Abstract

In today's data-driven era, optimization in a wide range of real-world scenarios is founded on data acquired from a broad range of actual processes, such as physical experiments, computer simulations, and historical records. Also, in the age of artificial intelligence, evolutionary computations (nature- and evolution-inspired computational methods) have become increasingly popular because of their ability to provide adequate approximate solutions to complex optimization problems. Thus, data-driven evolutionary optimization (DDEO) is an emerging research field that involves the construction of data-driven models to address challenging real-world optimization problems from data. Depending on the accessibility of the data, DDEO can be broadly divided into offline and online subfields, each of which demands a distinct set of approaches. Hence, this thesis aims to develop novel offline and online data-driven models to address single, multi-, and many-objective data-driven optimization problems applicable to various real-world issues.

In this thesis, the first work introduces a fast nature-inspired ant colony algorithm (CHUI-AC) that maps the feasible solution space to a directed graph with quadratic space complexity to mine closed high-utility itemsets from transactional datasets as an offline DDEO task. It provides proof of convergence for the proposed framework and exhibits superiority in terms of execution time and convergence rate. The second work introduces novel active learning with a reliability sampling-based evolutionary framework (ALeRSa-DDEA) that utilizes an ensemble of heterogeneous radial basis function neural networks (RBFNs) to actively learn valuable insights from unlabeled data and address offline single-objective DDEOs. In addition, ALeRSa-DDEA proposes a novel approach for selecting the most reliable query individuals based on reliability sampling to update the model. Furthermore, the generic framework's effectiveness is verified by aerodynamic airfoil design optimization from offline data. The third work proposes an adaptive model selection method with a reliable individual-based model management-driven multi-objective evolutionary algorithm (AdaMoR-DDMOEA) to address offline multi-objective DDEOs. In particular, the framework selects between deep neural network (DNN) and extreme gradient boosting (XGBoost) according to their k-fold cross-validation accuracy and utilizes a reliable individual selection strategy to update the chosen surrogate model. Finally, the last work proposes a cheaper Gaussian mixture model (GMM) clustering-based surrogate-assisted evolutionary algorithm (GMAEA) that learns to classify the candidate solutions into good and bad classes to solve online many-objective DDEOs. Also, it introduces a novel infill criterion that has been explicitly created to account for classification uncertainty. In each work, several benchmark and real-world problems are solved to validate the effectiveness of the proposed models.

Keywords: Ant colony system, Nature-inspired algorithms, Closed high utility itemset, Meta-heuristic, Offline data-driven evolutionary algorithm, Active learning, Surrogate models, Radial basis function neural networks, Deep neural network, Extreme gradient boosting, Offline data-driven multi-objective evolutionary algorithm, Data-driven multi-objective optimization, Clustering, Pareto dominance, Surrogate-assisted evolutionary algorithm, Infill criterion, Expensive many-objective optimization, Engineering optimization, Surrogate modeling, Data mining, Machine learning, Computational intelligence, Artificial intelligence