

Streszczenie

Application of machine learning methods in astrophysics and cosmology

Artem POLISZCZUK

Celem przedstawionej pracy było opracowanie nowych technik selekcji aktywnych jąder galaktyk opartych na algorytmach uczenia maszynowego, które pozwoliłyby na efektywne przeszukiwanie dużych zbiorów danych fotometrycznych z skutecznością, a niedostępną dla tradycyjnych metod selekcji. Zastosowania takiej metody pozwala stworzyć wysokiej jakości katalogi aktywnych jąder galaktyk do zastosowań w astronomii pozagalaktycznej i kosmologii obserwacyjnej. Badania przedstawione w pracy pokazują, że możliwe jest stworzenie modelu opartego na algorytmach uczenia maszynowego, który jest w stanie naśladować selekcję aktywnych jąder galaktyk w zakresie średniej podczerwieni, używając jedynie danych fotometrycznych z zakresu optycznego i bliskiej podczerwieni. Taka metoda zapewnia wysoką efektywność selekcji charakterystyczną dla technik stosowanych w średniej podczerwieni. Jednocześnie nowa metoda pozwala uniknąć znacznego zmniejszenia rozmiaru katalogu wynikającego z wymogu pomiaru w średniej podczerwieni.

W pracy wykorzystano dane z głębokiego przeglądu nieba w polu AKARI NEPWide. Wprowadzono szereg rozwiązań nie stosowanych dotąd w astronomii. Mechanizm naśladowania przez model selekcji opartej na średniej podczerwieni został uzyskany poprzez dwa podstawowe mechanizmy. Pierwszym z nich było pośrednie dostarczenie informacji na temat technik selekcji w średniej podczerwieni zawartych w konstrukcji próbki treningowej. Drugim było ograniczenie ryzyka ekstrapolacji w danych spoza próbki treningowej poprzez zastosowanie algorytmu Najmniejszego Wyznacznika Kowariancji. Zastosowana technika pozwoliła efektywnie ograniczyć obszar w wielowymiarowej przestrzeni cech do regionu reprezentowanego przez dane treningowe.

Ponadto przeprowadzono badania nad efektywnością i zastosowania różnych technik logiki rozmytej w selekcji aktywnych jąder galaktyk na podstawie różnych nadzorowanych algorytmów klasyfikacyjnych. Następnie zbadano efektywność zastosowania metod wyszukiwania anomalii oraz technik niskowymiarowej wizualizacji danych do znajdowania zanieczyszczeń katalogu wynikowego. Pozwoliło to zidentyfikować przypadki niepoprawnej fotometrycznej estymacji przesunięcia ku czerwieni, a także potencjalne grupy nieprawidłowo sklasyfikowanych obiektów.

Metody wprowadzone w pracy pozwalają na ominięcie trudności wynikających z ograniczeń instrumentów pomiarowych, a także umożliwiają precyzyjną kontrolę nad jakością katalogu wynikowego oraz rozpoznanie potencjalnych źródeł zanieczyszczeń. Pozwala to na dopasowanie katalogu wynikowego do potrzeb konkretnych zastosowań i tworzy efektywny zestaw narzędzi dla współczesnej i przyszłej astrofizyki.