

Posudek oponenta diplomové práce

Autor/Autorka

Bc. Zuzana Loudová

Název práce

Metody identifikace odlehlých pozorování

Studijní obor

Finanční informatika a statistika

Oponent práce

Ing. Tomáš Ťoupal, Ph. D.

Splnění cílů práce:

nadstandardně velmi dobře splněny s výhradami nebyly splněny

Odborný přínos práce:

nové výsledky netradiční postupy zpracování výsledků z různých zdrojů shrnutí výsledků z různých zdrojů bez přínosu

Matematická (odborná) úroveň:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Věcné chyby:

téměř žádné vzhledem k rozsahu přiměřený počet méně podstatné, větší množství podstatnější, větší množství závažné

Grafická, jazyková a formální úroveň:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Slovní hodnocení a dotazy:

Předložená diplomová práce se zabývá problematikou týkající se detekcí odlehlých hodnot ve vybraných souborech dat z důvodu „eliminace“ zkreslených výsledků při následujícím statistickém šetření. Nejprve jsou zde uvedeny metody vhodné pro detekci odlehlých pozorování v jednorozměrných i vícerozměrných souborech dat a následně je problematika rozšířena i na časové řady při detekci v regresi. Použité metody jsou jak parametrické, tak i neparametrické. V další části práce jsou pro zvolené metody provedeny numerické experimenty (generování požadovaných souborů dat), odhalující výhody a nevýhody vybraných postupů. Závěr práce se zabývá aplikací zvolených metod na souboru reálných dat.

Práce působí uceleným a promyšleným dojmem, který je podmíněn šíří obsahu získaných metod z různých zdrojů a následně završen použitím popsání modelů na jednorozměrné i dvourozměrné soubory dat včetně získaných reálných hodnot. Bohužel jsou zde zároveň obsaženy i některé nedostatky či nepřesnosti, ze kterých jsou patrné zejména:

- Překlepy.
- Nepřesné formulování a používání pojmů, např. střídání pojmu „odlehlá pozorování“ v anglické i české verzi.
- Nepřesnosti vzniklé při přepisování teorie z odborných metodik, např. lepší odlišení vektorů či konkrétních hodnot, odlišení indexového značení pro seřazené a neseřazené pozorování atd.
- Interpunkce za vzorci, chybí označení konkrétní verze použitého softwaru a odkazy na syntaxe konkrétních funkcí, sjednocení značení vzhledem k širokému spektru použitých zdrojů.

- Grafické zpracování získaných výsledků, občas chybí popis os, přehlednost, parametry a typ vygenerovaného rozdělení alespoň v popisku atd.
- Nepřesné formulování pojmů, kdy jsou použity formulace typu „pozorování s vysokou hodnotou“, „skutečná vzdálenost“ atd.
- Detailnější rozbor problematiky, na které jsou modely stavěny např. popis či odkaz na zdroj pro použití metrik, které se objeví až v testovací části např. Euklidovská metrika, používání pojmu „stopa“ a následné využívání zkratky bez předchozího popisu.
- Zároveň jsou v textu obsaženy některé chybné vzorce např. (10), kde jsou použity různé indexy a špatné značení pro odhad a tvrzení, které nejsou ani dokázány ani uveden zdroj použití, např. u „klasických reziduí“.
- Volba odlehklých hodnot ve vygenerovaném souboru dat, tj. hodnoty neovlivňující daný typ rozdělení.
- Atd.

Domnívám se, že cíle práce jsou splněny a autorka prokázala pochopení dané problematiky, a proto doporučuji práci k obhajobě.

Otázky:

Stručně popište použité metriky v práci a zároveň navrhněte další, které by mohly být použity?

Práci doporučuji – ~~nedoporučuji~~ uznat jako kvalifikační (*nehodící se škrtněte*).

Navrhuj hodnocení známkou:

VELMI DOBŘE

Datum, jméno a podpis: 11. 6. 2015, Tomáš Ťoupal

