Trypucie, 22 marca 2026

Marek J. Drużdżel
Politechnika Białostocka
Wydział Informatyki
m.druzdzel@pb.edu.pl

Recenzja rozprawy doktorskiej
pana Mohd Faizan Ansariego
zatytułowanej
Using a Camera to Determine Human Gaze Point

Układ rozprawy i jej części składowe

Rozprawa składa się z siedmiu rozdziałów, z których pierwsze trzy (1-3) mają charakter wprowadzający, kolejne trzy (4-6) przedstawiają trzy różne kierunki pracy, a ostatni rozdział (7) zawiera wnioski i plany dalszych prac. Rozprawa zawiera materiał z dwóch publikacji (rozdziały 4 i 5) oraz materiał, który prawdopodobnie stanie się trzecią publikacją (rozdział 6). Dodatkowe, nienumerowane części rozprawy to: streszczenie (w języku angielskim i polskim), bibliografia, spis rysunków i spis tabel. Jest to typowy układ spotykany w rozprawach doktorskich i tym samym jest zadowalający.

Piśmiennictwo

Język angielski w rozprawie pana Mohda Faizana Ansariego jest bardzo dobry i nie wymaga wielu poprawek. W dalszym ciągu tej recenzji zamieszczam kilka drobnych błędów i poprawek.

Cel pracy

Rozprawa doktorska pana Mohda Faizana Ansariego koncentruje się na problemie estymacji ruchu gałek ocznych. Problem ten jest istotny i ma wiele zastosowań, takich jak zrozumienie intencji i uwagi użytkownika, sposobu myślenia i nastroju.

Tradycyjną metodą, opracowaną i stosowaną w psychologii poznawczej, jest wykorzystanie okulografu. Problemem tego urządzenia jest jego koszt i konieczność zatrudnienia wyspecjalizowanego personelu do jego obsługi, chociaż obecnie dostępne są okulografy obsługiwane przez użytkownika. Ich cena jest nadal dość wysoka. Rozprawa doktorska pana Ansariego koncentruje się na śledzeniu ruchu gałek ocznych za pomocą prostej kamery komputerowej – urządzenia, które jest zarówno niedrogie, jak i łatwe w obsłudze. Autor poświęca szczególną uwagę wyzwaniom wynikającym z parametrów technicznych standardowych kamer komputerowych, w tym ich niskiej

rozdzielczości, położenia względem głowy użytkownika oraz zmiennych warunków oświetleniowych.

Istnieje mnóstwo publikacji z zakresu śledzenia ruchu gałek ocznych (proste wyszukiwanie w Google Scholar wskazuje ponad cztery miliony!), śledzenia spojrzenia (ponad milion!), śledzenia ruchu gałek ocznych za pomocą kamery internetowej (ponad 20 tysięcy!), śledzenia ruchu gałek ocznych za pomocą kamery internetowej opartej na sieciach neuronowych (ponad 20 tysięcy!) i śledzenia ruchu gałek ocznych za pomocą kamery internetowej opartej na sieciach neuronowych konwolucyjnych (ponad cztery tysiące!). To sugeruje, że jest to obszar wysokiego ryzyka, ponieważ konkurencja jest ogromna. Stwierdzenie autora: „Istnieją prace podobne do naszych" (strona 35) wydaje się nie oddawać sprawiedliwości ogromnej konkurencji w tej dziedzinie. Mimo to wydaje się, że pan Ansari znalazł niszę w tej dziedzinie. Moje badania nie są nawet w najmniejszym stopniu związane z tą dziedziną i nie śledziłem literatury na tyle uważnie, aby móc wiarygodnie stwierdzić, czy wkład pana Ansariego jest nowatorski. Ocenę unikalności wkładu Pana Ansariego pozostawiam jego promotorowi, prof. Pawłowi Kasprowskiemu, anonimowym recenzentom dwóch opublikowanych artykułów, których materiały zostały włączone do rozprawy, oraz moim kolegom, pozostałym recenzentom rozprawy. W mojej recenzji skupię się na konstruktywnej krytyce zaprezentowanych prac w tych obszarach, w których czuję się w miarę kompetentny.

Spośród trzech hipotez zaproponowanych w rozprawie, H1 wydaje się być przetestowana przez innych badaczy – istnieje wiele innych prac, w których zastosowano konwolucyjne sieci neuronowe do śledzenia ruchu gałek ocznych za pomocą kamer internetowych. Nie znam aktualnej literatury w tej dziedzinie, aby ocenić, czy dopasowanie modelu CNN do konkretnej osoby (hipoteza H2) zostało wystarczająco przetestowane i czy niniejsza rozprawa oferuje istotne nowe spostrzeżenia na ten temat.

Hipoteza H3 (dotycząca techniki zwanej „uczeniem transferowym") jest interesująca i przypomina mi technikę stosowaną w bayesowskich podejściach do uczenia się. Zanim nastąpi faktyczne uczenie się, model można wytrenować na danych populacyjnych, a nawet na parametrach uzyskanych od ekspertów (bayesowskie podejście do prawdopodobieństwa pozwala na uzyskanie i przetwarzanie ocen ekspertów). To, co w uczeniu maszynowym nazwano „uczeniem transferowym", wygląda jak po prostu przykład aktualizacji bayesowskiej – statystyka ogólnego przeznaczenia jest dostosowywana do problemu lub badanej osoby. Nie dziwi mnie, że ta korekta prowadzi do dokładniejszych modeli.

Podsumowując, cel rozprawy został osiągnięty i postawione hipotezy zostały pomyślnie przetestowane.

Zastosowane metody badawcze

Autor połączył projektowanie, implementację i testowanie proponowanych przez siebie rozwiązań. Chociaż mam kilka pytań, które zadaję poniżej, uważam, że ogólnie rzecz biorąc podejście jest poprawne.

Interesujące byłoby poznanie, jak proponowane metody sprawdzają się w przypadku kamer o wysokiej rozdzielczości i lepszym oświetleniu. Trend technologiczny zmierza w kierunku produkcji lepszych i tańszych urządzeń (na przykład o wyższej rozdzielczości) z biegiem czasu, a więc definicja „niskiej rozdzielczości" będzie się z czasem zmieniać. Można by pokusić się o inne podejście do testowania hipotez: wykorzystanie jednego ze zbiorów danych o wysokiej rozdzielczości, takich jak te wymienione w rozprawie, i wstępne przetworzenie ich w celu systematycznego obniżenia rozdzielczości. Rozpoczęcie badania od zbioru danych o wysokiej rozdzielczości pozwoliłoby na bardziej miarodajne porównanie wyników z wynikami uzyskanymi przez innych badaczy korzystających z tych samych zbiorów danych.

Jedną ze słabości przeprowadzonych eksperymentów jest to, że przeprowadzono je na danych zebranych od jednej osoby. W zależności od cech fizycznych tej osoby (np. kształtu i koloru oczu), wyniki mogą być specyficzne dla tej osoby. To właśnie zmienność między osobami daje nam wariancję, kluczową dla wyprowadzenia istotności statystycznej lub przedziałów ufności dla wyników.

Zbiór danych wykorzystany w „transferze edukacyjnym" (rozdział 6) obejmuje pomiary różnych osób (Azjatów, Afrykanów i Europejczyków). Bardzo interesujące byłoby sprawdzenie dokładności w każdej z tych trzech grup osób oddzielnie. Cechy fizyczne każdej z trzech grup (takie jak kolor lub kształt oczu) mogą wpływać na dokładność.

Zauważyłem, że autor używa różnych miar dokładności w różnych rozdziałach (błąd kątowy, błąd minimalny, średni błąd bezwzględny). Czy nie byłoby lepiej użyć jednej miary (błędu kątowego?) lub być może dwóch miar jednocześnie, ponieważ błąd kątowy może być zbyt przybliżony, na przykład przy zaledwie 20 predefiniowanych obszarach ekranu? Czasami trudno było zrozumieć w tekście, co dokładnie zostało zmierzone. Na przykład tabela 4.3 przedstawia wyniki eksperymentu. Jaka mierzona jest tu dokładność? Czy jest to trafne przewidywanie jednego z 20 regionów? Dobrym pomysłem może być przedstawienie każdego z rodzajów błędów. Zakładam, że wiele błędów dotyczyło wskazania sąsiednich regionów. Z tej perspektywy błąd kątowy byłby bardzo interesujący. Sugeruję przedstawienie obu błędów. W rozdziale 5, tabele 5.7 i kolejne zawierają minimalny błąd. Czy maksymalny błąd nie byłby bardziej interesujący? Biorąc pod uwagę zamieszanie, jakie pojawia się wokół różnych miar błędu, znalazłem następujące zdanie (strona 73): „Na podstawie naszych ustaleń oczywiste jest, że nasze wyniki przewyższyły wyniki poprzednich badań" niezbyt oczywistym.

Omówienie wyników badań

Dyskusja uzyskanych wyników wydaje się prawidłowa i nie stwierdziłem żadnych poważniejszych problemów dotyczących wewnętrznej i zewnętrznej spójności w pracy eksperymentalnej poza uwagami zawartymi w poprzedniej sekcji.

Praktyczne zastosowania pracy

Wyniki pracy są obiecujące. W szczególności modele dobrze dostosowane do konkretnego użytkownika działają lepiej niż modele trenowane na podstawie danych pochodzących z interakcji z wieloma użytkownikami. Autor wykazuje, że modele oparte na transferze danych są dokładniejsze w sytuacjach ograniczonej liczby danych niż modele z wagami wyuczonymi na ich podstawie. Wyniki uzyskane za pomocą prostej kamery internetowej o niskiej rozdzielczości są porównywalne z wynikami uzyskanymi za pomocą droższego i bardziej skomplikowanego sprzętu. Z tej perspektywy można oczekiwać, że prace te znajdą praktyczne zastosowanie w wielu obszarach, takich jak gry, reklama czy wywiady i przesłuchania.

Nieprawidłowości wykryte w pracy

Nie znalazłem żadnych nieprawidłowości w tej pracy. Moje uwagi i sugestie powyżej i poniżej powinny ulepszyć pracę, ale sama w sobie praca zasługuje na doktorat.

Ocena oryginalności rozprawy

Jak wyjaśniłem powyżej, uważam, że rozprawa rozwiązuje oryginalny problem naukowy. Chociaż nie jestem ekspertem w tej dziedzinie i nie mogę z całą pewnością stwierdzić, że proponowane rozwiązania są nowatorskie, istnieją silne przesłanki, które na to wskazują. Dwa z trzech głównych artykułów zostały opublikowane i zrecenzowane przez niezależnych, anonimowych recenzentów. Kandydat był pod opieką prof. Pawła Kasprowskiego, eksperta w tej dziedzinie. Zakładam, że pozostali dwaj recenzenci rozprawy przedstawią wiarygodną opinię na temat jej oryginalności.

Ogólna ocean

Ogólnie rzecz biorąc, moje wrażenie na temat rozprawy doktorskiej pana Mohda Faizana Ansariego jest pozytywne. Kandydat wykazał, że jest zdolny do samodzielnego prowadzenia wysokiej jakości badań i moim zdaniem zasługuje na doktorat. Zalecam umożliwienie kandydatowi przejścia do kolejnych etapów przewodu doktorskiego.

Mniej istotne uwagi i pytania

W akapicie na dole strony 86 kandydat wspomina o kilku przeprowadzonych testach istotności statystycznej. Należy być świadomym zagrożeń związanych z testami wielu hipotez. Ile testów tam wykonano?

Nie rozumiałem, dlaczego niepoprawnie sklasyfikowane obrazy zostały ręcznie usunięte (strona 60). Czy nie wpływa to bezpośrednio na miarę dokładności metody?

Po co badać wyniki dla lewego i prawego oka oddzielnie? Sądzę, że poza kilkoma klinicznie interesującymi przypadkami, oba oczy koncentrują się na tym samym kierunku.

Powód gorszych wyników w tabeli 5.8 autor wyjaśnia następująco: „liczba obrazów treningowych (1880) jest mniejsza w porównaniu z liczbą obrazów testowych (2360)". Czy nie jest to właśnie problem nadmiernego dopasowania?

Być może lepiej byłoby połączyć część materiału wprowadzającego z rozdziałów 4 (ostatni akapit na stronie 41 i pierwsze dwa akapity na stronie 42), 5 (trzy pierwsze akapity na stronie 58) i 6 z rozdziałami 1-3. Zdaję sobie sprawę, że wprowadzenia do artykułów zawierały ogólny materiał wprowadzający, ale w przypadku rozprawy lepiej, aby rozdziały 4, 5 i 6 były wolne od materiału wprowadzającego, chyba że jest on bezpośrednio związany z przeprowadzanymi eksperymentami.

Nie zrozumiałem szczegółów konfiguracji przy akwizycji danych, ale zakładam, że użytkownicy byli proszeni o kliknięcie w obszar, który emitował sygnał wizualny, i zakładam, że z założenia patrzyli na punkt lub obszar, w który klikali? Dobrym pomysłem może być pokazanie w pracy zrzutu ekranu przedstawiającego rzeczywistą interakcję użytkownika podczas gromadzenia danych, np. w sekcji 4.1.1.

Drobne uwagi i poprawki

Język angielski rozprawy jest bardzo dobry i nie wymaga wielu poprawek. Poniżej wskażę tylko kilka problemów.

W technicznym angielskim lepiej jest unikać strony biernej. Nie ma zbyt wielu przypadków strony biernej w rozprawie, ale zauważyłem kilka, zwłaszcza w tekście zaczerpniętym z dwóch opublikowanych artykułów.

Zwyczajowo unika się form krótkich (takich jak „let's", „it's", „there's", itp.) w artykułach i rozprawach. W rozprawie jest ich zaledwie kilka, co jest dobre. Sugeruję, aby zastąpić je ich pełnymi odpowiednikami, tj. „let us", „it is", „there is", itp.

Wzory matematyczne należy ogólnie traktować jako części zdań, a zatem zazwyczaj stawia się po nich przecinek lub kropkę, w zależności od zdania je otaczającego. Na przykład równanie (2.6) powinno kończyć się kropką, a następne zdanie powinno zaczynać się wielką literą (np. „Minimize..."). Równanie (2.8) również powinno kończyć się kropką, natomiast równanie (2.9) przecinkiem, ponieważ istnieje tekst kontynuujący jego opis (przed słowem „Where" nie powinno być myślnika, a słowo „where" nie powinno zaczynać się wielką literą). Podobnie jest ze wszystkimi innymi równaniami.

W przypadku cytowania kilku artykułów z rzędu, np. strony 42, [101,66], zazwyczaj sortujemy cytowania, tj. w tym przypadku wpisujemy [66,101].

Rysunek 3.1 jest nieco sprzeczny z intuicją, ponieważ jego celem jest pokazanie osi czasu, podczas gdy przedstawia on to, co nazwałbym „zawijaniem czasu", gdzie rok 2024 jest poprzedzony rokiem 1908. Myślę, że bardziej odpowiedni byłby wykres liniowy, być może zawinięty na kilka linii.

Strona 19: Słowo „data" w języku angielskim jest w liczbie mnogiej, dlatego należy pisać „data are" zamiast „data is".

Strona 25, Sekcja 3.2.1: „Model based" => „Model-based"

Strona 31: Brakuje spacji przed cytatem [53].

Strona 31: „X Zhang et al." => Zhang et al." Inicjał imienia nie jest potrzebny.

Strona 32: Cytat Dubay et al. jest błędny – powinien brzmieć [36]. Usunąłbym imię („Neeru") i umieściłbym cytat zaraz po nim, tj. „et al. [36]".

Strona 35: "advance in term of complexity" => "advanced in terms of complexity"

Strona 35: "maybe be" => "may be"

Strona 35: "a wide range of subjects and exhibiting various environmental conditions" => "a wide range of subjects under various environmental conditions"

Cyfry od zera do dziesięciu należy w języku angielskim zapisywać słownie, a cyfry wyższe cyframi. Na przykład na stronie 36 powinniśmy napisać „Specifically, dataset has five head poses for each subject, with 21 gaze directions per head pose." zamiast „Specifically, dataset have 5 head poses for each subject, with 21 gaze directions per head pose."

Strona 36: "This dataset aims to tackle the challenge" => "This dataset aims at tackling the challenge"

Strona 36: "It include" => "It includes"

Strona 36: "boasts" => "includes"

Strona 39: Niepotrzebne spacje przed cytowaniami [175] i [35].

Strona 40: Błąd formatowania w cytowaniu „citezhang2021eye".

Strona 41: Dobrym pomysłem może być podanie pełnego cytowania artykułu, na którym oparty jest rozdział.

Strona 43: Wyrażenie „in the wild" wydaje się tu nieodpowiednie. Może po prostu je usunąć? Otaczające wyrażenia wyjaśniają, w jaki sposób zebrano dane.

Podczas zapisywania dużych liczb dobrym pomysłem jest umieszczenie przecinka jako separatora tysięcy. Na przykład na stronie 47 napisałbym „6,000 images".

Strona 52: "allocated for the test set of Dataset" => "allocated for the test set"

Strona 95: “from the scratch” => “from scratch”

Strona 95: “Impact of Dataset Amount on Convergence” => “Impact of Dataset Size on Convergence”

Strona 95: “This behavior demonstrates the efficiency of transfer learning in adapting diverse data sizes.” => “This behavior demonstrates the efficiency of transfer learning in adapting to diverse data sizes.”

Strona 95: “particularly in scenarios where datasets can be limited, noisy, or challenging to obtain” => “particularly in scenarios where data can be limited, noisy, or hard to obtain”

Strona 99: “In this thesis, multiple approaches were developed” => “In this thesis, three approaches were developed”