Vyjadrenie vedúceho na dizertačnú prácu Petra Hoffmanna "Machine Learning of Analysis by Reduction"

Redukčná analýza je metóda kontroly syntaktickej správnosti vety jej postupným zjednodušovaním. Zjednodušovanie vstupnej vety sa robí postupnými lokálnymi úpravami, ktoré zachovávajú prípadné chyby. Zjednodušovaním sa veta skracuje až na jednoduchú (napr. holú) vetu, o správnosti ktorej sa dá ľahko rozhodnúť. Formálnym modelom na implementáciu redukčnej analýzy je reštartovací automat. Reštartovací automat reprezentuje jednak samotný jazyk ako množinu správnych viet, ale tiež reláciu redukcie popisujúcu všetky možné zjednodušovania viet. Reštartovací automat je v práci reprezentovaný ako zoznam tzv. prepisovacích meta-inštrukcií, ktoré popisujú množinu všetkých možných redukcií, a zoznam prijímacích meta-inštrukcií, ktoré popisujú slová prijímané automatom priamo bez redukovania. Obecne je vstupné slovo daným reštartovacím automatom prijímané, ak existuje konečná postupnosť zjednodušení pomocou prepisovacích meta-inštrukcií, ktorá prepíše slovo na kratšie slovo prijímané niektorou prijímacou meta-inštrukciou.

Cieľom dizertačnej práce Petra Hoffmanna bolo navrhnúť algoritmus na učenie redukčnej analýzy z príkladov slov a redukcií. Slová a symboly vo formálnom jazyku zodpovedajú vetám a slovám v prirodzenom jazyku. Vstupom je teda

1. množina slov, ktoré patria do jazyka a nazývajú sa pozitívne príklady slov,
2. množina slov, ktoré do jazyka nepatria a nazývajú sa negatívne príklady slov,
3. množina dvojíc slov tvaru (u, v), kde slovo v môže vzniknúť redukciou slova u; táto množina sa nazýva množina pozitívnych príkladov redukcií a
4. množina dvojíc slov tvaru (u, v), kde slovo v nesmie vzniknúť redukciou slova u; táto množina sa nazýva množina negatívnych príkladov redukcií.

Výsledkom učenia je vhodný model reštartovacieho automatu, ktorý by mal prijímať všetky pozitívne príklady slov a vykonávať všetky pozitívne príklady redukcií, zamietať negatívne príklady slov a nerobiť negatívne príklady redukcii. Zadanie práce obsahovalo okrem návrhu algoritmu učenia aj skúmanie zložitosti problému učenia redukčnej analýzy a aj otestovanie navrhnutého algoritmu.

Doktorand vo svojej práci riešený problém formálne popísal a postupne sa venoval zložitosti problému učenia redukčnej analýzy, návrhu vlastného algoritmu učenia a nakoniec testovaniu navrhnutého algoritmov.

Jedným z hlavných výsledkov práce je, že problém inferencie reštartovacieho automatu z pozitívnych a negatívnych príkladov slov a redukcií je NP-ťažký, keď chceme minimalizovať buď počet meta-inštrukcií, alebo celkovú dĺžku zápisu automatu. Na druhú stranu inferencia čo najmenšieho reštartovacieho automatu iba z pozitívnych príkladov slov a redukcií je možná v polynomiálnom čase, ale výsledný automat prijíma príliš veľký jazyk.

Pretože pre NP-ťažké úlohy nie sú známe žiadne efektívne algoritmy, autor v práci navrhuje heuristický algoritmus učenia redukčnej analýzy, ktorý dokáže k zadanej množine pozitívnych príkladov slov a redukcií nájsť konzistentný reštartovací automat špeciálneho typu, tzv. single k-reverzibilný reštartovací automat (skrátene S-kR-RRWW-automat). Obmedzenie S-kR-RRWW-automatu oproti obecnému RRWW-automatu spočíva v tom, že jeden typ prepisu popisuje práve jedna meta-inštrukcia, podmienky na kontexty v meta-inštrukciách sú obmedzené na

k-reverzibilné jazyky a automat má práve jednu prijímaciu meta-inštrukciu, tiež obmedzenú na k-reverzibilný jazyk. Aj keď je S-kR-RRWW-automat oproti obecným reštartovacím automatom obmedzený, tak je v práci dokázané, že trieda jazykov je dosť bohatá. Ak S-kR-RRWW-automaty môžu používať pomocné symboly, tak trieda nimi rozpoznávaných jazykov zahrnuje triedu rastúcich kontextových jazykov. Tiež je ukázaná nekonečná rastúca hierarchia tried rozpoznávaných jazykov vzhľadom k rastúcemu stupňu reverzibilnosti k.

"Ručne" zostavené testy pre jazyky používané inými autormi na testovanie učenia jazykov ukázali, že navrhnutá metóda funguje výborne. Autor však chcel zistiť, aké výsledky dáva jeho metóda na širšej triede jazykov. Kvôli tomu, že existujúce benchmarkové testy sa ukázali byť pre testovanie učenia redukčnej analýzy nevhodné, tak Petr Hoffmann navrhol a implementoval vlastné testy. Pomocou rozsiahlych testov ukázal, že jeho učiaci algoritmus funguje s vysokou úspešnosťou pri učení redukčnej analýzy podľa náhodne generovaných reštartovacích automatov. I keď tieto testy sa obmedzili na porovnávanie cieľového jazyka a naučeného jazyka. Vlastné relácie redukcie naučeného automatu a cieľového automatu autor neporovnával.

Pri riešení hlavných problémov práce doktorand dosiahol zaujímavé výsledky i pre mnoho ďalších vedľajších problémov. Špeciálne by som vyzdvihol analýzu prístupov k testovaniu učenia jazykov a ako príklad analýzu rozhodnuteľnosti niektorých problémov pre reštartovacie automaty na EL-jazykoch. EL-jazyky sú jazyky generované tzv. párnymi lineárnymi gramatikami. Aj keď sa EL-jazyky spočiatku ukazovali ako vhodná trieda jazykov na testovanie učenia jazykov, práca dokazuje, že aj pri obmedzení na EL-jazyky niektoré základné problémy ako napr. overenie vlastnosti zachovania chyby pri redukciách sú algoritmicky nerozhodnuteľné.

Celkovo ciele práce boli splnené. Autorovi podarilo preskúmať zložitosť zadaného problému učenia redukčnej analýzy, navrhnúť vlastný algoritmus učenia pre vhodne zvolenú novú triedu reštartovacích automatov, preskúmať silu tejto novej triedy automatov a ukázať na testoch s vybranými jazykmi i jazykmi rozpoznávanými náhodne generovanými automatmi, že vyvinutý algoritmus má vysokú úspešnosť. Doktorand ukázal, že sa dokáže systematicky venovať danému problému, navrhnúť vlastné riešenia a ukázať ich vzťah k známym výsledkom. Časť výsledkov jeho práce už bola publikovaná vo forme konferenčných článkov. Jeho práca môže v budúcnosti prispieť k riešeniu zložitých problémov v učení formálnych ale i prirodzených jazykov. Preto doporučujem, aby práca Petra Hoffmanna bola uznaná ako dizertačná práca.

V Prahe, 23.5.2013

RNDr. František Mráz, CSc.
KSVI MFF UK