

Streszczenie

Wspomagane metodami uczenia maszynowego budowanie katalogów kwazarów oraz galaktyk na potrzeby badań kosmologicznych

Szymon NAKONECZNY

Prezentujemy dwa katalogi kwazarów zbudowane na podstawie danych z trzeciego oraz czwartego wydania (DR3, DR4) przeglądu Kilo-Degree Survey (KiDS). Prezentujemy dwie metodologie, pierwszą do klasyfikacji źródeł w KiDS DR3, oraz drugą, bardziej zaawansowaną, w której dodajemy dane z bliskiej podczerwieni, estymujemy fotometryczne przesunięcia ku czerwieni, oraz ekstrapolujemy modele uczenia maszynowego (ML) na magnitudy ciemniejsze niż te znane z przeglądów spektroskopowych. Analiza danych z KiDS DR4 produkuje ostateczny katalog kwazarów, którego wyniki opisują na tej stronie. Katalog oparty jest o optyczne filtry *ugri*, oraz bliską podczerwień *ZYJHK_s*, a modele ML trenowane są na danych znanych ze spektroskopowego przeglądu Sloan Digital Sky Survey (SDSS). W 45 milionach obiektów ograniczonych do detekcji we wszystkich 9 pasmach definiujemy podzbiory inferencyjne na podstawie przestrzeni cech zbudowanej z magnitud oraz ich kombinacji. Pokazujemy, że projekcje wysoko wymiarowej przestrzeni cech na dwa wymiary mogą być użyte, zamiast standardowych wykresów kolor-kolor, w celu wizualizacji wyników. Kalibracji modeli uczenia maszynowego dokonujemy za pomocą dwóch podzbiorów walidacyjnych: losowo wybranego z całego zakresu magnitud, oraz ekstrapolacyjnego, do którego wybieramy najciemniejsze magnitudy, które celowo wyłączamy z treningu. Takie podejście pozwala nam na prawidłowe dopasowanie kompromisu między obciążeniem a wariacją. Testujemy trzy modele ML: las losowy (RF), XGBoost (XGB), oraz sztuczna sieć neuronowa (ANN). Pokazujemy, że XGB dostarcza najlepszych wyników klasyfikacji, zaś ANN pozwala osiągnąć najlepsze wyniki dla klasyfikacji i przesunięcia ku czerwieni. Ostateczny katalog, zbudowany za pomocą ANN, testujemy metodą zliczeń, paralaksami z przeglądu Gaia, oraz zewnętrznymi katalogami kwazarów. Na podstawie tych testów, znajdujemy minimalne prawdopodobieństwo klasyfikacji, które kalibruje kompromis między czystością oraz kompletnością, i wynosi: $p(\text{QSO}_{\text{cand}}) > 0.9$ dla $r < 22$ oraz $p(\text{QSO}_{\text{cand}}) > 0.98$ dla $22 < r < 23.5$. W katalogu znajduje się 158,000 kandydatów na kwazary w bezpiecznym zakresie $r < 22$, oraz kolejne 185,000 kandydatów w zasięgu ekstrapolacji $22 < r < 23.5$. Na podstawie danych testowych szacujemy czystość oraz kompletność katalogu na odpowiednio 97% oraz 94%, a w wyniku ekstrapolacji o jedną magnitudę, kompletność maleje o 3%. Przesunięcia ku czerwieni modelujemy za pomocą rozkładu gaussowskiego, i błąd tej estymacji (średnia oraz rozrzut) wynosi 0.009 ± 0.12 w bezpiecznym zakresie, oraz -0.0004 ± 0.19 w ekstrapolacji, policzone jako średnia w dla $0.14 < z < 3.63$. Sukces ekstrapolacji stawia wyzwanie temu jak modele uczenia maszynowego są kalibrowane oraz stosowane do ciemnych obiektów. Katalog wynikowy jest gotowy do badań kosmologicznych oraz nad aktywnymi jądrami galaktyk (AGN), i w niniejszej pracy przedstawiamy pierwsze wyniki oszacowania funkcji biasu z użyciem korelacji krzyżowej z soczewkowaniem mikrofalowego promieniowania tła (CMB). Otrzymujemy wynik w postaci $b_q(z) = 0.57_{-0.03}^{+0.03}(1+z)^2 + 0.07_{-0.13}^{+0.06}$, co dla $z = 1.5$ daje wartość biasu $3.63_{-0.85}^{+0.25}$. Detekcja korelacji krzyżowej wynosi 15σ , co jest obecnie jedną z najsilniejszych detekcji tego sygnału.