Clustering of Cross-Referenced Astronomical Data Sets

Vera Abaimova — stormseecker@gmail.com
Mark Wells — mwellsa@gmail.com

December 9, 2014

Abstract

The large amount of astronomical data that is now available from a multitude of missions creates a need for machine learning methods that can analyze it and glean information about our universe. One challenge in particular is to analyze data about the same observations made by different missions, *i.e.*, analyze cross-referenced data, especially data in different formats. One such format is time series data provided by the *Kepler* mission, which results in additional difficulties. Our proposed method extracts features from cross-referenced data sets, including time series data, and applies a hierarchical clustering model in order to aid stellar classification.

1 Introduction

2 Related Work

3 Problem Definition

The input of our clustering problem is a list of $x_{i,j}$ where x_i is the star object, numbered from 1 to 222,158 and j is the feature, numbered from 1 to 14. Our feature space includes features such as fuv_mag, nuv_flux, and $kepmag$, among others, where fuv_mag is the far ultraviolet magnitude, nuv_flux is

the near ultraviolet, and *kepmag* is a measure of brightness of an object in the *Kepler* pass band.

Our output is a list of $x_{i,j} \in C_k$ where C is cluster and k denotes cluster number.

4   Method

5   Experiments

6   Summary