Dr hab. Przemysław Spurek, prof. UJ
Uniwersytet Jagielloński
Wydział Matematyki i Informatyki

RECENZJA
wniosku w sprawie nadania
mgr. Konradowi Czechowskiemu

stopnia doktora w dziedzinie nauk ścisłych i przyrodniczych
w dyscyplinie informatyka

Osiągnięciem przedłożonym przez mgr. Konrada Czechowskiego jest zbiór publikacji zatytułowany "Deep Learning for Planning and Reinforcement Learning". Jest to cykl powiązanych tematycznie artykułów naukowych w recenzowanych materiałach z konferencji międzynarodowych. Na cykl ten składa się pięć artykułów naukowych opublikowanych między innymi w materiałach pokonferencyjnych renomowanych konferencji ICLR oraz NeurIPS.

Sumaryczna liczba punktów MEiN przypisanych do publikacji wchodzących w skład cyklu wynosi 880, przy czym punktacja poszczególnych prac waha się w zakresie od 140 do 200, osiągając średnią na poziomie 176 punktów. Prace te doczekały się też dużej ilości cytowań, w szczególności praca [P1] posiada 807 cytowań, [P4] posiada 20 cytowań, [P5] posiada 20 cytowań na podstawie scholar google. Ponadto trzy prace opublikowane są na topowych konferencjach *ICLR i NeurIPS,* w których mgr Konrad Czechowski w pracy [P4] jest pierwszym autorem, a w [P5] trzecim autorem z adnotacją contributed equally. Ponadto prace [P1] [P5] były prezentowane jako *spotlight presentation* oraz *long presentation,* co jest ogromnym sukcesem naukowym. Pan Konrad jest też pierwszym autorem prac z konferencji IJCNN. Należy zauważyć, że dwie prace [P2] i [P3] są obecnie klasyfikowane jako konferencje typu B i posiadają tylko 70 punktów, ale nie wpływa to na ocenę dorobku.

Od strony bibliometrycznej, ocena przedstawionego przez mgr. Konrada Czechowskiego osiągnięcia naukowego jest więc **bardzo dobra**. Konferencje, w których opublikowane zostały prace wchodzące w skład cyklu są uznane w świecie naukowym, a niektóre z nich klasyfikowane są jako najlepsze konferencje w dziedzinie.

Pan mgr Konrad Czechowski nie jest samodzielnym autorem żadnej publikacji wchodzącej w skład cyklu, a prace posiadają od 5 do 14 współautorów. W czterech pracach doktorant jest wskazany jako osoba

o największym wkładzie. **Indywidualny wkład merytoryczny mgr. Konrada Czechowskiego w prace wchodzące w skład cyklu jest wiodący, jasno zdefiniowany i nie budzi wątpliwości.**

Praca doktorska jest dobrze napisana. Autor wprowadza wszystkie pojęcia potrzebne do zrozumienia prac wchodzących w skład cyklu, przejrzyście przedstawiając zagadnienia. **Poszczególne prace tworzą cykl powiązanych tematycznie artykułów.**

Poszczególne publikacje są dobrze napisane merytorycznie i technicznie. Problemy rozważane w poszczególnych tytułach są umotywowane i sformułowane bardzo naturalnie. Natomiast same rozwiązania problemów są intuicyjne i dobrze opisane.

Należy zwrócić uwagę, że prace [P1, P4, P5] są napisane bardzo poprawnie pod kątem publikowania na konferencjach naukowych. We wspomnianych pracach autor wychodzi od punktu dobrze przedstawionej motywacji. Następnie opisuje wprowadzone rozwiązanie. Duże fragmenty pracy dotyczące szczegółowych rozwiązań umieszczone są w Appendix. Ponadto prace zawierają liczne eksperymenty, wykonywane na dużych zbiorach danych i przejrzyście prezentujące przewagę wprowadzanych metod nad rozwiązywaniami referencyjnymi. Sekcje eksperymentalne zawsze stanowią jedną z najmocniejszych stron pracy.

Autor dobrze współpracuje w dużych zespołach badawczych, w których znajduje się kilku młodych naukowców jak i doświadczonych profesorów. Dzięki zaangażowaniu większej liczby osób, prace posiadają znacznie więcej skomplikowanych eksperymentów. Pozwala to doktorantowi publikować prace na najlepszych konferencjach w obszarze sztucznej inteligencji, cechującej się dużą złożonością modeli oraz wymagającymi eksperymentami.

W pracy [P1] autorzy badają, w jaki sposób modele wykorzystujące przewidywanie wideo mogą umożliwić agentom rozwiązywanie zadania polegającego na graniu w gry Atari przy mniejszej liczbie interakcji niż w algorytmach nie wykorzystujących modelu środowiska (model-free reinforcement learning). W pracy zaproponowany został model o nazwie **SimPLe** wykorzystujący modele predykcji wideo.

W zadaniu Atari celem jest znalezienie odpowiedniej polityki. W metodzie **SimPLe** oprócz środowiska (emulatora Atari), używane jest środowisko modelowane przez sieć neuronową, która nazywana jest modelem środowiska (world model). Model środowiska współdzieli przestrzeń akcji i przestrzeń nagród z środowiskiem. Głównym celem **SimPLe** jest nauczenie polityki przy użyciu symulowanego video.

Jednym z kluczowych aspektów przedstawionego podejścia jest zbudowanie modelu środowiska. W pracy opisane jest kilka istnicjących podejść do tego zagadnienia. Jednak ostatecznie autorzy decydują się na przedstawieniu swojego rozwiązania wykorzystując model stochastyczny z dyskretną przestrzenią ukrytą.

Praca jest doskonale napisana i z łatwością się ją czyta. Sekcja eksperymentalna jest bardzo dobrze przygotowana i stanowi najmocniejszą stronę niniejszej pracy.

Mgr Konrad Czechowski jest wymieniony jako szósty autor z dwunastu. To pokazuje, że wkład doktoranta nie jest kluczowy, ale nadal bardzo wysoki. Ponadto wydaje się, że niniejsza praca, o ile nie jest najważniejsza w kontekście przedstawionej rozprawy, o tyle wydaje się kluczowa w rozwoju kariery naukowej badacza. Pozwoliła ona autorowi zdobyć doświadczenie w pracy w dużej międzynarodowej grupie naukowej. Wydaje się, że doktorant świetnie się w niej odnalazł i wykorzystał zdobyte doświadczenia w dalszej pracy naukowej.

Pytanie:
Praca zyskała bardzo dużą ilość cytowań. Zastanawiam się, czy autor widzi możliwość użycia większej ilości modeli środowiska. Czy poprawa modelu środowiska może pomóc w uczeniu?

Motywacja pracy [P2] oparta jest na analizie błędów modelu w algorytmach uczenia ze wzmocnieniem. Powyższe błędy można podzielić na dwie kategorie: optymistyczne i pesymistyczne. W pracy autorzy zauważają, że ta druga kategoria błędów powoduje większe problemy, ponieważ model może nie odwiedzać pewnych ważnych obszarów eksplorowanej przestrzeni. W pracy opisany został jeden z błędów pesymistycznych nazywany *false loops*, który stanowi duży problem dla algorytmów pracujących w środowiskach dyskretnych. Motywacja pracy jest ciekawa i dobrze opisana. Stanowi ona punkt wyjścia do prezentowanej metody.

W pracy zaprezentowany został mechanizm **trust-but-verify (TBV)**, który wykorzystuje niepewność modelu do poprawienia eksploracji. W praktyce model wybiera dodatkowe akcje pod wpływem oszacowanej niepewności.

Podejście **trust-but-verify (TBV)** można dodać do wszystkich algorytmów planowania wykorzystujących wyszukiwania w grafach. Ogólna idea TBV opiera się na ocenie niepewności w stanach, aby zdiagnozować, czy model nie popełnia błędu typu pesymistycznego. Po wykonaniu predykcji za pomocą modelu planowania, jeżeli niepewność modelu jest duża, możemy podejrzewać wystąpienie błędu. Wtedy model wykonuje akcje w celu szerszej eksploracji.

Tak sformułowane rozwiązanie jest intuicyjne i logiczne. Ponadto zostało potwierdzone również eksperymentalnie. Wyniki numeryczne zawarte w pracy potwierdzają wydajność zaproponowanej metody.

Pytanie:
W pracy zaproponowany został sposób estymacji niepewności za pomocą testów statystycznych. Niestety nie są one doprecyzowane. Zastanawiam się, jak wygląda matematyczne sformułowanie testu. Ponadto wydaje się, że jest to bardzo prosty sposób estymacji niepewności. Czy autor próbował innych podejść do tego zagadnienia? Czy istnieje możliwość zwiększenia wydajności algorytmu przy wykorzystaniu bardziej zaawansowanych metod estymacji niepewności?

W pracy [P3] autorzy rozważają problemy związane z planowaniem w zadaniach posiadających dużą przestrzeń stanów. W szczególności zwracają uwagę na zależność między głębokością, a szerokością przeszukiwania, który w literaturze nazwany jest *depth and breadth of the search trade-off*. Ma on kluczowy wpływ na wydajność algorytmów uczenia ze wzmocnieniem. W pracy zaprezentowana jest metoda **Shoot Tree Search (STS)**, która pozwala kontrolować tą zależność.

Praca rozpoczyna się od opisu dwóch istniejących rozwiązań: *Temporal Difference (TD)* oraz *Monte Carlo Tree Search (MCTS)*. Powyższe podejścia realizują dwie strategie. Pierwsza z nich wykorzystuje struktury danych takie jak drzewa lub grafy. Druga opiera się na losowości, gdzie algorytm planowania wykonuje predykcje na losowych trajektoriach. W praktyce przestrzeń stanów może być ogromna co powoduje komplikacje podczas przeszukiwania. W związku z tym problem zrównoważenia głębokości i szerokości przeszukiwania jest dobrze uzasadniony. W pracy przedstawiona została dokładna analiza powyższego problemu w kontekście istniejących rozwiązań. Na tej podstawie wprowadzone zostało autorskie rozwiązanie, które można rozumieć jako kombinacje wcześniej przedstawionych podejść.

Głównym komponentem zaprezentowanej metody jest *multi-step expansion*, która rozważa predykcje o stałej długości. W przeciwieństwie do MCTS wprowadzone rozwiązanie wykorzystuje sieć neuronową do kierowania wyborem kolejnych akcji i dodaje odwiedzane węzły do drzewa przeszukiwania.

Struktura pracy jest przejrzysta, a wprowadzenie problemu i jego motywacja dobrze opisana. Natomiast sam opis metody opiera się na serii pseudokodów, które pojawiają się bardzo szybko i opisują całe podejście od ogólnych schematów do konkretnej metody. Same opisy algorytmów w pracy są cząstkowe i czytanie rozdziału *III Methods* opiera się w większości na analizie pseudokodów, oznaczeń w nich użytych oraz zależności między nimi. Taki sposób opisu jest ciekawy, ale z mojego punktu widzeniami algorytmy powinny być opisywane bardziej dokładnie. Pseudokod powinien stanowić dodatkową informację dla czytelnika, a nie główny trzon opisu metody.

Taki sposób opisu jest trudny w czytaniu, ale nie zmienia to faktu, że zaproponowana metoda jest ciekawa i daje dobre wyniki eksperymentalne. Zaproponowane rozwiązanie powoduje istotną poprawę względem rozważanych metod referencyjnych.

W pracy [P4] autor zaprezentował metodę generowania zadań pośrednich (*subgoals*) w metodach uczenia ze wzmocnieniem. Praca rozpoczyna się od podania intuicji prezentowanego rozwiązania. Motywacja biologiczna jest ciekawa i bardzo przekonująca.

W obliczu skomplikowanego zadania badacz często wyznacza sobie zadania pośrednie, które mają prowadzić do rozwiązania ostatecznego problemu. W takim procesie przechodzimy od jednego pomysłu do drugiego, co powoli zbliża nas do rozwiązania pełnego zadania. Kluczowe w tym procesie jest wyznaczanie zadań pośrednich i ich realizacja. Motywacja pracy jest bardzo przekonująca i dobrze

obrazuje działanie metody. Autor popiera swoje argumenty publikacjami z zakresu nauk neurobiologicznych.

Powyższa intuicja stoi u podstaw metody ***Subgoal Search (kSubS)***. Kluczowym elementem zaproponowanej metody jest stworzenie generator celów cząstkowych (*subgoal generator*), którego zadaniem jest tworzenie prostszych zadań, które prowadzą do rozwiązania zadanego problemu. W konsekwencji muszą one być na tyle proste, by algorytm mógł je rozwiązać relatywnie szybko i jednocześnie muszą nas zbliżać do rozwiązania głównego zadania.

Metoda *Subgoal Search (kSubS)* wykorzystuje model generatywny do przewidywania k kolejnych kroków. Takie podejście wykorzystuje uczenie nadzorowane, co powoduje, że model jest relatywnie prosty i intuicyjny. Metoda *kSubS* składa się z czterech komponentów: algorytmu planowania, generatora celów, polityki i value function. Generator tworzy potencjalne nowe podzadania. Metoda ogranicza się tylko do zadań, do których można dotrzeć relatywnie prosto. Procedura trwa do momentu znalezienia rozwiązania lub wyczerpania budżetu obliczeniowego. W głównej pracy przedstawione zostało rozwiązanie opierające się na *BestFS* natomiast w Appendix znalazła się metoda oparta na *MCTS*. Jako generator został użyty model *sequence-to-sequence transformer*.

Głównym wnioskiem pracy jest stwierdzenie, że model autoregresyjny oparty na transformatorze uczony w sposób nadzorowany jest skutecznym narzędziem do generowania zadań pośrednich, które umożliwiają efektywne rozwiązywanie głównego problemu.

Praca jest poprawnie napisana. Czyta się ją bez trudności, a struktura oparta na właściwej biologicznej motywacji zachęca do analizy całego problemu. Sekcja eksperymentalna, podobnie jak we wszystkich poprzednich pracach jest dobrze przygotowana, Zawiera dużą ilość ciekawych eksperymentów pokazujących wydajność zaproponowanego rozwiązania.

Pytanie:
Metoda wykorzystuje pruning w celu kontrolowania przeszukiwanych stanów. Sam proces jest opisany bardzo ogólnie. Proszę wyjaśnić jak duże znaczenie ma wydajność tego komponentu w całym procesie uczenia.

W pracy [P5] autorzy zaprezentowali uogólnienie metody z pracy [4]. Motywacją do rozszerzenia algorytmu *Subgoal Search (kSubS)* jest spostrzeżenie, że główne zadanie może być podzielone na mniejsze zadania o różnej trudności. W pracy zaprezentowany jest bardzo dobry przykład prowadzenia pojazdów. Gdy samochód jedzie wąskimi uliczkami musimy rozwiązywać dużo szybko pojawiających się problemów, jak pokonywanie zakrętów czy omijanie przeszkód. Natomiast jadąc prostym odcinkiem drogi możemy się skupić na dłuższym horyzoncie czasowym. Podobnie jak w poprzednich pracach, motywacja jest bardzo dobrze sformułowana i przemawia do wyobraźni czytelnika.

W oparciu o powyższą motywację autorzy wprowadzają algorytm *Adaptive Subgoal Search (AdaSubS)*, który wykorzystuje zmienny horyzont planowania. Dzięki takiej modyfikacji model może wykorzystywać dłuższe horyzonty czasowe, co daje przewagę w przypadku bardzo skomplikowanych zagadnień. Metoda **AdaSubS** preferuje dłuższe horyzonty, które redukuje do krótszych tylko gdy utknie.

AdaSubS składa się z podobnych komponentów jak *kSubS,* Autorzy dodali sieć weryfikującą *verifier* oraz zmodyfikowali *politykę* do *conditional low-level policy (CLLP).* Podobnie jak w *kSubS* generator przewiduje kandydatów na cele pośrednie, a *verifier* oraz *CLLP* oceniają czy predykcje są poprawne i osiągalne. W praktyce, autorzy proponują wykorzystanie kilku *subgoal generators* z różnymi horyzontami. Sieć weryfikująca odpowiada na binarny problem klasyfikacyjny. Musi ocenić czy na podstawie *CLLP* możliwe jest osiągnięcie stanu końcowego.

Zaprezentowana metoda jest prosta i intuicyjna. Ponadto jest ciekawym uogólnieniem z pracy [P4]. Podobnie jak wcześniejsze prace, sekcja eksperymentalna jest doskonale przygotowana, Wykorzystuje kilka różnych scenariuszy i dobrze pokazuje wyższość zaprezentowanej metody nad konkurencyjnymi rozwiązaniami.

Pytanie:
W praktyce *subgoal generators* jest bardzo podobny do rozwiązania z pracy [P4]. Czy możliwa jest modyfikacja tego komponentu w taki sposób, aby przewidywał dowolną wartość, a nie wybierał jedną z zaproponowanych przez użytkownika?

Pod koniec recenzji chciałbym zaznaczyć, że prace wchodzące w skład cyklu dobrze świadczą o warsztacie naukowym mgr. Konrada Czechowskiego. Ponadto pokazuje dogłębne zrozumienie tematyki i obiecująco rokuje na dalszy rozwój naukowy. Ponadto trzy prace zostały opublikowane na najlepszych konferencjach naukowych. Ponadto prace [P1], [P5] były prezentowane jako *spotlight presentation* oraz *long presentation,* co jest ogromnym sukcesem autora.

Podsumowując, rozprawa spełnia wszelkie zwyczajowe i ustawowe kryteria stawiane rozprawom doktorskim, w związku z czym wnioskuję o dopuszczenie doktoranta do dalszych etapów postępowania. Ponadto wnoszę o wyróżnienie pracy.

Przemysł Spurek