

Posudek disertační práce

Uchazeč: Ing. Miroslav Vaniš
Název disertační práce: Optimization of Bayesian networks and their prediction properties
Školící pracoviště: Ústav aplikované matematiky, FD ČVUT v Praze
Školitel: doc. Ing. Ivan Nagy, Ph.D.
Oponent: doc. Ing. Zdeněk Lokaj, Ph.D.

Aktuálnost tématu disertační práce

Zvolené téma disertační práce je velice aktuální, neboť zpracování a vytěžování dat je současným trendem v mnoha oborech. Právě doprava je jedna z vědních disciplín, pro kterou je získávání informací z datových souborů zcela krucální, stejně jako propojování expertních znalostí a datových souborů a přínos takto získaných informací a jejich aplikace v praxi může být signifikantní a pro udržitelnou dopravu rovněž zásadní. Znalost kauzalit jednotlivých proměnných a rovněž jejich vliv na chování celého systému bude do budoucna stále důležitější a v praxi významně uplatnitelná.

vynikající	<input type="checkbox"/>	nadprůměrný	<input checked="" type="checkbox"/>	průměrný	<input type="checkbox"/>	podprůměrný	<input type="checkbox"/>	slabý	<input type="checkbox"/>
------------	--------------------------	-------------	-------------------------------------	----------	--------------------------	-------------	--------------------------	-------	--------------------------

Splnění cílů disertační práce

Stanovené cíle disertační práce spočívaly v návrhu slučovacího algoritmu, který by umožňoval vytvořit lepší Bayesovskou síť pro analýzu dat, kombinující expertní a algoritmický způsob tvorby, vytvořený tento algoritmus matematicky popsat a ověřit jeho fungování na reálných datech.

vynikající	<input type="checkbox"/>	nadprůměrný	<input checked="" type="checkbox"/>	průměrný	<input type="checkbox"/>	podprůměrný	<input type="checkbox"/>	slabý	<input type="checkbox"/>
------------	--------------------------	-------------	-------------------------------------	----------	--------------------------	-------------	--------------------------	-------	--------------------------

Metody a postupy řešení

Postup i metoda řešení disertační práce je logická, vychází z teoretické tvorby síťových grafů, teorie pravděpodobnosti a popisu Bayesovských sítí jako pravděpodobnostních grafických modelů, reprezentujících vztahy mezi datovými proměnnými prostřednictvím struktury sítě a podmíněných pravděpodobnostních tabulek a popis expertního a algoritmického přístupu jejich tvorby.

Na teoretickou část navazuje analýza rozdílů mezi přístupem disertanta v této práci a již existujících přístupů a realizovaných výzkumů slučování sítí, popis slučovacího algoritmu včetně uvedení příkladu, matematický popis slučovacího algoritmu a jeho vyhodnocení. V závěru disertant provádí aplikaci vytvořeného slučovacího algoritmu

na reálných datech o dopravních nehodách v hl.m. Praze za časové období 1 kalendářního roku a vyhodnocení kvality dosažených výsledků.

vynikající		nadprůměrný		průměrný	X	podprůměrný		slabý
------------	--	-------------	--	----------	---	-------------	--	-------

Výsledky disertace - konkrétní přínosy disertanta

Dosažené přínosy disertační práce spatřuji zejména v návrhu nového slučovacího algoritmu tvorby Bayesovských sítí, kombinující expertní a algoritmický způsob tvorby, jeho matematickém popisu a ověření na reálných datech, jako nástroj pro efektivní hledání kauzalit v datových sadách a vyhodnocování příčin sledovaných jevů a jejich pravděpodobností, což jsou zejména v dopravě velmi důležitá zjištění, navíc do budoucna s obrovským potenciálem využití.

Velmi kladně hodnotím, že autor postupoval logicky od teoretické analýzy, přes návrh vlastního algoritmu, matematický popis a ověření na reálných datech, které představují ucelený soubor dat o dopravních nehodách v hl.m. Praze.

Vyvinutý algoritmus, resp. výsledky při jeho aplikaci na datové soubory však nebyl podroben srovnání s alternativními způsoby zpracování dat, v práci je uvedeno pouze teoretické srovnání odlišností od již existujících přístupů ke slučování sítí, proto není možné dostatečně zhodnotit úspěšnost vyvinutého řešení.

Na druhou stranu je nutné zmínit, že průběžné výsledky byly publikovány v několika publikacích, dosažené výsledky tak byly parciálně podrobeny oponentuře odborné komunity. Bohužel celkové výsledky doposud publikovány nebyly.

vynikající		nadprůměrný	X	průměrný		podprůměrný		slabý
------------	--	-------------	---	----------	--	-------------	--	-------

Význam pro praxi a pro rozvoj vědního oboru

Dosažené výsledky disertační práce mohou významným způsobem posunout problematiku zpracování dat, vyhodnocování kauzalit jednotlivých parametrů a jejich vliv na sledované jevy a pravděpodobnosti jejich výskytu, a to nejen v dopravě, ale v zásadě v jakýchkoliv rozsáhlých datových souborech. Realizovaná metoda tvorby Bayesovských sítí by mohla přinést podstatně lepší výsledky při vyhodnocování příčin a pravděpodobností jevů na základě rozsáhlých datových souborů, například v logistice, plánování výroby či při optimalizaci procesů u obchodních společností.

vynikající		nadprůměrný	X	průměrný		podprůměrný		slabý
------------	--	-------------	---	----------	--	-------------	--	-------

Formální úprava disertační práce a její jazyková úroveň

Formální úprava práce je na dobré úrovni, zejména logické a přehledné členění jednotlivých kapitol a grafické zpracování. Čtenář se v práci dobře orientuje, jazyková

úroveň práce je na akceptovatelné úrovni, místy však autor užil spíše románový styl, ačkoliv by práci spíše prospěl styl více formální. V práci jsou místy překlepy a gramatické chyby, což je s ohledem na využití angličtiny celkem pochopitelné.

vynikající

nadprůměrný

průměrný

X

podprůměrný

slabý

Připomínky:

- 1) Uvedené výsledky disertační práce jsou poutavé a pro praxi velice zajímavé, bohužel doposud nebyly podrobeny praktickému porovnání s jinými způsoby zpracování dat a hodnocení kauzalit, což celkovou úroveň dosažených výsledků částečně snižuje. Na druhou stranu mohu konstatovat, že parciální výsledky autor publikoval během svého doktorského studia, čímž tento nedostatek částečně eliminoval.
- 2) Datová základna, která byla zpracovávána, byla omezená, proto se domnívám, že by bylo vhodnější, aby ověření proběhlo na větším množství různorodých druhů dat, aby výsledky byly více reprezentativní a autor prokázal obecnou využitelnost dosažených výsledků.

Otázky:

- 1) Jak by se navržený algoritmus choval v případě, že by datový soubor vykazoval vysokou míru chybovosti, resp. nekonzistence dat?
- 2) Jaká jsou omezení pro rozsah zpracovaných dat, resp. množství uzlů, aby navržené řešení poskytovalo kvalitní výsledky v akceptovatelném čase?
- 3) Jak je možné navržený algoritmus dále rozvíjet, aby bylo možné jej uplatnit jako univerzální nástroj pro zpracování dat a zkoumání kauzalit?

Závěrečné zhodnocení disertace:

Předložená disertační práce splnila stanovené cíle a doporučuji ve smyslu Zákona č. 111/1998 Sb., o vysokých školách po její úspěšné obhajobě udělení akademického titulu „philosophiae doctor“ (Ph.D.).

Udělení titulu Ph.D. doporučuji

ano

ne

V Praze dne 20.10.2021

.....
doc. Ing. Zdeněk Lokaj, Ph.D.