[구GC-27] QSO Selections Using Time Variability and Machine Learning

Dae-Won Kim[1,2], Pavlos Protopapas[1], Yong-Ik Byun[2], Charles Alcock[1], Roni Khardon[3]

[1]*Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA*
[2]*Department of Astronomy, Yonsei University, Seoul, South Korea*
[3]*Computer Science Department, Tufts University, Medford, MA, USA*

We present a new quasi-stellar object (QSO) selection algorithm using a Support Vector Machine, a supervised classification method, on a set of extracted time series features including period, amplitude, color, and autocorrelation value. We train a model that separates QSOs from variable stars, non-variable stars, and microlensing events using 58 known QSOs, 1629 variable stars, and 4288 non-variables in the MAssive Compact Halo Object (MACHO) database as a training set. To estimate the efficiency and the accuracy of the model, we perform a cross-validation test using the training set. The test shows that the model correctly identifies ~80% of known QSOs with a 25% false-positive rate. The majority of the false positives are Be stars. We applied the trained model to the MACHO Large Magellanic Cloud (LMC) data set, which consists of 40 million lightcurves, and found 1620 QSO candidates. During the selection, none of the 33,242 known MACHO variables were misclassified as QSO candidates. In order to estimate the true false-positive rate, we crossmatched the candidates with astronomical catalogs including the Spitzer Surveying the Agents of a Galaxy's Evolution (SAGE) LMC catalog and a few X-ray catalogs. The results further suggest that the majority of the candidates, more than 70%, are QSOs.