



Wintersemester 2006/07

Fundamente der Computational Intelligence
(Vorlesung)

Prof. Dr. Günter Rudolph

Fachbereich Informatik

Lehrstuhl für Algorithm Engineering





Nachtrag zur Genotyp-Phänotyp-Abbildung $\mathbb{B}^n \rightarrow [L, R] \subset \mathbb{R}$

- Standardkodierung für $b \in \mathbb{B}^n$

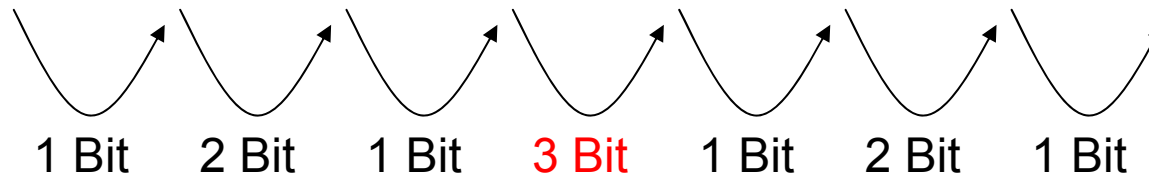
$$x = L + \frac{R - L}{2^n - 1} \sum_{i=0}^{n-1} b_{n-i} 2^i$$

→ Problem: Hammingklippen (*hamming cliffs*)

000	001	010	011	100	101	110	111
0	1	2	3	4	5	6	7

$L = 0, R = 7$

$n = 3$



↑
Hammingklippe



Nachtrag zur Genotyp-Phänotyp-Abbildung $\mathbb{B}^n \rightarrow [L, R] \subset \mathbb{R}$

- Graykodierung für $b \in \mathbb{B}^n$

Sei $a \in \mathbb{B}^n$ standardkodiert. Dann $b_i = \begin{cases} a_i, & \text{falls } i = 1 \\ a_{i-1} \oplus a_i, & \text{falls } i > 1 \end{cases}$

$\oplus = \text{XOR}$

000	001	011	010	110	111	101	100	← Genotyp
0	1	2	3	4	5	6	7	← Phänotyp

OK, keine Hammingklippen mehr ...

⇒ kleine Änderungen im Phänotyp „bewirken“ kleine Änderungen im Genotyp

Wir wollen jedoch wegen Evolution a la Darwin (nicht Lamarck):

⇒ kleine Änderungen im Genotyp bewirken kleine Änderungen im Phänotyp

aber: 1-Bit-Änderung: $000 \rightarrow 100 \Rightarrow \text{☹}$



Evolutionäre Algorithmen haben viele **Freiheitsgrade**:

- **Mutation**: Mutationsregel, Mutationsw'keit, Mutationsverteilung?
- **Rekombination**: 1-Punkt-Crossover, Uniform Crossover, Intermediär?
- **Selektion**: Turnierselektion ($q=?$), $(\mu + \lambda)$ -Selektion, ... ?
- **Populationsgröße**
- **Kodierung**
- ...

Welche EA-**Parametrisierung** ist besser für mein Problem?

Ergebnis eines EA-Laufs ist zufällig ...

... und Vergleich zufälliger Größen ist mathematisch sinnlos!

⇒ also mehrere Läufe und Anwendung statistischer Verfahren!



Nur mühselig vorankommend:

Vernünftiger / sinnvoller Einsatz statistischer Methoden beim Testen

Typisches Beispiel

Ausgangslage:

- Optimierung unter einfacher Zielsetzung (\rightarrow min!)
- 2 randomisierte Suchverfahren: A und B
- Welches ist besser?
- Qualitätsmaß: skalarer Zielfunktionswert
- Wie testen?



Typische Vorgehensweise:

$A(t)$ sei ZF-Wert von A nach t Iterationen, $B(t)$ analog

Entscheidungsregel:

$$\text{„A besser als B“} \Leftrightarrow E[A(t)] < E[B(t)]$$

Also: fixiere max. Laufzeit t und Versuchszahl N

$$\text{Wenn} \quad \frac{1}{N} \sum_{i=1}^N A_i(t) < \frac{1}{N} \sum_{i=1}^N B_i(t) \quad ,$$

dann A besser als B.



Getestetes Szenario:

repeat

Algorithmus 1x laufen lassen

verwende beste gefundene Lösung $A(t)$

until Prozessunterbrechung

Realistisches Szenario:

repeat

Algorithmus M mal laufen lassen

verwende beste gefundene Lösung $A^*(t)$ der M Läufe

until Prozessunterbrechung



Passender Test:

$$A^*(t) = \min \{ A_1(t), A_2(t), \dots, A_M(t) \}, B^*(t) \text{ analog}$$

Entscheidungsregel:

$$\text{„A besser als B“} \Leftrightarrow E[A^*(t)] < E[B^*(t)]$$

Also: fixiere max. Laufzeit t , Wiederholungen M und Versuchszahl N

$$\text{Wenn} \quad \frac{1}{N} \sum_{i=1}^N A_i^*(t) < \frac{1}{N} \sum_{i=1}^N B_i^*(t) \quad ,$$

dann A besser als B.



⇒ kann zu **gegensätzlichen Empfehlungen** führen!

Beispiel: (Gleichverteilung)

$X \sim U(a, b)$	$E[X] = \frac{a + b}{2}$	$E[X^*] = a + \frac{b - a}{M + 1}$
$A \sim U(10, 14)$	$E[A] = 12$	$E[A^*] = 10 + \frac{4}{M + 1}$
$B \sim U(8, 18)$	$E[B] = 13$	$E[B^*] = 8 + \frac{10}{M + 1}$

⇒ $E[A] < E[B]$ **aber** $E[A^*] > E[B^*]$ sobald $M \geq 3$



Parametrischer Test:

doppelter t-Test (Vergleich von Mittelwerten)

Voraussetzungen für Anwendung:

1. Grundgesamtheiten normalverteilt
2. Gleiche Varianz

normalverteilt? → bestenfalls approximativ

gleiche Varianz? → erst Testen durch F-Test!

Falls F-Test zur Ablehnung der Gleichheitshypothese führt,
dann können wir den t-Test nicht anwenden!

Falls wir es trotzdem tun, dann sind die Aussagen **statistisch wertlos!**



Was dann?

Aus Theorie der Extremwertverteilungen:

Für $V_i \geq L > -\infty$ mit $V_{1:k} = \min \{ V_1, V_2, \dots, V_k \}$:

$F_{V_{1:k}}(x) \rightarrow F_W(x)$ für große k (Konvergenz in Verteilung)

d.h. $V_{1:k}$ ist für große k approximativ Weibull-verteilt

also:

A^* und B^* approximativ Weibull-verteilt, aber mit verschiedenen Parametern

nicht-parametrischer Test:

robuster \nearrow Hypothese $F_{A^*}(x) = F_{B^*}(x)$ vs. Hypothese $F_{A^*}(x) = F_{B^*}(x - \theta)$ \uparrow Lageparameter



Mann/Whitney- bzw. Wilcoxon- bzw. U-Test:

nichtparametrischer Test für 2 unabhängige Stichproben

Hypothesen $H_0 : F_X = F_Y$ vs. $H_1 : F_X(x) = F_Y(x - \theta)$

bzgl. Lagealternative $\theta < 0$

Durchführung: ($n \geq 10$)

1. x_1, \dots, x_n und y_1, \dots, y_n sortieren
2. $u = \sum_i \text{Rang}(x_i) - n(n+1)/2$
3. $t = (u - n^2/2) / (n^2(2n-1)/12)^{1/2}$
4. $t \in K_\alpha ? \Rightarrow H_0$ ablehnen!

K_α ist $(1-2\alpha)$ -Quantil der $N(0,1)$ -Vert.



Alternativen:

- **van-der-Werden** bzw. **X-Test**:

mehr Rechenaufwand als U-Test,
aber höhere asymptotische Wirksamkeit

- **Siegel-Tukey-Test**: (Variabilitätsalternative)

$H_0 : F_X = F_Y$ vs. $H_1 : F_X(x) = F_Y(\theta x)$ für $0 < \theta \neq 1$

- **Mood-Test**: (Variabilitätsalternative)

asymptotisch wirksamer als Siegel-Tukey-Test

- > 2 unabh. Stichproben ! **Kruskal-Wallis-Test** (verallg. U-Test)



Ausgangslage

Kein Vorwissen (*a priori* Wissen) über Problem

Initiale Version 1

1. Wahl einer Repräsentation / Kodierung
2. Konstruktion einer Mutationsverteilung
3. (1+1)-EA anwenden!

ad 1.) Wähle die „natürlichste“ Repräsentation (nicht alles auf Bitstrings abbilden)

ad 2.) Verwende Verteilung maximaler Entropie

→ Verteilung mit geringstem eingebauten Vorwissen

ad 3.) Kann schon gut funktionieren; Rechner rechnen lassen

Während Rechner rechnet \Rightarrow **Nachdenken!** Könnte Rekombination oder andere Kodierung besser sein?



Version 2

Versuche mal $(1+\lambda)$ -EA und $(1, \lambda)$ -EA

noch immer keine
Rekombination nötig!

Version 3

Versuche mal $(\mu+\lambda)$ -EA und (μ, λ) -EA
sowie andere Selektionsformen

noch immer keine
Rekombination nötig!

Version 4

Rekombination einbauen

welche? stark abhängig von Repräsentation!

→ **vergleichen mit statistischen Tests!**



Verteilungen maximaler Entropie

Sei $(p_k)_{k \in I}$ die W'keitsfunktion einer diskreten Verteilung.

Dann ist $H(p) = -\sum_{k \in I} p_k \log(p_k)$ die **Entropie** der Verteilung.

Die Verteilung, für die $H(p)$ unter gleichen Nebenbedingungen maximal wird, heißt **Verteilung maximaler Entropie**. ■

Interpretation von Verteilungen **maximaler Entropie**:

- Verteilung maximaler Unvoreingenommenheit
- Verteilung mit geringstem einbauten „Vorwissen“
- Verteilung geringster Information



Beispiele:

Träger $\{ 1, 2, \dots, n \}$

\Rightarrow diskrete Gleichverteilung

zusätzl. $E[X] = \theta$

\Rightarrow Boltzmann-Verteilung

zusätzl. $V[X] = \eta^2$

\Rightarrow N.N. (berechenbar; **nicht** Binomial-Verteilung)

Träger \mathbb{N}

\Rightarrow nicht def.

zusätzl. $E[X] = \theta$

\Rightarrow geometrische Verteilung

zusätzl. $V[X] = \eta^2$

\Rightarrow ?

Träger \mathbb{Z}

\Rightarrow nicht def.

mit $E[|X|] = \theta$

\Rightarrow bi-geometrische Verteilung

mit $E[|X|^2] = \eta^2$

\Rightarrow N.N. (berechenbar)



Träger $[a,b] \subset \mathbb{R}$ \Rightarrow Gleichverteilung

Träger \mathbb{R}^+ mit $E[X] = \theta$ \Rightarrow Exponentialverteilung

Träger \mathbb{R}
mit $E[X] = \theta$, $V[X] = \eta^2$ \Rightarrow Normalverteilung $N(\theta, \eta^2)$

für Permutationsverteilungen ?

Erst wenn man etwas über das Problem weiß oder gelernt hat
 \Rightarrow einbauen in Verteilung



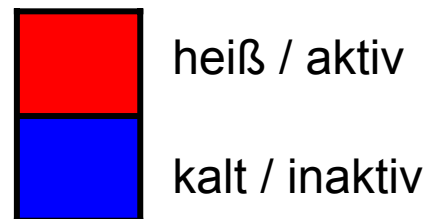
- Wissensintegration im Bereich EC wird immer wichtiger
- **Garantie:**
Je mehr Wissen über das zu optimierende Problem in irgendeiner Komponente eines EA kodiert werden kann, desto besser sind die erzielbaren Resultate!
- **Caveat:**
Suchraum / Alternativenmenge wird reduziert:
Kann zum Ausschluss guter Lösungen führen!
- Integration auf verschiedenen Ebenen:
 - Modellierung von Benutzerpräferenzen
 - Fitness-Funktionen
 - Mensch-Maschine-Interaktion
 - Genetische Operatoren:
 - Initialisierung
 - Rekombination
 - Mutation
 - Selektion



Frühes Beispiel: Brennstabwechselproblem (1994f.):
ICD, HUB, Siemens, KWU

A565	A321	0232	B121	D099	A111	A121
A226	0321	C139	A982	C321	C021	A222
A987	A553	B111	B112	A002	A144	0128
A009	B454	0287	A801	B071	B522	
C343	D762	C424	0999	B991		
0292	D393	A632	0020			
A233	B987	C112				

bedingt durch Interface:
Viertelsymmetrie



Problemwissen: Abbrand homogener je symmetrischer die Stabaktivität
⇒ Variationsoperatoren verändern, um Achtelsymmetrie zu erzwingen!