



WIRTSCHAFT
HOCHSCHULE MAINZ
UNIVERSITY OF
APPLIED SCIENCES

HIER STEHT DER TITEL DER ARBEIT
MÖGLICHER UNTERTITEL

Bachelorarbeit

eingereicht im: Januar 2022

von: Martin Mustermann
geboren am 23. März 1982
in Mainz

Matrikelnummer: 012345

Studiengang: Wirtschaftsinformatik, dual

Betreuer: Prof. Dr. Martin Huschens

Inhaltsverzeichnis

Stichwortverzeichnis	iii
Abbildungsverzeichnis	iv
Tabellenverzeichnis	v
Abkürzungsverzeichnis	vi
1 Einleitung	1
1.1 Ziel der Arbeit	2
1.2 Aufbau der Arbeit	2
2 Ein Kapitel	3
2.1 Bilder	3
2.2 Zitate	3
2.3 Editor	3
2.4 Mathematik	4
3 Summary, Conclusions, and Further Work	8
3.1 Summary	8
3.2 Conclusions	8
3.3 Further Work	9
Literaturverzeichnis	10

Stichwortverzeichnis

Genetische Algorithmen

Redundanz

Komponentenbasierte Softwareentwicklung

Abbildungsverzeichnis

2.1	Das hier steht im Abbildungsverzeichnis.	4
2.2	Dichten univariater Normalverteilungen	5

Tabellenverzeichnis

2.1	Eine sehr einfache Tabelle	6
2.2	Performance of GAs using different types of representations for deceptive tree problems of different sizes and with different T_{opt} (arbitrary tree, MST, and star)	7

Abkürzungsverzeichnis

GA Genetische Algorithmen, Genetischer Algorithmus

1 Einleitung

Eines der zentralen Gebiete der Wirtschaftsinformatik ist die Entwicklung und Nutzung von Planungs- und Dispositionssystemen. Derartige Systeme unterstützen Entscheider in Unternehmen bei der Erstellung von kurz-, mittel- und langfristigen Plänen für die Durchführung von betrieblichen Aufgaben und Prozessen (Mertens u. a. 2001; Scheer 1997; Goldberg 1989). In den letzten Jahren ist ein zunehmendes Interesse an der Erweiterung der bloßen Planung von Ressourcen und dem Finden von zulässigen Lösungen für komplexe Planungsprobleme hin zum Ermitteln von optimalen Lösungen für Planungsprobleme erkennbar. In diesem Rahmen werden die in Planungs- und Dispositionssystemen eingesetzten Planungsverfahren, welche eine zulässige Lösung (z.B. eine zulässige Struktur eines Kommunikationsnetzes) ermitteln, zunehmend um Metaheuristiken ergänzt, welche versuchen, eine optimale Lösung (z.B. die Struktur des Kommunikationsnetz mit den geringsten Kosten) zu finden. Dies führt entscheidungstheoretisch dazu, dass das Prinzip der Satisfizierung (Finden zulässiger Lösungen mit Hilfe von heuristischen Verfahren) durch das Prinzip der Extremierung (Finden optimaler Lösungen mit Hilfe von Metaheuristiken oder analytischen Verfahren) abgelöst wird.

Metaheuristiken entstanden an der Schnittstelle zwischen der Informatik und Künstlichen Intelligenz und stellen Methoden zur Verfügung, welche dem Nutzer in die Lage versetzen komplexe Probleme in kurzer Zeit optimal (bzw. annähernd optimal) zu lösen. Derartige Methoden zeichnen sich dadurch aus, dass sie einfache, oft der Natur entlehnte Prinzipien („naturinspiriert“) verwenden, wodurch eine zügige Modellierung und effiziente Lösung von Planungsproblemen ermöglicht wird.

In den letzten Jahren hat einhergehend mit der zunehmenden Vernetzung zwischen Unternehmen die Planung des Aufbaus von Kommunikationsnetzwerken zunehmend an Bedeutung gewonnen. Kommunikationsnetzwerke sind mittlerweile für Unternehmen ein integraler Bestandteil der Unternehmensinfrastruktur, und es muss sichergestellt werden, dass deren Aufbau und Struktur an die entsprechenden Anforderungen angepasst sind. Darüber hinaus stellen für Unternehmen im Telekommunikationsbereich (z.B. Netzanbieter, „Application Service Provider“, „Internet Service Provider“ etc.) die Planung von Kommunikationsnetzwerken Kernaufgaben dar, welche entscheidend für deren Ertragskraft und Wettbewerbsfähigkeit sind...

Bei der Lösung von komplexen Planungsproblemen mit Hilfe von Metaheuristiken muss allerdings generell berücksichtigt werden, dass die verwendeten Verfahren auf die Charakteristika und Eigenschaften des zu lösenden Problems abgestimmt werden müssen. Metaheuristiken liefern nur dann gute Ergebnisse, wenn sie entsprechend der Eigenschaften der zu lösenden Planungsprobleme ausgewählt werden (Wolpert u. Macready 1995); falls die verwendeten

Lösungsverfahren nicht zum Problem passen bzw. nicht an das zu lösende Problem angepasst werden können, führt dies unweigerlich zu schlechten Lösungen.

1.1 Ziel der Arbeit

Der vorliegende Beitrag beschäftigt sich mit der Entwicklung von problemangepassten Entscheidungsunterstützungssystemen (EUS) auf der Basis von Metaheuristiken für die optimale Lösung von Kommunikationsnetzwerkproblemen. Es wird demonstriert, wie sich durch eine vor der eigentlichen Problemlösung durchgeführten Problemanalyse Handlungsempfehlungen für den geeigneten Aufbau von EUS ableiten lassen. EUS (und als zentrales Element die darin enthaltenen Metaheuristiken), welche die aus der Problemanalyse abgeleiteten Handlungsempfehlungen verwenden, können so wesentlich bessere Lösungen liefern als vergleichbare EUS, welche kein problemspezifisches Wissen verwenden. Der Beitrag zeigt, wie ein problemangepasstes EUS für die Planung und Optimierung von baumförmigen Kommunikationsnetzwerken („Optimal Communication Spanning Tree“-Problem) entwickelt werden kann. Im Rahmen der Problemanalyse wird untersucht, welche Eigenschaften optimale Lösungen des OCST-Problems aufweisen. Hierbei zeigt sich, dass optimale Lösungen des OCST-Problems eine große Ähnlichkeit zum minimal spannenden Baum (MST) aufweisen. Basierend auf dieser Erkenntnis kann eine schon vorhandene Problemrepräsentation („link-and-node-biased“-Kodierung) so angepasst werden, dass Lösungen ähnlich dem MST mit größerer Wahrscheinlichkeit repräsentiert werden. Ein Vergleich der Leistungsfähigkeit von unterschiedlichen Metaheuristiken zeigt, dass die Verwendung einer derart problemangepassten Repräsentation die Leistungsfähigkeit von EUS deutlich steigern kann.

1.2 Aufbau der Arbeit

Hier den Aufbau der Arbeit erläutern.

2 Ein Kapitel

Hier kommt eine kurze Zusammenfassung des Kapitels hin.

2.1 Bilder

In Abbildung 2.1 findet man ein schönes Beispiel für das Paket subfig. Subfig ist sehr praktisch beim Setzen von mehreren Bildern in ein Bild. Man beachte, dass man mit dem Befehl `ref` auf Bilder verweisen kann, wenn diesen mit dem Befehl `label` eine Marke zugewiesen wurde.

Zum Zeichnen von plots ist gnuplot sehr zu empfehlen (vgl. Abbildung 2.1a). Ist perfekt für zwei oder dreidimensionale Plots und schaut um Welten besser aus als Excel.

Zum Malen von Bildern ist das Programm xfig oder jfig sehr gut geeignet (vgl. Abbildung 2.1b). Schaut zwar am Anfang ein etwas seltsam aus, ist aber sehr mächtig.

2.2 Zitate

Es gibt mögliche Arten von Verweisen. In Kotz u. a. (2000, S. 120) wurde gezeigt, dass DA schreiben schön ist. Weiterhin sind manche Sachen ganz toll (Kotz u. a. 2000, S. 120). Eine Dokumentation zum Gebrauch des natbib Styles (ermöglicht DIN-konforme Zitate) findet sich im Ordner literatur.

2.3 Editor

Als Editor für Fortgeschrittene ist xemacs gut geeignet. Bei Verwendung von xemacs ist der Gebrauch der refTeX und auctex-pakete zu empfehlen. Als Rechtsschreibprogramm sollte aspell verwendet werden und folgender Code in init.el eingefügt werden.

```
(require 'iso-cvt)
(add-hook 'LaTeX-mode-hook
(function (lambda ()
;; Setze Anführungszeichen etc. fuer Style german
(TeX-run-style-hooks "german")
;;
;; Lade Buffer und wandle nach ISO Latin-1:
(format-encode-buffer 'plain)
)))
```

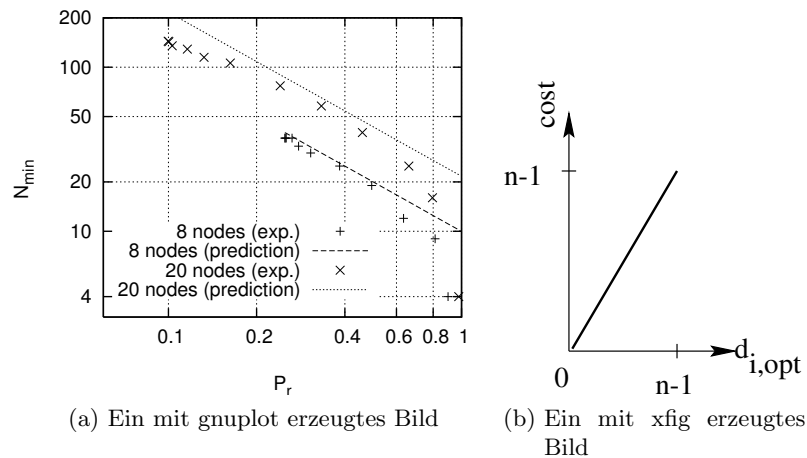


Abbildung 2.1: Beispiele für schöne Bilder.

```
(setq ispell-silently-savep t) ;save new words in pdict without questioning
(setq ispell-help-in-bufferp 'electric) ;get a better help buffer
(setq ispell-program-name "ispell")
(setq ispell-extra-args '(W2"))
```

```
(column-number-mode t)
```

```
(custom-set-variables
'(paren-mode 'sexp nil (paren)))
```

```
(setq reftex-plugin-into-AUCTeX t)
(require 'reftex "reftex" t)
(turn-on-reftex) ; use reftex
```

Neulinge können auch einen sonstigen Editor oder Sachen wie Texnic Center, etc verwenden.

2.4 Mathematik

In latex kann man Formeln wunderschön setzen.

As the following stuff is in English, we must change the hyphenation style.

The OCST problem is defined as follows. Let $G = (V, E)$ be a connected, undirected graph with $n = |V|$ nodes and $m = |E|$ edges. There are communication, or transportation demands, between every pair of nodes. An $n \times n$ demand matrix $R = (r_{ij})$ specifies the demands, where r_{ij} is the amount of traffic required between location v_i and v_j . Similarly, an $n \times n$ distance matrix

$W = w_{ij}$ determines the distance weights between each pair of sites. A tree $T = (V, F)$ where $F \subseteq E$ and $|F| = |V| - 1$ is called a *spanning tree* of G if it connects all the nodes. The weight $w(T)$ of the spanning tree is the weighted sum over all pairs of vertices of the cost of the path between the pair in T . The communication cost over the tree T is defined as

$$w(T) = \sum_{i,j \in V} w_{ij} b_{ij}, \quad (2.1)$$

where $B = b_{ij}$ denotes the traffic flowing directly and indirectly across the edge connecting nodes i and j . It is calculated according to the structure of T . T is the optimal communication spanning tree if $w(T) \leq w(T')$ for all other spanning trees T' . The OCST problem becomes the minimum spanning tree (MST) problem if $w(T) = \sum_{i,j \in V} w_{ij}$. Then, T is the simple minimum spanning tree if $w(T) \leq w(T')$ for all other spanning trees T' .

Cayley's formula identifies the number of spanning trees on n nodes as n^{n-2} (Cayley 1889). Furthermore, there are n different stars on a graph of n nodes. The similarity between two spanning trees T_i and T_j can be measured using the distance $d_{ij} \in \{0, 1, \dots, n-1\}$ as

$$d_{ij} = \frac{1}{2} \sum_{u,v \in V, u < v} |l_{uv}^i - l_{uv}^j|,$$

where l_{uv}^i is 1 if an edge from u to v exists in T_i and 0 if it does not exist in T_i . The number of edges that two trees T_i and T_j have in common is $n - 1 - d_{ij}$.

Ab jetzt wieder deutsche Trennung.

In Abbildung 2.2 gibt es noch ein Beispiel für ein Bild und in Tabelle 2.1 gibt es ein einfaches Bild. Tabelle 2.2 stellt eine etwas kompliziertere Tabelle dar.

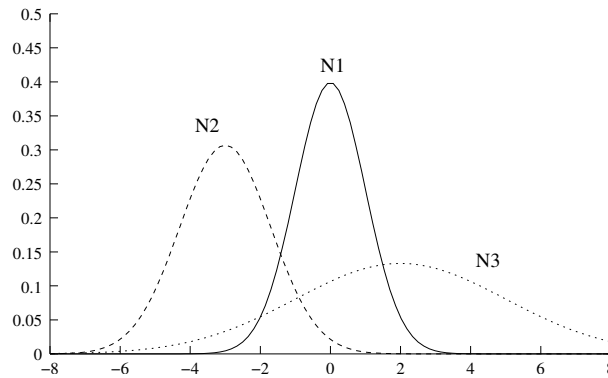


Abbildung 2.2: Dichten univariater Normalverteilungen

Beim Aufbau von Kommunikationsnetzen stehen Unternehmen vor dem Problem, eine Menge von unterschiedlichen Standorten so durch Leitungen zu verbinden, dass bestimmte Qualitätsanforderungen (z.B. Sicherheit, Bandbreite, Zuverlässigkeit, Skalierbarkeit, etc.) eingehalten werden und gleichzeitig die

Tabelle 2.1: Eine sehr einfache Tabelle

linksbündig	zentriert	rechtsbündig
1	2	3,141

Netzstruktur so gewählt wird, dass die Kosten des Netzaufbaus minimiert werden. Unternehmen, für welche der Aufbau von Kommunikationsnetzwerken relevant ist, lassen sich in zwei Gruppen unterteilen:

Auf der einen Seite sind Unternehmen zu finden, für die Kommunikationsnetzwerke ein wichtiger Teil ihrer Unternehmensinfrastruktur sind (z.B. Banken, Verlagshäuser, Bahn, Bundeswehr, Unternehmen mit mehreren Niederlassungen, etc.) und welche daher die Kommunikationsnetzwerke oft selbst betreiben. Ein Beispiel hierfür ist die DATEV eG, welche über ein eigenes „Genossenschaftsnetz“ mehrere Dutzend Niederlassungen an die Zentrale in Nürnberg anbindet. über dieses Kommunikationsnetz wird sowohl der gesamte Daten- als auch Telefonverkehr innerhalb des Unternehmens abgewickelt....

Beim Aufbau von Kommunikationsnetzwerken kann unterschieden werden zwischen vermaschten und baumförmigen Netzwerken (vergleiche Abbildung 2.2). In vermaschten Netzen existieren mehrere unterschiedliche Wege für den Transport von Daten zwischen zwei Standorten. Daher muss bei vermaschten Netzwerken das Routing der Daten durch das Netzwerk gesteuert werden (wie z.B. im Internet). In der Regel existieren mehrere Wege zwischen zwei Standorten und es muss mit Hilfe von Routingmechanismen festgelegt werden, welchen Weg die Daten durch das Netz nehmen sollen (wie z.B. durch Routingtabellen für den IP-Verkehr im Internet). Diese Entscheidung wird unter anderem durch die Entfernung von Standorten als auch durch die Auslastung der unterschiedlichen Leitungen beeinflusst. Die Ermittlung von Routingtabellen ist in der Praxis recht aufwendig und es besteht die Gefahr, dass Nachrichten Umwege durch das Netz nehmen oder sogar im Kreis laufen. Vermaschte Netzwerke bieten allerdings eine erhöhte Ausfallsicherheit, da beim Vorhandensein von mehreren Wegen zwischen zwei Knoten Daten auch dann noch ausgetauscht werden können, wenn ein Teil des Netzwerkes ausfällt.

Baumförmige Netze stellen einen Sonderfall von vermaschten Kommunikationsnetzwerken dar. Hierbei existiert genau ein möglicher Weg zwischen den Knoten im Netz. Daher muss keine Entscheidung über das Routing getroffen werden und Routingmechanismen zur Steuerung des Datenflusses entfallen. Allerdings haben baumförmige Netzwerke das Problem, dass durch den Ausfall einer Leitung ein Teil der Knoten nicht mehr erreichbar sind. Baumförmige Kommunikationsnetzwerke werden oft bei einer überschaubaren Anzahl von Standorten wie z.B. in Local Area Networks (LAN) oder Unternehmensnetzwerken („Corporate Networks“) eingesetzt. Insbesondere Unternehmen, welche eigene Netzwerke betreiben (z.B. die DATEV eG), verwenden oft baumförmige Kommunikationsnetzwerkstrukturen wegen ihrer geringeren Komplexität.

Tabelle 2.2: Performance of GAs using different types of representations for deceptive tree problems of different sizes and with different T_{opt} (arbitrary tree, MST, and star)

T_{opt}		oder 3				
		P_{succ}	fitness		t_{conv}	
			μ	σ	μ	σ
arbitrary tree	Prüfer number	0.54	0.49	(0.6)	26.0	(8.6)
	NetKey	0.78	0.23	(0.4)	23.0	(6.0)
	LB ($P_1=1$)	0.09	1.69	(0.8)	19.5	(7.6)
	LB ($P_1=20$)	0.82	0.18	(0.4)	23.3	(6.1)
	$P_1=P_2=1$	0.12	1.24	(0.6)	27.4	(8.0)
	heur. xover	0	2.63	(0.5)	8.7	(2.4)
	h. ini & xover	0	3.88	(0.1)	0.4	(0.4)

Das Problem des kostenminimalen Aufbaus von vermaschten bzw. baumförmigen Kommunikationsnetzen wurde schon sehr frühzeitig in der Literatur formalisiert und beschrieben. Beim „network design problem“ (Johnson u. a. 1978) soll diejenige Netzwerkstruktur bestimmt werden, welche alle Standorte miteinander verbindet, alle Kommunikationsanforderungen zwischen den einzelnen Standorten erfüllt und gleichzeitig minimale Kosten besitzt. Aus dem allgemeinen „network design problem“ lässt sich das „optimal communication spanning tree“ (OCST)-Problem ableiten, bei welchem die Struktur des gewünschten Netzwerkes baumförmig ist. Im folgenden Abschnitt soll näher auf das OCST Problem eingegangen werden.

3 Summary, Conclusions, and Further Work

The purpose of this book is to understand the influence of representations on the performance of genetic and evolutionary algorithms. This chapter summarizes the work contained in this study and lists its major contributions.

3.1 Summary

We started in Chap. 1 by providing the necessary background for examining representations for GEAs. Researchers recognized early that representations have a large influence on the performance of GEAs. Consequently, after a brief introduction into representations and GEAs, we discussed how the influence of representations on problem difficulty can be measured. The chapter ended with prior guidelines for choosing high-quality representations. Most of them are mainly based on empirical observations and intuition and not on theoretical analysis.

Therefore, we presented in Chap. 2 three aspects of a theory of representations for GEAs. We investigated how the locality, scaling, and locality of an encoding influences GEA performance. The performance of GEAs is determined by the solution quality at the end of a run and the number of generations until the population is converged. Consequently, for redundant and exponentially scaled encodings, we presented population sizing models and described how the time to convergence is changed. Furthermore, we were able to demonstrate that high-locality encodings do not change the difficulty of a problem; in contrast, when using low-locality encodings, on average, the difficulty of problems changes. Therefore, easy problems become more difficult and difficult problems become easier by the use of low-locality encodings. For all three properties of encodings, the theoretical models were verified with empirical results.

3.2 Conclusions

We summarize the most important contributions of this work.

Framework for design and analysis of representations (and operators) for GEAs. The main purpose of this study was to present a framework which describes how genetic representations influence the performance of GEAs. The performance of GEAs is measured by the solution quality at the end of the run and the number of generations until the population is converged. The proposed framework allows us to analyze the influence of existing representations on GEA performance and to develop efficient new representations in a theory-guided way. Furthermore, we illustrated that the framework can also be used for the design

and analysis of search operators, which are relevant for direct encodings. Based on the framework, the development of high-quality representations remains not only a matter of intuition and random search but becomes an engineering design task. Even though more work is needed, we believe that the results presented are sufficiently compelling to recommend increased use of the framework.

Redundancy, Scaling, and Locality. These are the three elements of the proposed framework of representations. We demonstrated that these three properties of representations influence GEA performance and presented theoretical models to predict how solution quality and time to convergence changes. By examining the redundancy, scaling, and locality of an encoding, we are able to predict the influence of representations on GEA performance.

The theoretical analysis shows that the redundancy of an encoding influences the supply of building blocks (BB) in the initial population. r denotes the number of genotypic BBs that represent the best phenotypic BB, and k_r denotes the order of redundancy. For synonymously redundant encodings, where all genotypes that represent the same phenotype are similar to each other, the probability of GEA failure goes either with $O(\exp(-r/2^{k_r}))$ (uniformly scaled representations) or with $O(\exp(-\sqrt{r/2^{k_r}}))$ (exponentially scaled representations). Therefore, GEA performance increases if the representation overrepresents high-quality BBs. If a representation is uniformly redundant, that means each phenotype is represented by the same number of genotypes, GEA performance remains unchanged in comparison to non-redundant encodings.

The analysis of the scaling of an encoding reveals that non-uniformly scaled representations modify the dynamics of genetic search. If exponentially scaled representations are used, the alleles are solved serially which increases the overall time until convergence and results in problems with genetic drift but allows rough approximations of the expected optimal solution after a few generations.

We know from previous work that the high locality of an encoding is a necessary condition for efficient mutation-based search. An encoding has high locality if neighboring phenotypes correspond to neighboring genotypes. Investigating the influence of locality shows that high-locality encodings do not change the difficulty of a problem. In contrast, low-locality encodings, where phenotypic neighbors do not correspond to genotypic neighbors, change problem difficulty and make, on average, easy problems more difficult and deceptive problems easier. Therefore, to assure that an easy problem remains easy, high-locality representations are necessary.

3.3 Further Work

What are the open questions? What should be done next?

Literaturverzeichnis

Cayley 1889

CAYLEY, A.: A Theorem on Trees. In: *Quarterly Journal of Mathematics* 23 (1889), S. 376–378

Goldberg 1989

GOLDBERG, D. E.: *Genetic algorithms in search, optimization, and machine learning*. Reading, MA : Addison-Wesley, 1989

Johnson u. a. 1978

JOHNSON, D. S. ; LENSTRA, J. K. ; KAN, A. H. G. R.: The complexity of the network design problem. In: *Networks* 8 (1978), S. 279–285

Kotz u. a. 2000

KOTZ, S. ; BALAKRISHNAN, N. ; JOHNSON, N. L.: *Wiley Series in Probability and Statistics*. Bd. 2: *Continuous Multivariate Distributions*. John Wiley and Sons, 2000

Mertens u. a. 2001

MERTENS, Peter ; BODENDORF, Freimut ; KÖNIG, Wolfgang ; PICOT, Arnold ; SCHUMANN, Matthias: *Grundzüge der Wirtschaftsinformatik*. 7. Berlin u. a. : Springer, 2001

Scheer 1997

SCHEER, A.-W.: *Wirtschaftsinformatik: Referenzmodelle für industrielle Geschäftsprozesse*. 7. Berlin et al. : Springer-Verlag, 1997

Wolpert u. Macready 1995

WOLPERT, D. H. ; MACREADY, W. G.: No free lunch theorems for search / Santa Fe Institute. Santa Fe, NM, 1995 (SFI-TR-95-02-010). – Tech. Rep. No.

Eidesstattliche Erklärung

Ich versichere, dass ich meine Diplomarbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen oder anderen Quellen entnommen sind, sind als solche eindeutig kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form noch nicht veröffentlicht und noch keiner Prüfungsbehörde vorgelegt worden.

Ort, den Datum

Martin Mustermann