

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

UMELÁ INTELIGENCIA

Akademický rok 2018/19

**Strojové učenie sa – evolučný algoritmus a evolučné
programovanie**

2019

Bratislava

riešiteľ

Daniel Minárik

rok štúdia: **druhý**

Obsah

Zadanie	3
1 Genetický algoritmus	4
1.1 Reprezentácia záhrady	4
1.2 Algoritmus „hrabania“ záhrady	4
1.3 Riešenie genetického algoritmu	5
2 Porovnanie metód a testovanie	7
2.1 Testovanie.....	7
2.2 Metódy výberu.....	8
2.3 Metódy kríženia	8
2.4 Metódy mutácie	8
3 Zhodnotenie	9

Zadanie

Zenová záhradka je plocha vysypaná hrubším pieskom (drobnými kamienkami). Obsahuje však aj nepohyblivé väčšie objekty, ako napríklad kamene, sochy, konštrukcie, samorasty. Mních má upraviť piesok v záhradke pomocou hrablí tak, že vzniknú pásy. Mních pri hrabaní nemôže samovoľne zmeniť smer. Zmena smeru je možná iba v prípade narazenia na prekážku, resp. na pohrabanú časť záhrady. V prípade, ak nie je možné pokračovať v hrabaní, t.j. nastalo uviaznutie, tak tento prípad sa považuje na nepodarený. Mních pri hrabaní musí vyjsť vždy von zo záhrady a taktiež musí platiť, že pohrabe všetky políčka záhrady.

Uvedenú úlohu riešte pomocou evolučného algoritmu. Pri riešení je potrebné použiť minimálne dve metódy výberu jedinca z populácie.

1 Genetický algoritmus

1.1 Reprezentácia záhrady

Záhrada je reprezentovaná v triede s názvom *Garden*. V tejto triede sa nachádza dvojrozmerné pole integerov, ktoré reprezentuje záhradu, resp. jej jednotlivé políčka. Pri inicializácii sa na týchto políčkach nachádza 0 a nie miestach, kde sa nachádzajú kamene sa nachádza číslo -1. Parametrami tejto triedy sú taktiež informácie o počte kameňov/piesku, výška, šírka záhrady a ďalšie údaje, ktoré slúžia na definovanie údajov, pre genetický algoritmus ako je napríklad dĺžka chromozómu. Maximálna dĺžka chromozómu zo zadania je polovička obvodu záhrady + počet kameňov. Číslo tohto chromozómu predstavujú rôzne možné začiatky vstupu do záhrady. Každá súradnica začiatku predstavuje nie len polohu x/y súradnicu v záhrade, ale taktiež aj smer pohybu, ktorý je možné z tohto políčka vykonať. Túto reprezentáciu aj na obrázku nižšie.

	0	1	2	3	4	5	6	7	8	9	10	11	
32													12
33													13
34													14
35													15
36													16
37													17
38													18
39													19
	20	21	22	23	24	25	26	27	28	29	30	31	

1.2 Algoritmus „hrabania“ záhrady

Tento algoritmus slúži hlavne na zistenie fitness funkcie, ktorá zodpovedá danému chromozómu, ktorý bol zadán ako vstupný parameter. Na základe tohto údajov sa pomocou metódy vytvorí objekt súradnica, ktorý zodpovedá vstupu do záhrady a možnom pohybe v záhrade. Následne je potrebné skontrolovať možný vstup do záhrady, pre prípad či nie je blokovaný. V prípade, ak vstup je blokovaný a je povolené nájdenie nového začiatku, tak sa vykoná tento proces nájdenia nového možného začiatku, ktorý je validný. V prípade, ak sa nepodari nájsť žiaden algoritmus končí. V opačnom prípade prebieha hrabanie.

Po vstupe do záhrady je súradnica inkrementovaná a týmto sa zabezpečuje pohyb vpred. Tento pohyb sa vykonáva, pokiaľ sa nevyjde von zo záhrady. Pri tomto procese môžu nastať tieto situácie:

1. Narazenie na prekážku, pričom je možná zmena smeru.
2. Narazenie na prekážku, pričom nie je možná zmena smeru.

V prvom prípade sa vykoná zmena smeru hrabania. Ak je táto zmena možná v oboch smeroch, tak sa zmena vykoná na základe kladnosti/zápornosti príslušného génu v chromozóme. V prípade, ak je možný len jeden z možností vykoná sa práve tento. Zmena smeru je reprezentovaná vytvorením novej súradnice s príslušnými hodnotami a smerom pohybu.

V druhom prípade, t.j. nastalo uviaznutie sa vykoná spätné mazanie miesta v záhrade, ktoré boli pohrabané v danom kroku. Toto mazanie prebieha pomocou dekrementovania súradnice, prípadne zmeny súradnice na predchádzajúcu, ak došlo k zmene pohybu, pokiaľ sa nedostane na počiatočné miesto hrabania. Následne daný gén v súradnici nahradený novým, ktorý je výsledkom metódy hľadania nového začiatku.

Skončenie cyklu hrabania môže nastať, v prípade, ak sme vyčerpali všetky gény v chromozóme alebo nie je možné vykonať žiadne hrabanie zo žiadneho počiatočného bodu. Následne po skončení tohto cyklu sa vykoná vypočítanie fitness funkcie, t.j. počet pohrabaných políčok záhrady. Výstupom tejto metódy je jedinec, ktorý obsahuje nový zmenený chromozóm a taktiež fitness funkciu.

Táto metóda hľadania nového začiatku spôsobuje to, že opätovné pohrabanie záhrady nezodpovedá pôvodnej fitness hodnote. Preto pre tento prípad je opätovne volaná táto metóda, avšak bez hľadania nových začiatkov. Týmto sa zabezpečí, aby chromozóm zodpovedal danej fitness hodnote.

1.3 Riešenie genetického algoritmu

Genetický algoritmus sa nachádza v triede s názvom *Genetic*. Algoritmus pracuje najmä s objektami *population* a *individual*. Objekt *individual* reprezentuje jednotlivca, jeho chromozóm/gény a taktiež jeho fitness hodnotu. Objekt *population* reprezentuje

populáciu, t.j. skupinu jednotlivcov. Títo jednotlivci, ktorý patria danej populácii sú uchovávaný v poli jednotlivcov.

Algoritmus začína s náhodne vygenerovanou populáciou, resp. náhodne vygenerovaným chromozómom. Následne, ak sa v tejto populácii nenachádza riešenie, tak sa pokračuje vo vytvorení novej populácie. V tomto riešení je aplikovaný takzvaný elitizmus. To znamená, že N najlepších jedincov sa priamo dostane do nasledujúcej generácie. Na druhú stranu je taktiež aplikovaný princíp, že každých x generácií sa y najhorších jedincov nahradí náhodne vygenerovanými. Ostatní jedinci sú do novej populácie pridaný krížením.

Pre možné kríženie je potrebné vybratie jedincov, ktorý sa budú krížiť. Pre výber jedincov sú vytvorené dve rôzne spôsoby. A to ruleta a turnaj. Princíp turnaja je taký, že sa vytvorí dočasná populácia vo veľkosti turnaja(počet jedincov v turnaji). Následne sa do tejto populácie pridajú náhodný jednotlivci. Výsledný jednotlivec turnaja je ten, ktorý ma najväčšiu hodnotu fitness. V metóde výberu ruletou sa snaží o to, aby najlepší jedinec mal čo najlepšiu šancu. Táto metóda je riešená pomocou vygenerovania náhodného čísla z rozsahu 0 až súčet fitness všetkých jednotlivcov. Následne sa od tejto hodnoty odčítavajú fitness hodnoty jednotlivcov. V prípade, že číslo nadobudne zápornú hodnotu, výhercom rulety je jednotlivec, ktorý sa posledný zúčastňoval na odčítaní.

Po výbere dvoch jednotlivcov je možné vykonať kríženie. V riešení sú implementované dve metódy kríženia. Prvá metóda je založená na vygenerovaní jedného náhodného bodu z rozsahu dĺžky chromozómu. Výsledný jednotlivec je tvorený prvou časťou chromozómu prvého jedinca(po náhodný bod) a druhou časťou z prvého jedinca. Druhá metóda využíva dve náhodne vygenerované body z rozsahu dĺžky chromozómu. Výsledný jednotlivec ma prvú časť z druhého jednotlivca a druhú z prvého jednotlivca.

Po vytvorení všetkých nových jedincov je vykonávaná mutácia týchto jedincov. Táto mutácia sa vykonáva len s určitou pravdepodobnosťou. Pre mutáciu sú vytvorené taktiež dve rôzne spôsoby. Prvým spôsobom je pridanie na náhodne miesta v chromozóme náhodné číslo. Druhým spôsobom mutácie je obrátenie poradia génov v určitej pod postupnosti v chromozóme.

Tento proces sa opakuje do chvíle, kedy sa nájde riešenie problému alebo nie je dosiahnutý maximálny počet generácií.

2 Porovnanie metód a testovanie

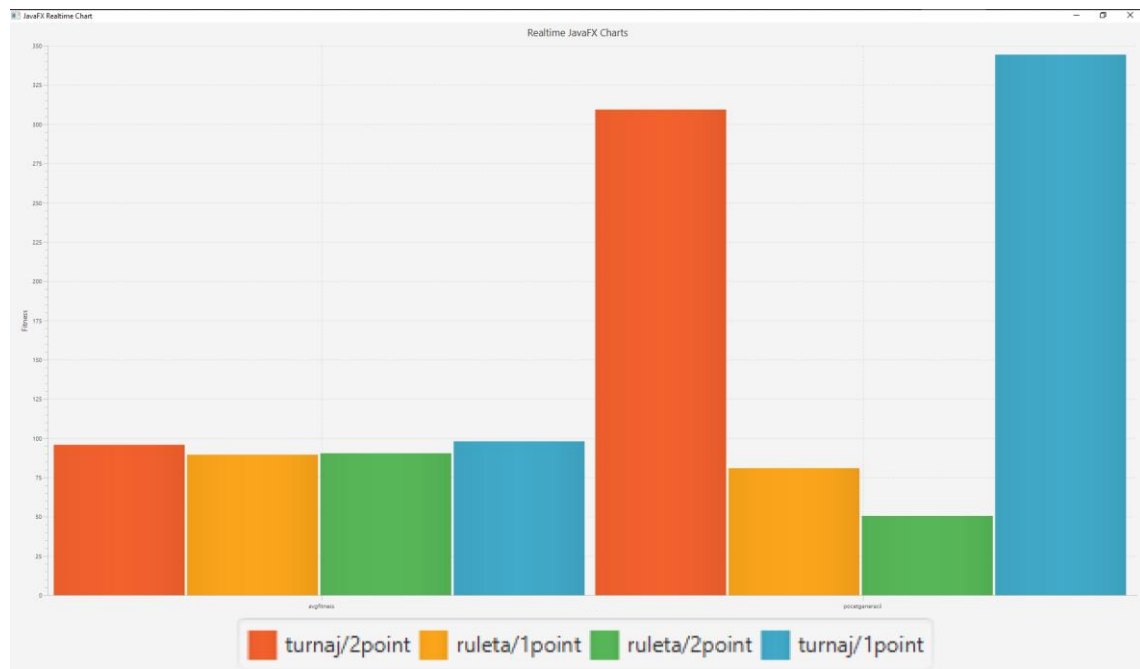
2.1 Testovanie

Na testovanie sú vytvorené 3 súbory, ktoré obsahujú v prvom riadku výšku, šírku, počet kameňov a v nasledujúcich riadkoch pozície jednotlivých kameňov.

Program je spustiteľný z triedy *Chart*, avšak názov súboru sa uvádza v triede *Runner*. Program sa spustí pre 40 opakovaní pričom v každých 10 opakovaní sa menia parametre výberu a kríženia. Výsledky sú zobrazené do konzoly a taktiež je vytvorený graf, ktorý znázorňuje priemerné výsledky týchto hľadání riešenia pomocou jednotlivých spôsobov.

V implementácii je taktiež možné vlastnoručne nastavenie hodnôt a to: veľkosť populácie, veľkosť turnaja, maximálny počet generácií, metóda kríženia, veľkosť elitizmu, metóda výberu, metóda mutácie, pravdepodobnosť mutácie.

Tento test bol vykonávaný na priemere z 10 hľadání pre každý jeden prípad dvojice metód kríženia a výberu jedincov. Výsledky tohto testovania sú analyzované v jednotlivých porovnaniach metód.



2.2 Metódy výberu

Pri testovaní riešenia výsledky ukazovali na výraznú výhodu metódy výberu ruletou. Táto skutočnosť môže byť spôsobená tým, že ruleta viac zabezpečuje, aby vybraný jednotlivec mal čo najvyššie fitness. V prípade turnaja o veľkosti 2 je možné že sa dostanú práve dvaja jedinci s veľmi nízkym fitness, pretože tento výber je náhodný.

2.3 Metódy kríženia

Výsledky testovania nenaznačovali žiadne výrazne odlišnosti týchto dvoch metód, ktoré boli implementované. Avšak priemerne sa metóda založená na dvoch bodoch vyznievala o niečo lepšie. Tento jav môže byť spôsobený tým, že pri riešení často nie je použitý celý chromozóm, teda nie je použitá jeho koncová časť, ktoré je pri krížení tvorená zväčša len z jedného predchádzajúceho jedinca. Avšak pri metóde 2-point je prostredná časť tvorená jedným jednotlivcom a zvyšok druhým zabezpečený fakt, že pri riešení budú pravdepodobné použité gény od oboch rodičov.

2.4 Metódy mutácie

Výsledky týchto pokusov nie sú zaznačené pomocou grafu, avšak z vlastného pozorovania výsledkov sa aplikovaním metódy mutácie inverziou výrazne znížil počet generácií na nájdenie výsledného riešenia. Tento fakt je podložený tým, že pri pridávaním náhodného čísla sa mohlo pridať číslo, ktoré sa v chromozómu už nachádzalo. Avšak pri metóde inverzie sa hodnoty génov nijako nemenia, len sa upravuje ich poradie.

3 Zhodnotenie

Genetické algoritmy predstavujú postup na riešenie takmer každého problému. Avšak tento postup nie je optimálny, resp. veľmi efektívny, preto je potrebné zabezpečiť, aby sme nemuseli využívať veľa technických prostriedkov a dosiahli výsledky týchto algoritmov v čo najmenšom počte generácií.

Riešenie tohto zadania, by bolo možné riešiť taktiež bez hľadania nových začiatkov pri hrabaní záhrady, avšak týmto spôsobom môžeme znížiť počet generácií, tým že až v prvej generácii dostávame veľmi vysoké fitness hodnoty. Z pozorovania sa niekedy pri vyšších číslach veľkosti populácie stávalo, že riešenie bolo nájdené už v prvej generácii.

Toto riešenie ma však stále miesta na zlepšenie hľadania. Tieto zlepšenia vidím hlavne v spôsobe hrabania záhrady, kde pri súčasnej implementácii sa pracuje s hashsetom na zaevidovanie vrcholov, z ktorých nie je možné opätovné začatie, avšak pri opätovnom sporení s upravených chromozómom nie sú výsledne fitness hodnoty totožné. Tento jav je daný tým že každý ďalší posun je daný taktiež obsahom tohto hashsetu. Ďalšou oblasťou zlepšenia by mohli byť nové metódy kríženia resp. mutácie, ktoré by zabezpečovali, aby sa v chromozóme neopakovali žiadne gény.