

Hodnocení bakalářské práce – oponent

Autor hodnocení:	Ing. Petr Kotas
Vedoucí bakalářské práce:	Ing. Pavel Praks, Ph.D.
Oponenti:	Ing. Petr Kotas RNDr. Pavel Jahoda, Ph.D.
Téma:	Simulační algoritmus SVD pro efektivní analýzu rozsáhlých dat
Verze ZP:	1
Student:	Bc. Lukáš Pištělák

1. Splnění požadavků zadání.

Dle mého názoru student naplnil zadání bakalářské práce pouze z části. Tématu zadání se práce věnuje pouze okrajově a neposkytuje dostatečný teoretický podklad. Výsledná metoda aproximace matice pomocí Monte-Carlo je uvedena pouze ve stručnosti pomocí zkácených výpisů zdrojových kódů, ze kterých není zcela zřejmé, že by student danému problému rozumně. Algoritmické výpisy navíc neposkytují dostatečné množství informací pro laika k pochopení problému.

Naopak student zadání značně přesáhl výběrem aplikace pro demonstraci potenciálního využití zadané metody. Student zvolil metodu vyhledávání podobných dokumentů, konkrétně metodu Latentního sémantického indexování (LSI). Je pravdou, že pro tuto aplikaci je zadaná metoda vhodná, protože implicitně zahazuje data s nízkou informační hodnotou. Nízká informační hodnota je v dokumentech obvykle přisuzována šumu.

2. Hodnocení formální stránky závěrečné práce.

Bakalářská práce je logicky členěna do samostatných kapitol. K členění mám jedinou výtku, podkapitola 3.1.5 "Úprava SVD rozkladu na částečný SVD rozklad (VD rozklad)" by spíše měla následovat přímo za kapitolou o SVD.

Po odborné stránce je bakalářská práce značně odbyta. V daném textu zcela chybí teoretický popis propojení metody Monte-Carlo a SVD. Neznalý čtenář je tedy odkázán na studium původního článku, který je v textu citován v omezené míře jako zdroj algoritmu. Taktéž chybí jakékoliv numerické výsledky ověřující teorii uvedenou v původním článku. Celkově práce působí dojmem, že student danému problému nerozumí a pouze slepě opsal algoritmus.

V části popisující LSI student uvádí jako hlavní výhodu aplikace SVD redukci dimenze matice dokumentů. Z daného textu lze snadno nabýt dojmu, že místo SVD se dá použít jiný maticový rozklad, což ale není pravda. Chybí totiž popis hlavní vazby SVD a LSI a sice vytvoření ortogonální báze dokumentů, která umožňuje vyhledání podobnosti. V popisu zadávání dotazu jsou jednotlivé vzorečky vykopírovány z různých zdrojů bez další kontroly. Vznikly tak chyby, kdy student násobí ve skalárním součinu matici a vektor. Tento výsledek pak prezentuje jako úhel mezi vektory.

Ve stejné kapitole se také objevuje zmínka o částečném SVD rozkladu, který je ovšem definován v navazující kapitole špatně. Student zaměnil pojem částečný SVD rozklad a vazbu mezi SVD a rozkladem na vlastní čísla. Využití pseudoinverze bez předchozího vysvětlení také není zcela ideální.

V části zabývající se výběrem náhodných čísel, student nevedl jakým způsobem se volí koeficienty a , c , m pro lineární kongruentní operátor. Volba těchto parametrů je přitom zásadní a ovlivňuje kvalitu generovaných náhodných čísel.

V kapitole vysvětlující Monte-Carlo simulaci je špatně uveden příklad. Náhodný výběr je zmíněn pouze v první větě, dále však chybí jakékoliv vysvětlení a propojení. Formátování samotného příkladu je voleno dosti nešťastně, vede ke zmatení čtenáře, který by mohl považovat vzorec (22) za jeden. Přitom se správně jedná o dva vzorce.

Kvalitu bakalářské práce také značně snižuje množství jazykových chyb.

3. *Hodnocení výsledků závěrečné práce.*

Práce prezentuje velmi zajímavou myšlenku propojení SVD a metod Monte-Carlo pro redukci dimenze problému za pomoci náhodného výběru. Na tuto myšlenku však čtenář přijde až z prostudování citovaného článku.

V textu je zmíněno převzetí hotového softwarového řešení pro LSI, které bylo rozšířeno o metodu Monte-Carlo. Pro přehlednost by neškodilo, kdyby byl celý pseudokód uveden s vyznačenou částí samostatné práce studenta. Ze stávajícího textu totiž není patrné, kolik práce student vykonal samostatně a kolik jí převzal.

V rámci bakalářské práce nebyly odevzdány zdrojové kódy. Tyto zdrojové kódy byly dodány až dodatečně na výžádání. Bohužel však nešly spustit, uváděná funkčnost se tedy nedala ověřit.

V rámci zrychlení samotné implementace student uvádí aplikaci paralelního konstruktu "parfor", který údajně urychlil výpočet norem. Při prostudování zdrojových kódů se však ukázalo, že parfor je aplikován nesprávně (bez inicializace paralelního toolboxu). V tomto případě se parfor chová dokonce pomaleji než původní sekvenční for, protože kontroluje existenci paralelního toolboxu. Student také zmiňuje efektivní využití potenciálu Matlabu, o vektorizaci samotného kódu se však ani nepokusil.

Prezentované výsledky z aplikace na vyhledávání podobnosti multimediálních dat jsou zajímavé a student prokázal samostatnost při získání dat z bezpečnostních kamer.

4. *Hodnocení práce z hlediska přínosu nových poznatků.*

Práce využívá metodu, která je popsána v doporučeném článku. Tuto metodu aplikuje na vyhledávání podobnosti v dokumentech. Z věcného hlediska, však nepřináší nic nového.

5. *Charakteristika výběru a využití studijních pramenů.*

Student čerpal z doporučené literatury, kterou cituje v rozumné míře. V úvodu je zmíněn souhlas s využitím obrazové databáze pro účely této práce, tento souhlas by měl být následován citací na původ dané databáze.

6. *Otázky k obhajobě.*

1. Vysvětlíte metodu Monte-Carlo a Latentní sémantické indexování. Jaký vztah má singulární rozklad k datům v datové matici?
2. Popište celý algoritmus prezentovaný v bakalářské práci včetně přejatého kódu. Mělo by být jasné, co je vlastním přínosem dané práce. Důraz klad'te na spojení SVD a Monte-Carlo, jak daná metoda funguje a proč funguje.
3. Jaké byly výsledky při vyhledávání barevných obrazů v porovnání s odstíny šedi? Zmiňujete testování různých barevných prostorů, není však zcela jasné jakým způsobem byly využity.

7. *Souhrnné hodnocení.*

Bakalářská práce působí odbytým dojmem a je škoda, že se student více věnoval aplikaci než samotnému tématu. Student po sobě práci evidentně nečetl, jinak by musel odhalit zjevné nedostatky prezentovaného textu.

Celkové hodnocení: nevyhověl