

Lernen von Wissensintensiven Ähnlichkeitsmaßen im Case-Based Reasoning

Armin Stahl

1 Case-Based Reasoning

In den vergangenen Jahren hat sich vor allem eine KI-Technik zunehmend auch im kommerziellen Umfeld erfolgreich durchsetzen können, nämlich die des *Fallbasierten Schließens* oder engl. *Case-Based Reasoning (CBR)*. Während die klassischen Anwendungsgebiete Klassifikations-, Diagnose- und Help-Desk-Systeme sind, hat sich CBR mittlerweile zunehmend auch im Bereich von eCommerce- [6] und Wissensmanagement-Aufgabenstellungen durchgesetzt.

Die Grundidee von CBR ist die Wiederverwendung von Situations-spezifischem Erfahrungswissen zur effizienten Lösung gegebener Problemstellungen [2]. Dieses Erfahrungswissen wird dabei typischer Weise durch strukturiert dokumentierte¹ Problem-Lösungs-Paare - die sog. *Fälle* - repräsentiert, die in einer speziellen Datenbank - der sog. *Fallbasis* - abgelegt werden.

Um dieses gesammelte Wissen erfolgreich für die Lösung eines gegebenen Problems einsetzen zu können, müssen in einem ersten Schritt - dem sog. *Retrieval* - Fälle ausgewählt werden, die aussichtsreich erscheinen zur aktuellen Problemlösung beizutragen. Die darin enthaltenen Lösungsvorschläge müssen dann evtl. in einem zweiten Schritt an die neue Problemstellung angepasst werden. Das klassische Prozessmodell des CBR [1] sieht darauf folgend zwei weitere Schritte vor, um die adaptierte Lösung zu validieren und im Erfolgsfall als neues Erfahrungswissen in der Fallbasis abzulegen.

Der Fokus der hier beschriebenen Dissertation [8] liegt jedoch auf der ersten, und damit grundlegenden Phase der Problemlösung, nämlich der Auswahl nützlicher Fälle. Da die tatsächliche Nützlichkeit eines Falles erst nach der Problemlösung sicher bestimmt werden kann, wird für das Retrieval eine Heuristik benötigt. Im CBR basiert diese Heuristik auf der Annahme, dass *ähnliche Probleme ähnliche Lösungen* haben. Um die Ähnlichkeit zwischen einer gegebenen Problemstellung - der sog. *Anfrage* - und denen in den Fällen dokumentierten Problemen zu bestimmen, werden sog. *Ähnlichkeitsmaße* eingesetzt.

Während in frühen CBR-Anwendungen vor allem einfache, generell einsetzbare Abstandsmaße (z.B. Hamming oder Euklidischer Abstand) eingesetzt wurden, hat sich mittlerweile gezeigt, dass in vielen neueren Anwendungen aufwendigere, auf die Anwendungsdomäne spezialisierte Maße benötigt werden, um zufriedenstellende Retrievalergebnisse zu erhalten. Wie bei allen Heuristiken, kann auch bei Ähnlichkeitsmaßen die Einbeziehung von Hintergrundwissen die Qualität der Heuristik deutlich verbessern. Der Nachteil solch *wissensintensiver Ähnlichkeitsmaße* ist jedoch der deutlich erhöhte Wissensakquisition- und -modellierungsaufwand. Die Reduzierung dieses Aufwands durch die Verwendung von relativ einfach verfügbarem Erfahrungswissen, ist jedoch gerade ein entscheidender Vorteil

¹ Wir setzen hier eine Attribut-Wert basierte Repräsentation voraus.

von CBR im Vergleich zu klassischen KI-Techniken, wie z.B. Regel-basierten Ansätzen.

Um die Definition wissensintensiver Ähnlichkeitsmaße zu vereinfachen, liegt es nahe, Techniken des Maschinellen Lernens einzusetzen. Vor allem für Klassifikationsaufgaben wurden schon früh Ansätze entwickelt, um Attributgewichtungen, welche auch eine Form von Ähnlichkeitswissen darstellen, aus vor-klassifizierten Datensätzen zu erlernen [4][10]. Für viele neueren CBR-Anwendungen sind diese Ansätze jedoch aus zwei Gründen nicht mehr ausreichend:

- Stehen keine klassifizierte Daten zur Verfügung (z.B. bei der Produktsuche im eCommerce), sind die existierenden Verfahren aufgrund dem Fehlen geeigneter Trainingsdaten nicht anwendbar.
- Neben Attributgewichten werden heute von vielen kommerziellen CBR-Werkzeugen sog. *lokale Ähnlichkeitsmaße* zur Kodierung von Domänenwissen bereitgestellt. Ein solches Maß dient dazu, die Ähnlichkeit bzgl. eines einzelnen Attributs zu berechnen. Basierend auf diesen lokalen Ähnlichkeitswerten und den Gewichten ergibt sich dann die Gesamtähnlichkeit zwischen Anfrage und Fall, z.B. durch die Bildung einer gewichteten Summe. Je nach Datentyp des Attributs werden lokale Ähnlichkeitsmaße sehr unterschiedlich repräsentiert, z.B. durch Tabellen oder mathematische Funktionen. Gerade deren Definition ist jedoch mit einem erheblichen Wissensakquisitions- und -modellierungsaufwand verbunden.

Das Ziel dieser Arbeit bestand daher in der Entwicklung eines alternativen Lernansatzes, der zum einen flexibel genug ist, um in sehr unterschiedlichen Anwendungsszenarien einsetzbar zu sein, und zum anderen auch das Lernen lokaler Ähnlichkeitsmaße ermöglicht.

2 Lernen durch Nützlichkeits-Feedback

Hat man es mit klassifizierten Daten zu tun, kann man diese direkt dazu nutzen, die Problemlösekompetenz eines KI-Systems zu messen, indem man einzelne Datensätze als Anfrage an das System stellt, und die Ausgabe mit der bekannten Lösung vergleicht (*leave-one-out Test*).

2.1 Nützlichkeits-Feedback

In anderen Domänen fällt es jedoch erheblich schwerer die Qualität von Systemausgaben zu beurteilen. Bei der Auswahl von geeigneten Produkten im eCommerce oder anderen generellen Retrieval-Aufgabenstellungen, hängt die Nützlichkeit der Retrieval-Ergebnisse z.B. oft zu einem großen Teil auch von individuellen Präferenzen der Benutzer ab. Hierbei ist weniger bedeutend, wie nützlich die einzelnen Fälle absolut sind, denn eine geringe Nützlichkeit kann auch mit der Qualität der Fallbasis zusammenhängen. Für das Retrieval interessant ist vielmehr die *relative Nützlichkeit* der retrieveden Fälle untereinander.

Der hier beschriebene Lernansatz basiert daher auf der Annahme, dass es einen *Ähnlichkeitslehrer* gibt, der für gegebene Beispielanfragen die Nützlichkeit der zurückgelieferten Fälle analysiert und gegebenenfalls das Ranking dieser (teilweise) korrigiert. Das Ergebnis ist eine partiell, bzgl. ihrer Nützlichkeit geordnete Menge von Fällen, die zusammen mit der zugrundeliegenden Anfrage als ein Trainingsbeispiel abgespeichert werden kann. Wie der Ähnlichkeitslehrer dabei genau realisiert wird, hängt vom jeweiligen Anwendungsszenario ab:

- Nahe liegend ist ein Domänenexperte, der aufgrund seiner Erfahrung die tatsächliche Nützlichkeit von Fällen schnell einschätzen kann.
- Auch die Benutzer selbst können in manchen Anwendungen die Rolle des Lehrers einnehmen. Hierbei kann man aktive Ansätze, bei denen der Benutzer explizit zu Feedback aufgefordert wird, und passive Ansätze, bei denen das Feedback aus dem Verhalten (z.B. Kaufverhalten) des Benutzers abgeleitet wird, unterscheiden.
- Neben menschlichen Lehrern können aber genauso Softwareagenten, z.B. mit Hilfe von Hintergrundwissen oder durch Simulation, die Nützlichkeit von Fällen automatisiert bewerten.

2.2 Lernansatz

Stellt sich nun die Frage, wie man mit Hilfe des beschriebenen Nützlichkeits-Feedback die relevanten Elemente eines Ähnlichkeitsmaßes, d.h. Attributgewichte und lokale Ähnlichkeitsmaße, erlernen kann. Prinzipiell beschreibt dieses Feedback ja partiell die gewünschten Retrievalergebnisse für eine Menge von Beispielanfragen. Diese optimalen Retrievalergebnisse kann man nun mit den tatsächlichen, unter Verwendung eines initialen Ähnlichkeitsmaßes erzielten, Ergebnissen vergleichen.

Durch die Definition einer geeigneten *Fehlerfunktion* $E: SIM \rightarrow R$ (wobei SIM die Menge aller repräsentierbaren Ähnlichkeitsmaße bezeichnet) lässt sich der Grad der Abweichung mathematisch durch einen numerischen Wert beschreiben. Auf der einen Seite sollte eine solche Funktion einen Fehler von 0 zurück liefern, genau dann, wenn die mit einem gegebenen Maß SIM erzielten Retrievalergebnisse mit den Trainingsdaten übereinstimmen. Auf der anderen Seite, je größer die Abweichung der beiden Partialordnungen ist, desto höher sollte auch der Fehlerwert ausfallen.

Wenn eine Fehlerfunktion E mit diesen Eigenschaften gegeben ist, kann die eigentliche Lernaufgabe als ein Optimierungsproblem aufgefasst werden. D.h. es wird ein optimales Ähnlichkeitsmaß Sim_{opt} gesucht, dass zu einem minimalen Retrieval-Fehler führt, oder anders ausgedrückt, es wird ein globales Minimum von E gesucht.

2.3 Lernalgorithmen

Prinzipiell existieren unterschiedliche Ansätze zur Lösung solcher Optimierungsprobleme. Für das Lernen von Attributgewichten haben sich z.B. Gradientenabstiegsverfahren [4][10][7] bewährt. Im Hinblick auf die Zielsetzung dieser Arbeit sind diese Verfahren jedoch nicht geeignet, da sie von der Fehlerfunktion E erfordern, dass diese partiell ableitbar in Bezug auf die zu optimierenden Parameter ist. Während die Definition einer solchen Funktion für das Lernen von Gewichten noch eingeschränkt möglich ist (vorausgesetzt man toleriert mögliche ungewollte Minima), gestaltet sich dies für lokale Ähnlichkeitsmaße als praktisch unmöglich.

Für die in dieser Arbeit vorausgesetzten lokalen Ähnlichkeitsmaße, nämlich Ähnlichkeitstabellen für ungeordnete Da-

tentypen, und differenz-basierte Ähnlichkeitsfunktionen für geordnete Datentypen, ist daher ein flexiblerer Lernalgorithmus notwendig. Hier bieten sich vor allem genetische oder evolutionäre Algorithmen [5] an. Zum einen erfordern diese Verfahren keinerlei besondere Eigenschaften von der Fehlerfunktion E (hier auch „*Fitnessfunktion*“ genannt) und zum anderen haben sie sich in ähnlichen Problemstellungen vielfach bewährt.

Um eine komplexe Abbildung auf Bitstring-Repräsentationen zu vermeiden, wurde im Rahmen dieser Arbeit ein evolutionärer Algorithmus entwickelt, der mehr oder weniger direkt auf den Repräsentationen der lokalen Ähnlichkeitsmaße aufsetzt. Zum einen optimiert der evolutionäre Prozess direkt die reellwertigen Einträge der Ähnlichkeitstabellen. Ähnlichkeitsfunktionen werden hingegen durch einen Vektor reellwertiger Stützstellen interpoliert, um möglichst flexibel sehr unterschiedliche Funktionen repräsentieren zu können. Für die Realisierung eines evolutionären Prozesses zur Optimierung dieser Tabellen und Vektoren war es notwendig spezialisierte *Crossover-* und *Mutationsoperatoren* zu definieren. Für eine genauere Beschreibung dieser Operatoren sei auf [9][8] verwiesen. Neben den lokalen Ähnlichkeitsmaßen berücksichtigt der vorgestellte Algorithmus auch noch die Attributgewichte, wobei hierfür jedoch auch eine Kombination mit einem Gradientenabstiegsverfahren möglich ist.

3 Anwendungsszenarien

Wie bereits beschrieben, war es ein wichtiges Ziel der Arbeit einen möglichst allgemein anwendbaren Lernansatz zu entwickeln. Im folgenden sollen jedoch einige typische Anwendungsszenarien kurz vorgestellt werden.

3.1 Unterstützung von Domänenexperten

Naheliegender ist eine Unterstützung eines Domänenexperten/Wissensingenieurs bei der Modellierung wissensintensiver Ähnlichkeitsmaße. Der Vorteil des Lernansatzes ist hierbei, dass der Experte sich nicht mit der detaillierten mathematischen Repräsentation der Ähnlichkeitsmaße auseinandersetzen muss, sondern sein Wissen über die Domäne anhand von Beispielen erklären kann. D.h., anstatt einzelne Einflüsse auf die Nützlichkeit bzw. Ähnlichkeit von Fällen zu isolieren und zu beschreiben, reicht es, wenn der Experte in der Lage ist, die relative Nützlichkeit der Fälle untereinander für gegebene Beispielanfragen abzuschätzen.

3.2 Akquisition von verteiltem Wissen

In vielen Anwendungsbereichen von CBR-Systemen sind die Benutzer oft selbst Experten in der jeweiligen Domäne (z.B. Diagnose-Systeme). Die Systeme haben hier im wesentlichen einen entscheidungsunterstützenden Charakter, und daher sind die Benutzer prinzipiell auch oft selbst in der Lage, offensichtliche Unzulänglichkeiten beim Retrieval zu erkennen. Bietet man den Benutzern hier die Möglichkeit für Feedback, könnte ein solches System Wissen all seiner erfahrenen Nutzer sammeln und zusammenzuführen. Dies sollte zu einer deutlichen Steigerung seiner Kompetenz führen, wovon wiederum alle Nutzer profitieren würden.

3.3 Lernen von Benutzerpräferenzen

Bei vielen Retrieval-Fragestellungen lassen sich in der Praxis nicht alle Suchkriterien explizit durch die Anfrage formulieren. Zum Beispiel wird ein Kunde in der Regel zwar die gewünschten Eigenschaften eines Produktes beschreiben, aber er wird

sicherlich nicht alle Alternativen aufzählen, für den Fall, dass das Wunschprodukt nicht verfügbar ist. Auf der anderen Seite wird jemand der eine Anfrage an eine Suchmaschine stellt, nur sein aktuelles Informationsbedürfnis beschreiben, aber nicht den kompletten Kontext seiner Anfrage (z.B. für was wird die Information benötigt, welche Anfragen hat der Nutzer in der Vergangenheit gestellt, etc.). Oft müssen beim Retrieval daher implizite Annahmen über die Wünsche und Anforderungen des Benutzers gemacht werden. Zur Repräsentation dieser Annahmen sind wissensintensive Ähnlichkeitsmaße sehr gut geeignet, wobei sich jedoch die Frage stellt, wie das notwendige Wissen akquiriert werden kann. Hier bietet sich wiederum der vorgestellte Lernansatz an, mit dessen Hilfe initiale Ähnlichkeitsmaße durch entsprechendes Feedback der Benutzer optimiert werden können.

3.4 Berücksichtigung von Adaptionmöglichkeiten

Stellt ein CBR-System die Möglichkeit bereit, Fälle nach dem Retrieval zu adaptieren (z.B. durch spezielle Adaptionenregeln), wird die Definition eines adäquaten Ähnlichkeitsmaßes zusätzlich erschwert. Um wirklich die nützlichsten Fälle ermitteln zu können, muss das Ähnlichkeitsmaß im Prinzip über alle Adaptionmöglichkeiten informiert sein. Denn ein vorab scheinbar sehr ungeeigneter bzw. „unähnlicher“ Fall, kann durch eine erfolgreiche Adaption im Endeffekt nützlicher sein, wie ein scheinbar ähnlicher aber „unadaptierbarer“ Fall. Der vorgestellte Ansatz stellt eine elegante Möglichkeit dar, vorhandenes Adaptionwissen automatisiert ins Ähnlichkeitsmaß zu transferieren. Hierzu müssen in einer Trainingsphase für eine gegebene Menge von Beispielanfragen (welche auch zufällig erzeugt werden können) alternative Fälle adaptiert und die Ergebnisse entsprechend ausgewertet werden.

3.5 Wartung

Wie in anderen wissensbasierten Systemen auch, kann auch das in Ähnlichkeitsmaßen kodierte Wissen in der Regel nicht als statisch vorausgesetzt werden, d.h. Veränderungen in der Anwendungsumgebung, den Retrieval-Anforderungen oder der Fallbasis können eine regelmäßige Wartung der Ähnlichkeitsmaße notwendig machen. Auch hier können Lernverfahren helfen, den erforderlichen Aufwand deutlich zu reduzieren.

4 Experimentelle Evaluation

Um die prinzipielle Anwendbarkeit des vorgestellten Lernansatzes zu demonstrieren, wurde eine experimentelle Evaluation in zwei der im vorherigen Abschnitt beschriebenen Anwendungsszenarien durchgeführt. Im ersten Experiment wurde versucht, die Präferenzen einer gewissen Klasse von Kunden bei dem Kauf von Gebrauchtwagen zu simulieren, und das Ähnlichkeitsmaß dahingehend zu optimieren. Das zweite Experiment zielte darauf ab, ein Ähnlichkeitsmaß im Hinblick auf verfügbare Adaptionmöglichkeiten (hier Regeln zum Umkonfigurieren von Personalcomputern) anzupassen.

In beiden Experimenten konnte die Qualität eines initialen Ähnlichkeitsmaßes durch Einsatz des vorgestellten Lernverfahrens deutlich gesteigert werden. Weiterhin wurde auch der Einfluss von verrauschten Trainingsdaten untersucht, denn die Gefahr hierfür scheint gerade im ersten Szenario recht hoch. Auch hier zeigte sich das Lernverfahren bei einer gewissen Mindestmenge an verfügbaren Daten relativ robust. Im Vergleich zum reinen Gewichtelernen besteht jedoch bei kleineren Trainingsdatensätzen ein hohes Risiko des Overfitting, was mit der,

durch die lokalen Maße bedingten, enormen Größe des Suchraums erklärt werden kann.

5 Ausblick

Der in dieser Arbeit entwickelte Ansatz hat das Potential, über das Lernen von Gewichten und die Anwendung in Klassifikationsdomänen hinaus, die Wissensakquisition und -modellierung beim Einsatz wissensintensiver Ähnlichkeitsmaße im CBR deutlich zu vereinfachen. Inwieweit dieses Potential von praktischem Nutzen ist, kann nur eine noch ausstehende Anwendung im kommerziellen Umfeld zeigen. Um auch bei kleineren Trainingsdatensätzen zuverlässige Lernergebnisse zu erzielen, scheint außerdem die Einbeziehung von leicht verfügbarem Hintergrundwissen in den Lernprozess aussichtsreich [3]. Hier ergeben sich sicherlich interessante Fragestellungen für zukünftige Forschungsarbeiten.

Literatur

- [1] A. Aamodt und E. Plaza. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications*, 7(1):39-59, 1994.
- [2] R. Bergmann. *Experience Management*. Springer, 2002.
- [3] T. Gabel. *Learning Similarity Measures: Strategies to Enhance the Optimisation Process*. Diplomarbeit, Technische Universität Kaiserslautern, 2003.
- [4] D. Lowe. Similarity Metric Learning for a Variable-Kernel Classifier. *Neural Computation*, 7, 1993.
- [5] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, 1996.
- [6] Sascha Schmitt und Bernd Schneider. Einsatzpotentiale der KI im Electronic Commerce. *Künstliche Intelligenz*, (1):5-11, 2001.
- [7] A. Stahl. Learning Feature Weights from Case Order Feedback. In *Proceedings of the 4th International Conference on Case-Based Reasoning (ICCB'2001)*. Springer, 2001.
- [8] A. Stahl. *Learning of Knowledge-Intensive Similarity Measures in Case-Based Reasoning*. Dissertation, Technische Universität Kaiserslautern, 2003.
- [9] A. Stahl und T. Gabel. Using Evolution Programs to Learn Local Similarity Measures. In *Proceedings of the 5th International Conference on Case-Based Reasoning (ICCB'2003)*. Springer, 2003.
- [10] D. Wettschereck und D.W. Aha. Weighting Features. In *Proceedings of the 1st International Conference on Case-Based Reasoning (ICCB'95)*. Springer, 1995.

Kontakt

Armin Stahl
Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH, AG Bildverstehen und Mustererkennung
Erwin-Schrödinger-Str. 57, 67663 Kaiserslautern
email: stahl@informatik.uni-kl.de



Armin Stahl (geb. 1974) studierte 1993-2000 Informatik (NF Elektrotechnik) an der Technischen Universität Kaiserslautern. Von 2000-2003 arbeitete in der AG „Künstliche Intelligenz – Wissensbasierte Systeme“ geleitet von Prof. Dr. Michael M. Richter, bei dem er im Oktober 2003 auch promovierte. Seit Anfang 2004 ist er im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH in der AG „Bildverstehen und Mustererkennung“ geleitet von Prof. Dr. Thomas Breuel tätig.