

POSUDOK HABILITAČNEJ PRÁCE

Téma: **Hierarchical cluster analysis of categorical data**

Typ záverečnej práce: Habilitačná práca

Autor: Ing. Zdeněk Šulc, Ph.D.

Oponent: doc. Ing. Mária Vojtková, Ph.D.

Kritériá hodnotenia záverečnej práce:

1. Stanovenie cieľa a miera jeho splnenia

Autor v úvode habilitačnej práce stanovil 3 základné vedecko-výskumné ciele:

- Preskúmať miery podobnosti pre kategoriálne dáta (predovšetkým nominálne premenné s viac ako dvomi kategóriami) a odporučiť použitie vhodných mier podobnosti v závislosti od použitej metódy zhlukovania ako aj vlastností dátového súboru. Výsledky dokazujúce splnenie tohto cieľa sú prehľadne zosumarizované v kapitole 5 - podkapitole 5.3.5.
- Porovnať bežne používané hodnotiace kritériá určené na posúdenie kvality zhlukovania pre kategoriálne dáta, analyzovať ich vzájomné vzťahy z rôznych uhlov pohľadu a skúmať vzťah medzi nimi. Autorovi sa nielen, že podarilo naplniť uvedený cieľ, ktorý zosumarizoval v kapitole 6 - podkapitole 6.3.4, ale navyše navrhol 2 nové hodnotiace kritériá pre kategoriálne dáta modifikáciou Hartiganových kritérií založených na variabilite a entropii.
- Predstaviť druhú generáciu balíka jazyka R *nomclust* a ilustrovať jeho použitie. Posledný cieľ je predovšetkým praktický a jeho naplnenie autor ilustroval podrobným opisom jednotlivých krokov analýzy prostredníctvom dvoch experimentov. Prvý experiment analyzoval 16 mier podobnosti na 2700 dátových súboroch a druhý experiment porovnával 11 hodnotiacich kritérií kvality zhlukovania na 8100 dátových súboroch.

2. Vhodnosť použitých metód, metodológia

Problematiku metodológie a použitých metód skúmania hierarchickej zhlukovej analýzy pre kategoriálne údaje autor podrobne opísal v kapitole 1, 2 a 3, pričom schopnosť vhodne aplikovať jednotlivé metódy a postupy spomínaných metód preukázal uskutočnením dvoch experimentov založených na generovaných dátových súboroch s dopredu stanovenými parametrami s využitím programového jazyka R. Otvára sa tu výzva na porovnanie výsledkov aplikácie iných softvérových riešení na túto problematiku.

3. Zhodnotenie poznatkovej bázy

Nevyhnutným predpokladom spracovania habilitačnej práce bolo naštudovanie poznatkov z oblasti použitých metód ako aj stavu skúmania v oblasti hierarchického zhlukovania kategoriálnych údajov. Autor pri spracovaní využil aktuálnu prevažne zahraničnú literatúru z predmetnej oblasti, pričom v predloženej práci sa odvoláva a cituje veľký počet zdrojov uvedených na str. 99 až 103. Predloženou prácou potvrdzuje, že sa vie dobre orientovať v citovanej literatúre, vie formulovať problém a vhodnými matematicko-štatistickými metódami s využitím zvoleného softvéru nájsť aj jeho riešenie.

4. Vedecký prínos a originalita práce

Predkladaná práca sa metodologickým aparátom s aplikáciou dvoch experimentov môže zaradiť medzi vedecké monografie a práve v tom vidím vedecký prínos habilitačnej práce. Publikovanie práce ako monografie by študentom, ale aj výskumným pracovníkom umožnilo hlbšie vniknúť do problematiky hierarchického zhukovania so zameraním na kategoriálne údaje, ktorá nie je až tak prebádaná. Zhukovaniu kategoriálnych údajov sa autor v niektorých súvislostiach venuje ako prvý. Súčasne by som chcela oceniť zvolený jazyk pre písanie práce, ktorý prácu predurčuje na zverejnenie nielen v domácich podmienkach, ale predovšetkým v zahraničí.

5. Aplikačné prínosy práce pre prax

Habilitačná práca poskytuje poznatky, odporúčania a postupy aplikácie hierarchickej zhukovej analýzy pre kategoriálne dáta, ktoré sú v súčasnosti obvyklým výsledkom zisťovania napríklad v marketingu, ale aj v mnohých iných oblastiach spoločenského života. Praktická časť práce je výsledkom aplikácie dvoch experimentov, pričom jednotlivé kroky pre ich realizáciu sú podrobne opísané a aplikovateľné v praxi prostredníctvom „open source“ softvéru v programe R. Okrem uvedených prínosov môžeme vnímať aj prínos práce v rovine vzdelávacej činnosti.

6. Otázky pre autora pri obhajobe práce

- a) Autor sa v práci venuje hierarchickým metódam zhukovej analýzy pre kategoriálne údaje. Pri zhukovaní číselných údajov sa v súčasnosti veľmi často stretávame s využitím nehierarchických zhukovacích metód. Aká je skúsenosť autora s použitím nehierarchických metód pre kategoriálne údaje?
- b) Autor podrobne ilustroval aplikáciu hierarchickej zhukovej analýzy pre kategoriálne údaje pomocou balíka *nomclust* jazyka R. Aký je Váš názor na použitie programovacieho jazyka Python resp. niektorého z komerčných balíkov programov ako napríklad SPSS alebo SAS pri aplikácii spomínaných metód?

7. Záverečné odporúčanie

Konštatujem, že predložená práca spĺňa požiadavky kladené na habilitačné práce z hľadiska obsahu, formy, prínosu a stanoveného cieľa. Predloženú prácu odporúčam na obhajobu pred vedeckou radou Fakulty informatiky a štatistiky VŠE a po jej úspešnom obhájení navrhujem udeliť Ing. Zdenkovi Šulcovi, Ph.D. vedecko-pedagogický titul „docent“

V Bratislave, dňa 5.2.2024

Podpis Vojtková