Advanced Topics in Computational Intelligence
(MAI-ATCI)
June 2018
Practical Project 3: FIR and GFS

This is the third practical project of the ATCI course that should be developed in groups of **three** students and should be delivered by **June 22nd**

Introduction

In this project it is proposed to compare different hybrid fuzzy logic techniques to predict human white wine test preferences based on physicochemical properties from wine analyses. The white wine data used for this study comes from the Portugal north-west region, named Minho, and it is available from the UCI machine learning repository and the Keel repository. It has been proposed for both, regression and classification, by Cortez et al. (2009).

In this work, you should analyze both problems: classification and regression. For each kind of problem, you should work, at least, with two methodologies. As a regression problem, it is mandatory to work with the FIR methodology. As a classification problem, it is mandatory to work with the Slave algorithm that KEEL software provides. You can chose, for both problems, other hybrid fuzzy algorithms, the ones that you prefer. For example, other Iterative Rule Learning approaches like MOGUL or Takagi-Sugeno-Kang (TSK-IRL) or Fuzzy Rule Learning approaches, as Wang and Mendel (WM-R), that we have seen in the CI course last semester (although it is not hybrid), all of them available in the KEEL software. You also can use Neuro-Fuzzy Systems as ANFIS that is available in Matlab and that we have also seen in the CI course.

Data

You can find the description of the data for the **classification** problem in http://sci2s.ugr.es/keel/dataset.php?cod=31 and for the **regression** problem in https://archive.ics.uci.edu/ml/datasets/Wine+Quality

The data sets are composed of the more common wine physicochemical tests and correspond to a set of 11 input variables for the regression problem and 13 variables for the classification problem. You can find the description of each input variable as well as the corresponding output variable interpretation in the links mentioned above.

The data has been already prepared as both Visual-FIR and KEEL software require. You will use a 5 fold cross-validation for the regression problem (as proposed in Cortez et al., 2009) and also a 5 fold cross-validation for the classification problem.

For the regression problem, you can compare your hybrid fuzzy techniques results with the ones shown in Cortez paper, i.e. multiple regression, neural networks and support vector machine techniques.

The data sets for FIR are provided in the FIRdata folder. The data sets for KEEL software are already included in the KEEL platform, so it will appear in the list of possible data sets to be studied when you open the KEEL software and you start:

1) a new classification experiment. Its name is *Wine* (in the KEEL datasets)
2) a new regression experiment. Its name is *WineWhite*.

Remember that when you prepare your experiment in the KEEL platform, you should select first the type of partition, then the dataset and afterwards the algorithm/s and a postprocessing method to have more information of the results achieved. For example, the Vis-Clas-Check gives you a global classification/regression error in training and test sets. The Vis-Clas-General gives you the p-value matrix and the results for each fold, etc. If you like, you can also include statistical tests in your experiment. You should look to the KEEL user's manual to see all the possibilities available.

Take into account that there are algorithms, such as MOGUL, that are much slower than other GFS, such as SLAVE. Therefore, I recommend you to perform separate experiments for those algorithms that are slower.

Expectations

It is expected that you analyse both problems to be modelled and study the different hybrid fuzzy techniques that you have chosen to perform these tasks. To this end, you should compare different configuration parameters of the selected methodologies in order to arrive to some conclusions. You must design the experiments that you consider appropriate to draw your own conclusions of this study.

You should for example:

- Compare different discretization parameters in FIR methodology
- Compare different masks an get conclusions out of it about the feature selection performed by FIR
- Compare different values of the parameters of the GFS approaches studied (number of labels, population size, number of iterations, ...)
- etc.

It is not expected that you perform a large analysis of different parameter values for those algorithms that are really very time consuming (i.e. MOGUL).

It is expected that you compare the wine quality prediction results obtained by using at least an error measure (the one that you prefer, for example the mean absolute deviation MAD) and, it would be nice if you compute the regression error characteristic (REC) curve. You can find a description of both in (Cortez et al., 2009).

It would be also interesting to compare the computational time needed for each methodology and the number of rules obtained in the different approaches.

It is expected that you report the experiments that you have performed, i.e. goals, decisions, results, comparisons and conclusions, in a document being it clear and reasoned.

Project delivery

The deadline for delivering your project report is June 22nd. You should send it to angela@cs.upc.edu.

The final report should include:

1. Full name of the three students that performed the practical work.
2. A description of the work and its goals.
3. The description of the different experiments performed for each of the problems (classification and regression), reasoning the choice.
4. The results obtained with each method, along with the best set of parameters for each of them.
5. Comparison of the results obtained for each method and their significance.
6. Discussion and conclusions of the results emphasising the strengths and weaknesses of the methodologies involved.

Contents of the zip file:

- **FIRdata** directory. This directory contains the WineWhite data of the regression problem, already prepared in 5-fold cross validation, in the format needed for FIR methodology.
- The paper (Cortez et al. 2009) referenced in this text.

References

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., Reis, J., 2009. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4), 547-553.