A unified approach to photometric redshift evaluation for galaxies and QSOs

O. Laurino[1], R. D'Abrusco[2], G. Riccio[1], M. Brescia[3], S.G. Djorgovski[4], G. Longo[1]

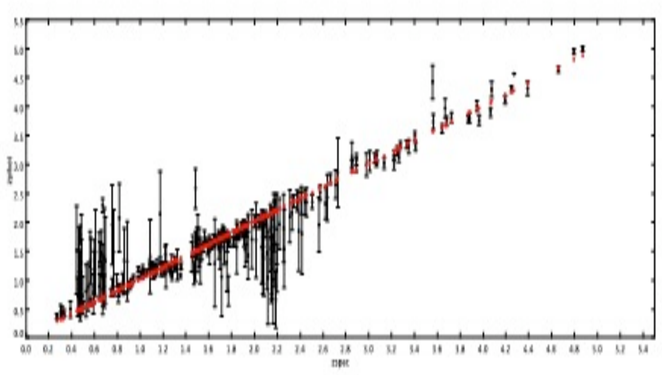
1-Dept. Physical Sciences, Univ. Napoli 2-Dept. Astronomy, Univ. Padova 3-INAF OACN 4-Caltech

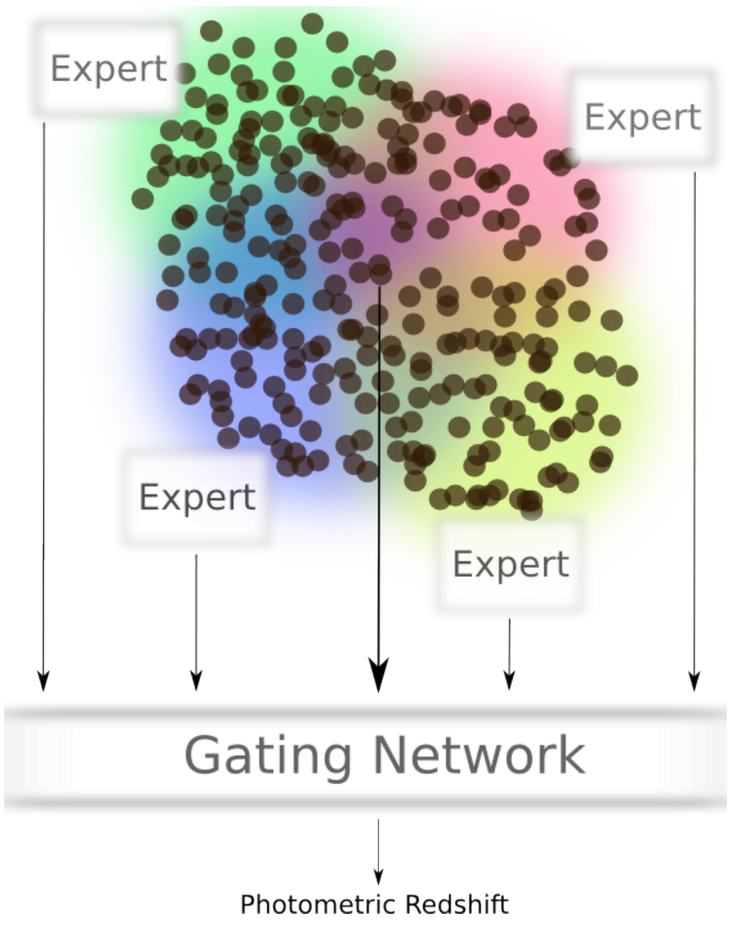
Abstract

In the era of massive astronomical datasets, the reconstruction of the three dimensional distribution of extragalactic sources in the redshift space is of crucial importance for constraining some of the main issues regarding the formation and evolution of such sources. A general method - Weak Gated Expert (WGE) - for the accurate determination of the photometric redshifts of galaxies and QSOs, based on multiwavelength photometry and a combination of data mining techniques, is discussed in this poster. The method employs specific tools developed under the EuroVO and NVO frameworks for data gathering, pre-processing and mining while relying on the scaling capabilities of the computing grid. It is being applied to the SDSS-DR7 galaxies catalogue and to the catalogue of candidate QSOs produced by D'Abrusco et al. 2009, and it is available to the astronomical community through the DAME infrastructure.

Outline

- The minimum N_{min} and maximum N_{max} numbers of clusters to be produced during clustering in the parameter space are fixed.
- A clustering algorithm is applied $N = N_{max}-N_{min}$ times in the photometric parameter space. Each partition contains n clusters, with n ranging from N_{min} to N_{max}. A fuzzy k-means algorithm is used so that each source has a non-zero membership probability to belong to each cluster. A threshold T is set so that each cluster will contains only objects with membership m > T.
- A distinct neural network (MLP) is trained on each cluster, so that each network becomes "expert" at associating colours with redshifts in a certain region of the parameter space.
- As clusters share a certain fraction of sources, the experts are correlated. A new MLP, called "gating network", is trained on both the outputs from the single experts and the set of original features in order to give the best estimate of the photometric redshift based on the different "opinions" of the experts. The overall error for a specific clustering pattern is determined using a validation sample.
- The smallest validation error indicates the optimal number of clusters providing the most accurate results. A test set is used to determine the generalization capability of the algorithm and the overall error.





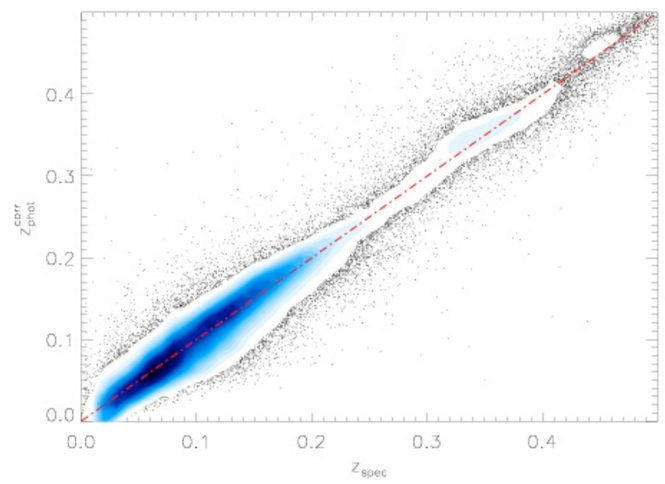
The uncertainties

Particular attention has been devoted to the characterization and evaluation of the possible sources of uncertainty in the estimation of the photometric redshifts:

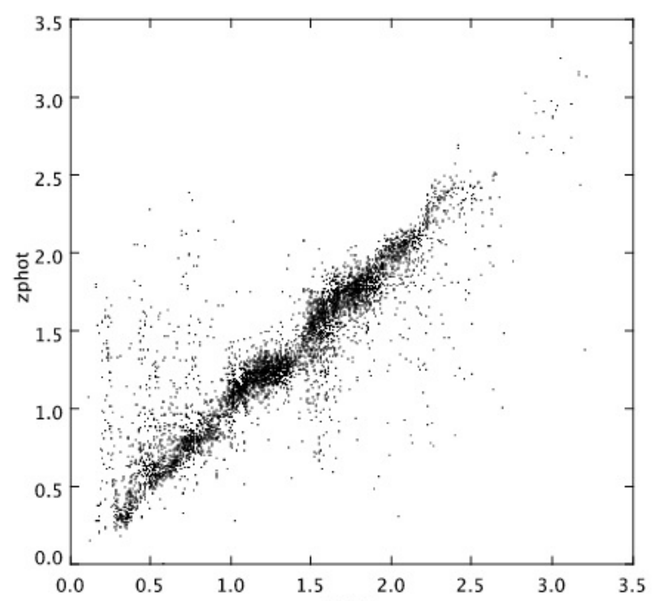
- Input noise: error propagation on the input parameter (Ball et al. 2008)
- Model variance: different models make differing predictions (Collister & Lahav 2004)
- Model bias: different models may be affected by different biases.
- Target noise: in some regions of the parameter space, data may represent poorly the relation between featured and targets (this paper).

Results

The WGE algorithm has been applied to optical SDSS data to evaluate redshifts of galaxies in a 4 dimensional space using a BoK covering the redshift interval [0.02, 0.5]. The accuracy of the z_{phot} reconstruction is σ_{rob}= 0.017.



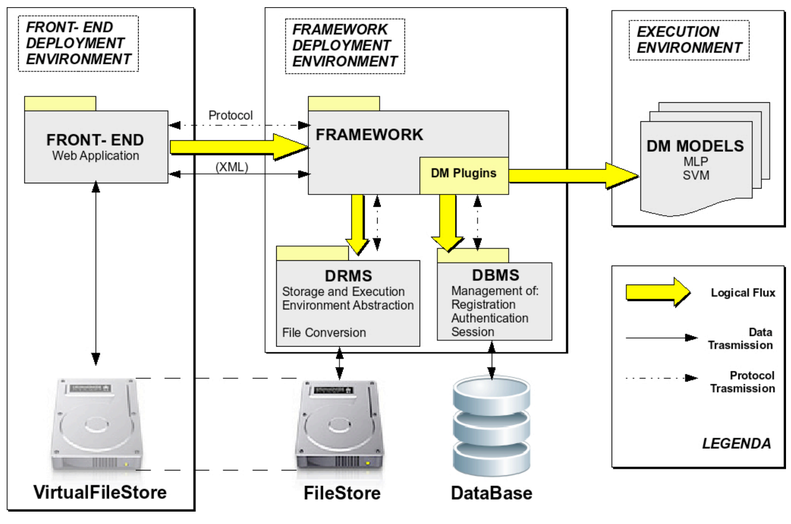
The algorithm has also been used to evaluate the photometric redshifts of candidate QSOs extracted by D'Abrusco et al. 2009 using optical and optical + ultraviolet photometric information, the BoK consisting of all spectroscopic confirmed QSOs observed in SDSS-DR7 (σ_{rob}= 0.057).



The DAME framework

DAME is an integrated, web oriented, platform independent, extensible framework for data mining:

- Integrated: the user can store his data and save his experiments and sessions. The developer can download a dedicated set of APIs which allow him to develope and contribute his own models.
- Web oriented: the framework is based on a service oriented architecture with a dedicated client web application.
- Platform independent: different "drivers" will be available, e.g. GRID, stand-alone,...
- Extensible: data mining models, complex tasks, middleware drivers and even the client application can all be extended and augmented by means of a consistent, object oriented plug-in architecture.



Contact: **laurino@na.infn.it**