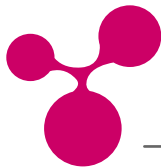


Technische Universität Dresden – Fakultät Informatik
Professur für Multimedialechnik, Privat-Dozentur für Angewandte Informatik

Prof. Dr.-Ing. Klaus Meißner
PD Dr.-Ing. habil. Martin Englien
(Hrsg.)



GENEME '09

GEMEINSCHAFTEN IN NEUEN MEDIEN

an der
Fakultät Informatik der Technischen Universität Dresden

mit Unterstützung der

3m5. Media GmbH, Dresden
GI-Regionalgruppe, Dresden
Communardo Software GmbH, Dresden
Kontext E GmbH, Dresden
Medienzentrum der TU Dresden
nubix Software-Design GmbH, Dresden
objectFab GmbH, Dresden
SALT Solutions GmbH, Dresden
Saxonia Systems AG, Dresden
T-Systems Multimedia Solutions GmbH

am 01. und 02. Oktober 2009 in Dresden

<http://www-mmt.inf.tu-dresden.de/geneme/>
geneme@mail-mmt.inf.tu-dresden.de

B.8 Explorative Suchstrategien am Beispiel von flickr.com

Birgit Wenke, Ulrike Lechner
Universität der Bundeswehr München

1 Motivation

Bei der Web2.0-Anwendung flickr.com [4] handelt es sich um ein Internetportal, in welchem Anwender ihre persönlichen Fotos einstellen können, um sie für andere Anwender im Internet erreichbar zu machen. Eine wesentliche Eigenschaft von flickr.com ist, dass die Anwender jedes ihrer Fotos mit frei gewählten Tags bezeichnen können, anhand denen die Fotos mit der portaleigenen Suchmaschine gefunden werden.

Aufgrund dieser freien Tagvergabe ist eine enorme Tagwolke entstanden, deren nachfolgend genannte Eigenschaften die Suche nach den richtigen Tags bzw. richtigen Bildern erschweren. Rechtschreibfehler, Begriffsungenauigkeiten durch Singular und Plural, unrelevante Zeichen vor den eigentlichen Tags, um das Ablagesystem zu beeinflussen und Synonymen oder Homonymen [6, 7, 9, 13] sind charakteristisch für die Tagwolke von flickr.com. Darüber hinaus besteht das sogenannte basic-level Problem: die Anwender können bei der Bestimmung ihrer Tags zwischen sehr speziellen und sehr allgemeinen Begriffen wählen [6, 9]. Unter anderem ist davon auszugehen, dass bei flickr.com rund 40% der Tags fehlerhaft sind [13].

In diesem Papier wird eine interaktive und explorative Vorgehensweise vorgestellt, die den Anwender in der Suche nach Tags unterstützt, die zu Bildern führen, die seinen subjektiven Präferenzen genügen. Die Besonderheit ist dabei, dass der Anwender seine Präferenzen nicht explizit formulieren muss. Statt dessen nutzt die Vorgehensweise das Feedback des Anwenders zu bereits vorgestellten Bildern. Grundlage der Vorgehensweise ist ein Evolutionärer Algorithmus zur Bildung von Klassifikationsregeln. Von Vorteil ist, dass das Vorgehen bei der Bildersuche für den Anwender weitestgehend unverändert bleibt, er aber von Anfang an von der Vorgehensweise profitiert, da keine vorhergehende Trainingsphase nötig ist. Der Suchvorgang selbst wird durch die stetige Verarbeitung seines Feedbacks, nachfolgend auch als Bewertung bezeichnet, beschleunigt und verbessert.

Evaluationsergebnisse zeigen, dass die Vorgehensweise innovative Anwendungsmöglichkeiten für moderne Web2.0-Anwendungen bietet und dem Anwender eine Unterstützung bei der Suche nach Tags bietet, die seinen subjektiven Interessen genügen [19].

Das Papier gliedert sich wie folgt. In Abschnitt 2 werden die theoretischen Grundlagen von Klassifikationsregeln und Evolutionären Algorithmen erläutert. Kapitel 3 befasst sich mit der Vorgehensweise. In Kapitel 4 werden die Evaluationsergebnisse erläutert und Kapitel 5 gibt eine abschließende Zusammenfassung.

2 Klassifikationsregeln und Evolutionäre Algorithmen

In den folgenden Abschnitten werden die Begriffe Klassifikationsregeln und Evolutionäre Algorithmen erläutert.

Klassifikationsregeln

Die hier vorgestellte Vorgehensweise verwendet Klassifikationsregeln zur Abbildung der Interessen des Anwenders bei der Suche nach Fotos auf flickr.com.

Allgemein gesprochen sind Klassifikationsregeln Muster, die in großen Datenmengen gefunden werden können. Klassifikationsregeln werden dazu verwendet, Objekte in Klassen einzuordnen. Damit die Klassifikation möglichst fehlerarm ist, wird nach Beziehungen zwischen Objekteigenschaften und den Klassenzugehörigkeiten der Objekte gesucht [2].

Im betrachteten Anwendungsfall von flickr.com handelt es sich bei den Tags der Fotos, um Objekte, welche den Klassen „interessant“ und „uninteressant“ zugeordnet werden. Die Klassifikationsregeln werden anhand der Tagmengen von Fotos gelernt, zu welchen der Anwender bereits ein Feedback gegeben hat. Diese Klassifikationsregeln werden dazu verwendet, noch nicht bewertete Fotos aufgrund ihrer Tags als interessant oder uninteressant zu klassifizieren.

Beispiel: Ein Anwender hat ein Foto, welches unter anderem die Tags „sunrise“ und „flower“ enthält, als interessant bewertet. Daraus wurde die Klassifikationsregel *sunrise, flower* → *interessant* gebildet. Sie ordnet neue Fotos der Klasse „interessant“ zu, wenn die Tags „sunrise“ und „flower“ in deren Tagtupeln enthalten sind, andernfalls werden sie der Klasse „uninteressant“ zugeordnet.

Die Begriffe auf der linken Seite der Klassifikationsregel werden im Folgenden als Bedingungen bezeichnet und der Begriff auf der rechten Seite als Klasse.

Evolutionäre Algorithmen

Evolutionäre Algorithmen sind stochastische, intelligente Suchverfahren, die sich an die Methoden der natürlichen Evolution anlehnen. Ziel ist es, ein globales Optimum eines Optimierungsproblems zu finden. Dieses ist definiert als eine Lösung eines Problems, die alle Nebenbedingungen der Problemstellung erfüllt und im Vergleich zu anderen Lösungsalternativen die beste Qualität aufweist [5]. In der vorgestellten Vorgehensweise wird nach Klassifikationsregeln bezüglich Tags gesucht, welche die Interessen des Anwenders bezüglich der gesuchten Bilder abbilden.

Ein Evolutionärer Algorithmus ist durch den folgenden Ablauf gekennzeichnet:

Nach der initialen Erstellung einer Anfangspopulation von Lösungen, hier Klassifikationsregeln, und deren Bewertung ist das Herzstück eines Evolutionären Algorithmus der iterative Durchlauf der folgenden Phasen: Bildung neuer Klassifikationsregeln anhand der genetischen Operatoren Rekombination und Mutation, Bewertung der neuen Klassifikationsregeln und Umweltselektion. Diese werden so lange durchlaufen, bis das globale Optimum gefunden ist oder ein Abbruchkriterium erreicht ist [16].

Im Detail heißt das: in jedem Durchlauf wird aus bereits bekannten Klassifikationsregeln (Eltern) mit Hilfe von genetischen Operatoren (Rekombination und Mutation) eine Menge neuer Klassifikationsregeln gebildet. Diese Klassifikationsregeln werden hinsichtlich ihrer Qualität bewertet und anhand eines Selektionsverfahrens (Umweltselektion) werden diejenigen Klassifikationsregeln bestimmt, welche im nächsten Durchlauf die Eltern sind. Für jede Phase eines Evolutionären Algorithmus gibt es zahlreiche Methoden, die nahezu beliebig miteinander kombiniert werden können, weswegen es verschiedenste Varianten von Evolutionären Algorithmen gibt [5].

In der Vorgehensweise werden als Rekombinationsverfahren das Diagonal Crossover und das Single-Point Crossover eingesetzt. Das Diagonal Crossover ist das explorativste Rekombinationsverfahren. Im Idealfall stammt bei einer neuen Klassifikationsregel jede Bedingung von einem anderen Elter. Hingegen wird beim Single-Point Crossover, welches das am wenigsten explorative Rekombinationsverfahren ist, der Bedingungsteil aus der Kombination der Bedingungen von genau zwei Eltern gebildet.

Die Mutation wird anhand von zwei verschiedenen Mutationsraten vorgenommen. Zum Einen beträgt die Mutationsrate zufallsbedingt 0 oder 1 und zum Anderen beträgt sie stets 1, d. h. es kommt immer zu einer Veränderung im Bedingungsteil.

Als Selektionsverfahren werden in der Vorgehensweise das Stochastic Universal Sampling und die Truncation Selection verwendet. Das Stochastic Universal Sampling ist ein probabilistisches Auswahlverfahren. Die Wahrscheinlichkeit einer Klassifikationsregel, ausgewählt zu werden ist abhängig von ihrer Qualität, wodurch auch Klassifikationsregeln mit geringer Qualität eine Selektionsmöglichkeit haben. Es ist das schnellste probabilistische Verfahren, da es in einem Selektionsvorgang eine Vielzahl von Klassifikationsregeln auswählen kann. Die Truncation Selection dagegen ist ein deterministisches Verfahren, d.h. die Selektion wird streng nach Rängen vorgenommen. Nur Klassifikationsregeln hoher Qualität, d.h. mit hohen Rängen, haben die Möglichkeit ausgewählt zu werden, wodurch sich die Gefahr des Steckenbleibens in einem lokalen Maximum erhöht [16].

3 Die Vorgehensweise

Wie bereits erläutert, navigiert der Nutzer bei flickr.com durch eine enorme Tagwolke, indem er Bilder und Tags beurteilt und ggf. den Tags von als interessant bewerteten Bildern weiter folgt. Vorab die richtigen Tags zu formulieren ist sicherlich in vielen Fällen durch den Anwender nicht leistbar – zu unterschiedlich sind individuelle Gewohnheiten und Sorgfalt, Bilder mit Tags zu versehen. Die entwickelte Vorgehensweise (vgl. Abbildung 1) ermöglicht dem Anwender eine explorative und interaktive Suche nach den zielführenden Tags.

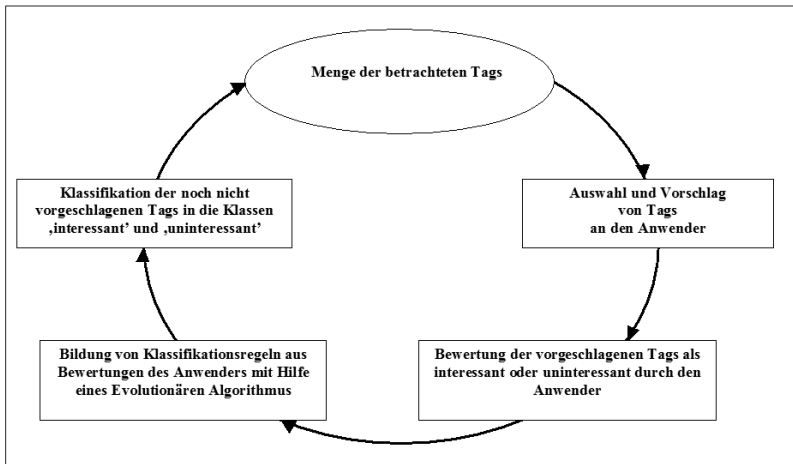


Abbildung 1: Schematischer Ablauf der Vorgehensweise [19]

Ein Evolutionärer Algorithmus erstellt mit Hilfe der Bewertungen des Anwenders Klassifikationsregeln. Durch die sofortige Anwendung und fortlaufende Verbesserung dieser Klassifikationsregeln anhand neuer Bewertungen während des Suchvorgangs werden dem Anwender nach und nach überwiegend Tags vorgeschlagen, die mit seinen Interessen übereinstimmen.

In jedem Durchlauf werden dem Anwender 5 Tagmengen vorgeschlagen. Für jede dieser Tagmengen gibt der Anwender anhand einer einfachen Eingabe an, ob die mit den vorgeschlagenen Tagmengen verbundenen Fotos für ihn von Interesse sind oder nicht. Diese Informationen werden zur Bildung von Klassifikationsregeln verwendet. Diese Klassifikationsregeln werden auf die Menge der noch nicht vorgeschlagenen Tags angewendet, wodurch diese in interessante und uninteressante Tags klassifiziert werden.

Im nächsten Durchlauf werden dem Anwender wieder 5 Tagmengen präsentiert, wobei es sich dabei um Tags handelt, die im vorhergehenden Lauf als interessant klassifiziert wurden. Auch diese Tagmengen sind vom Anwender wieder zu bewerten und diese Informationen werden zur Verbesserung der Klassifikationsregeln verwendet. Dieser Vorgang wird so lange wiederholt, bis ein Abbruchkriterium erreicht wird oder der Anwender den Programmlauf beendet [19].

Der Anwender profitiert von Anfang an von der Vorhergehensweise, da keine vorhergehende Trainingsphase nötig ist. Die Interessen des Anwenders werden erlernt, ohne dass dieser seine Interessen explizit formulieren muss. Dies ist von entscheidendem Vorteil, da die Formulierung der Interessen oftmals schwierig ist und sich die Interessen während des Suchvorgangs verändern können.

Die entwickelte Vorgehensweise unterscheidet sich von anderen Arbeiten dieses Gebiets ([1], [8], [10], [14], [15], [18]), da in dieser Arbeit Evolutionäre Algorithmen zur Bildung von Klassifikationsregeln aus dem interaktiven und iterativen Feedback des Anwenders eingesetzt werden.

Der Einsatz Evolutionärer Algorithmen für die betrachtete Aufgabenstellung bringt die folgenden Vorteile mit sich: mit einfachen Zielfunktionen kann der komplexe Suchraum mit sich verändernden Problemstellungen bearbeitet werden. Zudem können sowohl große Suchräume als auch mehrkriterielle Problemstellungen miteinbezogen werden [16].

4 Evaluation

Nachfolgend wird Einblick in die Evaluation gegeben, wobei das Evaluationsvorgehen, die zur Evaluation verwendeten Daten und die wichtigsten Evaluationsergebnisse dargestellt werden.

Evaluationsvorgehen

Bei der Evaluation wurden verschiedene evolutionäre Methoden und Anwenderinteressen miteinbezogen, um deren Einflüsse auf die Leistungsfähigkeit der Vorgehensweise zu untersuchen.

Die Evaluation der Vorgehensweise wurde durch simulierte Anwender an einer Menge von Tags des Fotoportals flickr.com vorgenommen, d.h. es wurde im Vorhinein festgelegt, welche Tags bzw. Tagmengen der Anwender bevorzugen würde. In der Evaluation wurde untersucht, inwieweit die Vorgehensweise geeignet ist, diese Tags zu identifizieren.

Die Evaluation der entwickelten Vorgehensweise wurde anhand von 150 Läufen vorgenommen. In jedem dieser 150 Läufe wurden 50 Bewertungsvorgänge durchgeführt, d. h. dem fiktiven Anwender wurden 50mal Tagmengen zur Bewertung vorgestellt. Danach wurde der Lauf abgebrochen, um eine Vergleichbarkeit der Ergebnisse zu gewährleisten.

Bewertet wurde die entwickelte Vorgehensweise anhand der durchschnittlichen Voraussagegenauigkeit, d. h. je öfter die vorgeschlagenen Tagmengen mit den im Vorhinein vom fiktiven Anwender bevorzugten Tags übereingestimmt haben, desto besser wurde die Vorgehensweise bewertet.

Dabei wurde die Vorgehensweise anhand von acht verschiedenen Verfahrensvarianten und sechs unterschiedlichen Interessensszenarien evaluiert [19].

Verfahrensvarianten

Die acht Verfahrensvarianten, die evaluiert wurden, ergeben sich aus der Kombination der bereits in Kapitel 2 genannten Rekombinations-, Mutations- und Selektionsverfahren (Tabelle 1). Es wird untersucht, inwiefern sich die Verfahrensvarianten für die betrachtete Aufgabenstellung eignen.

Tabelle 1: Verfahrensvarianten [19]

	Rekombination	Mutation	Selektion
Variante 1	Diagonal Crossover	Mutationsrate ≤ 1	Stochastic Universal Sampling
Variante 2	Diagonal Crossover	Mutationsrate ≤ 1	Truncation Selection
Variante 3	Diagonal Crossover	Mutationsrate = 1	Stochastic Universal Sampling
Variante 4	Diagonal Crossover	Mutationsrate = 1	Truncation Selection
Variante 5	Single-Point Crossover	Mutationsrate ≤ 1	Stochastic Universal Sampling
Variante 6	Single-Point Crossover	Mutationsrate ≤ 1	Truncation Selection
Variante 7	Single-Point Crossover	Mutationsrate = 1	Stochastic Universal Sampling
Variante 8	Single-Point Crossover	Mutationsrate = 1	Truncation Selection

Für jede der drei evolutionären Methoden wurde zum Einen das in der Literatur allgemein empfohlene Verfahren [16] gewählt (Diagonal Crossover, Mutationsrate ≤ 1 , Stochastic Universal Sampling) und zum Anderen das jeweils möglichst gegensätzliche Verfahren (Single-Point Crossover, Mutationsrate = 1, Truncation Selection).

Durch die Anwendung dieser jeweils gegensätzlichen Verfahren wird untersucht, ob sich im Anwendungsfall flickr.com die Empfehlungen der Literatur bestätigen.

Interessensszenarien

Zudem wurde die Vorgehensweise anhand von sechs verschiedenen Interessensszenarien evaluiert (Tabelle 2).

Tabelle 2: Interessensszenarien [19]

Szenario	Interesse
1	Das Tag ‚trip‘ muss enthalten sein.
2	Das Tag ‚blue‘ darf nicht enthalten sein.
3	Das Tag ‚old‘ oder das Tag ‚wood‘ muss enthalten sein.
4	Das Tag ‚Ireland‘ darf nicht enthalten sein oder das Tag ‚park‘ muss enthalten sein.
5	Sowohl das Tag ‚sun‘ als auch das Tag ‚light‘ dürfen nicht enthalten sein.
6	In der ersten Hälfte der Durchläufe muss das Tag ‚sand‘ enthalten sein und in der zweiten Hälfte das Tag ‚road‘.

Die ausgewählten Interessensszenarien variieren in ihrer Komplexität (Anzahl und Genauigkeit der Bedingungen). Zum Beispiel führt bei Interessensszenario 1 schon ein bestimmtes Tag („trip“) zu den interessanten Bildern. Hingegen wechselt bei Szenario 6 das Interesse, d.h. in der zweiten Hälfte der Durchläufe wird ein anderes Tag („road“) bevorzugt als in der ersten Hälfte („sand“).

Evaluationsdaten

Für die Evaluation wurde ein Ausschnitt von flickr.com gewählt. Dieser Ausschnitt wurde erstellt, indem verschiedene Suchbegriffe in das flickr-Portal mit einer ODER-Verknüpfung eingegeben wurden. Bei den Suchbegriffen wurden Begriffspaare gewählt, wobei die Begriffe der Paare zueinander aus inhaltlich unterschiedlichen Bereichen stammen und die Begriffe der Paare selbst inhaltlich große Gemeinsamkeiten aufweisen. Folgende Suchbegriffe wurden verwendet, wobei die Begriffspaare mit Schrägstrich gekennzeichnet sind: ocean/sea, flower/trees, sand/beach, summer/sun, winter/ice, mountain/nature, camera/light. Es wurden Begriffspaare gewählt, um dem Suchraum einen schwachen Zusammenhang zu geben. Die inhaltlichen Unterschiede zwischen den Begriffspaaren wurden gewählt, um die freie Tagwahl der Anwender abzubilden. Mit dieser Suche wurden insgesamt 68 verschiedene Tags ausgewählt [19]. (flickr.com 22.05.2007)

Evaluationsergebnisse

Nachfolgend werden die Evaluationsergebnisse der verschiedenen Verfahrensvarianten und Interessensszenarien veranschaulicht.

Evaluationsergebnisse der Verfahrensvarianten

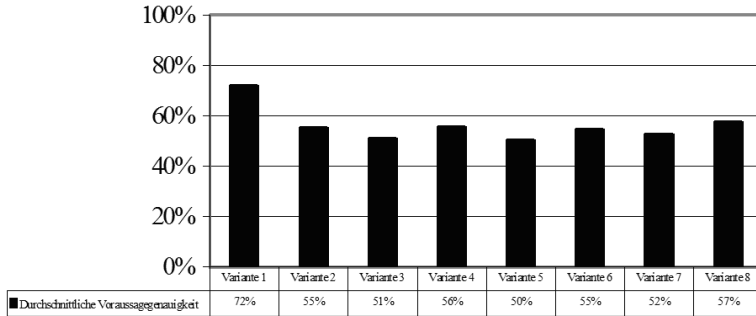


Abbildung 2: Durchschnittliche Voraussagegenauigkeit der acht Verfahrensvarianten [19]

Die durchschnittliche Voraussagegenauigkeit, die sich für jede der acht Verfahrensvarianten ergeben hat, ist in Abbildung 2 veranschaulicht. Die höchste durchschnittliche Voraussagegenauigkeit in Höhe von 72% wurde von der Variante 1 erzielt. Diese Variante verwendet als Rekombinationsverfahren das Diagonal Crossover, als Mutationsverfahren ein Verfahren mit einem Mutationsschritt von 0 oder 1 und als Selektionsverfahren das Stochastic Universal Sampling [19]. Damit bestätigen diese Evaluationsergebnisse die Erwartungen der Literatur, da die Variante 1 die in der Literatur empfohlenen Verfahren einsetzt.

Evaluationsergebnisse der Interessensszenarien

Die durchschnittliche Voraussagegenauigkeit, die sich für jedes der sechs Interessensszenarien und der acht Verfahrensvarianten ergeben hat, ist in Abbildung 3 veranschaulicht, wobei die Szenarien aufsteigend nach ihrer Komplexität aufgeführt sind.

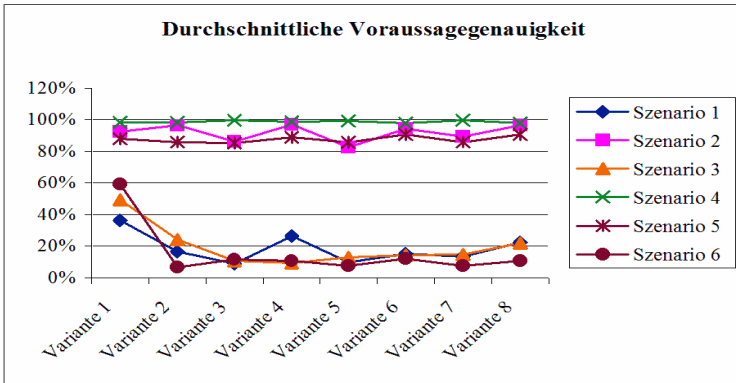


Abbildung 3: Durchschnittliche Voraussagegenauigkeit je Interessenszenario und Verfahrensvariante [19]

Der Vergleich der Ergebnisse zeigt, dass das Interessenszenario 4 mit Ergebniswerten im Bereich [98%;100%] für alle acht Verfahrensvarianten die besten Werte aufweist. Die Interessensszenarien 2 und 5 weisen mit Werten im Bereich [83%;97%] ähnlich gute Werte auf, wohingegen die Szenarien 1, 3 und 6 deutlich niedrigere Ergebniswerte erzielt haben. Diese liegen im Bereich [7%;59%] [19].

Ein Einfluss der Komplexität der Interessen spiegelt sich wie erwartet nicht in den Ergebniswerten nieder. Hierfür hätte beispielsweise das Szenario 1 einen sehr hohen Wert aufweisen müssen und das Szenario 5 einen vergleichsweise hohen. Beides ist nicht der Fall..

5 Zusammenfassung

Die vorgestellte Vorgehensweise bietet eine innovative Unterstützung des Anwenders bei der explorativen Suche nach den gesuchten Tags. Der entscheidende Unterschied zu anderen Verfahren liegt darin, dass die Suche nach subjektiven Präferenzen des Anwenders vorgenommen wird, ohne dass der Anwender diese explizit formulieren muss. Dies wurde durch eine individuelle Anpassung eines Evolutionären Algorithmus erreicht, mit welchem Klassifikationsregeln erstellt werden, die die Interessen des Anwenders aus seinem Feedback lernen.

Die maximale durchschnittliche Voraussagegenauigkeit, welche im Rahmen der Evaluation erreicht wurde, beläuft sich auf 73% und erzielt damit eine vergleichbare Qualität wie andere Klassifikationsalgorithmen (Ripper [3], CMAR [11], CBA [12] und C4.5 [17]). Es ist jedoch anzumerken, dass die Klassifikation in der vorgestellten Vorgehensweise erschweren Bedingungen ausgeliefert ist, da der

Evolutionäre Algorithmen mit den Methoden Rekombination, Mutation und Selektion ständig versucht ist, den Suchraum mit neuen Klassifikationsregeln, die von bereits bewährten Klassifikationsregeln abweichen, zu diversifizieren. Dies bedeutet für die Vorgehensweise, dass eine Voraussagegenauigkeit von 100% bei der Verwendung von evolutionären Algorithmen nur schwer denkbar wäre.

Bezüglich der Verfahrensvarianten hat sich gezeigt, dass mit der Wahl der evolutionären Methoden die Leistungsfähigkeit der Vorgehensweise beeinflusst werden kann und die Empfehlungen der Literatur haben bei allen untersuchten evolutionären Methoden zu den besten Ergebnissen geführt.

Die Evaluation verschiedener Interessenszenarien hat gezeigt, dass die Komplexität der Interessen kein Einfluss auf die Leistungsfähigkeit hat.

Der Einsatz der entwickelten Vorgehensweise ist in all den Bereichen denkbar, in denen große Mengen bisher nicht ausreichend ausgewerteter Daten in Form von klassischen Datenbanken vorliegen, wie beispielsweise in der Meteorologie, der Medizin, der Statistik, der Astronomie oder der Wirtschaft. Außerdem kann sie einen Nutzer beim interaktiven Browsen bzw. Suchen in Community Anwendungen unterstützen.

Literatur

- [1] Alvarez, J. L. / Mata, J. / Riquelme, J. C. (Oblic 2001): Oblic: Classification Systems using Evolutionary Algorithm, in: Proc. of the 6th Int'l. Work-Conference on Artificial and Natural Neural Networks 2001, S. 644-651
- [2] Bäck, Thomas / Schütz, Martin, Evolutionäre Algorithmen im Data Mining, in: Handbuch Data Mining im Marketing, hrsg. v. Hippner, Hajo / Küsters, Ulrich / Meyer, Matthias / Wilde, Klaus, Braunschweig, Wiesbaden 2001, S. 403-426
- [3] Cohen, W., Fast effective rule induction, in: ICML 1995, S. 115-123
- [4] flickr, www.flickr.com, Zugriff am 04.01.2007
- [5] Freitas, Alex A., Data Mining and knowledge discovery with evolutionary algorithms, Berlin u.a. 2002
- [6] Golder, Scott A. / Huberman, Bernardo A., The structure of collaborative tagging systems, in: Journal of Information Science (32) 2006, S. 198-208
- [7] Guy, Marieke / Tonkin, Emma, Folksonomies, tidying up tag?, in: D-Lib Magazine (12/1) 2006
- [8] Isasi, P. / Fernandez, F. (Evolutionary Approach 2003): Evolutionary approach to overcome initialization parameters in classification problems, in: Proc. of the 7th Int'l. Work-Conference on Artificial and Natural Neural Networks 2003, S. 254-261
- [9] Kroski, Ellyssa, The hive mind: folksonomies and user-based tagging, infotangle.blogspot.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-taggaing/, Zugriff am 23.05.2007

-
- [10] Kshetrapalapuram, K. K. / Kirley, M. (Mining Classification Rules 2005): Mining classification rules using evolutionary multi-objective Algorithms, in: Proc. of the 9th Conference on Knowledge-based Intelligent Information and Engineering Systems 2005, S. 959-965
- [11] Li, W. / Han, J. / Pei, J., CMAR: Accurate and efficient classification based on multiple-class association rule, in: Proc. of ICDM 2001
- [12] Liu, Bing / Hsu, Wynne / Ma, Yiming, Integrating Classification and association rule mining, in Proc. of the KDD 1997, S. 80-86
- [13] Mcgregor, George / McCulloch, Emma, Collaborative tagging as a knowledge organisation and resource discovery tool, in: Library Review 55/5, S. 291-300
- [14] Niimi, Ayahiko / Tazaki, Eiichiro (Extended Genetic Programming 2001): Extended genetic programming using Apriori algorithm for rule discovery, in: Post-Proc. of the Joint JSAI Workshop on New Frontiers in Artificial Intelligence 2001, S. 525-532
- [15] Niimi, Ayahiko / Tazaki, Eiichiro (Using Genetic Programming 2000): Rule discovery technique using genetic programming combined with Apriori algorithm, in: Proc. of the 3rd International Conference on Discovery Science 2000, S. 273-277
- [16] Pohlheim, Hartmut, Evolutionäre Algorithmen: Verfahren, Operatoren und Hinweise für die Praxis, New York u. a. 2000
- [17] Quinlan, J. R., C4.5: Programs for machine learning, San Mateo 1993
- [18] Sikora, M. (Fuzzy Rules Generation 2005): Fuzzy rules generation method for classification problems using rough sets and genetic algorithms, in: Proc. of the 10th Int'l. Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing 2005, S. 383-391
- [19] Wenke, Birgit, Anwenderspezifische Reduzierung von Mengen interessanter Assoziationsregeln mittels Evolutionärer Algorithmen, Aachen 2008