KEEL: A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems

http://www.keel.es

Research Groups:

SCI²S

Metrology and Models

Intelligent Systems

AYRNA

GRSI

KEEL: A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems

1. INTRODUCTION
2. KEEL
3. EXPERIMENTAL EXAMPLE
4. CONCLUSIONS AND FURTHER WORK

Introduction

- Evolutionary Algorithms (EAs) requires a certain programming expertise along with considerable time and effort to write a computer program for implementing algorithms that often are sophisticated.

Introduction

Software	Language	Graphical Interface			Input / Output			Pre-processing Variety				Learning Variety				Run Types		Advanced Features			
		Graph representation	Data visualization	Data management	ARFF data format	Other data formats	Data Base connection	Discretization	Feature Selection	Instance Selection	Missing values imputation	Classification	Regression	Clustering	Association Rules	On-line run	Off-line run	Post-processing	Meta Learning	Statistical tests	EAs
ADaM	C++	N	N	I	Y	N	N	N	A	B	N	I	N	A	B	Y	N	N	N	N	B
D2K	Java	Y	A	I	Y	Y	Y	I	A	B	B	A	A	A	A	Y	N	N	N	N	I
KNIME	Java	Y	A	A	Y	Y	Y	I	A	B	B	A	A	A	A	Y	N	N	N	I	B
MiningMart	Java	Y	B	A	N	N	Y	I	A	B	I	B	B	N	N	Y	N	N	N	N	B
Orange	C++	Y	A	A	N	Y	N	A	I	B	B	I	N	I	I	N	Y	N	N	N	N
Tanagra	C++	N	A	A	Y	Y	N	B	A	B	N	A	I	A	A	Y	N	N	I	A	N
Weka	Java	Y	A	A	Y	Y	Y	I	A	B	B	A	A	A	A	Y	N	N	I	N	B
RapidMiner	Java	N	A	A	Y	Y	Y	I	A	B	B	A	A	A	A	Y	N	N	A	B	I

- In the last few years, many software tools have been developed to reduce this task.
- We develop a non-commercial Java software tool named KEEL (Knowledge Extraction based on Evolutionary Learning).

Introduction

- This tool can offer several advantages:
  - It includes a big library with EAs algorithms based on different paradigms (Pittsburgh, Michigan, IRL and GCCL) and simplifies their integration with different pre-processing techniques.
  - It extends the range of possible users applying EAs.
  - This can be used on any machine with Java.

KEEL: A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems

1. INTRODUCTION
2. KEEL
3. EXPERIMENTAL EXAMPLE
4. CONCLUSIONS AND FURTHER WORK

KEEL : Functionality

- KEEL is a software tool to assess EAs for DM problems including regression, classification, clustering, pattern mining and so on.
- KEEL allows us to perform a complete analysis of any learning model in comparison to existing ones, including a statistical test module for comparison.
- Moreover, KEEL has been designed with a double goal: research and educational.

http://www.keel.es

KEEL : Main features

- EAs are presented in predicting models, pre-processing and postprocessing.
- It includes data pre-processing algorithms proposed in specialized literature: data transformation, discretization, instance selection and feature selection.
- It contains a statistical library for analyzing results
- Some algorithms have been developed by using Java Class Library for Evolutionary Computation (JCLEC).
- It provides a user-friendly graphical interface in which experimentations containing multiple data sets and algorithms connected among themselves can be easily performed.
- KEEL also allows creating experiments in on-line mode, aiming an educational support in order to learn the operation of the algorithm included.

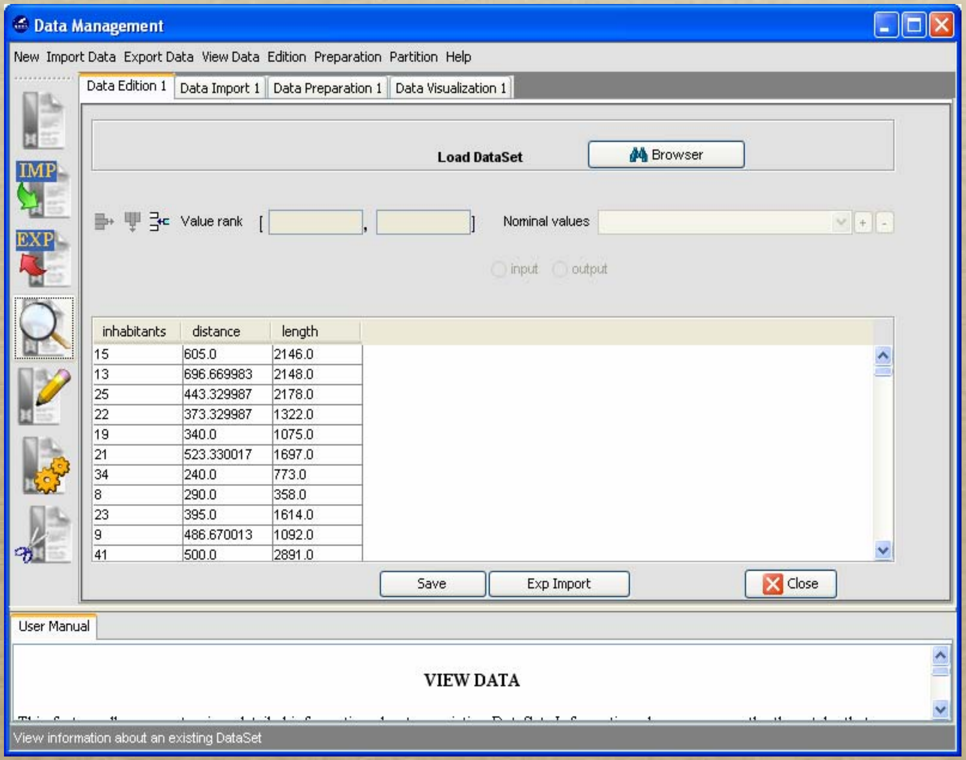
KEEL : Blocks

It is integrated by three main blocks:

- Data Management.
- Design of Experiments (off-line module).
- Educational Experiments (on-line module).

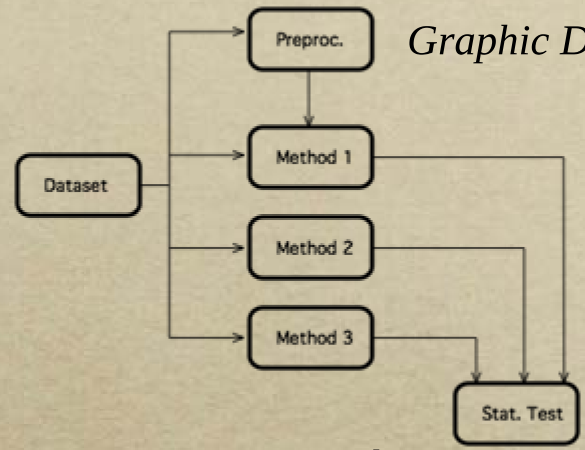
KEEL : Data Management

- This part is made up of a set of tools that can be used
  - to build new data
  - to export and import data in other formats
  - data edition and visualization
  - to apply transformations and partitioning to data.
  - etc.



KEEL : Design of experiments

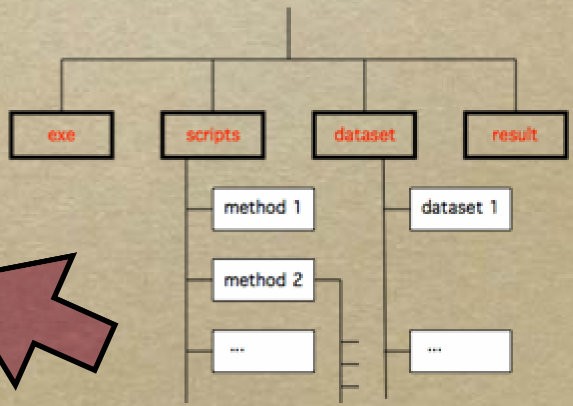
- It is a Graphical User Interface that allows the design of experiments for solving different machine learning problems.



Graphic Design

Execute in a remote machine

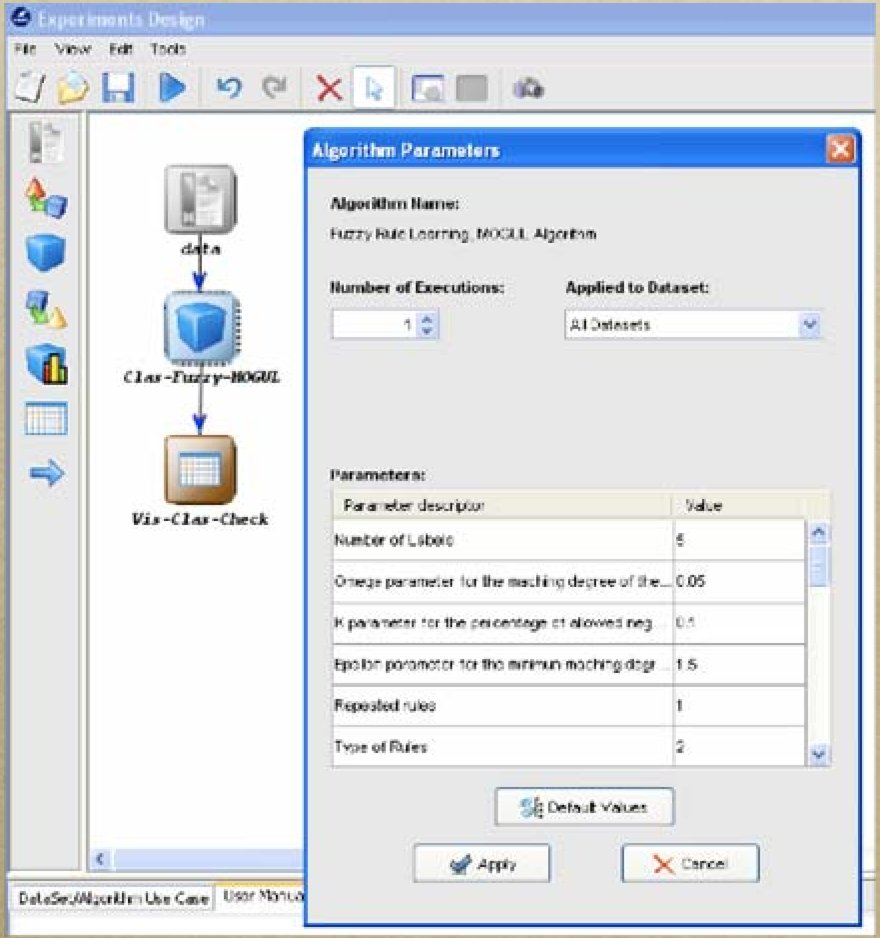
- Once the experiment is designed, it generates the directory structure and files required for running them in any local machine with Java.



Directory Structure and xml-based scripts

KEEL : Design of experiments

- The experiments are graphically modeled. They represent a multiple connection among data, algorithms and analysis/visualization modules.
- Aspects such as type of learning, validation, number of runs and algorithm's parameters can be easily configured.
- Once the experiment is created, KEEL generates a script-based program which can be run in any machine with JAVA Virtual Machine installed in it.

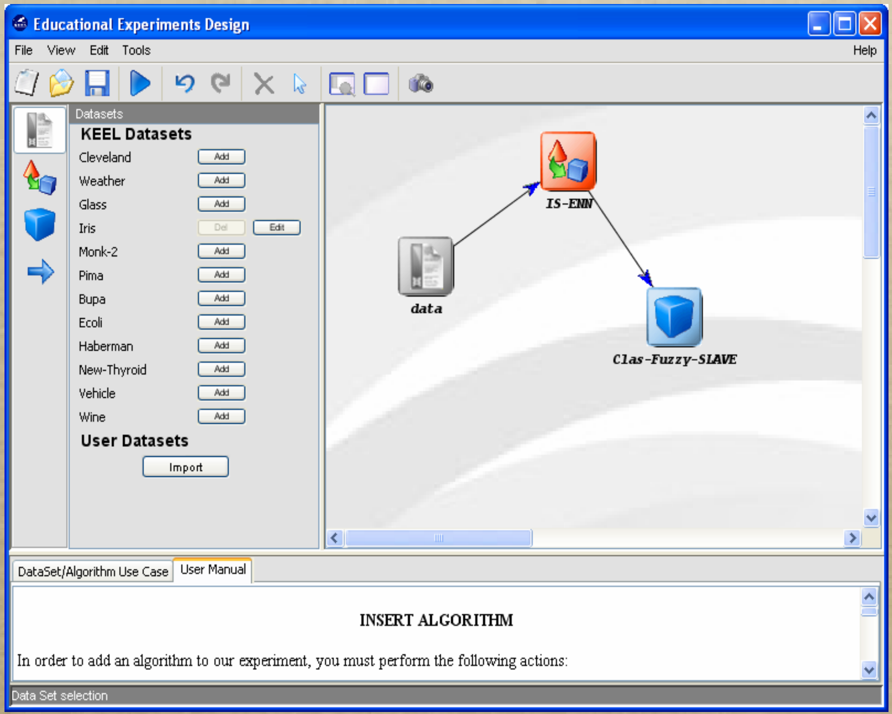


KEEL : Educational Module

- Similar structure to the design of experiments

- This allows for the design of experiments that can be run step-by-step in order to display the learning process of a certain model by using the software tool for educational purposes.

- Results and analysis are shown in on-line mode.

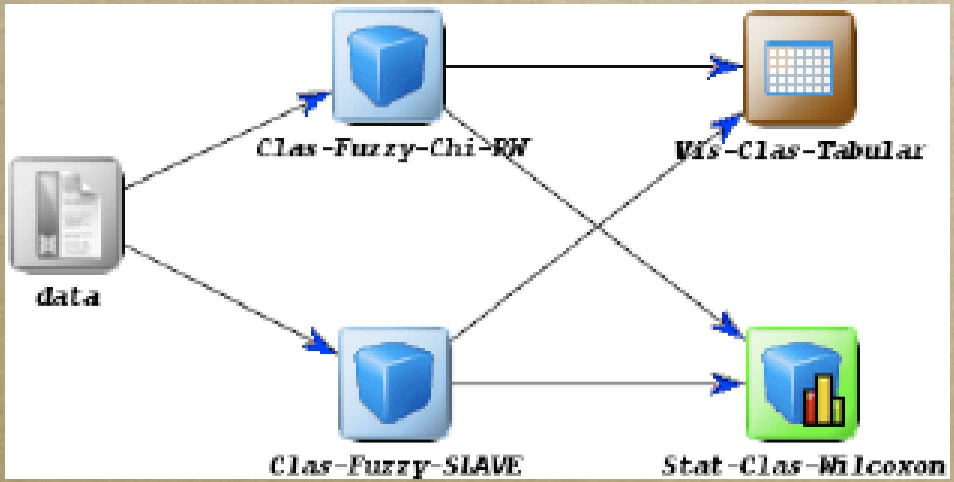


KEEL: A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems

1. INTRODUCTION
2. KEEL
3. EXPERIMENTAL EXAMPLE
4. CONCLUSIONS AND FURTHER WORK

Experimental example

- Type of learning: Classification
- Methods considered: SLAVE algorithm (Clas-Fuzzy-Slave) and Chi et al. algorithm with rule weights (Clas-Fuzzy-Chi-RW).
- Type of validation: 10-folder cross-validation model. SLAVE has been run 5 times per data partition (a total of 50 runs).
- Statistical Analysis: Wilcoxon test (Stat-Clas-Wilcoxon)



Experimental example

- 12 problems for classification:

Data set	#Examples	#Atts.	#Classes
Bupa	345	6	2
Cleveland	297	13	5
Ecoli	336	7	8
Glass	214	9	7
Haberman	306	3	2
Iris	150	4	3
Monk-2	432	6	2
New-thyroid	215	5	3
Pima	768	8	2
Vehicle	846	18	4
Wine	178	13	3
Wisconsin	683	9	2

Experimental example

Average Results:
(Vis-Clas-Tabular)

Dataset	CHI-RW			SLAVE		
	Acc_{Tr}	Acc_{Tst}	#Rules	Acc_{Tr}	Acc_{Tst}	#Rules
Bupa	59.87	57.87	43.3	60.60	58.28	3.9
Cleveland	91.25	39.09	230.5	79.05	53.52	39.6
Ecoli	79.53	78.33	43.5	82.75	79.10	11.7
Glass	65.99	60.04	27.1	71.70	62.16	12.8
Haberman	74.26	73.19	16.7	74.90	74.35	2.9
Iris	93.78	94.00	14.7	96.98	94.86	3.3
Monk-2	100.0	48.84	301.8	67.36	67.23	1.3
New-thyroid	85.94	84.24	18.4	89.82	87.99	3.9
Pima	75.62	72.40	105.2	75.45	74.44	4.7
Vehicle	65.92	60.77	227.8	66.31	60.18	20.8
Wine	98.75	92.68	121.1	94.60	90.42	4.3
Wisconsin	98.08	91.21	224	97.16	95.72	5.1
Average	**82.42**	71.06	114.51	79.72	**74.85**	**9.53**

Statistical Results:
(Stat-Clas-Wilcoxon)

		N	Mean Rank	Sum of Ranks
SLAVE vs. Chi-RW	Positive Ranks	10	6.9	69.0
	Negative Ranks	2	4.5	9.0
	Ties	0		
	Total	12		

Comparison	R^+	R^-	p-value
SLAVE vs. Chi-RW	69.0	9.0	0.019607

KEEL: A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems

1. INTRODUCTION
2. KEEL
3. EXPERIMENTAL EXAMPLE
4. CONCLUSIONS AND FURTHER WORK

Concluding Remarks

- ⇨ KEEL relieves researchers of much technical work and allows them to focus on the analysis of their new models in comparison with the existing ones

- ⇨ The tool enables researchers with a basic knowledge of evolutionary computation to apply EAs to their work.

Future work

⇨ A new set of EAs and a test tool that will allow us to apply parametric and non-parametric tests on any set of data

⇨ Data visualization tools for the on-line and offline modules.

⇨ A data set repository that includes the data set partitions and algorithm results on these data sets, the KEEL-dataset

KEEL: A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems

http://www.keel.es

Research Groups:

SCI²S · **Metrology and Models** · **Intelligent Systems**

AYRNA · **GRSI**