

POSUDOK VEDÚCEHO BAKALÁRSKEJ PRÁCE

Názov: Krabicový diagram pro vícerozměrná data

Autor: Tomáš Brabenec

ZHRNUTIE OBSAHU PRÁCE

V práci sú zavedené, a porované, tri prístupy k rozšíreniu krabicových diagramov na viacerozmerné dáta známe z literatúry. Prvá časť textu je venovaná diskusii o Mahalanobisovej vzdialenosti bodov od stredu elipticky symetrických rozdelení, ktorej vlastnosti sú kľúčové pre správne pochopenie geometrie viacerozmerných distribúcií.

V sekcii 1.2 je zavedené najjednoduchšie rozšírenie krabicových diagramov, tzv. Range-finder. Krátka diskusia poukazuje na radu nedostatkov tohto značne naivného prístupu.

V kapitole 3 autor zavádza, a skúma, tzv. Relplot, ktorý dáta aproximuje vhodne volenými elipsoidmi. Prístup z originálneho článku Goldberg a Iglewicz (1992, ďalej [GI]) je však sledovaný iba čiastočne. Tu autor ihneď pracuje s d -rozmernými rozdeleniami, a s využitím znalostí o Mahalanobisovej vzdialenosti až v druhej časti kapitoly ukazuje, že ním volený flexibilný prístup k elipsoidom generovaným Mahalanobisovou vzdialenosťou je totožný s neprehľadnými predpismi pre Replot z [GI].

V poslednej časti práce je skúmaný tzv. Bagplot, ktorý tvar náhodného výberu popisuje konvexným obalom istých význačných bodov.

Všetky tri prístupy sú porovnávané na rade príkladov, ktoré vhodne ilustrujú teoretický výklad. Diagramy sú porovnané na základe ich pokrytia (počtu bodov vo vnútornej krabici), robustnosti, a schopnosti odhaľovať odľahlé pozorovania.

CELKOVÉ HODNOTENIE PRÁCE

Vlastný príspevok. Dôkazy tvrdení v kapitolách 2 a 3 sú originálnym príspevkom autora. Za najdôležitejší príspevok považujem detailný popis, a zovšeobecnenie, konštrukcie Replotu z kapitoly 3. V [GI] sa autori zameriavajú iba na dvojrozmerné dáta, a Relplot konštruujú algoritmicky bez ďalších komentárov a náhľadov. Voľba niektorých voľných parametrov v algoritme je komentovaná iba vágne, a v praxi nie je vôbec zrejmé ako tieto parametre voliť (alebo interpretovať). V kapitole 3 predloženej práce autor podrobne analyzuje postup [GI], zdôvodňuje nejasné kroky konštrukcie, a ponúka vlastné interpretácie Relplotu v reči dobre známej Mahalanobisovej vzdialenosti. Zaujímavým pozorovaním je, že zdanlivo sofistikovaný Relplot nie je nič iné ako dva (robustné) odhady úrovňových množín Mahalanobisovej vzdialenosti s požadovaným asymptotickým pokrytím platným pre normálne rozdelenie.

Matematická úroveň. Úroveň práce je veľmi dobrá. Autor pracuje systematicky a dôkladne. Všetky tvrdenia sú riadne podložené teoretickými výsledkami.

Práca so zdrojmi. Zdroje sú citované vhodne a správne.

Formálna úprava. Pokiaľ viem, práca neobsahuje formálne chyby.

ZÁVER

Autor našťudoval a zrozumiteľne skompiloval značné množstvo netriviálnej odbornej literatúry. Ukázal, že je schopný kriticky porovnávať rôzne prístupy k riešeniu problému, a samostatne odvodzovať dôkazy jednoduchších tvrdení.

Prácu považujem za veľmi vydarený úvod do problematiky vizualizácie viacrozmerných dát a detekcie odľahlých pozorovaní. Určite ju odporúčam uznať ako bakalársku prácu na MFF UK.

Stanislav Nagy
KPMS MFF UK
19. mája 2019