Fallbasiertes und symbolisches Lernen*

Christoph Globig und Stefan Weß
Fachbereich Informatik
Sonderforschungsbereich 314
Universität Kaiserslautern
W-6750 Kaiserslautern

Zusammenfassung

In den letzten Jahren wurden Methoden des fallbasierten Schließens häufig in Bereichen verwendet, in denen traditionell symbolische Verfahren zum Einsatz kommen, beispielsweise in der Klassifikation. Damit stellt sich zwangsläufig die Frage nach den Unterschieden bzw. der Mächtigkeit dieser Lernverfahren. Jantke [Jantke, 1992] hat bereits Gemeinsamkeiten von *Induktiver Inferenz* und *fallbasierter Klassifikation* untersucht. In dieser Arbeit wollen wir einige Zusammenhänge zwischen der Fallbasis, dem Ähnlichkeitsmaß und dem zu erlernenden Begriff verdeutlichen. Zu diesem Zweck wird ein einfacher symbolischer Lernalgorithmus (der Versionenraum nach [Mitchell, 1982]) in eine äquivalente, fallbasiert arbeitende Variante transformiert. Die vorgestellten Ergebnisse bestätigen die Äquivalenz von symbolischen und fallbasierten Ansätzen und zeigen die starke Abhängigkeit zwischen dem im System verwendeten Maß und dem zu lernenden Begriff.

1 Einführung

Mit dem Begriff des fallbasierten Schließen ist der Vorgang des Lernens direkt verknüpft. Wird eine weitere Klassifikationsaufgabe korrekt gelöst, so wird sie als ein neuer Fall in die Fallbasis aufgenommen [Aha *et al.*, 1991]. Lernen und die Anwendung des gelernten Wissens sind dabei nicht strikt voneinander getrennt, sondern eng miteinander verzahnt. Aus der Sicht des Maschinellen Lernens kann fallbasiertes Schließen (fallbasiertes Lernen) daher als eine Art *Begriffsbildungsaufgabe* verstanden werden (vgl. [Aha & Kibler, 1989; Salzberg, 1991]). Damit stellt sich zwangsläufig die Frage, wie die gelernten Begriffe in fallbasierten Systemen repräsentiert werden. Im Gegensatz zu symbolischen Lernverfahren, bei denen ein gelernter Begriff *explizit* durch eine entsprechende Formel repräsentiert wird, beschreiben fallbasierte Verfahren Begriffe *implizit* als ein Tupel (sim, FB) (vgl. [Richter, 1992]), d.h. durch ein Ähnlichkeitsmaß sim und eine Menge FB von Fallbeispielen. Das Verhältnis von Fallbasis und Maß zu dem zu erlernenden Begriff kann daher durch folgende Gleichung charakterisiert werden:

*Die hier vorgestellte Arbeit wurde zum Teil gefördert durch die Deutsche Forschungsgemeinschaft, SFB 314: "Künstliche Intelligenz - Wissensbasierte Systeme"

$$\boxed{\text{Begriff} = \text{Fallbasis} + \text{Ähnlichkeitsmaß}}$$

Diese Gleichung hat mehr als nur einen symbolischen Wert. Sie zeigt erstens, daß es sehr viele unterschiedliche Möglichkeiten gibt, einen bestimmten Begriff fallbasiert zu repräsentieren[1], d.h. es existieren unter Umständen mehre Tupel $(sim_1, FB_1), \ldots, (sim_k, FB_k)$ für einen bestimmten Begriff B. Zweitens verdeutlicht sie die Möglichkeiten die Klassifikationsfähigkeit fallbasierter Systeme durch Lernen zu verbessern. Lernen kann in fallbasierten Systemen auf zwei Ebenen stattfinden: Einerseits durch die *Aufnahme neuer Fallbeispiele* in die Fallbasis und andererseits durch eine *Änderung des verwendeten Maßes* [2]. Es ist ein nach unserer Meinung wesentlicher Vorteil des fallbasierten Ansatzes, daß Defizite in einem Bereich, durch den jeweils anderen Bereich ausgeglichen werden können. Ein weniger gut informiertes Maß kann so durch geeignete Fallbeispiele kompensiert werden und umgekehrt. Werden in einem fallbasierten System unterschiedliche Begriffe gelernt, so wird jeder Begriff B_j durch ein *spezifisches Tupel* (sim_j, FB_j) repräsentiert.

Für unsere weiteren Betrachtungen gehen wir zunächst von folgendem Szenario aus (vgl. [Richter, 1992]):

1. Das Universum U der betrachteten Objekte sei endlich.
2. d ist total und es gilt $\forall a, b, x \in U \; [d(a,b) = 0 \Rightarrow d(x,a) = d(x,b)]$
3. Das verwendete Distanzmaß d sei fest.
4. Die Hypothesen bzw. die zu lernenden Begriffe sind binär.

Die Bedingung 2 garantiert, daß die Relation $\sim \subseteq U \times U$ definiert durch $x \sim y \Leftrightarrow d(x,y) = 0$ eine Äquivalenzrelation ist. $|U/\sim|$ gibt die Anzahl der Äquivalenzklassen an, die durch $\sim$ gebildet werden. Als erstes halten wir fest, daß mit d genau die Begriffe B gelernt werden können, für die gilt $d(x,y) = 0 \Rightarrow B(x) \equiv B(y)$, d.h. die Elemente einer Äquivalenzklasse müssen immer einheitlich klassifiziert sein.

Die einzelnen Äquivalenzklassen können unabhängig voneinander belegt werden, ohne die angegebenen Bedingung zu verletzen. Damit erhält man sofort, daß mit d $2^{|U/\sim|}$ Begriffe unterschieden werden können, die durch $|U/\sim|$ (geeignete) Fälle eindeutig festgelegt werden können. Das Distanzmaß d spannt also den Raum der erlernbaren Begriffe $B(x)$ auf, aus dem mit der Fallbasis FB ein bestimmter Begriff B ausgewählt wird.

Fallbasierte Lernverfahren nutzen Fallbeispiele einerseits dazu Äquivalenzklassen zu belegen. Andererseits wird Wissen, das aus den präsentierten Fällen gezogen wird, auch dazu verwendet, die Anzahl der Äquivalenzklassen so zu verkleinern, daß das Zielbegriff in ihnen mit möglichst wenigen Fallbeispielen identifiziert werden kann. Gleichzeitig nimmt die Zahl der Begriffe, die mit dem modifizierten Maß überhaupt unterschieden werden können, ab.

[1] Wie auch bei der Addition eine bestimmte Summe durch sehr viele unterschiedliche Summanden gebildet werden kann.

[2] Wir verwenden im folgenden den Begriff Maß für *Ähnlichkeits-* und *Distanzmaße*. Beide Darstellungsformen sind äquivalent [Richter & Wess, 1991] und wir verwenden im folgenden je nach Kontext die eine oder die andere Darstellungsart.

2 Eine fallbasierte Variante eines symbolischen Lernverfahrens

Um die Überlegungen aus dem letzten Abschnitt zu belegen und die Äquivalenz von symbolischen und fallbasierten Ansätzen zu verdeutlichen, wollen wir in diesem Abschnitt ein bekanntes symbolisches Lernverfahren (den Versionenraum (VS) nach [Mitchell, 1982]) in eine entsprechende fallbasiert arbeitende Variante VS_CBR transformieren.

2.1 Der symbolisch arbeitende Versionenraum

Das Universum U der Objekte besteht aus endlichen Vektoren über endlichen Wertemengen W_i ($U = W_1 \times \ldots \times W_n$). Die Begriffe, die gelernt werden sollen, legen die Ausprägungen von bestimmten Attributen fest. Wir können die Begriffe ebenfalls als Vektoren $(B_1, \ldots, B_n)$ beschreiben, mit $B_i = *$ oder $B_i = a_i$ ($a_i \in W_i$). Ein Objekt $a = (a_1, \ldots, a_n)$ erfüllt den Begriff B, wenn für alle $1 \leq i \leq n$ gilt: $B_i = *$ oder $B_i = a_i$.

Der symbolische Algorithmus *Versionenraum* [Mitchell, 1982] löst das Lernproblem, indem er zwei Mengen S und G von Begriffen verwaltet. In S ist der speziellste Begriff enthalten, der mit den bisher bekannten Fällen konsistent ist, in G die allgemeinsten Begriffe, die konsistent zu den bekannten Fällen sind. Der symbolische Algorithmus besteht darin, während der Präsentation weiterer Fälle die Mengen S und G so zu verändern, daß die angegebenen Eigenschaften erhalten bleiben. Der Begriff ist eindeutig bestimmt, wenn $S = G$ gilt. Wir verzichten hier auf eine genauere Beschreibung des symbolischen Algorithmus (vgl. [Mitchell, 1982]). Wichtig ist, daß zu jedem Zeitpunkt alle Objekte, die von S abgedeckt werden, sicher positiv sind und diejenigen, die nicht von einem Begriff aus G subsumiert werden, sicher negativ.

2.2 Eine fallbasierte Variante des Versionenraums

Wenn man sich den Versionenraum-Algorithmus ansieht, fällt auf, daß die wesentliche Lernleistung darin besteht, relevante Attributausprägungen von irrelevanten zu unterscheiden. Diese Beobachtung verwenden wir zur Konstruktion einer fallbasierten Variante VS-CBR des Grundalgorithmus.

Eine Attributausprägung heiße *relevant*, wenn sie im zu erlernenden Begriff verwendet wird. Für jedes Attribut i wird eine Funktion f_i definiert, die die Menge W_i auf $\{0, 1\}$ abbildet. Dabei wird gelten

$$f_i(x_i) = \begin{cases} 1 & : \quad B_i = x_i \text{ ist möglich} \\ 0 & : \quad \text{sonst} \end{cases}$$

Die so entstehenden Funktionen f_i werden zu einer Funktion $f : U \rightarrow \{0, 1\}^n$ zusammengefaßt. Die Distanz zwischen zwei Objekten a und b wird über die City-Block-Metrik definiert als

$$d_f(a, b) := |f_1(a_1) - f_1(b_1)| + \ldots + |f_n(a_n) - f_n(b_n)|$$

Jede Veränderung der Funktionen $f_1, f_2, \ldots, f_n$ bewirkt damit eine Änderung des der Klassifikation zugrundeliegenden Maßes d_f.

2.2.1 Algorithmus zum Lernen von f für VS-CBR

Die gewünschte Funktion f läßt sich durch den folgenden Algorithmus erlernen. Der Algorithmus erwartet (wie VS) als erstes einen positiven Fall.

1. Setze alle $f_i(x_i) = 0$ für alle $i, x_i \in W_i$
2. Der erste positive Fall sei $(a_1, \ldots, a_n)$. Setze $f_i(a_i) = 1$ und $FB = \{a\}$.
3. Lese einen neuen Fall $b = (b_1, \ldots, b_n)$ ein.
4. Wenn b negativ ist, speichere b in der Fallbasis FB.
5. Wenn b positiv und $f_i(b_i) = 0$, setze $f_i(x_i) = 0$ für alle x_i aus W_i (f_i ist jetzt die Nullfunktion).
6. Gilt $d_f(a, b) = 0$ und ist die Klassifikation von a, b unterschiedlich, dann **ABBRUCH**. Der zu lernende Begriff liegt nicht im Versionenraum und kann daher nicht gelernt werden.
7. Entferne redundante Fälle aus der Fallbasis FB.
8. Wenn der Begriff eindeutig bestimmt ist, gehe zu Schritt 9 sonst gehe zu Schritt 3.
9. **STOP:** Der Begriff ist eindeutig bestimmt. Gebe die Fallbasis FB und das verwendete Maß d_f als Klassifikator für den gelernten Begriff aus.

Wird ein Objekt c vorgelegt, wird die Menge $F = \{fb \,|\, d_f(fb, c) \leq d_f(fb', c)\}$ für alle Fälle fb' aus der Fallbasis FB berechnet. Enthält F mehr als ein Element, wird dasjenige mit der kleinsten Klassifikation als Vergleichsobjekt zurückgegeben. Diese Vereinbarung dient lediglich dazu, eine Semantik festzulegen, bevor der Begriff eindeutig bestimmt ist. Man kann leicht zeigen, daß der fallbasierte Algorithmus VS_CBR die Objekte, die der symbolische VS-Algorithmus sicher klassifizieren kann, ebenfalls korrekt behandelt. Ein Unterschied ist zunächst, daß die fallbasierte Variante VS_CBR zu jedem Objekt eine Klassifikation berechnet, während der symbolische Algorithmus VS bei unsicherer Klassifikation (d.h. das Objekt fällt unter einen Begriff aus G, aber nicht unter S) keine Entscheidung trifft. Durch einen Test auf die Sicherheit der durch den fallbasierten Algorithmus getroffene Klassifikation, kann man dieses Verhalten jedoch leicht erreichen. Die Fälle, die dem Algorithmus präsentiert werden, führen bei VS_CBR zu zwei unterschiedlichen Veränderungen im Tupel (d, FB):

- Positive Fälle werden zur Veränderung der f_i verwendet, d.h. zur Adaption des Distanzmaßes d und nicht in die Fallbasis FB aufgenommen. Der Abstand d aller positiven Fälle, b, b' zueinander ist nach der Anpassung von f $d_f(b, b') = 0$.

- Negative Fälle werden hingegen in die Fallbasis FB aufgenommen, ohne das Distanzmaß d_f zu verändern.

3 Wahl des Distanzmaßes

Wir haben oben argumentiert, daß unter den angegebenen Bedingungen mit einem Abstandsmaß d genau die Begriffe B lernbar sind, für die gilt: $d(x,y) = 0 \Rightarrow B(x) \equiv B(y)$. Ob ein Begriff fallbasiert lernbar ist, hängt also nur von der Definition des Abstandes 0 ab. Das Distanzmaß kann für Abstände größer 0 beliebig definert sein, z.B. konstant 1.

Welchen Sinn haben dann Distanzmaße, die einen Wertebereich haben, der größer ist als $\{0,1\}$? Der einzige Grund für die Wahl eines komplexeren Distanzmaßes ist (in dem hier betrachteten Szenario vgl. dazu [Wess, 1993]) die Hoffnung auf Zwischenhypothesen, die hohe Erkennungsraten haben, obwohl der Begriff durch die vorliegenden Fälle noch nicht eindeutig bestimmt ist. Dieser Fall liegt vor, wenn noch nicht alle Äquivalenzklassen belegt sind, die das Maß bildet.

Um dies zu illustrieren, betrachten wir zwei Ähnlichkeitsmaße, die dieselben Begriffe erlernen können. Das Universum bestehe aus Objekten mit vier Attributen. Jedes Attribut habe als Wertemenge $\{0,\ldots,15\}$. Dabei soll der folgende Begriff $B(x),\ x := (x_1,\ldots,x_n)$ vom fallbasierten System gelernt werden:

$$B(x) = 1 \Leftrightarrow (x_2 \geq 8 \wedge x_4 < 8) \vee (x_4 \geq 8 \wedge x_2 < 8)$$

und als Distanzmaße verwenden wir:

$$\begin{aligned} d_1(a,b) &= \begin{cases} 1 & : \quad a = b \\ 0 & : \quad sonst \end{cases} \\ d_2(a,b) &= \sum_i |a_i - b_i| \end{aligned}$$

Beide Maße können offensichtlich alle Begriffe über dem Universum U erlernen, da $d_i(a,b) = 0 \Leftrightarrow a = b$. Abbildung 1 gibt Erkennungsraten der Maße für den angegebenen Begriff $B(x)$ bei verschieden großen Fallbasen an. Man kann für d_1 nachrechnen, daß für eine Erkennungsrate von 90% mehr als 52000 Fälle notwendig sind.

Der Unterschied zwischen den Maßen d_1, d_2 ist, daß bei d_2 ein kleiner Abstand zwischen Objekten auf eine hohe Wahrscheinlichkeit für eine gleiche Klassifikation hinweist. Bei d_1 besagt ein Abstand größer 0 nichts über die Klassifikation der Objekte. Beim Maß d_2 verbessert sich daher die Klassifikationsgüte mit der Anzahl der betrachteten Fallbeispiele, während sie bei d_1 nahezu konstant bleibt.

Es ist aber ohne weiteres möglich, einen Begriff anzugeben, bei dem die Distanzen von d_2 ähnlich bedeutungslos sind, wie die von d_1 bei B.

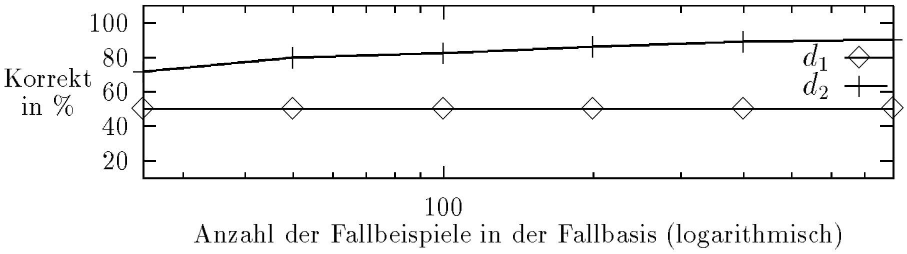

Abbildung 1: Angegeben ist die Erkennungsrate der Maße d_1 und d_2 auf Fallbasen unterschiedlicher Größe.

3.1 Universalität und Minimalität

Man kann fallbasierte Systeme (sim, FB) anhand der Dimensionen *Minimalität* und *Universalität* vergleichen.

Die Dimension Minimalität bezieht sich auf das im verwendeten Maß bereits implizit kodierte Wissen. Da wir dieses Wissen nicht direkt bestimmen können, verwenden wir dazu die Anzahl der für eine bestimmte Klassifikationsgüte minimal notwendigen Fallbeispiele.

Definition 1 (Minimalität) *Das in einem fallbasierten System (sim_1, FB_1) verwendete Maß sim_1 heißt besser informiert als das Maß sim_2 eines fallbasierten Systems (sim_2, FB_2), falls bei gleicher Klassifikationsgüte gilt: $|FB_1| < |FB_2|$ mit FB_1 und FB_2 sind minimal.*

Die zweite Dimension bezieht sich auf die Menge der mit dem Maß *sim* erlernbaren Begriffe B.

Definition 2 (Universalität) *Ein Maß sim_1 heißt universeller als ein Maß sim_2, falls die Menge der mit sim_1 erlernbaren Begriffe eine echte Obermenge der mit sim_2 lernbaren ist.*

Universalität und *Minimalität* sind Ziele, die einander widersprechen. Um die Minimalität eines Systems zu verbessern, muß im allgemeinen eine Verkleinerung der Menge der erlernbaren Begriffe in Kauf genommen werden, d.h. die Universalität nimmt ab. Man kann bei den Ähnlichkeitsmaßen zwei Extremfälle unterscheiden:

Identität der Objekte: Hier wird die Distanz nur dann minimal, wenn die verglichenen Objekte identisch sind. Unter dieses Extrem fällt auch das Maß d_1. Seine Universalität kann jedoch nur ausgenutzt werden, wenn die Fallbasis sämtliche Objekte des Universums U umfaßt. Vorher kann man den durch (d_1, FB) beschriebenen Begriff durch das Hinzufügen weiterer Fälle in die Fallbasis verändern.

Identität der Klassifikation: Die Ähnlichkeit wird dann maximal, wenn die Klassifikation der verglichenen Objekte übereinstimmt. Sämtliches Wissen über den zu erlernenden Begriff B ist bereits im Maß durch die Begriffsdefinition $B(x)$ kodiert. Die Fallbasis muß nur noch verwendet werden, um Trivialfälle auszuschließen. Das Maß

$sim(x,y) := B(x) \equiv B(y)$ kann jedoch nur noch zwischen vier Begriffen unterscheiden (B, $not(B)$, TRUE – d.h. alles positiv, FALSE – d.h. alles negativ).

Man kann sich die Gegenläufigkeit von *Minimalität* und *Universalität* anhand der folgenden Tabelle verdeutlichen. Sie listet zu verschiedenen Distanzmaßen auf, wieviele Begriffe unterschieden werden können und wieviele Fälle die Fallbasis mindestens umfassen muß, um eine eindeutige Festlegung zu treffen. In der Tabelle wird von einem Universum mit vier Dimensionen ausgegangen, die jeweils eine von 16 Ausprägungen annehmen können, also 65536 Objekte umfaßt. Der zu erlernende Begriff besteht in der Festlegung von zwei Ausprägungen.

Maß	min. Fallbasis	Größe des Hypothesenraums
$sim(x,y) := (x = y)$	$65536 = 16^4$	2^{65536}
VS_CBR_1	16	$65536 = 2^{16}$
VS_CBR_2	3 (4)	$16 = 2^4$
$sim(x,y) := K(x) \equiv K(y)$	2	2^2

Abbildung 2: Minimalität und Universalität bei verschiedenen Maßen

Die beiden anderen in der Tabelle angegebenen Maße liegen zwischen diesen Extremen. VS_CBR_1 und VS_CBR_2 haben weder die maximale Universaltität noch kommen sie mit einer minimalen Fallbasis aus. VS_CBR_1 bezeichnet das Distanzmaß, das nach der Präsentation des ersten positiven Beispiels von VS_CBR errechnet wird. Es legt in jeder Dimension eine Ausprägung auf eins. VS_CBR_2 ist das Maß, gegen das VS_CBR konvergiert. Es unterscheidet nur noch die beiden relevanten Ausprägungen voneinander, bildet also nur noch vier Äquivalenzklassen. Um alle mit diesem Maß erlernbaren Begriffe unterscheiden zu können, sind mindestens vier Fälle nötig. Für einen Begriff aus dem Versionenraum sind nur drei Fälle nötig.

4 Zusammenfassung

Unsere bisherigen Ergebnisse lassen fallbasierte und symbolische Klassifikatoren in einem sehr ähnlichen Licht erscheinen. Sie bestätigen damit die in [Jantke, 1992] präsentierten Ergebnisse. Die in dieser Arbeit exemplarisch am Beispiel des Versionenraums durchgeführte Transformation eines symbolischen Lernalgorithmus in eine fallbasierte Variante läßt sich auch für andere symbolsiche Algorithmen zur Begriffsbildung zeigen, z.B. ID3, AQ, CN2 etc. (vgl. [Globig, 1993]).

Beide Ansätze benötigen neben den eigentlichen Fällen weitere Informationen über den zu erlernende Begriff, um effizient lernen zu können. Zwar lassen sich für fallbasierte Lerner Ähnlichkeitsmaße konstruieren, die jeden binäre Begriff erkennen können, indem ein Ähnlichkeitsmaß verwendet wird, für das gilt: $sim(x,y) = 1 \Leftrightarrow x = y$. Nur gibt es dann Begriffe, die erst mit dem gesamten (oder nahezu dem gesamten) Universum als Fallbasis erkannt werden können.

Distanzmaße mit zum erlernenden Begriff passenden Umgebungen haben insbesondere dann einen Vorteil, wenn die Rahmenbedingungen des Lernens verändert werden – beispielsweise durch verrauschte Attributwerte. Wenn das Rauschen die Distanzbeziehungen im wesentlichen bestehen läßt, der Effekt des Rauschens also klein ist in Bezug auf das Distanzmaß, wird der nächste Nachbar des verrauschten Objektes immer noch die korrekte Klassifikation haben. Es ist aber festzuhalten, daß die Toleranz nicht eine Eigenschaft des fallbasierten Ansatzes sondern eine Eigenschaft des Distanzmaßes ist.

Mit anderen Worten ist das Ähnlichkeitsmaß der Bias des fallbasierten Lernens. Mit seiner Wahl und den möglichen Modifikationen wird die Menge der erkennbaren Begriffe festgelegt. Wir wollen den Vergleich der Ansätze weiterführen (vgl. [Haffner, 1993; Globig, 1993]). Es ist insbesondere geplant, das Verhalten der Algorithmen unter veränderten Rahmenbedingungen (Rauschen in den Daten, fehlende Werte und unscharfe Begriffe) zu untersuchen.

5 Danksagung

Die Autoren danken Michael M. Richter und der gesamten Arbeitsgruppe in Kaiserslautern für viele anregende Diskussionen und ein sehr gutes Arbeitsklima.

Literaturverzeichnis

AHA, DAVID W., KIBLER, DENNIS, & ALBERT, MARC K. 1991. Instance-Based Learning Algorithms. *Machine Learning*, **6**, 37–66. March 1991.

AHA, D.W., & KIBLER, D. 1989. Noise-Tolerant Instance-Based Learning Algorithms. *Pages 794–799 of: Proceedings of the 11th International Conference on Artificial Intelligence IJCAI-89.* IJCAI. Detroit, Michigan, USA.

GLOBIG, CHRISTOPH. 1993. *Fallbasiertes und Symbolisches Lernen.* Diplomarbeit, Universität Kaiserslautern. (in Vorbereitung).

HAFFNER, ERNST-GEORG. 1993. *Analyse dynamischer Lernregeln für Case-Based Learning Systeme.* Diplomarbeit, Universität Kaiserslautern.

JANTKE, KLAUS P. 1992. Case-Based Learning in Inductive Inference. *In: Proc. COLT-92.*

MITCHELL, T.M. 1982. Gerneralization as Search. *Artificial Intelligence*, **18**(2), 203–226.

RICHTER, M. M., & WESS, S. 1991. Similarity, Uncertainty and Case Based Reasoning in PATDEX. *Pages 249–265 of:* BOYER, ROBERT S. (ed), *Automated Reasoning - Essays in Honor of Woody Bledsoe.* Kluwer Academic Publishers.

RICHTER, MICHAEL M. 1992. Classification and Learning of Similarity Measures. *In: Proc. der 16. Jahrestagung der Gesellschaft für Klassifikation e.V.* Springer Verlag.

SALZBERG, STEVEN. 1991. Distance Metrics for Instance-Based Lerning. *Pages 399–408 of:* RAS, Z. W., & ZEMANKOVA, M. (eds), *Proceedings of the 6th International Symposium on Methodologies for Intelligent Systems (ISMIS'91).* Lecture Notes in Artificial Intelligence, vol. 542. Berlin: Springer-Verlag. Charlotte, North Carolina, USA, October 1991.

WESS, STEFAN. 1993. PATDEX - Inkrementelle und wissensbasierte Verbesserung von Ähnlichkeitsurteilen in der fallbasierten Diagnostik. *In: Tagungsband 2. deutsche Expertensystemtagung XPS-93.* Hamburg: Springer Verlag.