

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Conditional Probability Models in Transportation
Jméno autora:	Petr Šinkovec
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra kybernetiky
Oponent práce:	Ing. Jiří Kubalík, Ph.D.
Pracoviště oponenta práce:	CIIRC, ČVUT

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Cílem práce bylo prozkoumat možnosti použití genetického programování jako nástroje pro modelování podmíněných pravděpodobností a porovnat jeho efektivitu s řešením založeným na umělých neuronových sítích. Úspěšné zpracování tohoto tématu vyžaduje hlubší znalosti z oblasti neuronových sítí, genetického programování, přehled v oblasti pravděpodobnosti a statistiky a schopnost efektivně naprogramovat netriviální algoritmy. Proto hodnotím toto zadání jako náročnější.	

Splnění zadání	splněno
<i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Student splnil zadání ve všech bodech.	

Zvolený postup řešení	správný
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Student pracoval s pevnou strukturou modelů podmíněné pravděpodobnosti ve formě směsi Gaussových rozdělení pravděpodobnosti. Neuronové sítě používá pro učení parametrů těchto modelů. Genetické programování používá pro hledání výrazů generujících optimální parametry modelu pro daný vstup. Kromě jednoduchého genetického programování (GP) vyzkoušel i pokročilou variantu nazvanou <i>Genetic Programming with Explicit Fitness Sharing</i> (GPEFS) ve spojení se dvěma zajímavými rozšířeními - automaticky definovanými funkcemi a optimalizací konstant vyvíjených výrazů pomocí metody zpětného šíření chyby. Všechny varianty genetického programování důkladně otestoval mnoha experimenty, výsledky statisticky vyhodnotil a porovnal s řešením založeným na neuronových sítích. K postupu řešení mám jen jednu výhradu. Student úlohu prezentuje jako problém symbolické regrese. V práci mi chybí rešerše metod symbolické regrese založených na genetickém programování, kterých je celá řada.	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Odbornou úroveň práce hodnotím velice pozitivně.	

Formální a jazyková úroveň, rozsah práce	B - velmi dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Po formální a typografické stránce nemám výhrad. Výhradu mám pouze k tomu, že text je místy příliš stručný, některé důležité věci nejsou dostatečně popsány. Například v odstavci 2.2.1 bych uvítal alespoň jednoduchý popis algoritmu GPEFS doplněný pseudokódem. V kapitole 3, věnované experimentům, mi chybí popis datových sad, nad kterými byly experimenty provedeny, ve smyslu velikosti, zastoupení jednotlivých skupin, atd.	

Výběr zdrojů, korektnost citací

B - velmi dobře

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Student uvádí 11 referencí, což považuji za dostačující počet zdrojů. Jen s ohledem na řešené téma mohl student zahrnout navíc ještě několik prací o využití genetického programování pro symbolickou regresi.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Celkově je tato práce velmi zajímavá. Je zřejmé, zejména z rozsahu implementovaného kódu i z množství provedených a zpracovaných experimentů, že práce na tomto tématu pana Šinkovce bavila. Dosažené výsledky jsou pro danou oblast, tedy generování modelů podmíněných pravděpodobností v dopravě, určitě přínosné.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

I přes uvedené připomínky patří tato práce mezi velmi dobré. Pan Šinkovec ukázal, že dokáže systematicky řešit zadaný problém a analyzovat dosažené výsledky.

Do diskuze mám následující otázky:

- Novější metody symbolické regrese založené na genetickém programování, např. *Evolutionary Feature Synthesis* a *Multiple Regression Genetic Programming*, využívají různé metody regularizované lineární regrese pro sestavení výsledného modelu z evolučně vyvinutých příznaků. Dal by se podobný přístup použít i pro tuto úlohu?
- V odstavci 2.2.1 popisujete tzv. *nicning*, použitý v GPEFS. Pro jeho správnou funkci je klíčová metrika pro posuzování podobnosti dvou výrazů. Jakou metriku jste použil?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **B - velmi dobře**.

Datum: 17.6.2016

Podpis: