Information Sciences 465 (2018) 505–522

Contents lists available at ScienceDirect

Information Sciences

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

UnPART: PART without the 'partial' condition of it

Igor Ibarguren[1,*], Jesús M. Pérez[1], Javier Muguerza[1], Ibai Gurrutxaga[1], Olatz Arbelaitz[1]

Department of Computer Architecture and Technology, University of the Basque Country UPV/EHU, Manuel Lardizabal 1, 20018 Donostia, Spain

ARTICLE INFO

Article history:
Received 8 September 2017
Revised 1 March 2018
Accepted 8 July 2018

Keywords:
Comprehensible classifiers
Interpretable models
Rule sets
Full decision trees
Partial decision trees
Machine learning

ABSTRACT

The PART rule-induction algorithm creates rulesets by iteratively creating partial decision trees and extracting a rule from each tree. A recent study showed that growing trees further and combining it with pruning created classifiers with better discriminating capacity and less structural complexity. In this work we propose an algorithm that works in a similar way to PART, but building decision trees to their full extent, dropping the 'partial' condition of PART. We call this algorithm UnPART. We also propose using a different decision tree as base for PART-like algorithms. We choose CHAID* as replacement for C4.5, and propose CHAID*-based UnPART, PART and BFPART algorithms. We compare the six PART-like algorithms and their base decision trees from different points of view: discriminating capacity, structural complexity and computational cost. Results show that C4.5-based UnPART creates the best classifying models whereas CHAID*-based UnPART creates the simplest classifiers. We then compare these eight algorithms to a wider set of 21 comprehensible decision tree and rule-induction algorithms over 96 datasets from the perspective of discriminating capacity. Results show that C4.5-based UnPART has the best discriminating capacity among the compared algorithms.

1. Introduction

In the supervised machine-learning area, algorithms use training data to build models that correctly label unseen data. When the output to be predicted is limited to a known set of values, these models are called classifiers and their goal is to correctly predict the class of unseen examples. In many real world applications, specially those meant to be decision support systems for human users (such as medicine, fraud detection and finance), knowing the reasoning behind a classification is almost as important as the correct classification itself, as the reasoning helps gauge the output of the classifier.

In recent years the comprehensibility of the knowledge extracted with classification algorithms has become an active issue in the machine-learning community. In [15], we recently compiled a set of academic references and quotes by highly influential people highlighting the importance of understability.

Not all classifier models are understandable to humans. Some widely-used algorithms such as Artificial Neural Networks and Support Vector Machines are too complex due to their structure. Decision trees and rulesets provide comprehensible

* Corresponding author.
E-mail addresses: igor.ibarguren@ehu.eus (I. Ibarguren), txus.perez@ehu.eus (J.M. Pérez), j.muguerza@ehu.eus (J. Muguerza), i.gurrutxaga@ehu.eus (I. Gurrutxaga), olatz.arbelaitz@ehu.eus (O. Arbelaitz).
[1] http://www.aldapa.eus/

https://doi.org/10.1016/j.ins.2018.07.022
0020-0255/

models because the decisions within the rules are simple conditions about the values the data can take. The most-common approach to improving the discriminating capacity of rulesets and decision trees is to create multiple classifier systems by combining multiple simple classification models. However, as Domingos' argued "while a single decision tree can be easily understood by a human as long as it is not too large, fifty such trees, even if individually simple, exceed the capacity of even the most patient user" [7].

Among the most widely used rule-induction algorithms there is PART [10], proposed by the creators of the well-known WEKA software [14], an algorithm that combines the two major rule-induction paradigms: extracting rules from decision trees [20] and the separate-and-conquer technique [11] used by some rule-induction algorithms such as RIPPER [4]. PART works by partially constructing a C4.5 tree, and using the partial tree's most populous treated leaf node to generate a rule. The training examples covered by this rule are removed from the training set and the process is repeated until no more partial trees can be built.

A recently proposed variant of PART, BFPART [15], creates rulesets of greater discriminating capacity and lower structural complexity compared to PART. BFPART proposed multiple changes to the PART algorithm. Most notably, the use of the Best-First global search algorithm instead of the Hill-Climbing algorithm PART originally uses. As Best-First searches do not restrict to the current tree branch when choosing the next node to be developed in the partial C4.5 tree, this results in BFPART partial trees developing more than PART trees. The experiments comparing BFPART and PART seem to suggest that developing partial trees to a greater degree creates better classifiers. In fact, the best performing variants of PART were those almost fully developing the partial trees.

Based on that, we propose the following research questions:

1. Is using 'partial' trees key to the good discriminating capacity of PART-like algorithms? Or, can better rules be found if decision trees are fully developed?
2. How would that affect the structural complexity and the computational cost of the algorithm?
3. How would changing the base decision tree of PART-like algorithms affect their performance? Could CHAID* [15], an algorithm with similar discriminating capacity that generates simpler classifiers, be used as base?
4. How would these PART-like algorithms place in a wider comparison against state of the art algorithms for rule-induction?

In order to answer those questions, in this work we propose the following contributions. We first propose UnPART (1), a rule-induction algorithm that uses the same principle as PART, but using fully grown decision trees, and we compare it to BFPART, PART, and C4.5 (2). We also propose CHAID*-based variants for UnPART, BFPART and PART (3). Moreover, we place all of these algorithms in a wider study with 23 genetics-based and classical rule-induction and decision tree algorithms over 96 datasets (4). Our results show that C4.5-based UnPART ranks best among all compared algorithms, creating the simplest classifiers among C4.5-based algorithms. On the other hand, CHAID*-based UnPART creates even simpler models, sacrificing some of the discriminating capacity.

The results of the genetics-based and classical algorithms come from a study originally published in [9], a reference work that at the time of writing, has been cited 59 times in the ISI Web of Knowledge and 94 times in *Google Scholar*. The authors of that study published their train/test partitions to encourage other researchers to use their data, and their results to benchmark new algorithms.

The experiments in this work are divided into three parts. First we compare C4.5-based PART-like algorithms and C4.5 over 96 datasets. Then, we compare the results of the C4.5-based and CHAID*-based algorithms. In these two parts of the study, the algorithms are compared from six different points of view by using six criteria: Kappa, the Geometric Mean (GM), and the Area Under the ROC Curve (AUC) for discriminating capacity; the average rule or branch length (Length), and the product of Length by the number of rules or leaves for structural complexity; and time as computational cost. In [9], the authors created a taxonomy of genetics-based rule-induction algorithms and carried out an extensive study comparing 16 genetics-based and other six classical rule-induction and decision tree algorithms. Finally, the discriminating capacity of the six PART-based algorithms and CHAID* is compared to the results of that study. In order to distinguish the same variants with a different base decision tree algorithm, we use the *_C45* suffix for C4.5-based algorithms and *_CHD* for CHAID*-based algorithms.

The results of our first study, comparing UnPART_C45, BFPART_C45, PART_C45, and C4.5 show that UnPART_C45 creates the best classifying and simplest models. When also using CHAID* as the base for the PART variants, C4.5-based algorithms generate classifiers with a better discriminating capacity than CHAID*-based algorithms, whereas CHAID*-based algorithms create simpler classifiers. When comparing these algorithms with the genetics-based and classical algorithms, UnPART_C45 performs best among the 31 compared algorithms. Both UnPART versions rank better than their base decision tree algorithm.

The results presented in this article have been tested for significance using the state-of-the-art techniques proposed by Demšar in [5], Garcia et al. in [12,13], and Derrac et al. in [6]. Also, in order to properly combine the different performance measures used in this work, we have applied a cost-conscious methodology [21] to combine discriminating capacity and complexity/cost metrics.

The paper proceeds as follows. In Section 2 we give an insight into previous work. Section 3 explains the UnPART algorithm and compares it to the PART and BFPART algorithms. In Section 4 we define the experimental setup. In Section 5 we present the empirical results of the experiments. Finally, in Section 6, we make some concluding remarks for this work and propose future work.