



Obserwatorium Astronomiczne Uniwersytetu Warszawskiego

Al. Ujazdowskie 4, 00 478 Warszawa
tel: (22) 55-30-507, fax: (22) 629-49-67
www.astrouw.edu.pl

Warszawa, 31.08.2022 r.

Prof. dr hab. Szymon Kozłowski
Obserwatorium Astronomiczne
Uniwersytetu Warszawskiego
Al. Ujazdowskie 4
00-478 Warszawa
e-mail: simkoz@astrouw.edu.pl

Recenzja rozprawy doktorskiej mgr. Szymona Nakonecznego pt. „Machine learning based catalogs of quasars and galaxies for cosmological studies”

Rozprawa doktorska przygotowana przez Pana magistra Szymona Nakonecznego jest dysertacją, którą oceniam zdecydowanie pozytywnie. Jest to praca stanowiąca dowód wysokiej ekspertyzy analitycznej, a przede wszystkim informatycznej, umiejętności prowadzenia badań, analizy danych i konstruowania wniosków prowadzących do kolejnych ulepszeń w działaniu prezentowanych metod. Co więcej, podjęty temat jest aktualny i istotny dla naszego zrozumienia powstawania i ewolucji struktur we Wszechświecie.

Przedłożona mi do recenzji rozprawa doktorska podejmuje problematykę stworzenia dużych katalogów kwazarów z danych projektu Kilo-Degree Survey (KiDS) wymaganych w badaniach dotyczących rozkładu kwazarów w przestrzeni i ich relacji do niejednorodności mikrofalowego promieniowania tła.

W rozdziale pierwszym doktorant prezentuje bardzo krótkie wprowadzenie do zagadnień poruszanych w pracy. Jest ono stanowczo za krótkie i jest to wg mnie główny mankament tej dysertacji. W tym miejscu spodziewałbym się od doktoranta przekazania swojej wiedzy z podstaw podejmowanej tematyki astronomicznej. Obecna 2,5 stronicowa wersja przypomina raczej wstęp publikacji. Spodziewałbym się tutaj wyjaśnienia czym są AGNy i kwazary. Co „świeci” i „jak świeci” w AGNach/kwazarach, galaktykach spiralnych/eliptycznych, czy gwiazdach – obiektach dyskutowanych na dalszych stronach dysertacji. Jak wyglądają widma/SED tych obiektów? Jak te widma/SED mają się do użytych filtrów w funkcji przesunięcia ku czerwieni. Jakie są przepuszczalności tych filtrów?

W rozdziale drugim doktorant przechodzi do przedstawienia danych użytych w doktoracie oraz zastosowanej metodologii. Omówione zostają projekty Kilo-Degree Survey, w tym Data Release 3 i 4 (DR3, DR4), Viking, WISE, czy SDSS. Szczegółowo zaprezentowane zostało tworzenie zestawów danych (inference, training i validation sets) dla metod uczenia maszynowego (Random Forest (RF), XG-boost (XGB) i Artificial Neural Network (ANN)). Przedstawiona zostaje

także metoda auto i cross korelacji pomiędzy materią i kwazarami (quasar overdensity), oraz funkcja biasu (bias function) zależna od przesunięcia ku czerwieni (redshift).

W rozdziale trzecim doktorant przechodzi do zasadniczej analizy danych bazujących na obserwacjach z projektu KiDS DR3. Wybranych zostaje 17 cech: wielkości gwiazdowe w filtrach ugri, ich kolory i stosunki oraz cecha CLASS_STAR. Następnie po testach następuje wybór najlepszego modelu uczenia maszynowego do klasyfikacji kwazarów (Random Forest) oraz wizualna prezentacja wyników. Ciekawym dodatkiem jest weryfikacja stworzonego katalogu kwazarów poprzez ich analizę w danych z satelity Gaia (ruchy własne oraz paralaksy). Rozdział kończy się analizą kompletności próbki, porównaniem ze spektroskopowym katalogiem kwazarów 2DF oraz fotometrycznymi katalogami na bazie SDSS i WISE. Katalog kandydatów na kwazary zawiera 190 000 obiektów. Dokładność klasyfikacji wyniosła 97%, czystość próbki 91%, a kompletność to 87%. Wyniki tego rozdziału opublikowane zostały w renomowanym czasopiśmie astronomicznym *Astronomy & Astrophysics*, 2019, 624, A13, artykuł pt. „Catalog of quasars from the Kilo-Degree Survey Data Release 3” i zostały dobrze odebrane przez środowisko astronomiczne – praca ta cytowana była 26 razy wg NASA ADS (sierpień 2022).

W rozdziale czwartym doktorant analizuje dane z projektu KiDS DR4, które to są powiększone w stosunku do tych rozważanych w rozdziale trzecim. Poza danymi optycznym (filtry ugri), dane te zawierają także obserwacje podczerwone w filtrach ZYJHK_s. Kombinacja wielkości gwiazdowych, ich kolorów i stosunków oraz pomiar „gwiazdowości” łącznie tworzą zbiór o 83 cechach. Doktorant testuje modele uczenia maszynowego (RF, XGB i ANN) i dochodzi do wniosku, że dwie sztuczne sieci neuronowe dają najlepsze wyniki klasyfikacji i pomiaru przesunięcia ku czerwieni. Zaprezentowana jest jakość klasyfikacji, kompletności i czystości próbki w funkcji redshiftu. W tym rozdziale stworzony zostaje katalog zawierający około 300 000 kandydatów na kwazary. Wyniki tej pracy opublikowane zostały w renomowanym czasopiśmie astronomicznym *Astronomy & Astrophysics*, 2021, 649, A81, artykuł pt. „Photometric selection and redshifts for quasars in the Kilo-Degree Survey Data Release 4” i zostały dobrze odebrane przez środowisko astronomiczne – praca ta cytowana była 12 razy wg NASA ADS (sierpień 2022).

W rozdziale piątym doktorant przechodzi do zastosowań astrofizycznych stworzonych przez siebie katalogu kwazarów (DR4), a mianowicie wyznaczenia funkcji korelacji i funkcji biasu. Zaprezentowana zostaje auto-korelacja katalogu oraz cross-korelacja katalogu z soczewkowaną mapą mikrofalowego promieniowania tła w funkcji skali dla kilku zakresów prawdopodobieństwa ($p(\text{QSO})$) i/lub limitów wielkości gwiazdowych. Wyznaczona zostaje także funkcja biasu oraz jej parametry (A,B). Porównana zostaje ona do pracy innych autorów.

Dysertacja podsumowana i przedyskutowana zostaje w rozdziale szóstym. Pokróćce przedstawione zostają wcześniej prezentowane wyniki. Doktorant porównuje swoje wyniki do innych prac naukowych, przeprowadza dyskusję ograniczeń użytej metodologii oraz sposoby jej ulepszenia. Ostatnią częścią pracy doktorskiej jest bibliografia z ponad 100 referencjami.

Poza wstępem i bibliografią praca zawiera pięć zasadniczych rozdziałów o łącznej objętości około 60 stron. Struktura pracy jest klasyczna, czyli w pierwszej kolejności mamy przedstawioną teorię, omówione metody, zaprezentowane badania empiryczne, zakończone dyskusją i wnioskami. Autor jest zaznajomiony z literaturą przedmiotu, a dysertacja zawiera odnośniki (cytuje) ponad 100 prac. Praca jest napisana w języku angielskim. Wiedza zawarta w pracy jest dość mocno skondensowana.

Moje nieliczne uwagi krytyczne są głównie natury merytorycznej:

Strona 1. Już na pierwszej stronie “broad emission lines” to wg doktoranta [OIII]5007/Hbeta i [NII]6584/Halpha. Dodatkowo odnośniki do prac w tym miejscu są nie na temat. Gdyby w pracy znalazł się wstęp w pełnej formie (zawierający typowe widmo kwazara, np. Vanden Berk i in. 2001), zapewne obyłyby się bez takich błędów.

Strona 7. Do czego znormalizowane są osie y na rysunku 2.1?

Strona 24, równanie (2.15). We wzorze występuje kowariancja Cov^{-1} . Nie jest wyjaśnione czego to jest kowariancja.

Strona 57, sekcja 5.2. Wyniki funkcji biasu: Brakuje rysunku konturów parametrów AB. Czy te parametry są skorelowane? Jaki był sposób dopasowywania (w jednym miejscu widzę MCMC)? Same wartości χ^2 nie mówią nic o dobroci dopasowania – jaka była liczba stopni swobody?

Rozprawa doktorska (w tym publikacje Autora) stanowi w mojej ocenie oryginalny wkład w badania nad ewolucją struktur we Wszechświecie. Autor doskonale opanował techniki informatyczne (uczenie maszynowe), umiejętności otwierające wiele drzwi w karierze astronoma, nie tylko w dziedzinie badań kosmologicznych.

Autor dysertacji jest współautorem czterech recenzowanych prac astronomicznych (wszystkie w *Astronomy & Astrophysics*), cytowanych ponad 80 razy. Fragmenty rozdziałów 2-6 opublikowane zostały w renomowanym, recenzowanym czasopiśmie astronomicznym w postaci dwóch pierwszo-autorskich prac doktoranta, co świadczy o wysokiej jakości zaprezentowanych badań.

Wymienione powyżej uwagi krytyczne nie zmieniają mojej pozytywnej opinii o pracy. Rozprawa doktorska Pana mgr Szymona Nakonecznego spełnia formalne i zwyczajowe wymogi stawiane rozprawom doktorskim. Oceniam ją jako bardzo dobrą.

prof. dr hab. Szymon Kozłowski