

Posudek oponenta bakalářské práce

Autor: **Jazmína Kreanová**

Název práce: **Aplikace stochastických metod na egyptologická data**

Posudek vypracoval: **Ing. Martin Oharek**

Bakalářská práce se zabývá analýzou egyptologických dat z archeologického výzkumu pomocí různých druhů shlukovacích algoritmů. Dlouhodobým cílem práce je s pomocí matematického aparátu a metod strojového učení zkoumat starověký Egypt a hledat zajímavé vztahy a poměry ve fungování tehdejší společnosti.

Práce je rozdělena do několika částí. V první části je popsán úvod do shlukování, ve kterém studentka důkladně zavádí a popisuje obecné informace o shlukování, porovnává rozdíly mezi jednotlivými přístupy včetně grafických ilustrací. V druhé části jsou popsány konkrétní algoritmy včetně stručných pseudokódů. Text neobsahuje pouze strohý výčet metod, ale také zamyšlení a porovnání algoritmů včetně jistých výhod i nevýhod. Po formální stránce shledávám jako mírný problém překlep ve vzorci (2.4) na straně 28 (objevil se už v odvození (2.3)), kdy je psán index j místo správné varianty l . Dále v sekci 2.4.2 by bylo vhodné zadefinovat hojně používané pojmy z teorie grafů (strom, les, úplný ohodnocený graf,...), kde by nezasvěcený čtenář mohl obtížně chápat uvedený popis algoritmu.

Závěrečná část práce je věnována experimentům. Třetí kapitola je zaměřena na porovnání popsaných metod v teoretické části na uměle vygenerovaných 2D datech. Je uvedeno velké množství experimentů, které jsou detailně popsány a výsledky slovně i graficky vysvětleny. V samotném textu kapitoly jsou uvedeny pouze ilustrativní grafy, což napomáhá plynulejší četbě a pochopení. Kladně hodnotím uvedení velkého množství dodatečných výsledků do přílohy. K úplné dokonalosti této kapitoly chybí snad jen porovnání uváděných algoritmů na vícedimenzionálních datech z důvodu existujících nedostatků uvedených metod (např. k-means s euklidovskou metrikou), což by bylo jistě přínosné i z hlediska budoucí aplikace na vícerozměrná egyptologická data.

Poslední kapitola se zabývá experimenty na reálných egyptologických datech z databáze Maat-base, jejichž výsledky a závěry jsou prezentovány v podobném stylu jako v předchozí

kapitole. Studentka rozdělila tuto část na dvě hlavní podčásti, kde porovnává účinnost shlukovacích metod na základním datasetu s příznaky tvořenými zastoupením titulů u jednotlivých osob a následně tento dataset rozšíří o příznaky tvořenémi typem hrobky. Bylo provedeno velké množství experimentů a kvantitativních vyhodnocení doplněných o zajímavá zamyšlení nad validitou obdržených výsledků a porovnání mezi 5. a 6. periodou.

Jednotlivé části práce na sebe dobře navazují, z nichž každá je psána důkladně včetně autorských zamyšlení. Práce obsahuje místy jen drobné překlepy a nedostatky.

Bakalářskou práci doporučuji k obhajobě a hodnotím ji známkou:

A - výborně

.....

19.leden 2022

Otázky k obhajobě:

1. U některých uvedených algoritmů v této práci (např. k-means s euklidovskou metrikou) může být problém ve vyšších dimenzích (tzv. curse of dimensionality). Můžete zkusit vysvětlit, v čem problém spočívá a jak se dá případně řešit?
2. V sekci 4.1.2 uvádíte potíže se vznikem prázdných shluků. Můžete zmínit možné příčiny vzniku prázdných shluků?