(-: (-: SMILES :-) :-)

A Multi-purpose Learning System

Vicent Estruch, Cèsar Ferri,
José Hernández-Orallo, M.José Ramírez-Quintana
`{vestruch, cferri, jorallo, mramirez}@dsic.upv.es`
Dep. de Sistemes Informàtics i Computació,
Universitat Politècnica de València,
Valencia, Spain

8th European Conference on Logics in Artificial Intelligence
JELIA'02, System Presentation Session
Cosenza, Italy, September 23-26, 2002

Introduction

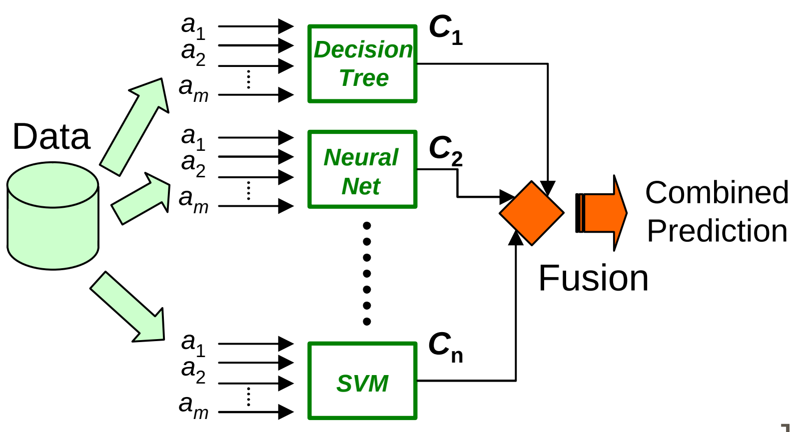
- **SMILES**:
  - integrates many different and innovative features in machine learning techniques.
  - extends classical decision tree learners in many ways:
    - new splitting criteria
    - non-greedy search
    - new partitions
    - extraction of several and different solutions
  - anytime handling of resources
  - sophisticated and quite effective handling of costs.

Motivation

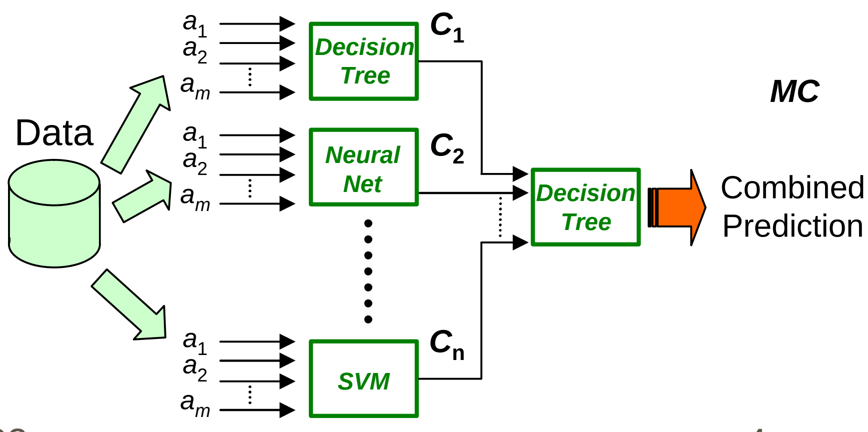
- Some hindrances for a wider applicability of Machine Learning:
  - Generation:
    - **Computational costs**: powerful methods in ML systems require huge amounts of memory and time to generate accurate hypotheses.
  - Application:
    - **Prediction error costs**: not all the errors have the same consequences: Cost matrices and ROC analysis necessary.
    - **Test costs**: not all the attributes can be tested economically. Especially in medical applications.
    - **Intelligibility**: the comprehensibility of the extracted models is critical for their validation, acceptance, diffusion and ultimate use.
    - **Throughput** (response time): complex models are difficult to be applied efficiently in real-time applications, such as fraud detection.

Ensemble Methods (1/2)

- Ensemble Methods (Multi-classifier or hybrid systems):
  - Aim at obtaining higher accuracy than single methods.
  - Generate multiple and possibly heterogeneous models and then combine them through *voting* or other fusion methods.
  - Good results related to the number and variety of classifiers.
  - Different topologies: simple, stacking, cascading, ...



Simple Combination



Stacking

Ensemble Methods (2/2)

- Main drawbacks of Ensemble Methods:
  - **Computational costs**: lots of memory and time are required to obtain and store the set of hypotheses (ensemble).
  - **Prediction error costs**: most ensemble methods are based on the maximisation of accuracy and not other cost-sensitive measures.
  - **Test costs**: the use of several (and diverse) hypotheses forces the evaluation of (almost) all the attributes.
  - **Intelligibility**: the combined model is a black box.
  - **Throughput**: the application of the combined model is slow.

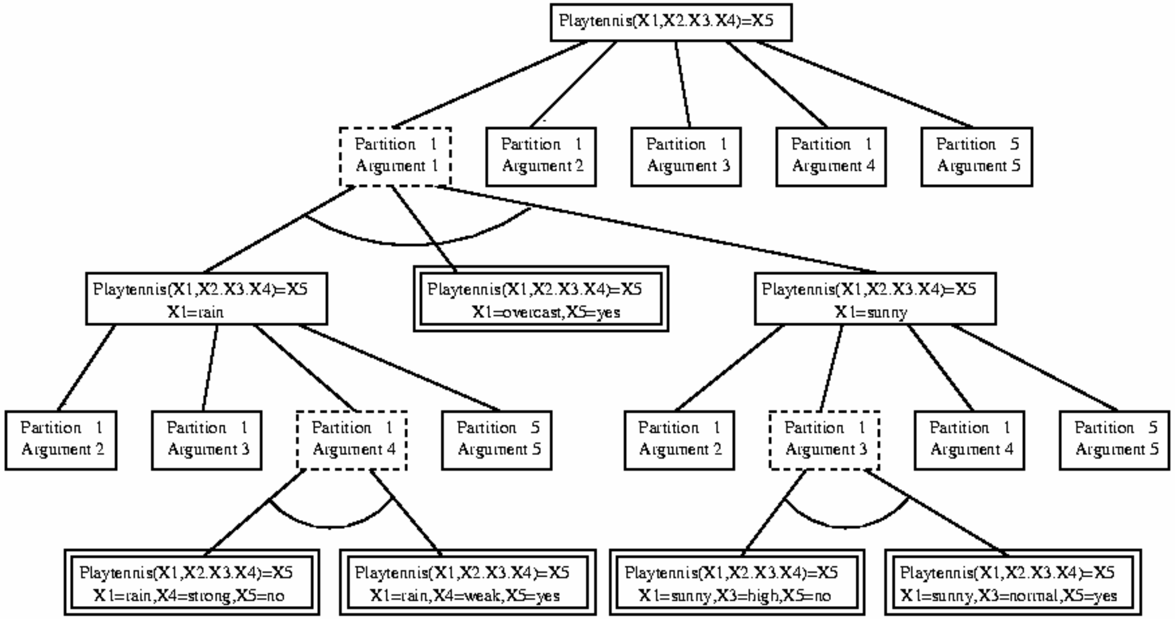
- The resolution of these drawbacks would boost the applicability of ensemble methods in machine learning applications.

Addressing Computational Costs

- Many ensemble solutions have common parts.
- Traditional ensemble methods repeat those parts: memory and time ↑↑↑

- **SMILES** is based on the construction of a *shared ensemble:*
  - **Common parts are shared in an AND/OR tree structure**.

DECISION MULTI-TREE



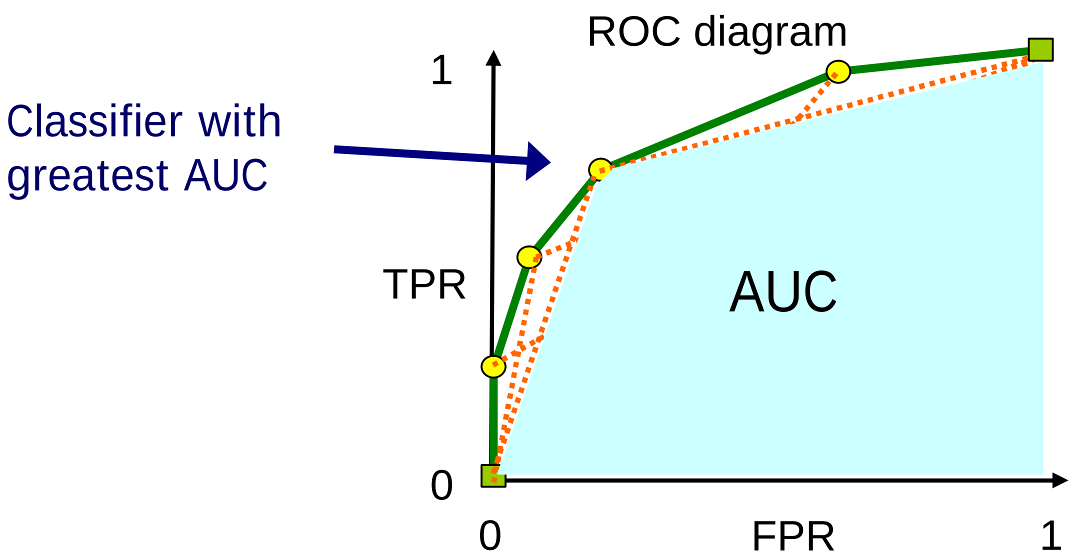
- Throughput is also improved by this technique.

Addressing Misclassification & Test Costs (1/2)

- Many ensemble methods aim at increasing accuracy.

AUC (Area Under the ROC Curve)

  - better measure when classification costs may be variable.
  - can be used as a metric for comparing classifiers:



- MAUC: Multi-class extension of the AUC measure (Hand & Till 2001).

Addressing Misclassification & Test Costs (2/2)

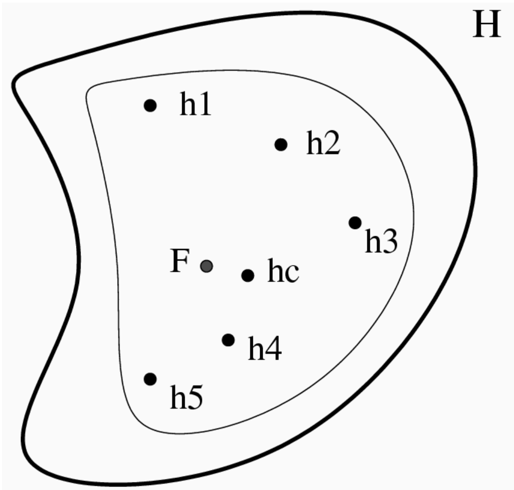
- **SMILES** has splitting criteria based on the maximisation of the AUC
  - MAUCsplit: Adaptation of Multi-class extension of AUC.
  - MSEsplit: Adaptation of Minimum Squared Error as splitting criterion.
- Splitting criteria can also be modified to minimise the test cost.

Addressing Test Cost and Intelligibility

- Ensemble methods (and many other ML methods) are:
  - Black boxes: no insight given by the model (ensembles, ANN, SVM...).
  - Attribute exhaustive: all or nearly all the attributes must be examined (ensembles, ANN, SVM, Bayes, ...).
  - Slow in real-time applications: all the classifiers must be evaluated.
- The Multi-tree structure (our shared ensemble) has also these problems.
- **SMILES** introduces the notion of “ARCHETYPE” of the ensemble.

Archetype

The archetype is the representative <u>single hypothesis</u> that is closer to the combined hypothesis.

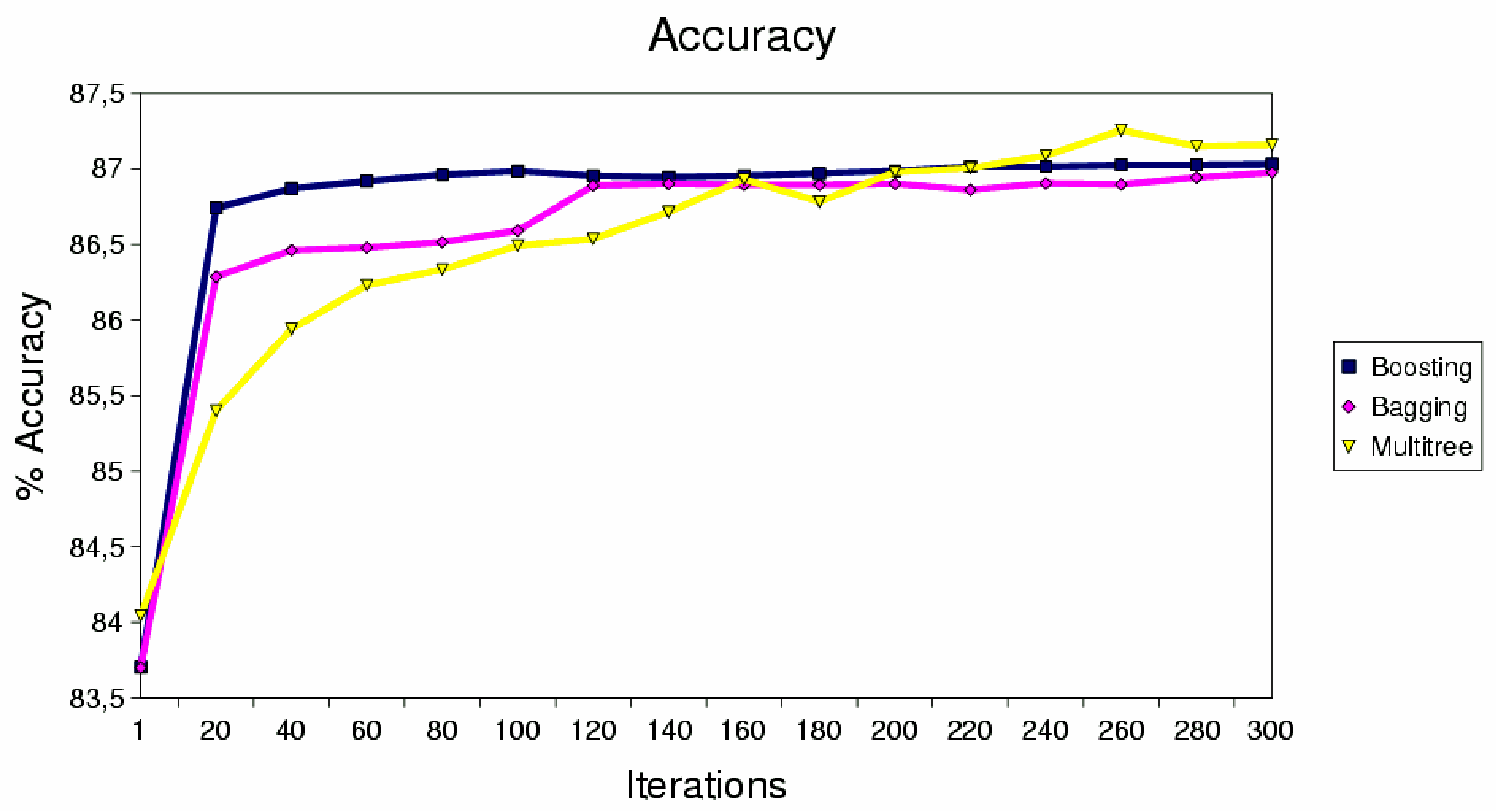


- H: hypothesis space
- h_i: hypotheses in the ensemble.
- F: combined hypothesis.
- h_c: archetype.

- **SMILES** extracts the archetype from the multi-tree structure without the need of a validation dataset.
- Comprehensibility, test cost and throughput problems solved.

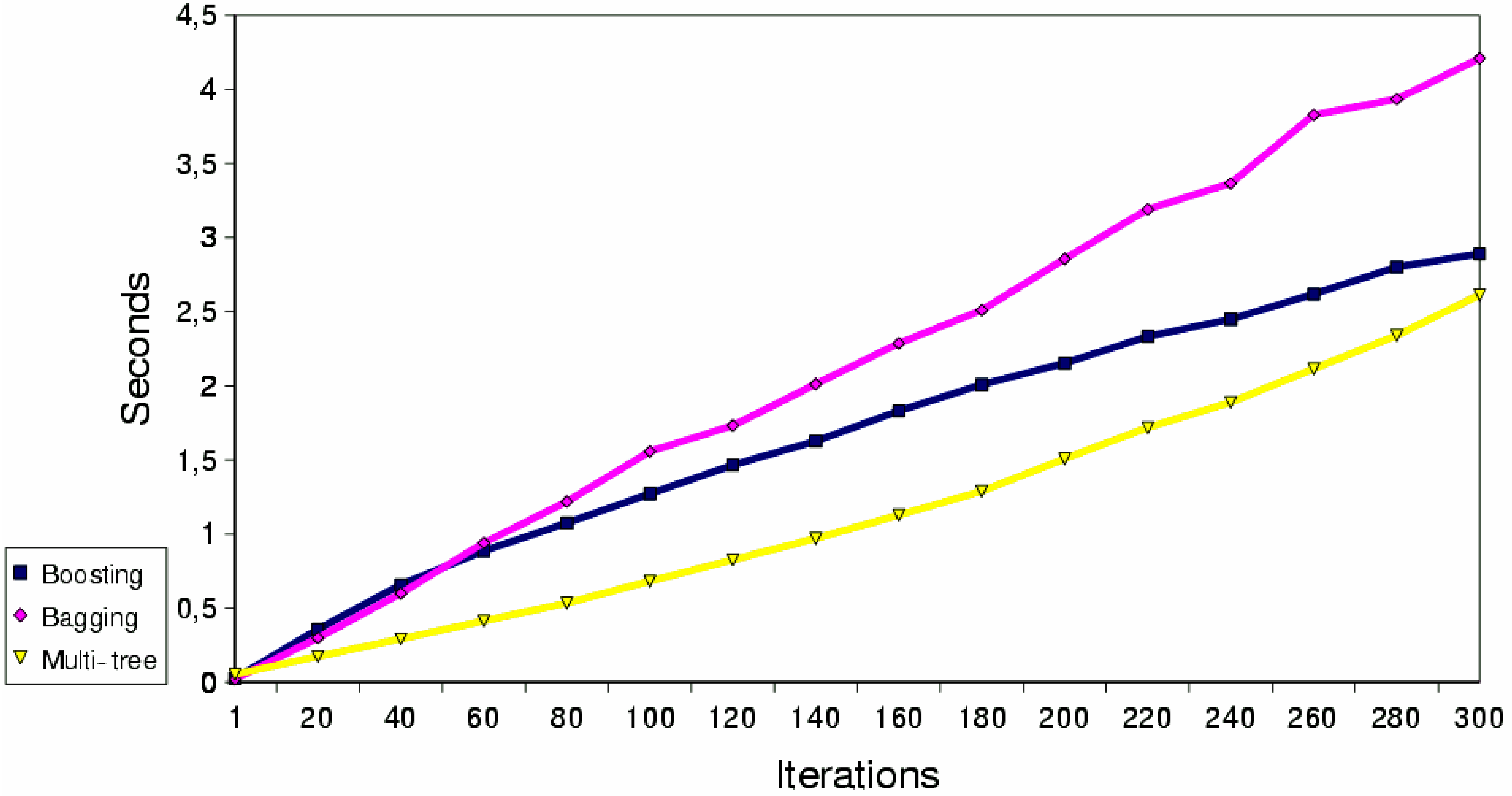
Some Experiments (1/4)

- Combination Accuracy compared to other Ensemble Methods:



Some Experiments (2/4)

- Combination Resources compared to other Ensemble Methods:



Some Experiments (3/4)

- Evaluation of splitting criteria wrt.:
  - accuracy
  - AUC
  - number of rules

GEOMEANS	GAINRATIO	MAUCSPLIT	MSESPLIT
Accuracy	**87.45**	87.19	87.05
M-AUC	87.42	**88.08**	87.98
Rules	23.27	**21.19**	22.99

25 Two-class datasets from UCI repository. Pruning enabled.

GEOMEANS	GAINRATIO	MAUCSPLIT	MSESPLIT
Accuracy	80.90	80.29	**83.12**
M-AUC	89.30	**90.18**	90.09
Rules	74.49	75.62	**68.26**

14 Multi-class datasets from UCI repository. Pruning enabled.

Some Experiments (4/4)

- Evaluation of the Archetype:

#	Dataset	Size	1	10				100				1000			
			1st	Comb	Arc	Occ	#Sol	Comb	Arc	Occ	#Sol	Comb	Arc	Occ	#Sol
1	monks1	566	92.3	96.1	96.0	96.5	107	100	100	100	8.7×10^{8}	100	100	100	1.6×10^{19}
2	monks2	601	74.8	74.9	74.3	74.3	148	77.4	76.1	72.5	2.6×10^{10}	82.3	82.1	70.4	3.2×10^{20}
3	monks3	554	97.5	97.7	97.7	97.6	46	97.5	97.6	97.5	80×10^{4}	97.7	97.7	97.6	7.1×10^{14}
4	tic-tac	958	78.2	79.0	78.1	78.3	257	82.7	78.2	78.6	2.7×10^{12}	84.6	79.8	79.5	3.1×10^{38}
5	house-votes	435	93.6	94.9	94.2	93.9	63	96.0	94.4	93.6	26×10^{5}	95.7	94.1	93.9	5.6×10^{11}
6	post-operative	87	60.9	63.8	61.8	60.0	55	66.3	63.8	62.3	59674	68.5	65.9	62.1	2.1×10^{9}
7	balance-scale	625	76.8	77.9	77.2	76.8	131	83.1	80.1	76.7	3.4×10^{8}	88.0	83.5	76.8	1.2×10^{18}
8	soybean-small	35	97.3	97.0	98.0	97.5	23	96.5	96.5	96.8	38737	95.0	93.3	96.3	1.8×10^{18}
9	dermatology	358	89.8	91.3	90.6	90.1	92	93.6	90.6	90.2	3.3×10^{7}	93.8	91.1	90.8	1.2×10^{10}
10	cars	1728	89.0	89.6	89.1	89.0	151	91.0	89.6	89.1	1.7×10^{9}	91.6	90.0	89.1	2.8×10^{24}
11	tae	151	62.9	62.5	62.3	61.9	97	64.5	61.9	62.1	1.5×10^{6}	64.5	60.9	61.1	4.6×10^{14}
12	new-thyroid	215	92.6	93.2	92.6	92.6	26	92.6	92.8	93.0	3392	90.7	92.6	93.7	6.1×10^{7}
13	ecoli	336	77.5	79.1	77.6	77.8	57	79.9	79.4	78.4	1134750	80.3	78.2	77.0	3.8×10^{8}
			82.41	83.49	82.85	82.55	78.31	85.45	83.78	82.91	4.3×10^{7}	86.44	84.49	82.65	6.2×10^{14}

- The accuracy gets close to the combined solution, and much better than the first single tree:

Availability

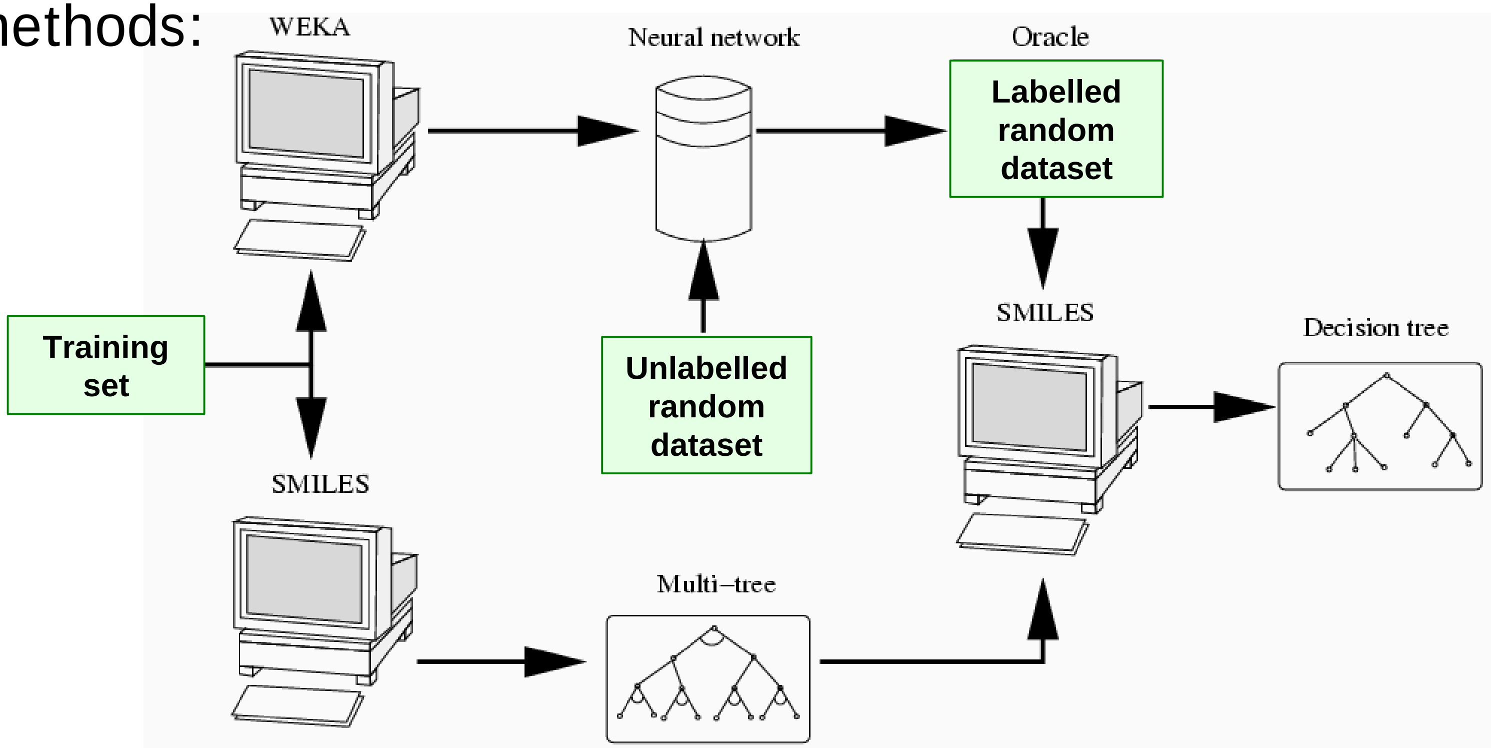
- **SMILES** is freely available at:

  http://www.dsic.upv.es/~flip/smiles/

- C++ sources.
- UNIX (Linux) and Windows versions.
- Many Examples (more than 30 datasets) adapted to **SMILES** format.
- Complete User Manual (90 pages).

Additional Applications

- **SMILES** can be used as a ‘by-pass’ for non-comprehensible ML methods:



- It’s different from stacking. The resulting model is semantically “similar” to the ANN but it is a comprehensible DT defined in terms of the original attributes.

Conclusions and Future Work

- **SMILES**:
  - combines and improves hypotheses combination and cost-sensitive learning (ROC analysis, AUC, test cost).
  - The *archetyping* technique provides a novel and different way to take advantage of classifier ensembles, especially shared ensembles.
  - Well suited for applications requiring high accuracy/AUC, low cost and high comprehensibility with flexible handling of resources.
- Future work:
  - Inputs and outputs in XML. (PMML standard)
  - Graphical interface.
  - Incremental extension.
  - Expressiveness extension (functional-logic, higher-order, ...)