

Der Artbegriff in der Genetischen Programmierung

Vom Fachbereich Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte

D i s s e r t a t i o n

von

Dipl.-Ing. Eva Brucherseifer

Geboren am 02.02.1973 in Worms

Referent:	Prof. Dr.-Ing. Jürgen Adamy
Korreferent:	Prof. Dr.-Ing. Klaus Hofmann
Tag der Einreichung:	11.09.2010
Tag der mündlichen Prüfung:	20.01.2011

D 17

Darmstadt 2010

Berichte aus der Automatisierungstechnik

Eva Brucherseifer

Der Artbegriff in der Genetischen Programmierung

D 17 (Diss. TU Darmstadt)

Shaker Verlag
Aachen 2011

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Zugl.: Darmstadt, Techn. Univ., Diss., 2010

Copyright Shaker Verlag 2011

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der Übersetzung, vorbehalten.

Printed in Germany.

ISBN 978-3-8322-9942-2

ISSN 0945-4659

Shaker Verlag GmbH • Postfach 101818 • 52018 Aachen

Telefon: 02407 / 95 96 - 0 • Telefax: 02407 / 95 96 - 9

Internet: www.shaker.de • E-Mail: info@shaker.de

Vorwort

Diese Arbeit begann am Fachgebiet Regelungstheorie und Robotik an der Technischen Universität Darmstadt, an der ich zunächst von Dez. 1998 bis 2003 als wissenschaftliche Assistentin beschäftigt war.

In dieser Zeit habe ich zunächst die Software SMOG von Dr.-Ing. Peter Marrenbach mit einer umfangreichen Vorarbeit übernommen. Seine Arbeit war für mich immer wegweisend. Während meiner Forschung am Fachgebiet und in der Arbeit mit Studenten hat mir mein Betreuer Prof. Dr.-Ing. Adamy großen Freiraum gewährt und ich konnte vieles lernen und ausprobieren. Er hat sicherlich die ein oder anderen Abwege akzeptieren müssen, nicht ohne immer wieder geduldig in Richtung der Dissertationsschrift zu weisen. Ich danke ihm für die schöne und lehrreiche Zeit am Fachgebiet.

Bei dem Betreuer meiner Studien- und Diplomarbeit sowie späteren Kollegen und Freund Herrn Dipl.-Ing. Karl-Heinz Hohm möchte mich für seine stetige immer große Hilfsbereitschaft in allen Fragen bedanken. Ebenso danke ich meinem ehemaligen Kollegen Dr. rer. nat. Roland Kempf für vielfältige intensive Diskussionen zu fachlichen Themen und mehr.

Als Teil der praktischen Umsetzung der Software unterstützte mich Prof. em. Dr.-Ing. H.-J. Hoffmann, der mich zu den Methoden der Softwarearchitektur geführt hat. Von ihm habe ich die Liebe zu schöner und guter Software gelernt, die letztlich zur Gründung meiner Softwareentwicklungsfirma basysKom geführt hat.

In der Folge wurde Forschung und Promotion zu einem Hobby, das mich nun über 7 weitere Jahre begleitet hat. An zahllosen Wochenenden und Urlauben habe ich gelernt, immer wieder neugierig zu sein und aus der Vielfalt der Möglichkeiten einen roten Faden herauszuarbeiten. Nach jeder Pause stellte sich mit etwas Abstand ein weiterer Knoten als lösbar heraus. Dabei war das Thema immer wieder fesselnd.

Dr. Stefan Werden möchte ich danken, dass er mich gelehrt hat, auch das vermeintlich offensichtliche zu formulieren und weiterzudenken. Manchmal sind die Dinge einfacher als man denkt und doch liegt die Kunst und die Arbeit genau in der Einfachheit.

Meinen Eltern danke ich für ihre stetige Unterstützung und die vielfältigen Hinweise und Vorschläge.

Nicht zuletzt aber danke ich meinem Mann Matthias, der über alle Jahre hinweg mich stetig unterstützt und mein Verschwinden an den Computer akzeptiert hat.

So bleibt am Ende eine wertvolle Erkenntnis: Forschung und wissenschaftliche Arbeit kann nur der eigenen Neugier dienen und aus tiefem innerem Antrieb erfolgen. So wünsche ich anderen Forschenden, dass sie die nötige Ruhe und Fokussierung finden und sich aus der wachsenden Erkenntnis heraus motivieren und daran erfreuen können.

Darmstadt, im Januar 2010

Eva Brucherseifer

Inhaltsverzeichnis

1	Einleitung	1
2	Struktur und Aufbau der Genetischen Programmierung	7
2.1	Evolutionäre Algorithmen: Eine Nachbildung der natürlichen Evolution	8
2.2	Genetische Programmierung zur Modellierung komplexer Systeme	14
2.2.1	Die datenbasierte Modellstruktursuche mit SMOG	16
2.2.2	Evaluierung von Prozessmodellen mit <i>emucalc</i>	22
2.2.3	Aufbau der Software: <i>smog</i> , <i>emucalc</i> und <i>emulize</i>	26
2.3	Testbeispiel: Symbolische Regression	28
2.4	Zur Analyse verfügbare Datenbasis	38
2.5	Zusammenfassung und Bewertung	40
3	Analyse und Visualisierung	42
3.1	Ziel der Analyse	42
3.2	Evolutionäre Optimierung: Pfade durch das Fitnessgebirge	45
3.3	Aktuelle Analysetechniken für die Genetischen Programmierung	49
3.4	Beitrag der Arbeit	60
4	Die Systematik der Lebewesen in der Biologie	62
4.1	Die Evolution von Individuen	63
4.2	Arten und ihre Entstehung	65
4.3	Systematik, Taxonomie und Phylogenetik	71
4.4	Moderne Phylogenetische Methoden (Tree Of Life)	73
4.4.1	Molekulare Phylogenetik	74
4.4.2	Vom Alignment zum Stammbaum	75
4.5	Anwendbarkeit auf Genetische Programmierung	79
4.5.1	Evolutionäre Optimierung als Makroevolution	79
4.5.2	Der Artbegriff: Entwicklung einer Vorgehensweise	80

5 Ähnlichkeitsmaße und Merkmale für den Vergleich von Individuen	84
5.1 Vergleich von Individuen anhand von Merkmalen und Distanzen	85
5.1.1 Gemeinsame Anforderungen	87
5.1.2 Distanzen im metrischen Raum	87
5.1.3 Merkmale im Vektorraum	88
5.1.4 Zusammenhang zwischen Ähnlichkeit und Distanz	90
5.2 Strukturelle Distanzmaße	92
5.2.1 Der Genotyp als Graph	92
5.2.2 String-Editierdistanz	100
5.2.3 Baum-Editierdistanz	104
5.2.4 Qualitative strukturelle Distanzmaße	109
5.2.5 Bewertung der strukturellen Distanzmaße	110
5.3 Semantische Distanzmaße	116
5.4 Phylogenetische Distanzmaße	118
5.5 Ausgewählte Merkmale	121
5.5.1 Strukturelle Merkmale	121
5.5.2 Semantische Merkmale	122
5.5.3 Phylogenetische Merkmale	124
5.5.4 Bewertung der Merkmale	125
5.6 Zusammenfassung und Bewertung	126
6 Klassifikation in Arten	127
6.1 Der Artbegriff für Evolutionäre Algorithmen	128
6.2 Clusteranalyse zur Artenbildung konkurrierender Individuen	129
6.2.1 Hierarchische Cluster	130
6.2.2 Visualisierung von Clusterstrukturen	137
6.3 Strukturierte Vorgehensweise zur Bildung von Arten	141
6.3.1 Vorbereitung	142
6.3.2 Clustering	144
6.3.3 Validierung und Vergleich	145
6.4 Experimenteller Vergleich der Methoden zur Artenbildung	154
6.5 Zusammenfassung und Bewertung	162
7 Visualisierung von Arten	164
7.1 Charakteristika einer Art	164
7.2 Visualisierung der Entwicklung von Arten: Der Artengraph	170
7.3 Anwendungsbeispiel: Reglerentwurf Antennenarm	178
7.3.1 Aufgabenstellung	178
7.3.2 Konfiguration des GP und verwendete Werkzeuge	181

7.3.3	Regleroptimierung mit Genetischer Programmierung unter Einsatz des Artengraphen	184
7.4	Zusammenfassung und Bewertung	198
8	Zusammenfassung	201
A	Pseudocode-Syntax	206
B	Artengraphen zum Anwendungsbeispiel: Reglerentwurf Antennenarm	207
	Literaturverzeichnis	213
	Lebenslauf	224

Kurzfassung

In der Automatisierung wird die Genetische Programmierung zur Modellbildung von komplexen Systemen und zum Reglerentwurf eingesetzt. Die Genetische Programmierung nimmt die biologische Evolution als Vorbild und ahmt sie nach. Mathematische Modelle werden dazu als Individuen kodiert und über eine Anzahl von Generationen hinweg evolutionär selektiert und verändert.

Bei einem Evolutionslauf über mehrere Generationen fällt mit vielen Individuen eine kaum zu überblickende, große Datenmenge an. Deshalb sind die Auswirkungen einer für die Genetische Programmierung gewählten Konfiguration für den Anwender nur schwer zu kontrollieren. Für den Modellbildner ist es wichtig, den Ablauf und die Ergebnisse der Genetischen Programmierung schnell analysieren und bewerten zu können. Diese Arbeit bestimmt, erprobt und bewertet daher Methoden und Verfahren, die eine tiefergehende Analyse des Evolutionsvorgangs durch eine geeignete Visualisierung ermöglichen.

Das neuartige Artenkonzept fasst ähnliche Individuen zu Arten zusammen, um sie so visualisieren zu können. Die Arten ermöglichen eine konzentrierte Analyse, die mit allen Individuen aufgrund deren großer Anzahl nicht möglich wäre. Die Artenbildung basiert auf der Bestimmung der Ähnlichkeit von Individuen. Bekannte und neue Merkmale und Distanzmaße zum Vergleich von Individuen werden daher vorgestellt und bezüglich ihrer Anwendbarkeit für die Artenbildung neu bewertet.

Zur Darstellung der evolutionären Abstammungslinien der Arten wurde der Artengraph entwickelt, der die aus den zwischen Individuen berechneten Distanzen, Abstammungen und Fitnesswerten aggregiert und so die Divergenz von Populationen visualisiert. Der Artengraph ist dem aus der Biologie bekannten Stammbaum nachempfunden und ermöglicht eine konzentrierte, effiziente und zeitnahe Analyse eines Optimierungslaufs.

Anhand einer Reglerentwicklung für einen Antennenarm konnte die Anwendung der entwickelten Visualisierungstechniken erfolgreich gezeigt werden. Die Analyse mittels Artengraph der evolutionären Entwicklung von Individuen während eines Optimierungslaufs führte zu einer verbesserten Parametrisierung der Genetischen Programmierung. Dies konnte mit nur wenigen Optimierungsläufen und damit geringem Zeitaufwand erreicht werden.