Offline Data-Driven Evolutionary Optimization Using Selective Surrogate Ensembles

Offline data-driven evolutionary optimization means evolutionary algorithm where historical data is used to build surrogate models that can be used to measure individuals' fitness instead of expensive objective and/or constraint functions. Contrary to online data-driven evolutionary optimization, no new data can be added during optimization.

Traditionally, in offline data-driven optimization, surrogate models are built using only the given historical data. This creates an issue where algorithm can only search limited decision space during the optimization process. To make better use of the offline data, ensemble learning based surrogate management strategy can be used. This means building a large number of surrogates before the optimization and using adaptive method to select some of them during the optimization.

Ensemble learning means machine learning method where multiple base learners are first constructed and then combined to create strong learner. Bagging (which is also known as bootstrap aggregating) and boosting are two most popular methods for generating ensemble. Algorithm described in the paper uses bagging, more precisely selective bagging which means bagging algorithm with model selection strategy.

In selective bagging first phase is the model generation which contains the bootstrap sampling and model training. Bootstrap sampling means creating number of subsets from original data by picking random points from the original data. After the subsets are created, each subset is used to train single model.

Next step in selective bagging is model combination. Model combination contains model selection and averaging phases. Model selection means selecting only some of the models are that are going to be used for producing the ensemble output. Different strategies can be used for model selection phase, and it is an important parameter of bagging algorithm.

Current model selection algorithms can be divided into two categories, local- and global-search based algorithms. Global search algorithms use sparse optimization, genetic algorithms and clustering. On the other hand, local search algorithms are greedy and either successively add models to an empty set or successively delete models from full set of models.

After models have been selected, their values are used to build the ensemble output. In algorithm proposed by the paper, ensemble output is built by averaging the outputs of selected models.