Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

ADAPTIVE RANKING AND SELECTION BASED GENETIC ALGORITHMS FOR DATA-DRIVEN PROBLEMS

Kimia Vahdat

Edward P. Fitts Department of Industrial and Systems Engineering
NC State University
915 Partners Way
Raleigh, NC 27606, USA

ABSTRACT

We present ARGA, the Adaptive Robust Genetic Algorithm, for optimizing simulation problems with binary variables affected by input uncertainty and Monte Carlo noise. In this method, a population evolves as more information about the high-dimensional, stochastic problem becomes available. ARGA conducts ranking and selection with a debiasing mechanism of fitness values using fast iterated bootstraps economized with control variates. Debiasing reduces the model risk due to input uncertainty bias, leading to a more accurate ranking of designs. Given the double loop of function evaluations, we incorporate adaptive budget allocation throughout the search only if the current population's proximity to optimality signals the need for a smaller standard error. In that case, we allocate replications to the input model of the design most responsible for risk. Empirical results with a fixed optimization budget show that ARGA obtains significantly better solutions in feature selection problems across various datasets.

1 INTRODUCTION

Simulation models serve as vital tools for evaluating, comparing, and optimizing system designs based on estimated performance. For efficient comparison, ranking and selection (R&S) methods are employed, distributing simulation efforts across designs to attain predetermined stochastic confidence levels. In this paper, we employ an optimal computing budget allocation (OCBA) during optimization that, though reminiscing the fixed-budget procedures in R&S for statistical guarantees, can stop before reaching that maximum budget guided by an adaptive sampling philosophy. A salient feature of our proposed method is its goal of performing R&S robustly, i.e., with consideration for model risk. Inaccurately estimating unknown input distributions from historical or input data can introduce input uncertainty (IU), a key contributor to model risk that significantly impacts output inferences and decisions. This complexity renders correct system identification unattainable even with infinite computation. To address these challenges, we combine the genetic algorithm (GA), a global binary optimization engine, with R&S techniques, addressing simulation issues affected by both stochastic uncertainty (SU) and IU.

We first and foremost hedge against IU by computing and reducing the estimation bias, when using empirical CDFs for input distributions, from the estimated quantities. The proposed *debaising* procedure comes at an increased computational cost. The adaptive approach strategically allocates resources considering SU and IU nuances, maximizing efficiency by determining optimal design-input model pairs for budget allocation within each GA iteration. The inner dynamics of adaptive budget allocation in each iteration of GA, nonparametric bias estimation, and optimization lead to a new stochastic GA that is more effective, i.e., more robust and efficient, for decision-making in noisy simulation environments.

We call this method Adaptive Robust GA (ARGA), which has the goal of improving the effectiveness of GA for optimization in a high-dimensional binary search space. ARGA capitalizes on three key elements to enhance its performance within the GA framework: (i) the iterative design generation and selection operations within GA, (ii) a variance-reduced fast-iterated bootstrapping (FIB) technique to reduce the risk caused by IU bias in the estimators, and (iii) an adaptive sampling scheme that increases the budget only *when* and *where* necessary. Notably, this work pioneers the integration of IU and adaptive sampling within an optimization framework. This work is a continuation of our previous robust estimation work (Vahdat and Shashaani 2021; Shashaani and Vahdat 2022) to handle the stochasticity of data-driven problems. The application of this method is also on machine learning (ML), where the ML model can be viewed as a black-box simulation. A deviation of the debiasing with nonparametric input models established in (Vahdat and Shashaani 2023) can reduce the risk when building ML models.

2 THE ADAPTIVE ROBUST GA

ARGA can be summarized in three main components:

1. **Robust R&S within GA:** To identify in each population and across input models, which case should receive additional budget, we leverage the robust OCBA (R-OCBA) (Gao et al. 2017), which suggests a way to link the computation and utilization of IU with budget allocation.
2. **Variance-reduced debiasing:** Inside the R&S procedure, we devise a debiasing procedure applying an FIB technique with control variates to efficiently calculate the induced bias during estimation given a fixed budget.
3. **Adaptive sampling strategy:** We then use a novel adaptive sampling rule that examines the proximity of the current population to optimality and informs whether the debiased estimated values in the current iteration require more precision. If that is the case, we allocate more budget to a design with an input model that we expect is more likely to strike a balance between the statistical error and optimality gap. We repeat this inspection until obtaining sufficient precision or exhausting the total per-iteration per-design budget.

3 CONCLUSION

Introducing ARGA, a novel genetic algorithm variant, enhances solution selection accuracy and decreases computational costs in stochastic optimization. ARGA offers substantial potential for cost reduction while maintaining or even surpassing solution quality. The adaptive approach ensures prudent allocation of resources, targeting crucial near-optimal solutions with considerable estimation errors. R-OCBA incorporation produces dual benefits. It secures selection probability during R-OCBA activation, guiding optimization effectively within resource limits. Moreover, valid design comparisons amplify exploration efficiency within the feasible region. Our empirical results demonstrate enhanced effectiveness in ARGA applied to a particular data-driven optimization problem, namely, feature selection, that searches for the most informative features in a large dataset.

REFERENCES

Gao, S., H. Xiao, E. Zhou, and W. Chen. 2017. “Robust Ranking and Selection with Optimal Computing Budget Allocation”. *Automatica* 81:30 – 36.

Shashaani, S., and K. Vahdat. 2022. “Improved Feature Selection with Simulation Optimization”. *Optimization and Engineering*:1573–2924.

Vahdat, K., and S. Shashaani. 2021. “Non-parametric Uncertainty Bias and Variance Estimation via Nested Bootstrapping and Influence Functions”. In *Proceedings of the 2021 Winter Simulation Conference*, edited by S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, 1–12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Vahdat, K., and S. Shashaani. 2023. “Robust Prediction Error Estimation with Monte Carlo Methodology”. *arXiv preprint arXiv:2207.13612*.