Computers and Chemical Engineering 109 (2018) 96–111

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Scenario tree reduction methods through clustering nodes

Zhiping Chen*, Zhe Yan

Department of Computing Science, School of Mathematics and Statistics, Xi'an Jiaotong University, 710049 Xi'an, Shaanxi, PR China

ARTICLE INFO

Article history:
Received 6 May 2017
Received in revised form 16 October 2017
Accepted 16 October 2017

Keywords:
Stochastic programs
Scenario tree reduction
Clustering
Stability
Computational efficiency
Portfolio selection

ABSTRACT

To develop practical and efficient scenario tree reduction methods, we introduce a new methodology which depends on clustering nodes, and thus an easy-to-handle distance function to measure the difference between two scenario trees is designed. On the basis of minimizing the new distance, we construct a multiperiod scenario tree reduction model which is supported theoretically by the stability results of stochastic programs. By solving the model, we design a stage-wise scenario tree reduction algorithm which is superior to the simultaneous backward reduction method in terms of both computational complexity and solution results of stochastic programming problems, the corresponding reduction algorithm especially for fan-liked trees is also presented. We further design a multiperiod scenario tree reduction algorithm with a pre-specified distance by utilizing the stability results of stochastic programs. A series of numerical experiments with real trading data and the application to multiperiod portfolio selection problem demonstrate the practicality, efficiency and robustness of proposed reduction model and algorithms.

1. Introduction

Stochastic programming is a very useful approach for coping with decision making problems under uncertainty in areas such as production, energy, transportation and finance. In general, stochastic parameters in these decision problems are expressed via continuous probability distributions or stochastic processes, and the original forms of these decision problems are infinite-dimensional optimization problems. To make these problems tractable, one usually resorts to the scenario tree technique (Høyland and Wallace, 2001) to transform the original stochastic optimization problem into a finite-dimensional programming problem. In order to approximate the underlying random data process as precisely as possible, the size, which usually corresponds to the number of all the nodes, of the scenario tree tends to be very large, and the deterministic transformation of the original decision problem is large-scale and hard to solve. Due to this, the scenario tree reduction method is proposed (Dupačová et al., 2003), which can be used to decrease the dimension of the decision problem. How to appropriately reduce a scenario tree has become an active research area since then.

Scenario tree reduction methods aim at reducing the node number of a scenario tree. For example, the k-means clustering method is used to partition a given set of nodes into a number of clusters (Arthur and Vassilvitskii, 2007). As a result of this partition, nodes with similar features are assigned to the same cluster. Each cluster is then replaced by a node, and the node number is thus reduced. When a scenario tree is reduced, the corresponding programming problem based on the tree is changed and so the corresponding optimal solution and value. A good scenario tree reduction method should ensure that the reduced tree's size is small enough while keeping the resulting optimal solution and value close to the original ones as much as possible. For this reason, most scenario tree reduction algorithms are designed by using the quantitative stability results of stochastic programs.

On the basis of an upper bound of the Fortet-Mourier probability metrics from the stability analysis of convex stochastic programs, backward reduction and forward selection methods are designed to reduce a scenario tree (Dupačová et al., 2003). These two methods are further developed to the simultaneous backward reduction method and the fast forward selection method to improve the scenario reduction performance (Heitsch and Römisch, 2003). To enhance the accuracy of the reduction process, the preceding reduction methods are improved by relying directly on Fortet-Mourier metrics instead of their upper bounds (Heitsch and Römisch, 2007). Some heuristics for forward and backward scenario tree reduction algorithms (Heitsch and Römisch, 2009a) are developed by applying the new stability results in Heitsch et al.

* Corresponding author.
E-mail addresses: zchen@mail.xjtu.edu.cn (Z. Chen), xyqm12@stu.xjtu.edu.cn (Z. Yan).

https://doi.org/10.1016/j.compchemeng.2017.10.017

Nomenclature

$\mathbb{P}$	the original scenario tree
τ_t	a node at stage t of the scenario tree $\mathbb{P}$
Ω_t	set of nodes at stage t of the scenario tree $\mathbb{P}$
N_t	number of nodes at stage t of the scenario tree $\mathbb{P}$
C_{τ_t}	set of all the son nodes of the node τ_t
ξ_{τ_t}	the associated value at node τ_t
p_{τ_t}	the occurring probability of node τ_t
$\phi(\cdot)$	the mapping which maps a node into its father node
$\tilde{\mathbb{P}}$	the reduced scenario tree
$\tilde{\tau}_t$	a node at stage t of the reduced scenario tree $\tilde{\mathbb{P}}$
$\tilde{\Omega}_t$	set of nodes at stage t of the reduced scenario tree $\tilde{\mathbb{P}}$
$\tilde{N}_t$	number of nodes at stage t of the reduced scenario tree $\tilde{\mathbb{P}}$
$\tilde{C}_{\tilde{\tau}_t}$	set of all the son nodes of the node $\tilde{\tau}_t$
$\tilde{N}^{\tilde{\tau}_t}$	number of the son nodes of the node $\tilde{\tau}_t$
$\tilde{\xi}_{\tilde{\tau}_t}$	the associated value at node $\tilde{\tau}_t$
$\tilde{p}_{\tilde{\tau}_t}$	the occurring probability of node $\tilde{\tau}_t$
$I_{\tilde{\tau}_t}$	the cluster of nodes in $\mathbb{P}$ which are replaced by the single node $\tilde{\tau}_t$
$S_{\tilde{\tau}_t}$	the son node set of the node $\tilde{\tau}_t$ before the son nodes are reduced
pred	the pre-specified distance
$W_1(W_2)$	the Wasserstein distance with order $r=1$ ($r=2$)
D_2	the D_r-distance with order $r=2$
SBRM	the simultaneous backward reduction method
SOST	branching structure of the original scenario tree
SNN	stage-wise node number of a scenario tree
SNFS	scenario number of the fan-liked scenario tree
SRST	branching structure of the reduced scenario tree
TNN	total number of nodes in the reduced scenario tree

(2006). Scenario tree reduction methods have been extended to chance constrained and mix-integer two-stage stochastic programs (Henrion et al., 2008, 2009). The sophisticated stability results for multistage stochastic programs in Heitsch et al. (2006) tell us that the so-called filtration distance should be considered during the multistage scenario tree reduction process. On the basis of the new stability result, a single node reduction method is designed (Heitsch and Römisch, 2009b), taking both the probability distribution distance and the filtration distance into account. All the above scenario tree reduction methods rely heavily on the solution of facility location problems which are NP-hard. A random search procedure that can quickly solve facility location problems is proposed (Armstrong et al., 2013), and thus an efficient scenario tree reduction method can be designed. Based on the transportation distance (i.e., Kantorovich distance), a mixed integer linear optimization and a linear programming problem are established, respectively, for scenario reduction (Li and Floudas, 2014, 2015; Li and Li, 2016). Recently, the notions of the nested distribution and the nested distance are proposed for stochastic programs (Pflug, 2009; Pflug and Pichler, 2012), with which the filtration distance can be avoided. These notions have been used for designing scenario tree generation methods and can also be used as a potential way for designing new scenario tree reduction methods (Pflug and Pichler, 2015; Kovacevic and Pichler, 2015).

Scenario tree reduction methods in Dupačová et al. (2003) and Heitsch and Römisch (2003) are designed by minimizing probabilistic distances which are similar to the Wasserstein distance. The Wasserstein distance between the probability measure P and the probability measure $\tilde{P}$ can be formulated as

$$W_r(P,\tilde{P}) := \inf\left\{ \left(\iint_{\Xi\times\tilde{\Xi}} d(\xi,\tilde{\xi})^r \eta(d\xi, d\tilde{\xi})\right)^{1/r} : \eta \text{ is a probability measure on } \Xi\times\tilde{\Xi} \text{ with marginal distributions } P \text{ and } \tilde{P} \right\}.$$

Here, $r\geq 1$, Ξ and $\tilde{\Xi}$ are sample spaces and $d(\xi,\tilde{\xi})$ is the distance between ξ and $\tilde{\xi}$. If $P(\tilde{P})$ stands for a multiperiod scenario tree, $\xi(\tilde{\xi})$ then corresponds to an arbitrary scenario. In this case, we can see that the Wasserstein distance treats scenarios of a scenario tree as independent paths and does not take into account the structure information (or filtration, interstage dependence information) of the scenario tree. This results in that the reduced scenario tree obtained through minimizing the Wasserstein distance to the original scenario tree will lose much structure information of the original scenario tree, and the solutions of the stochastic programming problem based on the reduced scenario tree might significantly deviate from those obtained based on the original scenario tree, even though the two trees are very close to each other in terms of the Wasserstein distance. This argument is supported by numerical results in this paper. Therefore, methods in Dupačová et al. (2003) and Heitsch and Römisch (2003) are not really appropriate for multiperiod scenario tree reduction. To overcome this, the filtration distance and the nested distance are introduced, the two distances consider the structure information of a scenario tree. However, they are complicated and difficult to handle which makes the corresponding scenario tree reduction algorithms (Heitsch and Römisch, 2009; Kovacevic and Pichler, 2015) intricate and inefficient in real applications.

To develop rational and efficient multiperiod scenario tree reduction methods, we introduce a new method which clusters nodes of a scenario tree. Concretely, we divide the nodes having the same father node into several clusters and replace each cluster by a new single node. By doing this, the node number of the scenario tree is decreased. Meanwhile, the dependence information among stages is kept because we only do the clustering procedure among the nodes with the same father node. An easy-to-handle distance function measuring the difference between the original scenario tree and the corresponding reduced tree is then derived. By minimizing this distance, we construct a multiperiod scenario tree reduction model which is supported theoretically by the stability results in Pflug and Pichler (2015). It is proved that in the single-period case, the reduced tree obtained by our model is the one which is closest to the original tree among the trees with the same size in terms of the Wasserstein distance. Two methods are designed to solve the model in the single-period case: the recursive-type method and the r-cluster method. We demonstrate that both methods consume less time than the simultaneous backward reduction method (Heitsch and Römisch, 2003).

By combining the recursive-type method and the r-cluster method, we design a stage-wise hybrid algorithm to solve the constructed multiperiod scenario tree reduction model. The resulting algorithm can generate a reduced scenario tree with the pre-specified structure. Moreover, to find a smaller scenario tree which is the closest one to the original tree in terms of a pre-specified distance, we design a multiperiod scenario tree reduction algorithm with a pre-specified distance by utilizing the stability result in Pflug and Pichler (2015). Finally, all the proposed algorithms are tested through numerical experiments to demonstrate their efficiency. And the reduction algorithms are applied to solve a multiperiod portfolio selection problem. The results show that it is necessary and significant to use the designed reduction methods when we want to solve stochastic programming problems on the basis of the scenario tree.