Dr hab. Bartosz Zieliński, prof. UJ
Wydział Matematyki i Informatyki
Uniwersytet Jagielloński
bartosz.zielinski@uj.edu.pl

Kraków, 3 lutego 2024 r.

Recenzja rozprawy doktorskiej
mgra Konrada Czechowskiego
zatytułowanej:
Deep Learning for Planning and Reinforcement Learning

1. Problem badawczy i jego znaczenie

W pracy doktorant rozważa integrację metod głębokiego uczenia z algorytmami planowania i uczenia ze wzmocnieniem. Proponuje on autorskie metody, a także dokonuje ich ewaluacji poprzez eksperymenty komputerowe. Opracowane metody są niewątpliwie nowatorskie i mają pozytywny wkład w rozwoju metod uczenia ze wzmocnieniem.

2. Wkład autora

Celem pracy było zaproponowanie szeregu metod związanych z algorytmami planowania i uczenia ze wzmocnieniem. O ile zakres oraz zawartość rozprawy oceniam pozytywnie, to zwrócić należy uwagę, że teza rozprawy nie została sformułowana, a zamiast niej przedstawiono ogólną motywację pracy sformułowaną następująco (tłumaczenie własne) "Obecne algorytmy uczenia głębokiego są wciąż dalekie od inteligencji wykazywanej przez ludzi lub zwierzęta. Dlatego, w tej pracy badane są metody algorytmiczne, integrujące dwie kluczowe zdolności potrzebne do zatarcia tej różnicy: zdolność do interakcji z otoczeniem i samodoskonalenia na podstawie zebranych danych oraz zdolność do tworzenia, oceny i realizacji planów."

Uwaga o „zatarciu różnicy między uczeniem głębokim a inteligencją człowieka" jest ryzykowna i może budzić wątpliwości u specjalistów z zakresu kognitywistki. Do tego, motywacja taka jest niezwykle ogólna. Zamiast niej, powinna pojawić się teza pracy, w której należałoby zawrzeć jakimi pozytywnymi cechami (najlepiej ilościowymi) będą się odznaczały zaproponowane metody w stosunku do metod znanych z literatury uczenia ze wzmocnieniem.

Z uwagi na niestandardowo dużą liczbę pierwszych autorów z równorzędnym wkładem (średnio trzech) i ogólnikowość sekcji „Significance and Author Contributions Summary", precyzyjna ocena wkładu autora do poszczególnych publikacji była niemożliwa. Dlatego, na moją prośbę przygotował on dokument, w którym doprecyzowuje swój wkład (**dokument ten załączam do niniejszej recenzji**).

Biorąc pod uwagę informację z tego dokumentu, do najważniejszych osiągnięć doktoranta prezentowanych w rozprawie zaliczam:

- **P1:** Wkład w implementację eksperymentów metody Simulated Policy Learning (SimPLe). SimPLe jest kompletnym algorytmem pasywnego uczenia ze wzmocnieniem (ang. model-based), opartym na modelach przewidywania wideo. Algorytm ten jako pierwszy uzyskiwał satysfakcjonujące wyniki na testach ALE (ang. Atari Learning Environment), przy znacznie mniejszej liczbie interakcji niż w przypadku algorytmów aktywnego (ang. model free). Był on wielokrotnie cytowany.
- **P2:** Zaproponowanie metody Trust-But-Verify (TBV), jej częściowa implementacja i nadzorowanie eksperymentów. Metoda TBV używa estymacji niepewności modelu do kierowania eksploracją.
- **P3:** Wkład w zaproponowanie metody Shoot Tree Search (STS), przeprowadzenie jej eksperymentów i przygotowanie lematu matematycznego. Metoda STS rozszerza algorytm Monte Carlo Tree Search (MCTS) poprzez wprowadzenie wieloetapowego rozrostu (ang. multi-step expansion), ukierunkowanego na przeszukiwaniu w głąb (ang. Depth-First Search).
- **P4:** Zaproponowanie tematyki badawczej i metody Subgoal Search (kSubS), wraz z nadzorowaniem eksperymentów. Metoda kSubS inspirowana jest procesem myślowym człowieka, polegającym na przechodzeniu pomiędzy powiązanymi ze sobą ideami. Przechodzenie to nie odbywa się jednak za pomocą pojedynczych akcji, lecz z użyciem podcelów (ang. subgoals), złożonych z kilku akcji. Pracę tę uznaję za największe osiągnięcie doktoranta z uwagi na znaczny wkład w jej powstanie, bycie pierwszym autorem i rangę konferencji.
- **P5:** Zaproponowanie tematyki badawczej i wkład w zaproponowanie metody Adaptive Subgoal Search (AdaSubS), wraz z nadzorowaniem eksperymentów. Metoda AdaSubS rozszerza metodę kSubS o automatycznie dobieraną długość podcelów. Tę pracę uznaję za drugie największe osiągnięcie doktoranta z uwagi na znaczny wkład w jej powstanie i rangę konferencji.
- Wnikliwą ocenę zaproponowanych metod na drodze eksperymentów komputerowych.

Wyniki uzyskiwane w trakcie pracy nad rozprawą zostały zawarte w materiałach konferencyjnych dwóch bardzo dobrych konferencji naukowych, ICLR (CORE A*) i NeurIPS (CORE A*), a także konferencji IJCNN (CORE B), które odbyły się w latach 2020-2023.

3. Poprawność

Praca bazuje na wspomnianych artkułach autora, opublikowanych na renomowanych konferencjach i stanowi przewodnik po wspomnianych publikacjach. Sekcja 2 rozprawy zredagowany jest starannie i wprowadza czytelnika w podstawowe zagadnienia związane z rozważaną tematyką. Jednakże Sekcja 3.1, opisująca wkład autora, pozostawia wiele do życzenia. Paragraf „Scientific Contributions Summary" miał opisać każde osiągnięcie za pomocą jednego zdania, ale niestety zdania te są niegramatyczne i ciężkie w odbiorze. Kolejny

paragraf „Significance and Author Contributions Summary" powinien zawierać szczegółowy opis wkładu doktoranta, jednak opis ten jest zbyt ogólnikowy. Po tym następują Sekcje 3.2-3.6, opisujące kolejne artykuły. Opisy te są jednak niedopracowane i brak im przejrzystej struktury. W rezultacie, recenzent był zmuszony bazować prawie wyłącznie na tekstach artykułów podczas oceny metod i narzędzi ich ewaluacji opartej na eksperymentach komputerowych.

Lektura rozprawy prowadzi do sformułowania poniższych pytań:

- Jaką jest teza pracy?
- Czy porównanie systemu drugiego z pozycji [13] do mechanizmu planowania jest uzasadnione? Jak zdefiniowano system drugi z pozycji [13] i jakie są przesłanki oraz obiekcje do utożsamiania go z mechanizmem planowania?
- W ograniczeniach pracy P5 napisane jest „Our algorithm is not guaranteed to find a solution. We found it is not problematic in practice." Dlaczego nie jest to problematyczne w praktyce?
- Jak wygląda kwestia rzeczywistych zastosowań prezentowanych metod, np. w autonomicznych robotach mobilnych? Czy prezentowane systemu mogą być wykorzystywane przy użyciu wbudowanych kart GPU, tj. Jetson Nano?
- Czy prezentowane algorytmu są interpretowalne? Do jakiego stopnia jesteśmy w stanie sprawdzić przesłanki stojące za daną akcją?
- Jakie są dalsze kierunki rozwoju w obszarze związanym z rozprawą?

4. Wiedza kandydata

Na podstawie lektury uważam, że doktorant posiada ugruntowaną wiedzę z zakresu informatyki, w szczególności w zakresie algorytmów planowania i uczenia ze wzmocnieniem. Doktorant posługuje się poprawnie zaawansowanym aparatem matematycznym, a także potrafi zaplanować i przeprowadzić eksperyment komputerowy w celu oceny jakości zaproponowanych metod.

Przegląd literaturowy dotyczący zagadnień przedstawionych w rozprawie pozwala stwierdzić, że doktorant posiada aktualną wiedzę z zakresu tematyki rozprawy, a także potrafi dokonać krytycznego przeglądu źródeł w celu wskazania ciekawych kierunków badań. Zawarty w dysertacji spis źródeł literaturowych, zawierających 59 pozycji, jest aktualny i kompletny oraz uzupełniony o pozycje cytowane w dołączonych do rozprawy artykułach.

5. Podsumowanie

Doktorant wykazał się w recenzowanej rozprawie właściwie stosowanym podejściem analitycznym i eksperymentalnym oraz dobrą znajomością aktualnej problematyki związanej algorytmami planowania i uczenia ze wzmocnieniem. Zostało to poparte bardzo dobrymi studiami literaturowymi, obejmującymi aktualne piśmiennictwo związane z problematyką rozprawy, co świadczy o bardzo dobrej wiedzy doktoranta z tego zakresu. Doktorant

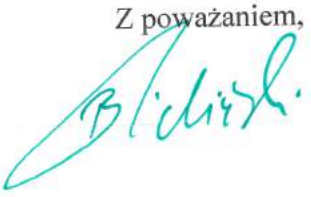
przeanalizował wyniki przeprowadzonych badań eksperymentalnych, a także ocenił jakość zaproponowanych metod na tle algorytmów znanych z literatury.

Recenzowana dysertacja przedstawia rozwiązanie ważnego i oryginalnego problemu, wzbogacając wiedzę dotyczącą algorytmów planowania i uczenia ze wzmocnieniem. Zawarte w niej wyniki badań eksperymentalnych wskazują również na możliwość wykorzystania otrzymanych metod w praktyce.

Przedstawione w punkcie 3 recenzji uwagi mają jedynie charakter dyskusyjny i w żaden sposób nie deprecjonują osiągniętych przez doktoranta rezultatów oraz nie wpływają na pozytywne wrażenie o przedłożonej rozprawie.

Dlatego, biorąc pod uwagę powyższe stwierdzam, że rozprawa mgra Konrada Czechowskiego pt. „Deep Learning for Planning and Reinforcement Learning" **spełnia wymagania** stawiane pracom doktorskim, dlatego wnoszę o jej przyjęcie i dopuszczenie mgr Konrada Czechowskiego do publicznej obrony.

Z poważaniem,



ZAŁĄCZNIK DO RECENZJI.

Konrad Czechowski contributions summary

P1 Ł. Kaiser*, M. Babaeizadeh*, P. Miłoś*, B. Osiński*, R. Campbell, **K. Czechowski**, D. Erhan, C. Finn, P. Kozakowski, S. Levine, A. Mohiuddin, R. Sepassi, G. Tucker, H.Michalewski *Model-based Reinforcement Learning for Atari* International Conference on Learning Representations (ICLR) 2020 awarded with **spotlight presentation**.

My contributions:

- Significant contribution to implementation of experiments - refactoring, debugging, and extending the core implementation of SimPLe, see https://github.com/tensorflow/tensor2tensor/commits?author=konradczechowski for details. (this was a complex project from engineering perspective, this contribution alone required months of full time engagement)
- Contribution to data analysis and presentation of results (including preparation of Tables 2, 3, and 4)
- Creation of an internal tool for investigation of SimPLe environment model quality used for debugging and published analysis (see below)
- Qualitative analysis of SimPLe interactions with the environments (including preparation of most of the examples described in section B of the appendix)
- Implementation of experiments replacing the inner loop RL algorithm (using Q-learning instead of PPO)

I estimate my contributions in the publication to 5%. (This was a large two year project, now with >800 citations)

P2 K. Czechowski*, T. Odrzygozdz*, M. Izworski*, Marek Zbysiński, Ł. Kuciński, and P. Miłoś *Trust, but Verify: Alleviating Pessimistic Errors in Model-Based Exploration.* International Joint Conference on Neural Networks (IJCNN) 2021.

I led the research effort in project which ended with this publication. This includes following contributions:

- I performed literature overview and research planning over the course of entire project.
    - Including prioritisation and coordination of work.
- I designed main algorithm (and partially implemented it)
- I supervised the execution of all experiments (often implemented by other team members)

I estimate my contributions in the publication to 35%.

K. Cz.

P3 K. Czechowski*, P. Januszewski*, P. Kozakowski*, Ł. Kuciński, P. Miłoś *Structure and Randomness in Planning and Reinforcement Learning.* International Joint Conference on Neural Networks (IJCNN) 2021.

- I co-designed the primary STS algorithm (I think I was the person which proposed exact STS formulation, though it was an effect of many internal discussions)
- I designed, implemented and conducted experiments in the Sokoban domain
- I formulated and proved a lemma explaining tree traversal efficiency of STS. (Lemma 3.1 in the PhD thesis)

I estimate my contributions in the publication to 30%.

P4 K. Czechowski*, T. Odrzygóźdź*, M. Zbysiński, M. Zawalski, K. Olejnik, Y. Wu, Ł. Kuciński, P. Miłoś *Subgoal Search For Complex Reasoning Tasks.* Advances in Neural Information Processing Systems (NeurIPS), 2021

I led the research effort in project which ended with this publication. This includes following contributions:

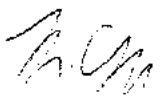
- I performed literature overview and research planning over the course of entire project.
    - Including prioritisation and coordination of work.
- I proposed the scope of research - using the subgoal based planning for combinatorially complex domains
- I designed kSubS algorithm
- I supervised the execution of all experiments published (often implemented by other team members)

I estimate my contributions in the publication to 35%.

P5 M. Zawalski*, M. Tyrolski*, **K. Czechowski***, D. Stachura, P. Piekos, T. Odrzygóźdź, Y. Wu, Ł. Kuciński, P. Miłoś *Fast and Precise: Adjusting Planning Horizon with Adaptive Subgoal Search.* International Conference on Learning Representations (ICLR) 2023 - distinguished as the **top-5%-most-notable publication** (long presentation).

I led the research effort in project which ended with this publication. This includes following contributions:

- I performed literature overview and research planning over the course of entire project.
    - Including prioritisation and coordination of work.
- Literature overview
- I proposed the scope of research (verifier and adaptivity)
- I designed the subgoal verification (implemented by Michał Tyrolski)
- I co-designed the approach for adaptive subgoal search.



    - Specifically, I proposed an initial set of ideas. Together with Michał Zawalski we considered and experimented with many different possible approaches to adaptiveness.
- I supervised the process of standardizing our algorithm, developing a unified formulation that enabled its consistent application across various domains.
- I supervised the standarization of our algorithm formulation across domains. To this end:
    - I proposed a common framework for subgoal generation (reformulated subgoal generation in sokoban domain in way allowing to effectively use beam-search)
    - I supervised evaluation of common approach to traversal between subgoals (replacing local exhaustive search in sokoban with low-level policy rollout)
- I supervised the execution of nearly all experiments published (except for some ablations performed for the ICLR resubmission)

I estimate my contributions in the publication to 25%.