

Beitrag zum Proseminar *Komplexe Adaptive Systeme*

Discovering Fuzzy Classifiers by Genetic Algorithms

Christoph Knopp <ckwon@gmx.de>

7. Januar 2005

PD Dr. Rüdiger Brause, Arbeitsgruppe Komplexe Systemarchitektur

J. W. Goethe-Universität Frankfurt / M., WS 2004/05



Discovering
Fuzzy
Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionenprozeß

Ergebnisse

Überblick

- **Einführung**
 - Konzepte der Fuzzy Logik
 - Entwicklung der Fuzzy Klassifikatoren
 - Evolutionsprozeß
 - Ergebnisse
-
- Problemstellung der Klassifikation
 - Lösungsvorschläge
 - Zielsetzung



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Ergebnisse

Überblick

- Einführung
 - **Konzepte der Fuzzy Logik**
 - Entwicklung der Fuzzy Klassifikatoren
 - Evolutionsprozeß
 - Ergebnisse
-
- Warum Fuzzy Logik?
 - Der Fuzzy Klassifikator
 - Fuzzy C-Mean Clustering



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Ergebnisse

Überblick

- Einführung
 - Konzepte der Fuzzy Logik
 - **Entwicklung der Fuzzy Klassifikatoren**
 - Evolutionsprozeß
 - Ergebnisse
-
- Genetische Algorithmen
 - Codierung von Regeln
 - Genetische Operatoren
 - Fitness-Funktion



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Ergebnisse

Überblick

- Einführung
 - Konzepte der Fuzzy Logik
 - Entwicklung der Fuzzy Klassifikatoren
 - **Evolutionsprozeß**
 - Ergebnisse
-
- Ablauf der Evolution
 - Bestimmung der Klassifikatoren



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Egebnisse

Überblick

- Einführung
- Konzepte der Fuzzy Logik
- Entwicklung der Fuzzy Klassifikatoren
- Evolutionsprozeß
- **Ergebnisse**

- Auswertung mit Testdatensätzen **IRIS**, **WINE** und **VOTE**
- Vergleich mit ähnlichen Verfahren



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Egebnisse

1 Einführung

Probleme in Anwendungen können oft auf Klassifikationsprobleme zurückgeführt werden.

Was ist Klassifikation?

- Einteilung von Elementen oder Objekten in Klassen
- korrekte Erkennung und Zuordnung von bekannten und unbekanntem Mustern

Welche Lösungen gibt es?

- Statistische Verfahren
- Methoden aus dem maschinellen Lernen und dem Data-Mining (Datengewinnung)



**Discovering
Fuzzy
Classifiers**

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionenprozess

Ergebnisse

2 Konzepte der Fuzzy Logik

Von Menschen formulierte Regeln sind für Computer nicht anwendbar.

Beispiel Heizung

- Wenn es ungefähr 19°C ist, dann muss die Heizung ein wenig aufgedreht werden

Menschen können diese Regel verstehen und danach handeln.

PC kann mit den unscharfen Begriffen

- ungefähr 19°C
- ein wenig aufdrehen

nicht arbeiten.



Discovering
Fuzzy
Classifiers

Titelseite
Einführung

Fuzzy Logik

Entwicklung
Evolutionsprozeß
Ergebnisse

Fuzzy Logik

In der Fuzzy (unscharfen) Mengen Theorie kann ein Element einer Menge auch nur zu einem bestimmten Grad angehören.

Beispiel:

Menge der angenehmen Raumtemperaturen 19°C - 24°C

- Mit der klassischen (scharfen) Mengenlehre: $18,9^{\circ}\text{C}$ gehört nicht mehr zu dieser Menge
- Mit der Fuzzy (unscharfen) Mengenlehre: $18,9^{\circ}\text{C}$ gehört noch mit 80% zu dieser Menge



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionenprozeß

Ergebnisse



Discovering Fuzzy Classifiers

Titelseite

Einführung

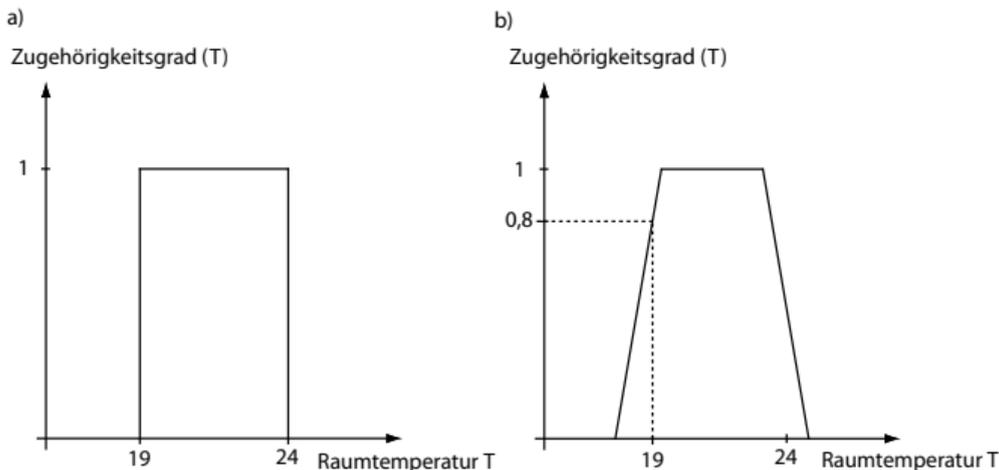
Fuzzy Logik

Entwicklung

Evolutionprozeß

Ergebnisse

Angenehme Raumtemperaturen im scharfen a) und unscharfen b) Intervall



Vorteil unscharfer Mengen:

- Mit Fuzzy Logik kann man Computern den Umgang mit unscharfen Begriffen wie „angenehm“, „warm“, „kalt“ ermöglichen

Der Fuzzy Klassifikator

Ein Fuzzy Klassifikator ist ein auf Fuzzy Logik basiertes Klassifikationsverfahren und kann durch Fuzzy Regeln dargestellt werden:

Fuzzy Regel:

IF Bedingung THEN Konsequenz [Gewicht]

Bedingung ist ein Fuzzy Ausdruck, *Konsequenz* ein atomarer Ausdruck, *Gewicht* ist eine reelle Zahl und gibt die Vertrauenswürdigkeit der Regel an.



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Ergebnisse

Fuzzy C-Mean Clustering

Eine Cluster-Analyse teilt eine Datenmenge in Klassen ein.

Das Fuzzy C-Mean Clustering arbeitet mit Fuzzy Logik und ordnet den Daten Zugehörigkeiten zu.

Dazu ermittelt der Fuzzy C-Mean Algorithmus ein Minimum der Bewertungsfunktion:

$$J(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m \|x_k - v_i\|^2$$

mit

$$u_{ik} \in [0, 1], \quad \sum_{i=1}^c u_{ik} = 1 \quad \forall k$$



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionenprozeß

Ergebnisse

3 Entwicklung der Fuzzy Klassifikatoren

Genetische Algorithmen

Genetische Algorithmen funktionieren ähnlich dem natürlichen Evolutionsprozeß:

- zufällige Erzeugung einer Grundpopulation aus Individuen
- Individuen enthalten Lösungen, die in den Chromosomen codiert sind
- eine Fitness-Funktion bewertet diese Lösungen
- Individuen mit guten Lösungen werden sich bevorzugt fortpflanzen (Survival of the Fittest)



Discovering
Fuzzy
Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Ergebnisse

Codierung der Regeln

Um die genetischen Algorithmen auf die Regeln der Fuzzy Klassifikatoren anwenden zu können, müssen diese in den Chromosomen codiert werden.

- *Michigan*-Ansatz:
 - jedes Individuum codiert eine Regel
- *Pittsburgh*-Ansatz:
 - jedes Individuum codiert eine Menge an Regeln



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Ergebnisse



Codierung des *Bedingung-Teils*

Verschiedene Möglichkeiten der Codierung:

- feste Kontroll-Strukturen
- lineare Bäume
- komplette Ausdrucks-Bäume

Hier wird die Disjunktive Normalform (DNF – k)
verwendet:

$$(A_1 \wedge A_2 \wedge \dots \wedge A_k) \vee (A_{k+1} \wedge A_{k+2} \wedge \dots \wedge A_{2k}) \vee \dots \\ \dots \vee (A_p \wedge A_{p+1} \wedge \dots \wedge A_{p+(k-1)})$$

Beispiel:

1	0	1	1	0	1	0	1	1	0	0	1
---	---	---	---	---	---	---	---	---	---	---	---

Genetische Operatoren



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Ergebnisse

- **Selektion**
 - Kreuzung
 - Mutation
-
- Selektion bestimmt die Individuen, die sich kreuzen dürfen
 - Individuen mit guter Fitness werden mit höherer Wahrscheinlichkeit ausgewählt
 - *Wettkampfselektion*: jeweils n Individuen werden miteinander verglichen

Genetische Operatoren



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Ergebnisse

- Selektion
 - **Kreuzung**
 - Mutation
-
- Chromosomen zweier Individuen werden vermischt
 - die neuen Individuen erben die Eigenschaften der Eltern – es entstehen optimierte Lösungen
 - *Single Point Crossover*: die Chromosomen werden an einem zufällig bestimmten Punkt getrennt und aufgeteilt

Genetische Operatoren



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Ergebnisse

- Selektion
 - Kreuzung
 - **Mutation**
-
- zufällige Änderung an den Chromosomen sorgt für einen großen Chromosomenpool
 - tritt nur mit geringer Wahrscheinlichkeit ein, um gute Lösungen nicht zu verlieren
 - es entstehen Lösungen, die durch Kreuzung nicht entstanden wären

Fitness Funktion

Die Fitness-Funktion bewertet die Effizienz der in den Individuen codierten Lösung.

Die Vorhersage-Wahrscheinlichkeit kann mit Hilfe der *Konfusionsmatrix* berechnet werden:

		Aktuelle Klasse	
		C	Nicht C
Vorhergesagte Klasse	C	TP	FP
	Nicht C	FN	TN



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionprozeß

Ergebnisse

Die Einträge der Konfusionsmatrix werden wie folgt berechnet:

$$TP = \sum_{i=1}^p \text{predicted}(\text{class}_1 \text{ data}_i)^{\frac{1}{2}}$$

$$TN = \sum_{i=1}^q (1 - \text{predicted}(\text{class}_2 \text{ data}_i))^{\frac{1}{2}}$$

$$FP = \sum_{i=1}^q \text{predicted}(\text{class}_2 \text{ data}_i)^{\frac{1}{2}}$$

$$FN = \sum_{i=1}^p (1 - \text{predicted}(\text{class}_1 \text{ data}_i))^{\frac{1}{2}}$$



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Ergebnisse



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Ergebnisse

Mit der Konfusionsmatrix kann man folgende Verhältnisse definieren:

Konfidenzfaktor (CF)

$$CF = \frac{TP}{(TP + FP)}$$

Vollständigkeitsfaktor (Comp)

$$Comp_1 = \frac{TP}{(TP + FN)}$$

$$Comp_2 = \frac{TN}{(TN + FP)}$$



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Ergebnisse

Die Fitness-Funktion wird nun wie folgt definiert:

$$\begin{aligned}\text{Fitness} &= \text{Comp}_1 \cdot \text{Comp}_2 \\ &= \frac{\text{TP}}{(\text{TP} + \text{FN})} \cdot \frac{\text{TN}}{(\text{TN} + \text{FP})}\end{aligned}$$

Damit die Regeln möglichst verständlich sind, läßt man einen weiteren Faktor einfließen – die *Einfachheit* (Simp)

$$\text{Simp} = \frac{1}{C_c} + \frac{R_c}{C_c}$$

Die endgültige Fitness-Funktion sieht dann wie folgt aus:

$$\text{Fitness} = w_1 \cdot (\text{Comp}_1 \cdot \text{Comp}_2) + w_2 \cdot \text{Simp}$$

4 Evolutionsprozeß

Darstellung des *THEN*-Teils

- *THEN*-Teil wird mit in die Chromosomen des Individuums codiert
- Zusammenfassen aller Individuen einer Population, die die gleiche Klasse vorhersagen
- den Regeln werden im Evolutionsprozeß die passenden Klassen zugeordnet

Hier wird die zweite Methode verwendet, da die Test-Datensätze nur wenig Klassen haben.



Discovering
Fuzzy
Classifiers

Titelseite
Einführung
Fuzzy Logik
Entwicklung

Evolutionsprozeß

Ergebnisse

Erkennungsmethode



Discovering Fuzzy Classifiers

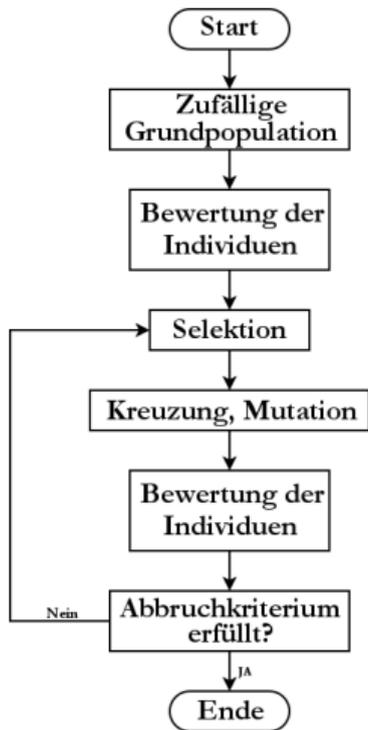
Titelseite
Einführung
Fuzzy Logik
Entwicklung

Evolutionsprozeß

Ergebnisse

- Die Bestimmung der Klasse wird durch das Produkt aus Konfidenzkoeffizient der Regel und der Vorhersagegenauigkeit der Regel festgelegt.
- Der Grenzwert zur erfolgreichen Erkennung wird auf 0,5 festgelegt.
- Werden alle Klassifikatoren betrachtet, dann legt die zutreffendste Regeln die Klasse fest.

Ablauf



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionsprozeß

Egebnisse

5 Ergebnisse

Um die Performance des vorgestellten Verfahrens (System1) zu ermitteln, werden die Test-Datensätze *IRIS*, *WINE* und *VOTE* verwendet.

Datensätze	Datensatz Eigenschaften		
	Größe	Klassen	Attribute
WINE	178	3	14
IRIS	150	3	4
VOTE	435	2	16



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionenprozeß

Ergebnisse

Testparameter

- numerische Attribute werden normalisiert, damit sie zwischen 0 und 1 liegen
- Zugehörigkeitsfunktion wird als 3-Dreieckfunktion definiert
- Mutationsrate/Wettkampfgröße für *IRIS* (0,5/2), *WINE* (0,9/4) und *VOTE* (0,65/4)
- Populationsgröße: 200 Individuen
- Anzahl Iterationen: 200
- Parameter für die Fitness-Funktion: $w_1 = 0,999$ und $w_2 = 0,001$



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionprozeß

Ergebnisse



Zum Vergleich werden zwei weitere evolutionäre Systeme ausgewertet:

- System2 verwendet komplette Ausdrucks-Bäume
- System3 setzt auf Co-Evolution

Erkennungsraten der verschiedenen Systeme:

System	IRIS	WINE	VOTE
System1	97,11	94,06	95,33
System2	94,84	92,22	95,42
System2 : DNF	93,3	90,55	95,43
System3	95,3	–	–

Fazit

Die vorgestellte Methode zur Bestimmung von Fuzzy Klassifikatoren liefert gute Ergebnisse und übertrifft damit sogar vergleichbare Systeme.

Ausblick:

- Verbesserung der Fitness-Funktion
- Berechnung der Konfusionsmatrix durch *Aggregations* Funktionen
- Fortschritte in der binären Klassifikation durch Co-Evolution



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionprozeß

Ergebnisse



Discovering Fuzzy Classifiers

Titelseite

Einführung

Fuzzy Logik

Entwicklung

Evolutionprozeß

Ergebnisse

Danke

Diese Präsentation wurde mit \LaTeX erstellt.