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

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

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

<table>
<thead>
<tr>
<th>Data set</th>
<th>#Examples</th>
<th>#Atts.</th>
<th>#Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bupa</td>
<td>345</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Cleveland</td>
<td>297</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>Ecoli</td>
<td>336</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Glass</td>
<td>214</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Haberman</td>
<td>306</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Iris</td>
<td>150</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Monk-2</td>
<td>432</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>New-thyroid</td>
<td>215</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Pima</td>
<td>768</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Vehicle</td>
<td>846</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Wine</td>
<td>178</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>683</td>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>
Experimental example

Average Results: (Vis-Clas-Tabular)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>CHI-RW</th>
<th>SLAVE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\overline{Acc}_{TR}$</td>
<td>$\overline{Acc}_{TS}$</td>
</tr>
<tr>
<td>Bupa</td>
<td>59.87</td>
<td>57.87</td>
</tr>
<tr>
<td>Cleveland</td>
<td>91.25</td>
<td>39.09</td>
</tr>
<tr>
<td>Ecoli</td>
<td>79.53</td>
<td>78.33</td>
</tr>
<tr>
<td>Glass</td>
<td>65.99</td>
<td>60.04</td>
</tr>
<tr>
<td>Haberman</td>
<td>74.26</td>
<td>73.19</td>
</tr>
<tr>
<td>Iris</td>
<td>93.78</td>
<td>94.00</td>
</tr>
<tr>
<td>Monk-2</td>
<td>100.0</td>
<td>48.84</td>
</tr>
<tr>
<td>New-thyroid</td>
<td>85.94</td>
<td>84.24</td>
</tr>
<tr>
<td>Pima</td>
<td>75.62</td>
<td>72.40</td>
</tr>
<tr>
<td>Vehicle</td>
<td>65.92</td>
<td>60.77</td>
</tr>
<tr>
<td>Wine</td>
<td>98.75</td>
<td>92.68</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>98.08</td>
<td>91.21</td>
</tr>
<tr>
<td>Average</td>
<td>82.42</td>
<td>71.06</td>
</tr>
</tbody>
</table>

Statistical Results: (Stat-Clas-Wilcoxon)

<table>
<thead>
<tr>
<th>SLAVE vs. Chi-RW</th>
<th>Positive Ranks</th>
<th>Negative Ranks</th>
<th>Ties</th>
<th>Total</th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Ranks</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>12</td>
<td></td>
<td>6.9</td>
<td>69.0</td>
</tr>
<tr>
<td>Negative Ranks</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>4.5</td>
<td>9.0</td>
</tr>
</tbody>
</table>

Comparison

$$R^+ = 69.0$$  
$$R^- = 9.0$$  
$$p-value = 0.019607$$
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