



POSUDEK OPONENTA DISERTAČNÍ PRÁCE

Jméno studenta: Ing. Bc. Ondřej Rozinek

Název práce: Similarity Space and Its Applications

Autor posudku: doc. Mgr. Josef Horálek, Ph.D.

Studijní program: Electrical Engineering and Informatics

Vedoucí práce: doc. Ing. Jan Mareš, Ph.D.

Odborný konzultant: Mgr. Ing. Pavel Kříž, Ph.D.

Předložená disertační práce se zabývá problematikou prostoru podobnosti a jejími aplikacemi. Jedná se oblast založenou na operacích s množinami využívající matematický aparát pro její popis. Řešenou problematiku autor rozsáhle a odpovídajícím způsobem zkoumal s cílem nalézt a popsat matematické prostory a funkce vhodné pro modelování podobnosti definovaných případů uživatelem této teorie.

Samotná struktura předložené práce odpovídá kritériím a požadavkům kladených na disertační práci. Autor nejprve vysvětluje motivaci k řešení komplexního problému prostoru podobnosti, a to formou jeho využitelnosti v různých oblastech výzkumu a aplikovatelnosti. V neposlední řadě je pak nutné zmínit, že autor se od roku 2010 věnuje výzkumu, který se zaměřuje na optimalizaci algoritmů pro vyhledávání v databázích a algoritmů fuzzy porovnávání pro tzv. nečistá data. V následujících částech práce autor rozvíjí koncept prostoru podobnosti včetně uceleného přehledu současného stavu výzkumu této problematiky a představuje nový koncept prostoru podobnosti a diskutuje jeho základní teorie. Zde se zároveň zabývá dualitou mezi prostorem podobnosti a metrickým prostorem. Ve třetí kapitole autor podrobně představuje nový pohled na lineární regresi v prostoru podobnosti a nové poznatky a metodiky. Logicky pak navazuje část zabývající se novými pohledy na teorie pevných bodů v prostoru podobnosti, a to včetně popisu současného stavu v příbuzných metrických prostorech a představuje některé aplikace diskutované teorie v prostoru podobnosti. Konkrétní nasazení prostoru podobnosti pro přibližné porovnávání řetězců, porovnávání záznamů a deduplikaci záznamů je pak náplní obsáhlé páté kapitoly, která obsahuje i demonstrace praktické využitelnosti uvedených konceptů. Na závěr pak autor shrnuje výsledky práce a diskutuje význam a dopad výsledků v rámci širší vědecké komunity s ohledem na jeho publikační činnost a shrnuje hlavní zjištění, přínosy a navrhuje směry budoucího výzkumu.

Při hodnocení cílů a přínosu předložené disertační práce je vhodné si celou problematiku rozdělit na konceptní, teoretickou, simulační a praktickou část. V teoretické části byl kladen důraz na sjednocení a další rozvoj teorie prostoru podobnosti. Ambiciózní cíl disertační práce definovaný jako návrh na sjednocení a zejména pak rozvoj teorie prostoru podobnosti a jeho



aplikovatelnost na prakticky využitelné oblasti, jako jsou zejména přibližné porovnávání řetězců, porovnávání záznamů a deduplikace v prostoru podobnosti, byly na základě kapitoly 2.6 a 2.7 naplněny. Jedná se o výzkum normalizačních transformací, které zachovávají podobnost, identifikace a důkaz podobnostních funkcí patřících do třídy funkcí, které splňují axiomy podobnostního prostoru, označeného $C(A)$ a zavedení nových popisů a definic topologie, konvergence a spojitosti v prostoru podobnosti. V neposlední řadě pak je nutné zmínit i nové teorie v oblasti pevných bodů v prostoru podobnosti, důkaz její existence, jednoznačnosti a konvergence.

Uvedená a zkoumaná teoretická východiska bylo nutné ověřit v praxi a autor tedy aplikoval teoretické poznatky v oblastech přibližného porovnávání řetězců, přibližného porovnávání záznamů a deduplikace v rámci prostoru podobnosti v rámci implementace jím navržených a testovaných algoritmů, jejichž výsledky uvedené v šesté kapitole dokládají přínos autora ve zkoumané oblasti podobnostních funkcí v úlohách NLP. Autorovi implementace teoretických návrhů v oblasti přibližného porovnávání řetězců, zejména pomocí ConvJ a ConvJW (uvedených v kapitole 5.3), překonávají stávající nejmodernější metody jak v účinnosti, tak v efektivitě, kdy autorem navržené metody dosahují skóre F1 87,95 % s ConvJW a využívají konvoluční přístupy, což umožňuje teoretickou časovou složitost $O(n \log n)$. Tento přístup zdůrazňuje potenciál pro efektivní zpracování velkých souborů dat.

Předložená disertační práce plně naplňuje smysl a cíle disertační práce. Autor na základě podrobné studie teoretických předpokladů matematického aparátu navrhl inovativní a optimalizační přístupy v oblasti prostoru podobnosti, které byly ověřeny jejich publikováním v rámci odborných konferenčních příspěvků a článků v impaktovaných časopisech. Autor však nezůstává jen u teoretického ověřování výsledků, ale své matematické postupy implementuje, jak uvádí v příloze A. Předložené a diskutované výsledky implementace nad testovacími sadami dat prokazují správnost teoretických předpokladů, a autorovi se tak podařilo prokázat správnost navržených konceptů.

Předloženou práci doporučuji k obhajobě.

V Hradci Králové dne 17.5.2024

doc. Mgr. Josef Horálek, Ph.D.