



UNIVERZITA KARLOVA V PRAZE
MATEMATICKO-FYZIKÁLNÍ FAKULTA

KATEDRA PRAVDĚPODOBNOTI A MATEMATICKÉ STATISTIKY

186 75 PRAHA 8, SOKOLOVSKÁ 83

Tel : 221913275

Karolína Rezková

Archetypální analýza jako segmentační nástroj

Posudek bakalářské práce

Předložená bakalářská práce je věnována nepříliš známé metodě z oblasti mnohorozměrné analýzy, tzv. archetypální analýze, a jejímu použití pro segmentaci zákazníků.

Autorka nejprve popisuje základní optimalizační problém uvažovaný v daném kontextu a poté podrobně ukazuje, že řešení, tj. archetypy, leží na hranici konvexního obalu dat. Hlavním obsahem práce je analýza reálných dat převzatých z literatury. Data jsou stručně popsána v kapitole 2. Kapitola 3 nejprve krátce rozebírá pojem segmentace, a poté analyzuje data popsána v předchozí kapitole.

Práce je napsána v TeXu, který by si na řadě míst zasloužil mnohem větší pozornost, například použití programu vlna.exe. Jednopísmenné předložky a spojky na koncích řádků velmi ruší a znepríjemňují čtení. Také odkazování na vzorce není vždy korektní. Ve výtkách formálního typu by se dalo pokračovat. Použitá čeština je na mnoha místech značně kostrbatá, používané anglikanismy na řadě míst velmi ruší (surová data, dataset, ...).

Autorka cituje pouze základní zdroje. Z obsahu je jasné, že mimo ně žádnou další literaturu nepoužila, což je spíše škoda.

K obsahu práce mám následující připomínky:

1. Autorka ne vždy jasně rozlišuje mezi náhodnými veličinami a jejich realizacemi.
2. Chtěl bych upozornit, že problematice segmentace dat byla v literatuře věnována mnohem větší pozornost, než se autorka domnívá. Přístup k segmentaci uvedený na straně 15 by si zasloužil mnohem větší pozornosti a pochopení ze strany autorky.
3. Místo mechanického výpočtu uvedeného na horní polovině strany 11 by bylo užitečnější, kdyby si autorka uvědomila, že v klasickém přístupu tvoří pořadí některou náhodnou permutaci čísel $\{1, \dots, n\}$ a r_5 vyjádřila ve tvaru mnohem zajímavější pro interpretaci, než je uvedeno v předložené práci.
4. Autorku bych rád upozornil, že systematizace dat (kvalitativní, kvantitativní, ...) byla v literatuře provedena dávno před citovanou literaturou, a běžně se používá.
5. V 10^6 jsou i, j indexy a ne proměnné.
6. V 10_{10-11} je použita dosti nešťastná formulace.
7. Zavedení minima a maxima v 11_{11} je poněkud neortodoxní.
8. Atd.

K obhajobě mám následující dotazy,

1. Jak autorka zjišťovala, viz strana 13, že někteří respondenti odpovídají nekonzistentně?
2. Jak autorka definovala pořadí při shodě dat, zvláště pak u binárních proměnných?
3. Jakou vypovídací hodnotu má korelační koeficient, ať již Pearsonův nebo Spearmanův, formálně spočtený mezi binární proměnnou BrI a nominální proměnnou ZnačkaProduktu, apod.? Blíže viz paragraf 2.2.1.
4. Jaký význam má znak procenta v popisu vertikální osy na obrázku 3.2?
5. Je opravdu zapotřebí nominální proměnné nahrazovat proměnnými binárními, jak je v práci vesměs uvažováno?
6. Došlo by v 12¹⁷ ke změně výsledků, kdybychom kódovali muže nulou a ženy jednotkou?

Téma samotné považuji za zajímavé a jsem přesvědčen, že je v něm stále otevřena řada zajímavých problémů. Tento potenciál autorka nikterak nevyužila. Celkově se jedná o podprůměrnou práci jak po stránce matematické, tak po stránce aplikační. Práci doporučuji k obhajobě.

Praha 21. srpna 2015

Prof. RNDr. Jaromír Antoch, CSc.