

Classification Methods for the Analysis of High Content Screening Data

The current paradigm for the identification of candidate drugs within the pharmaceutical industry typically involves the use of high throughput screens. A high throughput screen allows a large number of compounds to be tested in a biological assay in order to identify any activity inhibiting or activating a biological process. From each of the assays run through a high throughput screen a high content screen image is produced which can be analysed using advanced imaging algorithms to produce a set of variables which reflect the observed activity of the cells within the image.

Classification methods have important applications in the analysis of high content screening data where they are used to predict which compounds have the potential to be developed into new drugs. Statistical approaches have been developed that enable classification using a single parameter. However, approaches for multi-parametric selection are still in their infancy. Furthermore, proper exploitation of the information contained within each high content screen image will enable more refined compound selection.

A new classification technique for the analysis of data from high content screening experiments will be presented and the methodology illustrated on an example data set using a random forest classifier.