

## Kapitola 4

# Evolučné algoritmy

Charles Darwin bol Angličan, syn významného lekára. Vyštudoval teológiu, po štúdiu sa zaoberal geologickými formáciami v horách Walesu. Koncom roka 1831 však odišiel na 5-ročnú výskumnú cestu okolo sveta. Loď HMS Beagle ho zaviedla aj na Galapágy, ekvádorské súostrovie 19 sopečných ostrovov vo východnej časti Tichého oceánu, kde zhromaždil podľa jeho slov najcennejšiu časť prírodovedeckého materiálu, ktorý použil vo svojom najväčšom diele publikovanom v roku 1859 – *O vzniku druhov prírodným výberom alebo uchovávanie prospešných plemien v boji o život*. Hlavou plnou prírodovedných informácií, získaných z okružnej plavby okolo sveta, v ňom vrhá ucelený pohľad na vývoj druhov oslobodený od spirituality a náboženských predstáv svojej a predchádzajúcich dôb. Darwin vysvetľuje vznik rôznych druhov organizmov na základe prirodzeného výberu, teda schopnosti prežiť len tých najschopnejších. Významným argumentom pre jeho teóriu boli aj stratigrafické<sup>1</sup> výsledky geológa Charlesa Lyella, ktoré podporovali rodiacu sa evolučnú teóriu v oblasti „časovej zložitosti“. Aplikované princípy klasickej evolučnej teórie s mnohými „vylepšeniami“ hrajú významnú úlohu vo fonde vedomostí ľudstva.

Evolučné algoritmy (EA), ktoré sú postavené na myšlienkach evolučnej teórie, začali vznikať už v 50-tych rokoch 20. storočia. Výraznejší záujem však nastal až približne o 30 rokov neskôr, kedy David Goldberg významne rozšíril prácu Johna Hollanda o genetických algoritmoch (z roku 1975 [34]) v práci publikovanej v roku 1989 [30]. Značným impulzom pre popularizáciu EA bola prvá väčšia práca o genetickom programovaní, ktorej autorom je John Koza [44].

### 4.1 Biologické pojmy v kontexte evolučných algoritmov

Vo zvyšku tejto práce sa budú vyskytovať pojmy pochádzajúce z biológie, no sémantika nie všetkých je zhodná so sémantikou v kontexte EA. Pre ujasnenie pojmov je uvedený ich krátky prehľad.

Medzi základné pojmy Darwinovej evolučnej teórie patrí populácia. Populácia je množina jedincov, ktorí sú reprezentovaní svojím genetickým materiálom. V tejto práci budú voľne zamieňané pojmy genetický materiál, genóm a chromozóm, aj keď z pohľadu biológie tieto pojmy nie sú rovnocenné. Genotyp je vlastné zakódovanie genetickej informácie do určitej štruktúry. Spôsob, akým sa genotyp v danom prostredí interpretuje, ako dobre rieši nastolený problém, sa nazýva fenotyp. Jedinec s rovnakým genotypom môže mať v inom prostredí odlišnú schopnosť prežitia, inými slovami, odlišné prostredie spôsobí odlišnú inter-

<sup>1</sup> Stratigrafia je geologický vedný obor, ktorý študuje vek sedimentárnych vrstiev hornín.

pretáciu genotypu na fenotyp. Genetický materiál sa skladá z lineárne usporiadaných génov, v kontexte EA jeden gén kóduje jednu vlastnosť. Konkrétna vlastnosť, hodnota génu, sa nazýva alela. V rámci počítačovej terminológie môžeme povedať, že každý gén reprezentuje určitý dátový typ a alely sú hodnotami daného dátového typu, génu.

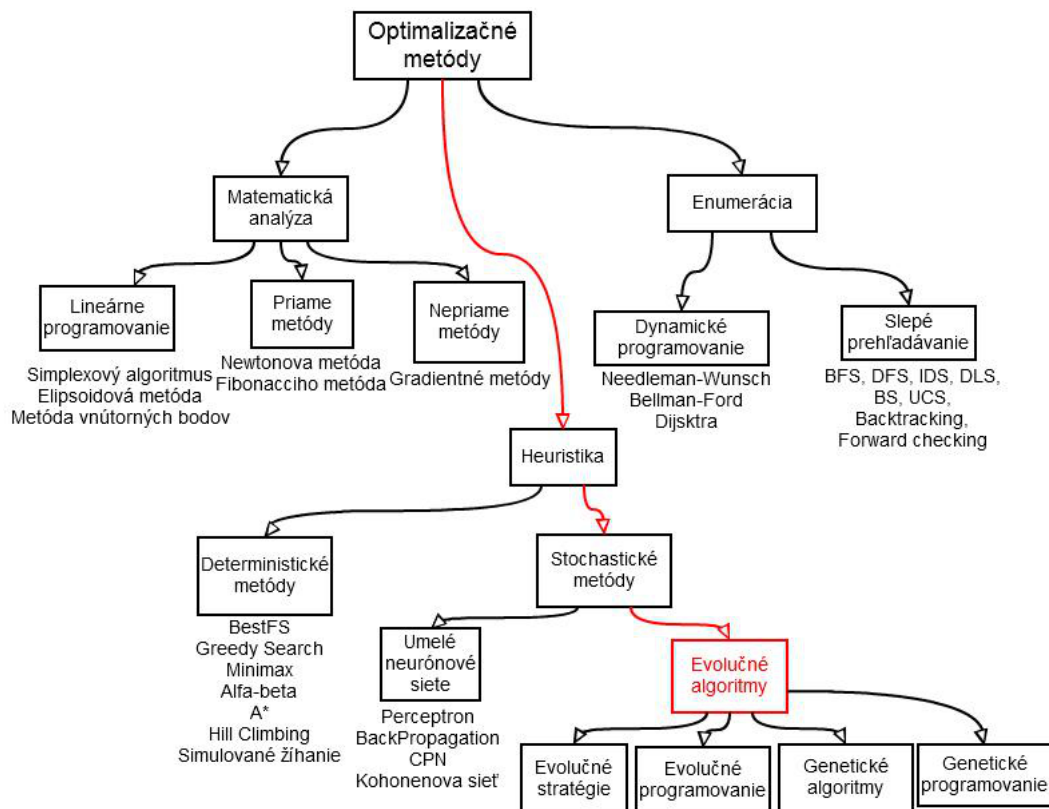
## 4.2 Klasifikácia evolučných algoritmov

Existujú problémy, ktorých exaktný model sa nedá, nevieme alebo je veľmi náročné zostaviť, a navyše počet riešení daného problému, resp. stavový priestor úlohy je obrovský. Kvôli takýmto problémom vznikol pojem *softcomputing* a všetko čo podneš spadá. Je to spôsob riešenia problémov „s rozumom“, teda nie hrubou (hard) silou ako je to prípade klasických algoritmov na prehľadávanie stavového priestoru (napr. BFS, DFS). Softcomputing rieši problémy pomocou určitej heuristiky resp. heuristickej funkcie, ktorou algoritmus „myslí“. Potrebne je dodať, že metódy softcomputingu väčšinou nenájdu najlepšie riešenie, ale len suboptimálne, ktoré však takmer vždy postačuje. Obrázok 4.1 zobrazuje zasadenie evolučných algoritmov do kontextu všetkých významných optimalizačných techník.

Evolučné algoritmy sú spoločným vyjadrením pre množinu moderných matematických postupov, ktoré využívajú modely evolučných procesov v prírode založených na Darwinovej evolučnej teórii popísanej na začiatku tejto kapitoly. Jednotlivé riešenia, ktoré tvoria populáciu, sa vyvíjajú na základe klasických evolučných a genetických operátorov ako je selekcia, kríženie či mutácia. EA stavajú a súčasne aj padajú na fitness funkciu, ktorá definuje schopnosť jedinca prežiť v danom prostredí resp. schopnosť riešenia riešiť daný problém. Evolučné algoritmy sú v poslednej dobe veľmi rozšírené. Prílišná popularizácia so sebou však prináša aj nerealistické očakávania. Podobne ako evolučné algoritmy, tak všeobecne aj všetky optimalizačné techniky nefungujú ako univerzálne metódy, ale každá z nich sa hodí na inú oblasť problémov. Zistiť, ktorá technika je najlepšia, prípadne s akými parametrami, je už úlohou inžinierskeho návrhu.

Podstatným rozdielom oproti klasickým optimalizačným metódam je práca nie s jedným, ale s množinou riešení, na ktorú je možno aplikovať genetické a iné operátory. Princíp evolúcie jednotlivých riešení v rámci populácie popisuje nasledovná všeobecná schéma evolučného algoritmu [36]:

1. Vynuluj hodnotu počítadla generácii  $t = 0$ .
2. Náhodne vygeneruj počiatočnú populáciu  $P(0)$ .
3. Vypočítaj ohodnotenie (*fitness*) každého jedinca v počiatočnej populácii  $P(0)$ .
4. Vyber dvojice jedincov z populácie  $P(t)$  a vytvor ich potomkov  $P'(t)$ .
5. Vytvor novú populáciu  $P(t+1)$  z pôvodnej populácie  $P(t)$  a množiny potomkov  $P'(t)$ .
6. Zväčši hodnotu počítadla generácii o jedna ( $t := t + 1$ ).
7. Vypočítaj ohodnotenie (*fitness*) každého jedinca v populácii  $P(t)$ .
8. Ak je  $t$  rovné maximálnemu počtu generácii alebo je splnené iné ukončovacie kritérium, vráť ako výsledok populáciu  $P(t)$ ; inak pokračuj krokom číslo 4.



Obrázek 4.1: Klasifikácia evolučných algoritmov v kontexte optimalizačných techník. Soft-computing spadá medzi stochastické heuristické metódy.

Keď v roku 1963 začali Hans-Paul Schwefel a Ingo Rechenberg na Technickej univerzite v Berlíne s napodobňovaním vývoja v prírode, boli presvedčení, že ich metóda najlepšie aproximuje evolúciu v živej prírode. Preto svoju metódu nazvali, celkom všeobecne, *evolučné stratégie* (ES). Postupom času sa však ukázalo, že tento spôsob rieši len určitý typ úloh, hlavne v stavebnom a strojnom inžinierstve. Genetické algoritmy nie sú teda podradené evolučným stratégiám, ako sa domnievali, ale naopak, svojou popularitou ich zatieňujú [40]. Genóm v rámci ES je zložený z génov, ktoré sú reprezentované reálnymi číslami, z čoho vyplýva implementačná diferenciácia genetických operátorov. Mutačný operátor je väčšinou implementovaný pomocou pripočítania hodnoty podľa Gaussovej funkcie. Takýto spôsob mutácie rieši jeden problém genetických algoritmov, ktorý znie: malé zmeny genotypu nemusia viesť k malým zmenám fenotypu<sup>2</sup>.

### 4.3 Evolučné operátory

Evolučné procesy z biológie boli aplikované a svojím spôsobom interpretované v teórii evolučných algoritmov. Ide o pekný príklad medzioborového transféru informácií. V evolučnom procese je nutná značná variácia genómu, ktorá je zabezpečovaná rekombinačnými (gene-

<sup>2</sup> Tento problém sa dá obísť napríklad pomocou Grayovho kódovania, v ktorom sa každé dve po sebe idúce hodnoty líšia v bitovom vyjadrení len na jednej pozícii.

tickými) operátormi. Následný výber najlepších jedincov reprezentuje operátor selekcie.

*Selekcia* je evolučný operátor, ktorý určuje, ktoré riešenie v populácii riešení prežije, a ktoré nie. Reprezentuje prirodzený výber popísaný Darwinom. Rozoznávame 3 najpoužívanější typy selekcie a ich varianty:

- **Koleso šťastia** – jednotlivým jedincom sa priradí pravdepodobnosť výberu do ďalšej generácie na základe hodnoty fitness funkcie, „lepší“ jedinci budú vybraný s vyššou pravdepodobnosťou.
- **Turnaj** – je založený na náhodnom výbere  $n$ -tíc jedincov a ich súboji, v ktorom sú zbraňami hodnoty fitness funkcie, víťaz je vo väčšine variant tejto selekcie vybraný do ďalšej generácie vždy, no môže byť vybraný s pravdepodobnosťou menšou než 1.
- **„Najlepší vyhráva“** – je najjednoduchším typom selekcie, v každej generácii preferuje len tých najstatnejších jedincov, jedincov umiestnených na čelných priečkach rebríčka zostavovaného komisiou, ktorá hodnotí statnosť jedinca na základe hodnoty fitness funkcie. Popisovaný spôsob výberu je vhodný v prípade, keď fitness funkcia nemá veľa extrémov, pretože evolučné algoritmy založené na tomto type selekcie nevedia riešiť tzv. klamné problémy a multimodálne funkcie, pretože v populácii sa nezachováva diverzita jedincov, hodnôt fitness funkcie. Evolučné algoritmy založené na tomto type selekcie častokrát predbežne konvergujú, uviaznu v lokálnom extréme fitness funkcie.

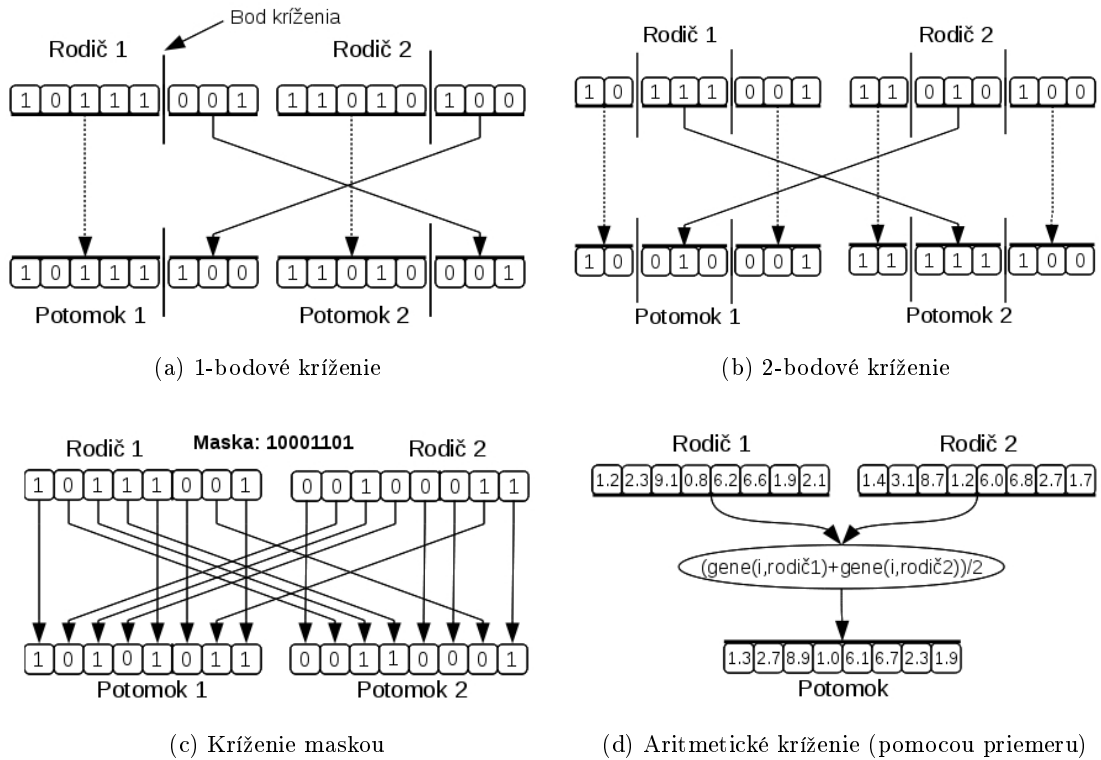
Pri praktických aplikáciach je len málokedy možné sa stretnúť s určitým typom selekcie v základnej podobe, takmer vždy sa používajú ich možné variácie a kombinácie. Pre priblíženie, existujú aj prístupy, ktoré pracujú s dvoma populáciami kvôli zachovaniu rôznorodosti populácie a dochádza k migrácii medzi populáciami. Každá populácia je však založená na inej fitness funkcii [53]. So selekciou úzko súvisí obnova populácie. Po vyhodnotení hodnôt fitness funkcie jedincov populácie a selekcii jedincov, ktorí „prežijú“, máme viac možností ako nahradiť aktuálnu populáciu. Rozlíšujeme 2 základné prístupy:

- **Úplná obnova populácie** – dochádza k vymieraniu rodičov, teda celá generácia je nahradená novou.
- **Čiastočná obnova populácie (steady state)** – potomkami sa nahradí len určitá časť jedincov.

*Kríženie* je základným rekombinačným operátorom, pomocou ktorého sa mieša genetická informácia 2 jedincov. Rôzne techniky kríženia majú spoločnú tú vlastnosť, že ide vždy o vzájomnú výmenu častí chromozómov. V niektorých prípadoch môže byť však potrebné a užitočné dať jedincom možnosť prežiť bezo zmeny a uchovať pre budúcu populáciu kópie rodičovských jedincov. Preto sa operátor kríženia aplikuje s istou pravdepodobnosťou a v ostatných prípadoch sú za potomkov rodičovského páru prehlásené ich priame kópie. Pravdepodobnosť použitia operátora kríženia je obvykle relatívne vysoká (0,75–0,95 [64]). Kríženie umožňuje rýchlu výmenu relatívne veľkého množstva genetickej informácie a do značnej miery ovplyvňuje efektívnosť evolučného algoritmu [36]. K tomuto operátoru možno pristupovať viacerými spôsobmi (vizuálna podoba je zobrazená na obrázku 4.2):

- **$N$ -bodové kríženie** – najpoužívanější typ kríženia, väčšinou sa používa kríženie jednobodové alebo dvojbodové, závisí samozrejme na danom probléme a veľkosti chromozómu.

- **Kríženie maskou** – náhodne sa vygeneruje bitová maska, ktorej dĺžka je zhodná s dĺžkou chromozómu a noví jedinci sa tvoria tak, že gény na pozíciách obsahujúcich „0“ zdedí prvý potomok a gény na pozíciách obsahujúcich „1“ zdedí potomok druhý.
- **Aritmetické kríženie** – využíva sa najmä pri volučných stratégiách, kde sú gény reprezentované reálnymi číslami, noví jedinci sa tvoria na základe aplikácie nejakého aritmetického operátora (väčšinou priemer) na gény rodičov.

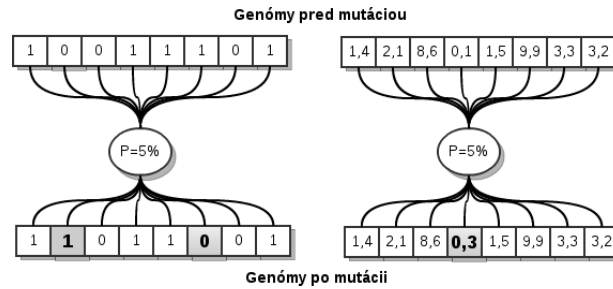


Obrázek 4.2: Rôzne typy evolučného operátora kríženia, prevzaté z [11].

Operátor *mutácie* väčšinou veľmi jednoduchým spôsobom a s relatívne malou pravdepodobnosťou (0,001–0,05 [64]) náhodne mení hodnotu jednotlivých génov. V prípade binárneho kódovania to konkrétne znamená, že vybraný gén v danom chromozóme zmení svoju hodnotu z nuly na jednotku a naopak. Slúži ako nástroj, ktorý bráni príliš rýchlemu „zjednotvárneniu“ vlastností v rámci populácie, strate užitočného genetického materiálu a predčasnej konvergencie populácie [36]. Rozlišujeme v zásade 2 typy mutácie (ilustruje obrázok 4.3):

- **Inverzia génu** – vhodné a použiteľné len pri binárnej reprezentácii génov, teda génov, ktoré môžu existovať len v 2 navzájom rôznych alelách
- **Pripočítanie hodnoty rozloženia pravdepodobnosti** – využíva sa pri reprezentácii génov reálnymi číslami, k hodnote génu sa pripočíta hodnota daná určitým rozdelením pravdepodobnosti

Špecificky definované chromozómy jedincov si vyžadujú špecifické operátory kríženia a mutácie. Rozšírené je tzv. *permutačné kódovanie* kandidátneho riešenia. Využíva sa väčšinou pri problémoch, kedy je riešením permutácia určitých hodnôt. Podrobne študovaným



Obrázek 4.3: Mutácia genómu, ľavá časť obrázku reprezentuje aplikáciu operátoru mutácie pomocou inverzie génu, pravá časť pomocou pripočítania hodnoty rozloženia pravdepodobnosti.

je napríklad problém obchodného cestujúceho<sup>3</sup>. Mutácia takto zakódovaných jedincov sa najčastejšie rieši prehodením hodnôt dvoch náhodne vybraných génov mutovaného chromozómu. Alternatívnym prístupom je obrátenie podsekvencie chromozómu (angl. Reverse Sequence Mutation), kedy sa vezme dvomi náhodne vygenerovanými pozíciami obmedzená podsekvencia chromozómu a poradie génov v sekvencii sa otočení [4]. Kríženie je zložitejšie, existujú 3 najpoužívanéjšie prístupy [4] (pre zamedzenie konfúzií sú ponechané ich originálne anglické názvy):

- **Order 1 crossover (OX)** – náhodne sa vygenerujú body kríženia, gény medzi bodmi kríženia 1. rodiča sa skopírujú do potomka, následne sa postupne prechádzajú jednotlivé gény 2. rodiča (od 2. bodu kríženia) a do potomka sa kopírujú len tie gény, ktoré sa nenachádzajú v časti medzi bodmi kríženia 1. rodiča.
- **Partially-mapped crossover (PMX)** – náhodne sa vygenerujú body kríženia, gény medzi bodmi kríženia 1. rodiča sú skopírované do potomka. Prechádzajú sa gény medzi bodmi kríženia 1. rodiča zľava doprava, aktuálny gén sa „spojí“ s protiahlým génom a s génom s rovnakou hodnotou v 2. rodičovi. Ak protiahlý gén nemá rovnakú hodnotu ako aktuálny, skopírujú sa tieto 2 gény a prehodia ich pozície, následne sa postupným posunom doprava tento algoritmus opakuje. Následne sa do potomka skopírujú len tie gény z 2. rodiča, ktoré sú mimo body kríženia, pričom sa zachováva pozícia génov. V poslednom kroku sa prechádzajú jednotlivé gény 2. rodiča, začína sa génom za 2. bodom kríženia a do potomka sa kopírujú tie gény, ktorých hodnota sa nenachádza v časti medzi bodmi kríženia 1. rodiča alebo v potomkovi.
- **Cycle crossover (CX)** – kríženie sa začína s génom na 1. pozícii 1. rodiča (alebo na inej štartovacej pozícii) a skopíruje sa na 1. pozíciu potomka. Potomok teda nemôže dediť 1. gén z 2. rodiča, takže tento gén musí byť vyhľadaný v 1. rodičovi a skopírovaný do potomka. Nech tento gén má pozíciu  $x$  v 1. rodičovi. Potom je zdedený potomkom na pozícii  $x$  a nemôže byť teda zdedený od 2. rodiča. Tento proces sa opakuje, kým sa nevytvorí cyklus, teda kým sa nedosiahne gén, ktorý už potomok zdedil. Potom sa vyberie gén z 2. rodiča, a vytvorí sa analogicky cyklus, tentokrát však potomok dedí gény 2. rodiča. Nevýhodou tohto prístupu môže byť to, že podsekvencie dedených génov nemusia byť súvislé.

<sup>3</sup> Problém obchodného cestujúceho (angl. Travelling Salesman Problem) je optimalizačný problém nájdenia najkratšej možnej cesty prechádzajúcej všetkými zadanými bodmi na mape. Matematicky ide o nájdenie Hamiltonovskej cesty v grafe  $G$  s najnižšou cenou  $C$ .