

# Accepted Manuscript

BFPART: Best-First PART

Igor Ibarguren, Aritz Lasarguren, Jesús M. Pérez, Javier Muguerza,  
Ibai Gurrutxaga, Olatz Arbelaitz

PII: S0020-0255(16)30506-0  
DOI: [10.1016/j.ins.2016.07.023](https://doi.org/10.1016/j.ins.2016.07.023)  
Reference: INS 12349



To appear in: *Information Sciences*

Received date: 27 April 2015  
Revised date: 30 June 2016  
Accepted date: 6 July 2016

Please cite this article as: Igor Ibarguren, Aritz Lasarguren, Jesús M. Pérez, Javier Muguerza, Ibai Gurrutxaga, Olatz Arbelaitz, BFPART: Best-First PART, *Information Sciences* (2016), doi: [10.1016/j.ins.2016.07.023](https://doi.org/10.1016/j.ins.2016.07.023)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# BFPART: Best-First PART

Igor Ibarguren\*, Aritz Lasarguren, Jesús M. Pérez, Javier Muguerza, Ibai Gurrutxaga, Olatz Arbelaiz

*Department of Computer Architecture and Technology, University of the Basque Country  
UPV/EHU, Manuel Lardizabal 1, 20018 Donostia, Spain  
<http://www.aldapa.eus/>*

## Abstract

In supervised classification, decision tree and rule induction algorithms possess the desired ability to build understandable models. The PART algorithm creates partially developed C4.5 decision trees and extracts a rule from each tree. Some of the criteria used by this algorithm can be modified to yield better results. In this work, we propose and compare 16 variants of the PART algorithm from the perspectives of discriminating capacity, complexity of the models, and the computational cost, for 36 real-world problems obtained from the UCI repository. The use of the Best-First optimization algorithm to find the next node to develop in a partial tree improves the results of the PART algorithm. The best-performing variant also ranks first when compared to the well-established C4.5 algorithm and a modified version of the CHAID decision tree induction algorithm that handles continuous features, which is also proposed in this paper. In order to study its performance in comparison to other rivals, this comparison of algorithms also includes the original PART algorithm. For all performance measures, we test the results for statistical significance using state-of-the-art methods.

*Keywords:* Local search algorithms, Best-First, comprehensible classifiers, interpretable models, rules sets, partial decision trees, machine learning.

## 1. Introduction

Interpretability is an added value sought in classifiers that are built within the supervised machine-learning area. Given a set of training data, a common task is to extract a model that predicts the class label of an unseen example. The generalization or discriminating capacity of a classifier measures the number of these predictions of unseen examples that are correct. Most efforts made in

---

\*Corresponding author

*Email addresses:* [igor.ibarguren@ehu.eus](mailto:igor.ibarguren@ehu.eus) (Igor Ibarguren), [aritz.lasarguren@ehu.eus](mailto:aritz.lasarguren@ehu.eus) (Aritz Lasarguren), [txus.perez@ehu.eus](mailto:txus.perez@ehu.eus) (Jesús M. Pérez), [j.muguerza@ehu.eus](mailto:j.muguerza@ehu.eus) (Javier Muguerza), [i.gurrutxaga@ehu.eus](mailto:i.gurrutxaga@ehu.eus) (Ibai Gurrutxaga), [o.arbelaitz@ehu.eus](mailto:o.arbelaitz@ehu.eus) (Olatz Arbelaiz)

Download English Version:

<https://daneshyari.com/en/article/6857291>

Download Persian Version:

<https://daneshyari.com/article/6857291>

[Daneshyari.com](https://daneshyari.com)