

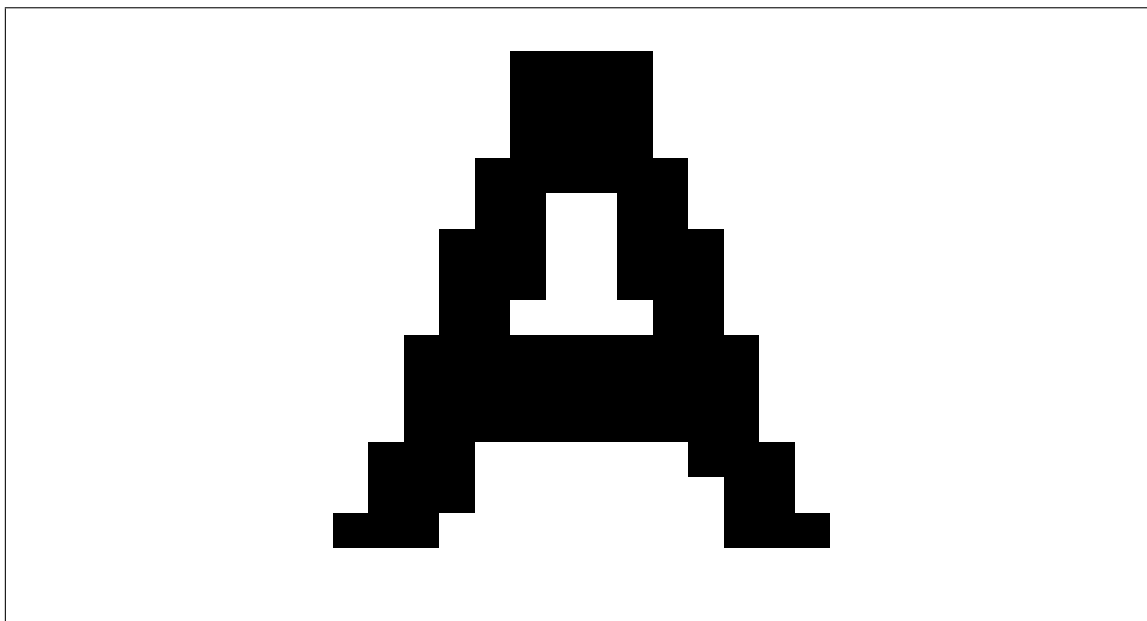
BI-ZUM - Rozeznávání znaků abecedy pomocí GA

Ondřej Perný

26. května 2019

1 Popis

Cílem této úlohy je pomocí genetického algoritmu vyšlechtit pětici bodů, které od sebe dokáží co nejlépe odlišit obrázky 16x16 pixelů v .bmp formátu. Tyto obrázky představují jednotlivé písmena anglické abecedy.



Obrázek 1: Ukázka písmene A ze vstupního datasetu (16x16 pixelů)

2 Řešení

2.1 Reprezentace

Každý jedinec je tvořen třídou "Individual". Tento jedinec je reprezentován 5 body-/souřadnicemi (x, y). Pro jednodušší práci je ale každá z těchto souřadnic reprezentována v paměti jen jedním číslem, které lze převádět na odpovídající souřadnice. Každý jedinec je při jeho vzniku kontrolován, jestli je validní, předcházím tím potřebě opravovat chybné jedince.

Hlavní třídou programu je "Population", která zastřešuje všechny jedince. Třída si pamatuje vždy jedince aktuální generace a obsahuje counter generace, aby bylo jasné o kterou generaci se jedná.

2.2 Inicializace

Inicializace jedince se provádí náhodným výběrem pěti bodů z množiny platných bodů.

2.3 Ohodnocení

Ohodnocení jedince zajišťuje fitness funkce. Čím větší hodnotu má, tím schopnější je daný jedinec. Jeden bod je připočítán za každou kombinaci obrázků, které od sebe dokáže jedinec rozpoznat bezkonfliktně. Maximální hodnota je 650.

2.4 Selektce

Selektce je prováděná klasickou "tournament selection". Tedy vyberu vždy k jedinců z kompletní populace a vyberu nejlepšího. Toto opakuji dokud nemám požadovaný počet jedinců.

2.5 Náhrada populace

Náhrada probíhá nejprve selekcí jedinců (viz. tournament) a následně křížením těchto vybraných jedinců, případně jejich mutací. Každý jedinec má danou pravděpodobnostní šanci mutovat. Jedinec mutuje ještě za předpokladu, že byl vytvořen křížením, ale stejný jedinec již v nové populaci je - nechci redundantní jedince. Mutace probíhá prohozením jednoho nebo dvou bodů. Křížení probíhá metodou "single-point crossover".

3 Experimenty

3.1 Elitismus

Řešení s elitismem, tedy zachováním nejlepších jedinců do další generace, dosahovalo podobných výsledků jako řešení bez, avšak s menším počtem generací, neboť toto řešení eliminuje klesání fitness nejlepšího jedince z populace. Je tedy použito ve finální verzi.

3.2 Fitness sharing

"Fitness sharing" je metoda pro škálování fitness jedinců, podle jejich blízkosti. Snaží se tedy bránit konvergenci jedinců do hustých oblastí, naopak podporuje jedince v řídkých oblastech. Vzdálenost byla počítána jako počet bodů, kterými se jednotliví jedinci liší. Výsledkem použití této metody bylo různorodější rozprostření jednotlivých jedinců, ale celkové fitness jedinců se nezlepšilo, naopak častá konvergence fitness k hodnotě 640 (asi velký počet různorodých jedinců, pro toto řešení) a bylo náročné najít lepší ohodnocení. Proto tato metoda nakonec používána není (v kódu implementována je, jen je zakomentované její užití).

3.3 Ostrovní model

Na začátku šlechtění si můžeme vybrat jestli použijeme jednu populaci, nebo dvě populace od sebe izolované - ostrovní model, které si občas vymění některé jedince.

3.4 Katastrofa

Katastrofa pobije většinu nejlepších jedinců, dává tedy možnost evoluce jinak slabším jedincům, kteří mají ale šanci se nakonec dostat k lepším výsledkům, např. když předchozí generace příliš konvergovala k nějakému řešení, které nebylo dostatečné.