



# A VNS algorithm for feature selection in hierarchical classification context

H. Costa <sup>a,1</sup> L. R. Galvão <sup>a,2</sup> L. H. C. Merschmann <sup>b,3</sup>  
M. J. F. Souza <sup>a,4</sup>

<sup>a</sup> *Federal University of Ouro Preto, Ouro Preto, MG, 35400-000, Brazil*

<sup>b</sup> *Federal University of Lavras, Lavras, MG, 37200-000, Brazil*

---

## Abstract

Feature selection, usually adopted as a preprocessing step for data mining, is used to select a subset of predictive features aiming to improve the performance of a predictive model. Despite of the benefits of feature selection for classification task, to the best of our knowledge, there is no work in the literature that addresses feature selection in conjunction with global hierarchical classifiers. Thus, in this paper, we fill this gap proposing a feature selection method based on Variable Neighborhood Search (VNS) metaheuristic for the hierarchical classification context. Computational experiments were carried out on five bioinformatics datasets to evaluate the effect of the proposed algorithm on classification performance when using a global hierarchical classifier. As result, we have obtained a classifier performance improvement for three datasets and a competitive result for a fourth dataset, which indicates the suitability of the proposed method for the hierarchical classification scenario.

*Keywords:* Feature Selection, Hierarchical Classification, Variable Neighborhood Search, Filter, Wrapper

---

# 1 Introduction

In the last decades, data analysis techniques have become the main focus of attention because of the huge increase in data generation and storage, mainly in biological area. The transformation of these data in useful, novel, valid and understandable information introduces new challenges to researchers. Since it is not a trivial task, automated strategies to analyze the data are required. Therefore, the Knowledge Discovery in Databases (KDD) process adopted for this purpose is composed by three main steps: data preprocessing, data mining and results evaluation.

Feature selection, usually adopted as a preprocessing step, aims at identifying as much relevant features as possible with the goal of improving the performance of data mining techniques. Among the benefits of feature selection are improvement of predictive accuracy of classifiers and reduction of the execution time of classification processes [4].

Classification aims to predict the class label(s) of an object based on its features. In flat classification problems, each dataset example is assigned to one or more classes, wherein there are no hierarchical relationships among the classes. However, there are more complex classification problems, known as hierarchical classification problems, where the classes are naturally organized into hierarchies [11] represented by a tree or a direct acyclic graph (DAG).

In hierarchical classification problems different approaches can be used to handle the class hierarchy. In the local model approach, the classification is performed using a set of flat classifiers, while in the global model approach, a single classifier, designed to consider the class hierarchy as a whole, performs the classification.

Secker *et al.* [9] solved a protein classification problem by performing feature selection before training a flat classifier for each node of the class hierarchy. Paes *et al.* [7] used the Information Gain measure [4] to select the best features prior to the training of the flat classifiers which form the hierarchical classifiers. In a different way, in this work, we have proposed a feature selection technique to be used in conjunction with global hierarchical classifiers, i.e., without decomposing the hierarchical problem in several problems of flat classification. To the best of our knowledge, there is no work in the literature that addresses feature selection in conjunction with global hierarchical clas-

---

<sup>1</sup> Email: helen.c.s.costa@gmail.com

<sup>2</sup> Email: leandrodvmg@gmail.com

<sup>3</sup> Email: luiz.hcm@dcc.ufla.br

<sup>4</sup> Email: marcone@iceb.ufop.br

Download English Version:

<https://daneshyari.com/en/article/8903346>

Download Persian Version:

<https://daneshyari.com/article/8903346>

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