Data-Driven Surrogate-Assisted Multiobjective Evolutionary Optimization of a Trauma System

In many real-life optimization problems, there is no analytic objective or constraint functions defined, instead optimization can be done based on data collected from experiments, real events or complex numerical simulations. This is often called data-driven optimization. Most of the current work in data-driven evolutionary optimization has been carried out in the area of surrogate-assisted evolutionary optimization where expensive objective and/or constraint functions are (partially) replaced with cheaper surrogate model. Data-driven optimization problems can be divided into two categories, offline and online data-driven optimization.

In offline data-driven optimization, all data that will be used is gathered before the optimization process. Algorithm can explore data from the original data during the optimization, but no new data will be added during the optimization process.

In online data-driven optimization, new data can be acquired during optimization. This includes interactive evolutionary optimization where decision maker (human) is consulted for new data. Online data-driven optimization can be further divided into two categories, one where surrogate-management strategy can actively sample new data at desired points and other where surrogate-management strategy does not have any control on the new data. These categories are called online data-driven optimization with uncontrolled incremental data and online data-driven optimization with controlled incremental data.

To reduce number of expensive function evaluations, surrogates, which usually are machine learning models such as Kriging or other neural networks, are used. To avoid errors and false optima introduced by surrogates, sometimes also original objective function is used to re-evaluate individuals. Strategy for selecting the individuals for re-evaluation is one of the key points in surrogate management strategy. Surrogate management strategy usually involves tradeoffs between accuracy and computational cost since re-evaluating large number of individuals leads to better accuracy but in the other hand it increases the number of expensive function evaluations.

In data-driven evolutionary optimization, there might be challenges related to various topics. These topics include computational cost, data quantity, data quality and heterogeneity of data. Computational cost and data quantity challenges affect surrogate building process since too large amount of data can lead to increased computational costs while building surrogates and in the other hand, too small amount makes it difficult to build accurate surrogates. Quality issues with the data may include problems like imbalanced, ill-distributed or incomplete data. Heterogeneous data may come from multiple sources in multiple formats which may introduce problems while constructing the surrogates since the sources and formats need to be combined somehow. Preprocessing the data can help with heterogeneity issues, quality issues and with certain data quantity issues where methods like clustering can be used to reduce the amount of data.