

## Rozhodovací stromy pro klasifikaci dat

Autor se ve své práci zabývá problematikou klasifikace dat pomocí rozhodovacích stromů. Cílem práce bylo nastudovat, implementovat a porovnat různé modely rozhodovacích stromů s ohledem na různá ukončovací kritéria, kritéria na dělení uzlu a velikost vzniklých stromů. Součástí práce mělo být i zhodnocení vlastních výsledků a zkušeností.

V úvodu práce autor nastiňuje proces klasifikace dat a stanovuje si cíle práce. V druhé kapitole jsou zavedeny základní pojmy. V kapitolách 3 a 4 jsou popsány jednotlivé modely rozhodovacích stromů, na které se student zaměřil. V páté kapitole jsou prezentovány získané experimentální výsledky. Kapitola také obsahuje srovnání a zhodnocení získaných výsledků. V této kapitole jsou také prezentována vlastní tvrzení (i s důkazy), které se týkají vlastností studovaných modelů. V závěru práce autor shrnuje získané výsledky a navrhuje možné směry dalšího výzkumu. K práci je také přiloženo CD s implementovanou aplikací, kterou autor použil pro získání výsledků experimentů.

Za velké klady této práce považuji zejména:

1. Student provedl velké množství experimentů s ohledem na různé metody budování stromů a ořezávací kritéria. K otestování modelů použil autor vlastní implementaci algoritmů.
2. Získané výsledky experimentů autor přehledně a srozumitelně zpracoval a analyzoval.
3. Autor v experimentech ověřoval nepodložená tvrzení autorů SDT stromu ohledně unimodality chybové funkce.
4. Autor navrhl dva vlastní algoritmy pro hledání optimálního dělení.
5. Studované SDT stromy používají k optimalizaci parametrů stromu Levengberg-Marquardtovu optimalizační metodu, která pracuje s Hessianou maticí. Autorovi se podařilo dokázat, že tato metoda není použitelná obecně na všechny SDT stromy díky singularitě Hessovské matice.

Vlastní práce je napsaná srozumitelně a má velmi dobrou grafickou úpravu s velkým množstvím obrázků, tabulek a grafů. Práce obsahuje minimálně překlepů a gramatických chyb (např. několik chybějících čárek na str. 9, 22, 36). Z formálního hlediska práce obsahuje jen drobné chyby - např. nekorektní formulace („...T je matice příznaků, kde T je trénovací množina ...“ na str. 12), nekonzistentní vztah ( $F_{out}(u) \subseteq G|_v$ , avšak  $F_{out}(u)$  je množina vrcholů a  $G|_v$  je graf na str. 14) nebo chyba ve vzorci ( $F_{out}(v)$  místo  $F_{out}(v_1)$  na str. 26). Na straně 95 došlo k částečnému poničení popisků v obrázku 5.7. Nicméně tyto nedostatky lze poměrně snadno odstranit.

Domnívám se, že předkládaná práce splňuje původní cíl. Student pronikl dostatečně hluboko do dané problematiky a prokázal schopnost samostatně řešit a analyzovat dané úkoly. Práce Jakuba Čechmánka splňuje nároky kladené na práci diplomovou, a proto ji doporučuji k obhajobě.

V Praze 19. 5. 2008

RNDr. Jana Štanclová, Ph.D.

