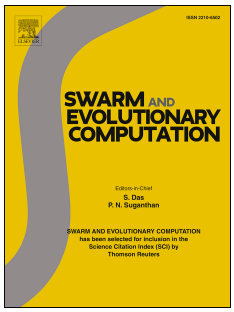
Accepted Manuscript



Combining three strategies for evolutionary instance selection for instance-based learning

Aida de Haro-García, Javier Pérez-Rodríguez, Nicolás García-Pedrajas

PII: S2210-6502(17)30592-8

DOI: 10.1016/j.swevo.2018.02.022

Reference: SWEVO 375

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

Received Date: 11 July 2017

Revised Date: 25 February 2018

Accepted Date: 26 February 2018

Please cite this article as: A. de Haro-García, J. Pérez-Rodríguez, Nicolá. García-Pedrajas, Combining three strategies for evolutionary instance selection for instance-based learning, *Swarm and Evolutionary Computation BASE DATA* (2018), doi: 10.1016/j.swevo.2018.02.022.

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.

ACCEPTED MANUSCRIPT

Combining three strategies for evolutionary instance selection for instance-based learning

Aida de Haro-García[a], Javier Pérez-Rodríguez[a], Nicolás García-Pedrajas[a,*]

[a] *University of Córdoba, Campus de Rabanales, 14011 Córdoba (Spain)*

Abstract

Instance-based learning methods, such as the k-nearest neighbor rule, are among the top-performing methods in any classification task. Despite their simplicity, they achieve comparable performance to much more complex methods. However, one of their problems is the necessity to store all of the training instances in memory. For large datasets, this might also affect the speed of the testing process. As a solution to this problem, instance selection methods, which remove redundant and noisy instances, have been proposed. However, a general side effect of these methods is that they produce a significant reduction in the accuracy of the instance-based learner. The classification performance is significantly worse compared to the application of the k-nearest neighbor rule using all instances.

In this paper, we propose an evolutionary instance selection algorithm that combines three strategies to avoid this negative side effect. First, it uses the framework of a CHC genetic algorithm, as it has been proved to be the best-performing method for this task. Second, it incorporates the possibility of selecting each instance more than once. This has also been proven useful in previous works. Finally, it uses a local value for k that depends on the nearest neighbor of every test instance. These three combined strategies are able to achieve better reduction than previous approaches while maintaining the same classification performance as the k-nearest neighbor rule.

In a large set of 150 real-world problems, our approach proves to be

[☆] This work was supported in part by Project TIN2015-66108-P of the Spanish Ministry of Science and Innovation

[*] Corresponding author

Email addresses: adeharo@uco.es (Aida de Haro-García), javier.perez@uco.es (Javier Pérez-Rodríguez), npedrajas@uco.es (Nicolás García-Pedrajas)

Download English Version:

https://daneshyari.com/en/article/8953868

Download Persian Version:

https://daneshyari.com/article/8953868

Daneshyari.com