

Posudek oponenta disertační práce

Název: Analysis of Client and Product Base: Specifics of Retail Transaction Data

Autor: Ing. Ondřej Sokol

Školitel: prof. RNDr. Ing. Michal Černý, Ph.D.

Oponent: doc. Ing. Jaromír Kukal, Ph.D.

Předložená disertační práce je věnována aplikaci matematických metod na velmi aktuální oblast analýzy transakčních dat z hlediska zákazníka a jeho chování ve vztahu k produktům a jejich kategoriím. Jsou zde v samostatných kapitolách navrženy tři různé přístupy k analýze dat. Práce je velmi přehledně členěna. Výklad jednotlivých přístupů vždy začíná uvedením do problematiky a rešerší současného stavu výzkumu. Pak následuje motivace a exaktní matematický popis nově navrženého modelu, analýza jeho vlastností včetně ukázek numerického chování na umělých a reálných datech.

První úloha se zabývá shlukováním produktů na základě dat o chování spotřebitelů. Autor zde navrhl originální účelovou funkci, pomocí které popisuje kvalitu disjunktního rozkladu množiny produktů na jednotlivé shluky. Podařilo se mu dokázat, že příslušná celočíselná optimalizační úloha je NP-hard. Pro nalezení jejího optima využívá standardní nástroj pro genetickou optimalizaci. Teorii uvedené metody se podařilo publikovat v impaktovaném časopise. Zde bych chtěl znát názor autora na citlivost účelové funkce (2.4) na malé hodnoty D_k ve jmenovateli, což je obecná nectnost průměrování poměrů hodnot. Bylo by možné zkonstruovat ze stejných veličin jiné účelové funkce?

Druhá úloha je zaměřena na segmentaci zákazníků na základě agregovaných hodnot z nákupních košů podle kategorií produktů. Autor zde navrhuje vektorový popis dat, na který navazuje standardní shlukovou analýzou metodou k-means prováděnou na dvou úrovních. Zde bych se rád zeptal na alternativní metody shlukové analýzy (např. DBSCAN, OPTICS) a na vhodnost jejich využití v dané aplikaci v závislosti na velikosti dat, délce příznakového vektoru a počtu shluků.

Třetí úloha se týká pokročilé analýzy chování zákazníků ve vztahu k jejich členství v zákaznickém klubu. S využitím podmíněných pravděpodobností je zde aplikována metoda maximální věrohodnosti, a tak je posléze formulována a řešena úloha nelineárního programování.

Závěrem konstatuji, že práce je realizována na velmi dobré matematické, ekonometrické a typografické úrovni, přináší nové vědecké poznatky a splňuje obecné požadavky kladené na disertační práci. Proto ji doporučuji k obhajobě před příslušnou komisí.

V Praze dne 31.3.2020

doc. Ing. Jaromír Kukal, Ph.D.

