

GENETICKÉ ALGORITMY (GA)

- * Holland 60. roky 20. stor.
- **Populácia** umelých chromozómov sa cyklicky podrobuje selektívnej reprodukcií preferujúcej výkonnejších jedincov a náhodným zmenám
- **Umelý chromozóm** (genotyp) = reťazec symbolov kódujúci vlastnosti jedinca (fenotyp)
 - napr. binárna hodnota premennej alebo postupnosť premenných
 - mnoho typov kódovaní
 - binárne, Greyov kód, reálne hodnoty
 - abecedy - napr. binárna, ternárna, ...

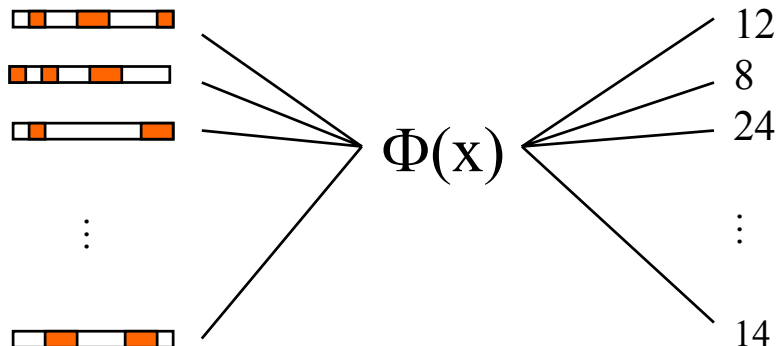
GA - fitness funkcia

- **Ohodnotenie [Fitness function]** - kritérium výkonnosti, zobrazenie: genotyp \rightarrow reálne(celé) číslo
 - čím vyššia hodnota, tým je jedinec lepší

$$\Phi: \text{genotyp} \rightarrow R$$

populácia

fitness



Genetické algoritmy

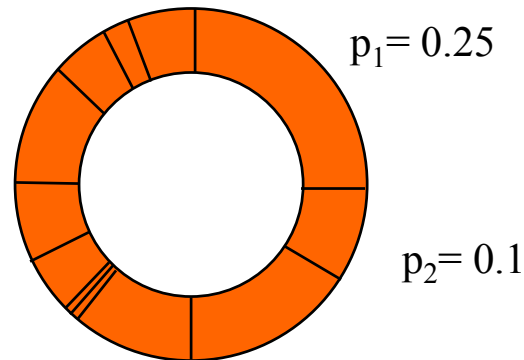
- **Jednoduchý genetický algoritmus** (Goldberg 1989)
 - vytvor populáciu N náhodne vygenerovaných chromozómov
 x_1, x_2, \dots, x_N
 - opakuj
 - dekóduj všetky chromozómy a spočítaj ich ohodnotenie $f_i = \Phi(x_i)$
 - vytvor novú populáciu selektívnou reprodukciou
 - rekombinuj chromozómy - kríženie
 - mutuj chromozómy
 - dokiaľ sa neobjaví hľadaný jedinec, alebo ohodnotenie najlepšieho nerastie

Jednoduchá selekcia

- Zo starej populácie vytvárame novú kopírovaním chromozómov tak, že čím lepší jedinec, tým viacej jeho kópií sa môže objaviť v novej populácii
- reprodukcia **ruletou**:
 - každý jedinec dostane na ruletovom kole priehradku veľkosti p_i úmernej jeho ohodnoteniu - to bude pravdepodobnosť, že bude reprodukovaný

$$p_i = \frac{f_i}{\sum_{j=1}^n f_j}$$

- ruletou sa otočí N krát



Varianty selekcie 1

■ Problémy jednoduchej selekcie

a) keď všetci jedinci majú podobné ohodnotenie
náhodné prehľadávanie s genetickým posunom

b) keď jeden až dvaja jedinci majú ohodnotenie o mnoho vyššie než zvyšok populácie

takmer všetci jedinci v novej generácii budú kópiou toho istého jedinca,
predčasná konvergencia

■ Riešenie:

1. **škálovanie [scaling]** (typicky lineárna transformácia)

pre a) zväčšiť rozdiely, pre b) zmenšiť

2. **selekcia podľa poradia [rank based]** - jedinci sa zoradia podľa ohodnotenia a pravdepodobnosť reprodukcie je úmerná poradiu jedinca, nie jeho ohodnoteniu

Varianty selekcie 2

3. **S orezávaním [truncation]** - jedincov usporiadame podľa veľkosti ohodnotenia, M najlepších jedincov okopírujeme O krát tak, že

$$N = M \times O$$

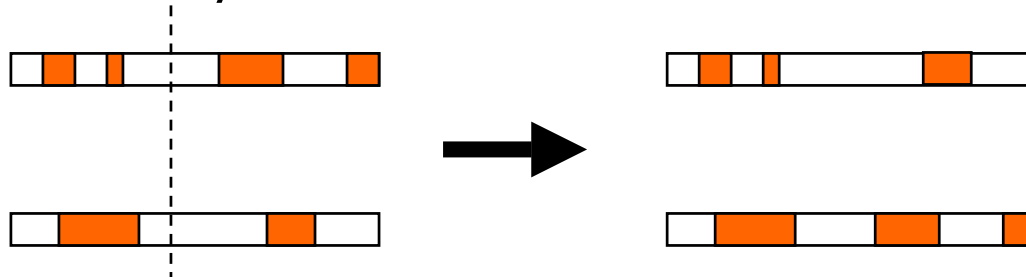
4. **Turnajom [tournament]** - náhodne sa vyberú 2 jedinci a vygeneruje sa náhodné číslo $r \in \langle 0,1 \rangle$, ak je $r < T$, kde $T \in \langle 0,1 \rangle$ je preddefinovaný parameter, tak bude okopírovaný jedinec s vyšším ohodnotením inak ten druhý

- Pri ďalších genetických operáciách sa môže doteraz najlepší jedinec stratiť. Pomoc: **elitizmus [elitism]**- S najlepších jedincov je bez zmeny okopírovaných do novej generácie

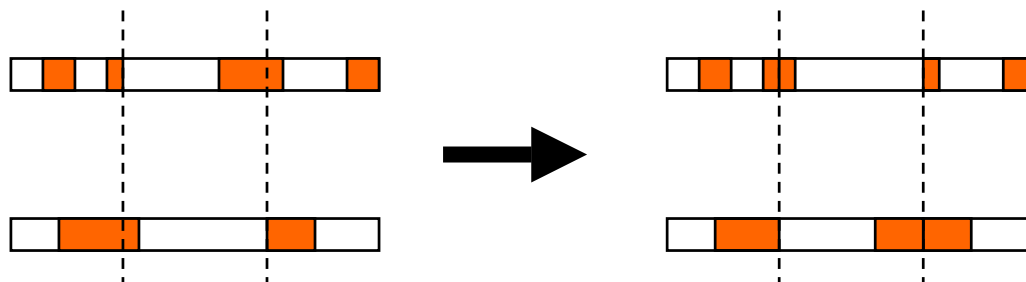
Kríženie

- Jedinci sú náhodne spárovaní a každý pár je s danou pravdepodobnosťou skrížený

Jednobodové
kríženie



Viacbodové
kríženie



Uniformné
kríženie

Náhodný výber bitu od jedného z rodičov

Mutácia, distribuované GA

- Každý prvok chromozómu je s danou pravdepodobnosťou zmenený
 - bit sa neguje, v prípade iných oborov hodnôt sa nahradí náhodnou hodnotou z daného oboru hodnôt, alebo sa pričíta náhodná hodnota podľa nejakého rozdelenia so stredom 0
- jedinci v populácii sú rozmiestnení napr. v dvojrozmernom priestore (napr. toroid); selekcia a kríženie sa deje iba lokálne, subpopulácie sa prekrývajú – tým je možné šírenie „dobrých“ vlastností cez celú populáciu

Nedeterminizmus

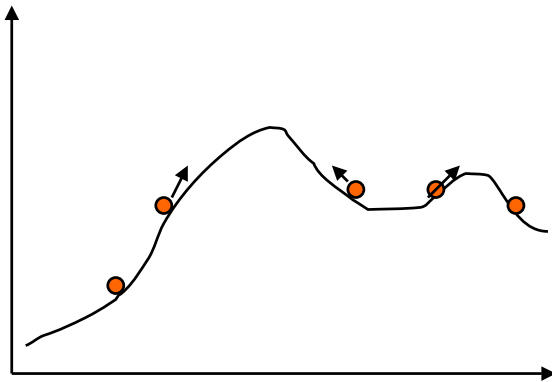


- Celý GA je nedeterministický, používa náhodné veličiny.
- Typicky sa sleduje priemerné a maximálne ohodnotenie, GA sa zastaví, keď je dosiahnutá dopredu zadaná hodnota ohodnotenia, alebo zadaný počet generácií, alebo sa ohodnotenie v priebehu niekoľkých generácií nemení.
- Následne sa GA spúšťa opakovane aj s pozmenenými parametrami.
- Výsledkom je najlepší jedinec vybraný z finálnych generácií zo všetkých behov GA.

Základné bloky a teória schém

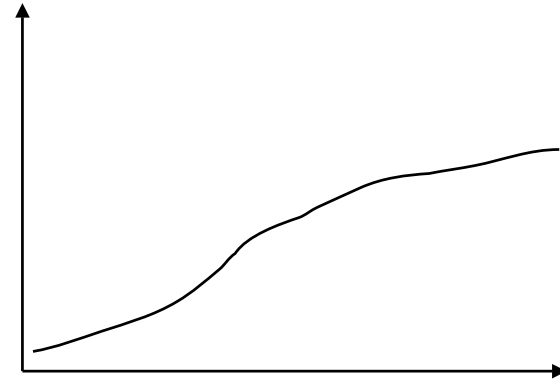
- Ohodnotenie si môžeme predstaviť ako mnohorozmernú nadrovinu udávajúcu ohodnotenie vo všetkých možných hodnotách genotypu – krajina s kopcami a údoliami [fitness landscape]; pre chromozóm dĺžky 1 je to funkcia jednej premennej

Ohodnotenie



Priestor genetic. vlastností

Max. ohodnotenie



generácie

Schémy

- Schéma je šablóna popisujúca nejakú skupinu reťazcov
 - $1^{**}1 = \{1001, 1011, 1101, 1111\}$ * zastupuje ľubovoľnú prípustnú hodnotu
- binárne reťazce dĺžky l umožňujú až 3^l rôznych schém
- populácia s N jedincami s reťazcami dĺžky k obsahuje niečo medzi 2^k a $N2^k$ schémami
- schéma môže popisovať komponentu chromozómu, ktorá zaručuje vysoké ohodnotenie
- potenciálne umožňuje preskúmať viacej reťazcov než ich je v populácii
- Holland 1975 - GA spracuje v jednom kroku až N^3 schém, aj keď má iba N reťazcov (implicitný paralelizmus GA)

Schémy

- niektoré schémy majú vyššie priemerné ohodnotenie;
- dlhé schémy sa ľahko rozbijú $1*****1$ $***01**$
- Najdôležitejšie sú schémy s krátkou dĺžkou. Schémy s krátkou dĺžkou a vysokým ohodnotením sa objavujú v exponenciálne mnohých potomkoch v priebehu GA.
- Schémy sú považované za základné bloky evolúcie a kríženie je hlavný operátor, pretože umožňuje preskúmať kombinácie schém
- Funguje to však iba pri vhodnom kódovaní, kde základné bloky majú krátku dĺžku.
- POZOR: niekedy kríženie „kazí“ výsledky a pravdepodobnosť kríženia sa preto nastavuje na 0, alebo na veľmi malú hodnotu.

Veta o schémach I

- $o(H)$ ozn. **rád schémy** H = počet pevných pozícií v schéme (s hodnotou 0 alebo 1 pre binárnu abecedu). Napr.
 - $o(011*1**)= 4$
 - $o(1*****)= 1$
- $\delta(H)$ ozn. **dĺžku schémy** H = vzdialenosť medzi prvou a poslednou pevnou pozíciou (s hodnotou 0 alebo 1 pre binárnu abecedu)
 - $\delta(011*1**)= 4$
 - $\delta(1*****)= 0$
- analyzujeme vývoj GA – vplyv reprodukcie, kríženia a mutácie na počet reťazcov zodpovedajúcich danej schéme

Vplyv reprodukcie na očakávaný počet reťazcov danej schémy v populácii

- Predpokladajme, že v kroku t obsahuje populácia m reťazcov odpovedajúcich schéme H ($m = m(H, t)$)
- pri reprodukcii je reťazec A_i vybraný do ďalšej populácie (podľa jeho ohodnotenia $f_i = \Phi(A_i)$) s pravdepodobnosťou

$$p_i = \frac{f_i}{\sum_{j=1}^n f_j}$$

- v ďalšej populácii veľkosti n bude očakávaný počet reťazcov odpovedajúcich schéme H ($m = m(H, t+1)$) určený pomocou

$$m(H, t+1) = m(H, t) \cdot n \cdot \frac{f(H)}{\sum_{j=1}^n f_j}$$

- kde $f(H)$ odpovedá priemernému hodnoteniu reťazcov odpovedajúcich schéme H

Vplyv reprodukcie na očakávaný počet reťazcov danej schémy v populácii

- $f(H)$ odpovedá priemernému hodnoteniu reťazcov odpovedajúcich schéme H v kroku i

- potom pre

$$\bar{f} = \frac{\sum_j f_j}{n}$$

- dostávame

$$m(H, t+1) = m(H, t) \cdot \frac{f(H)}{\bar{f}}$$

- **⇒ pri prostej reprodukcii počet reťazcov odpovedajúcich určitej schéme v populácii rastie, resp. klesá, podľa ich priemerného ohodnotenia**

Vplyv kríženia na očakávaný počet reťazcov danej schémy v populácii

- Pre každú schému je možné určiť pravdepodobnosť jeho pretrvania po krížení p_s podľa

$$p_s = 1 - \frac{\delta(H)}{l-1}$$

- (teda schéma "prežije", ak bude bod kríženia "mimo jeho dĺžky")
- Ak kríženie nastáva náhodne – s pravdepodobnosťou p_c – bude pravdepodobnosť pretrvania schémy po krížení

$$p_s = 1 - p_c \frac{\delta(H)}{l-1}$$

- \Rightarrow vplyv kombinácie reprodukcie a kríženia na počet reťazcov schémy H v populácii bude

$$m(H, t+1) \geq m(H, t) \cdot \frac{f(H)}{\bar{f}} \cdot \left[1 - p_c \frac{\delta(H)}{l-1} \right]$$

Vplyv kríženia na očakávaný počet reťazcov danej schémy v populácii

$$m(H, t+1) \geq m(H, t) \cdot \frac{f(H)}{\bar{f}} \cdot \left[1 - p_c \frac{\delta(H)}{l-1} \right]$$

- Pri reprodukcií a krížení počet reťazcov schémy H rastie, resp. klesá, v závislosti na
 - ohodnotení schémy
 - dĺžke schémy

Vplyv mutácie na očakávaný počet reťazcov danej schémy v populácii

- Náhodná zmena na jednej pozícii s pravdepodobnosťou p_m
- aby "prežila" schéma H , tak sa musí zachovať každá z jeho pevných pozícií
 - | každá pozícia prežíva s pravdepodobnosťou $1 - p_m$
 - | všetky mutácie sú navzájom nezávislé
- schéma H "prežije", ak prežije každá z jeho $o(H)$ pevných pozícií
- pravdepodobnosť, že schéma H "prežije" mutáciu je

$$(1 - p_m)^{o(H)}$$

- aproximácia pre malé hodnoty p_m ($\ll 1$)

$$(1 - p_m)^{o(H)} \approx 1 - o(H)p_m$$

Vplyv reprodukcie, kríženia a mutácie na očakávaný počet reťazcov danej schémy v populácii

- očakávaný počet reťazcov po reprodukcii, krížení a mutácii

$$m(H, t+1) \geq m(H, t) \cdot \frac{f(H)}{\bar{f}} \cdot \left[1 - p_c \frac{\delta(H)}{l-1} - o(H) \cdot p_m \right]$$

- **⇒ najväčšiu "šancu na prežitie" majú krátke schémy s malým počtom pevných pozícií a nadpriemerným ohodnotením**