INF FAKULTÄT FÜR INFORMATIK

Evolutionary Algorithms

The Schema Theorem, No Free Lunch Theorem

Prof. Dr. Rudolf Kruse **Pascal Held**

{kruse,pheld}@iws.cs.uni-magdeburg.de

Otto-von-Guericke University Magdeburg

School of Computer Science

Department of Knowledge and Language Processing

Outline

Why Do Evolutionary Algorithms Work?

Approach from [Holland, 1975]:

- consider schemata (that is, partly specified binary chromosomes)
- investigate how the number of chromosomes matching a schema evolve over several generations

Objective: rough stochastic statement that describes how a genetic algorithm explores the search space
To keep things **simple**: Confinement to

- bit strings (chromosomes of zero and one) of a fixed length L
- fitness-proportionate selection (Roulette-wheel selection)
- bit-mutation (using the mutation probability p_m)
- one point-crossover

Algorithm 1 Genetic Algorithm

Input: target function F

1: $t \leftarrow 0$
2: $P(t) \leftarrow$ create population with μ individuals /* μ must be even */
3: evaluate $P(t)$ with F
4: **while** termination criterion is not fulfilled {
5: $P'(t) \leftarrow$ select μ individuals $A^{(1)}, \ldots, A^{(\mu)}$ from $P(t)$ with *roulette wheel selection*
6: $P'' \leftarrow \emptyset$
7: **for** $i \leftarrow 1, \ldots, \frac{\mu}{2}$ {
8: $u \leftarrow$ choose random number from $U([0, 1))$
9: **if** $u \leq p_x$ { /* recombin. prob. p_x */
10: $B, C \leftarrow$ *one-point crossover*$(A^{(2i-1)}, A^{(2i)})$
11: } **else** {
12: $B \leftarrow A^{(2i-1)}$
13: $C \leftarrow A^{(2i)}$
14: }
15: $B \leftarrow$ *Bit-Mutation*(B)
16: $C \leftarrow$ *Bit-Mutation*(C)
17: $P'' \leftarrow P'' \cup \{B, C\}$
18: }
19: evaluate P'' with F
20: $t \leftarrow t + 1$
21: $P(t) \leftarrow P''$
22: }
23: **return** best individual from $P(t)$

Schemata

Definition (Schema)

A **schema** h is a character string of length L over the alphabet $\{0, 1, *\}$, that is $h \in \{0, 1, *\}^L$.
The character $*$ is called **wildcard character** or **Don't-Care-Symbol**.

Definition (Matching)

A chromosome $c \in \{0, 1\}^L$ **matches a schema** $h \in \{0, 1, *\}^L$, written as: $c \triangleleft h$, if and only if it coincides with h at all positions where h is 0 or 1.
Positions at which h is $*$ are not taken into account

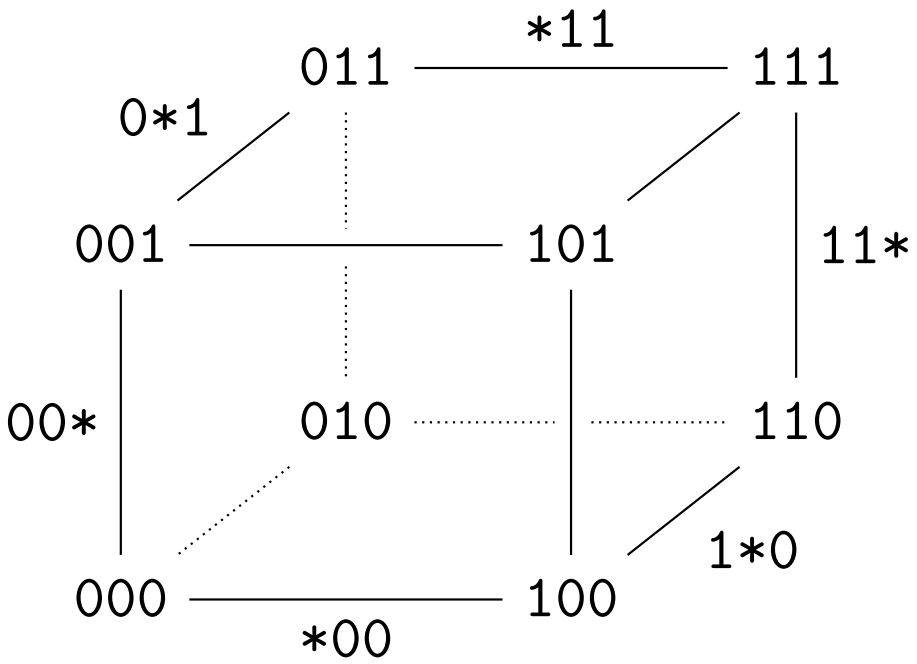
Schemata: an Illustration

$h =$	$**0*11*10*$	schema of length 10
$c_1 =$	1100111100	matches schema h, that is $c_1 \triangleleft h$
$c_2 =$	1111111111	does not match h, that is $c_2 \not\triangleleft h$

- there are 2^L possible chromosomes and 3^L schemata
- every chromosome matches $\sum_{i=0}^{L} \binom{L}{i} = 2^L$ schemata
- population of size μ can match close to $\mu 2^L$ different schemata (usually a lot smaller due to similar chromosomes)
- observation of a chromosome $\widehat{=}$ Observation of many schemata at the same time
- **implicit parallelism**

Schemata: Hyperplanes

Geometrically, a schema can be seen as describing a hyperplane in a unit hypercube (but only hyperplanes that are parallel or orthogonal to the sides of the hypercube).



Examples:

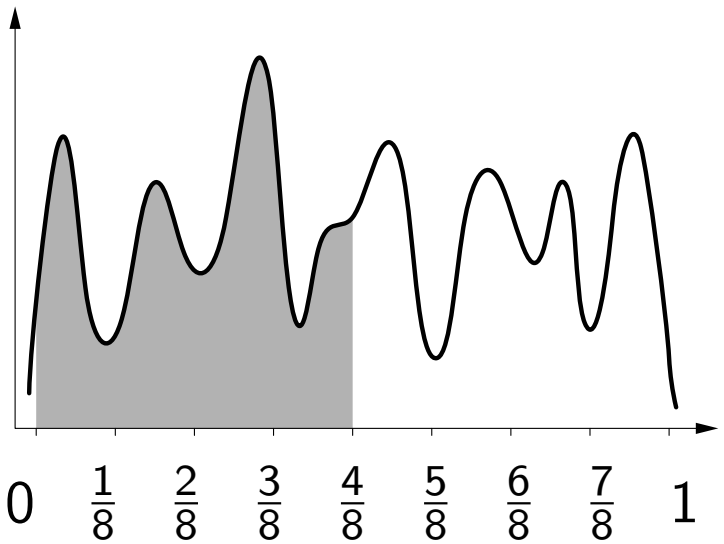
$*00 \quad \widehat{=} \quad$ edge connecting the corners 000 and 100 (bottom front)

Schemata: Domain of the Fitness Function

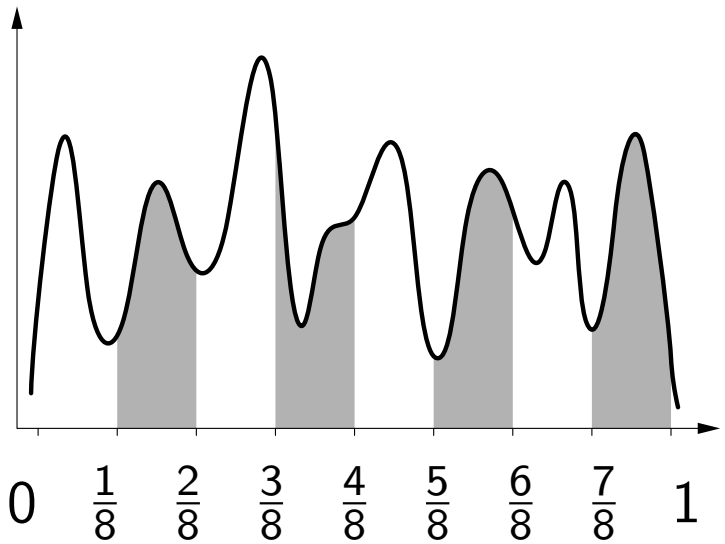
given: real function $f : x \in [0, 1] \rightarrow \mathbb{R}$

suppose: binary enconding of x (no Gray code)

schema $\widehat{=}$ „strip pattern“ (periodical Fct.) in $\text{dom}(f) = [0, 1]$



schema 0**...*



schema **1*...*

schemata with Gray-Encoding: see also exercise sheet

Influence of Selection

- Examination how **selection** and applying **genetic operators** (Mutation und Crossover) influence the chromosomes
- Tracing the breeding of chromosomes which match the schema
- Effect of selection: what fitness the chromosomes have that match a schema h?
  Approach: defining the mean relative fitness of chromosomes

Definition (Mean Relative Fitness)

The **mean relative fitness** of chromosomes that match schema h in the population $P(t)$ is

$$f_{\text{rel}}(h) = \frac{\sum_{A \in P(t), A.G \triangleleft h} A.F_{\text{rel}}}{|\{A \in P(t) \mid A.G \triangleleft h\}|}.$$

Influence of Selection

The average number of chromosomes in the next generation of a Schema h matching chromosome is

$$f_{\text{rel}}(h) \cdot |P|$$

The expected number of chromosomes that match schema h after selection, is

$$(\text{Zahl vorher passender Chromosomen}) \cdot f_{\text{rel}}(h) \cdot |P|$$

Influence of Selection

Further observations of the relative fitness of a schema:

$$
\begin{aligned}
f_{\text{rel}}(h) \cdot |P| &= \frac{\sum_{A \in P(t), A.G \lhd h} A.F_{\text{rel}}(c)}{|\{A \in P(t) \mid A.G \lhd h\}|} \cdot |P| \\
&= \frac{\sum_{A \in P(t), A.G \lhd h} \frac{A.F}{\sum_{B \in P(t)} B.F}}{|\{A \in P(t) \mid A.G \lhd h\}|} \cdot |P| \\
&= \frac{\frac{\sum_{A \in P(t), A.G \lhd h} A.F}{|\{A \in P(t) \mid A.G \lhd h\}|}}{\frac{\sum_{B \in P(t)} B.F}{|P|}} = \frac{\overline{f_t(h)}}{\overline{f_t}}
\end{aligned}
$$

$\overline{f_t(h)}$ average fitness of the chromosomes matching h in $P(t)$

$\overline{f_t}$ average fitness of all chromosomes of the t-th generation

The average number of offsprings can be written by the ratio of the average fitness of a schema and the total average fitness.

Influence of Mutation

We need measures with which we can compute probabilities that the match to a schema is preserved

Definition (Order (for 1-Bit- and Bit-Mutation))

The **order** of a schema h is the number of zeroes and ones in h, that is $\text{ord}(h) = \#0 + \#1 = L - \#*$ (#: number of occurences of).

For instance: $\text{ord}(**0*11*10*) = 5$

Influence of Mutation

Match to schema h is preserved...

- with Bit-Mutation using prob. $(1 - p_m)^{\text{ord}(h)}$
- with 1-Bit-Mutation with
  Prob. $1 - \frac{\text{ord}(h)}{L}$, if bit is inverted,
  Prob. $1 - \frac{\text{ord}(h)}{2L}$, if new bit is determined by random

Explaination:

- Bit-Mutation inverts a bit with prob. p_m and with prob. $(1 - p_m)$ otherwise
- 1-Bit-Mutation chooses one of the L genes of a chromosome of length L with same probability

Influence of Crossover

Definition (Defining Length (for one-point crossover))

The **defining length** of a schema h is the difference between the position of the last 0/1 and the first 0/1 in h.

Example: $\mathrm{dl}(**0*11*10*) = 9 - 3 = 6$

Influence of Crossover

- One-point crossover: probability that the cut point splits a chromosome in such a way that some of the fixed characters of a schema lie on one side of the cut and some on the other is $\frac{\mathrm{dl}(h)}{L-1}$

Explaination:

- One-point crossover: $L-1$ possible cut points on chromosomes of length L (all equally likely)
- with $\mathrm{dl}(h)$ of these cut points, genes specified by the schema are exchanged between individuals
- matching might (or might not) get lost

Outline

Definitions

Definition (expected value of matching chromosomes)

$N(h, t)$ is the expected value of the number of chromosomes that match the schema h during the t-th generation.

Definition (expected value after selection)

$N(h, t + \Delta t_s)$ is the expected value of the number of chromosomes that match the schema h during the t-th generation after selection.

Definitions

Definition (expected value after crossover)

$N(h, t + \Delta t_s + \Delta t_x)$ is the expected value of the number of chromosomes that match the schema h during the t-th generation after selection and crossover.

Definition (expected value after mutation)

$N(h, t + \Delta t_s + \Delta t_x + \Delta t_m) = N(h, t + 1)$ is the expected value of the number of chromosomes that match the schema h during the t-th generation after selection, crossover and mutation (and thus during the $t + 1$-th generation).

- **searched:** relation between $N(h, t)$ and $N(h, t + 1)$

Selection

approach: analyze the impact of selection, crossover and mutation step by step, based on average fitness, order and defining length of a schema.

- **impact of selection:** can be described by average fitness

$$N(h, t + \Delta t_s) = N(h, t) \cdot f_{\text{rel}}(h) \cdot |P|$$

$N(h, t) \cdot f_{\text{rel}}(h)$	probability that chosen chromosome matches schema h
$f_{\text{rel}}(h) \cdot \|P\|$	average number of offsprings for one chromosome matching schema h

- *note:* relative fitness $f_{\text{rel}}(h)$ cannot be determined exactly, as the number chromosomes matching schema h is only an approximation

Crossover

impact of Crossover: described by

$$N(h, t + \Delta t_s + \Delta t_x) = \underbrace{(1 - p_x) \cdot N(h, t + \Delta t_s)}_{E_A} + \underbrace{p_x \cdot N(h, t + \Delta t_s) \cdot (1 - p_{\text{loss}})}_{E_B} + C$$

p_x	probability of crossover
p_{loss}	probability of a chromosome matching schema h losing its matching during 1-point-crossover
E_A	expected value of the number of chromosomes matching schema h and *not* taking part in crossover
E_B	expected value of the number of chromosomes taking part in crossover without losing its matching to schema h
C	gained number of chromosomes matching schema h, won by... (see exercise)

Considering the Probability p_{loss}

examples:

$$
\begin{aligned}
h &= \texttt{**0*}\,|\,\texttt{1*1*} && && \texttt{**0*1*1*} && = h \\
h \triangleright c_1 &= \texttt{0000}\,|\,\texttt{1111} && \rightarrow && \texttt{00000000} && = c_1' \not\lhd h \\
h \not\triangleright c_2 &= \texttt{1111}\,|\,\texttt{0000} && \rightarrow && \texttt{11111111} && = c_2' \not\lhd h
\end{aligned}
$$

$$
\begin{aligned}
h &= \texttt{**0*}\,|\,\texttt{1*1*} && && \texttt{**0*1*1*} && = h \\
h \triangleright c_1 &= \texttt{0000}\,|\,\texttt{1111} && \rightarrow && \texttt{00001010} && = c_1' \lhd h \\
h \triangleright c_2 &= \texttt{1101}\,|\,\texttt{1010} && \rightarrow && \texttt{11011111} && = c_2' \lhd h
\end{aligned}
$$

thus:

$$
p_{\text{loss}} \leq \underbrace{\frac{\text{dl}(h)}{L-1}}_{Pr_A = p_x^*(h)} \cdot \left(1 - \underbrace{\frac{N(h, t+\Delta t_s)}{|P|}}_{Pr_B} \right)
$$

Pr_A probability of cut position between fixed genes

Pr_B probability of 2nd chromosome matching schema h

quesion: why does only $\leq$ hold, and $=$ not? (see exercise)

Crossover

substitution of the expression for p_{loss} yields:

$$
\begin{aligned}
& N(h, t + \Delta t_s + \Delta t_x) \\
& \geq (1 - p_x) \cdot N(h, t + \Delta t_s) \\
& \quad + \quad p_x \cdot N(h, t + \Delta t_s) \cdot \left(1 - \frac{\mathrm{dl}(h)}{L-1} \cdot \left(1 - \frac{N(h, t + \Delta t_s)}{|P|}\right)\right) \\
& = N(h, t + \Delta t_s) \left(1 - p_x + p_x \cdot \left(1 - \frac{\mathrm{dl}(h)}{L-1} \cdot \left(1 - \frac{N(h, t + \Delta t_s)}{|P|}\right)\right)\right) \\
& = N(h, t + \Delta t_s) \cdot \left(1 - p_x \frac{\mathrm{dl}(h)}{L-1} \cdot \left(1 - \frac{N(h, t + \Delta t_s)}{|P|}\right)\right) \\
& \overset{(*)}{=} N(h, t) \cdot f_{\text{rel}}(h) \cdot |P| \cdot \left(1 - p_x \frac{\mathrm{dl}(h)}{L-1} \cdot (1 - N(h, t) \cdot f_{\text{rel}}(h))\right)
\end{aligned}
$$

step $(*)$: twice use of the previously derived relation
$N(h, t + \Delta t_s) = N(h, t) \cdot f_{\text{rel}}(h) \cdot |P|$

Mutation

impact of binary mutation: described by order

$$\begin{aligned} N(h, t+1) &= N(h, t + \Delta t_s + \Delta t_x + \Delta t_m) \\ &= N(h, t + \Delta t_s + \Delta t_x) \cdot (1 - p_m)^{\text{ord}(h)} \end{aligned}$$

explaination: in order to not lose matching, none of the ord(h) genes fixed in schema h must be altered

alternative models possible, e.g.:
exactly one bit altered per chromosome $\Rightarrow$ 1 bit mutation

$$\begin{aligned} N(h, t+1) &= N(h, t + \Delta t_s + \Delta t_x + \Delta t_m) \\ &= N(h, t + \Delta t_s + \Delta t_x) \cdot \left(1 - \frac{\text{ord}(h)}{L}\right) \end{aligned}$$

The Schema Theorem

including binary mutation, following holds:

$$N(h, t+1) \geq f_{\text{rel}}(h) \cdot |P| \cdot \left(1 - p_x \frac{\text{dl}(h)}{L-1} \cdot (1 - N(h,t) \cdot f_{\text{rel}}(h))\right) \cdot (1-p_m)^{\text{ord}(h)} \cdot N(h,t)$$

substitution of the fitness relations yields

$$N(h, t+1) \geq \frac{\overline{f_t(h)}}{\overline{f_t}} \left(1 - p_x \frac{\text{dl}(h)}{L-1} \left(1 - \frac{N(h,t)}{|P|} \cdot \frac{\overline{f_t(h)}}{\overline{f_t}}\right)\right) \cdot (1-p_m)^{\text{ord}(h)} \cdot N(h,t)$$

interpretation: schemata with

- a score above average,
- short defining length and
- low order

do breed very heavily (approx. exponential)

Outline

Building Block Hypothesis

- a GA explores Ω particularly well in hyperplanes (schemata with good average fitness, low defining length and low order)
- chromosomes breed particularly well in these regions
- these schemata are called **building blocks**,
  thus the name **Building Block Hypothesis**
- **note:** this form of this hypothesis only holds for bit sequences, fitness proportional selection, binary mutation and 1 point crossover
- with other genetic operations the blocks can possibly be described by other characteristics
- a high average finess is always a good characteristic, as every selection method prefers chromosomes with high fitness

Building Block Hypothesis

GA operates best, if **short schemes with low order** (so-called blocks) are used to create schemata with higher Fitness.

- **but:** combining blocks leads to schemata of higher order and higher defining length of blocks
- thus new building blocks are more likely to be destroyed by an operation
- operation methods of GAs still aren't studied and explored sufficiently.

Outline

Summary

Schema Theorem only holds for **one schema** independently of other schemata within the population.

- other schemata will breed, too
- given time: convergence of the population, thus decreasing evolutionary pressure
- relative fitness of a schema converges towards $1/|P|$ (same fitness for all individuals)
- finally: expected number of copies decreases, because of genetic operations

Summary

- implicit assumption: only little interdependencies between genes (low *epistasis*), thus fitness of chromosomes that are matching the same scheme is quite similar
- implicit assumption: on chromosome, interacting genes are located close to each other for small blocks
  - this argument only concerns the restriction to 1 point crossover (not the approach itself)
  - other operation specific measures instead of defining length possible

Outline

No Free Lunch Theorem

Preconditions

search space Ω

$\mathcal{F}$ space of all optimization problems (objective function)

Ω and $\mathcal{F}$ are alike

lack of knowledge regarding optimization problem

- uniform distribution among all these problems
- every problem $F \in \mathcal{F}$ occurs with a probability of $\frac{1}{|\mathcal{F}|}$

- further simplifications
  - $\forall F \in \mathcal{F} \quad F : \Omega \mapsto \mathbb{R}$ holds
  - $\forall F \in \mathcal{F}$ are defined on the same search space Ω
- $\mathcal{L}$et A be the set of all optimization algorithms operating on Ω

No Free Lunch Theorem

Characterization of an algorithm

- which individuals are considered in which order on $F \in \mathcal{F}$
- only n evaluations possible for optimization
  - $\text{optimization}_{F,n} : \mathcal{A} \mapsto \Omega^n$
- with every optimization the algorithm assesses an individual 1x

$\Rightarrow$ $\text{optimization}_{F,n}(\text{Alg})$ contains n different individuals in total

- let every algorithm Alg be deterministically

$\Rightarrow$ $\text{optimization}_{F,n}(\text{Alg})$ unique, too

For a problem $F \in \mathcal{F}$, an optimization problem $\text{Alg} \in \mathcal{A}$ and $n \in \mathbb{N}$:

$$\text{optimization}_{F,n}(\text{Alg}) = (y_1, \ldots, y_n) \in \Omega^n$$

with $y_i \neq y_j$ for $i \neq j$ and y_k being the individual, which analyzes Alg with F as k-th element.

comparison of $\text{Alg}_1,\ \text{Alg}_2 \in \mathcal{A}$

- via performance schemata QuAlg (quality of an algorithm)
  - defined with $\text{optimization}_{F,n}(\text{Alg}) = (y_1, \ldots, y_n)$ by $q_n : \mathbb{R}^n \mapsto \mathbb{R}$ as
    $$\text{QuAlg}_{F,n}(\text{Alg}) = q_n\left(F(y_1), \ldots, F(y_n)\right)$$
  - e.g. average / best goodness or
  - number of needed evaluations until optimum is reached
- expected performance E of the first n evaluations of Alg on an arbitrary problem

$$E\left[\text{QuAlg}_{F,n}(\text{Alg}) \mid F \in \mathcal{F}\right] = \frac{1}{\#\mathcal{F}} \sum_{F \in \mathcal{F}} \text{QuAlg}_{F,n}(\text{Alg})$$

$\Rightarrow$ average on all possible problems

No Free Lunch

Theorem (No free lunch)

For two algorithms $Alg_1, Alg_2 \in \mathcal{A}$ and the class of all problems $\mathcal{F}$ the following applies to a performance schema QuAlg:

$$E\left[QuAlg_{F,n}(Alg_1) \mid F \in \mathcal{F}\right] = E\left[QuAlg_{F,n}(Alg_2) \mid F \in \mathcal{F}\right]$$

Consequences

no algorithm is superior to all others on average

if there was an algorithm superior on $\mathcal{F}' \subset \mathcal{F}$, thus

$$E\left[\mathrm{QuAlg}_{F,n}(\mathrm{Alg}_1) \mid F \in \mathcal{F}'\right] < E\left[\mathrm{QuAlg}_{F,n}(\mathrm{Alg}_2) \mid F \in \mathcal{F}'\right]$$

- then it immediately follows that

$$E\left[\mathrm{QuAlg}_{F,n}(\mathrm{Alg}_1) \mid F \in \mathcal{F} \setminus \mathcal{F}'\right] > E\left[\mathrm{QuAlg}_{F,n}(\mathrm{Alg}_2) \mid F \in \mathcal{F} \setminus \mathcal{F}'\right]$$

  - for every algorithm: $\exists$ niche within the space of all problems which it is particularly good for
  - application: which algorithm to use for which problem?
  - research: which class of problems is best for a particular algorithm?

Summary

if there is no previous knowledge regarding the problem

$\Rightarrow$ expected results of an EA not better than those of any other method

if there is knowledge regarding the problem

- e.g. assumptions about the goodness space

$\Rightarrow$ general application of certain algorithms is suggested

knowledge regarding structure of the problem should influence the choice or design of an optimization algorithm

Literature Regarding This Lecture I

Holland, J. H. (1975).
Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence.
University of Michigan Press.