SMILES: A Multi-purpose Learning System*

V. Estruch C. Ferri J. Hernández M.J. Ramírez

DSIC, UPV, Camino de Vera s/n, 46020 Valencia, Spain.
{vestruch,cferri,jorallo,mramirez}@dsic.upv.es

Abstract. A machine learning system is useful for extracting models from data that can be used for many applications such as data analysis, decision support or data mining. SMILES is a machine learning system that integrates many different features from other machine learning techniques and paradigms, and more importantly, it presents several innovations in almost all of these features, such as ensemble methods, cost-sensitive learning, and the generation of a comprehensible model from an ensemble. This paper contains a short description of the main features of the system as well as some experimental results.
Keywords: Decision tree learning, machine learning, data mining, cost-sensitive learning, comprehensibility, ensemble methods.

1 Introduction

SMILES (Stunning Multi-purpose Integrated LEarning System) is a machine learning system that contains features from some of the most recent directions in machine learning. In particular, it extends classical decision tree learners in many ways (new splitting criteria, non-greedy search, new partitions, extraction of several and different solutions, combination of solutions), and it can be configured to minimise misclassification costs.

In this paper we focus on the three main contributions of SMILES. First, we present the structure used in the system: a decision multi-tree. This kind of structure allows the generation of a set of hypotheses taking into consideration limited computational capabilities. We also combine the hypotheses in order to improve the overall accuracy, defining a new ensemble method. Secondly, we show an interesting capability of the system: the generation of a single representative solution from an ensemble of solutions. Finally, we introduce other features of the system on cost-sensitive learning.

2 The Structure of the SMILES System

The construction of decision trees is performed in two different steps: first, the whole decision tree is constructed by using a splitting criterion that selects the best split; secondly, unuseful parts of the tree are removed (*pruning*). Thus, a decision tree is built in a eager way, which allows a quick construction of a model. However, bad models may be produced because the alternatives that are not selected by the splitting criterion are immediately rejected. The main

* This work has been partially supported by CICYT under grant TIC2001-2705-C03-01 and by Generalitat Valenciana under grant GV00-092-14.

algorithm of SMILES is based on a new structure in which the rejected splits are not removed, but stored as *suspended* nodes. The further exploration of these nodes after the first solution has been built allows the extraction of new models from this structure. Since each new model is obtained by continuing the construction of the multi-tree, these models share their common parts. For this reason, a decision multi-tree can also be seen as an AND/OR tree [4], if one consider the alternative nodes as OR-nodes, and the nodes generated by an exploited OR-node as AND-nodes. Because each new solution is built following the construction of a complete tree, our method differs from other approaches such as *boosting* or *bagging* [6], which induce a new decision tree for each solution. The result is a multi-tree rather than a forest, with the advantage that a multi-tree shares the common parts and the forest does not. We perform a greedy search for each solution, but once the first solution is found the following ones can be obtained taking into account a limited computation time (the time and memory which are required to produce the next n solutions increases in a sublinear way). For more details on the structure, we refer to [5].

The number of trees (OR-nodes) to be explored and how to select them determines the resulting multi-tree. With all these opened branches we can do two things: select one solution or combine a set of solutions. For the former, we have implemented several selection criteria to specify which of all the possible solutions must be selected (*Occam*, *coverage best*,...). For the latter, one way to use the multiplicity in the multi-tree is to combine the results of different branches. SMILES implements several criteria to combine a set of solutions given in the multi-tree (*shared ensemble*) [1].

3 Selecting an Archetype from a Shared Ensemble

A combined hypothesis is usually a voting of many hypotheses and, as a result, the comprehensibility of the single solution is lost. For this reason, SMILES is also able to select one single solution (we call it the *archetype* of the ensemble [3]) that is semantically close to the combined hypothesis. This is done without the use of an extra validation dataset but the use of an invented random dataset.

The use of an archetype hypothesis is a way to obtain similar results to other methods such as *boosting* or *bagging*, but maintaining the comprehensibility of a declarative model.

4 Cost-sensitive Learning and ROC Analysis Features

Accuracy (or error), i.e., percentage of instances that are correctly classified (respectively incorrectly classified) has traditionally been used as a measure of the quality of classifiers. However, in most situations, not every misclassification has the same consequences. SMILES contains different methods for cost-sensitive learning: assignment of classes based on costs, cost-sensitive splitting criteria,...

The usefulness of cost-sensitive learning does not only apply when the misclassification costs or class distributions are known a priori. If they are not known, one or many classifiers can be generated in order to behave well in the widest range of circumstances or contexts. The Receiver Operating Characteristic (ROC) analysis provides tools to select a set of classifiers that would behave optimally and reject some other useless classifiers. In particular, SMILES is the first system which incorporates measures based on the area under the ROC curve to generate and select models which behave well in the presence of variable cost contexts or class distribution changes [2].

5 Experiments

In this section we present an experimental evaluation of the SMILES system. Here, we only present results of the combined hypothesis of our ensemble method w.r.t. other well-known methods. A deeper experimental evaluation of other features can be found in [1–3].

	Bagging			Boosting			Multi-tree		
#	10	100	300	10	100	300	10	100	300
balance-scale	82.24	82.76	82.87	78.72	76.00	75.60	77.89	83.02	85.50
cars	93.89	94.36	94.29	95.92	97.07	97.15	89.34	90.90	91.53
dermatology	96.40	97.19	97.30	96.62	96.51	96.65	91.43	94.00	95.71
ecoli	83.90	85.15	85.56	83.66	84.29	84.20	78.58	80.09	79.64
house-votes	95.40	95.73	95.77	95.19	95.24	95.42	94.56	95.93	96.21
iris	94.20	94.27	94.53	94.20	94.53	94.53	94.27	94.47	94.47
monks1	99.95	100.00	100.00	99.46	99.46	99.46	96.45	99.89	100.00
monks2	65.52	67.51	67.94	76.67	82.17	83.40	75.33	77.15	79.37
monks3	98.76	98.88	98.88	97.96	97.92	97.92	97.84	97.62	97.65
new-thyroid	94.33	94.66	94.81	94.98	95.31	95.22	93.43	92.57	92.71
post-operative	63.11	64.89	64.78	59.67	59.00	59.00	63.00	67.00	67.75
soybean-small	97.75	97.95	97.95	97.95	97.95	97.95	96.00	94.75	95.75
tae	60.10	61.11	61.05	64.81	64.81	64.81	65.00	65.13	65.40
tic-tac	83.06	84.05	83.91	82.11	82.62	82.55	79.23	82.68	83.72
wine	94.90	95.90	96.35	95.90	96.85	96.57	93.29	92.53	91.94
GeoMean	85.77	86.59	87.03	86.61	86.99	86.70	84.91	86.49	87.16

Table 1. Accuracy comparison between ensemble methods.

Table 1 presents a comparison of accuracy between our method (*multi-tree*), *boosting* and *bagging*, depending on the number of iterations. We have employed the Weka[1] implementation of these two ensemble methods. For all the experiments we have used *GainRatio* as splitting criterion, and we have chosen a simple random method for populating the multi-tree and a fusion strategy based on selecting the branch that gives a maximum cardinality for the majority class. The datasets have been extracted from the UCI repository[2].The experiments were performed with a Pentium III-800Mhz with 180MB of memory running linux 2.4.2. Since there are many sources of randomness, we have performed the experiments by averaging 10 results of a 10-fold cross-validation (1500 runs in total). The results present the mean accuracy for each dataset, and finally the geometric mean of all the datasets. Although initially our method obtains lower results with a few iterations, with a higher number of iterations it surpasses the other systems.

Nevertheless, the major advantage of the method is appreciated by looking at the consumption of resources. Figure 1 shows the average training time depending on the number of iterations (1-300) for the three methods. Note that the time increase of *Bagging* is linear, as expected. *Boosting* behaves better with high

[1] `http://www.cs.waikato.ac.nz/~ml/weka/`
[2] `http://www.ics.uci.edu/~mlearn/MLSummary.html`

values because the algorithm implemented in Weka trickily stops the learning if it does not detect a significant increasing of accuracy. Finally, SMILES presents a sub-linear increase of required time due to the sharing of common components of the multi-tree structure.

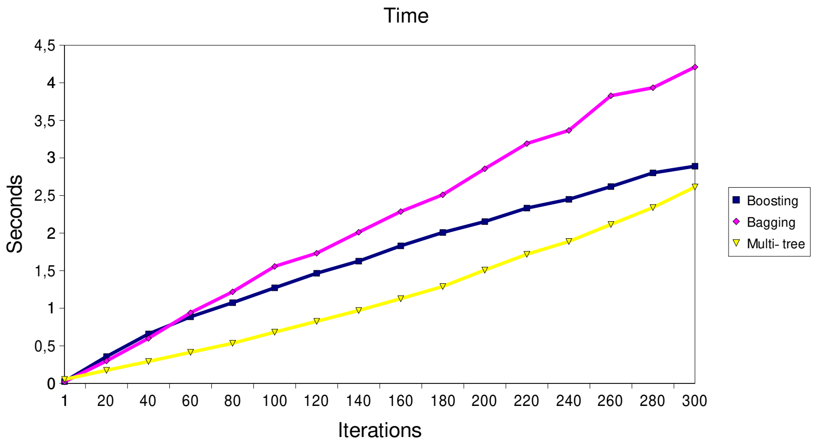

Fig. 1. Time comparison between ensemble methods.

6 Conclusions

Summing up, the SMILES system surpasses other ensemble learning methods in terms of accuracy and resource requirements by the use of shared ensembles. It also allows a wide range of applications in data mining, especially when cost-information can hazardously vary at application time or whenever a comprehensible model is required with high accuracy. These comprehensible models can be embodied into other kinds of AI systems: expert systems, knowledge-based systems and decision-support systems.

References

1. V. Estruch, C. Ferri, J. Hernández, and M.J. Ramírez. Shared Ensembles using Multi-trees. In *the 8th Ib. Conf. on A. I., Iberamia'02*. (Submitted), 2002.
2. C. Ferri, P. Flach, and J. Hernández. Learning Decision Trees using the Area Under the ROC Curve. In *Int. Conf. on Machine Learning, ICML'02*. (To appear), 2002.
3. C. Ferri, J. Hernández, and M.J. Ramírez. From Ensemble Methods to Comprehensible Models. In *the 5th Int. Conf. on Discovery Science*. (Submitted), 2002.
4. C. Ferri, J. Hernández, and M.J. Ramírez. Induction of Decision Multi-trees using Levin Search. In *Int. Conf. on Computational Science, ICCS'02*, volume 2329 of *LNCS*, pages 166–175, 2002.
5. C. Ferri, J. Hernández, and M.J. Ramírez. SMILES System, a multi-purpose learning system. `http://www.dsic.upv.es/~flip/smiles/`, 2002.
6. J. R. Quinlan. Bagging, Boosting, and C4.5. In *Proc. of the 13th Nat. Conf. on A.I. and the Eighth Innovative Applications of A.I. Conf.*, pages 725–730. AAAI Press / MIT Press, 1996.