheidungsbäume

- klassische Entscheidungsbaumverfahren
- fuzzy Entscheidungsbaumverfahren

Entscheidungsbäume



Was ist ein guter Mietwagen für einen Familienurlaub auf Kreta?

Entscheidungsbäume

- definiert für eine Menge von zulässigen Tests und eine Menge von Klassen
- innerer Knoten:
 - Test unter welchen Voraussetzungen zu einem Nachfolgeknoten verzweigt wird
 - ausgehende Kanten werden mit Bedingungen beschriftet
- Blattknoten:
 - repräsentiert eine Klassenzugehörigkeit

Entscheidungsbäume

- werden aus einer Menge klassifizierter Beispiele (Lernmenge) generiert \Rightarrow überwachtes Lernen
- Lernmenge charakterisiert durch:
 - Attribute (beschreiben Instanzen)
 - Attributwerte (diskret oder kontinuierlich)
 - Klassen (Zielattribut)
- dienen der schnellen und nachvollziehbaren Klassifikation unbekannter Beispiele

Entscheidungsbäume

- Anwendung:
 - wenn Knoten Blattknoten:
 - gib Klassenzugehörigkeit zurück
 - wenn Knoten innerer Knoten:
 - führe Test aus
 - rufe Algorithmus für passenden Unterbaum auf

klassische Entscheidungsbäume

- Lernen von Entscheidungsbäumen (Top-Down):
 - wähle Test der Unsicherheit bzgl. der Klassen minimiert
 - teile Lernmenge entsprechend des Tests auf und fahre rekursiv mit den Folgeknoten fort

- Minimierung der Unsicherheit:
 - mittels Information Gain

Minimierung der Unsicherheit

- Maß für Ungewissheit: (Entropie)

- $H = - \sum_v p(v) \log_2 p(v)$

- Information Gain:

- $IG = H - \sum_i P(a = v_i) H_{a = v[i]}$

ID3-Algorithmus

ID3(Instanzmenge M , Attributmenge A)

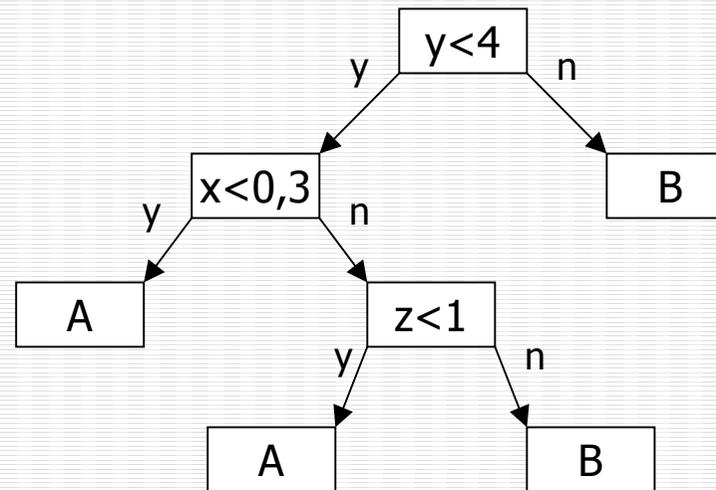
- wenn alle Instanzen aus M die gleiche Klasse C haben, gib C zurück
- sonst
 - kein Attribut mehr in $A \Rightarrow$ gib häufigste Klasse C zurück
 - wähle das Attribut a_i , das den Information Gain maximiert und teile damit den Knoten
 - rufe ID3 rekursiv für alle Unterbäume auf, wobei $A \setminus a_i$ gilt

Beispiel C4.5

- die Arbeitsweise von C4.5 ähnelt der von ID3, es werden allerdings kontinuierliche Attribute mittels binärerer Splits unterteilt

x	y	z	c
0.1	2	0	A
0.1	3	0	A
0.2	3	1	A
0.2	4	0	B
0.3	2	0	A
0.35	2	1	B
0.3	3	1	B
0.3	4	1	B

C4.5



Beispiel C4.5

$$H = -\sum_v p(v) \log_2 p(v)$$

$$IG = H - \sum_i P(a = v_i) H_{a=v_i}$$

Attribut:	Wert:	info(X):			info1(X):			info2(X):			IG:		
		As:	Bs:	H:	As:	Bs:	H:	As:	Bs:	H:	A%	B%	IG:
x	< 0,1	4	4	1,00	0	0	0,00	4	4	1,00	0,00	1,00	0,00
x	< 0,2	4	4	1,00	2	0	0,00	2	4	0,92	0,25	0,75	0,31
x	< 0,3	4	4	1,00	3	1	0,81	1	3	0,81	0,50	0,50	0,19
x	< 0,35	4	4	1,00	4	3	0,99	0	1	0,00	0,88	0,13	0,14
y	< 2	4	4	1,00	0	0	0,00	4	4	1,00	0,00	1,00	0,00
y	< 3	4	4	1,00	2	1	0,92	2	3	0,97	0,38	0,63	0,05
y	< 4	4	4	1,00	4	2	0,92	0	2	0,00	0,75	0,25	0,31
z	< 1	4	4	1,00	3	1	0,81	1	3	0,81	0,50	0,50	0,19

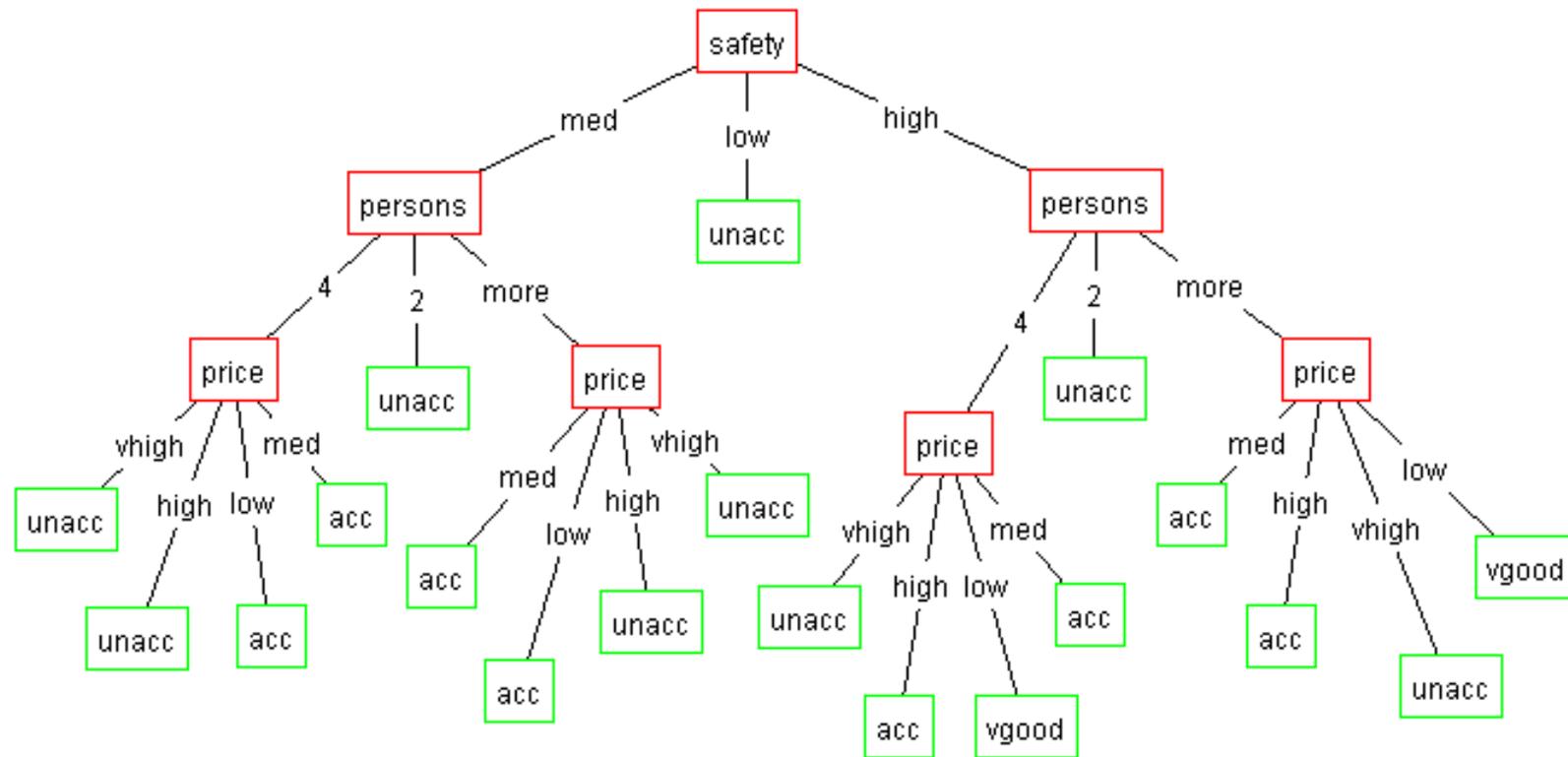
Attribut:	Wert:	info(X):			info1(X):			info2(X):			IG:		
		As:	Bs:	H:	As:	Bs:	H:	As:	Bs:	H:	A%	B%	IG:
x	< 0,1	4	2	0,92	0	0	0,00	4	2	0,92	0,00	1,00	0,00
x	< 0,2	4	2	0,92	2	0	0,00	2	2	1,00	0,33	0,67	0,25
x	< 0,3	4	2	0,92	3	0	0,00	1	2	0,92	0,50	0,50	0,46
x	< 0,35	4	2	0,92	4	1	0,72	0	1	0,00	0,83	0,17	0,32
y	< 2	4	2	0,92	0	0	0,00	4	2	0,92	0,00	1,00	0,00
y	< 3	4	2	0,92	2	1	0,92	2	1	0,92	0,50	0,50	0,00
y	< 4	4	2	0,92	4	2	0,92	0	0	0,00	1,00	0,00	0,00
z	< 1	4	2	0,92	3	0	0,00	1	2	0,92	0,50	0,50	0,46

Attribut:	Wert:	info(X):			info1(X):			info2(X):			IG:		
		As:	Bs:	H:	As:	Bs:	H:	As:	Bs:	H:	A%	B%	IG:
x	< 0,1	1	2	0,92	0	0	0,00	1	2	0,92	0,00	1,00	0,00
x	< 0,2	1	2	0,92	0	0	0,00	1	2	0,92	0,00	1,00	0,00
x	< 0,3	1	2	0,92	0	0	0,00	1	2	0,92	0,00	1,00	0,00
x	< 0,35	1	2	0,92	1	1	1,00	0	1	0,00	0,67	0,33	0,25
y	< 2	1	2	0,92	0	0	0,00	1	2	0,92	0,00	1,00	0,00
y	< 3	1	2	0,92	1	1	1,00	0	1	0,00	0,67	0,33	0,25
y	< 4	1	2	0,92	2	1	0,92	0	0	0,00	1,00	0,00	0,00
z	< 1	1	2	0,92	1	0	0,00	0	2	0,00	0,33	0,67	0,92

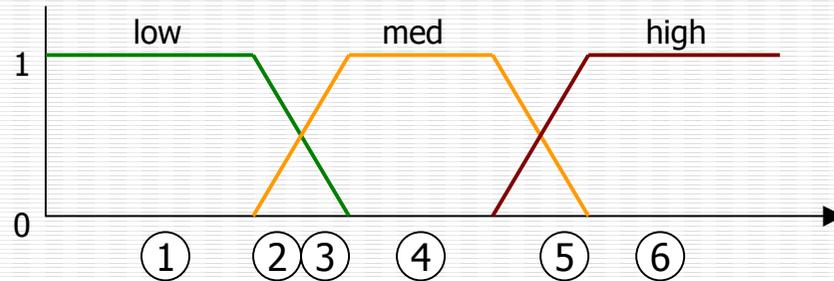
Warum fuzzy-Entscheidungsbäume?

- Lernmenge kann verrauscht sein
 - physikalische Messfehler
 - subjektive Beurteilung
- es können Attributwerte fehlen
 - unvollständige Datenbankeinträge
- Notwendigkeit sprachliche Unsicherheiten zu modellieren
 - linguistische Variablen (was ist "groß"?)

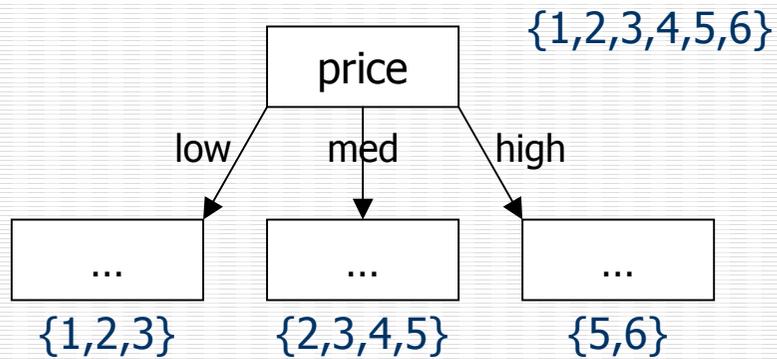
fuzzy-Entscheidungsbäume



fuzzy-Entscheidungsbäume



Beispiel:	Preis:	Klasse:
1	50,00	pos
2	57,00	neg
3	59,00	pos
4	66,00	pos
5	74,00	neg
6	79,00	neg



Linguistische Variablen

- Benutzerdefinierte Festlegung
 - fuzzy-Mengen werden aufgrund von Expertenwissen festgelegt
 - viel Hintergrundwissen nötig
 - mögliche Zusammenhänge in Beispielmenge werden nicht berücksichtigt
 - Zugehörigkeitsfunktionen u.U. unflexibel

Linguistische Variablen

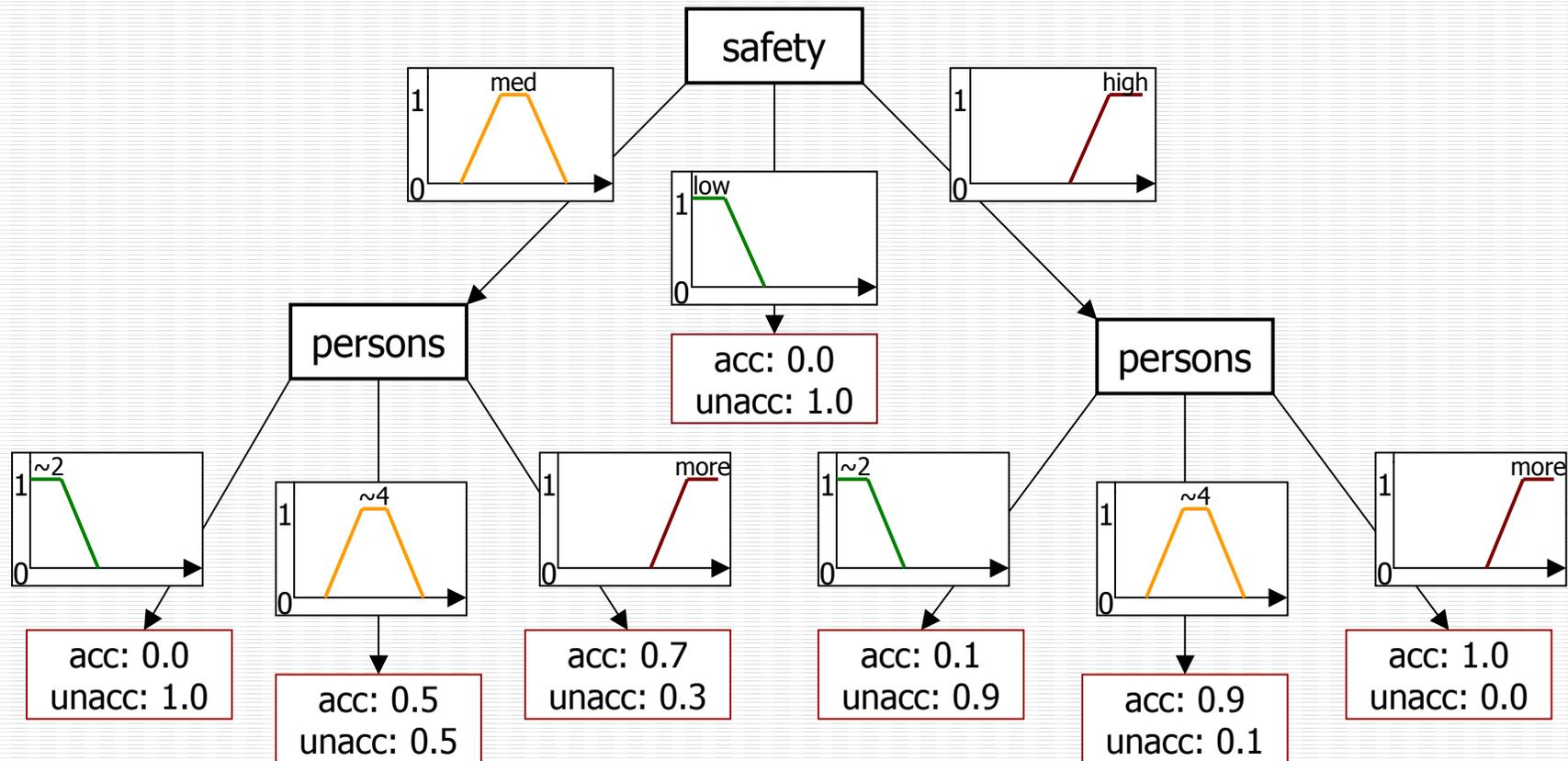
- Automatische Bestimmung
 - für kontinuierliche Werte werden besonders interessante Intervalle bestimmt und über diese Intervalle werden fuzzy-Mengen gelegt
 - keine Vorarbeit nötig – Definition der Zugehörigkeitsfunktionen erfolgt während des Lernprozesses
 - mögliche Zusammenhänge in Beispielmenge können berücksichtigt und gezielt genutzt werden

Anwendung FDT

- betrachte den Pfad durch Baum als Folge von Bedingungen
- verknüpfe die Zugehörigkeitsgrade zu den einzelnen Bedingungen mit einer t-Norm (das Ergebnis ist der Zugehörigkeitsgrad des Beispiels zu dem Blattknoten)
- wiederhole dies für alle weiteren Pfade im Baum
- verknüpfe die Klassenzugehörigkeits-Tupel der Blattknoten elementweise mit einer t-Conorm

- das Ergebnis-Tupel muss u.U. noch auf 1 normiert und kann anschließend defuzzifiziert werden

Beispiel



fuzzy-TDIDT

TDIDT(Beispielmenge M , Attributmenge A)

- wenn hoher Anteil an Beispielen in M die gleiche Klasse C haben, gib Klassenzugehörigkeit 1 für C zurück ($|M \cap C| / |M| > \theta_r$)
- sonst
 - wenn $A = \{\}$ oder $|M| \leq \theta_n$ gib Klassenzugehörigkeitsvektor für alle Klassen oder "unbestimmt" zurück
 - sonst
 - wähle bestes Attribut a_i mit den besten Werten
 - teile M nach Wert von a_i
 - rufe TDIDF($M_1, A \setminus a_i$), ... TDIDF($M_j, A \setminus a_i$) auf

FDT und verrauschte Daten

- **Meßfehler:**
 - fast immer bei kontinuierlichen Werten vorhanden
 - der Fehler wird an den Grenzen zwischen den Attributwerten durch Zugehörigkeitsgrade reduziert
 - fehleranfällige Objekte können geringere Klassenzugehörigkeiten zugewiesen bekommen
- **subjektive Urteile:**
 - Glaubwürdigkeit durch Zugehörigkeitswerte steuerbar
- **inkonsistente Daten lassen sich verarbeiten**

FDT und fehlende Attributwerte

- Wahrscheinlichkeit für bestimmten Attributwert:
 - Wahrscheinlichkeit kann Glaubwürdigkeit des gesamten Beispiels verringern
 - Wichtung der Klassenzugehörigkeit mit der Wahrscheinlichkeit für ergänzten Attributwert
- Attributwertbestimmung:
 - das Attribut mit fehlendem Attributwert wird Zielattribut, der FDT kann dann über die Klassenzugehörigkeit die Sicherheit für den Attributwert bestimmen
- Attributwert "*unbekannt*":
 - Klassenzugehörigkeit bleibt unverändert

Unterschiede FDT zu DT

- Beispiele können in mehreren Knoten auftreten, d.h. für jedes Beispiel wird ein Zugehörigkeitsgrad zu jedem Knoten des Baumes bestimmt
- die Berechnung des Informationsgehaltes muss u.U. angepasst werden, um die partiellen Zugehörigkeiten abzubilden (zusätzliche Normierung)
- größere Robustheit bzgl. inkonsistenter Daten

Literatur & Anwendungen

- CIspace: Tools for Learning Computational Intelligence
<http://www.cs.ubc.ca/labs/lci/CIspace/download.html>
- M. Guetova. Inkrementelle Fuzzy-Entscheidungs**ä**ume. Diplomarbeit TU Dresden, 2001.
http://www.wv.inf.tu-dresden.de/Publications/Diploma/diplom_guetova.ps.gz
- C.Z. Janikow. Fuzzy Decision Trees: Issues and Methods. 1996.
<http://www.cs.umsl.edu/~janikow/fid/>
- T. Scheffer. Entscheidungs**ä**ume. Vorlesungsfolien Data Mining, WS2002/2003.
<http://kd.cs.uni-magdeburg.de/wm2002ws.html>
- J. Zeidler. Unsch**ä**rf**ä**e Entscheidungs**ä**ume. Dissertation TU Chemnitz, 1999.
<http://www.tu-chemnitz/~jzei/STAFF/JENS/lernen.html>