Computers & Industrial Engineering 81 (2015) 78–89

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Robust data envelopment analysis approaches for evaluating algorithmic performance ☆

Chung-Cheng Lu *

Department of Industrial Engineering and Management, National Taipei University of Technology, 1, Section 3, Chung-Hsiao East Road, Taipei 10608, Taiwan

A R T I C L E I N F O

Article history:
Received 22 January 2013
Received in revised form 16 June 2014
Accepted 24 December 2014
Available online 2 January 2015

Keywords:
Meta-heuristics
Uncertainty modeling
Data envelopment analysis
Robust counterpart optimization

A B S T R A C T

Recent advances in state-of-the-art meta-heuristics feature the incorporation of probabilistic operators aiming to diversify search directions or to escape from being trapped in local optima. This feature would result in non-deterministic output in solutions that vary from one run to another of a meta-heuristic. Consequently, both the average and variation of outputs over multiple runs have to be considered in evaluating performances of different configurations of a meta-heuristic or distinct meta-heuristics. To this end, this work considers each algorithm as a decision-making unit (DMU) and develops robust data envelopment analysis (DEA) models taking into account not only average but also standard deviation of an algorithm's output for evaluating relative efficiencies of a set of algorithms. The robust DEA models describe uncertain output using an uncertainty set, and aim to maximize a DMU's worst-case relative efficiency with respect to that uncertainty set. The proposed models are employed to evaluate a set of distinct configurations of a genetic algorithm and a set of parameter settings of a simulated annealing heuristic. Evaluation results demonstrate that the robust DEA models are able to identify efficient algorithmic configurations. The proposed models contribute not only to the evaluation of meta-heuristics but also to the DEA methodology.

1. Introduction

This work focuses on determining relative efficiencies of a set of algorithmic configurations of a meta-heuristic or a set of different meta-heuristics for solving combinatorial optimization problems. Typically, a meta-heuristic consists of several algorithmic operators, each of which may be implemented in a number of distinct ways, thus resulting in various configurations (or combinations) with different performances. Comparing various combinations and identifying the most efficient one(s) are the critical tasks at the final stage of developing a meta-heuristic.

In the literature, the most commonly-used method is the empirical analysis which involves an extensive set of paired-*t* tests for comparing *average* algorithmic performances with respect to a number of criteria (such as computational time, objective values, robustness, and flexibility) on a wide range of problem instances (Bräysy & Gendreau, 2005a,b). More sophisticated methods may be based on the Design of Experiments (DOE) and Analysis of Variance (ANOVA), e.g., Coy (2000), Francois and Lavergne (2001), Rardin and Uzsoy (2001), Bartz-Beielstein (2006), Ruiz, Maroto, and Alcaraz (2006), Birattari (2009). While these methods help to decide best configurations, experimental design should be ideally and carefully used on a number of combinations of algorithmic operators, to arrive at conclusions that have meaning in a statistical sense. Inevitably, the complexity of making the decision based on the empirical analysis increases dramatically with the numbers of distinct operators, parameters values, and evaluation criteria, mainly due to a large number of paired-*t* tests. Furthermore, since ANOVA is parametric, when using the DOE with ANOVA, we have to check the three main hypotheses, which are normality, homoskedasticity, and independence of residuals; this is generally not a simple task. As also pointed out by Ruiz et al. (2006), the resulting ANOVA has many degrees of freedom; one has to be very careful when analyzing results of an experiment with such large sample sizes. In solving the meta-heuristics tuning problem, which aims to determine the best configuration of a meta-heuristic, Birattari (2009) proposed the *F*-Race algorithm, which adopts the Friedman two-way ANOVA in the racing algorithm inspired by Hoeffding race, introduced by Maron and Moore (1994), for solving the model selection problem in machine learning. Although this approach is statistically sound, it is computationally intensive and still has to satisfy the hypotheses of ANOVA.

To avoid the hassles of using the above empirical analysis methods, Lu and Yu (2012) proposed an alternative that adopted

☆ This manuscript was processed by Area Editor Imed Kacem.

* Tel.: +886 2 27712171x2306; fax: +886 2 27317168.
E-mail address: jasoncclu@gmail.com

http://dx.doi.org/10.1016/j.cie.2014.12.027
0360-8352/

data envelopment analysis (or DEA; e.g., Amin & Toloo, 2007; Banker, Charnes, & Cooper, 1984; Charnes, Cooper, & Rhodes, 1978; Toloo, 2012, 2013, 2014a,b; Toloo & Nalchigar, 2009) to evaluate relative efficiencies of a set of combinations of the genetic algorithm (GA) operators (i.e., selection, crossover and mutation) for solving the pickup and delivery vehicle routing problem with soft time windows (PDVRPSTW). In this approach, each possible combination of GA operators was considered as a decision making unit (DMU), and DEA was adopted to evaluate and compare the algorithmic efficiency of the distinct GA combinations under consideration. In addition, the cross-efficiency (CE) method (e.g., Doyle & Green, 1994) was employed to rank the combinations. The numerical results showed that DEA is well-suited for determining efficient combinations of GA operators (Lu & Wu, 2014).

While DEA represents a promising alternative for evaluating relative efficiencies of algorithms, one of its limitations is that input and output data of DMUs which are used as the coefficients in the corresponding linear programs (LPs) have to be precisely-known, *a priori* (Toloo & Nalchigar, 2011). This limitation may affect the efficiency evaluation of some state-of-the-art meta-heuristics that incorporate probabilistic operators with the aim to diversify search directions or to escape from being trapped in local optima. For example, GAs include a mutation rate that controls the probability of applying a mutation operator to chromosomes; simulated annealing (SA) approaches use Cauchy or Boltzmann functions (e.g., Lin, Lee, Lu, & Ying, 2011; Lin, Yu, & Lu, 2011; Ying, Lin, & Lu, 2011) in the annealing process to determine the probability of replacing current solution with a worse solution. In this type of meta-heuristics, incorporating probabilistic operators would result in non-deterministic or uncertain output in solutions that vary from one run to another. Consequently, both the average and variation of outputs over multiple runs have to be considered in evaluating algorithmic efficiency.

Previous research on DEA models with imprecise data may be classified into three main categories, namely, fuzzy DEA, interval DEA, and robust DEA. In the fuzzy DEA approach, Sengupta (1992) presented a fuzzy LP transformation to deal with DEA models with fuzzy input and output data. Based on the α-cut approach, Kao and Liu (2000) proposed a transformation of a fuzzy DEA model into a family of crisp DEA models, while Guo and Tanaka (2001) introduced an approach that changed a fuzzy DEA model to a bi-level LP model. Some other examples of fuzzy DEA models can be found in Soleimani-damaneh, Jahanshahloo, and Abbasbandy (2006) and Liu (2008). A common shortcoming of the fuzzy DEA approach is that fuzzy DEA models are computationally expensive (Soleimani-damaneh et al., 2006).

In the interval DEA approach, input and output values are selected from their respective intervals with prescribed lower and upper bounds so as to maximize a DMU's relative efficiency score. Because both the input and output data and weights are variables, the interval DEA approach results in nonlinear programs (NLP). Cooper, Park, and Yu (1999) developed an interval approach that permits mixtures of imprecise and precise data. To deal with the nonlinear interval DEA model, they proposed a two-stage transformation that involves scale transformations and variable alternations to transform the interval DEA model into an ordinary linear program. Despotis and Smirlis (2002) defined the upper and lower bounds for the efficiency scores of the DMUs, and proposed transformations of nonlinear interval DEA models to LP equivalents. They also used a post-DEA model and the endurance indices to discriminate among the efficient DMUs. Some other examples of interval DEA models can be found in Entani, Maeda, and Tanaka (2002), Kao (2006), and Toloo and Ertay (2014). In the interval DEA approach, some DMUs may be always efficient or inefficient for any combinations of values within given intervals, while others may be either efficient or inefficient depending on the values assigned. Moreover, the DMUs are no longer represented by points in the hyper-plane, and instead a number of efficient frontiers may exist. The efficiency of the DMUs may vary according to the efficiency frontier selected.

The robust DEA approach is based on the robust counterpart optimization (RCO) approach (e.g., Ben-Tal & Nemirovski, 1999; Bertsimas & Sim, 2004) which describes uncertain data using an uncertainty set, and aims to maximize a DMU's worst-case relative efficiency with respect to that uncertainty set. Sadjadi and Omrani (2008) proposed robust DEA models with consideration of uncertainty on output parameters for the performance assessment of electricity distribution companies. Shokouhi, Hatami-Marbini, Tavana, and Saati (2010) developed robust DEA models which consider uncertainty on both input and output parameters. Note that both of the works focused on the adaptation of Bertsimas and Sim's (2004) approach to the CCR model (Charnes et al., 1978). More recently, Sadjadi, Omrani, Abdollahzadeh, Alinaghian, and Mohammadi (2011) applied the RCO approach of Ben-Tal and Nemirovski (1999) to the super-efficiency DEA model of Andersen and Petersen (1993), which was also based on the CCR model.

To take into account the aforementioned output uncertainties (due to probabilistic operators) in using DEA models to evaluate algorithmic efficiency, this research develops two robust DEA models which take into account not only the average but also standard deviation of algorithm's output values. Particularly, our work adapts the RCO approach to the BCC model (Banker et al., 1984), which allows variable returns-to-scale (VRS) on production frontiers, whereas previous robust DEA approaches were based on the CCR model, which assumes constant returns-to-scale (CRS). The BCC model has seen a wider use than the CCR model, because the former was built based on the more general assumption, VRS. As a result, the proposed robust BCC models should receive greater attention than the robust CCR models presented in the literature (Sadjadi and Omrani, 2008; Sadjadi, Omrani, Abdollahzadeh, Alinaghian, and Mohammadi, 2011; Shokouhi et al., 2010). Moreover, this research represents the first to develop robust BCC models using distinct RCO techniques in the literature (i.e., Bertsimas & Sim, 2004; Ben-Tal & Nemirovski, 1999) and compared their results in algorithmic efficiency evaluation.

Using the same test dataset and GA algorithms provided by Lu and Yu (2012), we demonstrate the application of the robust DEA models to evaluate a set of GA algorithms for solving the PDVRPSTW. Additionally, the proposed models are applied to evaluate a set of parameter settings of a simulated annealing (SA) heuristic developed for solving the truck and trailer routing problem with time windows (Lin, Yu, et al., 2011; Lin, Lee, et al., 2011). To verify its effectiveness, the proposed approach is compared with the conventional empirical analysis method based on paired-*t* tests. It is important to note that a large number of paired-*t* tests have to be conducted, if the empirical analysis method is employed to determine the efficient combinations of GA operators or the efficient SA parameter settings. Furthermore, the solution variation among different runs of GA and SA due to the probabilistic feature in the mutation operator would add more complexity to the empirical analysis approach. Instead, by applying the proposed robust DEA models, which explicitly address (output) coefficient uncertainties in the BCC model, the relative efficiency of each algorithmic configuration can be easily determined.

The rest of this paper is structured as follows. Section 2 describes the robust DEA models for evaluating the relative efficiency of a set of algorithmic combinations. Section 3 presents the applications of the robust DEA models to evaluate a number of combinations of GA operators and a set of SA parameter settings. Concluding remarks are in Section 4.