

Diploidní genetický algoritmus

Petr Dvořáček – xdvora0n@stud.fit.vutbr.cz

19. dubna 2014

1 Úvod

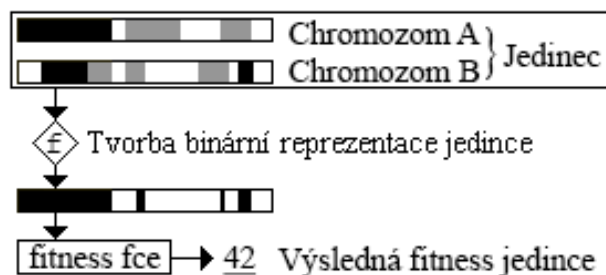
Tento dokument pojednává o diploidních genetických algoritmech a jejich využití v úlohách závislých na změně prostředí. Mezi takovéto úlohy patří například dynamický problém obchodního cestujícího. V tomto projektu se však zaměřuji na binárně reprezentované dynamické úlohy, jako je například One-max problém.

V druhé kapitole popisují diploidní genetické algoritmy, včetně tvorby fenotypů a tvorby dynamického prostředí. V třetí kapitole se věnuji experimentům, kde porovnávám DGA s GA. Na závěr shrnuji dosažené výsledky.

2 Princip algoritmu DGA

Princip algoritmu vychází z procesů biotické mitózy. Jedná se o tzv. pohlavní rozmnožování. Pro zjednodušení řekněme, že každý jedinec vlastní jeden pár chromozomů. První z nich zdědil od matky, druhý od otce. Před předáním genetické informace se tento pár chromozomů navzájem zkříží. Vzniknou tedy dva nové chromozomy, které se dědí do dalších generací.

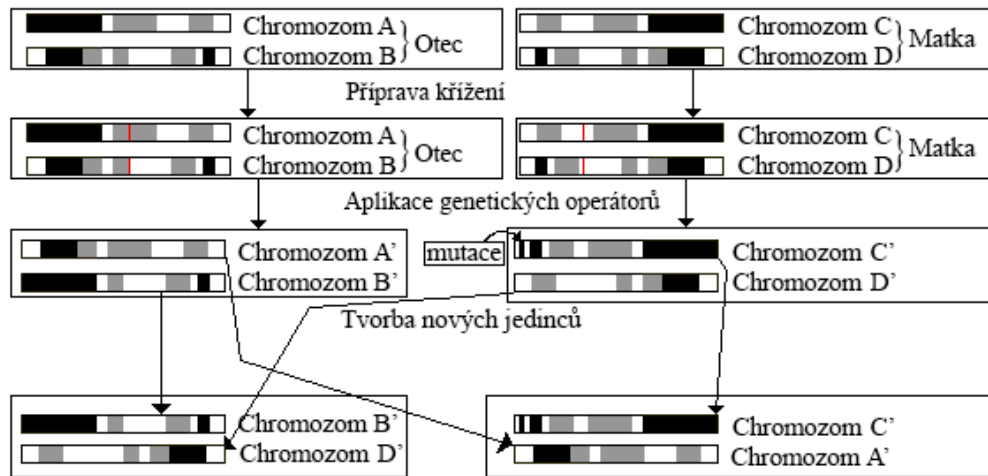
Na rozdíl od klasických genetických algoritmů (GA) jsou tedy diploidní genetické algoritmy (DGA) reprezentovány dvěma chromozómy. Z nich je potřeba vytvořit strukturu, se kterou se pak lépe pracuje. Této struktuře říkáme fenotyp. Fenotyp se ohodnocuje v závislosti na aktuálním prostředí. Tento princip tvorby je zobrazen na obrázku 1 níže.



Obrázek 1: Tvorba fenotypu.

Další změnou v DGA je lehká úprava reprodukce, která je rozdělena do dvou částí. V prvním kroku si oba jedinci vytvoří svou dočasnou kopii – haploidní chromozomy. V této kopii pak dochází ke křížení chromozomů daného rodiče a následné mutaci. Díky reprezentaci DGA mohou být některé

mutace neutrální – nemají žádný vliv na fenotyp jedince. V druhém kroku dojde k výměně jednoho chromozomu mezi oběma dočasnými kopiemi a vzniknou tak dva nové jedinci.



Obrázek 2: Tvorba nových jedinců.

2.1 Tvorba fenotypů pomocí dominantních schémat

2.1.1 Deterministická dominance

Existuje řada dominantních schémat, pomocí kterých se určuje fenotyp jedince. Asi nejznámější je tri-alelické schéma od Hollstein-Hollanda, které je reprezentováno množinou $\{0, 1, 2\}$. Prvky této množiny se nazývají alely. Hollandovo schéma je tri-alelické, neboť množina má mohutnost právě tři. Princip převodu genotypu na fenotyp je znázorněn v tabulce 1. Obsahuje-li jeden gen hodnotu 2, pak fenotyp nabude hodnotu 1. Mají-li oba geny hodnotu 1, pak i fenotyp získá 1. Ve zbylých případech fenotyp bude 0. Toto schéma můžeme vylepšit takovým způsobem, že zvětšíme jeho kardinalitu, jak je znázorněno v tabulce 1. Toto schéma může být zapsáno rovnicí:

$$P_i = \begin{cases} 0 & m + n < c + 1 \\ 1 & m + n \geq c + 1 \end{cases}$$

Kde m respektive n je hodnotou daného genu jedince z prvního resp. druhého chromozomu. Kardinalitu pak značí c . Těmto DGA říkáme deterministicky dominantní diploidní genetické algoritmy (DDDGA, či zjednodušeně DDGA).

Tabulka 1: Schéma deterministické dominance Hollstein–Holland

	0	1	2
0	0	0	1
1	0	1	1
2	1	1	1

Tabulka 2: Deterministické schéma s kardinalitou 4.

	0	1	2	3
0	0	0	0	1
1	0	0	1	1
2	0	1	1	1
3	1	1	1	1

2.1.2 Nedeterministická dominance

U nedeterministických schémat je to trochu odlišné. Vybírají se alely s větší dominancí $(0, 1)$ než ty, které ji mají menší $(0, i)$. Mají-li dva rozdílné geny alelu se stejnou dominancí, vybere se náhodně jedna z nich. Jsou-li tyto geny stejné, pak i gen fenotypu má stejnou hodnotu. Tento princip je znázorněn v tabulce 2.

Toto schéma, jehož autor je Ng-Wong, můžeme rovněž rozšířit o kardinalitu. Toho můžeme dosáhnout takovým způsobem, že přidáme dalších dvojic alel. Nedeterministické schéma můžeme rovněž znázornit rovnicí:

$$P_i = \begin{cases} m & m_c > n_c \\ n & m_c < n_c \\ n & m_c = n_c \wedge m = n \\ rand(1, 0) & m_c = n_c \wedge m \neq n \end{cases}$$

Kde m respektive n je hodnotou daného genu jedince z prvního resp. druhého chromozomu. Jejich dominanci pak určují hodnoty n_c, m_c . Těmto DGA říkáme nedeterministicky dominantní diploidní genetické algoritmy (NDDGA, či zjednodušeně NDGA).

Tabulka 3: Nedeterministické schéma s kardinalitou 4.

	0	1	0	1
0	0	0/1	0	1
1	0/1	1	0	1
0	0	0	0	0/1
1	1	1	0/1	1

2.2 Dynamické testovací prostředí

V úvodu bylo zmíněno, že DGA slouží především pro dynamické problémy. Můžeme zavést různé typy dynamických prostředí. Prostor může být měněno třeba náhodně nebo periodicky.

2.2.1 Periodické dynamické prostředí

Pro periodickou změnu prostředí zavedeme vektor prostředí T . Tento vektor T slouží pouze pro binárně reprezentované problémy (Onamax, RoyalRoad...) a jeho velikost je shodná s délkou chromozomu. Jednotlivé prvky vektoru vytvoříme při inicializaci algoritmu pomocí rovnice:

Tabulka 4: Deterministické schéma s kardinalitou 6.

	0	1	0	1	0	1
0	0	0/1	0	1	0	1
1	0/1	1	0	1	0	1
0	0	0	0	0/1	0	1
1	1	1	0/1	1	0	1
0	0	0	0	0	0	0/1
1	1	1	1	1	0/1	1

$T = rand(0, 1) < \phi?1 : 0$. Jak z rovnice vyplývá, prvky vektoru jsou závislé na ϕ , což udává pravděpodobnost, že se prostředí změní. Dále zavedeme vektor templátu M . Tento vektor je pak změněn vektorem T každou k -tou generaci a to následovně: $M = M \text{ xor } T$. Fitness funkce f se pak provádí nad výsledkem operace xor mezi prvky vektoru M a binárně reprezentovanými geny jedince X . Fitness hodnotu dostaneme následovně: $fit = f(X \text{ xor } M)$.

Díky vlastnostem operace xor a díky tomu, že generujeme vektor T při inicializaci algoritmu, můžeme tvrdit, že se prostředí mění periodicky. Jak je znázorněno v tabulce níže, poslední sloupec udává fenotyp jedince X , který se vůbec nemění. Hodnoty fenotypu jsou zobrazeny modrou barvou. Vidíme, že se jeho hodnoty periodicky opakují.

Změna prostředí	T	M	X	$T \text{ xor } M$	$T \text{ xor } M \text{ xor } X$
1.	01101010	00000000	11111111	01101010	10010101
2.	01101010	01101010	11111111	00000000	11111111
3.	01101010	00000000	11111111	01101010	10010101
4.	01101010	01101010	11111111	00000000	11111111
5.	01101010	00000000	11111111	01101010	10010101
6.

2.2.2 Náhodné dynamické prostředí

Chceme-li dosáhnout (pseudo)náhodné změny prostředí, pak postupujeme obdobně jak u periodické změny. Výjimku tvoří pouze inicializace vektoru T . Změní-li se prostředí změní se i daný vektor T . Jak je uvedeno na příkladě níže, hodnoty fenotypu X se mění pseudonáhodně. Fenotyp jedince je zobrazen modrou barvou.

Změna prostředí	T	M	X	$T \text{ xor } M$	$T \text{ xor } M \text{ xor } X$
1.	01101010	00000000	11111111	01101010	10010101
2.	00011001	01101010	11111111	01110011	10001100
3.	00110110	10001100	11111111	10111010	01000101
4.	11001001	01000101	11111111	10001101	01110010
4.	00011110	01110010	11111111	01101100	10010011
6.

2.2.3 Statické fitness funkce

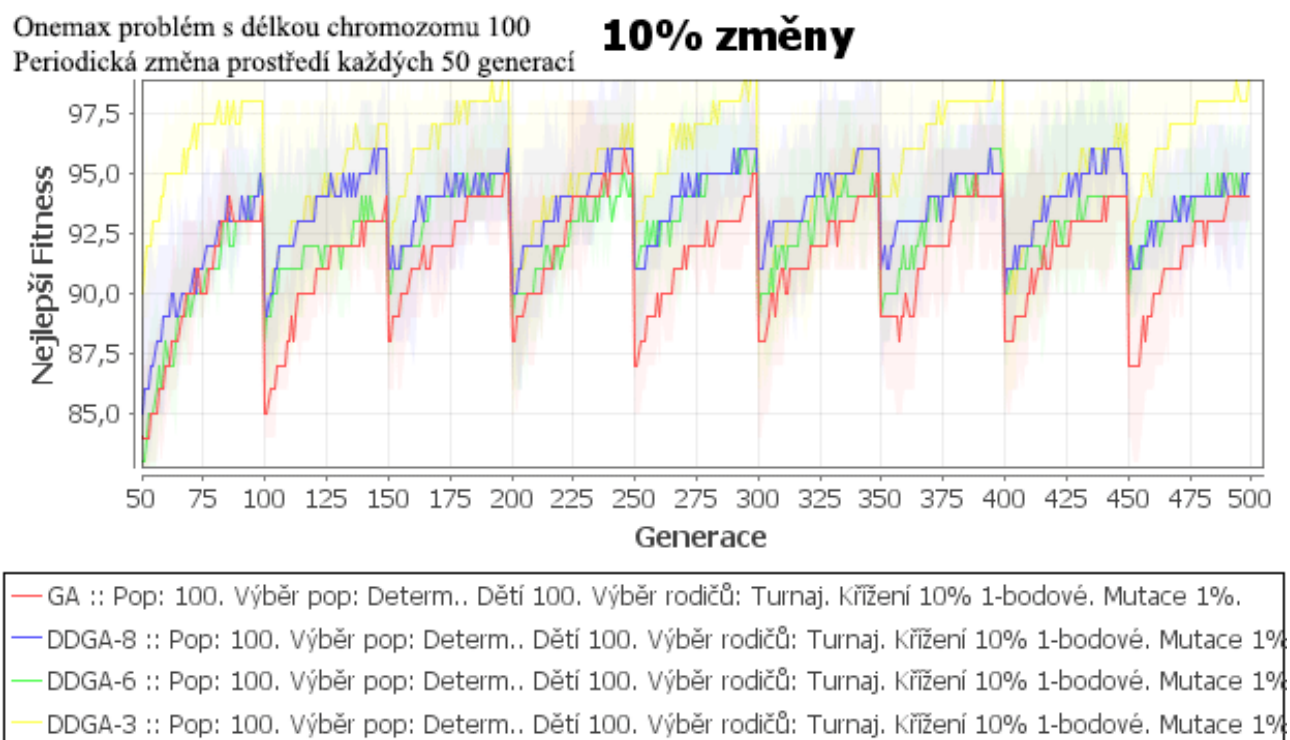
Mezi nejznámější fitness funkce patří One-max a Royal Road. Tyto byly rovněž implementovány. One-max problém se snaží vytvořit jedince, který bude obsahovat samé jedničky. Jeho fitness hodnota je tedy suma jedniček. Royal-road je podobný problému one-max s drobnou odlišností. Chromozom

je rozdělen do bloků po čtveřicích. Obsahuje-li tento blok samé jedničky, pak se k fitness přičte 4, jinak se přičte 0. Dále v tomto projektu byla použita klamná úloha nad problémem RoyalRoad.

3 Experimentální vyhodnocení

Cílem tohoto projektu bylo ukázat, že DGA mohou být vhodnější než klasické GA v dynamických problémech, které se chovají periodicky. Tj. mění své prostředí periodicky.

Jako první testovací úloha byl vytvořen onemax problém o délce chromozomu 100 s periodickou změnou prostředí každých $\tau = 50$ generací. Četnost změny činila $\phi = 10\%$. Pro každý GA bylo vytvořeno přibližně deset běhů se stejnými podmínkami. (Výběr rodičů byl prováděn turnajem, výběr nové populace deterministicky. Pravděpodobnost mutace byla 1% a pravděpodobnost jednobodového křížení byla zvolena na 10%. V každé generaci bylo vygenerováno 100 nových jedinců a velikost populace pak činila 100.) Výsledek, jak můžete vidět na grafu 3, odpovídá očekávání. Všechny DDGA jsou vhodnější než klasický GA. Nejlépe však dopadl DDGA s kardinalitou 3. To je dáno především menším počtem neutrálních mutací, než u DDGA s kardinalitou 8.

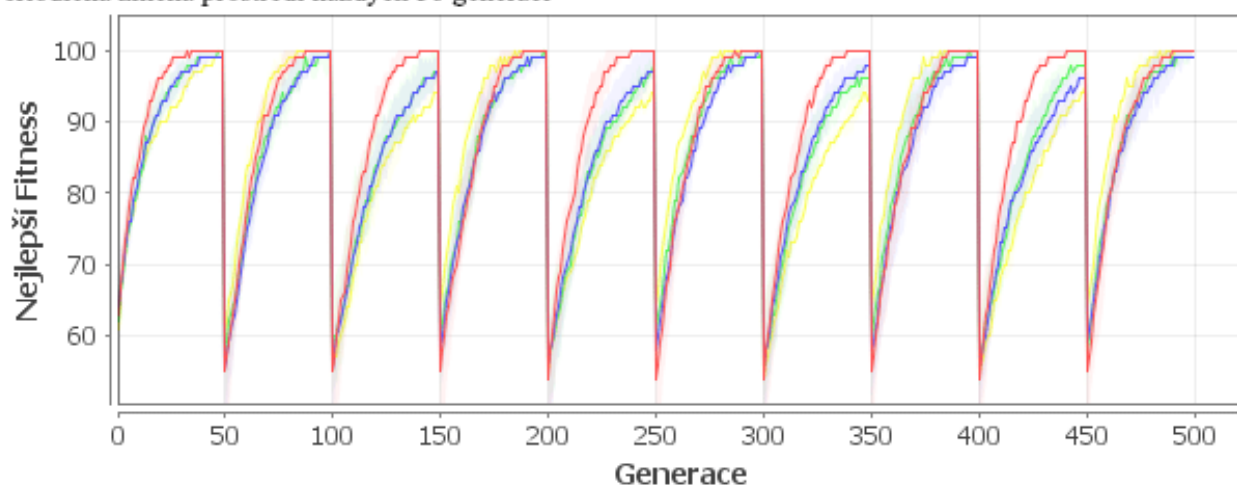


Obrázek 3: Vhodné použití DGA.

Z druhého experimentu plyne, že i klasické GA mohou být stejně dobré jako diploidní GA. To je zapříčiněno tím, že byly upraveny parametry. Zvýšili jsme pravděpodobnost křížení na 70% a četnost změny prostředí $\tau = 50\%$. Selektce populace se provádí turnajem. Jak lze vidět na grafu 4, klasické GA jsou stejně dobré jako DDGA. Tato úloha měla za úkol zdůraznit, že i GA mohou být stejně silné, jako DDGA.

Onemax problém s délkou chromozomu 100
Periodická změna prostředí každých 50 generací

50% změny



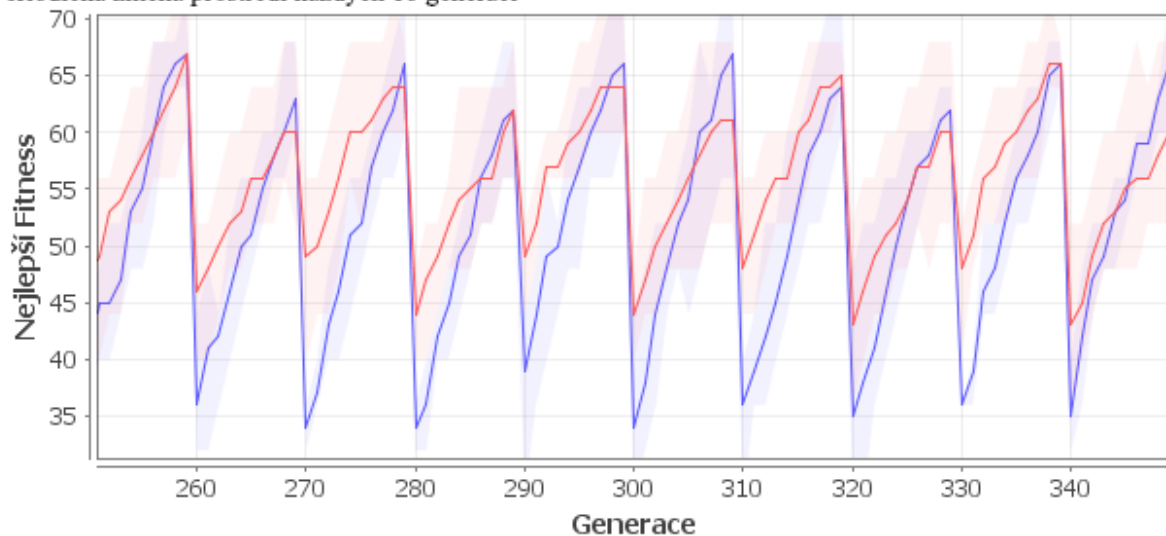
— GA :: Pop: 200. Výběr pop: Turnaj. Děti 200. Výběr rodičů: Turnaj. Křížení 70% 1-bodové. Mutace 1%.
— DDGA-8 :: Pop: 200. Výběr pop: Turnaj. Děti 200. Výběr rodičů: Turnaj. Křížení 70% 1-bodové. Mutace 1%.
— DDGA-6 :: Pop: 200. Výběr pop: Turnaj. Děti 200. Výběr rodičů: Turnaj. Křížení 70% 1-bodové. Mutace 1%.
— DDGA-3 :: Pop: 200. Výběr pop: Turnaj. Děti 200. Výběr rodičů: Turnaj. Křížení 70% 1-bodové. Mutace 1%.

Obrázek 4: DGA a GA mající podobný průběh fitness hodnoty.

Při dalším experimentu byl použit problém RoyalRoad s délkou chromozomu 100 s periodickou změnou prostředí každých 10 generací. Četnost změn činila 50%. Jak můžete na grafu 5 vidět, klasický GA se v tomto případě hůře vyrovnává se změnami než DGA. To je znázorněno menším pádem fitness funkce. Ovšem oproti DGA jeho fitness hodnota roste rychleji. Z toho plyne, že pokud by se prováděla změna prostředí méně frekventovaně, pak by byl klasický GA vhodnější.

Pro předposlední experiment bylo zvoleno náhodné dynamické prostředí. DGA se nevypořádá se změnami prostředí, které jsou náhodné. Neboť geny si pamatují změny, které nastaly. Na grafu 6 můžeme vidět, že všechny DGA se vypořádají se změnami prostředí mnohem hůře než klasický GA.

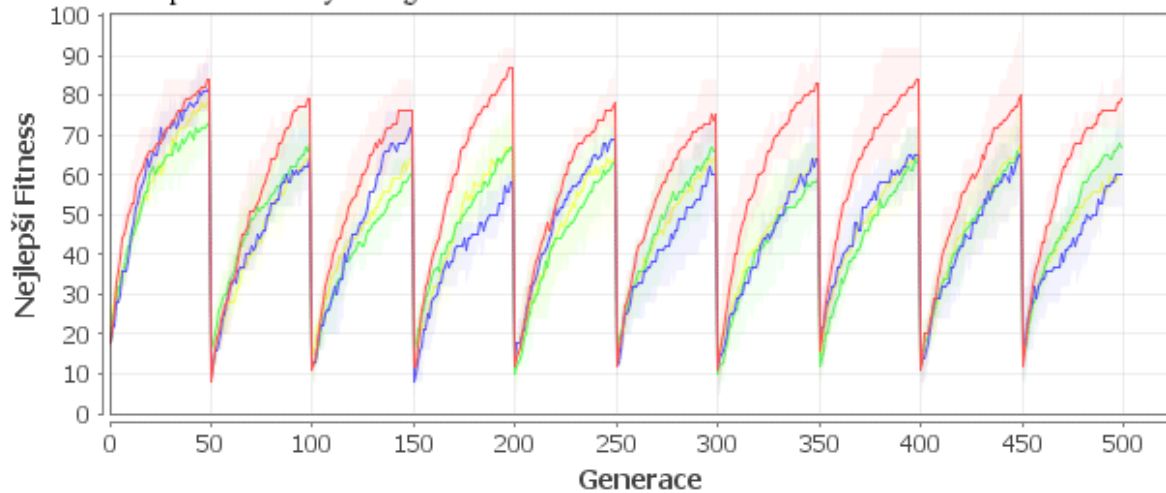
RoyalRoad problém s délkou chromozomu 100 **50% změny**
Periodická změna prostředí každých 10 generací



— DDGA-8 :: Pop: 100. Výběr pop: Turnaj. Děti 200. Výběr rodičů: Turnaj. Křížení 70% 1-bodové. Mutace 1%.
— GA :: Pop: 100. Výběr pop: Turnaj. Děti 200. Výběr rodičů: Turnaj. Křížení 70% 1-bodové. Mutace 1%.

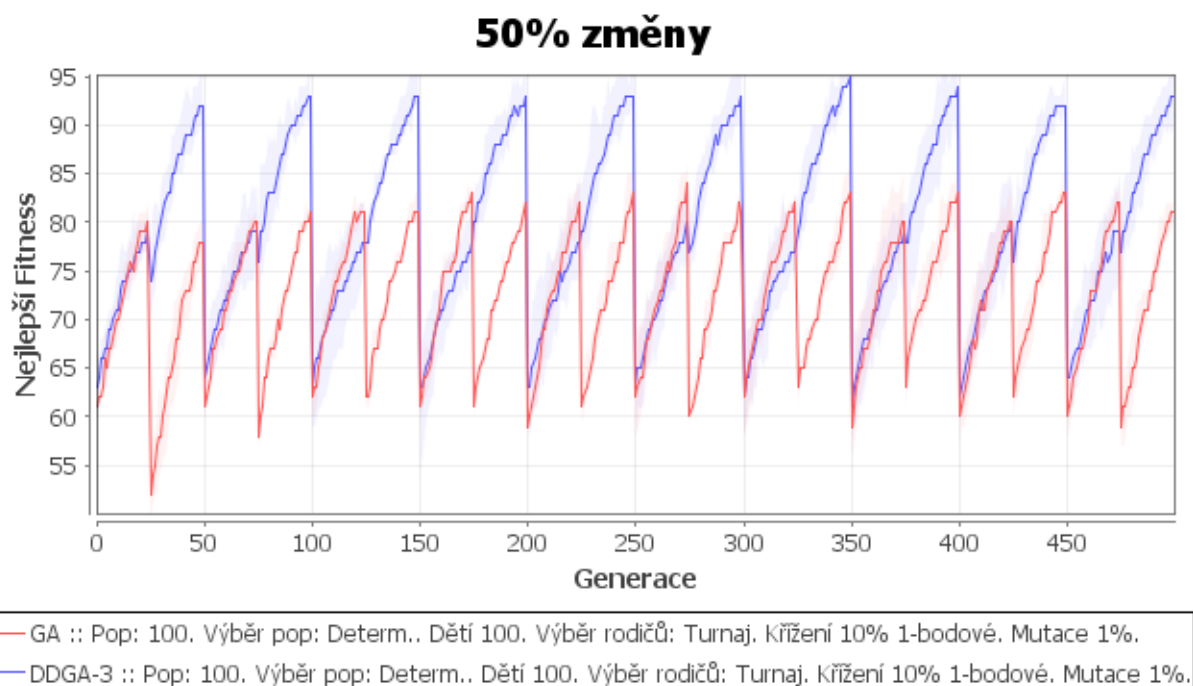
Obrázek 5: Vyrovnání DGA se změnami prostředí.

RoyalRoad problém s délkou chromozomu 100 **50% změny**
Náhodná změna prostředí každých 50 generací



— GA :: Pop: 100. Výběr pop: Turnaj. Děti 100. Výběr rodičů: Turnaj. Křížení 70% 1-bodové. Mutace 1%.
— DDGA-8 :: Pop: 100. Výběr pop: Turnaj. Děti 100. Výběr rodičů: Turnaj. Křížení 70% 1-bodové. Mutace 1%.
— DDGA-3 :: Pop: 100. Výběr pop: Turnaj. Děti 100. Výběr rodičů: Turnaj. Křížení 70% 1-bodové. Mutace 1%.
— NDGA-4 :: Pop: 100. Výběr pop: Turnaj. Děti 100. Výběr rodičů: Turnaj. Křížení 70% 1-bodové. Mutace 1%.

Obrázek 6: Nevhodně zvolené dynamické prostředí – DGA selhává.



Obrázek 7: Další vhodné použití DGA. Onemax problém s periodickou změnou prostředí 25.

4 Závěr

Cílem projektu bylo zjistit, kdy je vhodnější použití diploidního GA, než klasického GA. DGA selhávají na případech, kdy změna prostředí je náhodná (viz obr. 6) a výběr nové populace není deterministický (obr. 4, 5). Z experimentů (viz obr. 3, 7) vyplynulo, že DGA se snadněji vypořádávají se změnami než klasické GA.