Model Selection (MS) is an important aspect of machine learning, as necessitated by the No Free Lunch theorem for optimization. Given an ensemble of models, the task of MS is to identify a subset of models that are optimal in terms of pre-selected optimization criteria. There are many practical applications of MS, such as model parameter tuning, personalized recommendations, A/B testing, etc. Lately, some MS research has focused on trading off exactness of the optimization with somewhat alleviating the computational burden entailed. Recent attempts along this line include metaheuristics optimization, local search-based approaches, sequential model-based methods, portfolio algorithm approaches, and multi-armed bandits.

Racing Algorithms (RAs) are an active research area in MS, which trade off some computational cost for a reduced, but acceptable likelihood that the models returned are indeed optimal among the given ensemble of models. All existing RAs in the literature are designed as Single-Objective Racing Algorithm (SORA) for Single-Objective Model Selection (SOMS), where only one optimization criterion is considered for measuring the goodness of models. Moreover, they are offline algorithms, in which MS occurs before model deployment and the selected models are optimal in terms of their overall average performance on a validation set of problem instances.

This work aims at investigating racing approaches along two distinct directions: Extreme Model Selection (EMS) and Multi-Objective Model Selection (MOMS).

In EMS, given a problem instance and a limited computational budget shared among all the candidate models, one is interested in maximizing the final solution quality. In such a setting, MS occurs during model comparison in terms of maximum performance and involves no model validation. EMS is a natural framework for many applications. However, EMS problems remain unaddressed by current racing approaches. In this work, the first RA for EMS, named Max-Race, was developed, so that it optimizes the extreme solution quality by automatically allocating the computational resources among an ensemble of problem solvers for a given problem instance. In Max-Race, a parametric hypothesis test based on the Generalized Pareto Distribution (GPD) assumption is developed to identify significant difference between the extreme performances of two models. Experimental results have confirmed that Max-Race is able to identify the best extreme model with high accuracy and low computational overhead.

Furthermore in machine learning, as well as in many real-world applications, a variety of MS problems are multi-objective in nature. MS, which simultaneously considers multiple optimization criteria, is referred to as MOMS. Under this scheme, a set of Pareto optimal models is sought that reflect a variety of compromises between optimization objectives. So far, MOMS problems have received little attention in the relevant literature. Therefore, this work also developed the first Multi-Objective Racing Algorithm (MORA) for a fixed-budget setting, namely S-Race. S-Race addresses MOMS in the proper sense of Pareto optimality. Its key decision mechanism is the non-parametric sign test, which is employed for inferring pairwise dominance relationships. Moreover, the overall probability of falsely eliminating any non-dominated models is strictly controlled at a user specified level. Additionally, SPRINT-Race, the first MORA for a fixed-confidence setting, was also developed. In SPRINT-Race, pairwise dominance and non-dominance relationships are established via the Sequential Probability Ratio Test with an Indifference zone. Moreover, the overall probability of falsely eliminating any non-dominated models or mistakenly retaining any dominated models is controlled at a prescribed significance level. Extensive experimental analysis has demonstrated the efficiency and advantages of both S-Race and SPRINT-Race in MOMS.