

25-10-2023

PRZEWODNICZĄCY RADY DYSCYPLINY
Automatyka, Elektronika, Elektrotechnika
i Technologie Kosmiczne

WPŁYNEŁO

prof. dr hab. inż. Wojciech Szelaąg

dr hab. inż. Grzegorz Granosik, prof. uczelni
Politechnika Łódzka
Instytut Automatyki

Łódź, dn. 20.10.2023 r.

RECENZJA

rozprawy doktorskiej mgra inż. Szymona Sobczaka na temat „Restricted Boltzmann Machine as a binary image descriptors processor and its application in a mobile robot for scene recognition”

(Opinia niniejsza została przygotowana na zlecenie Przewodniczącego Rady
Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie kosmiczne
Politechniki Poznańskiej, zgodnie z Uchwałą Rady z dnia 21.06.2023)

1. Przedmiot rozprawy

Przetwarzanie obrazów, analiza otoczenia, planowanie ruchu robotów, omijanie przeszkód, nawigacja, planowanie chwytu, detekcja podłoża to bardzo aktualne zagadnienia naukowe, wszystkie coraz częściej wiążą się z wykorzystaniem sieci neuronowych do opisu i klasyfikacji danych. Wiele rozwiązań znanych ze światowej literatury naukowej, a nawet coraz częściej z produktów komercyjnych charakteryzuje się bardzo wysoką efektywnością, często zbliżoną do ludzkiej, a czasami przewyższającą człowieka w szybkości działania. Jednak zwykle wymaga to potężnych mocy obliczeniowych i zasobów sprzętowych, a przede wszystkim ogromnych ilości danych treningowych odpowiednio przygotowanych. Zadanie staje się trudne lub niemożliwe do wykonania w przypadku systemów wbudowanych, o limitowanych zasobach, jak to ma miejsce w autonomicznych robotach mobilnych. Takie właśnie problemy zauważa Autor rozprawy i aktywnie włącza się w ten obszar badań z propozycją modyfikacji procedury przetwarzania danych wizyjnych. Zastosowana metoda preprocesingu wydobywa cechy binarne z obrazu RGB, które dalej przetwarzane są przez sieć neuronową w wektory cech dla każdego piksela. Te tensory, a nie pierwotne obrazy, podawane są na zasadniczą sieć klasyfikującą, co daje lepsze wyniki i wymaga mniejszych zasobów obliczeniowych.

W obszarze narzędzi informatycznych Doktorant skupia swoją uwagę na ograniczonej maszynie Boltzmanna (RBM) – jest to narzędzie sztucznej inteligencji stosowane dość często, z uwagi na wspieranie uczenia bez nadzoru i efektywną obsługę dużych zbiorów danych – biorąc pod uwagę jedynie publikacje zgromadzone w bibliotece cyfrowej IEEE Explore można znaleźć prawie 30 artykułów opublikowanych na ten temat tylko w tym roku.

Podsumowując te wstępne uwagi, stwierdzam że rozprawa mgra inż. Szymona Sobczaka podejmuje bardzo ważną i trudną tematykę. Rozważany problem z całą pewnością można uznać za aktualne zagadnienie naukowe w obszarze automatyki i robotyki oraz w dyscyplinie automatyka, elektronika, elektrotechnika i technologie kosmiczne, a do jego rozwiązania potrzebne były studia literaturowe (przeprowadzone w sposób właściwy), samodzielne rozszerzenie znanych wcześniej wyników teoretycznych i wprowadzenie własnych rozwiązań oraz wykonanie szeregu badań eksperymentalnych nowo opracowanych algorytmów na ogólnodostępnych danych obrazowych oraz zebranych przez robota. Aspekt praktyczny jest

szczególnie wart podkreślenia, gdyż uwzględnia nie tylko teoretyczne rozwiązanie problemu, ale bierze pod uwagę cały szereg ograniczeń technicznych i trudności implementacyjnych.

2. Ocena zawartości rozprawy

Układ pracy jest logiczny i spójny. Podział treści jest właściwy i ściśle podporządkowany uzasadnieniu tez rozprawy, które zostały sformułowane następująco:

- it is possible to increase generalisation ability of a neural network with the use of the novel unsupervised preprocessing method based on binary descriptors and a recurrent neural network,
- this preprocessing may be applied to reduce the complexity of a classification neural network without a significant decrease in accuracy, thus minimising required computing resources,
- applying the proposed preprocessing may result in less image-distortion sensitivity of a classification neural network,
- this preprocessing may provide an additional metrics allowing to measure a distance between image datasets.

Dodatkowo Autor wskazał dwa istotne ograniczenia dla stosowania współczesnych metod analizy obrazów bazujących na sieciach neuronowych:

- Istnieje wiele systemów, zwłaszcza platform wbudowanych, które mają ograniczoną wydajność obliczeniową, ponieważ rozmiar, cena i zużycie energii są istotnymi czynnikami przy projektowaniu systemów. Robotyka, a w szczególności roboty mobilne, to działy przemysłu wykorzystujące tego typu systemy, dlatego istnieje znaczna potrzeba ograniczenia zasobów potrzebnych do rozwiązania danego problemu klasyfikacyjnego,
- Drugą ważną kwestią jest dostępność danych. Roboty mobilne są w stanie gromadzić dane obrazowe, ale nie mogą ich wykorzystać bez oznaczenia przez człowieka, dlatego bardzo ważne są badania nad metodami uczenia bez nadzoru, które nie wymagają dużej mocy obliczeniowej.

Tezy są oryginalne i świadczą o bardzo dobrej znajomości aktualnych badań naukowych dotyczących przetwarzania obrazów i klasyfikacji w czasie rzeczywistym. Wykazują też, że Autor potrafi formułować nowe, ambitne zadania badawcze oraz dążyć do ich co najmniej zadawalającego rozwiązania.

Praca została podzielona na 6 zasadniczych rozdziałów i wnioski, oraz zawiera 3 załączniki z materiałami dodatkowymi. Każdy rozdział ma dobrze zorganizowaną strukturę, jest przejrzysty graficznie i bogato ilustrowany co pomaga zrozumieć relacje pomiędzy etapami przetwarzania danych. Literatura obejmująca 199 pozycji jest bardzo trafnie dobrana, i niemal wszystkie zestawione publikacje są cytowane w rozprawie. Należy odnotować, że mgr Sobczak jest współautorem 3 przywołanych artykułów (w tym dwóch opublikowanego w czasopiśmie z IF).

W pierwszym rozdziale rozprawy Autor przedstawił ogólny przegląd aktualnej wiedzy w obszarze sztucznych sieci neuronowych, ich struktur i zastosowań. W szczególności poruszył sprawę trudności zastosowania sieci do analizy obrazów w robotyce, tam gdzie zasoby obliczeniowe są ograniczone, podobnie jak dostępność otagowanych danych (co ma miejsce w przypadku przeniesienia robota i sieci, które radzą sobie dobrze w jednym obszarze, do nowego obszaru). Autor zauważa, że oba problemy można rozwiązać poprzez odpowiedni preprocesing realizowany w sposób nienadzorowany i generujący najpierw deskryptory binarne a później wybierający najważniejsze cechy i przenoszący je do bardziej abstrakcyjnej przestrzeni. To oczywiście wymaga modyfikacji dalszych etapów przetwarzania, gdyż postać danych wejściowych dla sieci splutowej ulega zmianie. Doktorant wskazuje jeszcze dodatkowe zalety takiego podejścia, które razem zaprezentowane są w postaci tej rozprawy.

Obszar badawczy jest dalej doprecyzowany w rozdziale drugim – przedstawiono najczęściej używane ekstraktory cech obrazu, które stanowią zasadniczą część podejścia proponowanego w dysertacji. Autor skupia się na modelach wspierających uczenie nienadzorowane, omawia sieć Hopfielda i ograniczoną maszynę Boltzmanna. Prezentuje także deskryptory cech, w wersji o wartościach binarnych i rzeczywistych. Sekcja 2.5 przedstawia propozycję dodania informacji o kolorze i jasności do deskryptora kształtu, co odpowiada na słusznie zauważoną potrzebę włączenia tych informacji do algorytmów rozpoznawania otoczenia w robocie mobilnym. Autor proponuje rozszerzenie reprezentacji 8-bitowego deskryptora LBP8 działającego w przestrzeni szarości o dodatkowe 8 bitów kodujących uproszczoną informację o kolorze i jasności (CLBP).

W rozdziale trzecim przedstawiono ogólną koncepcję modelu ograniczonej maszyny Boltzmanna i jej działanie podczas uczenia nienadzorowanego. Opis jest uporządkowany, wychodzi od historycznie wcześniejszej maszyny Boltzmanna, wprowadza zależności analityczne na aktualizację prawdopodobieństwa oraz funkcję energii, podkreśla rolę uproszczeń w modelu RBM oraz zalety stochastycznej dynamiki pozwalającej na opuszczenie lokalnych minimów. Autor zwraca uwagę na dodatkowe ciekawe właściwości RBM pozwalające na określanie, który z aktualnych wzorów wejściowych pojawia się częściej w danych treningowych. Oddzielne podrozdziały (3.2 i 3.2) poświęcono zagadnieniu uczenia RBM – ponownie, historycznie uporządkowany opis pozwala łatwiej zrozumieć zawiłości procesu oraz zalety modyfikacji wprowadzanych przez kolejnych badaczy – pokazuje także znakomitą orientację Doktoranta w historycznej i aktualnej literaturze przedmiotu. Niezbędne wzory podane są w formie analitycznej, zaś algorytm uczenia (dywergencja kontrastowa – CD) w postaci pseudokodu.

Kolejny rozdział to szczegółowy opis zagadnień implementacyjnych, zastosowanego oprogramowania, bibliotek i narzędzi. Z uwagi na zaproponowaną własną formułę deskryptora konieczna była odpowiednia implementacja oraz integracja z algorytmem dywergencji kontrastowej. Dalej Autor umiejętnie korzysta z zalet języka Python i bogatej oferty bibliotek wspierających uczenie maszynowe – przy czym dokonuje wartościowych modyfikacji i uzupełnień, zapewniając wsparcie dla obliczeń równoległych w przypadku obecności GPU. Zaproponowano trzy architektury przetwarzania danych: RBM-CD i KNN jako końcowy klasyfikator, RBM-CD i CNN oraz wykorzystanie ukrytych warstw RBM jako ekstraktora cech dalej przetwarzanych przez sieć konwolucyjną. Autor powraca także do

problemu używania wytrenowanej sieci w nowych zastosowaniach oraz możliwej automatycznej oceny podobieństwa danych (lub podobieństwa rozkładu cech, w szczególności w rozważanym projekcie). Pokazuje możliwość użycia wspomnianej wcześniej cechy RBM obliczania prawdopodobieństwa krańcowego, proponuje rozszerzenie tych obliczeń na pewną grupę danych oraz zapisanie rozkładu w postaci histogramu, co dalej prowadzi do uproszczenia analizy do porównywania podobieństwa histogramów. Kolejne zastosowanie stworzonych przez Doktoranta algorytmów związanych z ograniczoną maszyną Boltzmann to ograniczenie wpływu zaszumienia danych wejściowych na końcowa klasyfikację. Autor zauważa możliwość wprowadzenia odszumienia do algorytmu przetwarzania RBM jako dodatkowy krok rekonstrukcji.

Rozdział 5 to prezentacja wyników przeprowadzonych badań co stanowi eksperymentalny dowód przydatności stosowania metod opisanych w rozdziale 4. Jest to największy rozdział, prezentujący aż 15 eksperymentów, walidujących 3 architektury oraz ich specyficzne właściwości; zastosowano 7 zbiorów danych o bardzo różnej liczebności i rozmiarach obrazów, liczbie kategorii i typach zdjęć. Większość zbiorów była otagowana, zaś część przeznaczona do uczenia nienadzorowanego. Kolejne eksperymenty są dobrze opisane i odwołują się do wcześniej omówionych teorii. Autor rozpoczyna od zbadania czasów przetwarzania danych z klasycznym deskrytorem LBP8 oraz zaproponowanym CLBP – wyniki pokazują, że nawet przy dość dużych rozmiarach obrazów czasy przetwarzania są poniżej 1ms, jednak wydłużają się znacząco przy wzroście liczby komórek wejściowych w modelu RBM. Kolejny eksperyment dotyczył rozpoznawania odrębnego zapisu cyfr i pokazał przewagę zaproponowanego preprocesingu nad przetwarzaniem surowych danych obrazowych. Eksperyment pozwolił także na dokładne strojenie parametrów modelu RBM. Eksperyment 4 wskazał wyraźnie lepsze działanie rozszerzonej reprezentacji deskryptora z kodowaniem koloru i jasności nad LBP8 dla kolorowych obrazów z bazy CIFAR-10. Kolejne eksperymenty pokazują zmiany w warstwach ukrytych RBM podczas uczenia – Autor pokazuje korelacje pomiędzy cechami z warstw ukrytych a obrazem treningowym. Eksperymenty 7 i 8 pokazują wsparcie dla procesu uczenia jakie daje zaproponowane przetwarzanie wstępne: poprawia się dokładność sieci (dla różnych rodzajów sieci splotowych) oraz wzrasta tolerancja na zmniejszenie liczby warstw konwolucyjnych oraz zmniejszenie liczby danych treningowych. Dalej Autor bada skuteczność mierzenia podobieństwa danych obrazowych na podstawie zaproponowanych metryk i algorytmów wykorzystujących właściwości RBM – pokazując obiecujące wyniki, zgodne z założeniami, przynajmniej dla kilku zbiorów danych. Podobnie sytuacja ma się z wykorzystaniem przetwarzania wstępnego do odszumiania, dla wielu rodzajów szumu i zakłóceń, także w przypadku szumu gradientowego specjalnie przygotowanego do zmylenia uczenia maszynowego. Rozdział pokazuje dużą sprawność Doktoranta w przygotowaniu i przeprowadzeniu rozbudowanej kampanii eksperymentów, a także w analizie i prezentacji wyników.

Rozdział 6 opisuje praktyczne zastosowanie przedstawionych wcześniej metod w urządzeniu wprowadzającym ograniczenia w zakresie zasobów obliczeniowych, o których była mowa we wstępie – czyli w robocie mobilnym. Autor wybiera nieduży pojazd typu differential drive (JetBot) z komputerem pokładowym klasy Jetson Nano i kamerą dostarczającą informacji o

otoczeniu. Tworzy dla tego zestawu odpowiednie aplikacje: pracującą na robocie (do przetwarzania obrazu, uczenia maszynowego, lokalizacji i komunikacji) oraz interfejs użytkownika pracujący na komputerze zewnętrznym. Tworzy także prostą infrastrukturę do pomiaru rzeczywistego błędu orientacji. Pięć kolejnych eksperymentów zostało przeprowadzonych i opisanych w tym rozdziale. Pokazały możliwość rozpoznawania i klasyfikacji kilku obiektów w typowym otoczeniu domowym z dokładnością nieco większą przy zastosowaniu przetwarzania wstępnego niż bez niego. Dalej zaś możliwość ustawiania robota naprzeciw danego obiektu wykorzystując metrykę podobieństwa obrazów, oraz możliwość tworzenia prostej mapy otoczenia robota zawierającej kątowne lokalizacje tych obiektów. Podobne testy zostały również przeprowadzone dla innych obiektów: otwartych i zamkniętych drzwi.

Praca jest zakończona zwięzłym podsumowaniem odnoszącym się do tezy rozprawy i wyników licznych eksperymentów potwierdzających wysoką efektywność stworzonych algorytmów w aplikacjach wbudowanych, oraz wskazującym także dalsze kierunki badań.

Za najważniejsze osiągnięcia Autora rozprawy uważam:

- Stworzenie algorytmu przetwarzania wstępnego danych wizyjnych, zawierającego przekształcenie obrazu RGB do postaci wyekstrahowanych binarnych cech, które dalej przetwarzane są w ograniczonej maszynie Boltzmanna w celu wydobycia z nich najważniejszych informacji i zależności między nimi, które można użyć dalej do klasyfikacji.
- Ulepszenie lokalnego deskryptora binarnego, który koduje nie tylko kształt danej części obrazu, ale także informacje o kolorze i jasności.
- Zaproponowanie metody modyfikacji danych wyjściowych z procesu przygotowania wstępnego w RBM, które zwiększają odporność dalszej sieci klasyfikującej (CNN) na zakłócenia w pierwotnych danych wejściowych.
- Zaproponowanie metody i miary porównania obrazowych danych wejściowych.
- Bardzo cenna jest weryfikacja praktyczna stworzonych algorytmów na docelowym urządzeniu wbudowanym jakim był autonomiczny robot mobilny. Eksperymenty wykazały, że dodanie przetwarzania wstępnego poprawia jakość rozpoznawania przez dalsze sieci neuronowe oraz pozwala na zmniejszenie ich rozmiaru.

3. Uwagi krytyczne i dyskusyjne

1. Drobne błędy w oznaczeniach: str. 28, 1-2 wiersz od góry, czas próbkowania oznaczony raz wielką a raz małą literą t.
2. Nasuwa się pytanie w jaki sposób kodowana jest informacja o kolorze, gdy we wzorze (2.29) porównywane wartości są równe (wzór stosuje ostre nierówności)? Które kodowanie jasności jest brane we wzorze (2.31)?
3. Brakuje informacji na jakim sprzęcie wykonano badania przedstawione w rozdziale 5.

4. Proszę o komentarz dotyczący statystycznej wiarygodności niektórych eksperymentów z rozdziału 5, gdzie rozpatrywanych było tylko kilka przypadków.
5. Czy stworzone oprogramowanie jest lub będzie udostępnione w formie otwartej, a może dodane do zasobów, z których korzystał Doktorant?
6. Czy charakterystyka z Fig. 6.6 jest identyczna dla obu silników?
7. Eksperymenty z rozdziału 6 dotyczą jedynie ruchu obrotowego, czy przemieszczenie liniowe może być rozpoznawane w podobny sposób? Jakie są spodziewane dokładności w zależności od kierunku przemieszczenia (styczne lub normalne do płaszczyzny kamery/obrazu)?
8. Bardzo często stosowanym obecnie środowiskiem implementacji aplikacji robotycznych jest Robot Operating System (ROS), oferujący bardzo wiele gotowych komponentów, zwłaszcza z obszaru lokalizacji, nawigacji i przetwarzania obrazów. Dla obszaru poruszanego w pracy odpowiednim komponentem może być visual odometry (często powiązane z konkretnymi czujnikami wizyjnymi). Czy możliwości visual odometry mogą być porównane z osiągniętymi przez Autora lub czy algorytmy bazujące na histogramach prawdopodobieństw mogą być ujęte w ogólną formę i włączone do ROS?
9. Jaka jest dokładność pomiaru wg metody opisanej w Załączniku C?

4. Podsumowanie i wnioski końcowe

Wymienione uwagi mają głównie charakter dyskusyjny i nie umniejszają podstawowych zalet rozprawy, które wymieniłem w pkt. 2. Z pełnym przekonaniem uważam, że mgr inż. Szymon Sobczak wykazał się znakomitymi umiejętnościami prowadzenia pracy naukowej w dyscyplinie automatyka i robotyka a w nowej klasyfikacji automatyka, elektronika, elektrotechnika i technologie kosmiczne. W mojej ocenie, zawartość merytoryczna przedstawionej rozprawy spełnia z pewnym nadmiarem wymagania stawiane przez art. 13 Ustawy z dnia 14 marca 2003 r o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. z dnia 21.06.2016r, poz. 882). Praca jest obszerna i bardzo starannie zredagowana, a jej przygotowanie w języku angielskim daje szansę na cytowania ważne dla rozwoju kariery naukowej Autora.

