Unsupervised photometric membership assignment in stellar clusters

Alberto Krone-Martins[1], André Moitinho de Almeida[1]

[1]SIM, Faculdade de Ciências da Universidade de Lisboa (SIM/FC-UL)
{algol, andre}@sim.ul.pt

One of the most ancient problems in the photometric study of stellar clusters is the assignment of stellar membership. Although several approaches are available for attacking this problem, they usually involve the adoption of complex theoretical models for the photometric data (called isochrones) and/or the selection and use of comparative control fields, possibly biasing some analysis.

In order to address this issue, we have developed a data-driven, automated and unsupervised method to perform membership assignment in Stellar Clusters using photometric and spatial data. Our method relies on the weakest assumption for a stellar cluster: the member stars shall be clustered in most spaces and also in the positional space, while even if field stars may be clustered in certain spaces, they are not expected to cluster also in the positional space.

This method is independent from complex models, as well as from observational control fields. It is based on an iterative solution, and principal component analysis (PCA), clustering algorithms and kernel density estimations (KDE). Optionally, it allows the user to take into account error models and missing data.

It relies on two iterative processes. In the first one, the original observables are transformed to the PCA space, where a randomly initialized k-means clustering analysis (with a large k) is performed on a certain number of the principal components (in this work, we selected the first four components due to physical motivations of the expected information content available from stellar photometry). Then, for each cluster detected in the PCA space, the method checks if the stars are also clustered on the positional space using KDEs and comparison with KDEs of random fields. If they are clustered, they are retained for the next iteration, otherwise they are classified as field stars and are not further considered. The processes iterates until convergence is attained: or the list of stars under analysis are not changing between successive iterations, or it is empty.

The second iterative process is based on running the first iterative process a large number of times. However, not on the reported values from the observables, but on a list created from random trials of the probability distribution function of each observable. Finally, all the results from all the runs are merged, and the frequency that each star was classified as cluster member over all the runs is considered as its cluster membership probability.

We will present a description of the method, and results obtained with the application of its R implementation on a set of realistic simulations produced with the MASSCLEAN and TRILEGAL codes. Finally, we will present results obtained on real data of selected Open Clusters, as Haffner 16, Haffner 10 and Czernik 29.