Prof. dr hab. inż. Cezary Zieliński
Instytut Automatyki i Informatyki Stosowanej
Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska
ul. Nowowiejska 15/19
00-665 Warszawa

Warszawa, 29 stycznia 2024 r.

WYDZIAŁ AUTOMATYKI, ROBOTYKI I ELEKTROTECHNIKI
POLITECHNIKA POZNAŃSKA
07 -02- 2024
WPŁYNĘŁO

PRZEWODNICZĄCY RADY DYSCYPLINY
Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne
prof. dr hab. inż. Wojciech Szeląg

Ocena rozprawy doktorskiej pt.:
Deep reinforcement learning for motion planning in man-made environments
autorstwa **mgr. inż. Piotra Kickiego**,
przygotowana dla Rady Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne Politechniki Poznańskiej.

Przedłożona do oceny rozprawa doktorska poświęcona jest efektywnym obliczeniowo metodom planowania ruchu robotów. Problem badawczy sformułowany w rozprawie jest istotny z punktu widzenia robotyki. Pomimo, że nie jest on nowy, to sposób jego rozwiązania stanowi nowy, ciekawy wkład do tej dziedziny wiedzy. Przedmiot rozprawy należy zakwalifikować do dyscypliny *Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne.*

Dysertacja zawiera 7 rozdziałów oraz spis cytowanej literatury. Rozdział 1. prezentuje motywację do podjęcia tematu, przegląd literatury oraz sformułowanie tez rozprawy. Zasadnicza część dysertacji została podzielona na dwa działy. Pierwszy, skomponowany z rozdziałów 2, 3 i 5, dotyczy algorytmów planowania ruchu pojazdów autonomicznych, bazujących na uczeniu maszynowym, natomiast drugi, w którego skład wchodzą rozdziały 4 i 6, koncentruje się na planowaniu ruchu manipulatorów i algorytmach o charakterze bardziej ogólnym. Rozdział 2. dotyczy planowania poprzez naukę, jak aproksymować idealną funkcję planowania za pomocą sekwencji lokalnych manewrów wykonywanych przez pojazdy samochodopodobne. Wspomniana idealna funkcja planowania przypisuje rozwiązalnemu problemowi planowania plan ruchu. O ile w rozdziale 2. korzystano z procesu decyzyjnego Markova oraz z sieci neuronowej iteracyjnie wyznaczającej kolejne segmenty ruchu, to w rozdziale 3. porzucono to podejście na korzyść „bandyckiego" sformułowania oraz reprezentacji zarówno ścieżki jak i profilu prędkości wzdłuż tej ścieżki za pomocą funkcji B-sklejanych. Metoda reprezentacji ścieżki brała pod uwagę ograniczenia narzucone na rozwiązanie, a w szczególności jego ciągłość wraz z odpowiednią liczbą pochodnych, omijanie przeszkód i łagodność skrętów. Rozdział 4. przesuwa swoje zainteresowanie z robotów nieholonomicznych, jakimi są samochody autonomiczne, na holonomiczne – manipulatory w tym przypadku. Co więcej, w tym rozdziale wprowadzono dodatkowo ograniczenia związane z dynamiką robota. Zwrócono uwagę na uczenie planowania w obliczu bliskości ograniczeń, akceptując drobne odchylenia. Dzięki zaproponowanym rozwiązaniom planowanie bądź przeplanowywanie można przeprowadzić w kilka milisekund. Rozdział 5. poświęcony jest ewaluacji zaproponowanych metod planowania ruchu wykorzystujących uczenie dla pojazdów oraz porównania ich z innymi znanymi z literatury planerami. Rozdział 6. przedstawia badania zachowania ogólnego algorytmu planowania zastosowanego

do manipulatora wykonującego zadania wymagające szybkich ruchów bądź przemieszczania ciężkich obiektów przy zachowaniu odpowiedniej ich orientacji przestrzennej. Ponadto porównano rezultaty planowania uzyskane za jego pomocą z tymi otrzymanymi dzięki zastosowaniu algorytmów znanych z literatury. Rozdział 7. stanowi podsumowanie wyników badań opisanych w rozprawie oraz wymienia główne dokonania Autora. Ponadto określa możliwe dalsze kierunki badań. Tekst rozprawy kończy lista cytowanej literatury, która zawiera 189 pozycji.

Motywacją do podjęcia badań była obserwacja, że brak jest szybkich metod planowania ruchu robotów w środowiskach narzucających istotne ograniczenia na dopuszczalność przemieszczeń. Opracowanie takich metod jest konieczne ze względu na chęć wykorzystywania robotów w środowisku stworzonym dla ludzi, co wymaga od tych urządzeń szybkich reakcji na zmieniającą się sytuację w ich otoczeniu. Główną tezą pracy jest stwierdzenie, iż szeroka klasa problemów związanych z planowaniem ruchu robotów może być rozwiązana w prawie stałym czasie dzięki zastosowaniu głębokiego uczenia ze wzmocnieniem. Teza ta została uszczegółowiona. Wskazano, że formalizacja problemu planowania ścieżki dla pojazdów samochodopodobnych poprzez zastosowanie procesów decyzyjnych Markova umożliwia sekwencyjne tworzenie rozwiązania poprzez wykorzystanie fragmentów ścieżek generowanych przez sieci neuronowe korzystające z uczenia przez wzmacnianie. Ponadto wykorzystanie krzywych B-sklejanych umożliwia efektywne planowanie przy pojedynczym wnioskowaniu sieci neuronowej. Podkreślono, że za pomocą sieci neuronowych możliwe jest rozwiązanie problemów planowania ruchu kinodynamicznego z ograniczeniami przy wykorzystaniu sieci neuronowych trenowanych przez wzmacnianie oraz reprezentacji trajektorii za pomocą funkcji B-sklejanych.

Analizę literatury tematu rozpoczęto od ogólnego przeglądu metod planowania ruchu. Następnie uszczegółowiono te elementy metod planowania ruchu, które były istotne dla rozważań zawartych w dysertacji. Opisano planowanie ruchu pojazdów autonomicznych. Dalej skoncentrowano się na metodach wprowadzania ograniczeń narzucanych na sposób wykonania zadania przez pojazd. Potem dodano ograniczenia związane z uwzględnieniem dynamiki robota. Na koniec zajęto się planowaniem ruchu z wykorzystaniem uczenia. Zwrócono szczególną uwagę na algorytmy wytwarzające plan ruchu w przewidzianym czasie. Pokazano, że planowanie ruchu jest problemem obliczeniowo wymagającym. Wymieniono, jakie algorytmy były dotychczas stosowane do rozwiązywania tego problemu, wskazując ich wady i zalety. Podkreślono, że głównym problemem jest wysoka wymiarowość problemu. Omówiono zarówno metody bazujące na próbkowaniu przestrzeni jak i metody optymalizacji ciągłej.

Rozprawa koncentruje się na dwóch najistotniejszych zagadnieniach związanych z planowaniem ruchu robotów:

- na reprezentacji generowanej trajektorii ruchu oraz
- na sposobie uczenia sieci neuronowej – w tym przypadku określeniu różniczkowalnej postaci złożonego funkcjonału, według którego dokonuje się wielokryterialnej optymalizacji wag sieci.

W obu przypadkach wypracowano bardzo interesujące propozycje. Co więcej, propozy-

cje te zweryfikowano symulacyjnie, w niektórych przypadkach również na rzeczywistym robocie. Istotnym jest, że w porównaniu z istniejącymi algorytmami planowania wykorzystującymi sieci neuronowe, efekty planowania, czyli wygenerowane trajektorie, były dużo lepsze. Wszakże najistotniejszym jest to, że czas planowania mieści się w granicach kilku do kilkudziesięciu milisekund, w zależności od złożoności środowiska i przyjętego modelu robota. To gwarantuje szybkie reakcje robota na zmiany zachodzące w jego otoczeniu. Ponadto, jak wskazują zaprezentowane przykłady, zaproponowane podejście może być zastosowane zarówno w przypadku robotów holonomicznych jak i nieholonomicznych.

Rozprawa poświęcona jest ciekawemu z punktu widzenia naukowego zagadnieniu planowania ruchu robotów i to zarówno pojazdów samochodopodobnych jak i manipulatorów. Temat został zbadany dogłębnie i wieloaspektowo. Struktura rozprawy jest właściwa. Cytowaną literaturę dobrano właściwie. Starano się przedstawić zarówno osiągnięte wyniki jak i niezbędny materiał wprowadzający umożliwiający zrozumienie przedstawianych algorytmów, a w szczególności ich aspektów matematycznych. Tekst zachowuje właściwą równowagę pomiędzy materiałem wprowadzającym, a prezentacją własnych dokonań. Do najbardziej wartościowych wyników naukowych należy zaliczyć:

- Iteracyjną metodę wykorzystania sieci neuronowej do wyznaczania segmentów ruchu pojazdu samochodopodobnego w obliczu przeszkód, wraz z określeniem różniczkowalnej funkcji kosztu umożliwiającej uczenie wag tej sieci metodą gradientową.
- Istotną modyfikację tej metody poprzez przekształcenie sieci neuronowej tak, aby w jednym kroku wyznaczała parametry funkcji B-sklejanych reprezentujących całą ścieżkę przejazdu robota oraz profil prędkości przejazdu wzdłuż tej trasy, przy jednoczesnym uproszczeniu funkcji kosztu i zachowaniu możliwości gradientowego uczenia wag sieci.
- Sposób konstrukcji funkcji kosztu używanej do modyfikacji wag w sieci neuronowej w trakcie jej uczenia. W szczególności uwzględniono zarówno ograniczenia narzucane przez środowisko jak i dynamikę robota.

Każda dobra rozprawa naukowa rozwiązuje jakąś klasę istotnych problemów w danej dziedzinie. I tak jest też w tym przypadku, ale każda dobra rozprawa rodzi też pytania, które wskazują dalsze kierunki badań. W zakończeniu rozprawy wskazano kilka takich kierunków. Wszakże w trakcie jej lektury nasunęły mi się dodatkowo poniższe pytania:

- Struktura układu sterowania manipulatorem na rys. 6.1 i 6.2 nie jest kompletna. W szczególności brakuje w tej strukturze sieci neuronowych, które są przedmiotem rozważań. Ze względu na podstawowy temat rozprawy należałoby pokazać, gdzie odbywa się przeplanowywanie i jak są dostarczane dane niezbędne do jego przeprowadzenia?
- W sekcji 6.5.4 pokazano, że nawet drobne zmiany wprowadzone do środowiska powodują konieczność powtórnego uczenia. Na ile można zmienić środowisko, by nie trzeba było ponownie uczyć sieci planującej? Co zrobić, by robot był w stanie radzić sobie ze znacznymi zmianami strukturalnymi środowiska bez konieczności powtórnego uczenia lub douczania? Czy istnieje jakaś miara dotycząca tego zagadnienia?

- Jak złożone są wzory wyrażające gradienty wykorzystywane do minimalizacji funkcjonałów dla poszczególnych zadań opisanych w rozprawie?
- W sekcji 2.3.2.2 został zdefiniowany kontekst $C = (E, \mathcal{T})$, gdzie E jest mapą, natomiast $\mathcal{T}$ zadaniem. Jakkolwiek w rozdziale 5. brak jest odniesienia do wyboru zarówno mapy jak i zadania, zarówno w fazie uczenia jak i planowania. Do jakiego stopnia sukces planowania jest zależny od wyboru zestawu map E oraz zestawu zadań $\mathcal{T}$ podczas uczenia? Na ile w trakcie planowania zarówno środowiska jak i zadania mogą być różne od tych użytych w trakcie uczenia?

Zdaję sobie sprawę, że rozprawa w obecnej formie liczy ponad 120 stron, więc Autor musiał ograniczyć jej zawartość do niezbędnego minimum, które i tak jest dość obfite. Zapewne przynajmniej na część z tych pytań odpowiedzi są mu znane, a nie zostały zamieszczone w rozprawie ze względu na brak miejsca. Spodziewam się, że przynajmniej na niektóre z tych pytań usłyszę odpowiedzi w trakcie obrony.

Ponadto sam tekst rozprawy można by ulepszyć biorąc pod uwagę poniższe wskazówki. Wymienię wpierw te dotyczące drobnych niewyjaśnionych kwestii:

- Na str. 29. wskazano, że wektor stanu aktualnego i pożądanego przekształcany jest przez sieć neuronową w wektor 256-elementowy. Wszakże nie wyjaśniono czemu niezbędnych jest aż tyle elementów?
- Dlaczego na wejściu procesora mapy (rys. 2.4 i 3.5) są dwie warstwy o rozmiarach 128×128, a nie jedna?
- Dlaczego w poszczególnych sieciach neuronowych jest akurat tyle warstw?
- W rozważaniach teoretycznych zakładano, że samochód zawsze porusza się do przodu (str. 21). Natomiast z rys. 5. i 8. wydaje się wynikać, że manewr parkowania wymaga cofania. Warto byłoby wyjaśnić tę kwestię.
- Z sekwencji zaprezentowanej na rys. 5.13 wydaje się, że niebieski samochód tylko nieznacznie wjechał na pas ruchu, po którym poruszał sie samochód własny. Oznaczałoby to, że jedynie nieznaczna modyfikacja trajektorii ruchu samochodu własnego była potrzebna. Co by się stało gdyby niebieski samochód zachowałby sie nieco bardziej agresywnie, tzn. do jakiego stopnia zaproponowany planer jest w stanie uniknąć kolizji? Czy istnieje jakaś miara mówiąca o stopniu pewności uniknięcia niebezpieczeństwa?
- Z rys. 6.1 wynika, że sterowniki ADRC wytwarzają wartości momentów siły zadanej dla silników poszczególnych złącz manipulatora. Wektor tych momentów podawany jest do kompensatora grawitacji, który wytwarza ostateczne wartości momentów sił zadanych dla poszczególnych silników. Kompensacja siły grawitacji oddziałującej na człony manipulatora musi się odbywać w przestrzeni operacyjnej (zadaniowej), natomiast momenty sił związane są z przestrzenią konfiguracyjną (złącz). W tej sytuacji, co konkretnie robi kompensator?
- W obserwatorze stanu z rys. 6.2 występują zmienne z_1, z_2, z_3. Nie wyjaśniono, co konkretnie one obrazują.

a następnie te natury redakcyjnej:

- Rozprawa zawiera listę symboli w niej wykorzystanych. Lista ta liczy ponad 4 strony. Przy tej liczbie lista ta jest niezbędna i dobrze, że została zamieszczona. Niestety nie ułożono jej w sposób systematyczny, np. alfabetycznie, co wydatnie ułatwiłoby znalezienie poszukiwanego symbolu.
- Przy tej liczbie symboli wskazanym byłoby zastosowanie jakiejś systematycznej konwencji notacyjnej. Przy jej braku nauczenie się znaczenia wszystkich symboli jest pewnym wyzwaniem.
- Jak wyprowadzono wzór (2.4)? Czy aby na pewno powinien on zawierać funkcję tan, a nie arctan? Jeżeli moje domniemanie jest słuszne, to należy również poprawić wzór (2.15).
- Na str. 28. Autor twierdzi, że mapa o przyjętych rozmiarach (128×128) może przyjmować 2^{14} stanów. Tyle to mapa zawiera komórek. W związku z tym możliwych stanów jest $2^{(2^{14})}$.
- We wzorach (4.7) i (4.8) kwadraty powinny być związane z c_i, a nie listą argumentów tej funkcji ujętą w nawiasy.

Powyższe uwagi mają na celu jedynie poprawę jakości przyszłych dzieł Doktoranta i nie mają wpływu na moją wysoką opinię o rozprawie.

Podsumowując, postawiony problem badawczy został rozwiązany. Uzyskane wyniki zostały opisane wyczerpująco, dbając o odpowiednie wprowadzenie matmatyczne w poruszane zagadnienia. Powyższe uwagi natury ogólnej stanowią jedynie wyraz ciekawości recenzenta, natomiast uwagi szczegółowe, uważam za mało istotne, bo ich przyczyny można zakwalifikować, jako usterki techniczne możliwe do łatwego naprawienia. Rozprawa jest starannie zredagowana, a powyższe drobne usterki są nieuniknione przy dziele pokaźnych rozmiarów. W związku z tym uwagi te nie wpływają na moją ogólnie pozytywną ocenę przedłożonej pracy doktorskiej, **którą uznaję za spełniającą z wyraźnym nadmiarem wymagania stawiane przed tego typu rozprawami** zgodnie z obowiązującymi przepisami. W mojej opinii **praca zasługuje na wyróżnienie**, co też rekomenduję Radzie Dyscypliny AEEiTK Politechniki Poznańskiej.

.....

podpis