

OPONENTSKÝ POSUDEK DISERTAČNÍ PRÁCE MGR. ONDŘEJE VENCÁLKA: WEIGHTED DATA DEPTH AND DEPTH BASED DISCRIMINATION

Autor se ve své práci zabývá problematikou jednoho z důležitých neparametrických přístupů k práci s mnohorozměrnými daty, který je založen na tzv. hloubce dat. Jde o téma, jemuž je v literatuře věnována větší pozornost až v průběhu posledních zhruba 10 let, a je to tudíž problematika vysoce aktuální.

Práce je rozčleněna do tří obsáhlých kapitol. Úvodní kapitola podává přehled problematiky s důrazem na různé funkce využívané pro měření hloubky dat. Kapitola obsahuje i nástin praktických problémů, které by bylo možno pomocí uvedené metodiky řešit. Druhá kapitola se věnuje vážené hloubce dat a jejím vlastnostem. Vychází ze zobecnění poloprostorové hloubky bodu na tzv. váženou poloprostorovou hloubku, která vykazuje některé lepší vlastnosti, zejména nulovost hloubky u bodů, které jsou v konvexním obalu nosiče rozdělení, ale mimo nosič samotný. Původní definice poloprostorové hloubky je založena na využití pravděpodobnosti poloprotstoru, zatímco modifikace navržená autorem pracuje s váhovou funkcí intergovanou na tomto poloprostoru dle pravděpodobnostní míry. Vážená hloubka může být uplatněna i u nekonvexních oblastí. Jsou probrány různé volby váhových funkcí a je studována problematika konzistence. Podkladem pro druhou kapitolu jsou vlastní výsledky, které autor se spolupracovníky publikoval v impaktovaném časopise. Třetí kapitola se zabývá úlohou diskriminace z pohledu řešení prostřednictvím hloubky dat. Po úvodním přehledu přináší vlastní autorovy výsledky, např. o modifikované metodě k nejbližších sousedů, jejíž teoretické vlastnosti jsou dokazovány v oddíle 3.4.3. Klasifikátor definovaný tímto způsobem má např. výrazně lepší chování než dříve zavedené klasifikátory v případě, kdy se uvažovaná rozložení liší variabilitou. V závěru práce jsou prezentovány výsledky simulační studie, která poskytuje velmi užitečný pohled na porovnání různých klasifikátorů.

Práce má jasnou a promyšlenou koncepci, s ohledem na složitost problematiky je psána přehledně, je dobře strukturována a po výkladové i formální stránce velmi pečlivě zpracována. Výklad je logický, doprovázený užitečnými příklady (např. oddíl 2.7) a názornými obrázky. Formálních nepřesností je minimálně (např. desetinné čárky místo teček na str. 20 dole).

V práci jsou zmíněny praktické aplikace, nicméně výpočty ukazují, že výsledky mohou být dosti odlišné v závislosti na volbě funkce hloubky dat, váhové funkce, resp. klasifikátoru v kapitole 3. Lze určit nějaká obecnější pravidla pro volbu v závislosti na povaze úlohy?

Autor na komplikovaném tématu naprosto jednoznačně prokázal schopnost samostatné tvořivé práce a aplikace znalostí z různých oblastí matematické statistiky a pravděpodobnosti. Zpracoval přehlednou rešerši, odvodil nové teoretické výsledky, věnoval pozornost výpočetním aspektům studovaného problému a simulační studií prokázal schopnost netriviálního použití statistického software. Cílů práce bylo určitě dosaženo a její výsledky nepochybně přispějí k rozvoji poznání v dané oblasti. Jádro třetí kapitoly by určitě bylo vhodné publikovat.

Souhrnně lze říci, že předložený text zcela jistě splňuje požadavky kladené na práce tohoto druhu. Doporučuji proto, aby na základě této disertační práce byl Mgr. O. Vencálkovi udělen titul PhD.

Praha, 3.11.2011

Marek Malý