

Playing Regex Golf with Genetic Programming

Projekat u okviru kursa Računarska inteligencija

Anđela Ilić

mi17105@alas.matf.bg.ac.rs

Mina Milošević

mi17081@alas.matf.bg.ac.rs

Matematički fakultet
Univerzitet u Beogradu

Februar 2021

Uvod

1 Opis problema

2 Implementacija

3 Rezultati

Opis problema

Cilj *Regex Golf* igre je pronaći najkraći regularni izraz kojim se mogu zapisati sve reči iz skupa M , ali kojim se ne može zapisati nijedna reč iz skupa U .

Za date skupove M i U ne možemo sa sigurnošću da tvrdimo da postoji rešenje koje zadovoljava prethodne uslove. Takođe, ako dobijemo regularni izraz koji zadovoljava navedene uslove, ne možemo za svaki primer znati da li postoji i bolje rešenje tj. kraći regularni izraz.

Implementacija

Terminal i Function skupovi

Svaka jedinka u Genetskom programiranju će biti predstavljena kao drvo. U listovima nalaze elementi koje ćemo jednim imenom zvati *Terminali* (terminal set), a u unutrašnjim čvorovima su elementi koje nazivamo *Funkcije* (function set).

Terminal i Function skupovi

Skup funkcija sadrži operatore koji se mogu javiti u regularnim izrazima. Primeri takvih operatora su: `.*`, `.+`, `.?`, `.{.,.}+`, `(.)`, `[.]`, `[^.]`, `..`, `.|..`. Tačka `.` je mesto na kome se nalaze deca u drvetu.

Skup terminala čine elementi koji zavise i koji ne zavise od ulaznih skupova M i U . Elementi koji su nezavisni - opsezi malih i velikih slova, brojeva u regularnim izrazima, karakteri `^` i `$`, wildcard karakter `'%'`. Elementi skupa terminala koji su zavisni - skup karaktera iz M , parcijalni opsezi karaktera iz M i n -grami.

Genetsko programiranje - jedinke

Svaka jedinka se predstavlja preko apstraktnog stabla. U korenu stabla se nalazi karakter '.' i koren uvek ima dva deteta. Elementi stabla se biraju random iz skupova *Function* i *Terminal*.

Od kreiranog drveta se dobija niska koja predstavlja validan regularni izraz. Za svaku jedinku računamo i *fitnes* funkciju po formuli:

$$f(x) = w_i * (n_m - n_u) - length(r)$$

Ovako definisanu funkciju maksimizujemo.

Genetsko programiranje - selekcija

Za *selekciju* koristimo turnirsku selekciju veličine 7. Jedinke za selekciju biramo random i uzimamo najbolju jedinku tj. onu koja ima najveći fitnes među odabranim.

Genetsko programiranje - ukrštanje

Genetsko programiranje - mutacija

Genetsko programiranje - nova populacija

Genetsko programiranje - parametri

Rezultati

Reference



Bartoli, A., Medvet, E., Lorenzo, A. D. & Tarlao, F. *Playing Regex Golf with Genetic Programming*. (2014).

KRAJ