RelEns 1.0

I. Introduction to RelEns interface

1. Tab Relational classification

- file/directory path (textfield and browse button)

is where you enter the path to your dataset file or directory containing your saved datasets (in case you want to work with saved datasets). You can enter the path directly by typing or you can browse it with the Browse button beside.

- checkbox: **Open and work with existing datasets**

By leaving the checkbox unchecked, you want to work with your original dataset. That means you can specify other parameters, such as the threshold (in case of multilabel problem) or number of folds for cross-validation. Otherwise you work with saved datasets.

- **User configuration**

+ checkbox: **Classification contains multilabel problem**

To indicate whether your dataset is multilabel.

+ threshold

Only available when you work with multilabel dataset, you have to choose a numeric value for this threshold.

+ checkbox: **Run Ensemble classification afterwards**

To indicate whether you want to run Ensemble Classification after several runs of Relational Classification. That number of runs range from 2 to 5.

+ Combobox: **Number of relation types**

Only available when you check “Run Ensemble classification afterwards”, where you need to choose a value between 2 and 5. This value is the number of Relational Classification runs before Ensemble Classification.

+ Combobox: **Number of folds for cross-validation**

Choose the number of folds your original dataset will be splitted into.

+ Combobox: **number of neighbors**

Only available when you work with PRN2Hop classifier. The default value can be changed in **Preferences**.

+ Combobox: **Techniques**

Only availabel when you work with either WeightedAvg or IndRVS classifier.

+ Checkbox: **Save test and training sets**

Indicates whether you want to save your test and training sets into a directory.

+ Listbox: **Classifier**

A list of availabel classifiers.

+ Button: **Start**

Start Relation Classification.

+ Button: **Cancel**

Cancel running Relational Classification.

2. Tab Ensemble classification

- **Fushion method**

You need to choose fushion method for Ensemble Classification. Default selection is Voting.

- **Other information**:

Configuration you are using in Relational Classification runs.

- Button: **Start**

Start Ensemble Classification.

- Button: **Cancel**

Cancel running Ensemble Classification.

There are 3 common buttons for both Relational and Ensemble Classification:

+ Button: **Clear**

Clear the text result.

+ Button: **Save**

Save text result to file.

+ Button: **Exit**

Exit application.

3. Menu File

- **Save**

Save the text result to file.

- **Preferences**

Where you choose default values for:

+ maximal number of iterations

+ number of neighbors to consider in WeightedNaiveBayes

+ Number of neighbors for PRN2Hop

+ Path to training and test sets

- **Exit**

Exit application.

4. **About**

Information about the application.

II. How-to questions

1. How to run Relational Classification?

A Relational Classification run starts with browsing your datasets and adjust your own configuration:

- Whether your datasets are multilabel, if yes, you have to choose a numeric value for your threshold.
- Whether you want to run Ensemble Classification afterwards, if yes, you have to choose how many Relational Classification run you want to carry out beforehand.
- You also need to specify the number of folds for cross-validation.
- For PRN2Hop classifier you can choose the number of neighbors as a parameter, or you might accept the default value.
- For WeightedAvg and IndRVS classifiers, you can choose the technique as a parameter, or you might accept the default value.
- Optionally you can save test and training datasets used by the classifier.

2. How to run Ensemble Classification?

You can only run Ensemble Classification when you choose "Run Ensemble Classification afterwards" on tab Relational Classification. After a certain number of Relational Classification run, you can choose fushion method (Voting or Stacking) before start classifying. (On tab Ensemble Classification you also see configuration information used in Relational Classification runs).

3. How to save test and training sets?

By checking "Save test and training sets" you will be prompted to save your datasets when running Relational Classification.

4. How to open and work with saved datasets?

After checking "**Open and work with existing datasets**", you can enter the dataset path in the textbox where you see "*Please select directory containing datasets*", after that, you will be prompted to browse the file path to your original dataset. Once the datasets are loaded successfully, you can start classifying.

A note on Number of relation types and number of Relational Classification (RC) runs:

+ If you run RC more than number of relation types, you will be informed about runs which is not counted. However you can change value of number of relation types to match your need.

Scenario 1: you selected number of relation types to be 5. You have run RC 3 times and want to run EC with 3 then you can change number of relation types from 5 to 3 and switch to tab EC.

Scenario 2: you selected number of relation types to be 2 and have run RC 2 times however you want to run Ensemble Classification (EC) with 3 runs of RC, then you can change number of relation types to 3 and run RC 1 more time before switch to tab EC.