

Posudok vedúceho na diplomovú prácu Lenky Kovářovej „Testování učení restartovacích automatů genetickými algoritmy“

Predmetom predkladanej práce je učenie jazykov reprezentovaných reštartovacími automatmi s obmedzeným kontextom. Tento model automatu je založený na prepisovacích pravidlách, aplikácia ktorých, na rozdiel od obecných reštartovacích automatov, závisí len na obmedzenom okolí prepisovaného podslova. Takéto inštrukcie sú dobre interpretovateľné pre počítač a zároveň zrozumiteľné pre človeka.

Strojovým učením jazykov (tzv. gramatickou inferenciou) reprezentovaných obmedzenými reštartovacími automatmi sa zaoberal vo svojej diplomovej práci už Mgr. Stanislav Basovník. Diplomantka mala nadviazať na Basovníkove výsledky a navrhnúť systém na testovanie učenia jazykov genetickými algoritmi. Druhým cieľom bolo navrhnúť vylepšenie Basovníkovej metódy, ktoré by malo zvyšovať úspešnosť učenia a prípadne tiež skracovať dobu potrebnú na naučenie daného jazyka.

V predloženej práci sú stručne popísané genetické algoritmy, rôzne reprezentácie formálnych jazykov medzi nimi i reštartovacie automaty s obmedzeným kontextom, problém učenia jazyka na základe príkladov slov z cieľového jazyka a slov, ktoré do jazyka nepatria. Ďalej autorka popísala konkrétnu metódu učenia pomocou genetických algoritmov z pozitívnych a negatívnych vzorov a navrhla vlastné rozšírenie použitých genetických operátorov o vkladanie nových prepisovacích pravidiel do jedincov v aktuálnej populácii.

Jej vylepšenie spočíva v tom, že vždy po niekoľkých generáciách sú do náhodne zvolených jedincov aktuálnej populácie vložené pravidlá generované síce náhodne, ale podľa vzorov, ktoré boli užívateľom dopredu zadané spolu s pravdepodobnosťami ich použitia. Takto generované pravidlá by mohli vzniknúť tiež pomocou mutácií jedincov, ale mohlo by to trvať veľmi dlho.

Autorka navrhla systém na testovanie učenia jazykov, ktorý dovoľuje prehľadne (v tabuľkovom procesore) popísať parametre experimentov, ktoré sa majú vykonať. Výsledky experimentov potom spracuje skript v systéme R a prezentuje výsledky testovania formou niekoľkých typov grafov.

Práca je napísaná prehľadne a stručne. Výsledky experimentov ukazujú, že navrhnuté rozšírenie skutočne pomáha. Samotné genetické učenie má niekoľko parametrov a ich počet s pridaným rozšírením narastá, preto autorka do práce zaradila iba časť vykonaných experimentov a k práci je priložený program, ktorý umožňuje tieto experimenty exaktne zopakovať a ľahko modifikovať.

K práci mám nasledujúce výhrady:

- Rozšírenie implementované v práci umožňuje špecifikovať aj zložitejšie vzory s opakovaním podslov. Ich zaradenie nie je v práci zdôvodnené a ani nie je použité v priložených testoch.
- Text práce je miestami nevyvážený. Napr. autorka dokazuje vzorec na sčítanie geometrickej postupnosti, ale zložitejšie problémy napr. s konvergenciu genetického algoritmu už nespomína.
- V texte sú občas zbytočné preklepy (napr. str. 14, r. 8: „je najprve uvedena motivaci restartovacích“, str. 22, r. -5: „mohl označit všechny slova“, str. 28, r. 16: „Tento řetězec je znakem | rozdělen na čtyři části“, ale v takom reťazci je sú 3 znaky |, atď).

Celkovo však konštatujem, že práca Bc. Lenky Kovářovej splnila všetky ciele zadania, a preto doporučujem, aby bola uznaná ako diplomová práca.

Praha, 23.1.2012

RNDr. František Mráz, CSc.

KSVI MFF UK