



HAL
open science

Data Science and AI-Based Optimization in Scientific Programming

Ricardo Soto, Juan Gómez-Pulido, Stéphane Caro, José Lanza-Gutiérrez

► **To cite this version:**

Ricardo Soto, Juan Gómez-Pulido, Stéphane Caro, José Lanza-Gutiérrez. Data Science and AI-Based Optimization in Scientific Programming. Scientific Programming, 2019, pp.1-3, 2019, 10.1155/2019/7154765 . hal-01976844

HAL Id: hal-01976844

<https://hal.science/hal-01976844v1>

Submitted on 18 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Editorial

Data Science and AI-Based Optimization in Scientific Programming

Ricardo Soto ¹, **Juan A. Gómez-Pulido** ², **Stéphane Caro**,³
and **José M. Lanza-Gutiérrez** ⁴

¹*Pontificia Universidad Católica de Valparaíso, Valparaíso 2362807, Chile*

²*Escuela Politécnica, Universidad de Extremadura, Cáceres 10003, Spain*

³*CNRS, Institut de Recherche en Communications et Cybernétique de Nantes, 44321 Nantes, France*

⁴*Centro de Electrónica Industrial, Universidad Politécnica de Madrid, Madrid 28006, Spain*

Correspondence should be addressed to Ricardo Soto; ricardo.soto@pucv.cl

Received 9 October 2018; Accepted 5 December 2018; Published 9 January 2019

Copyright © 2019 Ricardo Soto et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This special issue gives the opportunity to know recent advances in the application of intelligent techniques to data-based optimization problems in scientific programming.

Artificial intelligence is today supported for different powerful data science and optimization techniques. For instance, data science commonly relies on AI algorithms to efficiently solve classification, regression, and clustering problems. This fact is particularly interesting nowadays, when big data area gathers strength supplying huge amounts of data from many heterogeneous sources. On the other hand, complex optimization problems that cannot be tackled via traditional mathematical programming techniques are commonly solved with AI-based optimization approaches such as the metaheuristics. These approaches provide optimal solutions avoiding consumption of many computational resources.

Data science and AI-based optimization have also largely been used to solve problems related to scientific programming. Various examples are reported by the literature on task assignment in distributed/parallel systems, knowledge discovery, large-scale data mining, high-performance computing, big data, distributed/parallel search, text analysis/process/classification, and optimization for manufacturing, scheduling, and civil and financial engineering, among others. In this sense, this area provides a wide set of research lines and applications that deserves to be explored.

This special issue presents nine original, high-quality articles, clearly focused on theoretical and practical aspects of the interaction between artificial intelligence and data science in scientific programming, including cutting-edge topics about optimization, machine learning, recommender systems, metaheuristics, classification, recognition, and real-world application cases.

The first article in this special issue is entitled “Optimizing the Borrowing Limit and Interest Rate in P2P System: From Borrowers’ Perspective” by Z. Li et al. This article shows a good example of how artificial intelligence algorithms can optimize some parameters involved in problems characterized by data flows. The work elaborates on the advantages of using a three-layer BP neural network algorithm to predict the borrowing limit and interest rate when individuals take advantage of P2P online service to borrow money. This approach provides a novel focus from borrowers to predict and optimize the borrowing limit and interest rate given the limited information. In addition, both parameters are optimized by means of an algorithmic proposal where the neural network and a genetic algorithm work together to solve both single-target and double-target programming optimization problems. The proposal is tested on real-world data to check its goodness as a high-accuracy prediction method.

The second article is entitled “Leveraging Image Visual Features in Content-Based Recommender System” by F. Deng et al. In the line of this special issue, data sources and

intelligent algorithms for collaborative filtering are present in this work, where the authors tackle the problem of exploring the latent information in large databases from which recommender systems provide predictions or recommendations according to the users' preferences. The knowledge area of this work is very interesting nowadays since many online systems store useful information with regard to the users' behaviour when requesting items. These systems are not only e-commerce related platforms but also movies or academic databases, for example. Although recommender systems consider mainly user-item rating data, this work combines item hybrid features based on image visual features to propose a novel recommendation model, which can be applied for rating-based recommender scenarios too. This model is particularly useful in sparse data scenarios, where it has achieved better results than other conventional approaches.

The third article is entitled "An Artificial Bee Colony Algorithm with Random Location Updating" by L. Sun et al. The artificial bee colony algorithm (ABC) is a relatively modern metaheuristic inspired in the smart process of honey collection done by bees. Such an optimization technique has largely and successfully been used for solving complex optimization problems in several application domains. In this paper, the authors propose to modify the core of the ABC algorithm in order to strengthen the exploration phase and as a consequence to improve convergence speed as well as the quality of solutions. To this end, the original perturbation function is modified integrating a random location updating, which can expand the search range of new solutions and further improve the exploration ability of the algorithm. In addition, the chaos is employed to improve the quality of solutions, while the tournament selection strategy is adopted to maintain the population diversity in the evolutionary process. Different experiments are performed on the classic sphere, Rosenbrock, Rastrigin, Ackley, and Griewank functions for validating the proposed approach.

The fourth article is entitled "Cylindricity Error Evaluation Based on an Improved Harmony Search Algorithm" by Y. Yang et al. In this article, an improved version of the harmony search metaheuristic is employed for solving an interesting problem from the manufacturing domain called cylindricity error evaluation. The cylindricity error is one of the basic form errors in mechanical parts, which greatly influences assembly accuracy and service life of relevant parts. The authors propose to solve this problem by enhancing the original harmony search algorithm. On the one hand, as in the previous approach, the chaos is introduced to improve the quality of initial solutions. On the other hand, autonomous search features are integrated with the algorithm by making adaptive the *par* and *bw* parameters. This allows to properly balancing the global and local search capabilities of the algorithm. Interesting results are illustrated on classic benchmark optimization functions involving nonparametric statistical hypothesis tests.

The fifth article is entitled "Knowledge Graph Representation via Similarity-Based Embedding" by Z. Tan et al. In this work, the authors propose an interesting similarity-based knowledge embedding model, called SimE-ER. The

knowledge graph can be seen as a typical multirelational structure, while the knowledge graph embedding is a representation method that allows one to construct low-dimensional and continuous space to describe latent semantic information and also to predict the missing facts. In this paper, a new approach is proposed to calculate the entity and relation similarities between independent and associated spaces. Here, each entity (relation) is described as two parts. The entity (relation) features in independent space are represented by the features entity (relation) intrinsically owns, and in associated spaces, the entity (relation) features are expressed by the entity (relation) features they connect. Different experiments illustrate that the proposed approach is able to outperform existing competitors in terms of time and memory-space complexity.

The sixth article is entitled "A Novel Multimean Particle Swarm Optimization Algorithm for Nonlinear Continuous Optimization: Application to Feed-Forward Neural Network Training" by M. Hacibeyoglu and M. H. Ibrahim. Artificial neural networks are a crucial technique from the artificial intelligence sphere; they have largely participated in solving several classification, prediction, optimization, and identification problems. However, the correct operation of neural networks certainly depends on proper training, which is commonly conducted via the well-known back-propagation algorithm. This algorithm determines the weights of the network by computing explicit gradients of error such as sum square error. However, using this approach generally implies to have slow convergence and falling into local minima. To tackle this concern, metaheuristics are often employed. In this paper, a modified particle swarm optimization (PSO) is proposed to multilayer feed-forward artificial neural networks training. The main modification of the classic PSO is that the employed algorithm has multiple swarms instead of a single one such as in the classic PSO. This allows one to reduce the particles going out of search space and to reinforce the local search of each particle. Interesting experimental results demonstrate that the proposed approach improves the classification accuracy of multilayer feed-forward artificial neural networks.

The seventh article is entitled "Application of the Polyhedral Conic Functions Method in the Text Classification and Comparative Analysis" by N. Uylař Satı and B. Ordin. This work is within the text categorization field, whose goal is to classify documents into predefined classes. This field of knowledge is especially interesting today because of the heavy increase in online data during the last years. In the recent literature, many traditional supervised algorithms were proposed to solve the problem, such as logistic regression, support vector machines, and naive Bayes. On this basis, besides traditional supervised techniques, the authors explore the polyhedral conic function (PFC) methods as supervised classification functions. Specifically, the authors propose to solve binary and multiclass text classification problems by applying PCFs. The performance of the proposal is analyzed by solving real-world datasets from the literature while analyzing accuracy, f-measure, and execution time. As a conclusion, the authors assert that classification algorithms based on PFCs provide

promising results in comparison with traditional supervised algorithms.

The eighth article in this special issue is entitled “Railway Subgrade Defect Automatic Recognition Method Based on Improved Faster R-CNN” by X. Xu et al. This paper is within the railway subgrade detection field, which is a serious threat to train safety. Defect recognition is a challenging task because of the variety in defect shape and size and the amount of data provided by measurement systems, such the vehicle-mounted ground penetrating radar (GPR), which is the most relevant technology today. Most works on this line focus on traditional machine-learning techniques, where feature representation fails for subgrade defects because of such a variety. Moreover, although deep-learning methods were introduced in the railway field, they were not applied to subgrade defects recognition. On this basis, the authors propose a deep-learning approach to recognize defects from the GPR profile. To this end, they propose a method for applying faster R-CNN to automatic recognition of railway subgrade defects. The experiments done in a real-world setting show that the proposal provides better performance than a traditional approach using support vector machine and histogram of oriented gradients.

The ninth and final article in this special issue is entitled “High Frequency Trading: An Application to Emerging Chilean Stock Market” by B. Crawford et al. This paper seeks to design, implement, and test a fully automatic high-frequency trading system that operates in a small market with highly concentrated ownership, as is the Chilean stock market. A system that implements high-frequency trading is presented through advanced computer tools and modelled as an NP-Complete problem. The research performs individual tests of the algorithms implemented, reviewing the theoretical net return (profitability) that can be applied on the last day, month, and semester of real market data. The use of particle swarm optimization as an optimization algorithm is shown to be an effective solution since it is able to optimize a set of disparate variables but is bounded to a specific domain, resulting in substantial improvement in the final solution.

Conflicts of Interest

The editors declare that they have no conflicts of interest.

Ricardo Soto
Juan A. Gómez-Pulido
Stéphane Caro
José M. Lanza-Gutiérrez