Computers and Operations Research 88 (2017) 44–57

Contents lists available at ScienceDirect

Computers and Operations Research

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

Network design in scarce data environment using moment-based distributionally robust optimization

Hideaki Nakao, Siqian Shen*, Zhihao Chen

Department of Industrial and Operations Engineering University of Michigan, Ann Arbor, MI, USA

ARTICLE INFO

Article history:
Received 20 September 2016
Revised 30 June 2017
Accepted 2 July 2017
Available online 3 July 2017

Keywords:
Stochastic/robust network design
Distributionally robust optimization
Mixed-integer linear programming
Polynomial linearization
Cutting-plane algorithm

ABSTRACT

We consider a network design problem (NDP) under random demand with unknown distribution for which only a small number of observations are known. We design arc capacities in the first stage and optimize single-commodity network flows after realizing the demand in the second stage. The objective is to minimize the total cost of allocating arc capacities, flowing commodities, and penalty for unmet demand. We formulate a distributionally robust NDP (DR-NDP) by constructing an ambiguity set of the unknown demand distribution based on marginal moment information, to minimize the worst-case total cost over all possible distributions. Approximating polynomials with piecewise-linear functions, we reformulate DR-NDP as a mixed-integer linear program optimized via a cutting-plane algorithm. We test diverse network instances to compare DR-NDP with a stochastic programming approach, a deterministic benchmark model, and a robust NDP formulation. Our results demonstrate adequate robustness of optimal DR-NDP solutions and how they perform under varying demand, modeling parameter, network, and cost settings. The results highlight potential niche uses of DR-NDP in data-scarce contexts.

1. Introduction

Network design problems (NDPs) refer to an important class of problems arising in a wide range of applications, including traffic control, energy dispatch, disaster relief, and supply chain management. We focus on single-commodity NDPs with uncertain demand and design arc capacities of a given network to satisfy the demand through ad-hoc flows. In particular, we consider a distributionally robust optimization approach, when a small number of demand observations are known, but the data may not be sufficient for deriving an exact distribution that is necessary for applying traditional stochastic programming approaches.

In our problem, a network planner intends to build arcs and decide their capacities in a network before knowing actual realizations of random demand. The cost of building an arc is proportional to the capacity of the arc. Once the arcs are built, demand is realized and is met by flows generated from given supply locations to demand sites. The flow cost on each arc is proportional to the amount of flow on the arc. Additionally, any unmet demand is penalized proportional to the amount of demand shortage. The goal is to minimize the total cost of building arcs, flowing commodities, and unmet demand penalty. The latter two types of cost are random due to the uncertain demand. As the realized demand is not known a priori, the planner has to make decisions based on previously observed realizations. However, historical data may be scarce and may not describe the distribution accurately. To this end, we propose a distributionally robust NDP approach (DR-NDP), to conservatively allocate arc capacities as a precaution against observed data that may not be representative of the true demand distribution. Using empirical mean values and standard deviations, we construct an ambiguity set of possible demand distributions and seek optimal design of arc capacities to minimize the worst-case expected flow cost and demand shortage penalty for any distribution in the set. This approach provides more conservative solutions as compared to standard stochastic optimization methods that assume fully known demand distribution. On the other hand, the marginal moments used by DR-NDP can be easily obtained from the observed data. We show later via computational studies that using the moment information, DR-NDP also yields less conservative results than the ones of robust optimization approach relying on very little information about the uncertainty. Having a more complex formulation, it takes longer to solve DR-NDP than the stochastic NDP model constructed by using the Sample Average Approximation (SAA) method and the robust NDP formulation using a budget uncertainty set. But it takes very short amount of time (often within a few seconds) to implement all three approaches.

* Corresponding author.
E-mail addresses: nakaoh@umich.edu (H. Nakao), siqian@umich.edu, siqian.shen@gmail.com (S. Shen), czhihao@umich.edu (Z. Chen).

http://dx.doi.org/10.1016/j.cor.2017.07.002
0305-0548/

1.1. Applications

The above description reveals the opportunities that this research targets. Via DR-NDP, we aim to find relatively conservative arc design solutions, of which few data points are available to accurately determine the underlying true distribution of demand. The presence of the unmet demand penalty also limits its applications to problems in which demand shortage significantly impacts service quality and system reliability. We present two examples applications below.

- **Humanitarian relief supply network design:** When providing humanitarian relief after disaster occurs, little prior data about uncertain demand for aid is likely to be known due to the rare occurrence of disaster. Using DR-NDP, one can determine the most cost-effective way to supply relief to the affected region, and can conservatively design the supply network to provide high-quality relief operations.
- **Supply network design for new products:** Releasing new products or introducing existing products to new regions requires careful planning of the supply network. Historical data for forecasting consumer demand is likely to be uncertain and scarce. Thus, using DR-NDP to design supply networks makes sure that products are launched to satisfy most customers' needs and to keep demand shortage low.

1.2. Literature review

NDPs ubiquitously arise in network planning and operational management, and there is a significant amount of literature on the related theories and applications. Magnanti and Wong (1984) describe a unifying framework for deterministic single and multi-commodity NDPs, and specify the design of the framework for facility location and traffic network design problems. Specific NDPs are also studied in depth for different practical contexts, e.g., telecommunications and computer network design in Pióro and Medhi (2004) and Babonneau et al. (2013). Poss (2011) provides a detailed summary of models and algorithms for NDPs with diverse network and problem settings. Recently, parameter uncertainty becomes a major concern in NDP related studies, since the practical design of networks is usually in anticipation of future uncertain needs and circumstances. The literature of stochastic NDPs typically assumes fully known distributional information on the demand uncertainty, allowing to solve the problem by using large-scale stochastic programs with a finite, but large number of realizations (i.e., "scenarios") (see, e.g., Santoso et al., 2005; Ukkusuri and Patil, 2009). The scenarios can be generated either from certain assumed distributions (e.g., using the Sample Average Approximation (SAA) method (see Kleywegt et al., 2002)) or based on statistical information (e.g., using a moment-matching method (Høyland et al., 2003)).

In cases where the demand distribution is not fully known, another avenue for optimizing NDPs is through robust optimization (Ben-Tal and Nemirovski, 1998; Bertsimas and Sim, 2003,2004), to optimize the worst-case objective value for any uncertain parameter realization over a pre-defined uncertainty set. They are well-suited to conservatively solve problems without exact distributional information, but only having small amounts of data to derive beliefs of the uncertainty structure. Atamtürk and Zhang (2007) propose a two-stage robust optimization approach for solving network flow and design problems with uncertain demand. Like our paper, they model the network flow decisions as recourse variables decided after realizing the demand. This allows one to control conservatism of the solutions. They also apply a budget uncertainty set (see Bertsimas and Sim (2004)) under which they provide an upper bound on the probability of a robust solution being infeasible for any random demand vector. Altin et al. (2011), and Koster et al. (2013) describe a variety of formulations, complexities, valid inequalities, and computational results for robust NDP with uncertain (hose) demand. Álvarez-Miranda et al. (2014) discuss the complexity and propose heuristic algorithms for robust NDPs; Cacchiani et al. (2016) optimize single-commodity robust NDPs by using the branch-and-cut algorithm based on efficient separation procedures. They consider uncertainty sets of the hose demand modeled as a finite set of scenarios or as a polytope. Robust NDPs are often reformulated as two-stage mixed-integer linear programs, and can be optimized through decomposition, cutting-plane, and/or column-generation methods (see, e.g., Ayoub and Poss, 2016; Lee et al., 2013; Shen et al., 2016; Atamtürk and Zhang, 2007). Furthermore, in (Ben-Ameur and Kerivin, 2005; Mattia, 2013; Poss and Raack, 2013; Scutellà, 2009), the authors investigate robust NDPs with dynamic routing decisions made in multiple stages when sequentially realizing the demand, rather than using static network flows in the two-stage setting.

Distributionally robust (DR) optimization approaches recently receive much more attention for optimization under distributional uncertainty. They plan against the worst-case distribution in an ambiguity set that is typically built based on statistical information from historical data. Bertsimas et al. (2010); Delage and Ye (2010); Jiang and Guan (2016) propose different moment-based ambiguity sets for deriving tractable reformulations of DR stochastic or chance-constrained programs. In this paper, we consider a moment-based ambiguity set that is constructed based on the empirical mean values and standard deviations of demand. To the best of our knowledge, this approach is novel for designing network arc capacities with ad-hoc flows. Later, we will also numerically demonstrate that DR-NDP yields adequately robust but not over conservative solutions that are less sensitive to changes of distribution, demand, network, and cost parameters, compared with stochastic NDP or robust NDP.

1.3. Structure of paper

The remainder of the paper is organized as follows. In Section 2, we describe a generic NDP and the formulation of a marginal moment-based ambiguity set for DR-NDP. In Section 3, we reformulate DR-NDP as a single-level optimization model with mixed-integer linear programming constraints, and we approximate the quadratic objective by using a piecewise-linear function. In Section 4, we develop a cutting-plane algorithm for optimizing the reformulation. In Section 5, we demonstrate the computational results of DR-NDP and conduct sensitivity analysis by testing diverse network instances and parameters. In Section 6, we conclude the paper and discuss future research directions.

2. Problem formulation

Consider a directed network $G(N, A)$ with N being the set of nodes, and A being the set of arcs. The goal is to minimize the total cost of arc capacity allocation, network flows, and unmet demand penalty. Without loss of generality, the network has disjoint sets of supply nodes $S \subset N$ and demand nodes $D \subset N$. We assume that each supply node $i \in S$ has deterministic supply $s_i > 0$ and each demand node $i \in D$ has uncertain demand $d_i \geq 0$. The demands are assumed to be jointly distributed and, while the true distribution of the demand vector $d = (d_i,\ i \in D)^\mathsf{T}$ is not known, we are given a set of $|K|$ data observations of d, denoted by $\{d^k\}_{k \in K}$. We assume that d has a boxed-shaped support Ξ, given by $\Xi := \{d \in \mathbb{R}_+^{|D|} : \underline{d}_i \leq d_i \leq \overline{d}_i,\ \forall i \in D\}$, where $\underline{d}, \overline{d} \in \mathbb{R}_+^{|D|}$ represent the vectors of lower and upper bounds of d, respectively. These bounds can be obtained from historical demand data or via prior knowledge in specific problem contexts. (For example, the number of users at a