European Journal of Operational Research 265 (2018) 1094–1101

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Innovative Applications of O.R.

An efficient simulation optimization method for the generalized redundancy allocation problem

Kuo-Hao Chang*, Po-Yi Kuo

Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 17 November 2016
Accepted 25 August 2017
Available online 1 September 2017

Keywords:
Reliability
Generalized redundancy allocation problem
Simulation optimization
Importance sampling

ABSTRACT

The redundancy allocation problem (RAP) is concerned with the allocation of redundancy that maximizes the system reliability subject to constraints on system cost, or minimizes the system cost subject to constraints on the system reliability, has been an active research area in recent decades. In this paper, we consider the generalized redundancy allocation problem (GRAP), which extends traditional RAP to a more realistic situation where the system under consideration has a generalized (typically complex) network structure; for example, the components are connected with each other neither in series nor in parallel but in some logical relationship. Special attention is given to the case when the objective function, e.g., the system reliability, is not analytically available but has to be estimated through simulation. We propose a partitioning-based simulation optimization method to solve GRAP. Due to several specially-designed mechanisms, the proposed method is able to solve GRAP both effectively and efficiently. For efficacy, we prove that the proposed method can converge to the truly optimal solution with probability one (w.p.1). For efficiency, an extensive numerical experiment shows that the proposed method can find the optimal or nearly optimal solution of GRAP under a reasonable computational budget and outperforms the other existing methods on the created scenarios.

1. Introduction

Reliability optimization (RO) has been an active research area attracting a great deal of attention in recent decades due to its extensive real-world applications such as power systems, electronic systems, telecommunication systems and manufacturing systems. One excellent overview on system reliability optimization is provided in (Kuo, Parasad, 2000. The redundancy allocation problem (RAP), an important problem in the area of reliability optimization, corresponds to seeking the redundancy allocation that either maximizes the system reliability subject to constraints on system cost or minimizes the system cost subject to constraints on the system reliability. RAP is especially important during the crucial initial stages of designing a new system. Unfortunately, RAP is fundamentally a NP-hard combinatorial optimization problem (Chern, 1992), and the computational requirement grows exponentially as the number of nodes and links in the network increases.

As such, many heuristics have been developed for solving RAP efficiently in the literature, e.g., Coit and Smith (1996a), Dai, Xie, and Wang (2007) and Zou, Gao, Wua, Li, and Li (2010). In addition, Coit and Smith (1996b) proposed an approach that combines neural networks and the genetic algorithm to solve RAP, where the former serves as a tool for estimating the system reliability and the latter is applied to search for the components of a series-parallel system that has minimum cost, subject to a constraint on the minimum system reliability. Ouzineba, Nourelfatha, and Gendreau (2008) also proposed a tabu search heuristic to determine the minimal cost system configuration under availability constraints for multi-state series–parallel systems. Sahooa, Bhuniab, and Kapurc (2012) proposed a multi-objective reliability optimization problem in an interval environment and investigated the performance of the proposed techniques through sensitivity analyses. Tekiner-Mogulkoc and Coit (2011) further considered the problem where the reliability of most components are not known with certainty and proposed algorithms to minimize the coefficient of variation of the system reliability estimate. Zia and Coit (2010) proposed an optimization method based on column generation decomposition to maximize system reliability.

In a typical RAP, the problem is often assumed to have a series–parallel or k-out-of-n type network structure, making the objective function analytically available. In practice, however, many redundancy systems, such as telecommunication systems, can have a complex network topology whose components are connected with each other neither in series nor in parallel, but in some logical

* Corresponding author.
E-mail addresses: chang@mx.nthu.edu.tw, birdhow@gmail.com (K.-H. Chang).

http://dx.doi.org/10.1016/j.ejor.2017.08.049
0377-2217/

relationship. In this paper, we consider the generalized redundancy allocation problem (GRAP), which extends the traditional RAP to a more realistic situation where the system under consideration has a generalized (typically complex) network structure and thus the objective function, e.g., the system reliability, cannot be derived analytically but has to be estimated through simulation.

Zhao and Liu (2003) proposed an approach that combines stochastic simulation, the neural networks and the genetic algorithm to solve both parallel and standby redundancy optimization problems. Yeh, Lin, Chung, and Chih (2010) proposed a Monte-Carlo-simulation (MCS)-based particle swarm optimization to solve complex network reliability optimization problems. One drawback on the use of simulation is that it may require extensive (sometimes even unaffordable) computational time to produce statistically valid results, especially when the feasible region is very large. Moreover, the existing simulation-based methods are essentially heuristics that do not provide any convergence guarantee; that is, the optimal solution found by the algorithms is not necessarily the truly optimal solution of the problem.

In this research, we propose a simulation optimization (SO) method, called partitioning-based simulation optimization method for reliability optimization (PSORO), to solve GRAP. SO is a rapidly expanding field over the past decade due to its ability to solve many practical problems. It represents a class of systematic methodologies that can solve problems where the objective function is too complex to be analytical, often due to profound randomness and/or the complicated interaction between randomness and decision variables, and thus has to be estimated through stochastic simulation. One excellent survey about methods and applications of SO is given in Amaran, Sahinidis, Sharda, and Bury (2016).

In the literature, some efficient simulation optimization methods have been developed; for example, Shi and Ólafsson (2000) proposed a new randomized method, called the Nested Partitions (NP) method, for solving global discrete optimization problems. Chang, Hong, and Wan (2013) proposed a new Response-Surface-Method (RSM)-based framework, called Stochastic Trust-Region Response-Surface Method (STRONG) to solve continuous simulation optimization problems. Chang (2015) provided an improved version of STRONG that has better computational efficiency and can handle generally-distributed response variables. Chang, Li, and Wan (2014) further combined STRONG with efficient screening designs for handling large-scaled SO problems. Some metaheuristics have also been developed, including the genetic algorithm, tabu search, the Nelder–Mead simplex method (NM) and scatter search etc (Spall, 2003). Chang (2012) proposed a convergent NM-based framework, called the Stochastic Nelder–Mead simplex method (SNM), to handle gradient-free SO problems.

PSORO is a partitioning-based SO method specially designed for solving the GRAP. Its framework is outlined as follows. To begin with, PSORO partitions the feasible region into several subregions, samples a few solutions in each subregion, evaluates the promising index of each subregion, and identifies the most promising region. In later iterations, the whole process is repeated except that the feasible region is replaced by the most promising region. Throughout the whole process, the optimal solution is estimated with the "best" solution of all sampled solutions accumulated through the current iteration. In fact, the partitioning structure of PSORO allows the algorithm to identify and approach the region where the truly optimal solution is located, thereby reducing the search time and computations needed for locating the truly optimal solution. The concept of partitioning is not new and has been used in some methods; for example, branch-and-bound methods (Horst & Tuy, 2003), nested partitions (NP) method (Shi, Meyer, Bozbay, & Miller, 2004; Shi & Ólafsson, 2000), adaptive partitioned random search (APRS) (Tang, 1994), and adaptive global and local search (AGLS) (Chang, 2016). However, some specially-developed mechanisms are incorporated in PSORO to enable it to solve the GRAP both effectively and efficiently. We prove that PSORO can converge to the truly optimal solution with probability one (w.p.1), and, moreover, find the optimal or nearly optimal solution under a reasonable computational budget. Also, as will be shown later, PSORO is proved to significantly outperform the other existing methods in terms of efficiency and efficacy on selected test problems and is therefore worth further investigation. More details will be presented in later sections.

Table 1
Notations.

$R(t;\mathbf{x})$	System reliability at time t, depending on $\mathbf{x}$
$\mathbf{x}$	$(x_{11},\ldots,x_{1,m_1},x_{21},\ldots,x_{2,m_2},\ldots,x_{s,m_s})$
x_{ij}	Quantity of the jth available component used in subsystem i
m_i	Number of available components for subsystem i
s	Number of subsystems
C, W	System-level constraint limits for cost and weight
c_{ij}, w_{ij}	Cost and weight of the jth available component of subsystem i
t_0	Mission time (fixed)
λ_{ij}	Parameter for exponential distribution, $f_{ij}(t)=\lambda_{ij}\exp(-\lambda_{ij}t)$
n_i	Total number of components used in subsystem i, $n_i=\sum_{j=1}^{m_i}x_{ij}$
$n_{max,i}$	Upper bound for n_i ($n_i \le n_{max,i}$ $\forall i$)
N_k	The number of replications of each sampled solution in the kth iteration
$n(k)$	The number of solutions sampled by NPRO at iteration k

The remainder of this article is organized as follows. Section 2 describes the mathematical model of GRAP. Section 3 presents the IS-based MCS, followed by Section 4 where PSORO is introduced. Section 5 presents the results of numerical experiment and evaluates the efficacy and efficiency of PSORO. Section 6 concludes with general remarks and proposes directions for future research.

2. Generalized redundancy allocation problem

The optimal solution(s) to GRAP is the optimal redundancy allocation that maximizes the system reliability, subject to overall restrictions on the system cost, C, the system weight, W, and the maximum number of components for each subsystem, $n_{max,i}$. The notations are defined in Table 1 and the mathematical model of GRAP is given as follows:

$$\begin{aligned} &\max R(t_0;\mathbf{x}) \\ \text{subject to } &\sum_{i=1}^{s}\sum_{j=1}^{m_i} c_{ij}x_{ij} \le C, \\ &\sum_{i=1}^{s}\sum_{j=1}^{m_i} w_{ij}x_{ij} \le W, \\ &\sum_{j=1}^{m_i} x_{ij} \le n_{max,i}, \text{ for } i=1,\ldots,s, \\ &x_{ij}\in\{0,1,2,\ldots\}. \end{aligned} \tag{1}$$

GRAP assumes no particular structure in the network, i.e., the network structure is not restricted to be the series–parallel type. To generalize the methodology, we assume the system reliability is not analytically available, typically due to the complexity of the network structure, and instead has to be estimated by simulation. Unfortunately, the estimates provided by simulation inevitably contain noise, which poses challenges to evaluating whether one redundancy allocation is superior to another when comparing different choices of redundancy allocation. Clearly, the identification of the optimal solution to GRAP is even more challenging.

One way to address this issue is to increase the sample size, i.e., the number of replications, of each redundancy allocation to