

**POSUDEK DISERTAČNÍ PRÁCE MGR. MARTINA BETINCE
SHODNOST A SPOLEHLIVOST KLASIFIKÁTORŮ: VYBRANÉ POSTUPY
POMOCÍ ROC KŘIVEK A BOOTSTRAPU**

Autor se ve své práci zabývá problematikou klasifikátorů jejich shody a spolehlivosti. Jak zmiňuje na začátku práce, tato problematika je velmi široká a proto se zaměřuje na dvě konkrétní oblasti a to především na dvoutrídni klasifikaci s učitelem (přesněji na problematiku ROC křivek) ale i na hierarchické metody shlukové analýzy (hlavně na spolehlivost získaných evolučních stromů). Obě témata mají v praxi široké uplatnění. Autor ve své práci názorně seznamuje s problematikou a provádí rozsáhlé simulace hodnotící kvalitu klasifikačních postupů.

Práce je rozdělena do pěti kapitol. Úvodní kapitola seznamuje se základy klasifikace. Druhá kapitola se zaměřuje na vlastní popis problematiky ROC křivek, třetí a čtvrtá kapitola obsahuje hlavní výsledky práce v oblasti ROC křivek, zaměřuje se na návrh testů pro porovnání různých klasifikátorů s využitím parametrického přístupu. Tento parametrický přístup je v praxi často neprávem opomíjen a proto má práce pozitivní význam na zviditelnění parametrického přístupu. Třetí kapitola se zabývá popisem ROC křivek pro exponenciální rozložení klasifikátoru v porovnávaných skupinách. Čtvrtá kapitola je zaměřena na normální rozložení, tato kapitola obsahuje podstatné výsledky celé práce, na poli porovnání ROC křivek a provádí simulace ověřující vlastnosti křivek, jejich parametrů a navržených testů. S využitím simulací se podrobně zabývá i studiem síly testu a otázkou rozsahu dat. Simulace jsou tak rozsáhlé, že autor zvolil pro tištěnou část disertační práce jen ukázky a ostatní výpočty prezentuje na příloženém CD. Pátá kapitola se zabývá hierarchickými metodami shlukové analýzy včetně podrobné bootstrapové studie kvality získaných výsledků. Pro ilustraci reálného použití zvolil autor data z oblastí sociologie a pro pátou kapitolu z oblastí parazitologie.

Práce má jasnou a promyšlenou koncepci s cílem nejen hodnotit, ale i testovat kvalitu v praxi používaných klasifikátorů. Drobným nedostatkem v struktuře práce je skutečnost, že obsahuje dvě poměrně vzdálené problematiky - hlavní cíl práce je hodnocení ROC křivek a poslední kapitola o evolučních stromech pak řeší poněkud jinou problematiku. V závěru by bylo užitečné shrnout doporučení pro použití klasifikátorů, zejména ROC křivek v praxi.

Drobnou interpretační nepřesností z pohledu praxe je pohled na sociologická data jako na data, kde nejsme schopni se ptát na všechny detailní informace (viz. str. 21) v reálné situaci je ale často uvažována náhodnost odpovědi respondenta. V práci lze zaznamenat několik velmi drobných formálních nepřesností, které ale jsou vzhledem k obsahu práce nepodstatné (např. způsob stanovení minimálních bezpečných rozsahů v tab. 4.7 a spolehlivost této hranice).

Autor prokázal schopnost samostatné tvořivé práce a aplikace znalostí z různých oblastí matematické statistiky. Provedl množství technicky komplikovaných výpočtů a odvození. Bylo tak dosaženo cílů disertační práce, její výsledky přispívají k rozvoji problematiky klasifikace v oblasti ROC křivek i hodnocení spolehlivosti hierarchických metod shlukové analýzy.

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

Praha, 23.7.2007

