Cleaner Logistics and Supply Chain 4 (2022) 100062

Contents lists available at ScienceDirect

Cleaner Logistics and Supply Chain

journal homepage: www.journals.elsevier.com/cleaner-logistics-and-supply-chain

A global method from predictive to prescriptive analytics considering prediction error for "Predict, then optimize" with an example of low-carbon logistics

Shuaian Wang, Ran Yan *

Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

1. Introduction

Many real-world decision making problems for low-carbon logistics involve both prediction and optimization tasks, where the key unknown parameters in the optimization model for decision making are predicted by a prediction model by leveraging available auxiliary data, which are historical records related to the parameters. For example, to help a logistics company to choose the most suitable low-carbon technology to invest, we may first predict the carbon-emission performance of the candidate low-carbon technologies, and then optimize the decision. Common approaches treat the optimization model in a deterministic manner based on the predicted mean of the unknown parameters conditional on the auxiliary data. However, many factors that affect the value of the parameters are not collected or modeled, the amount of collected data is limited, and there are inevitable errors in the data collected. As a result, the machine learning model developed may not be perfect, i.e., the predicted conditional mean is not equal to the actual conditional mean value. Moreover, even if the machine learning model predicts the actual conditional mean value with no error, it is widely known that replacing the value of a random parameter with its mean may lead to suboptimal solutions to an optimization model (Shapiro et al., 2021).

Therefore, when the objective function of the optimization problem is nonlinear in the unknown parameters, the parameters should be modeled as a random variable. The decision problem of interest should be a stochastic optimization problem with imperfect observations to estimate the conditional distribution of its parameters (Bertsimas and Kallus, 2020). The key point is to capture the uncertainties presented in the optimization model by leveraging auxiliary data. This stream of problem is referred to as predictive prescription by Bertsimas and Kallus (2020). To capture the uncertainty, weight functions are derived for training data based on local-learning predictive methodology such as k-nearest-neighbors (kNN) regression, local linear regression, classification and regression trees, and random forests to estimate the distribution of the unknown parameters (Bertsimas and Kallus, 2020).

Motivated by the above research, this study proposes a global method to estimate the conditional distribution of the unknown parameters in the optimization model by leveraging auxiliary data. The conditional distribution of the unknown parameters is first estimated which relies on the prediction error of each record in the training set. Then, the estimated conditional distribution of the unknown parameters is input to the downstream optimization model to prescribe a decision. Our approach to estimate the conditional distribution of unknown parameters is inherently different from that in Bertsimas and Kallus (2020), and ours is more "global", whose meaning is to be explain later. The remainder of this study is organized as follows. The problem to be studied is presented in Section 2. Section 3 discusses the details of our method. Section 4 concludes this research.

2. Problem description

Consider an optimization model below for a low-carbon logistics problem that aims to prescribe an optimal decision z^*:

$$z^* \in \arg\min_{z \in Z} \mathrm{E}[f(C, z)] \quad (1)$$

where z denotes a decision (e.g., a combination of low-carbon technologies to purchase), Z denotes the set of all feasible decisions (e.g., decisions satisfying the total purchasing cost does not exceed the budget), C denotes a random parameter (e.g., the actual carbon-emission performance of a low-carbon technology, which depends on a number of factors such as temperature and humidity) with one realization denoted by c, f is a cost function associated with the parameter and the decision (e.g., total cost for the logistics company, including carbon tax and

* Corresponding author.
E-mail addresses: wangshuaian@gmail.com (S. Wang), ran-angel.yan@polyu.edu.hk (R. Yan).

https://doi.org/10.1016/j.clscn.2022.100062
Received 21 February 2022; Received in revised form 20 May 2022; Accepted 5 June 2022
Available online 9 June 2022

company reputation loss or gain), and E calculates the expected value with respect to the probability distribution of C. Denote the cumulative distribution function (CDF) of random variable C by $F_C(c)$, which is unknown. Since $F_C(c)$ is unknown, we cannot evaluate $\mathrm{E}[f(C, \quad z)]$ for a particular $z \in Z$ of interest and hence cannot solve the model. However, we have the associated auxiliary data denoted by set $S_n = \{(x^1, c^1), \ldots, (x^n, c^n)\}$, where x^1 to x^n are the features associated with the n records, and c^1 to c^n are their outputs. We also have the auxiliary data, denoted by x^{n+1}, which are the features of the record whose output is to be predicted, to predict the conditional distribution of C for our decision problem.

In traditional "predict, then optimize" scheme, for a regression problem, the mean squared error (MSE) is usually used as the metric to build a construction machine learning model to predict the value of the unknown parameters while leveraging the auxiliary data, i.e., $g: x \rightarrow c$. Under certain assumptions, the predicted value is the conditional mean of C given the auxiliary data. Then, the following optimization is regarded to be deterministic, and the optimal decision can be prescribed as follows.

$$z^{Deterministic} \in \arg\min_{z \in Z} f(g(x^{n+1}), \quad z) \tag{2}$$

In practical problems, $f(c, \quad z)$ can be either linear or nonlinear in c. For example, if f is the total amount of carbon emission and C is a vector of the amounts of carbon emission of different technologies, then $f(c, z)$ is linear in c. Bertsimas and Kallus (2020) pointed out that the traditional "predict, then optimize" scheme is reasonable in this case. However, if f includes the company reputation loss, which is positive when the total carbon emission exceeds a threshold and 0 otherwise, and C is a vector of the amounts of carbon emission of different technologies, then $f(c, \quad z)$ is nonlinear in c. In this case, Eq. (2) might be flawed. To remedy the above deficiency, Bertsimas and Kallus (2020) proposed a "local" prescriptive analytics method which first predicts the conditional distribution of C in a "local" manner, i.e., by using a few of the data in S_n that are similar to the sample of interest, and then input the predicted conditional distribution into the original stochastic model.

To extend the local method to a global one, Yan and Wang (2022) make an initial attempt to propose a "global" method based on quantile regression to estimate the conditional distribution of the unknown parameters in the optimization model. The authors consider two cases, one with an unknown parameter, and the other with multiple unknown parameters. When there is only one unknown parameter, i.e., the output of the machine learning model is of one-dimensional, the loss function is changed from the standard MSE to the one considering quantile $\alpha \in (0, 1)$. Then, the machine learning model is turned to predict the 100αth percentile of the target. By changing the value of α from 0 to 1, several scenarios of the predicted target can be obtained, and the conditional distribution of the target can thus be approximated. Then, the optimal decision can be obtained by finding the decision that leads to the least average cost over all the scenarios. If there are multiple unknown parameters, the values of all the dimensions can be predicted in one run by a multi-target regression model whose loss function considers quantile $\alpha \in (0, 1)$ or by several machine learning models whose loss functions consider quantile $\alpha \in (0, 1)$. Then, the sample average approximation (SAA) approach can be used to obtain the optimal decision in the second stage.

The basic idea of the above method based on quantile regression generally complies with that of bagging ensemble method: using different values of α in the loss function for machine learning model training is equivalent to drawing several samples from the whole dataset. The aim of the proposed method in this study and the bagging ensemble method is to take advantage of various base models to form a powerful ensemble so as to achieve more accurate prediction. The main difference from bagging and the proposed method in this study is that bagging is based on bootstrap sampling, i.e., conducting *random* sampling with replacement from the whole training set, while the samples are not drawn randomly in the first approach; instead, they are drawn by changing the value of α while considering the expected accuracy and the computation power intentionally.

In this study, another approach that predicts the conditional distribution of C in a "global" manner based on the errors in the training set is proposed.

3. A global prescriptive analytics method

We propose a global prescriptive analysis method in this section, which is based on the prediction error in the training set. Similar to Yan and Wang (2022), two cases are considered, where the first case has only one unknown parameter, and the second case has multiple unknown parameters.

3.1. One unknown parameter

We first examine the case when there is only one unknown parameter, that is, C is a scalar. A hypothesis, which is the best model mapping features to the target aiming to minimize the prediction error found by model training within the hypothesis space, i.e., $h^*: x \rightarrow y$ and $h^* \in \arg\min_{h \in H} \frac{1}{n}\sum_{i=1}^{n} \left[h(x^i) - c^i\right]^2$, is found in most traditional machine learning models. The predicted value given by h^* is the sample mean, ignoring the uncertainties in the unknown parameter.

To capture the uncertainties in the unknown parameter, we take the prediction error in the training set into account. The optimal decision function h^* is first calibrated on the training set, and it can be any machine learning model for regression tasks such as decision tree, random forest, support vector regressor, artificial neural network, and boosting models. For each record in the training set $S_n = \{(x^1, c^1), \ldots, (x^n, c^n)\}$, there is prediction loss $e_i = c^i - h^*(x^i), \quad i = 1, \ldots, n$. The prediction losses e_i can be treated as realizations of a random variable which takes a total of n values with equal possibility. Therefore, given the auxiliary data x^{n+1} and its predicted output $h^*(x^{n+1})$, the distribution of C takes a total of n possible values and can be approximated by $Pr(C = h^*(x^{n+1}) + e_i) = 1/n, \ i = 1, \ldots, n$. Then, the following stochastic programming model can be used to prescribe a decision.

$$z_{err}^{Global} \in \arg\min_{z \in Z} \frac{1}{n} \sum_{i=1}^{n} f(h^*(x^{n+1}) + e_i, \quad z) \tag{4}$$

We argue that this method is in a "global" manner as it uses all the samples in the training set to estimate the conditional distribution of C.

3.2. Multiple unknown parameters

If C is multi-dimensional (i.e., a multivariate random variable) whose dimension is denoted by d, $d \geqslant 2$, there are two possible approaches depending on how the machine learning model predicts the values in different dimensions.

First, the values of these d dimensions can be predicted in one run by one prediction model, which is the so-called multi-target regression. Then, different scenarios of C can be obtained by considering the prediction error in the training set. That is, the predicted output $h^*(x^{n+1})$ and the prediction loss e_i are vectors. Second, the values of C in each dimension can be predicted separately. Then, each of the d dimensions can also be first predicted by h^* and then be revised by the prediction error in the training set, and the total number of scenarios for C would be n^d.The problem of curse of dimension occurs for large d. We can thus use the SAA approach, where only a small subset of the 10^d scenarios is randomly selected. The total number of selected scenarios is denoted by U and the value of C in scenario $u = 1, \ldots, U$ is c^u; c^u is a vector. The stochastic optimization model in the second stage can be approximated as.

$$z_d^{\text{Global,SAA}} \in \arg\min_{z \in Z} \frac{1}{U} \sum_{u=1}^{U} f(c^u, \quad z) \tag{5}$$

Furthermore, several runs of SAA can be conducted, and the decision leading to the minimum objective function (over a much larger subset of the 10^d or n^d scenarios than $\{1, \ldots, U\}$) in all runs is selected as the optimal decision. We conjecture that if different dimensions of C are almost independent, the second approach outperforms the first one; if different dimensions of C are strongly positively correlated, the first approach is preferable.

4. Conclusion

A global predictive prescription method for the "predict, then optimize" framework is proposed in this study to prescribe a good decision from a good machine learning prediction model. Specifically, predictions given by the machine learning model are revised by considering the prediction error in the training set consisting of auxiliary data. Different from the predictive prescription proposed by Bertsimas and Kallus (2020) which is in a local manner, the proposed method in this study is in a global manner as all the auxiliary data are used to calibrate the machine learning model. It is also significantly different from the global extension of Bertsimas and Kallus (2020) proposed by Yan and Wang (2022) based on quantile regression, as the proposed model uses prediction errors to generate different scenarios of the target. The proposed model is actually a general one which can not only be applied to problems in low-carbon logistics but also to those in operations research and management science where optimal decisions need to be prescribed by using auxiliary data to predict the values of unknown parameters. In addition, the model can also be adjusted by e.g., attaching different weights to the samples in the training set to better capture tailored needs when addressing practical problems.

Two directions are worth exploring in the future. First, the theoretical properties of the machine learning models with the predicted output revised by the prediction error in the training set should be explored. Second, computational experiments should be conducted while using the performance of the following decision problem (which can be the one that minimizes the total cost or maximizes the total benefit) as the metric to help us understand the performance of our global method and the local method proposed by Bertsimas and Kallus (2020) over real-world problems.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bertsimas, D., Kallus, N., 2020. From predictive to prescriptive analytics. Management Science 66 (3), 1025–1044.

Shapiro, A., Dentcheva, D., Ruszczynski, A., 2021. Lectures on Stochastic Programming: Modeling and Theory. Society for Industrial and Applied Mathematics, Philadelphia, the USA.

Yan R., Wang, S., 2022. "Predict, then optimize" with quantile regression: A global method from predictive to prescriptive analytics and applications to multimodal transportation. Multimodal Transportation, accepted and in press.