Igor T. Podolak, dr hab.
Wydział Matematyki i Informatyki UJ
Telefon: 607 913 367
E-Mail: igor.podolak@uj.edu.pl

Igor T. Podolak

Igor T. Podolak, dr hab. • Wydział Matematyki i Informatyki UJ

Rada Dyscypliny Matematyki i Informatyki
Uniwersytetu Warszawskiego

Kraków, 14 sierpnia 2023

Recenzja pracy doktorskiej *Efficient methods for machine learning in sequential decision making* pana mgra Błażeja Bogumiła Osińskiego, przygotowana pod opieką promotora dr. hab. Piotra Miłosia, IMPAN, prof. uczelni

1 Tematyka rozprawy i oryginalność

Praca doktorska pana Błażeja Osińskiego jest poświęcona zagadnieniom optymalizacji kosztów przy uczeniu dla problemu planowania sekwencyjnego podejmowania decyzji. Próby rozwiązania takiego zagadnienia muszą złamać ciche założenie w uczeniu maszynowym, że dane mają niezależny i identyczny rozkład. Takie zadanie staje się więc znacznie trudniejsze. Ma to miejsce w szczególności w przypadku uczenia ze wzmocnieniem (*ang.* reinforcement learning, RL), któremu poświęcona jest przedstawiona dysertacja. Optymalizacja kosztów, omówiona w pracy, będzie, w dużym stopniu, polegać na minimalizacji kosztów pozyskania danych dla głębokiego uczenia.

W swojej pracy doktorant przedstawia kilka niezależnych podejść. Są nimi uczenie RL oparte na modelu środowiska, wykorzystanie danych z symulatorów dla problemów automatycznego kierowania samochodów, wykorzystanie uczenia imitacyjnego, w końcu użycie dużych modeli umożliwiające zintegrowania wielu sygnałów dla budowy zintegrowanego środowiska wiedzy dla uczonego modelu.

Rozpatrywane problemy stanowią ważne zagadnienia we współcześnie przeprowadzanych badaniach uczenia maszynowego (*ang.* machine learning, ML). Przedstawione przez doktoranta rozwiązania cechują się wysokim stopniem oryginalności i leżą w centrum aktualnych badań na świecie.

1.1 Wskaźniki bibliometryczne prac cyklu publikacji

Na pracę doktorską pana mgra Błażeja Osińskiego składa się cykl siedmiu artykułów opublikowanych w latach 2020–2022, wraz z wprowadzeniem oraz streszczeniem w języku polskim i angielskim. Tabela 1 zawiera wybrane wartości bibliometryczne oraz ważone przez realnie i uczciwie zadeklarowany udział własny doktoranta. Sumaryczne i ważone wartości w pełni spełniają, w moim przekonaniu, wymagania ustawy Prawo o szkolnictwie wyższym i nauce z 20 lipca 2018, z późniejszymi zmianami, art. 187, pkt. 3 i 4.

Wszystkie prace zostały napisane w kilku- do kilkunasto-osobowych zespołach. Ma to związek z tym, że wiele z nich powstało w trakcie staży (Google Brain, UC Berkeley) czy współpracy doktoranta z zespołami pracujących nad rozwiązaniami dla autonomicznych samochodów (Lyft Level 5, Woven). Wszystkie prace cyklu zostały opublikowane na konferencjach. Pierwsza na ICLR, konferencji (CORE A*, 200 p. MEiN), cztery na ICRA (CORE B, 70 p. MEiN). Dwie prace przedstawiono na CoRL (wartości CORE i MEiN nie są dostępne) o impact score IS rzędu, lub wyższym, temu dla ICRA.

n	nazwa	rok	%	MEiN	CORE	cyt.	IS	h5	$\sum$%h5	$\sum$%IS
1	ICLR	2020	20	200	A*	650	24.75	303	60.60	4.95
2	ICRA	2020	35	70	B	67	4.20	119	41.65	1.47
3	CoRL	2021	15	b.d.	b.d.	30	4.79	76	11.40	0.7185
4	ICRA	2021	5	70	B	5	3.96	119	3.57	0.198
5	ICRA	2021	10	70	B	45	3.96	119	11.90	0.396
6	CoRL	2022	50	b.d.	b.d.	65	4.80	76	38.00	2.40
7	ICRA	2022	5	70	B	22	3.24	119	3.57	0.162
Razem							49.70	931	171.69	10.2945

Tabela 1: Punktacje i cytowania: cyt. – szacowana na koniec lipca 2023 liczba cytowań, IS – Impact Score, h5 – Google Scholar h-5 index. $\sum$%h5 i $\sum$%IS – odpowiednie wartości ważone przez deklarowany udział doktoranta. Numeracja n jest podana jak w spisie we wstępie do dysertacji na stronach 9–10.

Chciałbym zwrócić uwagę, że prace cyklu były, w tym krótkim czasie od momentu ich publikacji, wielokrotnie już cytowane. Pierwsza, przedstawiona na ICLR, już prawie 700 razy, kilka po 60 razy, niektóre niewiele mniej. Stanowi to najlepszy wyznacznik jakości i oryginalności przedstawionych prac cyklu.

Wszystkie prace w cyklu zostały opublikowane w latach 2020–2022. Na pewno pomaga w tym praca zespołowa, a taka jest charakterystyka obecnej nauki, szczególnie w informatyce. Udział doktoranta w międzynarodowych zespołach autorskich stanowi jego istotne osiągnięcie, gdyż świadczy o uznaniu zdolności przez innych naukowców. Także liczba prac i szybkość ich publikacji stanowią o moim uznaniu dla doktoranta.

W trzech podstawowych publikacjach cyklu (pozycje 1, 2, 6 w tabeli 1) deklarowany udział doktoranta zawiera się w przedziale 20–50 procent, jest tam więc, biorąc pod uwagę liczbę autorów, jednym z podstawowych członków zespołu. W pozostałych udział jest mniejszy i stanowi około 10–15 procent. Uważam, że przedstawiony cykl publikacji spełnia wszelkie warunki ustawy dla prac doktorskich.

W czasie studiów doktoranckich pan mgr Błażej Osiński był także współautorem kilku innych publikacji, które nie są częścią przedstawionego cyklu. Współautorstwo doktoranta w nich wskazuje jednak na jego zdolności oraz intensywność pracy.

2 Metodyka i stopień realizacji celów

W pracach cyklu autor zaproponował kilka podejść do optymalizacji uczenia RL przez odpowiednią definicję sposobu uczenia czy dobór danych. Do pracy załączony jest wstęp opisujące i łączący wszystkie podjęte zagadnienia. Poniżej przedstawiam résumé poszczególnych artykułów cyklu wraz z pytaniami, które nasunęły mi się przy lekturze.

2.1 Uczenie RL oparte o model

W pracy *Model-based reinforcement learning for Atari*, ICLR 2020, doktorant przedstawia w jaki sposób oparcie uczenia RL o model pozwala na zmniejszenie liczby koniecznych iteracji, a więc i liczby przykładów dla skutecznego uczenia. Dla budowy modelu kluczowe jest zaproponowane wykorzystanie konwolucyjno-dekonwolucyjnej sieci dla przewidywania kolejnej klatki obrazu dla poprzedza-

jących kilku klatek w grze typu Atari. Takie rozwiązanie staje się równoznaczne z budową modelu dla uczenia RL.

Zaproponowane dodanie modelu dyskretnej przestrzeni ukrytej (*ang.* latent) latent pozwala na jednoczesne przewidywanie nagrody (*ang.* reward) całego modelu i staje się równoległą symulacją otoczenia. Przedstawione rozwiązanie umożliwia na jednoczesne symulowane uczenie się polityki agenta. Taka koncepcja daje możliwość bardzo efektywnego uczenia dla problemu RL a także ograniczenie, dla każdej z gier Atari, liczby interakcji ze środowiskiem do raptem 100 tysięcy dla osiągnięcia poziomu dostępnego dla uczenia RL bez modelu dopiero po wielu milionach iteracji. W pracy wykorzystane zostały bardzo zaawansowane metody, a wszystkie postawione cele zostały zrealizowane z sukcesem.

Zastosowane przez doktoranta podejście pokazuje w jaki, stosunkowo prosty, sposób można zbudować model gry jednocześnie zmniejszając zapotrzebowanie na interakcje ze środowiskiem (według doświadczeń, 5–10× w zależności od gry). Praca zyskała, w krótkim czasie od publikacji, prawie siedemset cytowań. Jednocześnie 100 tysięcy iteracji wyznaczyło nowy standard dla wielu benchmarków.

Uważam, że jest to jedna z najważniejszych prac przedstawionego cyklu, szczególnie przy wysokim udziale doktoranta.

2.2 Zbieranie danych z symulatorów

W pracy *Simulation based reinforcement learning for real-life autonomous driving*, ICRA 2020, doktorant postawił ważne pytanie czy jest możliwe nauczenie systemu automatycznego sterowania samochodem, SDV, w rzeczywistym środowisku przez transfer wiedzy z symulatora? Odpowiedź twierdząca dawałaby wskazówkę w jaki sposób można, stosunkowo niewielkim kosztem, zdobywać duże ilości danych uczących dla procesu RL.

Aby zbadać powyższą hipotezę o możliwości transferu danych, doktorant wykorzystał symulator jazdy CARLA, zmodyfikowany przez dodanie nowych map oraz nowych postrzeganych warunków pogodowych, a także augmentacje widzenia. Wyniki dla uczonej w schemacie RL sieci neuronowej pokazały, że przy wykorzystaniu szeregu dodatkowych rozszerzeń, w szczególności randomizacji sygnałów wejściowych, możliwe było uzyskanie dobrych wyników jazdy, zarówno w symulacji jak rzeczywistych (w doświadczeniach wykorzystano zabezpieczenie w postaci rzeczywistego kierowcy przejmującego sterowanie w sytuacjach zagro-

żenia). Moją uwagę zwróciło szerokie przebadanie w eksperymentach różnych schematów ablacji.

Nasunęły mi się przy lekturze pytania czy taki transfer danych symulowanych nie ogranicza uzyskanych informacji do wiedzy eksperta tworzącego symulator? A także jaka jest relacja jakości danych z symulatora do tych uzyskanych z czujników prawdziwych samochodów?

2.3 Uczenie imitacyjne

Dużą część prac cyklu obejmuje zagadnienie uczenia imitacyjnego (*ang.* imitation learning), w którym agent usiłuje sobie poradzić z rzadkimi sygnałami nagród przez efektywne śledzenie i naśladowanie akcji podejmowanych przez eksperta. Efektywne rozwiązanie tego problemu powinno pozwolić na ograniczenie liczby koniecznych danych. W *What data do we need for training an AV motion planner?*, ICRA 2021, autor zadał ważne pytanie o kompromis pomiędzy jakością a ilością danych dla uczenia SDV. Wyniki pokazują, że zwiększenie ilości danych, nawet niższej jakości, poprawia uczenie w porównaniu do mniejszej liczby danych wysokiej jakości.

W artykule dane z czujnikami o niższej jakości są określone jako "crowdsourced", jednak jako pochodzące od kierowców–ekspertów. Mam pytanie: czy wykorzystanie danych od kierowców będących jedynie "zwykłymi" kierowcami a nie ekspertami pogorszyłoby przewidywania? W artykule, dla testowania wpływu wartości danych, użyte są te same dane jedynie przez symulację obniżenia jakości czujników.

W *SimNet: learning reactive self-driving simulations from real-world*, ICRA 2021, autor sugeruje, że możliwe jest ułatwienie procesu weryfikacji systemów SDV przez wykorzystanie modeli uczenia maszynowego ML jako procesu Markowa. W zaproponowanym systemie zachowanie uczestników sytuacji jest opisane przez wiele modeli ML dla znanego SDV pokazując skuteczność takiego podejścia. Z lektury artykułu nie jest dla mnie jasne czy stany obiektów z_t^i obejmują jedynie położenie, zwrot, prędkość uczestnika i (zarówno obiektu SDV jak i innych uczestników ruchu) w chwili t, czy też, jak w późniejszej pracy *SafetyNet: safe planning for real-world ...*, wszystkie typowe dane czujników tego pojazdu?

Ważną pracą tej części doktoratu o uczeniu imitacyjnym jest *Urban driver: learning to drive from real-world demonstrations using policy gradient*, CoRL 2021, w którym doktorant przedstawia gradientowy model RL uczenia imitacyjnego

opartego na danych od eksperta. Dzięki ciekawemu uczeniu przez algorytm BPTT z ograniczoną sekwencją propagacji wstecz, uzyskiwane są wyniki zdecydowanie lepsze niż te dla porównywanych modeli. W tej pracy, dla pełnego testowania (patrz uwagi do poprzedniej pracy *SimNet: learning reactive self-driving* ...) rozwiązanie zostało zainstalowane na rzeczywistym samochodzie wyposażonym w pełen zestaw czujników. Pozwoliło to pokazać pełną skuteczność proponowanego rozwiązania. Chciałbym podkreślić, że w pracy wyczerpująco przedstawiono wytłumaczenie podstaw teoretycznych proponowanego rozwiązania.

Ostatnim artykułem w grupie prac zajmujących się uczeniem imitacyjnym jest *SafetyNet: Safe planning for real-world self-driving vehicles using machine-learned policies*, ICRA 2022. Opisuje on model predykcji dla SDV z zabezpieczeniem przed błędnymi akcjami, które zrealizowane jest jako dodatkowy model bezpieczeństwa (*ang.* fallback). Po wykryciu niebezpieczeństwa błędnej akcji, sterowanie pojazdem przekazywane jest do prostego systemu regułowego. Uzyskane wyniki pokazują wyraźnie, że użycie warstwy bezpieczeństwa znacznie zmniejsza liczbę ewentualnych kolizji. Przy lekturze nasunęło mi się pytanie czy należy rozumieć, że model ML jest odpowiedzialny jedynie za wyznaczanie trasy, a warstwa bezpieczeństwa całkowicie zabezpiecza przed błędami? Liczba możliwych kolizji bez warstwy zabezpieczającej jest w doświadczeniach stosunkowo wysoka, chociażby porównując z wynikami opisanymi w poprzedniej pracy *Urban driver: learning to drive* .... A więc, czy jeśli model ML będzie lepiej nauczony, to warstwa bezpieczeństwa będzie dalej konieczna? A może w ogóle nie warto uczyć go lepiej by zabezpieczał przed wypadkami?

2.4 Wykorzystanie modeli językowych

Bardzo ciekawe połączenie nowoczesnych dużych modeli językowego, asocjacji i nawigacji przedstawia praca *LM-Nav: Robotic navigation with large pre-trained models of language, vision, and action*, CoRL 2022. Jest to z pewnością praca z wielką przyszłością dla minimalizacji uczenia, przy jednocześnie wysokiej skuteczności, na co wskazują przedstawione doświadczenia. Pytanie, które nasunęło mi się w trakcie lektury, jest związane z ostatnio szeroko toczącą się dyskusją nad dużymi modelami językowymi wskazującą, między innymi, że modele te często konfabulują. Czy zaproponowany model jest na takie przypadki odporny? Czy, w związku z tym, jest skalowalny dla większych otoczeń pracy robota, na przykład zaplanowania podróży samochodowej z Krakowa do Lizbony?

2.5 Podsumowanie cyklu prac

Tematyka prac bardzo dobrze łączy się w cykl pokazując rozwój naukowy doktoranta. Pierwsza z prac pokazuje jego zainteresowanie uczeniem opartym na modelu uczenia ze wzmocnieniem RL, a kolejne artykuły przejście zainteresowań do szczegółowego zastosowania RL w problemie uczenia modeli automatycznego kierowania pojazdów. W tych pracach widać istotne przejście od zagadnień uczenia imitacyjnego, do łączenia grupy wstępnie nauczonych dużych modeli językowych, asocjacji wizji i języka oraz nawigacji co może pozwolić na osiągnięcie uczenia typu *few shot* w złożonych problemach sekwencyjnego podejmowania decyzji, któremu poświęcona jest cała dysertacja. Oceniam to połączenie bardzo wysoko.

Ważnym składnikiem przedstawionej dysertacji jest także prawie 20-stronicowy wstęp. Bardzo dobrze przedstawia on definicję problemu planowania sekwencyjnego, ważne problemy poszczególnych części, istotne pytania związane z poszczególnymi artykułami cyklu, a także plany na naukową przyszłość.

Mogę śmiało stwierdzić, że rozwój naukowy doktoranta jest bardzo dobrze ukierunkowany. Z pewnością temat nie jest wyczerpany i doktorant w przyszłej pracy będzie się tylko dalej rozwijał.

3 Silne punkty i uwagi

Cykl prac jest bardzo interesujący i dobrze łączy wszystkie zagadnienia.

1. Prace cyklu zostały napisane i opublikowane w bardzo krótkim czasie. Nie są to wszystkie opublikowane prace doktoranta. Świadczy to o jego zdolnościach i pracowitości.

2. Wszystkie prace w przedstawionym cyklu zostały napisane w wieloosobowych zespołach, wszystkie międzynarodowych, w silnych zespołach. Prace były zwykle napisane we współpracy z ważnymi uczelniami czy firmami badań naukowych (University of California, Berkeley, Google Brain, Lyft i Woven Planet Level 5), także w badaniach dla korporacji (Volkswagen czy Toyota). Pokazuje to na bardzo wysoki poziom zdolności doktoranta, umiejętność współpracy i docenienie przez międzynarodowe środowisko.

3. Wstęp do dysertacji jest napisany czytelnie i dobrze przedstawia problematykę przedstawionego cyklu prac, jak i plany na przyszłość. Świadczy to

o zdolności doktoranta do samodzielnego prowadzenia badań naukowych, spełniając wymóg zawarty w art. 187, pkt. 1 ustawy.

4. Prace cyklu mają bardzo wysokie liczby cytowań, zasadniczo niespotykaną na tym poziomie rozwoju naukowego, wpływając na rozwój dyscypliny.

5. Wszystkie prace, jak też wstęp, są bardzo dobrze zredagowane, a także poparte głębokim wglądem w literaturę. Zwracają uwagę bardzo dobrze wykonane doświadczenia, które są konieczne w tego typu badaniach.

Jedyne uwagi, które nasunęły mi się przy czytaniu dysertacji są następujące.

1. Wszystkie prace zostały opublikowane na konferencjach co jest naturalne przy obecnym szybkim rozwoju metod. Szkoda jednak, że któraś z nich nie została także opublikowana w którymś z ważnych czasopism.

2. Do pracy *LM-Nav: robotic navigation with large pre-trained ...* nie zostały dodane do składki załączniki, do których są odniesienia w tekście. Ale jest to bez znaczenia, gdyż prace są ogólnie dostępne.

Wszystkie te uwagi nasunęły mi się przy czytaniu dysertacji, i w żadnym stopniu nie wpływają na moją bardzo wysoką ocenę całego osiągnięcia.

4 Konkluzje

Po przeczytaniu dysertacji doktorskiej pana mgra Błażeja Osińskiego mogę, z całą odpowiedzialnością, stwierdzić, że jej zawartość potwierdza wiedzę teoretyczną kandydata w dyscyplinie informatyki i wykazuje jego zdolność do samodzielnego prowadzenia pracy naukowej, co jest zgodne z warunkami ustawy z 20 lipca 2018 Prawo o szkolnictwie wyższym i nauce, art. 187, ust. 1 i 2 (Dziennik Ustaw poz. 742, 2023). Przedstawiona dysertacja doktorska pana mgra Osińskiego pokazuje cały szereg oryginalnych rozwiązań problemów naukowych i zastosowania wyników własnych rozwiązań stanowiąc istotny wkład w rozwój dyscypliny, co jest również zgodne z wyżej wymienioną ustawą.

W związku z tym wnioskuję o dopuszczenie pana mgra Błażeja Osińskiego do następnych etapów postępowania o przyznanie stopnia naukowego doktora.

4.1 Dodatkowe wnioski

Ze względu na wysoką jakość zawartych w przedstawionej dysertacji wyników, bardzo dobre przedstawienie redakcję, zwracam się do Rady Naukowej Dyscyplin Matematyka i Informatyka Uniwersytetu Warszawskiego o jej wyróżnienie. Na poparcie tego wniosku, chciałbym zwrócić uwagę, z jednej strony, na bardzo dużą liczbę cytowań artykułów przedstawionego cyklu, a z drugiej na indywidualne osiągnięcia doktoranta uzewnętrzniające się we współpracy w międzynarodowych zespołach, wielu stażach w ważnych ośrodkach, a także fakt, że pan mgr Błażej Osiński był stypendystą fundacji Fulbrighta.

5 Zakwalifikowanie do dyscypliny

Prace w cyklu są w znacznej większości całkowicie nowymi modelami uczenia maszynowego, i jako takie należą do informatyki. Uważam jej obronę w dyscyplinie Informatyki za właściwą.

Z uszanowaniem,

Igor T. Podolak, dr hab.