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

ENHANCING PARALLEL LARGE-SCALE RANKING AND SELECTION USING CLUSTERING TECHNIQUES

Zishi Zhang

Guanghua School of Management
Peking University
5 Yiheyuan Road
Beijing 100871, CHINA

ABSTRACT

We explore the use of correlation-based clustering techniques to enhance large-scale R&S procedures under parallel computing environment. Both theoretical analysis and numerical experiments convincingly demonstrate that clustering techniques can significantly improve the sample efficiency of existing R&S methods.

1 CLUSTERING AND CONQUER

Ranking and selection (R&S) aims to find the alternative with the largest mean performance among a finite set of alternatives indexed by $\mathscr{P} = \{1,2,\ldots,p\}$ through simulation experiments. The large-scale R&S problem in parallel computing environments has emerged as an important research topic in recent years, where “large-scale" refers to a large number of alternatives. Existing procedures are divided into two branches: fully sequential procedures and stage-wise procedures. The former are sample-efficient but suffer from heavy communication and synchronization costs for parallel implementations. The latter is more suitable for parallel environments but requires a large amount of sample size.

To break this dilemma, we explore the utilization of correlation information among alternatives, rather than relying solely on mean information as in traditional approaches. We model the behaviors of alternatives as correlated random variables: $(X_1,X_2,\ldots,X_p) \sim N(\boldsymbol{\mu},\Sigma_{p\times p})$. Structurally similar alternatives are govern by several common latent factors and thereby exhibit high correlations. Let x_{ij} denote the jth observation of alternative i, and the jth observation vector $(x_{1j},x_{2j},\ldots,x_{pj}) \sim N(\boldsymbol{\mu},\ \Sigma_{p\times p})$ independent and identically. We assume $cov(x_{im},x_{jn}) = 0$ if $m \neq n$.

Some parallel R&S procedures, such as Zhong and Hong (2022) and Ni et al. (2017), use "divide and conquer" strategies. As shown in Figure 1(a), those strategies involve randomly distributing alternatives across parallel cores (or processors), with simulations conducted independently on each core. After each core decides a local best, the global best is selected in the final round of simulation. We improve upon the “divide and conquer" scheme and develop a parallel “clustering and conquer" procedure for large-scale R&S. As shown in Figure 1(b), in this new approach, we cluster together alternatives with strong correlations and allocate them to the same core, rather than random allocation.

We test the enhancements achieved through correlation-based clustering across several established methods, including KT procedure (Zhong and Hong 2022), Rinott’s procedure, EA (equal allocation) and OCBA (the versions incorporating clustering are denoted as Procedure-CC). Figure 2 shows the required sample size to achieve a PCS of 0.9 for different numbers of alternatives, p. The utilization of clustering techniques significantly reduces the required sample size.

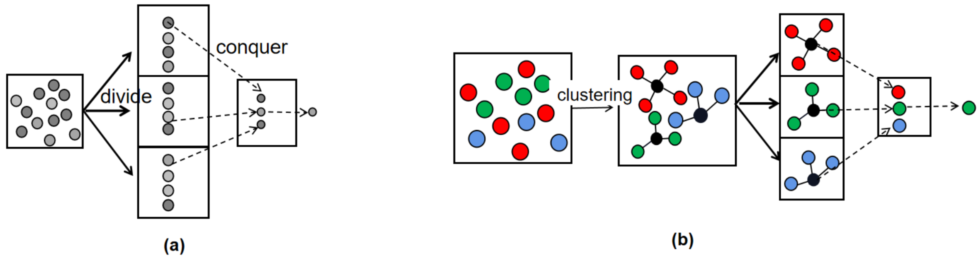

Figure 1: (a) “divide and conquer" scheme (b) “clustering and conquer" scheme.

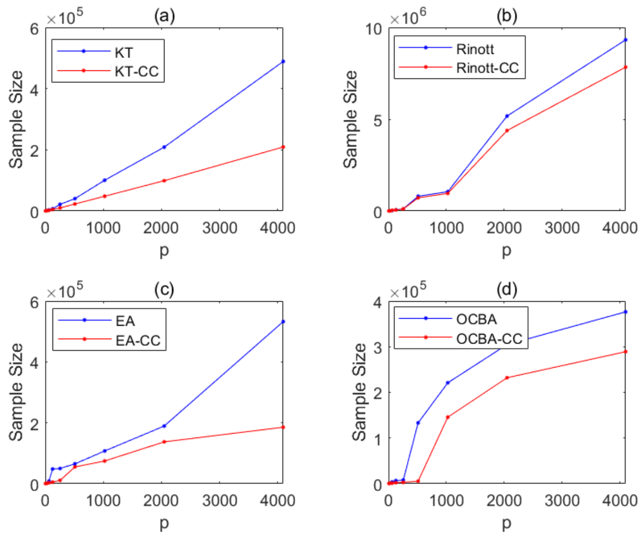

Figure 2: The required sample sizes to guarantee $PCS > 0.9$.

2 THE VALUE OF CORRELATION INFORMATION

In a future journal paper, we will present a comprehensive theoretical analysis on how the correlation-based clustering technique enhances the sample efficiency of R&S and quantifies the extent of improvement achieved. Here, due to space constraints in this poster, we provide a simplified version of one key corollary, omitting its mathematical assumptions: if we increase the correlation coefficients between alternative [1] and the other alternatives by Δr, the PCS is increased by:

$$PCS' - PCS = \sum_{i \in \mathscr{P} \setminus \{[1]\}} |I_i| \cdot \Delta r + o\Big(\sum_{i \in \mathscr{P} \setminus \{[1]\}} |I_i| \cdot \Delta r\Big),$$

where $|I_i| \sim O(\sqrt{N} e^{-A_i N})$, $A_i > 0$ and N is the total sample size. This corollary reveals the value of correlation information. Increasing correlations can enhance the probability of correct selection (PCS), which in turn leads to a significant reduction in the required sample size to achieve a certain level of PCS.

REFERENCES

Ni, E. C., D. F. Ciocan, S. G. Henderson, and S. R. Hunter. 2017. “Efficient ranking and selection in parallel computing environments”. *Operations Research* 65(3):821–836.

Zhong, Y., and L. J. Hong. 2022. “Knockout-tournament procedures for large-scale ranking and selection in parallel computing environments”. *Operations Research* 70(1):432–453.