Information Sciences 240 (2013) 1–20

Contents lists available at SciVerse ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

# An interpretable classification rule mining algorithm

Alberto Cano, Amelia Zafra, Sebastián Ventura *

*Department of Computer Science and Numerical Analysis, University of Córdoba, Spain*

A R T I C L E   I N F O

*Article history:*
Received 9 October 2011
Received in revised form 14 February 2013
Accepted 16 March 2013
Available online 6 April 2013

*Keywords:*
Classification
Evolutionary programming
Interpretability
Rule mining

A B S T R A C T

Obtaining comprehensible classifiers may be as important as achieving high accuracy in many real-life applications such as knowledge discovery tools and decision support systems. This paper introduces an efficient Evolutionary Programming algorithm for solving classification problems by means of very interpretable and comprehensible IF-THEN classification rules. This algorithm, called the Interpretable Classification Rule Mining (ICRM) algorithm, is designed to maximize the comprehensibility of the classifier by minimizing the number of rules and the number of conditions. The evolutionary process is conducted to construct classification rules using only relevant attributes, avoiding noisy and redundant data information. The algorithm is evaluated and compared to nine other well-known classification techniques in 35 varied application domains. Experimental results are validated using several non-parametric statistical tests applied on multiple classification and interpretability metrics. The experiments show that the proposal obtains good results, improving significantly the interpretability measures over the rest of the algorithms, while achieving competitive accuracy. This is a significant advantage over other algorithms as it allows to obtain an accurate and very comprehensible classifier quickly.



## 1. Introduction

Discovering knowledge in large amounts of data collected over the last decades has become significantly challenging and difficult, especially in large-scale databases. Data mining (DM) [60] involves the use of data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. Classification and regression are two forms of data analysis which can be used to extract models describing important data classes or to predict future data trends. Classification predicts categorical labels whereas regression models predict continuous-valued functions.

The data analysis tools used for DM include statistical models, mathematical methods, and machine learning algorithms. Classification is a common task in supervised machine learning with the search for algorithms that learn from training examples to produce predictions about future examples.

Classification has been successfully solved using several approaches [26]. On the one hand, there are approaches such as artificial neural networks (ANN) [46], support vector machines (SVM) [16], and instance-based learning methods [2]. These approaches obtain accurate classification models but they must be regarded as black boxes, i.e., they are opaque to the user. Opaque predictive models prevent the user from tracing the logic behind a prediction and obtaining interesting knowledge previously unknown from the model. These classifiers do not permit human understanding and inspection, they are not directly interpretable by an expert and it is not possible to discover which are the relevant attributes to predict the class of an example. This opacity prevents them from being used in many real-life knowledge discovery applications where both accuracy and comprehensibility are required, such as medical diagnosis [55], credit risk evaluation [42], and decision support systems [6], since the prediction model must explain the reasons for classification.

* Corresponding author.
*E-mail addresses:* acano@uco.es (A. Cano), azafra@uco.es (A. Zafra), sventura@uco.es (S. Ventura).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.03.038

On the other hand, there are machine learning approaches which overcome this limitation and provide transparent and comprehensible classifiers such as decision trees [62] and rule-based systems [49]. Evolutionary Algorithms [65], and specifically Evolutionary Programming (EP) [13,64] and Genetic Programming (GP) [25], have been successfully applied to build decision trees and rule-based systems easily. Rule-based systems are especially user-friendly and offer compact, understandable, intuitive and accurate classification models. To obtain comprehensibility, accuracy is often sacrificed by using simpler but transparent models, achieving a trade-off between accuracy and comprehensibility. Even though there are many rule based classification models, it has not been until recently that the comprehensibility of the models is becoming a more relevant objective. Proof of this trend is found in recent studies of issue [18,27,34,57], i.e, the comprehensibility of the models is a new challenge as important as accuracy. This paper focuses on the interpretability, trying to reach more comprehensible models than most of the current proposals and thus covering the needs of many application domains that require greater comprehensibility than the provided by current methods.

This paper presents an EP approach applied to classification problems to obtain comprehensible rule-based classifiers. This algorithm, called ICRM (Interpretable Classification Rule Mining), is designed to obtain a base of rules with the minimum number of rules and conditions, in order to maximize its interpretability, while obtaining competitive accuracy results. The algorithm uses an individual = rule representation, following the Iterative Rule Learning (IRL) model. Individuals are constructed by means of a context-free grammar [33,61], which establishes a formal definition of the syntactical restrictions of the problem to be solved and its possible solutions, so that only grammatically correct individuals are generated. Next, the most important characteristics of the algorithm are detailed. Firstly, the algorithm guarantees obtaining the minimum number of rules. This is possible because it generates one rule per class, together with a default class prediction, which is assigned when none of the available rules are triggered. Moreover, it is guaranteed that there are no contradictory or redundant rules, i.e., there is no pair of rules with the same antecedents and different consequents. Finally, it also guarantees the minimum number of conditions forming the antecedents of these rules, which is achieved by selecting only the most relevant and discriminating attributes that separate the classes in the attribute domains.

The experiments carried out on 35 different data sets and nine other algorithms show the competitive performance of our proposal in terms of predictive accuracy and execution time, obtaining significantly better results than all the other algorithms in terms of all the interpretability measures considered: the minimum number of rules, minimum number of conditions per rule, and minimum number of conditions of the classifier. The experimental study includes a statistical analysis based on the Bonferroni–Dunn [24] and Wilcoxon [59] non-parametric tests [28,29] in order to evaluate whether there are statistically differences in the results of the algorithms.

This paper is structured as follows. Section 2 briefly reviews the related background works. Section 3 describes the ICRM algorithm. Section 4 describes the experimental study whose results are discussed in Section 5. Finally, Section 6 draws some conclusions raised from the work.

## 2. Background

This section introduces the accuracy vs interpretability problem and discusses the interpretability definition and metrics. Finally, it briefly reviews the most important works related to genetic rule-based classification systems in recent years.

### *2.1. Accuracy vs interpretability*

Classification with rule-based systems comes with two contradictory requirements in the obtained model: the interpretability, capability the behavior of the real system in an understandable way, and the accuracy, capability to faithfully represent the real system. Obtaining high degrees of interpretability and accuracy is a contradictory purpose and, in practice, one of the two properties prevails over the other. To find the best trade-off between them is an optimization problem that is very difficult to solve efficiently [36].

In contrast, when looking for interpretability, fuzzy-based systems are usually considered [5,41]. These systems comprise very interpretable rules since they employ linguistic variables to address the vagueness of human language [44]. One of these systems is the algorithm by *González and Pérez*, SLAVE (Structural Learning Algorithm on Vague Environment) [30], which is a genetic learning algorithm that uses the iterative approach to learn fuzzy rules. Another fuzzy rule-based algorithm is GFS-GP [52], which combines genetic programming operators with simulated annealing search. The use of evolutionary programming, genetic programming and grammars to construct classification rules has been widely used [7,20,25,66]. However, this interpretability perspective is not related to the search for simpler classifiers in terms of fewer number of rules and conditions, in which we are really interested.

### *2.2. Interpretability metrics*

There is no well-established definition of the interpretability of a rule-based system. The interpretability of a rule set is very important, due to the fact that very large sets of rules or very complex rules are rather lacking in interest. In fact, studies have focused on reducing the complexity of rule-based classifiers [35,53]. Nevertheless, there are some indicators that allow us to estimate the interpretability and comprehensibility of a rule-based classifier, which are described by García et al. [27]: