

Oponentní posudek

disertační práce Martina Betince s názvem

Shodnost a spolehlivost klasifikátorů: Vybrané přístupy pomocí ROC křivek a bootstrapu.

Předkládaná práce se podrobně zabývá klasifikací, zejména její spolehlivostí. V kapitolách 2 až 4 jsou důkladně rozebírány vlastnosti ROC křivek a uvedeny testy pro jejich porovnávání (testy shody). V kapitole 4 (str. 60 a další) odvodil autor nový parametrický test pro shodu dvou ROC křivek pro situaci, kdy dva výběry nejsou nezávislé, což je v aplikacích velmi častá a důležitá oblast. Tento test považuji za nejdůležitější teoretický výsledek disertační práce. Kapitole 5 se zabývá evolučními stromy a je v ní navržena a realizována metoda odhadu spolehlivosti evolučních stromů, založená na bootstrapu. Ve stručném závěru jsou shrnuty dosažené výsledky a naznačeny další otevřené problémy. Dále tištěná verze práce obsahuje tři přílohy, ve kterých jsou postupně uvedeny vybrané vlastnosti matic, vlastnosti rozdělení užívaných v disertační práci a přehled důležitých výsledků teorie maximální věrohodnosti. Čtvrtá příloha popisuje obsah příloženého CD obsahujícího mimo jiné i skripty procedur v R pro testování ROC křivek a zdrojový text funkce v Matlabu pro odhad hladin spolehlivosti evolučních stromů.

Celá disertační práce je zpracována pečlivě, typografie je promyšlená, nadpisy tabulek jsou však oproti zvyklostem uváděny pod tabulkami místo nad tabulkami a číslování stránek je v některých částech práce dole uprostřed, jinde vpravo nahoře. Práce je napsána kultivovanou češtinou, vzhledem k místu vzniku občas až s usilovně přesným vyjadřováním, které kupodivu ztěžuje srozumitelnost. Počet překlepů a drobných chyb v interpunkci nepřesahuje míru nyní obvyklou v pracích podobného rozsahu. Z těchto nedopatření za zmínku stojí jen

- matoucí „uprostřed“ v popisu obrázku 2.7 na str. 29,
- podivný zápis matice $\mathbf{A}_{i,j} = \dots = \{a_{i,j}\}$ v odst. A.1 na str. 121,
- dva úsměvné překlepy (extrémistické „*politické*“ hnutí na str. 9 a „*návodný*“ výběr na str. 31 dole).

K některým částem práce mám následující výhrady nebo dotazy:

- Metoda podpůrných vektorů není tak běžná jako ostatní metody klasifikace porovnávané v disertační práci. Popis této metody v jednom odstavci na str. 7 je velmi kusý a ne zcela jasný. Navrhuji podrobněji vysvětlit při obhajobě.
- Na str. 12 se uvádí, že *TPR* a *FPR* jsou definovány jako funkce přežití. Pokud se nemýlím, tak funkce přežití předpokládá nezáporný argument (doba přežití). Tomu však neodpovídá druhá část vztahu na posledním řádku na str. 12.
- Co znamená „nominální velikost testu“ na str. 58 uprostřed?
- Výsledky simulací na str. 72 až 92 jsou rozsáhlé, ale prezentovány jsou velmi nepřehledně. Lze tyto výsledky nějak strukturovaně shrnout (podrobněji než v Závěru), aby vypovídaly jasně a srozumitelně?
- Smysl symbolu h v (5.9) není jasný, patrně jen zbytečná (a nesprávná?) formální komplikace zápisu.

Pro širší aplikaci výsledků disertační práce by bylo užitečné napsat podrobnější návod pro testování shody ROC křivek s doporučeními pro aplikace (srozumitelný i uživatelům-nestatistikům), s instrukcemi pro spouštění skriptů v R a vše zpřístupnit na internetu, případně i publikovat v aplikačně orientovaném časopisu.

Závěrem konstatuji, že téma práce bylo zvoleno vhodně a cíle disertační práce formulované v úvodu byly splněny. Autor osvědčil schopnost samostatné vědecké práce v této oblasti.

Disertační práci doporučuji k obhajobě a rovněž v případě úspěšné obhajoby doporučuji, aby v souladu s paragrafem 47 Zákona o vysokých školách č. 111/98 Sb. byl Mgr. Martinu Betincovi udělen akademický titul doktor (Ph.D.)



V Ostravě 13. srpna 2007

Doc. Ing. Josef Tvrđík, CSc.