Automatica 84 (2017) 117–127

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Simulation budget allocation for simultaneously selecting the best and worst subsets☆

Hui Xiao [a], Siyang Gao [b], Loo Hay Lee [c]

[a] Department of Management Science, School of Statistics, Southwestern University of Finance and Economics, Chengdu, China
[b] Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong, China
[c] Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore

ARTICLE INFO

Article history:
Received 23 June 2016
Received in revised form 19 January 2017
Accepted 27 May 2017

Keywords:
Simulation
Ranking and selection
Computing budget allocation
Optimization
Discrete event dynamic systems

ABSTRACT

Motivated by the practical needs in simulation optimization, this paper considers the problem of selecting the best m and worst n designs from a total of k alternatives based on their mean performance values, which are unknown and can only be estimated via simulation. In order to improve the efficiency of simulation, this research characterizes an asymptotically optimal allocation of simulation replications among the k designs such that the probability of correctly selecting the best m and worst n designs can be maximized, and develops a corresponding selection procedure for implementation purpose. The efficiency of the proposed procedure is demonstrated via numerical experiments.

1. Introduction

Discrete event dynamic systems (DEDS) are systems in which state evolution depends entirely on the occurrence of discrete events. They widely exist in manufacturing, communication, and service systems (Cassandras & Lafortune, 2008). The control decisions for these systems are only made when certain events happen. For example, actions are taken only when a new customer arrives at a queuing network in admission control. In material handling systems, the order actions are triggered only when the inventory level falls below the pre-specified threshold. A lot of work has been done to study the optimal control of the DEDS such as queuing systems and supply chain systems (Xia & Jia, 2015). However, with the rapid advance of technology, the system complexity increases significantly. DEDS in real industry rarely satisfy the assumptions of analytical models. As a result, the discrete event simulation has become a widely-used tool for evaluating the performance of stochastic systems since any level of details of the system can be modeled via simulation (Law, 2005). However, running the simulation model is usually economically expensive and time-consuming. In addition, a large number of simulation replications are typically needed in order for the sample mean to be a statistically significant estimation due to the slow convergence rate of it (Chen & Lee, 2010). As a result, the efficiency of simulation remains a big concern, especially when the number of alternatives for comparison is large or the unit cost of simulation is expensive. To improve the efficiency, ordinal optimization (OO) has emerged as an efficient technique for discrete event simulation (Ho, Sreenivas, & Vakili, 1992), and it has been successfully applied to several problems related to DEDS (Cassandras, Dai, & Panayiotou, 1998; Hsieh, Chen, & Chang, 2007; Patsis, Chen, & Larson, 1997). The efficiency of OO was further improved by Chen, Lin, Yücesan, and Chick (2000) using the optimal computing budget allocation (OCBA) framework. The problem considered by OO and OCBA falls under the well-established branch of statistics known as ranking and selection (R&S). In recent years, substantial research efforts have been devoted to studying R&S procedures with applications to simulation (Xu, Huang, Chen, & Lee, 2015).

There exist three main approaches to study the R&S problems in simulation when the number of competing designs is relatively small, i.e., the indifference-zone (IZ) formulation, the value of information procedure (VIP) and the OCBA framework. The IZ approach, which assumes that the mean performance of the best design is at least δ better than that of other designs, aims to find a feasible way such that the pre-specified probability of correct selection can be guaranteed (Dudewicz & Dalal, 1975; Kim & Nelson, 2001; Rinott, 1978). The VIP describes the evidence of correct selection using the

☆ This work has been supported in part by National Science Foundation of China (Grants 71601158 and 61603321) Hong Kong Research Grants Council (Grants 21204316 and 11214317), and City University of Hong Kong (Grants 7200419 and 7004464). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Bart De Schutter under the direction of Editor Christos G. Cassandras.

E-mail addresses: msxh@swufe.edu.cn (H. Xiao), siyangao@cityu.edu.hk (S. Gao), iseleelh@nus.edu.sg (L.H. Lee).

http://dx.doi.org/10.1016/j.automatica.2017.07.006
0005-1098/

Bayesian posterior distribution. The simulation budget allocation is determined via maximizing the value of information based on decision theory tools (Chick & Inoue, 2001). The OCBA approach allocates the simulation budget intelligently based on mean and variance of all competing designs in order to maximize the probability of correction (Chen & Lee, 2010; Chen et al., 2000; Jia, Zhou, & Chen, 2013). A comprehensive comparison of the performance among the above-mentioned R&S procedures was conducted by Branke, Chick, and Schmidt (2007) in which it was concluded that no R&S procedures can dominate in all scenarios.

When the number of alternatives for comparison is relatively large, the required simulation budget increases rapidly if the above-mentioned R&S procedures are used. The large-scale R&S problem has been addressed using different approaches in the literature. For example, Nelson et al. (2001) and Alrefaei and Alawneh (2004) proposed two-stage procedures based on the IZ formulation, where the noncompetitive designs were screened out in the first stage to avoid sampling for these designs during the second stage. The procedure was extended to be fully sequential by Frazier (2014). Besides, ordinal optimization (OO), which concentrates on finding a good enough design within a significantly reduced computing budget, is another popular method to address the large-scale R&S problems (Ho et al., 1992). Recently, OO has also incorporated with the IZ procedure and OCBA approach for selecting the best design and best subset for large-scale problems (Almomani & Alrefaei, 2016; Al-Salem, Almomani, Alrefaei, & Diabat, 2017). Furthermore, with the rapid development of technology, multi-core personal computers and many-core servers become easily accessible for ordinary users. Thus, the large-scale R&S problem is studied using parallel computing in recent years (Luo, Hong, Nelson, & Wu, 2015; Ni et al., 2013). Lastly, Metamodels such as kriging and neural network were built to approximate the objective values (Barton & Mechesheimer, 2006; Chen, Ackerman, & Nelson, 2013). Random search algorithms such as the simulated annealing (e.g., Alrefaei & Andradóttir, 1999), stochastic comparison (e.g., Gong, Ho, & Zhai, 2000), and nested partitions (Shi & Ólafsson, 2000) were applied in the simulation setting to find the optimal design.

This research adopts the OCBA approach to address a variant of small-scale R&S problems. In the literature, OCBA approach was extended in various ways to determine the simulation budget allocation in different contexts. For example, Chen, He, Fu, and Lee (2008) proposed a budget allocation rule for selecting the top m designs. The rule was further improved by Zhang, Lee, Chew, Xu, and Chen (2015) using a tighter bound on the probability of correct selection. Lee et al. (2010a, b) investigated the OCBA for multi-objective (MOCBA) optimization problems. The MOCBA procedure was suggested to select the optimal Pareto set when each design has multiple performance measures. Another variant of OCBA is to use the opportunity cost as the selection criterion instead of the correct selection probability (Gao & Chen, 2015; Gao & Shi, 2015; He, Chick, & Chen, 2007). Other variants of OCBA include selecting the best design subject to constraints (Lee, Pujowidianto, Li, Chen, & Yap, 2012), selecting the best design subject to stochastic time (Jia, 2013), selecting the best design considering resource sharing and allocation (Peng, Chen, Fu, & Hu, 2013), and complete ranking (Xiao, Lee, & Ng, 2014). These works are either selecting the single best or selecting the best subset from a finite number of alternatives. Motivated by the R&S problem in practice, Zhang, Zhang, Wang, and Zhou (2016) investigated the simulation budget allocation of selecting the best and worst designs simultaneously. The derived budget allocation rule can be applied to best-worst scaling, a discrete choice model that requires picking both the highest and lowest utility items at the same time (Flynn, Louviere, Peters, & Coast, 2007).

In practice, decision makers usually prefer to have several alternatives provided by computer models instead of trusting models unconditionally since models are only the abstraction of real systems. Final decisions might be made by considering other qualitative criteria or political feasibilities that are not incorporated in the computer models (Zhang et al., 2015). Thus, returning the single best alternative by computer models is sometimes insufficient for decision makers. Therefore, it is practically important to consider the problem of selecting the best subset and worst subset. Obtaining a subset of choices provides decision makers a more flexible and people-oriented way to support decision making.

This research aims to generalize the problem of selecting the single best and worst to that of selecting the best and worst subsets. The proposed model can be applied to best-worst scaling problems that having multiple performance measures. For example, selecting a subset of best items and a subset of worst items when applying best-worst scaling in health is more useful than providing only the best and worst items since there are multiple performance measures for health (Lancsar, Louviere, Donaldson, Currie, & Burgess, 2013). Certain items need to be compared and removed by considering other qualitative criteria that are not incorporated in the model. Similarly, the proposed procedure of this paper can be integrated with the best-worst scaling model to perform selection of choices in a variety of areas such as business ethics (Auger, Devinney, & Louviere, 2007), psychology (Lee, Soutar, & Louviere, 2007), personality assessment (Lee, Soutar, & Louviere, 2008), and food science (Jaeger, Jørgensen, Aaslyng, & Bredie, 2008) because the behaviors of choice in real life are usually affect by randomness and noise.

Development of efficient simulation procedures for selecting the best m and worst n designs is also beneficial to global simulation optimization that requires rewarding the elite subset and penalizing the worst subset in each iteration of the algorithm. For example, one of the selection schemes in genetic algorithm (GA) is to select the best individuals for retainment and worst ones for replacement (Goldberg & Deb, 1991). The entire elite subset and the worst subset are used to update the subsequent population that drives the search of additional candidate solutions. A poor selection of the best subset and worst subset leads the search in a possibly misleading direction. Similarly, one variant of the particle swarm optimization (PSO) requires replacing the worst particles by the elite particles (Fourie & Groenwold, 2002; Shi & Eberhart, 1998). The overall efficiency of the GA and PSO in simulation setting depends on how efficiently we simulate the candidate solutions and identify the elite subset and the worst subset.

In view of the above-mentioned applications, this paper develops a new procedure of selecting the best and worst subsets simultaneously. The objective of this research is to determine how the simulation replications should be allocated such that the probability of correct selection can be maximized. We use the large deviation theory to derive the asymptotically optimal simulation budget allocation rule for general underlying distributions and an easily implementable asymptotically optimal allocation rule in the case of normal underlying distributions. The rest of the paper is organized as follows. Section 2 formulates the problem. Sections 3 and 4 derive the asymptotically optimal allocation rule for general underlying distributions and normal underlying distributions respectively. A sequential simulation procedure is developed in Section 6. Numerical experiments are conducted to test the efficiency of the proposed simulation procedure in Section 6. Finally, this paper is concluded in Section 7.

2. Problem formulation

Consider the problem of selecting the best m and worst n designs simultaneously from k alternatives, where $m + n < k$ and k is finite and relatively small. The unknown mean performance $\mu_i \in R, \forall i \in \{1, \dots, k\}$ is used as the selection criterion, and