Zpráva k semestrální práci z předmětu BI-ZUM

Andrey Babushkin
babusand@fit.cvut.cz

2. května 2016

Abstrakt
Algoritmus běží ve dvou proudech plus jeden, který se zabývá vyměnou jedinců mezi dvěma proudy. Každý jedinec má operátory inicializace, mutace, křížení, opravy DNA a hill climbingu. Inicializace má heuristiku, která odstraňuje ze grafu ty uzly, které mají právě jednoho souseda. Oprava DNA, za prvé, přidává uzly, aby všechny hrany byly pokryty, za druhé, odstraňuje ty uzly, které lze odstranit. Po opravě DNA s malou pravděpodobností uzel projde hill climbing.
Když jedinci v populaci budou příliš podobní, dojde k katastrofě - náhradě 1/10 populace novými jedinci. Podobností je kosinusová vzdálenost mezi jedinci.

Parametry algoritmu
Každý uzel je nějaké reálné číslo na intervalu [— 0.1, 0.1]. Větší 0 - uzej byl vybrán. Hodnotou fitness je záporný počet vybraných uzlů.
Testoval jsem hodně vstupních parametrů, ale nejlepší jsou 2-3% mutace a 30% křížení, v populaci je 400 - 600 jedinců.

Operátor mutace
Algoritmus mutace je velmi jednoduchý. Vybírá náhodné číslo mezi 0 a (počet uzlů / 10), počet mutací. Dále zvětšuje náhodný uzel o náhodné číslo z normálního rozdělení děleno 8 (nějaké magické číslo, výsledky jsou desetkrát lepší s 8, než např. s 9 nebo 7). Po mutaci s velmi malou pravděpodobností jedinec bude odeslán do hill climbingu.

Operátor křížení
Náhodně vybírá počet částí, dál dělí každého rodiče na vybrány počet částí a kombinuje potomky z těchto částí. Ale Bůh randomu může nastavit jeden parametr na true a uzly v potomku budou spočitány jako aritmetický průměr uzlů rodičů. Tudíž existují dvě varianty křížení.

Selekce
Čistá turnajová selekce. Velikost turnaju je (počet uzlů / počet uzlů, které musejí být vybrány).

Shrnutí
Implementoval jsem několik technik, které měli nejaké výsledky, ale horší.
1) Simulované žíhání místo hill climbingu.
2) Tabu search uvnitř hill climbingu.
3) Algoritmus diferenciální evoluce (to byl důvod, proč jsem začal používat reálná čísla). Bohužel, byly horší výsledky.
4) A hóóóódně různých realizací už popsaných algoritmů.

Přidal bych ještě nějakou podobu simulovaného žíhání do operátoru mutace, aby mutace byla tím míň, čím je lepší řešení.

Také bych změnil logiku práce s váhou uzlů. Bylo by dobře, kdybych nějak hodnotil "správnost" volby a zvětšoval bych váhu tohoto uzlu, aby pravděpodobnost odstranění ho z řešení byla méně.

Výsledky

Bohužel, algoritmus nenašel globální optimum. Minimální výsledky pro první a druhou šablonu jsou 1265 a 1503. Na obrázcích jsou jiné výsledky, trochu horší, ale neměl jsem implementovany výpis do CSV souboru, když jsem dosahl těch nejlepších. Třetí výsledek je ještě undefined, ale už mám 5798. Možná se bude ještě zlepšovat.

Osobní poděkování

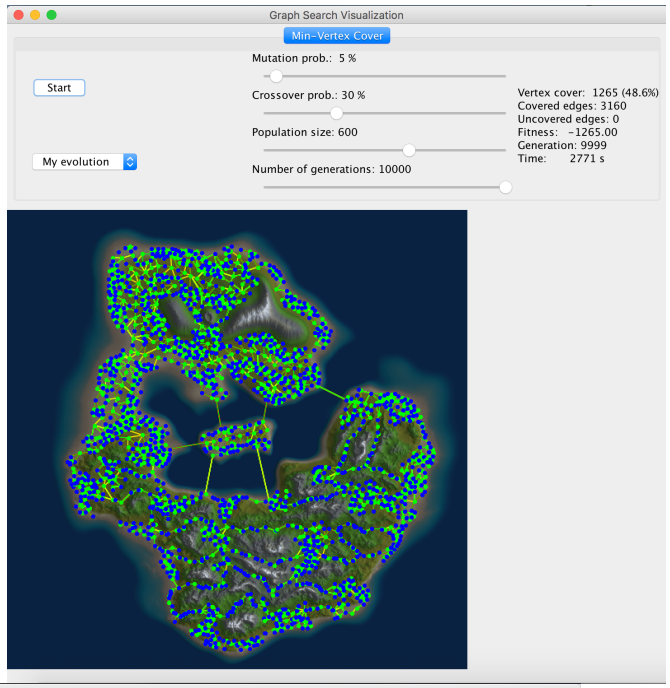
Chtěl bych poděkovat svému procesoru, který bez ustání pracoval ve dne v noci několik dní. :)

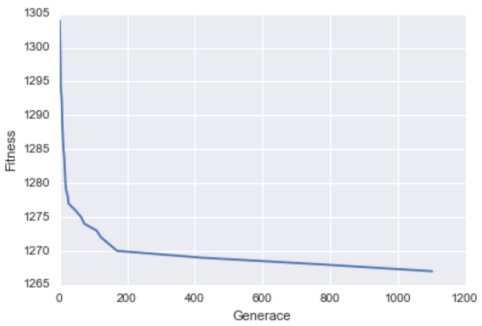
Odkazy

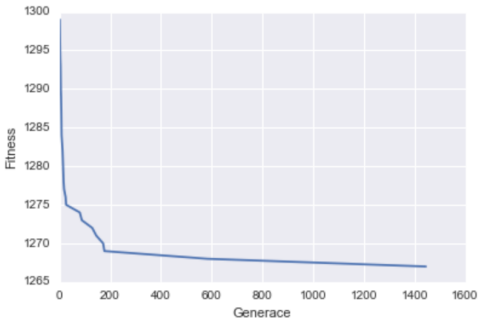
https://en.wikipedia.org/wiki/Differential_evolution
https://www.coursera.org/specializations/machine-learning-data-analysis

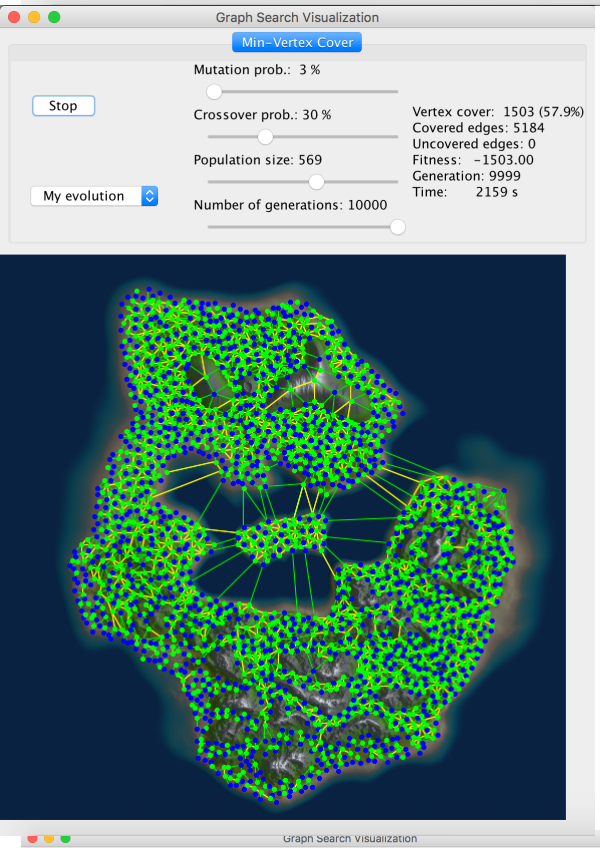
Obrázky

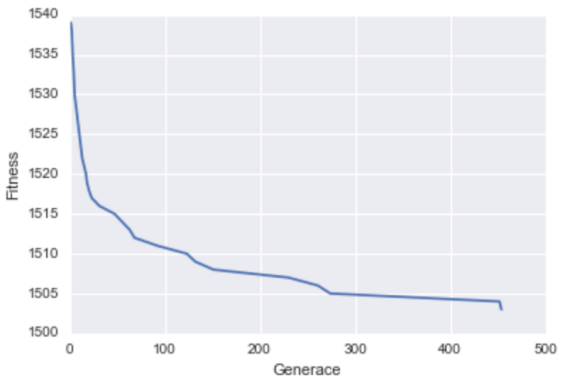
Na příští stránce jsou několik obrázků, na kterých je možno uvidět výsledky. Zleva jsou screenshoty finálních stavů GUI, zprava jsou dva grafy – závislost fitness na generaci pro oba proudy.

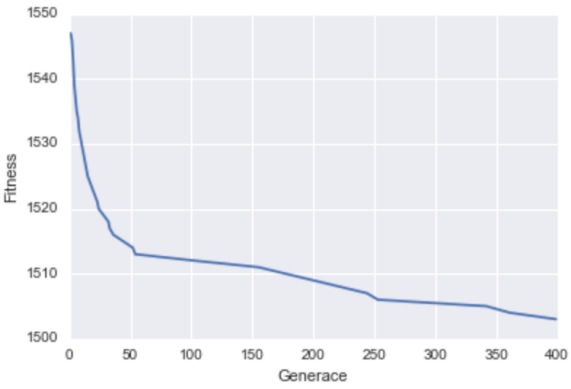

Graph Search Visualization
Min-Vertex Cover
Mutation prob.: 5 %
Start
Crossover prob.: 30 %
Population size: 600
My evolution
Number of generations: 10000
Vertex cover: 1265 (48.6%)
Covered edges: 3160
Uncovered edges: 0
Fitness: -1265.00
Generation: 9999
Time: 2771 s

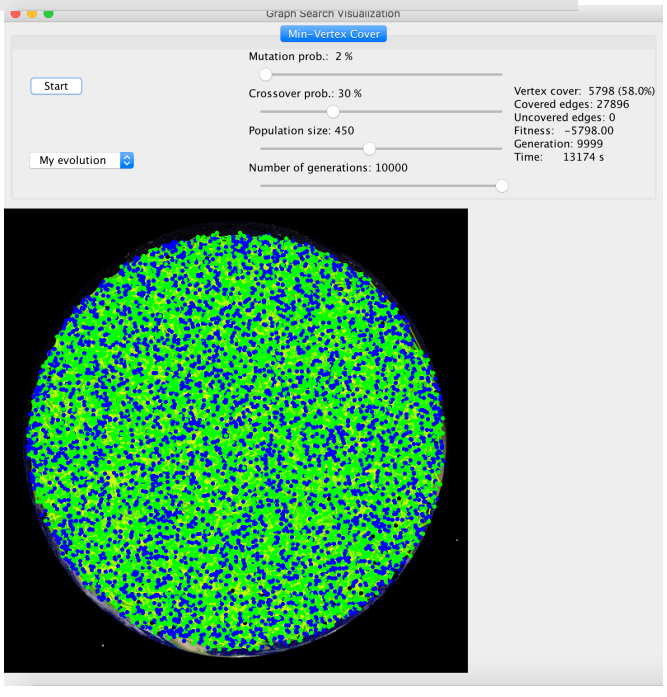

Fitness
Generace

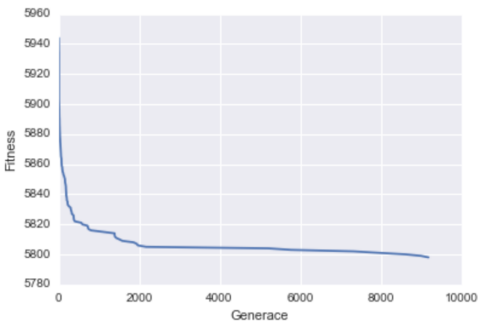

Fitness
Generace

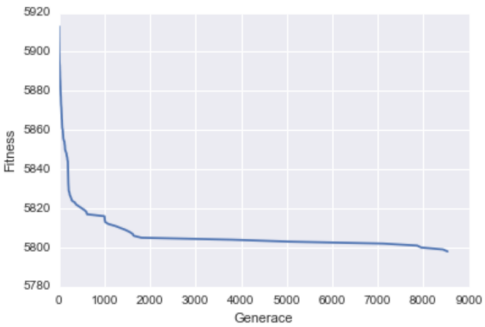

Graph Search Visualization
Min-Vertex Cover
Mutation prob.: 3 %
Stop
Crossover prob.: 30 %
Population size: 569
My evolution
Number of generations: 10000
Vertex cover: 1503 (57.9%)
Covered edges: 5184
Uncovered edges: 0
Fitness: -1503.00
Generation: 9999
Time: 2159 s


Fitness
Generace


Fitness
Generace


Graph Search Visualization
Min-Vertex Cover
Mutation prob.: 2 %
Start
Crossover prob.: 30 %
Population size: 450
My evolution
Number of generations: 10000
Vertex cover: 5798 (58.0%)
Covered edges: 27896
Uncovered edges: 0
Fitness: -5798.00
Generation: 9999
Time: 13174 s


Fitness
Generace


Fitness
Generace