

Fuzzy Pattern Trees: Ein alternativer Ansatz zur Fuzzy-Modellierung

Fuzzy Pattern Trees

Robin Senge, Thomas Fober, Philipps-Universität Marburg,
Maryam Nasiri, Universität Siegen,
Eyke Hüllermeier, Philipps-Universität Marburg

Zusammenfassung Dieser Beitrag stellt einen neuartigen Ansatz zur Modellierung von Fuzzy-Systemen vor, die so genannten Fuzzy Pattern Trees. Deren Motivation leitet sich vor allem aus vermeintlichen Schwächen klassischer regelbasierter Architekturen ab, die dieses Gebiet nach wie vor dominieren. Durch eine hierarchische, modulare Struktur sowie die Verwendung eines breiten Spektrums von (nichtlinearen) Aggregationsoperatoren ist ein Fuzzy Pattern Tree in der Lage, funktionale Abhängigkeiten zwischen Ein- und Ausgabegrößen auf sehr flexible Art und Weise und in sehr kompakter Form darzustellen. Hierdurch kann ein guter Kompromiss zwischen Modellgenauigkeit und Verständlichkeit erzielt werden. Neben der Erläuterung der wesentlichen Modellierungskonzepte geht der Beitrag speziell auf die Probleme der datenbasierten Modellanpassung sowie des automatisierten Lernens der Modellstruktur

ein. Illustriert werden die Modelle anhand eines Färbeprozesses aus der Textilindustrie. ►► **Summary** This paper advocates a novel approach to fuzzy systems modeling called Fuzzy Pattern Trees. The latter is largely motivated by alleged disadvantages of rule-based system architectures that still dominate this field. Due to its hierarchical, modular structure and the use of different types of (nonlinear) aggregation operators, a Fuzzy Pattern Tree has the ability to represent functional dependencies in a more flexible and more compact way, thereby offering a reasonable balance between accuracy and model transparency. Apart from giving an overview of the main modeling concepts, we specifically address the problems of data-driven model calibration and FPT structure learning. The involved model classes get illustrated by a coloring application from the textile industry.

Schlagwörter Fuzzy Logik, Systemmodellierung, Regelsysteme, Maschinelles Lernen, Polyesterfärbung ►►
Keywords Fuzzy logic, systems modeling, rule-based systems, machine learning, polyester dyeing

1 Einleitung

Unter Fuzzy-Systemen versteht man formale Modelle, die unter Verwendung mathematischer Konzepte der „unscharfen Logik“ (Fuzzy Logik) definiert werden. Modelle dieser Art finden vielfältige Anwendungen, beispielsweise in der Kontrolltheorie, wobei regelbasierte Architekturen nach wie vor den größten Bekanntheitsgrad besitzen.

Fuzzy-Systeme realisieren eine Schnittstelle zwischen Mensch und Maschine: Mathematische Konzepte wie unscharfe Mengen (Fuzzy Sets) sowie verallgemeinerte logische Operatoren ermöglichen es, vage kognitive Konzepte und sprachliche Begriffe wie „hohe Temperatur“

adäquat zu formalisieren und das Schlussfolgern mit solchen Begriffen nachzubilden. Auf diese Art und Weise gelingt es, menschliches Expertenwissen einer maschinellen Verarbeitung zugänglich zu machen [15]. Die besondere Bedeutung von regelbasierten Methoden erklärt sich dadurch, dass es Experten häufig leicht fällt, ihr Wissen in Form von WENN-DANN Regeln zu beschreiben, die typischerweise die Zustände einer Reihe von Eingabevariablen mit den Werten einer oder mehrerer Ausgabevariablen in Beziehung setzen.

Der regelbasierte Ansatz ist zu einer Art Paradigma geworden und wird unseres Erachtens nicht kritisch genug in Frage gestellt. Tatsächlich besitzt dieser Ansatz,

zumindest in seiner ursprünglichen Konzeption, neben unbestrittenen Vorzügen auch eine Reihe potentieller Nachteile:

- Regeln sind rein lokale Einheiten, die immer nur einen kleinen Teil des Eingaberaums abdecken. Um einen globalen Zusammenhang zwischen Ein- und Ausgaben zu beschreiben, benötigt man häufig eine sehr große Zahl solcher Regeln; tatsächlich kann man ein exponentielles Wachstum dieser Zahl in der Dimensionalität der Eingabe nicht nur theoretisch begründen sondern auch praktisch beobachten.
- Einzelne Regeln entsprechen achsenparallelen Rechtecken im Eingaberaum und sind von ihrer Struktur her nicht sonderlich flexibel. Um einen Zusammenhang ausreichend genau zu beschreiben, muss man ihn häufig durch eine Vielzahl solcher „Boxen“ approximieren.
- Die mit dieser Problematik einhergehende Beschreibungskomplexität geht auf Kosten der Interpretierbarkeit, die im Kontext der Fuzzy-Modellierung einen hohen Stellenwert besitzt: Selbst wenn ein Experte einzelne Regeln verstehen kann, wird er das Modell als Ganzes kaum erfassen können. Neben der reinen Größe des Modells spielt hierbei der logische Inferenzmechanismus eine wichtige Rolle, der dazu führt, dass die Regeln auf komplexe Art und Weise miteinander interagieren.
- Verschärft werden die genannten Probleme noch dann, wenn die Modelle nicht vom Experten selbst vorgegeben, sondern aus beobachteten Daten extrahiert werden; dieser *datenbasierte* Ansatz gewinnt gegenüber dem *wissensbasierten* Ansatz zusehends an Bedeutung. Beispielsweise werden automatisch extrahierte unscharfe Mengen nicht unbedingt mit sprachlichen Begriffen assoziiert werden können, die der kognitiven Konzeption des Experten entsprechen.

Einschränkend erwähnt werden sollte an dieser Stelle, dass sich die obigen Kritikpunkte im Wesentlichen auf einfache Fuzzy-Systeme wie Mamdani Modelle [11] beziehen, so wie sie in der Praxis meistens verwendet werden, und dass wir unter dem Begriff „Regelsystem“ lediglich klassische Regelsysteme dieser Art verstehen. In der Literatur wurden diverse Erweiterungen vorgeschlagen, inklusive hierarchischer Fuzzy-Systeme [20], die einige der oben genannten Nachteile vermeiden (dafür aber neue mit sich bringen). Eine detaillierte Diskussion solcher Erweiterungen kann im Rahmen dieses Artikels allerdings nicht geleistet werden.

Wesentlich zurückzuführen sind die diskutierten Schwächen auf die *flache* Struktur klassischer Regelsysteme, gepaart mit der Einschränkung auf recht einfache Erweiterungen der Konjunktion und Disjunktion als logische Operatoren. Die Modellklasse der Fuzzy Pattern Trees unterscheidet sich genau in diesen Punkten: Ein Fuzzy Pattern Tree (FPT) ist eine modulare, *hierarchische* Struktur und verfügt über ein breites Spektrum an Aggregationsoperatoren.

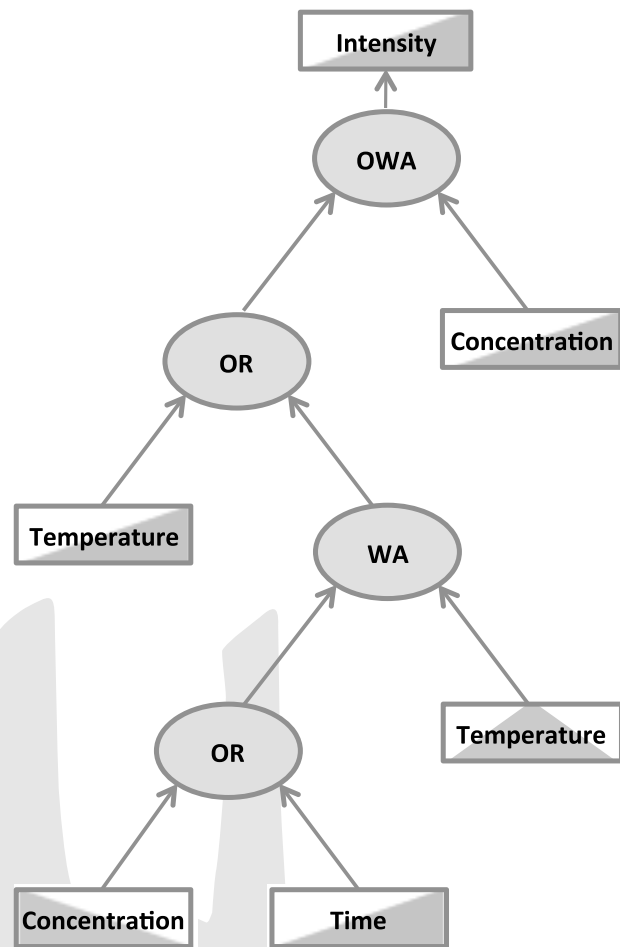


Bild 1 Beispiel eines FPT zur Modellierung der Farbintensität im Rahmen eines Färbeprozesses (vgl. Abschnitt 4).

Die charakteristische Eigenschaft des hierarchischen Ansatzes, nämlich die (rekursive) Zerlegung eines Problems in einfachere Teilprobleme mit anschließender Kombination der entsprechenden Lösungen, scheint eine zentrale Voraussetzung für die Beherrschbarkeit komplexer Systeme zu sein. Sie ermöglicht darüber hinaus das Betrachten des Systems auf unterschiedlichen Abstraktionsebenen.

Bild 1 zeigt ein (vereinfachtes) Beispiel eines FPT zur Modellierung der Farbintensität von Textilien im Rahmen eines Färbeprozesses (vgl. Abschnitt 4). Das Modell spezifiziert unscharfe Bedingungen für eine „hohe Intensität“ und ist so zu lesen, dass hierfür zwei Bedingungen gleichzeitig erfüllt sein müssen: eine hohe Farbkonzentration (im Färbebad) sowie ein weiteres Kriterium. Die erste Bedingung ist ein *Basiskriterium*, das durch eine unscharfe Teilmenge des Wertebereichs möglicher Konzentrationen modelliert wird. Die zweite Bedingung ist ein *komplexes Kriterium*, das wiederum in Teilkriterien zerlegt wird: Die Temperatur (des Färbebades) ist hoch oder eine weitere Bedingung ist erfüllt, die nochmals in zwei Teilkriterien zerfällt und diese mithilfe eines Mittelwertoperators kombiniert.

Die Modellklasse der Fuzzy Pattern Trees wird im nächsten Kapitel detaillierter erläutert. Kapitel 3 ist der datenbasierten Modellbildung gewidmet, die sich entweder auf die Identifikation von Parametern bei gegebener Modellstruktur beschränkt oder aber das Lernen dieser Struktur selbst beinhalten kann. In Kapitel 4 wird der FPT Ansatz anhand einer konkreten Fallstudie erläutert.

2 Fuzzy Pattern Trees

Ein FPT realisiert eine Abbildung von einem Eingaberaum $\mathbb{X} = \mathbb{X}_1 \times \dots \times \mathbb{X}_m$ in einen Ausgaberaum \mathbb{Y} . Eingaben sind also Vektoren $\mathbf{x} = (x_1, \dots, x_m)$, wobei die Komponente x_i Werte in \mathbb{X}_i annimmt; obwohl nicht zwingend notwendig, werden wir im Folgenden annehmen, dass alle Wertebereiche \mathbb{X}_i kontinuierlich (also Teilmengen von \mathbb{R}) sind. Für die Ausgabe werden wir zwei Fälle betrachten, den binären mit $\mathbb{Y} = \{0, 1\}$ und den kontinuierlichen mit $\mathbb{Y} = \mathbb{R}$ oder $\mathbb{Y} = [a, b] \subset \mathbb{R}$.

Ein FPT ist ein Binärbaum mit folgenden Eigenschaften:

- Jeder innere Knoten N_i ist mit einem (binären) Aggregationsoperator $A_i: [0, 1]^2 \rightarrow [0, 1]$ assoziiert.
- Die Wurzel N_0 ist zusätzlich mit einer Funktion $\nu: [0, 1] \rightarrow \mathbb{Y}$ assoziiert (Defuzzifizier).
- Jeder Blattknoten L_j ist mit einer Zugehörigkeitsfunktion $\mu_j: \mathbb{X}_{v(j)} \rightarrow [0, 1]$ assoziiert (Fuzzifizier).

Die Auswertung eines FPT für eine Eingabe $\mathbf{x} = (x_1, \dots, x_m) \in \mathbb{X}$ erfolgt rekursiv, und zwar gemäß der folgenden Regeln:

- Die Wurzel berechnet $\nu(A_0(y_l, y_r))$ und definiert damit die Ausgabe des Baumes; y_l und y_r sind die Ausgaben des linken bzw. rechten Nachfolgeknotens.
- Jeder innere Knoten N_i berechnet $A_i(y_l, y_r)$ als Ausgabe, wobei y_l und y_r wiederum die Ausgaben des linken bzw. rechten Nachfolgers sind.
- Jedes Blatt berechnet $\mu_j(x_{v(j)})$.

In einem FPT werden die Eingaben x_i also von unten (Blattknoten) nach oben (Wurzel) propagiert. Fasst man die μ_j und die A_i zu jeweils einem „Fuzzifizier“ μ und einer Aggregation A zusammen, dann kann man die durch einen FPT definierte Abbildung auch etwas salopp als $\nu(A(\mu(\mathbf{x})))$ schreiben. Wir werden die Komponenten eines FPT im Folgenden etwas genauer diskutieren.

2.1 Fuzzifizierung und Defuzzifizierung

Jeder der Wertebereiche \mathbb{X}_i wird mithilfe einer Fuzzy-Partition $\{F_{i,1}, \dots, F_{i,n_i}\}$ diskretisiert, wobei

$$F_{i,j}: \mathbb{X}_i \rightarrow [0, 1] \quad (j = 1, \dots, n_i) \quad (1)$$

und $F_{i,1}(x_i) + F_{i,2}(x_i) + \dots + F_{i,n_i}(x_i) > 0$ für alle $x_i \in \mathbb{X}_i$. Oft werden die unscharfen Mengen (1) mit sprachlichen Begriffen wie z. B. ‚groß‘, ‚mittel‘ oder ‚klein‘ markiert.

Die Definition der Funktion $\nu: [0, 1] \rightarrow \mathbb{Y}$ bzw. deren Inverse $\nu^{-1}: \mathbb{Y} \rightarrow [0, 1]$ hängen von \mathbb{Y} ab. Für $\mathbb{Y} = [a, b]$ kann man im einfachsten Fall $\nu^{-1}(z) = (z-a)/(b-a)$ setzen. Für den Fall $\mathbb{Y} = \mathbb{R}$ bietet sich, ähnlich wie bei

neuronalen Netzen, eine sigmoide Transformation an: $\nu^{-1}(z) = (1 + \exp(-\alpha \cdot z))^{-1}$. Im Fall einer binären Ausgabe setzen wir $\nu(z) = 1$ falls $z > \theta$ und $\nu(z) = 0$ sonst, wobei $0 < \theta < 1$ ein Schwellwert ist (ν^{-1} ist hier per Definition die identische Abbildung).

2.2 Aggregationsfunktionen

Aggregationsfunktionen sind Funktionen mit speziellen Eigenschaften, wie etwa Monotonie und Kommutativität; sie sind in der Regel zweistellig aber gleichzeitig assoziativ, lassen sich also kanonisch auf n -stellige Funktionen erweitern. In der Literatur unterscheidet man drei Klassen von Aggregationsfunktionen: konjunktive, kompensatorische und disjunktive [8]. Definiert man eine Ordnung \leq auf den Aggregationsfunktionen gemäß der klassischen (punktweisen) Ordnung auf Funktionen, dann sind konjunktive Aggregationen A solche mit $A \leq \min$, kompensatorische (Mittelwerte) solche mit $\min \leq A \leq \max$, und disjunktive solche mit $\max \leq A$.

Gilt $A \leq A'$ für zwei Aggregatoren A und A' , so ist A gewissermaßen „strenger“ als A' : Bei gleichen Eingaben (z. B. Grad der Erfülltheit zweier Kriterien) liefert A' immer mindestens einen Wert in Höhe des Wertes von A . Die konjunktive bzw. disjunktive Aggregation verallgemeinert das logische UND bzw. ODER, erstere ist also strenger als letztere. Zwischen diesen beiden Klassen liegen die Mittelwerte, die einen Kompromiss zwischen beiden Eingaben erzeugen und die „Lücke“ zwischen dem Minimum und dem Maximum füllen (siehe Bild 2).

Wie in der Fuzzy Logik üblich, verwenden wir zur Modellierung der verallgemeinerten Konjunktion die Klasse der so genannten T-Normen (engl.: *triangular norms*), also der monotonen, assoziativen und kommutativen Abbildungen $[0, 1]^2 \rightarrow [0, 1]$ mit 1 als neutralem Element [10]. Entsprechend dient die duale Klasse der T-Conormen (die sich lediglich durch das neutrale Element 0 statt 1 unterscheidet) zur Modellierung der logischen Disjunktion.

Besonders attraktiv aus Sicht der Modellierung ist die Verwendung von *parametrisierten* Familien solcher Normen, denn sie erlauben durch die Wahl eines Parameters eine kontinuierliche Variation der „Strenge“ der Aggregation. Als Beispiel erwähnen wir die Familie der Dubois-Prade T-Normen bzw. T-Conormen [5]:

$$\begin{aligned} T_\lambda(a, b) &= \frac{ab}{\max\{a, b, \lambda\}} \\ \perp_\lambda(a, b) &= \frac{a + b - ab - \min\{a, b, 1 - \lambda\}}{\max\{1 - a, 1 - b, \lambda\}} \end{aligned} \quad (2)$$



Bild 2 Spektrum von Aggregationsfunktionen.

Der Parameter λ nimmt Werte in $[0, 1]$ an und erzeugt für die Randwerte 0 und 1 die klassischen T-Normen (T-Conormen) Minimum (Maximum) und Produkt (algebraische Summe).

Als Durchschnittsoperator verwenden wir das so genannte *Choquet Integral* (CI), das die Lücke zwischen Minimum und Maximum schließt [8]. Um diesen Operator formal korrekt zu definieren, müsste man ihn als ein Integral einer Funktion bzgl. eines geeigneten zugrunde liegenden (nicht-additiven) Maßes schreiben. In dem für uns relevanten (diskreten) Fall mit lediglich zwei Argumenten kann man jedoch zeigen, dass sich das Choquet Integral zu folgendem Ausdruck vereinfacht:

$$CI(u, v) = \begin{cases} (1-\beta)u + \beta v & \text{falls } u \leq v \\ \alpha u + (1-\alpha)v & \text{falls } u > v \end{cases}, \quad (3)$$

wobei $\alpha, \beta \in [0, 1]$. Man beachte die folgenden Spezialfälle: Für $(\alpha, \beta) = (0, 0)$ bzw. $(\alpha, \beta) = (1, 1)$ ist $CI = \min$ bzw. $CI = \max$, für $\beta = 1 - \alpha$ ergibt sich das *gewichtete arithmetische Mittel*, und für $\beta = \alpha$ erhalten wir den *Ordered Weighted Average* (OWA) Operator [22].

2.3 Modellierung und Interpretation

Die unscharfen Mengen an den Blättern eines FPT, angewandt auf die entsprechenden Variablen, können als (unäre) Prädikate aufgefasst und ggf. sprachlich ausgedrückt werden, z. B. „Temperatur ist mittel“. Ebenso definiert die mit der Wurzel assoziierte Zugehörigkeitsfunktion ein unscharfes Prädikat, das wir bislang generisch als „die Ausgabe ist hoch“ angenommen haben. Der innere Teil des FPT definiert ein Kriterium für die Erfüllung dieses Prädikates, das sich rekursiv durch die Aggregation von Teilkriterien zusammensetzt. Letztlich spezifiziert ein FPT also (unscharfe) Bedingungen an die Eingabevariablen, die dazu führen, dass „die Ausgabe hoch ist“, oder allgemeiner, dass das unscharfe Prädikat auf dem Ausgabeattribut zu einem hohen Grad erfüllt ist.

Ein menschlicher Experte wird ein solches Modell typischerweise „top-down“ entwerfen, also durch das rekursive Zerlegen von Kriterien in Teilkriterien. Die Vorzüge eines solchen Ansatzes sind auch in vielen anderen Gebieten bekannt; beispielhaft genannt sei hier der *Analytic Hierarchy Process* im Bereich der Entscheidungsunterstützung [16]. Für jedes Kriterium muss zudem festgelegt werden, auf welche Weise die Teilkriterien aggregiert werden sollen (konjunktiv, kompensatorisch oder disjunktiv) und wie die entsprechenden Aggregationen parametrisiert sind (z. B. wie streng eine Konjunktion ist oder welche Gewichte die Teilkriterien beim arithmetische Mittel haben).

Bei der Interpretation eines FPT kann ebenfalls die hierarchische Struktur ausgenutzt werden, die eine gewisse Modularität des Ansatzes unterstützt: Jedes Kriterium wird „erklärt“ durch die jeweiligen Teilkriterien und deren Aggregation. Wiederum kann diese Erklärung

(bei Bedarf) rekursiv fortgesetzt werden, und zwar völlig unabhängig vom Rest des Modells. Auf diese Art und Weise kann nicht nur das Modell als solches analysiert, sondern auch eine konkrete Prädiktion nachvollzogen werden: Die numerische Ausgabe eines Knotens erklärt sich durch dessen Eingaben, also den Grad der Erfüllung der Teilkriterien, sowie deren Verknüpfung.

2.4 Vergleich mit Regelsystemen

In [17] wurde eine Erweiterung des FPT Ansatzes vorgestellt, die ein Modell mithilfe eines Ensembles von FPTs beschreibt. Statt in unserem Beispiel z. B. lediglich einen Baum zu verwenden, der die Bedingungen für „hohe Intensität“ spezifiziert (und automatisch eine niedrige Intensität impliziert, falls diese Bedingungen nicht erfüllt sind), könnte man etwa weitere Bäume für „mittlere Intensität“ und „geringe Intensität“ definieren. Für die endgültige Prädiktion ist dann eine weitere Aggregation der Ausgaben aller Bäume notwendig.

Dieser Ansatz kommt den klassischen regelbasierten Fuzzy-Systemen recht nahe, wobei jeder FPT mit einer Regel bzw. der Aggregation mehrerer Regeln mit der gleichen oder ähnlichen Konsequenzen assoziiert werden kann; die Aggregation der Ausgaben der Bäume entspricht dann im Wesentlichen dem Schritt der Defuzzifizierung in Regelsystemen. FPTs haben jedoch eine Reihe von Vorteilen, speziell die folgenden.

Durch Verwendung einer hierarchischen anstatt einer flachen Modellstruktur können Modelle evtl. kompakter formuliert werden. Dies liegt zum einen daran, dass innerhalb der Klasse der FPTs deutlich mehr äquivalente Umformungen möglich sind als in der Klasse der Regelsysteme; beispielsweise kann man den Ausdruck $\max\{\min\{A, B\}, \min\{A, C\}\}$ als Disjunktion zweier Regeln mit Antezedenzen $A \wedge B$ bzw. $A \wedge C$ betrachten, den logisch äquivalenten Ausdruck $\min\{A, \max\{B, C\}\}$ jedoch nicht. Zudem wird die Klasse der darstellbaren analytischen Ausdrücke bzw. der „Grad der Nichtlinearität“ durch die Möglichkeit der Rekursion extrem vergrößert. Unter Verwendung des Produkts als spezielle T-Norm lassen sich mit einem FPT der Höhe k z. B. sämtliche Monome vom Grad k darstellen, selbst wenn alle Zugehörigkeitsfunktionen nur linear sind.

In diesem Zusammenhang ist es interessant zu erwähnen, dass ähnliche Vorteile von „tiefen“ gegenüber „flachen“ Strukturen auch in anderen Bereichen beobachtet wurden, aktuell beispielsweise im Gebiet des *deep learning* [2]: Obwohl künstliche neuronale Netze mit nur einer mittleren Schicht von Neuronen universelle Approximationseigenschaften besitzen, kann es sein, dass die praktische Umsetzung dieser theoretischen Eigenschaft extrem viele Neuronen in dieser Zwischenschicht erfordert. Die gleiche Approximationsgüte kann oft mit insgesamt deutlich weniger Neuronen erreicht werden, wenn man diese auf mehrere Schichten verteilt und entsprechend vernetzt. Tatsächlich ist es auch bei Fuzzy-Regelsystemen so, dass deren universelle Appro-

ximationseigenschaften typischerweise mit einer beliebig großen Anzahl von Regeln erkaufte werden.

Der zweite wichtige Vorteil von FPTs ist die Verwendung einer größeren Klasse von Aggregationsfunktionen, speziell den Mittelwertoperatoren. Das Beispiel der gewichteten Summe macht dies recht deutlich: Dieser Operator erlaubt es, lineare oder stückweise lineare Funktionen (unter Verwendung nur eines Knotens) zu beschreiben, wohingegen die Approximation solcher Funktionen mit (achsenparallelen) Regeln sehr aufwändig ist.

Die Vermutung liegt nahe, dass die Interpretierbarkeit eines FPT unter der Verwendung unterschiedlicher Operatoren leiden könnte. Dem mag man widersprechen: Jeder Operator ist an sich gut interpretierbar, und kann aufgrund der Modularität eines FPT tatsächlich isoliert vom Rest des Modells betrachtet werden.

3 Datenbasierte Modellbildung

In vielen Anwendungen liegen (möglicherweise veräuschte) Daten über das Ein-/Ausgabeverhalten des zu modellierenden Systems vor:

$$\mathcal{T} = \{(\mathbf{x}^{(t)}, y^{(t)})\}_{t=1}^n \subset \mathbb{X} \times \mathbb{Y}, \quad (4)$$

wobei $\mathbf{x}^{(t)}$ einen Eingabevektor und $y^{(t)}$ die hierzu beobachtete Ausgabe bezeichnet. Diese Daten können genutzt werden, um die wissensbasierte Modellbildung eines Experten zu ersetzen oder zumindest zu unterstützen. Tatsächlich sind Experten häufig in der Lage, die Struktur eines Modells zu beschreiben, inklusive der Art der gewünschten Aggregation (konjunktiv, kompensatorisch, disjunktiv), tun sich aber schwer mit der Parametrisierung der Operatoren. Eine nahe liegende Idee besteht dann darin, die Parameter optimal an die Daten anzupassen – wir bezeichnen dies als *Modellkalibrierung*.

Ist keinerlei Vorwissen vorhanden, so kann man versuchen, nicht nur die Parameter sondern auch die Modellstruktur aus den Daten zu extrahieren. Dieses Problem fällt in den Bereich des maschinellen Lernens und ist vergleichbar etwa zum Lernen von Bayesschen Netzen, die ebenfalls aus einem qualitativen (nämlich der Graphstruktur) und einem quantitativen Teil (den bedingten Wahrscheinlichkeitsverteilungen) bestehen.

3.1 Modellkalibrierung

Wir gehen davon aus, dass die Modellstruktur bis auf die Parametrisierung festgelegt ist. Alle Modellparameter können zu einem Parametervektor θ zusammengefasst werden, der folgende Einträge enthält:

- Etwaige Parameter zur Spezifikation der Zugehörigkeitsfunktion (1) an den Blattknoten.
- Die Parameter der Aggregatoren an den inneren Knoten (beispielsweise λ im Fall von (2) oder α und β im Fall von (3)).

Sei $\mathcal{M}[\theta]$ das mit θ parametrisierte Modell und $\mathcal{M}[\theta](\mathbf{x})$ die Ausgabe dieses Modells für den Eingabevektor \mathbf{x} . Um

eine optimale Parametrisierung θ^* zu finden, wird auf die Daten (4) zurückgegriffen und folgendes Optimierungsproblem gelöst:

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^n L(\mathcal{M}[\theta](x^{(i)}), y^{(i)}),$$

wobei $L(\cdot)$ eine geeignete Fehlerfunktion ist, wie z. B. die quadrierte Abweichung zwischen Prädiktion und Zielwert:

$$L(\hat{y}, y) = (\hat{y} - y)^2 \quad (5)$$

Dieses Optimierungsproblem wird in der Regel hochgradig nichtlinear und multimodal sein. Zudem ist es nicht möglich, eine geschlossene, differenzierbare Formulierung zu finden. In [23] wurde daher die Verwendung von Evolutionsstrategien [3] vorgeschlagen, einer speziellen Form der evolutionären Optimierung, die sich für dieses Problem besonders gut zu eignen scheint.

3.2 Modellinduktion

Wir wenden uns nun dem Problem der Modellinduktion zu, worunter wir das gleichzeitige Lernen der Struktur und der Parameter eines FPT verstehen. Das Lernen der Modellstruktur erfordert eine systematische Suche im Raum der FPTs, die wir mithilfe eines heuristischen Suchverfahrens realisieren. Das Problem der Parameterschätzung wird in diesen Prozess eingebettet werden. Bevor wir Details dieses Ansatzes beschreiben, betrachten wir als ersten Schritt unseres Lernverfahrens die „Fuzzifizierung“ der Eingabevariablen durch die Definition der Fuzzy-Partitionen (1).

Definition der Fuzzy-Partitionen

Zur Spezifikation dieser Partitionen gibt es diverse Möglichkeiten. In der derzeitigen Implementierung beschränken wir uns auf die folgende, recht einfache Lösung. Für jeden Wertebereich \mathbb{X}_i wird eine Partition bestehend aus drei unscharfen Mengen HIGH, MEDIUM und LOW erzeugt. Die Zugehörigkeitsfunktion der ersten Menge ist gegeben durch

$$F_{i,high}(x) = \left[\frac{x - a_i}{b_i - a_i} \right]_0^1,$$

wobei a_i bzw. b_i der kleinste bzw. größte beobachtete Wert für x_i ist und $[z]_0^1 = \max\{0, \min\{1, z\}\}$. Die zweite unscharfe Menge ist definiert durch $F_{i,low} = 1 - F_{i,high}$. Da alle Operatoren an den inneren Knoten eines FPT monoton steigend sind, kann mit $F_{i,high}$ bzw. $F_{i,low}$ eine monoton steigende bzw. monoton fallende Abhängigkeit der Ausgabe y von der Variablen x_i modelliert werden.

Um auch nicht-monotone Abhängigkeiten erfassen zu können, wird eine dritte unscharfe Menge verwendet. Für MEDIUM wird ein „Idealwert“ $c \in [a, b]$ bestimmt, und $F_{i,med}$ dann durch eine Dreiecksfunktion mit Kern $\{c\}$ und Träger (a, b) definiert. Der Parameter c wird da-

bei so gewählt, dass die Pearson-Korrelation zwischen den Zugehörigkeitsgraden $F_{i,med}(x_i^{(t)})$ und den Zielwerten $v^{-1}(y^{(t)})$ auf den Trainingsdaten (4) maximiert wird.

Top-Down Induktion eines FPT

Das Grundgerüst unseres Lernverfahrens basiert auf einem *beam search*, einer heuristischen Suchstrategie, die eine Menge von Kandidatenlösungen (in unserem Fall FPTs) verwaltet. Diese Menge – genannt *beam* – hat eine feste Größe (typische Werte sind 5 oder 10) und speichert die bislang besten gefundenen Lösungen. Konkret wird das folgende iterative Verfahren implementiert:

1. *Initialisierung*: Der beam C_0 wird mit allen *primitiven Bäumen* initialisiert, die jeweils nur aus einem Knoten (markiert mit einer unscharfen Menge $F_{i,j}$) bestehen.
2. *Expansion der Kandidaten*: Ausgehend von den aktuellen Lösungen C_i wird eine neue Kandidatenmenge T_{i+1} konstruiert. Hierzu werden die Modelle $\mathcal{M} \in C_i$ expandiert, und zwar wie folgt:
 - a) Wähle einen der Blattknoten $F_{i,j}$ aus \mathcal{M} .
 - b) Erzeuge einen neuen Teilbaum $S[A_\theta]$, der aus drei Knoten besteht: Einem Aggregator A_θ als Wurzel, $F_{i,j}$ als linkes Blatt und einem weiteren primitiven Baum $F_{i',j'}$ (mit $i \neq i'$ oder $j \neq j'$) als rechtes Blatt.
 - c) Erzeuge einen neuen Kandidatenbaum \mathcal{M}^* , indem der Eingabeknoten $F_{i,j}$ in \mathcal{M} durch den Teilbaum $S[A_\theta]$ ersetzt wird.
 - d) Optimierte den Parameter θ und füge \mathcal{M}^* zu T_{i+1} hinzu. Zur Optimierung kann ein beliebiges direktes Suchverfahren eingesetzt werden; im konkreten Fall wurde eine Evolutionsstrategie verwendet [3].
 - e) Iteriere die ersten beiden Schritte für alle Kombinationen aus zur Verfügung stehenden Eingabeknoten und Aggregationsoperatoren.
3. *Bewertung der Kandidaten*: Bewerte die neuen Kandidaten aus T_{i+1} (anhand des mittleren Fehlers auf den Trainingsdaten) und selektiere die besten c für den neuen beam C_{i+1} .
4. *Abbruchbedingung*: Enthält C_{i+1} nicht wenigstens ein Modell, das signifikant besser ist als das beste Modell in C_i , dann breche die Suche ab; ansonsten fahre fort mit Schritt 2.

Die Kandidatenbäume wachsen also nach unten, da immer wieder Blätter des Baumes durch kleine Teilbäume ersetzt werden; wir sprechen daher von einer „top-down“ Induktion.

Die Optimierung von θ in Schritt 2(d) hängt vom Typ der Aggregationsfunktion A ab. Eine lokal optimale Parametrisierung des Choquet Integrals (3) kann z. B. durch Lösen eines Regressionsproblems (unter Nebenbedingungen) gefunden werden. Generell ist es auch möglich, die Kalibrierung aus Abschnitt 3.1 auf den lokalen Teilbaum $S[A_\theta]$ anzuwenden. Auch die Abbruchbedingung bedürfte weiterer Erklärung, zumal sie wesentlichen Einfluss auf die Modellkomplexität und damit auf die Generalisierungsfähigkeit (Vermeidung der Überanpassung an die Daten) hat. Da eine detaillierte Diskussion solcher

technischen Details den Rahmen dieses Artikels sprengen würde, verweisen wir an dieser Stelle auf [18].

4 Fallstudie: Polyesterfärbung

In diesem Abschnitt stellen wir eine Fallstudie vor, die aufgrund der überschaubaren Dimensionalität des Eingaberaumes recht handhabbar, aber dennoch von unmittlbarer praktischer (industrieller) Relevanz ist.

Die wichtigste künstlich erzeugte Faser ist das Polyethylenterephthalat (PET), auch bekannt als Polyester. PET enthält Estergruppen ($-\text{CO}-\text{O}-$) in ihrer Hauptmolekularkette und wird in einem Schmelzspinnverfahren hergestellt. Estergruppen resultieren aus einer Reaktion zwischen bifunktionellen Carbonsäuren und bifunktionellen Alkoholen [7]. Die Abwesenheit einer reaktionsfähigen Gruppe, fähig mit anionischen und kationischen Färbemitteln eine Reaktion einzugehen, hat dazu geführt, dass die Färbung und Bedruckung von (nichtmodifizierten) PET Fasern nur mit Dispersionsfarbstoffen möglich ist [21]. Die kompakte Struktur der PET Fasern erschwert unter normalen Bedingungen zudem ein Durchdringen des Dispersionsfarbstoffs bis in die Faser. Aber auch die Dispersionsfarben an sich führen zu Problemen: Die in diesem Artikel betrachteten *high energy* Farbstoffe enthalten zwar polare Gruppen wie $-\text{NHR}$, $-\text{OH}$ und NH_2 , jedoch keinerlei ionische Gruppen, was dazu führt, dass solche Farbstoffe nur sehr schlecht wasserlöslich sind [9; 14].

Die Färbung von PET Fasern setzt daher bestimmte Bedingungen voraus, wie eine gewisse Temperatur des Färbeades (ca. 130 °C), trockene Hitze (190–220 °C) und die Verwendung von Färbebeschleunigern sowie einem bestimmten Träger im Färbebad [12].

Unser Ziel ist es, die Abhängigkeit der Farbintensität von der Färbungszeit, der Temperatur und Farbkonzentration zu modellieren. Für eine gegebene Einstellung dieser Parameter soll das Modell das Verhältnis K/S bestimmen, das grob gesagt den Anteil des absorbierten Lichts zum reflektierten Licht angibt und sich somit als Indikator für die Intensität der Färbung eignet [1; 6]. Die genannten Parameter haben wesentlichen Einfluss auf die Intensität und eignen sich als Einstellgrößen zwecks Kontrolle und Optimierung des Färbeprozesses.

4.1 Datenerzeugung und Modellierung

Zur Erzeugung eines Datensatzes wurde die *Polymat Farbmaschine AHIBA 1000* verwendet. Durch systematische Variation der Farbkonzentration (%) in {0,75, 1,5, 3, 4,5, 6}, der Temperatur des Färbeades (°C) in {100, 110, 115, 120, 125, 130} und der Zeit (Minuten) in {12, 24, 36, 48} wurde jeweils ein experimentelles Färbungsergebnis (gemessen in K/S) für jedes der insgesamt 120 Kombinationen der Eingabevariablen erzeugt. Alle anderen Parameter, wie der pH-Wert, wurden fixiert. Da insgesamt sieben Farben zum Einsatz gekommen sind, nämlich *Disperse Blue 266*, *Brown 1*, *Blue 56*, *Red 60*, *Yellow 7*, *Yellow 23* sowie ein Gemisch aller

sechs Grundfarben, stehen sieben Datensätze mit jeweils 120 Beispielen zur Verfügung [19].

Wir betrachten zwei alternative Ansätze zur Modellbildung, die uns einerseits den Vergleich zwischen wissenschaftlicher und datenbasierter Modellierung erlauben, und andererseits zwischen klassischen regelbasierten Methoden und unserem neuen FPT Ansatz.

Der wissenschaftliche Ansatz wurde unter Verwendung der bekannten Mamdani-Inferenz umgesetzt, ein Fuzzy-Inferenzsystem basierend auf linguistischen Regeln [11; 13]. Tatsächlich besteht in der Literatur Einigkeit darüber, dass sich Mamdani-Systeme sehr gut zum Zweck der wissenschaftlichen Modellierung eignen, und hinsichtlich ihrer Interpretierbarkeit Vorteile gegenüber anderen Fuzzy-Inferenzsystemen (wie etwa Takagi-Sugeno) besitzen [4].

Unter Nutzung der physikalischen und chemischen Eigenschaften der PET Fasern wurden durch den Experten für jede der Farben 8 Regeln formuliert. Diese Regeln haben folgende Form:

1. If (temperature is LOW) and (time is LOW) and (concentration is LOW), then (K/S is VERY LOW).
2. If (temperature is MEDIUM) and (concentration is HIGH), then (K/S is HIGH).
3. ...

Die involvierten unscharfen Mengen wurden als Gauss-Funktionen modelliert; für weitere Details sei auf [13] verwiesen.

Alternativ hierzu wurden für jeden der Datensätze FPTs rein datenbasiert erzeugt, und zwar mithilfe des in Abschnitt 3.2 vorgestellten Ansatzes.

4.2 Ergebnisse

Die erzeugten Modelle wurden anhand von zwei Kriterien bewertet: Genauigkeit (Prädiktionsgüte) und Verständlichkeit. Um die Prädiktionsgüte pro Datensatz zu ermitteln, wurde das jeweilige Mamdani-System auf alle 120 Beispielinstanzen angewendet, und die Ausgaben mit den Zielwerten verglichen. Für die FPTs, die ja aus den Daten erzeugt wurden, muss eine faire Auswertung auf unabhängigen Testdaten erfolgen. Daher wurde eine 5-fache Kreuzvalidierung durchgeführt: Die 120 Beispiele werden zufällig in 5 gleich große Teilmengen unterteilt, und jeder der Teile wird zum Testen des Modells verwendet, das auf den restlichen 4 Teilen gelernt wurde. Der gesamte Vorgang wurde 10 mal wiederholt und die Ergebnisse gemittelt.

Die durchschnittlichen RMSD-Werte für alle sieben Farben sind in Tabelle 1 zusammengefasst. Interessanterweise schneiden die FPTs durchweg signifikant besser ab als die Mamdani-Systeme. Dies ist insofern bemerkenswert, als dass letztere tatsächlich nach bestem Wissen von der dritten Co-Autorin dieses Beitrags erstellt wurden, die sich mit diesem Anwendungsproblem hervorragend auskennt.

Im Hinblick auf die Verständlichkeit schneiden die Mamdani-Systeme natürlich sehr gut ab, denn sie wur-

Tabelle 1 Fehler in Form von RMSD \pm Standardabweichung und Anzahl der Knoten des FPTs pro Farbe.

Farbe	Mamdani	FPT	
	RMSD	RMSD	#Knoten
Blue 266	3,3663	2,23 \pm 0,16	15
Brown 1	4,3129	2,68 \pm 0,13	9
Blue 56	3,7985	2,74 \pm 0,28	13
Red 60	4,6461	3,03 \pm 0,14	3
Yellow 7	4,9089	2,29 \pm 0,08	11
Yellow 23	4,1213	2,84 \pm 0,16	9
Mixture	3,9760	1,92 \pm 0,04	5

den ja vom Experten selbst entworfen. Aber auch die automatisch generierten FPTs wurden vom Experten als durchweg gut verständlich und plausibel bewertet. Der in Bild 1 gezeigte FPT ist einer der erzeugten Bäume, und gehört innerhalb dieser Studie bereits zu den komplexeren. Auch die Mamdani-Systeme sind von der Größe her überschaubar, was jedoch wesentlich der geringen Dimensionalität des Eingaberaums zu verdanken ist. So gesehen mögen 8 Regeln bei nur 3 Variablen bereits als recht viel erscheinen, und tatsächlich würde diese Zahl bei Berücksichtigung weiterer Einflussgrößen signifikant steigen.

5 Zusammenfassung

In diesem Artikel wurden Fuzzy Pattern Trees als ein neuartiger Ansatz zur Fuzzy-Modellierung vorgestellt und mit herkömmlichen regelbasierten Techniken verglichen. Neben der Beschreibung der wesentlichen Modellierungskonzepte wurden Methoden zur Unterstützung einer datenbasierten Modellbildung vorgestellt. Schließlich wurde das Potential dieses Ansatzes anhand eines Praxisbeispiels, der Färbung von Polyester, illustriert.

FPTs zeichnen sich unserer Meinung nach speziell durch ihre hierarchische Modellstruktur sowie der Verwendung eines breiten Spektrums von Aggregationsfunktionen aus. Daneben besitzen FPTs eine Reihe von weiteren interessanten Eigenschaften, die leider nur gestreift oder gar nicht erwähnt werden konnten; hierzu gehören die integrierte Attributauswahl, die Garantie spezieller Monotonieeigenschaften, der Umgang mit fehlenden oder unpräzisen Attributwerten, sowie die Möglichkeit einer kosteneffizienten Modellauswertung. Abermals verweisen wir daher auf die ergänzende Literatur [17; 18; 23].

Software zu Fuzzy Pattern Trees ist verfügbar unter: <http://www.uni-marburg.de/fb12/kebi/research/>

Literatur

- [1] E. Allen. *Optical Radiation Measurements*. Academic Press New York, 1980.
- [2] Y. Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009.
- [3] H.-G. Beyer and H.-P. Schwefel. Evolution strategies: A comprehensive introduction. *Natural Computing*, 1(1):3–52, 2002.

- [4] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena, editors. *Interpretability Issues in Fuzzy Modeling*. Studies in Fuzziness and Soft Computing. Springer-Verlag, Berlin, 2003.
- [5] D. Dubois and H. Prade. New results about properties and semantics of fuzzy set-theoretic operators. In P. P. Wang and S. K. Chang, editors, *Fuzzy Sets: Theory and Applications to Policy Analysis and Information Systems*, pages 59–75. Plenum Press, New York, 1980.
- [6] W. Baumann et al. Determination of relative colour strength and residual color difference by means of reflectance measurement. *J. Soc. Dyers and Colour*, 103:100–105, 1987.
- [7] J. Goorden Cook. *Handbook of Textile Fibers, Man Made Fiber*. Merrow Technical Library, 1984.
- [8] M. Grabisch, J. L. Marichal, R. Mesiar, and E. Pap. *Aggregation Functions*. Cambridge University Press, 2009.
- [9] A. Johnson. *The Theory of Coloration of Textiles*. Bardford: S.D.C. Pub., 2nd edition, 1989.
- [10] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, 2002.
- [11] E. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7:1–13, 1975.
- [12] R. W. Moncrieff. *Man-Made Fibers*. Newnes-Butterworths, 1971.
- [13] M. Nasiri, E. Hüllermeier, R. Senge, and E. Lughofer. Comparing methods for knowledge-driven and data-driven fuzzy modeling: A case study in textile industry. In *IFSA-2011, World Congress of the International Fuzzy Systems Association*, 2011.
- [14] D. M. Nunn. *The Dyeing of Synthetic Polymer and Acetate Fibers*. 1979.
- [15] K. M. Passino and S. Yurkovich. *Fuzzy Control*. Addison-Wesley, 1998.
- [16] T. L. Saaty. *The Analytic Hierarchy Process*. McGraw-Hill, 1980.
- [17] R. Senge and E. Hüllermeier. Pattern trees for regression and fuzzy systems modeling. In *2010 IEEE International Conference on Fuzzy Systems (FUZZ)*, 2010.
- [18] R. Senge and E. Hüllermeier. Top-down induction of fuzzy pattern trees. *IEEE Transactions on Fuzzy Systems*, 19 (2):241–252, 2011.
- [19] H. Tavanai, S. M. Taheri, and M. Nasiri. Modeling of colour yield in polyethylene terephthalate dyeing with statistical and fuzzy regression. *Iranian Polymer Journal*, 14(11):954–968, 2005.
- [20] V. Torra. A review on the construction of hierarchical fuzzy systems. *Int. Journal of Intelligent Systems*, 17(5):531–543, 2002.
- [21] E. R. Trotman. *Dyeing and Chemical Technology of Textile Fibers*. Charles Griffin, 1970.
- [22] R. R. Yager. On ordered weighted averaging aggregation operators in multi criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18(1):183–190, 1988.
- [23] Y. Yi, T. Fober, and E. Hüllermeier. Fuzzy operator trees for modeling rating functions. *International Journal of Computational Intelligence and Applications*, 8(4):413–428, 2009.

Manuskripteingang: 2. Mai 2012

Dipl.-Inf. Robin Senge ist wissenschaftlicher Mitarbeiter und Doktorand am Fachbereich Mathematik und Informatik der Philipps-Universität Marburg. Hauptarbeitsgebiete: Maschinelles Lernen und Knowledge Discovery mit Fuzzy Systemen.

Adresse: Philipps-Universität Marburg, FB Mathematik und Informatik, Biegestraße 10, 35032 Marburg,
E-Mail: senge@mathematik.uni-marburg.de

Dipl.-Inf. Thomas Fober hat an der Universität Dortmund Informatik mit Schwerpunkt *Computational Intelligence* studiert. Zurzeit ist er wissenschaftlicher Mitarbeiter am Fachbereich Mathematik und Informatik der Philipps-Universität Marburg und schließt seine Promotion auf dem Gebiet der strukturellen Bioinformatik ab. Weitere Interessensgebiete sind die (multikriterielle) Optimierung, Fuzzy Systeme, maschinelles Lernen und Data Mining, sowie die statistische Auswertung großer Datenmengen.

Adresse: Philipps-Universität Marburg, FB Mathematik und Informatik, Biegestraße 10, 35032 Marburg,
E-Mail: thomas@mathematik.uni-marburg.de

M.Sc. Maryam Nasiri ist Doktorandin am Fachbereich Informatik und Elektrotechnik der Universität Siegen. Hauptarbeitsgebiete: Anwendungen von Fuzzy Logik im Bereich Textiltechnik.

Adresse: Universität Siegen, Fakultät IV, Department Elektrotechnik und Informatik, Praktische Informatik (PI), Hölderlinstr. 3, 57068 Siegen, E-Mail: nasiri@informatik.uni-siegen.de

Prof. Dr. Eyke Hüllermeier ist Professor an der Philipps-Universität Marburg, Fachbereich Mathematik und Informatik. Seine Hauptarbeitsgebiete liegen in den Bereichen Maschinelles Lernen, Fuzzy Logik und der Verarbeitung unsicheren Wissens. Neben theoretischen und methodischen Grundlagen interessiert er sich auch für Anwendungen in unterschiedlichen Gebieten, von den Bio- und Lebenswissenschaften über technische Disziplinen bis hin zu den Geisteswissenschaften. Er hat etwa 200 wissenschaftliche Beiträge publiziert, viele davon in hochrangigen Journalen und internationalen Konferenzen. Er ist Editor-in-Chief von *Fuzzy Sets and Systems* und Mitglied im Editorial Board einer Reihe weiterer internationaler Zeitschriften auf dem Gebiet der Computational Intelligence.

Adresse: Philipps-Universität Marburg, FB Mathematik und Informatik, Biegestraße 10, 35032 Marburg,
E-Mail: eyke@mathematik.uni-marburg.de