

dr hab. inż. Roman Zajdel, prof. PRz
Katedra Informatyki i Automatyki
Wydział Elektrotechniki i Informatyki
Politechnika Rzeszowska
al. Powstańców Warszawy 12
35-959 Rzeszów

Rada Naukowa Dyscypliny Rzeszów, 12 grudnia 2024
INFORMATYKA TECHNICZNA
I TELEKOMUNIKACJA
Sekretariat
Data wpływu 16.12.24 r.
Numer.....

RECENZJA

rozprawy doktorskiej dla
Rady Naukowej Dyscypliny
Informatyka Techniczna i Telekomunikacja
Politechniki Warszawskiej

Tytuł rozprawy: **Leveraging the distribution of collected samples in reinforcement learning**

Autor rozprawy: **mgr inż. Jakub Łyskawa**

Promotor: **dr hab. inż. Paweł Wawrzyński**

Uwagi wstępne

Recenzja została przygotowana na podstawie uchwały nr 742/2024 Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej, z dnia 17 września 2024 roku.

Łączna objętość recenzowanej pracy wynosi 103 stron.

1. Charakterystyka zagadnienia naukowego i charakter rozprawy

Rozprawa porusza zaawansowane zagadnienie optymalizacji algorytmów uczenia się ze wzmocnieniem (Reinforcement Learning, RL), będącego jednym z kluczowych nurtów badań nad sztuczną inteligencją i systemami autonomicznymi. Praca koncentruje się na algorytmach off-policy, które wykorzystują zgromadzone doświadczenia w celu przyspieszenia i stabilizacji procesu uczenia, oraz na hierarchicznym uczeniu się ze wzmocnieniem (Hierarchical Reinforcement Learning, HRL), umożliwiającym dekompozycję złożonych zadań na łatwiejsze do zarządzania cele pośrednie. Autor postawił sobie za cel poprawę efektywności algorytmów RL poprzez uwzględnienie zależności czasowych między kolejnymi akcjami oraz adaptacyjne ustalanie parametrów procesu uczenia.

W ramach rozprawy opracowano trzy autorskie algorytmy: Actor-Critic with Experience Replay and Autocorrelated Actions (ACERAC), Actor-Critic with Experience Replay and Sustained Actions (SusACER) oraz Adaptive Subgoal Required Distance (ASRD) w kontekście HRL. Każdy z tych algorytmów stanowi rozwinięcie istniejących metod, jednocześnie odpowiadając na



kluczowe wyzwania związane z praktycznymi zastosowaniami RL. Zastosowanie autokorelacji w ACERAC, zmiennej długości trwania akcji w SusACER oraz adaptacyjnej optymalizacji celu pośredniego w ASRD pozwala na zwiększenie wydajności algorytmów i ich przystosowanie do środowisk o złożonej dynamice czasowej.

Rozprawa ma charakter teoretyczno-empiryczny, z wyraźnym naciskiem na testowanie zaproponowanych metod w środowiskach symulacyjnych. Autor łączy dogłębną analizę teoretyczną z eksperymentami przeprowadzonymi na szeroką skalę, co znacząco zwiększa wiarygodność i możliwości zastosowania wyników. Praca podaje również potencjalne zastosowania algorytmów RL w robotyce oraz systemach autonomicznych, gdzie precyzja, stabilność i adaptacyjność mają kluczowe znaczenie. Całość ukazuje dogłębne przygotowanie merytoryczne autora oraz jego wysokie kompetencje badawcze.

2. Umiejscowienie problemu rozpatrywanego w rozprawie w kontekście literatury światowej

Problematyka badawcza poruszona w rozprawie wpisuje się w intensywnie rozwijający się obszar badań nad uczeniem ze wzmocnieniem, ze szczególnym uwzględnieniem algorytmów off-policy oraz hierarchicznego RL (HRL). Oba te zagadnienia mają kluczowe znaczenie dla nowoczesnych zastosowań sztucznej inteligencji. Dynamiczny rozwój badań nad RL, zapoczątkowany wprowadzeniem pamięci doświadczeń (replay buffer), umożliwia bardziej efektywne zarządzanie zebranymi danymi, stabilizację procesu uczenia oraz poprawę wyników w aplikacjach praktycznych. Fundamentalne prace, takie jak Deep Q-Network (Mnih et al., 2013) czy Actor-Critic with Experience Replay (Wawrzyński, 2009), odegrały kluczową rolę w ugruntowaniu tej dziedziny, a ich idee stały się podstawą wielu późniejszych badań. Rozwijając algorytm ACER, kandydat proponuje nowatorskie podejście do zarządzania pamięcią doświadczeń, w którym uwzględnia autokorelację w szumie akcji. Jest to unikalne rozwiązanie, które wyróżnia się na tle literatury, znacząco przyczyniając się do rozwoju istniejących metod.

Ważnym wkładem autora w badania nad RL jest również wprowadzenie adaptacyjnej długości trwania akcji w algorytmie SusACER. Tradycyjne algorytmy RL opierają się na stałej częstotliwości aktualizacji, co może być ograniczeniem w środowiskach o zmiennej dynamice. Analiza przeprowadzona przez autora, ukazująca wpływ zmiennego czasu trwania akcji na efektywność algorytmu, wypełnia lukę w dotychczasowej literaturze tematu. Dodatkowo metoda ASRD w HRL, automatycznie dostosowująca odległość celu pośredniego, wprowadza nowy poziom adaptacyjności do hierarchicznych metod RL. W literaturze, na przykład w pracach Hierarchical Actor-Critic (Levy et al., 2017) i HiTS (Gürtler et al., 2021), podkreślono znaczenie hierarchicznego modelowania, lecz bez uwzględnienia adaptacyjnej optymalizacji celów pośrednich. Propozycja ASRD wyróżnia się więc swoją innowacyjnością i potencjalnym zastosowaniem w systemach decyzyjnych wymagających dynamicznego dostosowywania parametrów.

Dodatkowym atutem rozprawy jest uwzględnienie najnowszych trendów badawczych, takich jak Quantum Reinforcement Learning (QRL). Włączenie technologii kwantowych do przyspieszenia procesu uczenia i poprawy eksploracji otwiera nowe możliwości badawcze, a kandydat wykazuje się dogłębną znajomością tego rozwijającego się kierunku badań. Szeroki przegląd literatury,

uwzględniający zarówno klasyczne prace, jak i najnowsze kierunki, świadczy o znakomitym rozeznaniu autora w aktualnym stanie wiedzy i badaniach nad RL.

3. Układ pracy, charakterystyka rozdziałów oraz opinia na temat ogólnej wiedzy teoretycznej kandydata z zakresu dyscypliny

Rozprawa doktorska Jakuba Łyskawy składa się z pięciu rozdziałów. Rozdział pierwszy oraz drugi stanowią odpowiednio wprowadzenie do tematyki pracy oraz do uczenia się ze wzmocnieniem. Kandydat wykazuje się głęboką znajomością literatury przedmiotu, omawiając istniejące podejścia i uzasadniając potrzebę opracowania nowych metod w środowiskach wymagających precyzyjnej dyskretyzacji czasu oraz hierarchicznego modelowania zadań. W rozdziale tym jasno sformułowane są także cele badawcze pracy. Rozdziały trzeci i czwarty podsumowują pięć publikacji (zamieszczonych w załącznikach do pracy) składających się dorobek doktoranta. Załączone publikacje naukowe tworzą logicznie powiązaną całość i stanowią rdzeń merytoryczny pracy. Każda z publikacji wnosi istotny wkład do tematyki uczenia się ze wzmocnieniem i jest odpowiednio umiejscowiona w strukturze rozprawy, co pozwala na przejrzystą prezentację osiągnięć naukowych autora. Kandydat odgrywał kluczową rolę w realizacji tych badań, co zostało określone procentowym udziałem w poszczególnych publikacjach.

Pierwsza z publikacji, „ACERAC: Efficient Reinforcement Learning in Fine Time Discretization” (IEEE Transactions on Neural Networks and Learning Systems), przedstawia algorytm ACERAC, który wprowadza autokorelację działań w celu poprawy stabilności i efektywności algorytmów RL. Kandydat odegrał wiodącą rolę w opracowaniu tego algorytmu, co zostało określone udziałem na poziomie 60%. Publikacja ta, opublikowana w czasopiśmie o wysokim współczynniku wpływu (impact factor 10.4), podkreśla innowacyjność podejścia autora.

Druga publikacja, „A Framework for Reinforcement Learning with Autocorrelated Actions” (27th International Conference on Neural Information Processing), rozwija ideę zarządzania autokorelacją działań w RL. Kandydat uczestniczył w tych badaniach w 30%, wnosząc znaczący wkład w eksperymenty i analizę wyników.

Trzecia publikacja, „Actor-Critic with Variable Time Discretization via Sustained Actions”, opisuje algorytm SusACER, który wprowadza zmienną długość trwania działań w RL, co pozwala na dynamiczne dostosowanie algorytmów do zmiennych środowisk. Kandydat był głównym autorem tej pracy, co zostało określone udziałem na poziomie 80%.

Czwarta publikacja, „Subgoal Reachability in Goal Conditioned Hierarchical Reinforcement Learning” (Applied Soft Computing), przedstawia adaptacyjne podejście do ustalania celów pośrednich w hierarchicznym RL poprzez algorytm ASRD. Autor wniósł 70% wkładu do tej publikacji, opracowując kluczowe elementy teoretyczne i przeprowadzając eksperymenty.

Piąta publikacja, „Influence of IQT on Research in ICT” (Journal of Theoretical and Applied Computer Science), skupia się na przeglądzie potencjalnych zastosowań RL i ich powiązań z innowacjami w innych dziedzinach informatyki. Kandydat uczestniczył w tych badaniach w 40%, wnosząc znaczący wkład do analizy zastosowań RL w praktyce.

Podsumowując, układ rozprawy jest logiczny i spójny, a publikacje wchodzące w jej skład pokazują szeroką wiedzę teoretyczną i praktyczną kandydata w dziedzinie RL. Kandydat wykazał się zaawansowanymi kompetencjami w zakresie projektowania algorytmów, przeprowadzania badań

empirycznych oraz krytycznej analizy wyników, co świadczy o jego wysokim poziomie przygotowania merytorycznego w dyscyplinie informatyka techniczna i telekomunikacja.

4. Ocena wartości naukowej rozprawy

Wartość naukowa rozprawy doktorskiej Jakuba Łyskawy jest wysoka i wynika z połączenia oryginalności, starannie przeprowadzonej analizy empirycznej oraz potencjalnych zastosowań wyników. Praca ta wnosi nowatorskie rozwiązania w zakresie optymalizacji algorytmów uczenia ze wzmocnieniem, zwłaszcza tych stosowanych w środowiskach o precyzyjnej dyskretyzacji czasu i hierarchicznym modelowaniu zadań. Kandydat, opracowując trzy autorskie algorytmy: ACERAC, SusACER i ASRD, nie tylko rozbudowuje istniejące metody, ale również przyczynia się do ich rozwoju w zakresie nowatorskich podejść opartych na adaptacyjności i autokorelacji.

4.1. Nowatorstwo proponowanych rozwiązań

Kluczową cechą podkreślającą wysoką wartość naukową tej rozprawy jest oryginalność zaproponowanych rozwiązań. Kandydat opracował nowatorskie podejście do zarządzania pamięcią doświadczeń poprzez wprowadzenie mechanizmu uwzględniającego autokorelację akcji w algorytmie ACERAC. Chociaż autokorelacja akcji była wcześniej badana w kontekście specyficznych systemów fizycznych, jej zastosowanie w metodach uczenia się ze wzmocnieniem pozostawało dotąd niewystarczająco rozwinięte. Wprowadzenie autokorelacji jako narzędzia wspierającego eksplorację i stabilność algorytmów off-policy stanowi unikalne rozwiązanie, które otwiera nowe możliwości badawcze w zakresie optymalizacji procesów uczenia w RL. Proponowane podejście rozszerza horyzonty badawcze i sprzyja praktycznym zastosowaniom RL w systemach autonomicznych, gdzie stabilność akcji ma fundamentalne znaczenie – przykładem mogą być robotyka i systemy sterowania.

Kolejną istotną innowacją przedstawioną w rozprawie jest algorytm SusACER, który wprowadza mechanizm dynamicznej zmiany długości trwania akcji w RL. Dotychczasowe algorytmy RL opierały się na stałej częstotliwości aktualizacji akcji, co często prowadziło do ich nieefektywności w środowiskach o zmiennej dynamice. Propozycja Jakuba Łyskawy pozwala na lepsze zarządzanie czasem trwania akcji, co umożliwia bardziej precyzyjne dostosowanie algorytmu do specyfiki zadania. Wprowadzenie dynamicznej dyskretyzacji czasu przynosi potencjalne korzyści w zastosowaniach wymagających elastyczności, znacząco poszerzając ramy dotychczasowych badań nad RL i oferując nowe podejście do równoważenia eksploracji oraz eksploatacji.

Metoda adaptacyjnego ustalania celu pośredniego (ASRD) stanowi kolejne przełomowe rozwiązanie w kontekście hierarchicznego uczenia ze wzmocnieniem (HRL). Dotychczasowe metody wymagały ręcznego doboru parametrów określających odległość celu pośredniego, co wiązało się z ryzykiem wyboru wartości suboptymalnych. ASRD eliminuje tę konieczność, wprowadzając dynamicznie dostosowujący się mechanizm, który umożliwia skuteczniejsze uczenie w złożonych środowiskach hierarchicznych. Literatura dotycząca HRL wskazywała na brak podobnych rozwiązań, co czyni metodę ASRD znaczącym wkładem w rozwój tej dziedziny, oferującym bardziej elastyczne i automatyczne podejście do zarządzania celami pośrednimi.

4.2. Staranność badań empirycznych

Badania empiryczne przeprowadzone przez kandydata są kompleksowe i potwierdzają skuteczność zaproponowanych algorytmów. Każdy z trzech autorskich algorytmów został przetestowany w zróżnicowanych środowiskach symulacyjnych, które różniły się częstotliwością aktualizacji,

poziomem dynamiki oraz wymaganiami dotyczącymi precyzyjnego sterowania. Dzięki szerokiemu zakresowi testów kandydat przedstawił wyniki dobrze udokumentowane, a analiza wykazuje, że proponowane metody przewyższają standardowe algorytmy, takie jak SAC czy PPO, w scenariuszach o wysokich wymaganiach czasowych. Staranność badań empirycznych widoczna jest także w precyzyjnym doborze metryk ewaluacyjnych, które pozwalają na szczegółowe porównanie osiągnięć różnych algorytmów. Kandydat zastosował metryki oceniające zarówno stabilność, jak i szybkość konwergencji, co jest kluczowe dla pełnej oceny skuteczności metod RL. Analiza danych przedstawiona w rozprawie jest rzetelna, a wyniki zaprezentowane w sposób przejrzysty, co ułatwia ich interpretację i ocenę wartości naukowej pracy.

4.3. Potencjalne zastosowania wyników w praktyce

Wartość naukowa rozprawy jest podkreślona przez potencjalne aplikacje wyników w praktycznych zastosowaniach, zwłaszcza w dziedzinie robotyki oraz systemów autonomicznych. Algorytmy ACERAC, SusACER i ASRD mogą znaleźć zastosowanie w systemach, w których precyzyjne zarządzanie czasem oraz stabilność akcji są kluczowe. Przykładowo, w robotyce przemysłowej i mobilnej, gdzie zmienne warunki środowiskowe wymagają strategii adaptacyjnych, dynamiczna długość trwania akcji może przyczynić się do znacznego usprawnienia procesów sterowania. Zastosowanie algorytmu ACERAC w systemach wymagających stabilnej eksploracji w środowiskach o precyzyjnej dyskretyzacji czasu może prowadzić do opracowania bardziej efektywnych rozwiązań w dziedzinach takich jak drony, pojazdy autonomiczne czy roboty kroczące. Z kolei algorytm ASRD, który automatycznie dostosowuje odległość celu pośredniego w zadaniach hierarchicznych, stanowi cenne narzędzie w systemach decyzyjnych, gdzie kluczowe jest rozbijanie złożonych problemów na mniejsze, zarządzalne podzadania. Zastosowanie ASRD w systemach takich jak autonomiczne nawigacje i roboty humanoidalne może umożliwić bardziej elastyczne i skalowalne podejście do złożonych zadań wymagających hierarchicznego planowania.

4.4. Wkład w rozwój wiedzy naukowej i otwarcie nowych kierunków badań

Rozprawa Jakuba Łyskawy nie tylko wnosi wartościowe innowacje do istniejących metod RL, ale także otwiera nowe kierunki badań w tej dziedzinie. Wprowadzenie autokorelacji działań oraz dynamicznej dyskretyzacji czasu pozwala na rozwój badań nad RL w kierunku bardziej elastycznych i adaptacyjnych metod, co może stanowić inspirację dla kolejnych prac. Metody takie jak ACERAC i SusACER wskazują potencjalne drogi rozwoju RL, w których kluczowe są zarówno stabilność, jak i elastyczność akcji, co stanowi wyzwanie w przypadku aplikacji w zmiennych i dynamicznych środowiskach. Ponadto adaptacyjność algorytmu ASRD w HRL jest innowacją, która może zainspirować dalsze badania nad dynamicznym zarządzaniem celami pośrednimi w hierarchicznych systemach decyzyjnych. Przegląd technologii kwantowych, choć zwięzły, wskazuje na potencjalne możliwości łączenia kwantowych metod z klasycznym RL, co stanowi nowatorski aspekt pracy i otwiera przestrzeń do przyszłych badań nad Quantum Reinforcement Learning (QRL).

Podsumowanie oceny wartości naukowej

Podsumowując, wartość naukowa rozprawy doktorskiej Jakuba Łyskawy jest wysoka, a jej znaczenie wynika z połączenia oryginalnych rozwiązań, starannie przeprowadzonych badań empirycznych oraz potencjalnych zastosowań w praktyce. Proponowane przez autora algorytmy rozwijają istniejące metody RL o nowe podejścia adaptacyjne i statystyczne, które mają znaczenie

nie tylko teoretyczne, ale także aplikacyjne. Wyniki osiągnięte przez autora mogą stanowić podstawę do dalszych badań w dziedzinie RL, a także inspirować rozwój metod RL w kierunku nowych obszarów zastosowań, co czyni tę pracę cennym wkładem do nauki.

5. Oryginalność rozprawy i wartość jej rezultatów na tle literatury światowej

Rozprawa doktorskiej Jakuba Łyskawcy cechuje się znaczną oryginalnością, co przejawia się poprzez zastosowanie innowacyjnych metod rozwiązania problemów, które nie były dotychczas szeroko rozwijane w literaturze światowej. Główną innowacją pracy jest rozwój algorytmów, które uwzględniają specyficzne właściwości środowisk wymagających precyzyjnego zarządzania czasem trwania akcji, dynamiką działań oraz adaptacyjności parametrów w procesie uczenia. Zaproponowano trzy unikalne algorytmy: ACERAC, SusACER oraz ASRD, które łącznie tworzą nową jakość w zakresie efektywnego wykorzystywania pamięci doświadczeń oraz adaptacyjnego ustalania celów pośrednich w hierarchicznym RL.

Algorytm ACERAC rozwija popularny Actor-Critic with Experience Replay (ACER), wprowadzając do procesu uczenia element autokorelacji szumu akcji. W literaturze badawczej koncepcja autokorelacji była rozważana głównie w kontekście zjawisk fizycznych, takich jak inercja systemów mechanicznych, i jej potencjalne zastosowanie w poprawie stabilności algorytmów RL pozostawało w dużej mierze niewykorzystane. Praca Jakuba Łyskawcy jest jedną z pierwszych, które szczegółowo badają ten aspekt oraz proponują bezpośrednie włączenie autokorelacji do procesu uczenia. Wprowadzenie tego mechanizmu pozwala na skuteczniejsze zarządzanie trajektoriami akcji oraz stabilizację procesu uczenia w środowiskach o wysokiej częstotliwości dyskretyzacji czasu. Algorytm ACERAC wyróżnia się tym na tle literatury, że wykazuje poprawę w stosunku do standardowych metod RL w symulacjach, co może mieć istotne znaczenie w praktycznych zastosowaniach, takich jak robotyka.

Algorytm SusACER rozwija metodologię RL poprzez zastosowanie zmiennej długości trwania akcji. Klasyczne metody RL zakładają stałą częstotliwość aktualizacji działań, co może być ograniczeniem w środowiskach o zmiennej dynamice lub tam, gdzie częstotliwość akcji wymaga dostosowania w zależności od zmieniających się warunków. Rozwiązanie zaproponowane przez autora pozwala na dynamiczne dostosowywanie częstotliwości aktualizacji działań, co ma na celu zoptymalizowanie równowagi między eksploracją a eksploatacją oraz zmniejszenie liczby niepotrzebnych obliczeń. W literaturze tematu stosunkowo niewiele uwagi poświęcono zmiennej długości trwania działań, dlatego propozycja Jakuba Łyskawcy wyróżnia się oryginalnością i wprowadza cenny wkład do badań nad metodami optymalizacyjnymi w RL.

Trzecia propozycja autora, algorytm ASRD w HRL, wprowadza adaptacyjność do hierarchicznego uczenia ze wzmocnieniem. Większość badań nad HRL, takich jak Hierarchical Actor-Critic (Levy et al., 2017), zakłada ręczne dostrajanie parametrów związanych z celami pośrednimi. Metoda ASRD umożliwia dynamiczne dostosowanie odległości celu pośredniego (Subgoal Reachability Distance) na podstawie wyników wcześniejszych prób, co pozwala na lepsze dostosowanie procesu uczenia do aktualnych warunków. Adaptacyjność ta, stosowana na poziomie celów pośrednich, jest nowatorskim podejściem, które eliminuje konieczność ręcznej optymalizacji parametrów i zwiększa stabilność oraz skuteczność algorytmu HRL.

Ostatecznie wartość rezultatów rozprawy doktorskiej Jakuba Łyskawcy na tle literatury światowej jest wysoka. Praca nie tylko rozwija istniejące metody, ale również oferuje innowacyjne podejścia,

które mogą stać się podstawą dla przyszłych badań. Wyniki eksperymentalne uzyskane przez autora są przekonujące, a proponowane rozwiązania mają potencjalne zastosowania w zaawansowanych systemach autonomicznych, gdzie stabilność i adaptacyjność algorytmów RL są kluczowe. W kontekście literatury światowej, praca wyróżnia się nowymi sposobami wykorzystania zależności statystycznych i adaptacyjnych mechanizmów, co czyni ją wartościowym wkładem do nauki.

6. Ocena umiejętności samodzielnego prowadzenia przez kandydata pracy naukowej

Umiejętności samodzielnego prowadzenia pracy naukowej przez Jakuba Łyskawę są na bardzo wysokim poziomie, co widać na każdym etapie realizacji rozprawy. Przede wszystkim autor wykazał się zdolnością do samodzielnego formułowania istotnych problemów badawczych, które w zasadniczy sposób rozbudowują istniejącą wiedzę z zakresu RL. Wyodrębnienie zagadnień takich jak autokorelacja akcji, zmienna dyskretyzacja czasu oraz adaptacyjna optymalizacja celów pośrednich wskazuje na inicjatywę autora oraz jego zdolność do dostrzegania ważnych wyzwań badawczych w tej dziedzinie. Kandydat potrafił nie tylko dostrzec istotne problemy, ale także zaproponować innowacyjne rozwiązania, co stanowi świadectwo jego samodzielności w prowadzeniu badań naukowych.

Kandydat wykazał się również wysokim poziomem umiejętności implementacyjnych. Algorytmy takie jak ACERAC, SusACER oraz ASRD, które wymagają zaawansowanej znajomości programowania oraz matematyki stosowanej, zostały przez autora w pełni zaimplementowane i przetestowane. Autor przeprowadził kompleksowe testy, które potwierdzają efektywność proponowanych metod. Planowanie oraz organizacja badań empirycznych, ich opis oraz szczegółowa analiza wyników są na wysokim poziomie, co świadczy o profesjonalnym podejściu kandydata do realizacji badań.

Dodatkowo, w swojej pracy autor wykazał zdolność do krytycznej oceny uzyskanych wyników, co jest niezbędne w samodzielnej pracy badawczej. Kandydat porównał osiągnięcia proponowanych metod z innymi popularnymi algorytmami RL, co pozwoliło na weryfikację ich skuteczności oraz wyciągnięcie wniosków na temat ich zastosowań i ograniczeń. Analiza wyników była przeprowadzona rzetelnie, a przedstawione wnioski są dobrze uzasadnione. Kandydat trafnie wskazał również kierunki przyszłych badań oraz możliwości zastosowania swoich metod w praktyce, co jest dodatkowym dowodem na jego zdolność do samodzielnego prowadzenia badań na poziomie eksperckim.

Podsumowując, Jakub Łyskawa posiada umiejętności niezbędne do samodzielnego prowadzenia badań naukowych. Jego zdolność do planowania i realizacji eksperymentów, krytycznego analizowania wyników oraz inicjatywa badawcza są na wysokim poziomie, co sprawia, że jest on w pełni przygotowany do samodzielnej pracy naukowej.

7. Słabe strony rozprawy

Choć rozprawa doktorskiej Jakuba Łyskawę cechuje się wysoką wartością naukową i znaczną oryginalnością, można wskazać kilka słabych stron, które mogłyby zostać rozwinięte lub poprawione, aby zwiększyć jej wartość merytoryczną oraz aplikacyjną.

Pierwszym istotnym ograniczeniem pracy jest przeprowadzenie eksperymentów wyłącznie w środowiskach symulacyjnych. Testowanie nowych algorytmów w symulacjach dostarcza

użytecznych danych na temat ich efektywności, jednakże ogranicza wartość wyników w kontekście ich rzeczywistego zastosowania. Przeprowadzenie testów w rzeczywistych środowiskach, np. w kontekście robotyki, mogłoby dostarczyć bardziej wiarygodnych dowodów na skuteczność zaproponowanych metod, a także wykazać ich skalowalność i odporność na niespodziewane zmienne. Wprowadzenie takich badań pozwoliłoby lepiej oszacować potencjalne wyzwania związane z implementacją zaproponowanych algorytmów w praktyce.

Kolejną słabością pracy jest ograniczona analiza hiperparametrów wprowadzonych w autorskich algorytmach. Zaproponowane metody, takie jak ACERAC i SusACER, wymagają dostosowania dodatkowych parametrów, np. stopnia autokorelacji czy długości trwania akcji. Praca nie zawiera szczegółowej analizy wrażliwości wyników na te parametry, co mogłoby dostarczyć bardziej kompletnych informacji na temat optymalnych ustawień i ułatwić ich aplikację w praktyce. Dodatkowa analiza wpływu hiperparametrów na stabilność algorytmu oraz jego wydajność mogłaby wzbogacić tę część pracy.

Autor przedstawia również ograniczoną dyskusję na temat potencjalnych ograniczeń metod adaptacyjnych. Adaptacyjność, choć przydatna, może wiązać się z dodatkową złożonością obliczeniową lub problemami wynikającymi z nieprzewidywalnych zmian parametrów w czasie rzeczywistym. Praca nie wprowadza analizy wpływu adaptacyjności na stabilność algorytmów ani potencjalnych problemów związanych z zastosowaniem metod ASRD. Dalsze rozwinięcie tej kwestii mogłoby pomóc zrozumieć ryzyka związane z zastosowaniem adaptacyjnych mechanizmów w praktyce i zwiększyć przejrzystość proponowanych metod.

Przegląd literatury dotyczący technologii kwantowych w RL, choć interesujący, jest dość ogólny i niezbyt silnie powiązany z głównymi zagadnieniami badawczymi rozprawy. Kandydat mógłby rozszerzyć tę sekcję, wskazując konkretne potencjalne połączenia technologii kwantowych z omawianymi problemami RL, jak np. możliwość wykorzystania kwantowych algorytmów do przyspieszenia procesów decyzyjnych w hierarchicznym RL. Rozwinięcie tej części w kontekście praktycznego zastosowania technologii kwantowych w proponowanych algorytmach mogłoby lepiej uzasadnić ich wprowadzenie i zwiększyć spójność pracy.

Ostatecznie, słabe strony rozprawy doktorskiej Jakuba Łyskawego dotyczą przede wszystkim obszarów, które mogłyby zwiększyć aplikacyjność i praktyczną wartość przedstawionych wyników. Mimo tych ograniczeń, praca charakteryzuje się wysokim poziomem merytorycznym i wnosi cenne innowacje do dziedziny RL, a wskazane uwagi stanowią potencjalne kierunki dalszego rozwoju badań autora.

8. Wniosek końcowy

Recenzowana rozprawa doktorska **Pana mgra inż. Jakuba Łyskawego spełnia w mojej opinii wymagania ustawy z dnia 20 lipca 2018 roku Prawo o szkolnictwie wyższym i nauce stawiane kandydatom do stopnia naukowego doktora w dyscyplinie informatyka techniczna i telekomunikacja.** Rozprawa proponuje istotne innowacje w obszarze uczenia się ze wzmocnieniem, a przeprowadzone badania mają dużą wartość naukową i aplikacyjną. Autor wykazał się wysokim poziomem samodzielności i inicjatywy badawczej, co uzasadnia dopuszczenie rozprawy do publicznej obrony.

Wniosek o wyróżnienie rozprawy doktorskiej

Na podstawie przeprowadzonej recenzji **wnoszę o wyróżnienie** rozprawy doktorskiej Pana mgr inż. Jakuba Łyskawy pt. „Leveraging the distribution of collected samples in reinforcement learning”, przygotowanej pod opieką naukową dr. hab. Pawła Wawrzyńskiego. W mojej ocenie praca nie tylko spełnia wszelkie wymagania stawiane rozprawom doktorskim ale wykracza poza standardy, wyróżniając się wysoką wartością naukową, innowacyjnością oraz profesjonalizmem w realizacji badań.

Rozprawa Jakuba Łyskawy charakteryzuje się wyjątkową oryginalnością. Kandydat zaproponował i wdrożył trzy nowatorskie algorytmy – ACERAC, SusACER i ASRD – które wnoszą istotne wartości do obszaru uczenia się ze wzmocnieniem. Każdy z algorytmów rozwija istniejące podejścia w unikalny sposób, wprowadzając do literatury światowej zupełnie nowe rozwiązania oraz odpowiadając na istotne wyzwania badawcze. Rozwiązania zaproponowane przez autora wzbogacają nie tylko wiedzę teoretyczną, ale także mają szeroki potencjał aplikacyjny w dziedzinach, które wymagają stabilnych, precyzyjnych i adaptacyjnych algorytmów, takich jak robotyka, systemy autonomiczne oraz przemysłowe systemy kontroli.

Przeprowadzone badania empiryczne zostały wykonane z dużą starannością i precyzją, co pozwala na wiarygodne potwierdzenie skuteczności zaproponowanych metod. Autor zaplanował i przeprowadził szerokie testy symulacyjne, a wyniki zostały rzetelnie przedstawione i oparte na wnikliwej analizie, co dodatkowo podkreśla profesjonalizm kandydata. Kandydat wykazał się również wyjątkowymi umiejętnościami badawczymi oraz zdolnością do prowadzenia samodzielnych badań naukowych. Praca dowodzi jego wysokich kompetencji w zakresie identyfikowania kluczowych problemów badawczych, projektowania innowacyjnych rozwiązań oraz ich precyzyjnej ewaluacji. Szeroka wiedza teoretyczna autora oraz umiejętność integracji wyników z aktualnymi trendami badawczymi czynią tę rozprawę wartościowym wkładem do rozwoju dyscypliny naukowej informatyka techniczna i telekomunikacja.

Biorąc pod uwagę powyższe argumenty, uznaję, że rozprawa ta wybitnie przyczynia się do rozwoju wiedzy w dziedzinie i reprezentuje poziom zdecydowanie wykraczający ponad wymagania stawiane pracom doktorskim. W związku z tym z pełnym przekonaniem wnoszę o jej wyróżnienie.



dr hab. inż. Roman Zajdel

Wydział Elektrotechniki i Informatyki
Politechnika Rzeszowska

