Computers and Chemical Engineering 106 (2017) 464–479

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Data-driven robust optimization based on kernel learning

Chao Shang[a], Xiaolin Huang[b], Fengqi You[a,*]

[a] Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, USA
[b] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200400, China

ARTICLE INFO

Article history:
Received 2 June 2017
Received in revised form 5 July 2017
Accepted 11 July 2017
Available online 18 July 2017

Keywords:
Robust optimization
Uncertainty set
Data-driven methods
Support vector clustering
Piecewise linear modeling

ABSTRACT

We propose piecewise linear kernel-based support vector clustering (SVC) as a new approach tailored to data-driven robust optimization. By solving a quadratic program, the distributional geometry of massive uncertain data can be effectively captured as a compact convex uncertainty set, which considerably reduces conservatism of robust optimization problems. The induced robust counterpart problem retains the same type as the deterministic problem, which provides significant computational benefits. In addition, by exploiting statistical properties of SVC, the fraction of data coverage of the data-driven uncertainty set can be easily selected by adjusting only one parameter, which furnishes an interpretable and pragmatic way to control conservatism and exclude outliers. Numerical studies and an industrial application of process network planning demonstrate that, the proposed data-driven approach can effectively utilize useful information with massive data, and better hedge against uncertainties and yield less conservative solutions.

1. Introduction

In science and engineering, optimization problems are inevitably encountered whenever one seeks to make decisions by maximizing/minimizing some certain criterion. However, real-world parameters are subject to randomness in various degrees, rendering deterministic optimization models unreliable in an uncertain environment (Sahinidis, 2004). It has been demonstrated that even a slight perturbation on parameters in an optimization problem could exert overwhelming effects on the computed optimal solutions, which results in suboptimality or even infeasibility of optimization problems (Mulvey et al., 1995). Motivated by the urgent requirement of handling uncertainties in decision making, stochastic optimization and robust optimization have received immense research attentions in recent decades (Kall et al., 1994; Ben-Tal et al., 2009; Birge and Louveaux, 2011; Bertsimas et al., 2011; Gabrel et al., 2014; Yuan et al., 2017), which hedge against uncertainties by bringing in conservatism. Stochastic optimization entails complete knowledge about the underlying probability distribution of uncertainties, which may be unrealistic in practice. As an effective alternative, robust optimization takes a deterministic and set-based way to model uncertainties, and balance between the modeling power and computational tractability, which has obtained considerable attentions recently in the realm of process systems engineering. Typical applications include process network design (Gong et al., 2016; Gong and You, 2017), supply chain management (Tong et al., 2014; Yue and You, 2016), and process scheduling (Shi and You, 2016).

A paramount ingredient in robust optimization is to construct an uncertainty set including probable realizations of uncertain parameters. The earliest attempts in formulating uncertainty sets date back to the 1970s, with the work of Soyster (1973), in which coefficients are perturbed by uncertainties distributed in a known box. Despite its computational convenience and guaranteed feasibility, the box uncertainty set tends to induce over-conservative decisions. Later, immense research effort has been made on devising more flexible robust models to ameliorate over-conservatism. Ellipsoidal uncertainty sets have been put forward independently by El Ghaoui et al. (1998); Ben-Tal and Nemirovski (1998, 1999), based on which the robust counterpart model simplifies to a conic quadratic problem in the presence of linear constraints. To enhance the modeling flexibilities, intersections of basic uncertainty sets have been designed, including the "interval + ellipsoidal" uncertainty set (Ben-Tal and Nemirovski, 2000) and the "interval + polyhedral" uncertainty set. Bertsimas and Sim (2004) robustify linear programs using a polyhedral uncertainty set adjustable with the so-called *budget*, which turns out to be

* Corresponding author.
E-mail addresses: chao.shang@cornell.edu (C. Shang), xiaolinhuang@sjtu.edu.cn (X. Huang), fengqi.you@cornell.edu (F. You).

http://dx.doi.org/10.1016/j.compchemeng.2017.07.004
0098-1354/

identical to the "interval + polyhedral" uncertainty set. Interested readers are referred to review papers (Bertsimas et al., 2011; Gabrel et al., 2014) and the monograph (Ben-Tal et al., 2009) for a comprehensive overview of uncertainty sets-induced robust optimization methods.

Despite the burgeoning prevalence of uncertainty set-induced robust optimization, a potential limitation is that uncertainties in each dimension are assumed as being independently and symmetrically distributed. The issue of data correlation is first addressed by Bertsimas and Sim (2004), in which correlated uncertainties are disentangled by means of an underlying source of independent uncertainties, expressed as:

$$\tilde{a}_{ij} = a_{ij} + \sum_{k \in K_i} \tilde{\eta}_{ik} g_{kj}. \tag{1}$$

where $\{a_{ij}\}$ stands for the nominal value, and $\{\tilde{\eta}_{ik}\}$ denotes the source of independent uncertainties. Ferreira et al. (2012) suggest determining the values of $\{\tilde{a}_{ij}\}$ and $\{g_{kj}\}$ by means of principal component analysis (PCA) and minimum power decomposition (MPD). Based on (1), classical symmetric uncertainty sets are generalized by Yuan et al. (2016), and explicit formulations of their robust counterparts are also provided. Jalilvand-Nejad et al. (2016) devise correlated polyhedral uncertainty sets using second-order statistical information from historical data. In regard to asymmetric uncertainties, Chen et al. (2007) adopt the forward and backward deviations to capture distributional asymmetry through the norm-induced uncertainty set while still preserving its tractability. In Natarajan et al. (2008), a modified value-at-risk (VaR) measure is developed and investigated by taking into account asymmetries in the distributions of portfolio returns.

A prominent and practical issue of uncertainty set-induced robust optimization is how to determine the set coefficients appropriately, which are in general assumed as known according to domain-specific knowledge. In the absence of first-principle knowledge, leveraging available historical data provides a practical way to characterize the distributional information. For instance, the lower and the upper bounds of an interval can be specified as the minimum and the maximum of observed data samples, albeit conservatively. More sophisticated approaches make use of variance and covariance of historical data, which renders the model statistically interpretable (Pachamanova, 2002; Bertsimas and Pachamanova, 2008; Ferreira et al., 2012). An alternative streamline of data-driven optimization is the *distributionally robust optimization*, which utilizes both data and hypothesis tests to construct the ambiguity set including $\mathbb{P}$ at a high confidence level (Delage and Ye, 2010; Jiang and Guan, 2016; Bertsimas et al., 2017). However, one still needs to specify the type of hypothesis tests to yield a reliable solution, and the induced optimization problem is generally difficult to solve (Hanasusanto et al., 2017). For example, the moment-based hypothesis test typically leads to reformulations in terms of linear matrix inequalities and bilinear matrix inequalities (Delage and Ye, 2010; Zymler et al., 2013), which erect obstacles to further tackle mixed-integer and large-scale problems that are commonly encountered in process systems engineering. In this work, we hence focus on uncertainty set-based robust optimization with better applicability and implementation convenience.

In real-world applications, the underlying distribution $\mathbb{P}$ of uncertainties may be intrinsically complicated and vary under different circumstances. When one is faced with high-dimensional uncertainties, it is rather challenging to choose the type of uncertainty sets by prior knowledge, tune the coefficients, and further evaluate its divergence with the true underlying distribution $\mathbb{P}$. The era of big data paves new way for decision-making under uncertainties by exploiting massive data available at hand (Bertsimas et al., 2011; Qin, 2014). A desirable uncertainty set shall flexibly adapt to the intrinsic structure behind data, thereby well characterizing $\mathbb{P}$ and ameliorating the suboptimality of solutions. From a machine learning perspective, constructions of uncertainty sets based upon historical data can be viewed as an *unsupervised learning* problem. There have been a plethora of effective unsupervised learning models, for example, kernel density estimation (KDE) and support vector machines (SVM), which could provide powerful representations of data distributions (Bishop, 2006). In principle, one could resort to such machine learning tools to estimate data densities with sufficient accuracies; nevertheless, it remains a challenging task to formulate an appropriate uncertainty set for modeling robust optimization problems. This is mainly because complicating nonlinear items, such as the radial basis function (RBF) $\exp\{-x^2/2\sigma^2\}$ and the sigmoid function $\tanh(\gamma x + r)$, dominate conventional machine learning models, which invariably prohibit an analytical treatment of robust optimization problems, especially a tractable robust counterpart reformulation. This may somewhat explain the scarce of applications of machine learning models in robust optimization all this time.

In this paper, we propose an effective data-driven approach for robust optimization that is tailored to uncertainty set constructions as well as computational implementations, thereby bridging machine learning and robust optimization directly. As an extended SVM technology, support vector clustering (SVC) has been extensively adopted to estimate the support of an unknown probability distribution from random data samples (Schölkopf et al., 1999; Müller et al., 2001). A particular merit of SVC is that it has theoretical underpinnings in statistical learning theory (Vapnik, 2013), and enjoys desirable generalization performance in face of real-world problems (Lee and Lee, 2005). Instead of using the ubiquitous RBF kernel involving intricate nonlinearities, a novel piecewise linear kernel, referred to as the generalized intersection kernel, is proposed in this work to formulate the SVC model, which entails solving a quadratic program (QP) only. Thanks to the kernel formulation, the SVC model could not only handle correlated uncertainties and lead to asymmetric uncertainty sets, but also enjoys an adaptive complexity, thereby featuring a *nonparametric* scheme. Most importantly, it leads to a *convex polyhedral* uncertainty set, thereby rendering the robust counterpart problem of the same type as the deterministic problem, which provides computational convenience. If the deterministic problem is an MILP, then the robust counterpart problem can be also cast as an MILP. In this way, a satisfactory trade-off can be achieved between modeling power of SVC and computational convenience of robust optimization. Moreover, we show that the parameters to be tuned bear explicit statistical implications, which allow one to easily control the conservatism as well as the complexity of the induced robust optimization problems. Numerical and application case studies are conducted to show the practicability and efficacy of the proposed data-driven approach in hedging against uncertainties and alleviating the conservatism of robust solutions.

The layout of this paper is organized as follows. Section 2 revisits the conventional uncertainty sets for robust optimization, and the SVC-based unsupervised learning schema. In Section 3, the piecewise linear kernel-based SVC is proposed, along with the induced data-driven uncertainty set. Its properties are also discussed from various aspects, including the parameter interpretation, asymptotic behavior, and computational tractability. Section 4 is devoted to a comprehensive numerical evaluation on the performance of the proposed approach compared with classical ones. In Section 5, case studies on an application of process network planning are performed, followed by concluding remarks in Section 6.