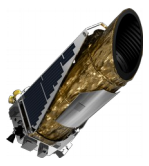
Obtaining a classification of A-F stars through clustering analyzing the morphology of the light curves

Rodón Ortiz, JR [1], Pascual-Granado, J [1], Moya, A [2]

1 Departamento de Física Estelar. Instituto de Astrofísica de Andalucía, Granada, Spain

2 Departament d' Astronomia i Astrofísica, Universitat de Valencia, Burjassot, Spain

DESCRIPTION



Asteroseismology is experiencing a revolution thanks to high-precision asteroseismic space missions (Kepler, K2 and TESS) and their large ground-based monitoring programs. Those instruments have provided an unprecedented wealth of information which allows us to study statistical properties and search for hidden relationships between pulsation and/or physical observables.

Obtaining a large database with well-defined parameters can help to the interpretation of the data. Based on such a DB, This work focuses on the automatic classification of stars depending on their type based on their relationship with physical parameters.

Previous works have already related morphology with stellar parameters (e.g. metallicity, Teff, luminosity or log g) classifying the observed frequency spectra according to their position in the HR diagram. The novelty of this work lies in the automation of the process and the search for groups with similar parameters around time and space.

Subsequently, once the morphology information is obtained, unsupervised classification machine learning techniques are applied. Finally, we will see the common characteristics of the groups that we find.

This approach could be particularly useful for stars whose pulsation content is difficult to interpret. This is the case for classical intermediate-mass pulsating stars (γ-Dor, δ-Scuti, hybrids) for which current theories do not adequately predict the observed oscillation spectra.

Here we use the light curves of stars that have been already studied, taking advantage of the most recent precise stellar characterizations carried out with Asteroseismology. Thus, we obtain a complete set of empirical relationships between morphological characteristics of the stellar light curves and the estimated values of temperature, metallicity, luminosity and surface gravities.

INPUT DATA

Objetive of clasification: Types of Stars

Astronomers observe the stars in order to determine their mass, age, chemical composition, luminosity and other properties. Obtaining a large database with well-defined parameters can help to the interpretation of the data. Based on such a DB, This work focuses on the automatic classification of stars depending on their type based on their relationship with physical parameters.

Target with 6 values: A, B, F, G, K, M
Numbers of star in DB: 934
8 features:

Mass	Ratio	Effective Temp.	Metallicity
luminosity	Surface gravities	Rotation	Cutoff

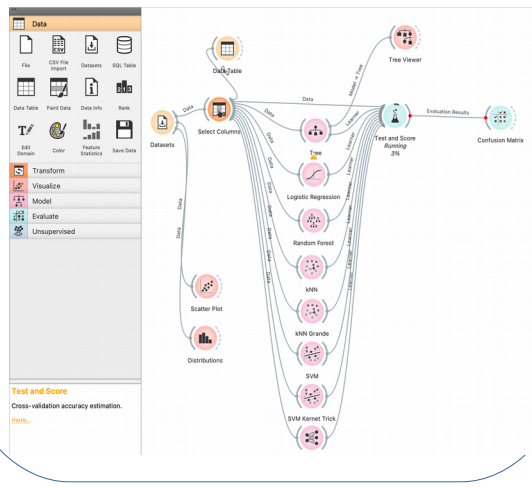
TOOLS

Orange is a machine learning and data mining suite for data analysis through Python scripting. To explore data with Orange, one requires no programming or in-depth mathematical knowledge. We believe that workflow-based data science tools democratize data science by hiding complex underlying mechanics and exposing intuitive concepts. Anyone who owns data, or is motivated to peek into data, should have the means to do so.

Free as in Freedom

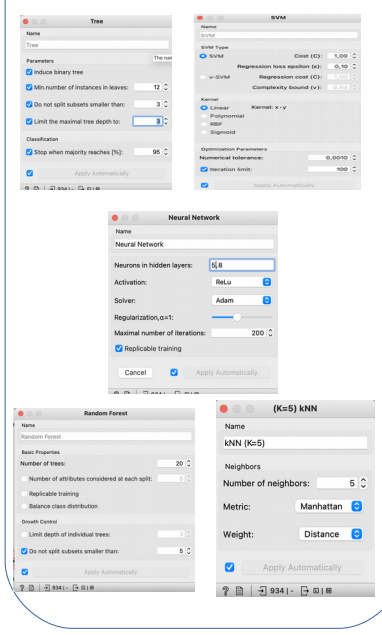
TECHNIQUES

- Logistic Regression.
- Classification Tree
- KNN (k=100 and k5)
- Ramdom Forest
- SVM with/without Kernel Trick
- Neural Network



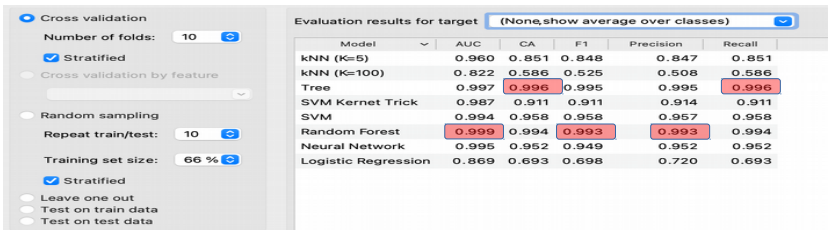
OPTIMIZATIONS

Looking for simpler structures with evaluation coefficients of the best models



CONCLUSIONS

Looking at all estimators, the two best methods of classifying this set of stars are "Classification Tree" and "Random Forest".



Model	AUC	CA	F1	Precision	Recall
kNN (K=5)	0.960	0.851	0.848	0.847	0.851
kNN (K=100)	0.822	0.586	0.525	0.508	0.586
Tree	0.997	0.996	0.995	0.995	0.996
SVM Kernet Trick	0.987	0.911	0.911	0.914	0.911
SVM	0.994	0.958	0.958	0.957	0.958
Random Forest	0.999	0.994	0.993	0.993	0.994
Neural Network	0.995	0.952	0.949	0.952	0.952
Logistic Regression	0.869	0.693	0.698	0.720	0.693

Some basic terms are Precision, Recall, AUC (Partial Area under the curve) and F1-Score. These relate to getting a finer-grained idea of how well a classifier is doing, as opposed to just looking at overall accurac

COLLABORATE WITH US

José Ramón Rodón: rodon@iaa.es Javier Pascual: javier@iaa.es Andrés Moya: andres.moya-bedon.uv.es