

Bc. Viktorie Nesrstová

na téma

Metody robustní vůči prvkovým odlehlým hodnotám

Diplomová práce o celkovém rozsahu 75 stran je členěna do čtyř kapitol. V první kapitole autorka čtenáře seznamuje se základními pojmy z oblasti robustní statistiky a vysvětluje rozdíl mezi řádkovými a prvkovými odlehlými pozorováními. V další kapitole následuje představení tří konkrétních metod pro práci s prvkovými odlehlými hodnotami. Zatímco první z nich, Detect Deviating Cells, je určena zejména k identifikaci těchto odlehlých pozorování, zbývající dvě, střelející S-odhad a tříkroková regrese, jsou variantami robustní regrese. Tato kapitola je, podobně jako ostatní, sepsána velmi přehledně a působí uceleným dojmem. Na druhou stranu se však autorka v některých pasážích až příliš striktně drží zdrojového textu. Některé postupy představené v druhé kapitole by si zasloužily důkladnější vysvětlení a vybrané nově se vyskytující pojmy (S-regrese, s. 25; EM algoritmus, s. 30; konzistence...) alespoň odkaz na relevantní literaturu.

Hlavním přínosem práce je dle mého názoru třetí a čtvrtá kapitola. Simulační studie poskytuje zajímavé srovnání představených metod z hlediska jejich citlivosti na „míru odlehlosti“ nestandardních pozorování, či výkonnosti při různých hodnotách počtu proměnných a pozorování. Schopnost praktického použití robustních metod pak autorka prokázala analýzou konkrétní datové sady. K praktické části práce mám několik otázek:

- V kapitole 1.2 jsou představeny dva kontaminační modely, souvisí s nimi nějak způsob přípravy kontaminovaných datových sad v simulační studii?
- Která metoda je pro analýzu dané reálné datové sady nejvhodnější? Lze o tomto vůbec rozhodnout?
- Z jakého důvodu byla použita log transformace vysvětlujících proměnných? Znamenala tato úprava nějaké vylepšení použitých modelů, a jak by nyní vypadala interpretace jejich parametrů?

Mimo tyto otázky mám k práci také několik drobnějších poznámek, které uvádím níže:

s. 9 Lze uvést konkrétní příklad bodu selhání, např. pro medián nebo průměr?

s. 11 Zde je uvedeno, že za účelem porovnání robustního odhadu variability a směrodatné odchylky, zavádíme tzv. normovanou MAD. Hodnoty 1,18608 a 1,11195 tedy odpovídají MAD nebo MADN?

s. 14 „Bod selhání robustních M-odhadů je 0 %, jelikož tyto odhady nejsou obecně robustní vůči vlivným hodnotám.“ Tomuto nerozumím, proč se jim tedy říká „robustní M-odhady“?

s. 16 Zápis zavádějící novou třídu (kontaminovaných) rozdělení není příliš jasný a zasloužil by si podrobnější komentář.

s. 24 Ve vztahu (2.11) se vyskytuje značení $ave F()$. Není však zřejmé, co funkce *ave*, která se zde objevuje poprvé, znamená.

Jakým způsobem probíhá destandardizace matice $\mathbf{Z}_{imp}$?

s. 30 V souladu se značením zavedeným v kapitole 1 by mělo být psáno *Med* namísto použitého *med*.

s. 37 Co znamenají jednotlivé parametry funkce DDCpars? Souvisejí nějak s parametry vyskytujícími se v algoritmu metody DDC, jež byl popsán v kapitole 2.1.1.?

s. 41 Interpretace FDR myslím není správná. Hodnota 0,75 neznamená, že se podařilo identifikovat 25 % (všech) skutečně odlehlých hodnot, nýbrž že z hodnot, které metoda označila za odlehlé, 25 % skutečně bylo odlehlými hodnotami. Jakou část všech outlierů to představuje, ale nevíme.

s. 54 Autorka uvádí, že největší směrodatná odchylka je spjatá s proměnnou wrist. Myslím, že toto zjištění není příliš překvapivé, uvědomíme-li si, že proměnné wrist odpovídá také vysoká hodnota regresního koeficientu.

Ani rozdíl ve znaménku regresního koeficientu u proměnné ankle mi nepřijde příliš zajímavý. U všech metod mu totiž přísluší poměrně vysoká směrodatná odchylka, a tento koeficient tak (bez ohledu na jeho bodový odhad) patrně není významně odlišný od nuly.

Co se mi naopak zdá být zajímavé, jsou záporné koeficienty u proměnných neck, chest, hip aj., lze tento vztah k celkové hodnotě tělesného tuku nějak vysvětlit?

Celkově považuji práci za zdařilou, navíc psanou s minimem gramatických a stylistických chyb. Navzdory výše uvedeným komentářům ji tak doporučuji k obhajobě a navrhuji hodnotit stupněm **B**.

V Olomouci dne 29. 4. 2019

Kamila Fačevicová