Multimodal Transportation 1 (2022) 100035

Contents lists available at ScienceDirect

Multimodal Transportation

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

Editorial

"Predict, then optimize" with quantile regression: A global method from predictive to prescriptive analytics and applications to multimodal transportation

1. Introduction

In real-world decision making problems, it is necessary in many cases to develop a holistic framework combining prediction with optimization where the prediction model aims to predict the unknown parameters using the available auxiliary data, and the predicted values of these parameters are input to the optimization model to generate decisions. For instance, we may first predict the travel demand and then optimize the traffic signal at intersection, or we first predict the travel time and then optimize the route choice for vehicles. This is because both prediction and optimization tasks are difficult in themselves (Poornima and Pushpalatha, 2020); consequently, they are usually treated sequentially in existing studies (Yan et al., 2021; Yan and Wang, 2022). However, such sequential method can be flawed, as it ignores the connection between the prediction and optimization problems and the uncertainties in the unknown parameters. In existing literature, two approaches are popular to integrate prediction with optimization more effectively, which we will introduce in the next two sections.

One popular approach to integrating prediction with optimization is called the smart "predict, then optimize" (SPO) framework (Elmachtoub et al., 2020), which is suitable to be applied to optimization models with a linear objective function. The core idea is that the prediction model in the first stage directly leverages the structure and property of the optimization model (including the objective function and the constraints) in the second stage. This is mainly achieved by developing proper SPO loss which takes the decision error induced by the predictions into account in developing the prediction model. Since the concept and framework of SPO was proposed, it has been receiving increasing attention, which can also be found in Balghiti et al. (2021) and Kallus and Mao (2020).

When the objective function is nonlinear in the unknown parameters, uncertainties in these parameters can have a very large influence on the decisions generated. Therefore, another popular approach aims to capture the uncertainty presented in the optimization model by leveraging auxiliary data. This stream of problem is referred to as predictive prescription by Bertsimas and Kallus (2020). The authors further proposed a framework that derives weight functions to model uncertain costs from data. The weight functions are motivated by local-learning predictive methodologies to estimate the conditional distribution of unknown parameters Bertsimas and Kallus (2020). In this way, the parameters are modeled as a random variable and the decision problem of interest as a stochastic optimization problem with imperfect observations to estimate the conditional distribution of parameters. Unlike common approaches in a deterministic manner, it improves the decision quality by overcoming the problem of not considering the auxiliary data in operations research and management science problems in transportation Bertsimas and Kallus (2020). Motivated by this approach, this study proposes a global method based on quantile regression to estimate the conditional distribution of the unknown parameters in the optimization model by leveraging auxiliary data.

2. Problem description

Consider an optimization model below that aims to prescribe an optimal decision z^*:

$$z^* \in \underset{z \in Z}{\operatorname{argmin}} \mathrm{E}[c(z; Y)] \tag{1}$$

where z denotes a decision, Z denotes the set of all feasible decisions, Y denotes a random parameter, c is a cost function associated with the parameter and the decision, and E calculates the expected value with respect to the probability distribution of Y. Note

https://doi.org/10.1016/j.multra.2022.100035
Received 7 January 2022; Received in revised form 23 March 2022; Accepted 4 April 2022

that here c is nonlinear in Y, otherwise the problem can be converted into $z^* \in \underset{z\in Z}{\operatorname{argmin}}\, c(z; \mathrm{E}[Y])$. Denote the cumulative distribution function (CDF) of random variable Y by $F_Y(y)$, which is unknown. Consequently, we cannot evaluate $\mathrm{E}[c(z;Y)]$ for a particular $z \in Z$ of interest and thus we cannot solve the model directly. However, we have the associated auxiliary data and the corresponding output, which are historical records related to the unknown parameter denoted by set $S_N = \{(x^1, y^1), ..., (x^N, y^N)\}$, where x^n is the auxiliary data and y^n is its output which is also the historical data on the uncertain quantities of interest Y, $n = 1, ..., N$. We also have the auxiliary data, denoted by x^{N+1}, to predict the conditional distribution of Y for our decision problem. Problems in this form can be quite common in multimodal transportation. For example, a retailer needs to decide the purchase capacity on day $N+1$ based on its conditional distribution to minimize the loss caused by the unsold quantities. He has the historical data from days 1 to N, which is $S_N = \{(x^1, y^1), ..., (x^N, y^N)\}$, where y^n is the demand on day n and x^n is the features of that day (e.g., whether there are major national and international events, whether there are disasters). Z is the set of candidate quantities to purchase, and z is a decision of the purchase quantity. Then, based on S_N, the conditional distribution of Y can be predicted using the auxiliary data x^{N+1}.

A traditional "predict, then optimize" scheme works as follows. Using dataset S_N, one can build a machine learning model $g : x \to y$ to predict the "value" of the parameter based on the auxiliary data. If the mean squared error (MSE) is used as the loss function to train the machine learning model, the "value" to be predicted is the conditional mean of Y given the auxiliary data. Then, we assume that $\Pr(Y = g(x^{N+1})) = 1$ and solve the resulting deterministic model to prescribe a decision

$$z^{Deterministic} \in \underset{z\in Z}{\operatorname{argmin}}\, c(z; g(x^{N+1})). \tag{2}$$

Bertsimas and Kallus (2020) pointed out that the traditional "predict, then optimize" scheme is reasonable when, for a particular realization of Y denoted by y, $c(z; y)$ is linear in y, but is flawed when $c(z; y)$ is nonlinear in y. For example, suppose $c(z; y) = \max(y - z, 2z - y)$, $Z = [-1, 1]$, and the conditional distribution of Y given x^{N+1} is $\Pr(Y = -1) = 0.5$ and $\Pr(Y = 1) = 0.5$. Suppose that $g(x^{N+1}) = 0$, i.e., the machine learning model g has successfully predicted the conditional mean. Then, the traditional "predict, then optimize" scheme will yield a solution of $z = 0$ (with an expected cost of 1). However, the optimal solution is $z = -0.5$ (with an expected cost of 0.75).

To remedy the above deficiency, Bertsimas and Kallus (2020) proposed a prescriptive analytics method. This method essentially predicts the conditional distribution of Y in a "local" manner and then inputs the predicted conditional distribution into the original stochastic model. Take the k-nearest neighbors (kNN) method as an example of the machine learning model. Denote the set of the k nearest neighbors of x^{N+1} in dataset S_N as $N_k(x)$ with an index set J, $|J| = k$. Then, instead of using $\sum_{j\in J} y^j/k$ as a point estimation of Y, Bertsimas and Kallus (2020) assumed that the conditional distribution of Y could be approximated by $\Pr(Y = y^j) = 1/k$, $j \in J$, and then used the following model to prescribe a decision

$$z^{Local} \in \underset{z\in Z}{\operatorname{argmin}}\, \frac{1}{k} \sum_{j\in J} c(z; y^j). \tag{3}$$

We argue that the conditional distribution of Y given a particular x^{N+1} predicted by Bertsimas and Kallus (2020) is in a "local" manner because they only used a few data in S_N that are similar to the sample of interest (i.e., $N_k(x)$). (Note that in addition to kNN, the authors also used other machine learning models.) Alternatively, our study proposes an approach that predicts the conditional distribution of Y in a "global" manner based on the idea of quantile regression, elaborated in the next section.

3. A global prescriptive analytics method

3.1. One unknown parameter

We first examine the case when there is only one unknown parameter, that is, Y is one-dimensional. Most of the traditional machine learning models (e.g., multiple linear regression and artificial neural networks) aim to find a function, denoted by $h^* : x \to y$, out of a set of decision functions, denoted by H, that minimizes the loss function, which is usually the MSE, over the historical dataset S_N, i.e., $h^* \in \underset{h\in H}{\operatorname{argmin}}\, \frac{1}{n} \sum_{i=1}^{n} [h(x^i) - y^i]^2$. The predicted value under this circumstance is the conditional sample mean which is deterministic. To capture the uncertainties in the unknown parameter, the global method proposed in our study uses the idea from quantile regression (Koenker and Hallock, 2001; Koenker, 2017) to generate conditional distribution. Given a quantile $\alpha \in (0, 1)$, the machine learning model that predicts the 100αth percentile of Y, which is denoted by h^α, can be estimated by simply changing the loss function to the following:

$$h^\alpha \in \underset{h\in H}{\operatorname{argmin}} \sum\nolimits_{i=1}^{n} \left\{(1-\alpha) \max\left\{h(x^i) - y^i, 0\right\} + \alpha \max\left\{y^i - h(x^i), 0\right\}\right\}. \tag{4}$$

By changing α between 0 and 1, e.g., by setting $\alpha = 0.05, 0.15, ..., 0.95$, the conditional distribution of Y given the auxiliary data x^{N+1} can be approximated by $\Pr(Y = h^\alpha(x^{N+1})) = 1/10$, $\alpha = 0.05, 0.15, ..., 0.95$. Then, the following model can be used to prescribe a decision

$$z^{Global} \in \underset{z\in Z}{\operatorname{argmin}}\, \frac{1}{10} \sum_{\alpha=0.05,0.15,...,0.95} c(z; h^\alpha(x^{N+1})). \tag{5}$$

Note that in our method, when estimating the conditional distribution of Y, all data in S_N are used as the training set. Hence, our method is "global". The interval or the selected values of α can largely influence the model performance. We propose that the interval should be smaller, or equivalently the number of selected values should be larger when dealing with more sophisticated machine learning models or when there are larger dataset.

3.2. Multiple unknown parameters

If Y is multi-dimensional (i.e., a multivariate random variable) whose dimension is denoted by d, $d \geq 2$, there are two possible outcomes 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, we can use a loss function similar to the one in Eq. (4) aiming to minimize the sum of MSE over all dimensions. Different prediction functions h^{α} (note that now h^{α} is a vector of functions) will then be obtained and used to predict different scenarios of Y in the optimization model. It should be noted that $h^{\alpha}(x^{N+1})$ should not be interpreted as the 100αth percentile of Y (because Y is multi-dimensional) but should only be understood as a scenario of Y.

Second, the values of Y in each dimension can be predicted separately. Then, we can predict the 100αth percentile of each of the d dimensions, where $\alpha = 0.05, 0.15, ..., 0.95$, and the total number of scenarios would achieve 10^d. Consequently, the problem of curse of dimension occurs for large d. We can thus use the sample average approximation (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 Y in scenario $u = 1, ..., U$ is $\hat{y}^u$), and the optimization decision in the second stage can be approximated as

$$z_d^{\text{Global,SAA}} \in \underset{z \in Z}{\operatorname{argmin}} \frac{1}{U} \sum_{u=1}^{U} c(z; \hat{y}^u). \tag{6}$$

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 scenarios than $\{1, ..., U\}$) in all runs is selected as the optimal decision.

4. Conclusion

A global predictive prescription method for data-driven optimization is proposed in this study aiming to go from a good prediction to a good decision. The method combines a machine learning model and an optimization model. The machine learning model uses the loss function in quantile regression to predict the distribution of the unknown parameters in the following optimization model based on auxiliary data in a global manner. The method is useful when the objective function in nonlinear in the parameter in the optimization model. It can be applied to operations research and management science problems in multimodal transportation with unknown parameters to prescribe optimal decisions. In order to compare the local and global methods used for prescriptive analytics, in future research, extensive numerical experiments using open data should be conducted, and conclusions and insights on their performance as well as application scenarios should also be provided.

Declaration of Competing Interests

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.

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

Ran Yan*
Research assistant professor, Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

*Corresponding author.
E-mail addresses: wangshuaian@gmail.com (S. Wang), angel-ran.yan@connect.polyu.hk (R. Yan)
Revised 23 March 2022

References

Balghiti, O.E., Elmachtoub, A.N., Grigas, P., Tewari, A., 2021. Generalization bounds in the predict-then-optimize framework. arXiv preprint arXiv:1905.11488.
Bertsimas, D., Kallus, N., 2020. From predictive to prescriptive analytics. Manage. Sci. 66 (3), 1025–1044.
Elmachtoub, A., Liang, J.C.N., McNellis, R., 2020. Decision trees for decision-making under the predict-then-optimize framework. Int. Conference Machine Learn. 2020, 2858–2867.
Kallus, N., Mao, X., 2020. Stochastic optimization forests. arXiv preprint arXiv:2008.07473.
Koenker, R., 2017. Quantile regression: 40 years on. Ann. Rev. Econ. 9, 155–176.
Koenker, R., Hallock, K.F., 2001. Quantile regression. J. Econ. Perspect. 15 (4), 143–156.
Poornima, S., Pushpalatha, M., 2020. A survey on various applications of prescriptive analytics. Int. J. Intell. Networks 1, 76–84.
Yan, R., Wang S., 2022. Integrating prediction with optimization: Models and applications in transportation management. Submitted to Multimodal Transportation, accepted and in press.
Yan, R., Wang, S., Zhen, L., Laporte, G., 2021. Emerging approaches applied to maritime transport research: Past and future. Commun. Transport. Res. 1, 100011.