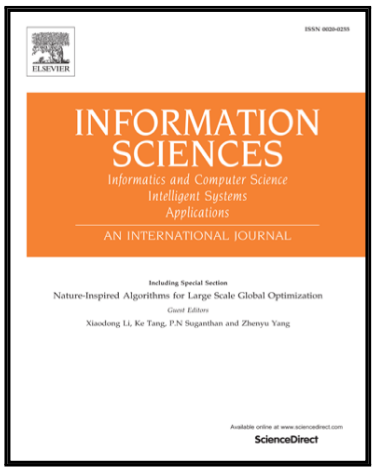
Accepted Manuscript



A Distributed Approach to Multi-Objective Evolutionary Generation of Fuzzy Rule-based Classifiers from Big Data

Andrea Ferranti, Francesco Marcelloni, Armando Segatori,
Michela Antonelli, Pietro Ducange

PII: S0020-0255(17)30809-5
DOI: 10.1016/j.ins.2017.06.039
Reference: INS 12961

To appear in: *Information Sciences*

Received date: 2 August 2016
Revised date: 24 June 2017
Accepted date: 26 June 2017

Please cite this article as: Andrea Ferranti, Francesco Marcelloni, Armando Segatori, Michela Antonelli, Pietro Ducange, A Distributed Approach to Multi-Objective Evolutionary Generation of Fuzzy Rule-based Classifiers from Big Data, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.06.039

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

A Distributed Approach to Multi-Objective Evolutionary Generation of Fuzzy Rule-based Classifiers from Big Data

Andrea Ferranti[a], Francesco Marcelloni[a,*], Armando Segatori[a], Michela Antonelli[b], Pietro Ducange[c]

[a] *Dipartimento di Ingegneria dell'Informazione: University of Pisa, Pisa, Italy*
[b] *Translational Imaging Group, University College of London, London, UK*
[c] *SMART Engineering Solutions & Technologies (SMARTEST) Research Centre, eCAMPUS University, Novedrate, Italy*

Abstract

In the last years, multi-objective evolutionary algorithms (MOEAs) have been extensively used to generate sets of fuzzy rule-based classifiers (FRBCs) with different trade-offs between accuracy and interpretability. Since the computation of the accuracy for each chromosome evaluation requires the scan of the overall training set, these approaches have proved to be very expensive in terms of execution time and memory occupation. For this reason, they have not been applied to very large datasets yet. On the other hand, just for these datasets, interpretablity of classifiers would be very desirable. In the last years the advent of a number of open source cluster computing frameworks has however opened new interesting perspectives. In this paper, we exploit one of these frameworks, namely Apache Spark, and propose the first distributed multi-objective evolutionary approach to learn concurrently the rule and data bases of FRBCs by maximizing accuracy and minimizing complexity. During the evolutionary process, the computation of the fitness is divided among the cluster nodes, thus allowing the designer to distribute both the computational complexity and the dataset storing. We have performed a number of experiments on ten real-world big datasets, evaluating our distributed approach in terms of both classification rate and scalability, and comparing it with two well-known state-of-art distributed classifiers. Finally, we have evaluated the achievable speedup on a small computer cluster. We present that the distributed version can efficiently extract compact rule bases with high accuracy, preserving the interpretability of the rule base, and can manage big datasets even with modest hardware support.

Keywords: Multi-objective Evolutionary Fuzzy Systems, Big Data, Fuzzy Rule-based Classifiers, Apache Spark

*Corresponding author, email: f.marcelloni@iet.unipi.it, Tel: +39 0502217678 Fax: +39 0502217600
Email addresses: andreaferranti@hotmail.it (Andrea Ferranti), f.marcelloni@iet.unipi.it (Francesco Marcelloni), armando.segatori@for.unipi.it (Armando Segatori), m.antonelli@ucl.ac.uk (Michela Antonelli), pietro.ducange@uniecampus.it (Pietro Ducange)

ACCEPTED MANUSCRIPT

Download English Version:

https://daneshyari.com/en/article/4944364

Download Persian Version:

https://daneshyari.com/article/4944364

Daneshyari.com