

# Accepted Manuscript

## A Greedy Search Tree Heuristic for Symbolic Regression

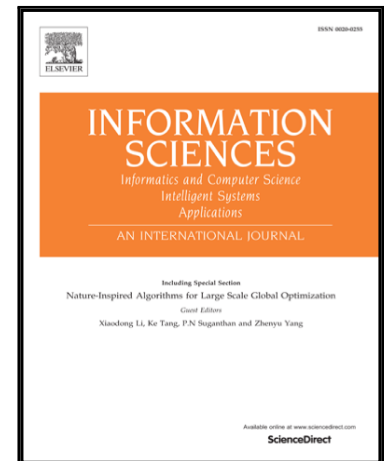
Fabrício Olivetti de França

PII: S0020-0255(16)30863-5  
DOI: [10.1016/j.ins.2018.02.040](https://doi.org/10.1016/j.ins.2018.02.040)  
Reference: INS 13445

To appear in: *Information Sciences*

Received date: 12 September 2016  
Revised date: 28 June 2017  
Accepted date: 17 February 2018

Please cite this article as: Fabrício Olivetti de França, A Greedy Search Tree Heuristic for Symbolic Regression, *Information Sciences* (2018), doi: [10.1016/j.ins.2018.02.040](https://doi.org/10.1016/j.ins.2018.02.040)



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.

# A Greedy Search Tree Heuristic for Symbolic Regression

Fabrício Olivetti de França<sup>a</sup>

<sup>a</sup>Center of Mathematics, Computing and Cognition (CMCC), Universidade Federal do ABC (UFABC) – Santo André, SP, Brazil. E-mail: folivetti@ufabc.edu.br

---

## Abstract

Symbolic Regression tries to find a mathematical expression that describes the relationship of a set of explanatory variables to a measured variable. The main objective is to find a model that minimizes the error and, optionally, that also minimizes the expression size. A smaller expression can be seen as an interpretable model considered a reliable decision model. This is often performed with Genetic Programming, which represents their solution as expression trees. The shortcoming of this algorithm lies on this representation that defines a rugged search space and contains expressions of any size and difficulty. These pose as a challenge to find the optimal solution under computational constraints. This paper introduces a new data structure, called Interaction-Transformation (IT), that constrains the search space in order to exclude a region of larger and more complicated expressions. In order to test this data structure, it was also introduced an heuristic called SymTree. The obtained results show evidence that SymTree are capable of obtaining the optimal solution whenever the target function is within the search space of the IT data structure and competitive results when it is not. Overall, the algorithm found a good compromise between accuracy and simplicity for all the generated models.

*Keywords:* Symbolic Regression, Regression Analysis, Greedy Heuristic

---

## 1. Introduction

Many decision making process can be automated by learning a computational model through a set of observed data. For example, credit risk can be estimated by using explanatory variables related to the consumer behavior [9]. A recommender system can estimate the likelihood of a  
5 given person to consume an item given their past transactions [1].

Download English Version:

<https://daneshyari.com/en/article/6856536>

Download Persian Version:

<https://daneshyari.com/article/6856536>

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