POSUDEK OPONENTA ZÁVĚREČNÉ PRÁCE

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	End-to-end řízení F1/10 autonomního auta s využitím neuronových sítí
Jméno autora:	Žuffa Marek
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Department of Computer Science
Oponent práce:	Paplhám Jakub
Pracoviště oponenta práce:	Skupina vizuálního rozpoznávání, Na Zderaze 269/4 Praha, G-105

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání

mimořádně náročné

Hodnocení náročnosti zadání závěrečné práce.

Ač to není z názvu práce ihned jasné, zadání požaduje využití zpětnovazebního učení. Úspěšné řešení práce proto vyžaduje znalost neuronových sítí, současných state-of-the-art metod zpětnovazebního učení, jejich implementaci a trénování agenta v simulátoru. Tuto část považuji za náročnější, ale nikoliv mimořádně náročnou, jelikož řada implementací je volně k dispozici a student je z oborové specializace Artificial Intelligence.

Práce dále vyžaduje úspěšné nasazení agentů ze simulátoru na reálném modelu auta. Tuto část považuji za mimořádně náročnou z dvou důvodů:

1. Vyžaduje práci s hardwarem a operačním systémem ROS, což ve studiu Otevřené Informatiky není běžné.
2. Vyžaduje nasazení agentů ze simulátoru na reálném modelu. Toto považuji za nejtěžší část zadání. Často vyžaduje bezpečnostní opatření, např. detekci nebezpečných stavů a přepnutí do řízení pomocí klasických metod. Jinak při prozkoumávání stavového prostoru velmi snadno dojde k rozbití fyzického systému.

Celkově zadání hodnotím jako mimořádně náročné.

Splnění zadání

splněno s menšími výhradami

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.

K práci samotné mám výhrady. Všechny body zadání jsou ale splněny, někdy ovšem v minimální přijatelné míře.

Zvolený postup řešení

vynikající

Posuďte, zda student zvolil správný postup nebo metody řešení.

Student měl za úkol použít alespoň jeden model-free a jeden model-based přístup. (Vysvětlení: model-based přístupy explicitně využívají model prostředí. Ten buď může být přesný, nebo naučený agentem z interakce s prostředím.)

Jako model-free metodu student zvolil TD3 (Twin Delayed Deep Deterministic policy gradient, 2018), což je jedna ze standardně používaných metod. Na složitých úlohách se za lepší přístup považují SAC (Soft Actor Critic) nebo PPO (Proximal Policy Optimization), tyto metody ovšem mají více hyperparametrů, jsou výrazně náročnější na implementaci a v případě PPO vyžadují diskretizaci prostoru možných akcí. TD3 považuji za bezpečnou a vhodnou volbu, která by měla stačit k úspěšnému řešení úlohy v simulátoru, jelikož prostor akcí není složitý (kartézský součin dvou intervalů, [-1,1]x[-1,1]).

Jako model-based metodu student zvolil Dreamer (2019/2020). Tato metoda je jedním z prvních úspěšných pokusů použití model-based přístupu. Nicméně na standardních úlohách (benchmark 55 her z konzole Atari) je tento přístup výrazně horší, než jednoduché model-free přístupy. Za vhodnější volbu bych považoval použít pozdější iterace DreamerV2 (2020/2021) nebo DreamerV3 (2023), které „porážejí" běžné model-free přístupy. Předpokládám, že Dreamer byl zvolen jelikož již byl na stejné úloze úspešně použitý, Latent Imagination Facilitates Zero-Shot Transfer in Autonomous Racing, Brunnbauer et al., 2022. Ač by použití DreamerV2/V3 mohlo dosáhnou lepších výsledků, volbu DreamerV1 nepovažuji za chybnou.

Celkově zvolený postup považuji za vynikající. Zvolené metody odpovídají zadání.

Odborná úroveň **C - dobře**

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

Student v rámci závěrečné práce prokázal schopnost dohledat informace v odborné literatuře, implementovat pokročilé algoritmy a nasadit je na reálném modelu auta. S matematickými vyjádřeními v práci nejsem spokojen. Místy jsem v práci narážel na části, kvůli kterým jsem pochyboval, zda student skutečně chápe metody, které používá. Např: „*In continuous space, the Q-values cannot be computed directly. Solving the optimization problem can be difficult and computationally too expensive.*" (str. 25) Spočítat Q-hodnoty přímo lze, ale zjistit pro kterou akci je Q-hodnota maximální přímo nelze (jelikož akce jsou spojité). Je možné, že student problematice skutečně rozumí. Je to ale bohužel skryto za neuceleným, těžko čitelným textem s řadou chyb, viz komentáře v **Formální a jazyková úroveň, rozsah práce**.

Teorie na které jsou použité algoritmy založeny je netriviální. Znalosti získané během studia student prokazuje tím, že je schopen tyto metody používat. Kvůli chybám v rovnicích a místy nejasným formulacím však nemohu odbornost práce hodnotit lépe.

Formální a jazyková úroveň, rozsah práce **D - uspokojivě**

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

TLDR:
Práce je psána formálním jazykem, ve většině případů srozumitelnou angličtinou. V některých případech není text pochopitelný. Místy se objevují překlepy nebo je text nevhodně formulován. Kapitola 2 (Literature Review) by mohla být lépe strukturována. V kapitole 3 (Problem Statement) je definován partially observable Markov decision process (POMDP), ale není definována úloha. POMDP je model prostředí, ale co s ním chceme dělat? Student zápasí s matematickou notací. Odkazuje na články s netriviální teorií, kde každý článek používá trochu jinou notaci. Student pak mezi notacemi článků volně přechází, někdy i uprostřed jednoho algoritmu. V rovnicích často chybí závorky nebo escape symbol \. Řada symbolů není definována. Ač mají tyto symboly standardní interpretaci, bylo by vhodné buď odkazovat na nějaký text a říct, že notace je identická, nebo poskytnout čtenáři seznam symbolů a jejich význam. *Detailní komentáře jsou k dispozici níže.*

Chyby v matematických vyjádřeních:

- Rovnice (4.3): r_t^{prog} není definované. Domnívám se, že se jedná o výraz (5.21).
- Strana 18, pod (4.5): Místo c_d je cd
- Strana 20, nad (4.7): [...] with the current estimate of current state $V(S_t)$
  - $V(S_t)$ je odhad state-value funkce, nikoliv odhad stavu. Ač se jedná o standardní notaci, měla by být definovaná.
- Strana 20, Algorithm 1: Skutečná state-value funkce v_π není definovaná. Policy π není definovaná. Ač se jedná o standardní notaci, mělo by to být definované, nebo poskynut odkaz např. na [*Sutton, Barto. Reinforcement learning, An Introduction]* a řečeno, že se používá stejná notace jako v této knize. Algoritmus sice tuto knihu správně cituje, ale bylo by vhodné v kapitole 3 (Problem Statement) na knihu odkázat s tím, že notace bude identická. V kapitole 3 se definuje partially observable Markov decision process, ale nedefinuje se jak s tímto procesem policy interaguje nebo jakou roli hraje value funkce.

- Strana 20, kapitola 4.5.4: The On-policy algorithms optimize a chosen policy q^π
  - Toto není policy, ale state-action-value funkce. Policy je π
  - Obdobně pro q^* o pár řádků níže
- Strana 21: This algorithm (Alg. 2) has proven optimality in the infinite horizon [37].
  - Zde si nejsem jistý, zda ja algoritmus optimální. Algoritmus je konzistentní estimátor optimální action-value funkce Q – myslím, že to je to co student myslí. Je ale možné, že je tento algoritmus asymptoticky optimální ve smyslu „big-O“ notace.
- Strana 26, Algorithm 3:
  - Řádek 9: μ_θ je funkce, měla by dostat argumenty.
  - Řádek 12: s_{t+1} místo $s_t + 1$
  - Řádek 13: s_i, a_i, r_{i+1} místo s, a, r nebo s' místo s_{i+1}
  - Řádek 16: Obdobně s řádkem 13, buď používat notaci se „spodním indexem“ nebo s „prime“, ale nepřepínat mezi nimi v rámci jednoho algoritmu.
  - Řádek 21,22: ρ je definováno až za algoritmem, za rovnicí (5.5)
- Strana 27, rovnice (5.3):
  - Q_{ϕ} místo Q_{phi}
  - Maximum v rovnici je přes a_{i+} ale ve výrazu je jen a'
- Strana 28, Algorithm 4:
  - Řádek 1: $Q_{\theta_{1}}$ a $Q_{\theta_{2}}$ místo $Q_{\theta 1}$ a $Q_{\theta 2}$
  - Řádek 11: Q je funkce, měla by obdržet argumenty. Zde je důležité, ale vůbec není jasné, jaké dostane.
- Strana 29, rovnice (5.10): ϕ místo phi
- Strana 30, nad rovnicí (5.12): Chybí závorka uzavírající $p(a_t|o_{\leq t}, a_{<t})$
- Strana 32, rovnice (5.15): Vypadá to jako nějaká varianta „generalized H-step truncated lambda-return“, ale nevím jaká. Význam symbolů v rovnici není vysvětlený.
- Strana 32, pod rovnicí (5.15): k není nikde definované a v žádné rovnici se nevyskytuje.
- Strana 33, rovnice (5.20): Druhý argument KL divergence je mimo závorku. Chybí || na oddělení argumentů.

Ukázky kde není text pochopitelný:

- *This phase uses Offline Reinforcement Learning with Implicit Q-Learning (IQL) to extract a critic for a readily available, diverse offline dataset collected on a different robot, using a similar task objective: goal-directed velocity toward check-points selected from a mix of future states and random points in space [21]. After that, the critic is discarded, as only the image encoder is extracted.* (str. 6)
  - Nerozumím tomu co je zde encoder. Nejprve se extrahuje critic, pak se zahodí, a zůstane encoder?
- *A slightly more complex reward can be seen in [13], where the rewards are based on the distance to the optimal race-line states to minimize lap times.*
  - Není jasné co jsou „optimal states“.
- *While in the original paper, the representation model is implemented as a combination of a Convolutional Neural Network (CNN) and a Recurrent State Space Model (RSSM), in [17], it is implemented as a Multi Layer Perceptron (MLP).*
  - Z tohoto textu se zdá, že celý model je MLP. Ve skutečnosti je CNN nahrazeno MLP, ale stále se jedná o RSSM

Drobné chyby (chybějící slova, překlepy, ...):

Legenda:

- [...] : Text pokračuje, ale je vynechán.
- **Tučné**: Zvýraznění problematické části věty.
- *Kurzíva:* Citace textu z práce.
- (str. X): Strana v textu.

Ukázky:

- *This leads to less computing in one step.* (str. 3)
- *[...] doing not plausible actions.* (str. 4)
- *They modified the network by not only predicting the actions in states but rather planning the whole trajectory with*

*an **expectancy value**.* (str. 5)
  - Nerozumím tomu, co je „expectancy value“. Pokud „expected“, tak nerozumím k čemu se vztahuje.
- ***An** F1TENTH platform [1] is a popular autonomous vehicle control research choice. (str. 7)*
  - „The“
- *Also, it has been stated that reinforcement learning converges to better results **as the immediate reward and environment and action space** [26].* (str. 17)
  - Nejspíš chybí část věty.
- ***One of the conditions is** that in continuous action space, the function $Q* (s, a)$ is differentiable with respect to the action argument [10].* (str. 26)
  - Touto větou začíná podsekce. Jedna z podmínek čeho?
- *In [40], **authors the issues**, “A common failure mode for DDPG is that the learned Q-function begins to dramatically overestimate Q-values, which then leads to the policy breaking because it exploits the errors in the Q-function.”* (str. 28)
  - Nejspíš chybí slovo.
- *The **likelihood** of repeating updates with respect to an unchanged critic is limited by sufficiently delaying the policy updates.* (str. 29)
  - Osobně chápu „likelihood“ jako $p(x; \theta)$, zde by bylo lepší použít slovo „probability“.
- *As **this algorithm is used in this thesis**, the implementation from the Stable Baselines 3 [42] library is utilized.* (str. 29)
  - Jelikož touto větou začíná kapitola, bylo by lepší říct: „As the algorithm from Section X.Y is used […]“ nebo „As TD3 is used […]“
- ***Value** of the imagined trajectories […] **have** to be estimated […]* (str. 32)
  - „Values“, nebo „has“
- *The variational **lover** bound […]* (str. 33)
  - „lower“
- *Both **D3** and Dreamer agents […]* (str. 51)
  - „TD3“

Výběr zdrojů, korektnost citací **C - 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.

TLDR:
Výběr pramenů je správný, zdroje jsou relevantní. Převzaté výsledky jsou od vlastních dostatečně odlišeny. Bibliografické citace nejsou plně v souladu s citačními zvyklostmi a normami, ve většině jsou však správné. Místy citace chybí, případně je citován špatný zdroj. V jednom případě chybí v citaci publisher/konference. Detailní komentáře jsou k dispozici níže.

Chybějící citace: (v závorce strana v práci)
DonkeyCar (1), SAC, D4PG, PPO (3), A2C, LSTM (4), Roborace Simulator, SVL Simulator, Carla, (6), ROS2 (11), DQN (27)

Chybné citace:
DDPG (str. 3) cituje implementaci algoritmu od OpenAI. Vhodnější by bylo citovat článek, který metodu navrhuje:
Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N.M., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). Continuous control with deep reinforcement learning. CoRR, abs/1509.02971.

[31] Felipe Codevilla et al. Exploring the limitations of behavior cloning for autonomous driving. 2019.
- Chybí konference / publisher. Článek je z ICCV 2019.

[36] Wah Loon Keng Laura Graesser. Foundations of Deep Reinforcement Learning: Theory and Practice in Python. Addison-Wesley Data & Analytics Series. Addison-Wesley, 2020. isbn: 0135172381; 9780135172384. url: libgen.li/file.php?md5=202217fd5fd079d3910c39a6db4d0598.
- Volný přístup k informacím považuji za důležitý, i přesto ale citovat libgen nepovažuji za vhodné.
- Libgen je stínová knihovna poskytující bezplatný přístup k akademickým článkům, knihám a dalším písemným materiálům, často obcházející autorská práva.

Diskutabilní:
Res**t**Net18 (str. 5) – cituje se článek, který používá ResNet18, ne původní „ResNet článek“. Ač již má ResNet článek citací dost, považoval bych za vhodné citovat ten. Dále je zde překlep RestNet namísto ResNet. Na podobné překlepy v citování je dobré dát si pozor, jelikož existují např. ResNext, ResNest, Res2Net, ResNetD, WResNet.

[22] Axel Brunnbauer and Luigi Berducci. racecar_gym. Version 0.0.1. url: https://github.com/axelbr/racecar_gym.
- U citací webových stránek je vhodné mít "Accessed: yyyy-dd-mm". Např. u citace [1] toto je "Accessed: 2024-30-04".

[25] Mark Towers et al. Gymnasium. Mar. 2023. doi: 10.5281/zenodo. 8127026. url: https : / / zenodo . org / record / 8127025 (visited on 07/08/2023).
- Formátování "visited on 07/08/2023" není konzistentní s "Accessed: yyyy-dd-mm" u jiných citací webových stránek.

[45] Tier IV. AWSIM. https://tier4.github.io/AWSIM/.
- Chybí "Accessed: yyyy-dd-mm".

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.

Viz **CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE.**

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.

Cílem práce bylo použít zpětnovazební učení k řízení modelu auta v simulátoru. Následně měl být natrénovaný agent přenesen na reálný model auta. K dispozici byly tři tratě v simulátoru a jedna fyzická. Natrénovaní agenti avšak nefungovali dobře ani v simulátoru – ani na tratích, na kterých byli učeni (např. Figure 7.14, str. 49). Sám student píše: „Neither of the agents managed to finish laps on any other track than the one they had been trained on.“ (str. 43) Ve výsledcích ale uvádí: „We have successfully implemented, trained, and evaluated two Reinforcement Learning agents.“ (str. 52) „Both algorithms used a reward signal defined in Section 5.3 to promote faster driving while not compromising safety.“ (str. 52) Nesouhlasím s tvrzením, že byli agenti úspěšně natrénováni. Některé z uvedených důvodů, které způsobily problémy s učením jsou: „The implementation from Stable Baselines 3 [42] did not work properly, and in both training and evaluation, it often returned actions that led to crashes.“ (str. 52) „A more accurate simulator, like Project Chrono [44] or AWSIM [45], should help with minimizing the sim-to-real gap.“ (str. 55) Nemyslím si, že by problémem byl nepřesný simulátor. Pokud algoritmy nedokázali úspěšně naučit agenta v jednoduchém simulátoru, v složitějším simulátoru by byl problém ještě značnější. Pokud byl problém s implementací TD3 v Stable Baselines, proč nepoužít vlastní implementaci? TD3 není implementačně složité.

Přesto že zadání práce je ve všech bodech splněné, nejsem s finální podobou této práce spokojen. Body zadání jsou často splněny v minimální přijatelné míře. Věřím, že textu mohlo být věnováno více času, experimenty mohly být rozšířeny a lépe popsány. Věřím, že s několika týdny intenzivní práce by tato práce mohla být hodnocena A – výborně, potenciál prokazuje. V současné podobě jsem se rozhodoval mezi stupni C – dobře a D – uspokojivě. Jako finální stupeň jsem zvolil D – uspokojivě. V hodnocení přihlížím k vysoké časové náročnosti nasazení agentů na reálném modelu auta. Pokud by byl v práci použit pouze simulátor, zvažoval bych horší klasifikační stupeň.

Otázky:

1. Proč jste zvolil jako reprezentaci stavu pouze sken z lidaru? Jakým způsobem pak může agent poznat, zda se vysokou rychlostí blíží k zatáčce a má brzdit, nebo zda je zastavený před zatáčkou a má se začít rozjíždět?
2. Jakým způsobem by jste mohl rozšířit stav agenta tak, aby : A) byl jako senzor i nadále použit pouze lidar, a zároveň B) agent měl k dispozici informaci (alespoň nějakou) o své rychlosti?
3. Agenti v simulátoru nefungovali obzvlášť dobře. Simulátor poskytuje údaje *pose*, *velocity*, *acceleration*, *lidar*, *rgb_camera*. Vy jste používal pouze lidar, jelikož ten je k dispozici na fyzickém modelu auta. Zkoušel jste zda se agenti úspěšně naučí projíždět v simulátoru, když mají k dispozici všechny tyto informace? Zkoušel jste, které senzory by byly potřeba pro úspěšné projíždění okruhů?
4. V práci jsem nenalezl žádné podrobnosti o učení agentů. Jaké nastavení hyperparametrů jste použil? Jaký jste použil optimizer? Jak jste nastavil počáteční hodnotu parametrů modelu? Jak jste nastavil počáteční hodnotu parametrů u dvou critic sítí pro TD3?
5. V práci jsem nenalezl podrobnosti o tom, jak vypadali neuronové sítě, které jste použil. Popište architekturu použitých modelů, s konkrétními velikostmi vrstev.
6. Ve textu se odkazujete na implementaci článku [Latent Imagination Facilitates Zero-Shot Transfer in Autonomous Racing, Brunnbauer et al., 2022]. Pokud se nemýlím, používáte stejný simulátor jako tento článek i stejnou implementaci učícího algoritmu. V čem se Vaše práce od tohoto článku odlišuje? Cynický pohled na práci by mohl být, že se jedná o pokus zopakovat výsledky tohoto článku.
7. Jak interpretujete rozdílné běhy učení TD3 ve Figure 7.11? S jakým nastavením hyperparametrů byly experimenty zkoušeny? Proč jste po „pádu“ TD-crashed nespustil učící algoritmus dál z uloženého checkpointu modelu?

8. V Figure 7.3, 7.7, 7.11, 7.15 je kolem „čar“ také „průhledná plocha“. Co tato plocha znázorňuje? Jedná se o min/max hodnotu nebo standardní odchylku mezi více běhy učení? Pokud ano, kolik takových běhů jste spouštěl?
9. Jaké rady by jste měl pro nasazení agentů naučených v simulátoru na reálném modelu (auta či jiném)?
10. Která část práce byla pro Vás nejobtížnější? Pokud by jste mohl změnit něco ve Vašem postupu, co by jste změnil? Jaké znalosti/dovednosti, které jste získal během psaní práce si nejvíce vážíte?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **D - uspokojivě.**

Datum: 5.6.2024 Podpis: