POSUDEK OPONENTA ZÁVĚREČNÉ PRÁCE

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	**Zrychlení exaktního algoritmu pomocí metod strojového učení**
Jméno autora:	**Bc. David Procházka**
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Oponent práce:	Mgr. Ing. Petr Mánek
Pracoviště oponenta práce:	Ústav technické a experimentální fyziky (ÚTEF)

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání **náročnější**

Hodnocení náročnosti zadání závěrečné práce.

Zadání předložené diplomové práce vyžadovalo propojení znalostí dvou témat. Autor musel pochopit rozvrhovací úlohy a jejich matematickou formulaci, porozumět metodám konvexní a kombinatorické optimalizace, provést rešerši relevantní literatury, porovnat nejnovější přístupy k řešení zvolené třídy úloh, jeden z nich zvolit a naimplementovat. Z druhé strany bylo třeba nastudovat metody strojového učení a metodologii jejich vývoje, navrhnout vhodný způsob jejich integrace do exaktního algoritmu, zvolený model aplikovat a experimentálně vyhodnotit jeho přínos.

Splnění zadání **splněno**

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.

Zadání práce bylo spněno ve všech bodech. Práce byla navíc v závěru rozšířena o vývoj a vyhodnocení heuristického algoritmu, který lze alternativně využít k řešení zvolené rozvrhovací úlohy.

Zvolený postup řešení **správný**

Posuďte, zda student zvolil správný postup nebo metody řešení.

Autor si z uvedené literatury vybral optimalizační metodu typu Branch & Bound, která je navržena k řešení zadané třídy rozvrhovacích úloh. Algoritmus byl implementován v téměř plném rozsahu, chybějící část (tvrzení 12) je přijatelně odůvodněna.

Strojové učení (ML) je do exaktního algoritmu zavedeno v bodě, kde se výpočet větví v prohledávaném prostoru. Je zřejmé, že dobře naučený model by na začátku výpočtu mohl algoritmus rychle navést k dostatečně kvalitnímu řešení, které by umožnilo významné prořezání prohledávaného prostoru a tím i celkové zrychlení. Tato aplikace je podle mého názoru správná. Jako model ML autor zvolil rekurentní neuronové sítě (RNN) s architekturou LSTM. Toto rozhodnutí je motivováno požadavkem na flexibilitu vzhledem k velikosti rozvrhovací úlohy a hlavně předchozím úspěšným využitím architektury v citované literatuře. Předložený postup je jedním z možných správných řešení zadané úlohy. Naučený model ML je vyhodnocen na simulovaných instancích úlohy. Autor kromě úspěšnosti a výpočetní rychlosti modelu věnuje pozornost i optimalizaci hyperparametrů.

Odborná úroveň **A - výborně**

Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.

Práce obsahuje kvalitní úvod do problematiky a vynikající přehled relevantní současné literatury. Autor s uvedenými zdroji pracuje v celém rozsahu textu a své řešení staví na výsledcích ostatních výzkumníků. Formalizace řešené úlohy, definice i vyhodnocení předloženého algoritmu využívá s ustálené pojmy a postupy.

Po implementační stránce je práce velmi dobře strukturovaná. Autor zvolil vhodnou architekturu, která umožňuje účinnou abstrakci často používaných objektů a snadnou rozšiřitelnost (dobrým příkladem jsou metriky a výběrové funkce). Zdrojový kód přiměřeně využívá prostředky zvoleného jazyka a standardní knihovny ML, které v několika případech vhodně rozšiřuje.

Oceňuji, že na několika místech se zdrojový kód odkazuje na popis algoritmu v kapitole 4. V příloze mi ovšem chybí technická dokumentace, která by formálně vysvětlila význam jednotlivých tříd, funkcí a parametrů, a případně obsahovala příklad jejich využití. Dále jsem v příloze marně hledal zdrojový kód, který byl využit k vykreslení grafů v kapitole 6 – toto značně ztíží práci komukoliv, kdo by se pokusil výsledky práce replikovat.

Formální a jazyková úroveň, rozsah práce **B - velmi dobře**

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.

Text práce je gramaticky na velmi dobré úrovni. Práce má očekávaný rozsah, je celkově vhodně rozčleněna a dobře se čte. Jedinou výjimkou je občasné kolísání formálního značení a terminologie, které se objevuje zejména v první polovině:

1. Na str. 1 a 17 je uvedeno *„space state“* místo *„state space“*.
2. V podmínce (1) formulace ILP 2.3 je uvedeno „∀x“ přestože místo „x“ v podmínce figuruje proměnná „x_{jt}“. V podmínkách (1), (2), (3) a (5) tohoto ILP jsou dále výskyty „x_{tj}“ a „x_{jt}“ volně zaměňovány.
3. V alg. 3 je proměnná *l* nastavena *„argmin(p_j)“*. O odstavec níže je ovšem tento proces popsán jako výběr úlohy s *„longest processing time“*.
4. V alg. 5 jsou obrazy *q* odkazovány jako *„q(S)“*, *„q_i“* a *„q_{Ji}“*. Rozdíl v tomto značení není přímo patrný. Dále algoritmus ověřuje na 4. řádku podmínku *„j = 0“*, která nikdy nenastane, protože proměnná *j > 0*.
5. Mezi citovanými tvrzeními 10, 11 a 12 není harmonizováno značení implikace (tvrzení 10 používá *„if ... then“*, zatímco tvrzení 11-12 používají operátor *„⇒“*).
6. Asymptotické složitosti na konci sekce 4.7 ve třech případech postrádají uzavírající kulatou závorku.
7. V alg. 6 je použit programátorský operátor rovnosti *„==“* přestože např. v alg. 5 podobnou roli zastává matematický operátor rovnosti *„=“*.
8. Na straně 24 je uvedeno *„DFS nature of the algorithm is assured by the use of a first in, first out queue“*.
9. Zatímco alg. 1-6 používají běžné značení množin, v alg. 7 jsou pro zdánlivě stejný účel využity jak složené {} tak hranaté závorky []. O několik odstavců níže autor navíc uvádí „it is preferred first to explore the node where the job is on time“, což není konzistentní s pseudokódem, kde na řádku 24 je nejprve prozkoumán stav, kdy je úloha zpožděna.
10. V sekcích 5.2, 5.4 je vektor pravděpodobností (ne nutně hodnot 0, 1) označován jako *„one-hot encoding“*.

Tyto nesrovnalosti považuji za drobné překlepy, které nejsou konzistentní s celkovým směřováním práce. Má domněnka je podpořena i pozorováním, že žádné z výše uvedených bodů (obzvlášť č. 8) nelze identifikovat ve zdrojovém kódu přiložené implementace.

Výběr zdrojů, korektnost citací **A - výborně**

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.

Autor čerpal z přiměřeného množství zdrojů, kromě aktuální odborné literatury se odkázal v kapitole 3 i přehledovou publikaci. Formát bibliografických referencí odpovídá normě. Výsledky práce a převzaté prvky jsou vhodně odlišeny a snadno rozpoznatelné.

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.

Předloženou práci považuji za velmi kvalitní. Pozitivně hodnotím, že po vyhodnocení výsledků autor v experimentální sekci navrhnul a otestoval alternativní přístup v podobě heuristického algoritmu, který využívá vybudovaného modelu. Za diskutabilní považuji optimalizaci jednotlivých hyperparametrů nezávisle na ostatních, viz otázky níže.

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.

Níže uvádím několik otázek, které by mohly přispět k diskusi u obhajoby:

1. V sekci 5.1 uvádíte, že ke každé simulované instanci rozvrhovací úlohy ukládáte pouze jedno optimální řešení. Jakým způsobem lze tento postup odůvodnit? Jsou-li k dispozici dvě nebo více optimálních řešení, jakým způsobem je uložené řešení vybráno?
2. Podle jakých kritérií bylo vybráno 26 složek charakteristického vektoru, který popisujete v sekci 5.2? Bylo možné po naučení určit, kterým složkám vektoru byly přiřazeny nejvýznamnější váhy?
3. Jak si můžeme být jistí, že nedošlo k přeučení modelu (overfitting)?
4. Výsledky optimalizace hyperparametrů naznačují, že některé součásti architektury neuronové sítě (např. *feature extraction, embedding, multi-head attention*) nemají na kvalitu výstupu významný vliv. Lze odhadnout výpočetní složitost těchto součástí? Jak moc bychom neuronovou síť zrychlili, pokud by tyto součásti byly z výpočtu vyloučeny?
5. V tabulkách 6.20, 6.21 a 6.29 uvádíte průměrné časy běhu algoritmů. Přestože zmiňujete použitý hardware, není z textu zcela jasné jakým způsobem byly tyto hodnoty naměřeny. Bylo měření provedeno jednou či opakovaně? Byla naměřena každá instance zvlášť či vše dohromady?
6. Optimalizaci hyperparametrů provádíte nezávisle na sobě, tj. všechny hyperparametry až na jeden zafixujete a zkoumáte vliv změny volného hyperparametru na přesnost modelu. Jak v tomto režimu zajistíte konvergenci k lokálnímu optimu? Jakým způsobem se tento přístup vypořádá se sedlovým bodem?

Za předpokladu, že autor dovede uspokojivě zodpovědět výše zmíněné otázky, předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně.**

Datum: 1. září 2022 Podpis: