



Posudek disertační práce

Uchazeč Ing. Gabriel GyöriNázev disertační práce LMS v geodetických aplikaciach vyrovnávacieho počtuStudijní obor Geodézie a kartografieŠkolitel Prof. Ing. Aleš Čepek, CSc.Oponent Doc. Ing. Tomáš Bayer, Ph.D.e-mail tomas.bayer@natur.cuni.cz

Aktuálnost tématu disertační práce

Metody robustní statistiky hrají stále významnější roli při zpracování geodetických dat - poskytují uspokojivé výsledky i v případech, kdy vstupní data nesplňují předpoklady modelu (např. normálního rozdělení). Metoda LMS je jedním z robustních odhadů, ve statistické praxi však nepříliš často používaným, a to z důvodu neexistence řešení v algebraickém tvaru (jedná se o NP těžký problém). Vzhledem k vysokému bodu selhání ($\varepsilon = 0.5$) by LMS mohla nalézt praktické použití při zpracování dat s vysokou mírou chybovosti. V dlouhodobém horizontu sice existuje poptávka po metodě řešení LMS vhodné pro i středně velké množiny, alternativně však lze použít jiné robustní metody s podobnými charakteristikami.

vynikající nadprůměrný průměrný podprůměrný slabý

Splnění cílů disertační práce

Cílem práce mělo být navržení nové metody řešení problému LMS, tento bod se však podařilo splnit pouze částečně. Řešení publikované v kap. 13.2 je vázáno na geodetické sítě a je otázkou, zda je v obecnější formě použitelné i pro jiný typ dat. Chybí upřesnění, jaké předpoklady jsou kladeny na vstupní data (nezávislost, normalita, velikost, ...), či požadavek přesnosti výsledného odhadu.

vynikající nadprůměrný průměrný podprůměrný slabý

Metody a postupy řešení

Řešení je založeno na heuristickém přístupu, při kterém jsou z množiny vstupních měření prováděny opakované náhodné výběry vzorků s cílem nalézt podmnožinu optimálních vlastností vzhledem k LMS. Prakticky je tento krok realizován iterativním převažováním (greedy strategie, podobně jako u LTS) s průběžným generováním odhadů kvality měření na základě diskrétního rozdělení pravděpodobností; "podezřelá" měření s nižší pravděpodobností použití mají menší váhu. Autor navrhuje 3 různé strategie volby vah, což považuji za zajímavé a přínosné.

Vzhledem k nedeterministické povaze metody předpokládáme, že bude poskytovat přibližné řešení v polynomiálním čase jako ostatní publikované metody, což navržený postup ne zcela splňuje (exponenciální čas).

Zatímco text práce je poměrně obsáhlý (> 200 stran), obsahuje řadu odvození více či méně souvisejících s prací, navrženou metodu řešení nalezneme až v kapitole 13.2 popsánu nezvykle stručně, a to na 4 stranách textu bez uvedení informace o tom, že se jedná o klíčovou část práce. Popis metody působí díky střídání textového a matematického zápisu poněkud nepřehledně a není jasné, zda by se ho případným čtenářům podařilo reprodukovat. Prosím autora, zda by mohl postup se stručným popisem jednotlivých kroků zařadit do prezentace při obhajobě.

Testování metody autor provádí na modelu geodetické sítě. Zaujalo mě, že pro praktickou realizaci výpočtu navíc využívá databázový stroj. Případný uživatel, který bude chtít spočítat

robustní odhad z několika měření, nebude na svůj stroj instalovat PostgreSQL. Daleko větší smysl než kombinace C++/PostgreSQL by mi dávalo použití skriptu v některém matematickém SW: Matlab, Octave,, což je uživatelsky vstřícnější i nepoměrně jednodušší postup.

K navrženému konceptu mám následující připomínky:

- 1) Při použití metody pro řídké matice má navržená reprezentace 1D polem jednu podstatnou nevýhodu, neumožňuje získat prvek matice v konstantním čase, avšak v $O(n)$, což je pro praktické použití nevhodné. Obdobně komplikované je i přidání prvku do matice. Autor sice dosáhl efektivnějšího uložení, avšak za cenu zvýšené režie při práci s prvky matice. Z kódu je patrné, že autor tuto myšlenku následně opustil, a maticové výpočty řešil klasicky.
- 2) Zajímavou, avšak již použitou myšlenkou (*Chong-wei et al, 1993*), představuje autorem navržená paralelizace výpočtu LMS. Otázkou je, zda pro tento konkrétní koncept řešení (problém maticových operací souvisejících s LMS není efektivně dekomponovatelný) nepředstavuje spíše samoučelné vylepšení; nemá totiž vliv na kvalitu odhadu.
- 3) Autor pro řešení soustavy lineárních rovnic zvolil metodu konjugovaných gradientů, proč pro operace s "malými" maticemi použil tuto iterativní metodu, která je navíc náročná na čtení a vkládání prvků do matice (bod 1)?

Při čtení práce jsem nabyl přesvědčení, že autor věnoval více času implementaci metody, než jejímu konceptuálnímu návrhu, popisu a analýze (4 strany vs. cca 50).

vynikající nadprůměrný průměrný podprůměrný slabý

Výsledky disertace - konkrétní přínosy disertanta

Hlavním přínosem práce je návrh a algoritmu pro řešení problému LMS, jeho implementace a testování. Algoritmus byl otestován nad modelem vyrovnání geodetické sítě; zda funguje i pro obecná data není z práce patrné. Pokud bude hlavním cílem hodnocení efektivity metody kvalita řešení, lze považovat implementační detaily a jejich popis za méně významné, byť tvoří většinu práce. Autor programuje s velkou rutinou, avšak v kontextu této práce je programovací jazyk pouze prostředkem a nikoliv cílem.

U předložené disertační práce je velmi obtížné rozlišit, které konkrétní dílčí kroky představují konkrétní přínos autora, a které jsou převzaty odborné literatury (autor ani v kapitole 13.2 s návrhem algoritmu neuvádí, že se jedná o klíčovou část jeho práce).

Autor provádí detailní testování navržené metody s různými váhovými schémata, výsledky jsou prezentovány ve formě tabulek a grafů, zamýšlí se nad praktickým použitím metody. Zaujalo mě, že při použití váhových schémat se hodnoty odhadu mediánu liší již na čtvrtém desetinném místě (Tab. 13.2, 13.3, str. 178, 179). Při aplikacích v geodézii, kde pracujeme s desetinnými vteřinami, by přesnost takového řešení nebyla dostatečná. Podobná připomínka se týká i Tab. 13.4 (str. 180), kde se v závislosti na počtu "převažovaných" řádků matice liší hodnota mediánu z jednotlivých kombinací cca o 20%. Prosím o vysvětlení při obhajobě.

Autora chválím za vzájemnou integraci jednotlivých modulů tvořících celek a za implementaci algoritmu; s tímto krokem autor strávil spoustu času a věnoval mu velké úsilí. Zdrojový kód podporuje normu C++11 a je dostupný pro případné zájemce na síťovém úložišti [github](#), což oceňuji. Ve formálním popisu algoritmu však chybí komentáře, což poněkud snižuje čitelnost.

vynikající nadprůměrný průměrný podprůměrný slabý

Význam pro praxi a pro rozvoj vědního oboru

Navržená metoda má spíše akademický význam, její využití má lokální charakter a bude se týkat specifických pasáží vyrovnávacího počtu souvisejících se zpracováním observačních dat malého rozsahu v geodézii. Bylo by přínosné, kdyby se metodu podařilo zobecnit pro jiné než geodetické použití. Za hlavní omezení při praktickém použití považuji asymptotickou složitost algoritmu (exponenciální), která vyplývá z Tab. 9.26 (str. 119). Pro soustavu lineárních rovnic ($m = 25, n = 9$) probíhá výpočet 489 s, což je zhruba 8 min, přitom se jedná o "malá" data. Běžně používané metody pracují s daty o 2 řády většími.

vynikající nadprůměrný průměrný podprůměrný slabý

Formální úprava disertační práce a její jazyková úroveň

Disertační práci tvoří 237 stran, z toho vlastní text představuje 197 stran. Text práce by bylo vhodné přiměřeně zkrátit a zpřehlednit: vypustit kap. 6, 7, spojit kapitoly 4+5, 10+11, některé pasáže jsou popsány příliš detailně včetně odvození, která lze najít v odborné literatuře, jiné jsou nezvykle stručné. Autor využívá systém LaTeX, což má pozitivní vliv na sazbu, zejména matematických vzorců, k jazykové úrovni nemám připomínky.

U grafů převažuje textový popis obou os místo uvádění zástupných proměnných (dtto u tabulek), v tabulkách navíc nejsou u záhlaví sloupců uváděny jednotky, střídá se zápis hodnot v pevné i plovoucí řádové čárce v jedné tabulce (napříč celým textem), přehlednosti by prospělo rozumné zaokrouhlení tabelovaných hodnot (Tab. 9.7, str. 99).

V textu se vyskytuje různý překlad termínu LMS:

“... *minimalizace mediánu čtverců oprav měřených veličin.*” (str. 17).

“... *minimalizace mediánu čtverců reziduí.*” (str. 17).

“... *minimalizace mediánu nejmenších čtverců.*” (str. 21).

V textu jsou mnohdy použity specifické termíny: design matrix (prvá matice plánu, str. 17), bod zvratu (str. 25), feasible řešení (str. 34), score funkce (str. 43), parciální kvadratická forma (59), smyčka (slučka, str. 61).

vynikající nadprůměrný průměrný podprůměrný slabý

Připomínky

K textu práce mám několik níže uvedených připomínek spíše formálního charakteru:

- str. 24: “Metoda MNČ ... dokáže poskytnout správný odhad, pokud obsahuje některou z metod testování odlehlých měření.” Kéž by tomu tak bylo ...
- str. 25: “Outliers ... nejsou vzdáleny od regresní přímky, ale jsou vzdáleny od skupiny měření”. Není pravda, jedná se o hodnoty vzdálené od regresní přímky; mohou mít různou leverage a impact (zohledňuje oba faktory). Záměna těchto pojmů na více místech v textu.
- str. 25: Vzorec 3.3: zvláštní forma zápisu, navíc místo „beta“ má být β .
- str. 27: Algoritmy pracující hrubou silou nemají polynomiální, ale exponenciální či faktoriální složitost.
- str. 34: “Feasible řešení”. Co to znamená? Znamé jsou termíny “feasible region”, “feasible set” odpovídající stavovému prostoru řešení.
- str. 43: Vzorec 4.4, zřejmě má být $|\epsilon_i|^{p-1}$.
- str. 43: Vzorec 4.6, symbol $|\cdot|$ představuje normu nebo absolutní hodnotu?
- str. 44: Vzorec 4.11, zřejmě má být $a_i x$, a_i je řádkový vektor.
- str. 45: Vzorec 4.15, pro první případ má být $|v| \text{sign}(v)$.
- str. 47: Co je $L_1 L_2$ norma?
- str. 53: Vzorec 5.1, nejsou vysvětleny symboly něm použité.
- str. 53: Vzorec 5.5 nedává smysl.
- str. 55: Vzorce 5.13, 5.14, levé strany nerovnosti rovny 1.
- str. 61: Vzorec 6.3 se vztahuje k čemu? Tvoří samostatnou větu.
- str. 64: Autor zmiňuje použití garbage collectoru při realokaci objektů na haldě (heapu) s cílem vytvořit souvislé oblasti v této části paměti. C++ však nepoužívá garbage collector a haldu tvoří nesouvislé alokované oblasti paměti.

Závěrečné zhodnocení disertace

Předložená disertační práce obsahuje originální autorské myšlenky vedoucí k řešení zadaného problému, které se opírají o teoretické zdůvodnění, byť vzhledem k exponenciální složitosti je praktická použitelnost metody sporná.

Práce Ing. Gabriela Györiho splňuje podmínky kladené na doktorskou disertační práci, představuje původní vědecké dílo s přínosem v oblasti teoretické i implementační, byť popis a rozbor navrhované metody měly být detailnější a metoda měla být koncipována i pro obecná

data.

Autor prokázal schopnost tvořivého myšlení, systematického, logického přístupu i samostatného řešení vědeckého problému.

Připomínky obsahového i formálního charakteru poněkud snižují úroveň této práce.

Doporučuji však, aby Ing. Gabrielu Györimu **bylo umožněno** obhájit jeho doktorskou disertační práci.

Doporučuji po úspěšné obhajobě disertační práce udělení titulu Ph.D. ano ne

Datum: 4. února 2018

Podpis oponenta:

A handwritten signature in blue ink, appearing to read 'Bajtar', with a horizontal line extending to the right.