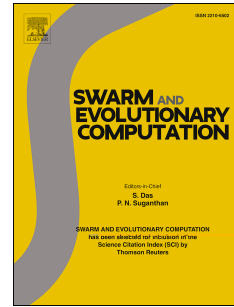


Accepted Manuscript

Large-dimensionality small-instance set feature selection: A hybrid bio-inspired heuristic approach

Hossam M. Zawbaa, E. Emary, Crina Grosan, Vaclav Snasel



PII: S2210-6502(16)30304-2

DOI: [10.1016/j.swevo.2018.02.021](https://doi.org/10.1016/j.swevo.2018.02.021)

Reference: SWEVO 374

To appear in: *Swarm and Evolutionary Computation BASE DATA*

Received Date: 26 September 2016

Revised Date: 2 November 2017

Accepted Date: 24 February 2018

Please cite this article as: H.M. Zawbaa, E. Emary, C. Grosan, V. Snasel, Large-dimensionality small-instance set feature selection: A hybrid bio-inspired heuristic approach, *Swarm and Evolutionary Computation BASE DATA* (2018), doi: 10.1016/j.swevo.2018.02.021.

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.

Large-dimensionality small-instance set feature selection: a hybrid bio-inspired heuristic approach

Hossam M. Zawbaa^{1,2,*}, E. Emary^{3,4}, Crina Grosan^{1,5}, Vaclav Snasel⁶

¹*Faculty of Mathematics and Computer Science, Babes-Bolyai University, Romania.*

²*Faculty of Computers and Information, Beni-Suef University, Egypt.*

³*Faculty of Computers and Information, Cairo University, Egypt.*

⁴*Faculty of Computer Studies, Arab Open University, Cairo, Egypt.*

⁵*College of Engineering, Design and Physical Sciences, Brunel University, United Kingdom.*

⁶*Faculty of Electrical Engineering and Computer Science, VSB-Technical University of Ostrava, Czech Republic.*

**hossam.zawbaa@gmail.com*

Abstract

Selection of a representative set of features is still a crucial and challenging problem in machine learning. The complexity of the problem increases when any of the following situations occur: a very large number of attributes (large dimensionality); a very small number of instances or time points (small-instance set). The first situation poses problems for machine learning algorithm as the search space for selecting a combination of relevant features becomes impossible to explore in a reasonable time and with reasonable computational resources. The second aspect poses the problem of having insufficient data to learn from (insufficient examples). In this work, we approach both these issues at the same time. The methods we proposed are heuristics inspired from nature (in particular, from biology). We propose a hybrid of two methods which has the advantage of providing a good learning from fewer examples and a fair selection of features from a really large set, all these while ensuring a high standard classification accuracy of the data. The methods used are antlion optimization (ALO), grey wolf optimization (GWO), and a combination of the two (ALO-GWO). We test their performance on datasets having almost 50,000 features and less than 200 instances. The results look promising while compared with other methods such as genetic algorithms (GA) and particle swarm optimization (PSO).

Download English Version:

<https://daneshyari.com/en/article/8953859>

Download Persian Version:

<https://daneshyari.com/article/8953859>

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