

Posudek oponenta diplomové práce

Autor: Jakub Šmíd
Název práce: Agent optimization by means of genetic programming
Vedoucí: Roman Neruda
Oponent: Ondřej Kazík

Autorova diplomová práce se zabývá vývojem programů pro odhad chybovosti a časových nároků konkrétních metod strojového učení na neznámých datech na základě předchozích zkušeností. Programy jsou evolvovány s pomocí genetického programování a mají sloužit pro výběr optimální metody pro neznámá data – úlohy nazývané metaučení. Metoda odhadu je pak implementována jako agent v multiagentním prostředí JADE a otestována na konkrétních problémech data-miningu.

Práce začíná úvodem do problematiky data-miningu a obecným představením cílů práce. V druhé kapitole jsou popsány problémové okruhy, přesněji stanoveny cíle a popsány stávající metody metaučení. Další kapitola obsahuje popis použité metody evolučních algoritmů – genetického programování. Kapitola čtyři ukazuje návrh užitých algoritmů a agenta provádějícího odhad výkonu metod data-miningu. Pátá kapitola popisuje implementační podrobnosti. Šestá kapitola uvádí popis provedených experimentů a jejich výsledku. V závěru je shrnutí celé práce a v poslední kapitole nalezneme možnosti jejího dalšího rozšíření.

Jako hlavní přínos práce vidím novou metriku určující podobnost metadat (informace o struktuře vstupních dat) stejně jako využití genetického programování při návrhu metod metaučení – tedy v současnosti studované oblasti pro podporu uživatelských experimentů nad předtím neznámými daty. Výsledky prokázaly slibné chování výsledných metod. Zároveň multiagentní přístup usnadňuje zařazení metody do širší aplikace data-miningu v multiagentních systémech.

Bylo by však vhodné více využít multi-agentní prostředí – zejména pro implementaci genetického programování. V stávající verzi je totiž agentní technologie využita pouze pro agenta odhadujícího výsledky výpočetních metod. Zde pak chybí popis protokolů, které určují konkrétní běh agenta, a skrze které dochází k výměně informací s jeho okolím. Dále by bylo zajímavé uvést pro výsledky v cílové aplikaci metaučení (například jak často odhadne nejlepší metodu pro daná data, nebo jaký je rozdíl chyby metody vybrané na základě algoritmu a nejlepšího otestovaného výsledku). Proto je vyhodnocení úspěšnosti výsledků spíše vágní.

Vzhledem k aplikačnímu významu, zajímavosti metody i úspěšnosti výsledků by bylo vhodné rozpracování diplomové práce pro další vědecký výzkum a publikaci výsledků.

Práci doporučuji k obhajobě.

V Praze, dne 26. 8. 2012, Ondřej Kazík