International Journal of Swarm Intelligence Research
Volume 12 • Issue 1 • January-March 2021

A Comparative Analysis of Evolutionary Algorithms for Data Classification Using KEEL Tool

Amrit Pal Singh, Jaypee Institute of Information Technology, India
https://orcid.org/0000-0001-9633-8601

Chetna Gupta, Jaypee Institute of Information Technology, India

Rashpal Singh, Citi, Canada

Nandini Singh, Nucleus Software, India

ABSTRACT

Evolutionary algorithms are inspired by the biological model of evolution and natural selection and are used to solve computationally hard problems, also known as NP-hard problems. The main motive to use these algorithms is their robust and adaptive nature to provide best search techniques for complex problems. This paper presents a comparative analysis of classification of algorithm's family instead of algorithms comparison using KEEL tool. This work compares SSMA-C, DROP3PSO-C, FURIA-C, GFS-MaxLogitBoost-Cand CPSO-C algorithms. Further, these selected evolutionary algorithms are compared against two statistical classifiers using the Wilcoxon signed rank test and Friedman test on following datasets—bupa, ecoli, glass, haberman, iris, monks, vehicle, and wine—to calculate classification efficiencies of these algorithms. Experimental results reveal some differences among these algorithms. Visualization module in the model has been used to give overall results as a summary while statistical test using Clas-Wilcoxin-ST compared the algorithms in a pair-wise fashion to conclude experimental findings.

KEYWORDS

Data Classification, Evolutionary Algorithms, Friedman Test, KEEL Tool, PSO

1. INTRODUCTION

With the immense growth in the amount of data storage, there has been a focus on effective use of Data Mining (DM) techniques to identify the information and patterns in these data storages. Research has shown how data mining techniques can help in extracting relevant information (with intelligent methods) from a data set (Alcalá-Fdez et al. 2008). This discovered information should be correct, comprehensible, relevant and interesting in order to consider the whole process of mining a successful one (Kamath & Cantu-Paz 2005).

DOI: 10.4018/IJSIR.2021010102

Copyright © 2021, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Data classification is a classical DM technique which is used to find rules and properties that classify the data into disjoint groups (Romero & Ventura 2010). The goal of classification is to correctly predict a category class for each input data in the data. Researchers in past have used several conventional and evolutionary tools and techniques for data mining. Conventional techniques suffer from drawback of producing non optimal results as they tend to increase computational cost by producing large number of features than actually required. These methods worked on assumption of a-priori knowledge about the data set, which in most cases is not available. On the other hand, evolutionary algorithms produce much better optimal results because they are domain independent and can manage attribute interaction much better than the conventional data mining techniques (Alcalá-Fdez et al. 2008).

Evolutionary Algorithms are genetic processes that are naturally evolved optimized algorithms (Sharma 2015). In past few years, different evolutionary algorithms have been developed such as genetic algorithms (GA), backtracking search optimization algorithms, differential search algorithms, multi objective bat algorithm, hybrid particle swarn optimization algorithm etc. (Engelbrecht 2007). For pre-processing and post-processing the discovered knowledge genetic algorithms are used whereas for rule discovery and data pre-processing genetic programming is used (Freitas 2003). Evolutionary algorithms needs to be predesigned by programming expertise which requires considerable time, effort, knowledge and experience (Alcalá-Fdez et al. 2009).

KEEL is an open source java software tool. It is specifically used for classification, pattern mining, regression (Alcalá-Fdez et al. 2008). It makes efficient use of these Evolutionary Algorithms with a strict implementation of object oriented approval. It completely manages data and experiment's design using dataset repository and number of inbuilt evolutionary algorithms. KEEL serves several features, firstly, it consists of several evolutionary algorithms well classified into different categories (Alcalá-Fdez et al. 2008, Alcalá-Fdez et al. 2009). Secondly, it includes pre-processing techniques, post processing methods, different visualization modules and so on for various purposes. Furthermore, this tool has a great advantage that extends the range of users (Alcalá-Fdez et al. 2008). It is due to user friendly interface that provides easy to use software which requires less knowledge and experience.

In this paper we undertake comparative analysis of five evolutionary algorithms using KEEL tool namely CPSO-C, SSMA-C, FURIA-C, GFS-MaxLogitBoost-C and DROP3PSO-C on eight datasets namely, bupa, ecoli, glass, haberman, iris, monk, vehicle and wine. The analysis is based on comparison of maximum and minimum efficiencies of selected algorithms on 8 datasets. Further Clas-Wilcoxin-ST (Wilcoxon 1992) is used as statistical technique for pair-wise comparison giving out positive rank, negative rank and p-value as output.

The main contribution of this work is to compare classification algorithm's family instead of classification algorithms comparison. Further, these selected evolutionary algorithms are compared against two statistical classifiers using the Wilcoxon signed rank test and Friedman test on following datasets: bupa, ecoli, glass, haberman, iris, monks, vehicle and wine.

Section 2 discusses related studies that correspond to our work. Section 3 describes methodology used in this paper: evolutionary algorithms used in the model, datasets used, experimental setup. Section 4 illustrates functional blocks and the experimental case study with the experimental results. Finally, section 5 frames conclusions and future work.

2. RELATED STUDY

This section presents research work related to use of Evolutionary Algorithms in data classification published between years 2009 to 2019. Alcal´a-Fdez et al. 2009 introduced KEEL tool to analyze Evolutionary Algorithms that are based on variety of approaches like Pittsburgh, Michigan etc. The most important feature of this tool is that it enables researchers with little knowledge. KEEL has been pictured as a big collection of genetic fuzzy system algorithms by Alcal´a-Fdez et al. 2008. A case study has been illustrated of the experiment prepared in KEEL which made use of wilcoxon's

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/a-comparative-analysis-of-evolutionary-algorithms-for-data-classification-using-keel-tool/272896

Related Content

3D Real-Time Reconstruction Approach for Multimedia Sensor Networks
Ahmed Mostefaouiand Benoit Piranda (2012). *Intelligent and Knowledge-Based Computing for Business and Organizational Advancements (pp. 321-339).*
www.irma-international.org/chapter/real-time-reconstruction-approach-multimedia/65801

Cinco-Based Approach for Agent Petri Net Models
Amel Dembriand Mohammed Redjimi (2022). *International Journal of Organizational and Collective Intelligence (pp. 1-17).*
www.irma-international.org/article/cinco-based-approach-for-agent-petri-net-models/305841

Applications in Bioinformatics and Medical Informatics
E. Parsopoulos Konstantinosand N. Vrahatis Michael (2010). *Particle Swarm Optimization and Intelligence: Advances and Applications (pp. 204-221).*
www.irma-international.org/chapter/applications-bioinformatics-medical-informatics/40636

From Swarm Art Toward Ecosystem Art
Stefan Bornhofen, Vincent Gardeuxand Andréa Machizaud (2012). *International Journal of Swarm Intelligence Research (pp. 1-18).*
www.irma-international.org/article/swarm-art-toward-ecosystem-art/70744

MO-TRIBES for the Optimal Design of Analog Filters
Mourad Fakhfakhand Patrick Siarry (2013). *Swarm Intelligence for Electric and Electronic Engineering (pp. 40-56).*
www.irma-international.org/chapter/tribes-optimal-design-analog-filters/72822