

GUASOM: Gaia Utility for Analysis and Knowledge Discovery based on Self-Organizing Maps

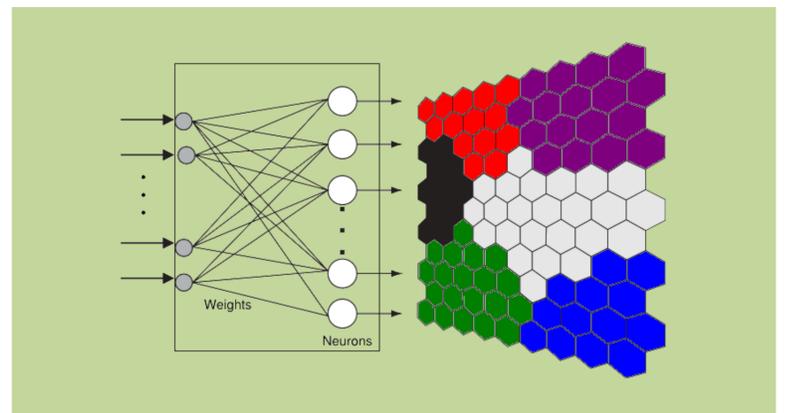
D. Fustes, M. Manteiga, C. Dafonte, B. Arcay, M.A. Álvarez, D. Garabato

Computer Science Faculty, University of A Coruña (Spain)

We present a method for knowledge analysis in large astronomical spectrophotometric archives. The method is based on a type of unsupervised learning Artificial Neural Networks named Self-organizing maps (SOMs). SOMs are used to organize the information in clusters of objects, as homogeneously as possible according to their spectral energy distributions (BP/RP spectra), and to project them onto a 2D grid where the data structure can be visualized. Our algorithm has been tested by means of simulated Gaia spectrophotometry, which is based on SDSS observations and theoretical spectral libraries covering a wide sample of astronomical objects. We demonstrate the usefulness of the method by analyzing the spectra that were rejected by the SDSS spectroscopic classification pipeline and thus classified as "UNKNOWN". REFERENCE: A&A 559, A7 (2013)

Artificial Neural Networks (ANNs) are computational models inspired by the human brain, raised in order to achieve a better understanding about its natural behavior and develop similar systems. Furthermore ANNs acquire knowledge through a learning process so as humans do.

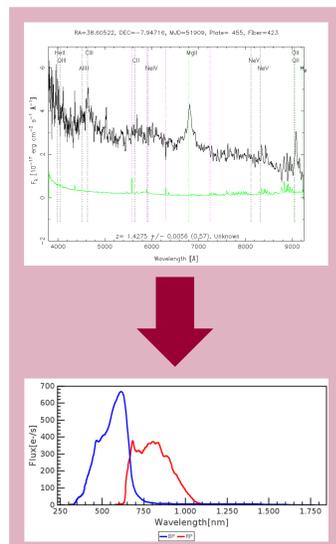
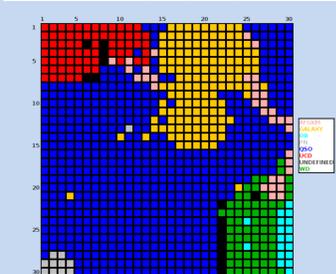
One of the main features of Self Organizing Maps (SOMs) is that they preserve the topological properties of the input space, making them useful for visualizing low-dimensional views of high-dimensional data, classifying them into different clusters according to their similarity.



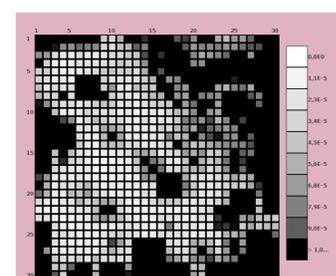
We show the distribution of astronomical object classes obtained with Gaia photometric simulations (150,000 objects) over the computed SOM. The color assigned to each of the clusters was set as a function of the predominant class in the objects belonging to it. The black color indicates that the cluster is empty. The confusion matrix allows to evaluate the clustering performance. In this case, the achieved compression rate is 167:1, with a mean class purity around 98.5% in the clusters.

	AFGKM	GALAXY	OB	PN	QSO	UCD	WD
AFGKM	95,28	1,3	0	0	0	2,7	0,72
GALAXY	0,67	98,4	0	0	0,94	0	0
OB	0	0	99,76	0	0	0	0,24
PN	0	0	0	99,33	0	0	0,67
QSO	0,44	0,83	0	0	98,7	0,10	0,10
UCD	0,51	0	0	0	0	99,49	0
WD	1,46	0	0	0	0	0	98,54
COUNT	5000	33 670	9999	748	70 554	9890	20 556

Class	Description
AFGKM	Main sequence stars, from PHOENIX model
OB	Very hot OB stars
WD	White dwarfs, both DA and DB
UCD	Ultra cool dwarfs stars
GALAXY	Semi-empirical galaxies
QSO	Semi-empirical quasars
PN	Semi-empirical planetary nebulae



We compiled a new library formed by spectra from SDSS that were classified as "UNKNOWN" by the SDSS spectroscopic classification pipeline. A dataset composed of 10,125 objects, mostly faint objects (mean G magnitude 19) and unsuccessful observations that we transformed to Gaia BP and RP format by the use of Gaia simulator GOG. Several visualization tools are available with which to unveil the data's physical nature and distribution. Photometric color distribution GRP - GBP can be used as a guide for effective temperatures. The U-Matrix displays the distances between clusters in a gray levels' scale, allowing to identify outlying clusters. A distance-based model (KNN) allows to retrieve the closer templates for each cluster prototype. External archives can contribute to the identification process, we opted for the SIMBAD catalog to perform crossmatching with the SDSS outliers, looking for more identifications. In this case, we retrieved those objects in SIMBAD within a radius of one arcsec from every SDSS outlier, obtaining its SIMBAD type in case it exists. Our method allows the potential discovery of hundreds of new objects, such as white dwarfs and quasars.



SOM Visualization Tool

SOM map selection

StarColor: Show plot

Parallax: Show plot

All the maps are 16x16 dimensions each one shows all the contained neurons

Information

Neuron (10, 13)

GMag: 14.9622
Parallax: 0.0002
PAbsolDelta: 0.0000
PAbsolDelta2: 0.0000
SlyRadious: 19.7326

Spectral Show plot

neuron graphic explanation

Available options

Select variable: Show

Drop outliers (N): Normal

0 20 40 60 80 100

Print plot

